Year : 2023

How social media data can complement opinion surveys for the study of public opinion – A social science perspective

Reveilhac Maud

UNIL | Université de Lausanne

FACULTÉ DES SCIENCES SOCIALES ET POLITIQUES

INSTITUT DES SCIENCES SOCIALES

How social media data can complement opinion surveys for the study of public opinion – A social science perspective

THÈSE DE DOCTORAT

Présentée à la :

Faculté des sciences sociales et politiques de l'Université de Lausanne

Pour l'obtention du grade de

Docteur en sciences sociales

par

Maud Reveilhac

Directrice de thèse
Prof. Stephanie Steinmetz

Co-directeur de thèse
Dr. Davide Morselli

Jury

Prof. Nicky Le Feuvre (Présidente)
Prof. André Berchtold (expert interne)
Prof. Gerold Schneider (expert externe)
Prof. Hajo Boomgaarden (expert externe)

Lausanne
2023

UNIL | Université de Lausanne

FACULTÉ DES SCIENCES SOCIALES ET POLITIQUES

INSTITUT DES SCIENCES SOCIALES

How social media data can complement opinion surveys for the study of public opinion – A social science perspective

THÈSE DE DOCTORAT

Présentée à la :

Faculté des sciences sociales et politiques de l'Université de Lausanne

Pour l'obtention du grade de

Docteur en sciences sociales

par

Maud Reveilhac

Directrice de thèse
Prof. Stephanie Steinmetz

Co-directeur de thèse
Dr. Davide Morselli

Jury

Prof. Nicky Le Feuvre (Présidente)
Prof. André Berchtold (expert interne)
Prof. Gerold Schneider (expert externe)
Prof. Hajo Boomgaarden (expert externe)

Lausanne
2023

UNIL | Université de Lausanne
Faculté des sciences
sociales et politiques

IMPRIMATUR

Le Décanat de la Faculté des sciences sociales et politiques de l'Université de Lausanne, au nom du Conseil et sur proposition d'un jury formé des professeurs

- Stéphanie STEINMETZ, Directrice de thèse, Professeure à l'Université de Lausanne
- Davide MORSELLI, Co-Directeur de thèse, Professeur à l'Université de Lausanne
- André BERCHTOLD, Professeur à l'Université de Lausanne
- Hajo BOOMGAARDEN, Professeur à l'Université de Vienne, Autriche
- Gerold SCHNEIDER, Professeur à l'Université de Zurich

autorise, sans se prononcer sur les opinions de la candidate, l'impression de la thèse de Madame Maud REVEILHAC, intitulée :

« How social media data can complement opinion surveys for the study of public opinion – A social science perspective »

Nicky LE FEUVRE
Doyenne

Lausanne, le 23 janvier 2023

ABSTRACT

The motivation for this thesis is to contribute to the scientific debate about whether opinion surveys will become obsolete to measure opinion and attitudinal trends and whether social media data can contribute to improving the scientific understanding of public opinion. While research into the potential value of social media as a survey supplement is ongoing, there are still questions about the reliability of these online data for analysing public opinion. Assessing potential overlaps between social media and survey results is frequently the major goal of current investigations. Theoretically, this thesis takes a different approach by outlining how both data sources can best support (rather than replace) one another. Methodologically, it investigates ways to measure opinions on social media so that they can best complement survey findings. Empirically, it offers empirical studies on a variety of societal and political themes where the complementary nature of the two data sources is crucial to the design of the study and the interpretation of the results. The thesis outlines the benefits and drawbacks of each data source for the study of public opinion and suggests directions for future investigation.

RESUME

Cette thèse participe au débat scientifique sur l’obsolescence des sondages d'opinion pour la mesure des opinions et attitudes, et sur la contribution des données des médias sociaux à la compréhension de l'opinion publique. Certes, la plus-value des médias sociaux en tant que complément possible aux enquêtes d’opinion est d’ores et déjà étudiée. Cependant, ce domaine est encore en développement et des doutes persistent quant à la validité des données en ligne pour étudier l'opinion publique. Les études existantes se concentrent en particulier sur la correspondance entre les tendances mesurées par les médias sociaux et celles issues des enquêtes d’opinion. Théoriquement, cette thèse adopte une approche différente proposant de complémentariser (et non remplacer) ces deux sources de données. Sur le plan méthodologique, elle étudie les moyens de mesurer les opinions exprimées sur les médias sociaux afin de compléter aux mieux les résultats des enquêtes. Elle propose également des études empiriques sur une variété de thèmes sociétaux et politiques où la complémentarité des deux sources de données est au cœur du dessin de recherche et de l'interprétation des résultats. La thèse aborde les avantages et inconvénients de chaque source de données relative à l'étude de l'opinion publique et propose des pistes de recherche futures.

ACKNOWLEDGMENTS

I certify that the work presented in this thesis is that of the author and the mentioned co-authors. Furthermore, the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program.

I would like to express my deepest gratitude to my co-directors, Stephanie Steinmetz and Davide Morselli, for their invaluable feedbacks and support. I also could not have undertaken this journey without my thesis committee, who generously provided knowledge and expertise. Major thanks should also go to my co-authors who impacted and inspired me.

Additionally, I am also grateful to the members of the FORS-SSP (Swiss Centre of Expertise in the Social Sciences and the Faculty of Social and Political Sciences) scientific programme, to the members of LINES (Life Course and Inequality Research Centre), and to the members of the GREC (Research Group on Elections and Political Citizenship) for their feedback sessions and research support. I also would like to thank Michael Amsler from Zurich University for his very valuable advice, especially with respect to word embeddings.

FOREWORD

This is an article-based dissertation. All of the work presented henceforth has been published in peer reviewed journals or peer reviewed book chapters. For all of the theoretical, methodological and empirical articles, I was the lead investigator responsible for all major areas of concept formation, data collection, method development and implementation, and analysis, as well as manuscript composition. There is, however, one exception, with the book chapter entitled "Assessing how Attitudes to Migration in Social Media Complement Public Attitudes Found in Opinion Surveys" for which I share perfect co-authorship with Prof. Gerold Schneider.

Stephanie Steinmetz and Davide Morselli were respectively the supervisor and co-supervisor of this dissertation project and provided valuable feedback throughout various manuscript edits and presentations that happened before their publication.

The reader should note that each of the articles introduced in this dissertation is distinct and intended to "stand alone" as published work. Although each article is self-contained, each of them addresses questions of theoretical and methodological importance within the common issue of this thesis.

Furthermore, my research project was also supported institutionally. During my time as a PhD student, I was allowed to be part of a team of researchers from the platform FORS-SSP which brings together researchers from the Faculty of Social and Political Sciences and collaborators from the Swiss Center of Competence in Social Sciences at the University of Lausanne.

TABLE OF CONTENT

CHAPTER 1. INTRODUCTION

1.1. Background and motivations

Social media, such as Twitter and Facebook, have become an essential source of data for taking the pulse of publicly expressed opinions, especially for journalists (Anstead & O'Loughlin, 2015; McGregor, 2019). Over the last decades, it served as a dominant force for the creation and spread of public opinion (Burgess & Bruns, 2013; van Dijck & Poell, 2013; Diaz, 2016). For instance, research has shown that twenty percent of Americans changed their opinion on a candidate, or an issue based on content they saw on social media (Pew Research Center, 2016). In line with the rapid expansion of available digital data and the capabilities of ready-to-use off-the-shelf software solutions, social media data have also become increasingly popular for conducting social science research. Researchers interested in public opinion have thus turned increasingly to new modes of data collection associated with this new data source, enabling them to investigate attitudes and behaviours among the population and sub-groups (Klašnja et al., 2016). Looking at recent developments this is supported by the continued increase of computational social science research centres, digital data research programs, as well as journals and conferences dedicated to computational social science research.

In the meantime, opinion surveys remain the main method of collecting opinions from random samples of the population in order to generate statistical inferences and derive public opinion, which can then be conceived as the expression of a collective opinion, such as a protest or a vote. However, survey instruments are losing credibility, in particular due to falling response rates (Groves, 2011; Brick & Williams, 2012; Czajka & Beyler, 2016). Even if weighting techniques exist to correct for possible biases in the representativeness of the surveyed population, there is a valid concern about whether the responses of surveyed citizens (or those who agree to take part in surveys) can truly reflect important societal trends. Furthermore, surveys are potentially poor at recruiting hard-to-reach demographics and do not always coincide with the emergence of important societal questions (Japec et al., 2015). Moreover, even given the long history of opinion survey research, there is still no definitive agreement on how to best measure public opinion from surveys as there is no straightforward strategies to decide which person to interview and what questions to ask them (Berinsky, 2017). However, despite high cost affecting the traditional surveys, where probability sampling is a prominent feature, the use of alternative (and sometimes less costly) data sources, most notably social media data, raises unresolved questions about the quality and validity of statistics produced in such ways.

From the 2010s, groups of social scientists were enthusiastic about the potential of large digital databases to address long-standing and fundamental issues of society with the help of sophisticated data analysis (King, 2011). This included the harvesting of million social media posts and the use of new natural language methods to extract relevant information. However, not all social scientists have embraced the shift of thinking about social media data as an interesting and useful source for studying public opinion. Until today, concerns are held about the validity of incomplete and messy social media data to draw conclusions of societal importance. Other concerns relate to the fact that these data are often used for descriptive research, but rarely for looking at causes of the observed patterns, while there is also a concern that computational (social) scientists are not sufficiently referring to social and political theories to motivate their approach and to interpret their findings (Ledford, 2020).

Despite these concerns, both research fields are increasingly merging, notably in shared study programs, conferences and publications. Analyses of (social) media texts are experiencing an unprecedent boom across the social sciences (Van Atteveldt et al., 2019). In parallel, computational methods are consistently evolving, and social media sites are also likely to change. Engaging in interdisciplinary research is critical to conveying complementary expertise for understanding when and how to complement surveys and social media data. For instance, social researchers have a longer-standing tradition of survey methodology to investigate societal questions, especially regarding questions of concept validity and robustness of findings. Furthermore, computational (social) scientists and computational linguists are much quicker to adopt and adapt new text analysis methods as they come along. These collaborations can encourage common agendas. For example, similar to the trend towards increasing internationally comparative survey research (Hanitzsch et al., 2019), (social) media analysis is increasingly conducted in a comparative fashion (Lind et al., 2019).

Against this background, the motivation for this thesis is to contribute to the scientific debate about whether opinion surveys become obsolete to measure opinion and attitudinal trends, and whether social media data can contribute to improve the scientific understanding of public opinion. Certainly, investigation into social media utility as a possible complement to survey is ongoing, but the field is still developing and doubts persist in the academic community about the validity of these online data for studying public opinion (Gayo-Avello, 2013; Jungherr, 2016). The difficulty in complementing both data sources remains an important challenge, most notably with respect to the assessment of what types of methods and data features are most suitable for any specific task (Schober et al., 2020). This thesis encourages the idea that the alignment of social media findings with social scientific knowledge should be included as a core evaluation criterion.

Scholars have looked at the reliability of social media data to assess opinions compared to classical surveys, especially in the field of election forecasts (O'Connor et al., 2010). The main aim of these studies is often to assess potential points of alignment between social media and survey findings, notably with a view to replacing surveys with social media data. This work is complicated by the fact that it is not easy to convert social media data into survey-like variables without losing a great deal of information, and because social media data generally lack socio-demographic information. The proposed thesis takes a different approach by providing insights about how both data sources can best complement (and not replace) each other. To do so, it consists of one theoretical, one methodological, and two empirical chapters, all of which can be read independently as they address specific research challenges.

1.2. Synthesis of each research challenge

Each chapter of this thesis thus addresses a specific challenge and makes an independent contribution to the study of public opinion. The outline of the thesis is based on a collection of theoretical, methodological, and empirical studies, which aim to address the following research questions:

Chapter 2 discusses how and whether social media data can complement traditional survey data to study public opinion. The research questions it answers reads as follows: How are social media data used for the study of public opinion? It proposes a theoretical framework for assessing the potential of social media data to strengthen the findings from existing survey research. The abundance, affordance and versatility of social media data creates opportunities for complementing traditional survey research. The decline in survey response rates makes it more complex and costly to study public opinion. In this respect, social media offer more flexibility. However, this advantage is sometimes outweighed by the efforts required to extract relevant and valid data for measuring opinions and attitudes, thus limiting the reliability of studying public opinion in this way. To date, we still lack a comprehensive framework informing researchers about the pros and cons of different approaches to complementing survey data and social media data for the study of public opinion. Therefore, chapter 5 provides guidelines for using social media data in the study of public opinion in relation to four main research objectives, namely the improvement of validity, sustainability, reliability and interpretability.

Chapter 3 provides two methodological works adopting a social science perspective on what reliable methods for extracting opinions from social media data are. It thus aims to answer the following research question: What are reliable methods for extracting opinions from social media data? The starting point of these methodological works is the fact that there is still little knowledge about best practice in constructing valid social media-based estimates of opinions (Schober et al., 2020). Measuring publicly available opinions validly requires data of very high quality. However, social media data are often messy (e.g., unusual writing conventions, improper spelling, ambiguous messages, etc.). The textual and spontaneous nature of social media texts – as opposed to the calibrated and well-designed survey questions – suggests the need for reflection on what might be best practice in adapting text analysis methods to answer social science research questions. Furthermore, there is little knowledge about which tools are better suited for specific measurement tasks and what pre-processing stages or tuning parameters are most reliable for the detection of specific textual properties and concepts (van Atteveldt et al., 2021; Baden et al., 2021). The first methodological work provides a case study for helping researchers navigate decisions when producing measures of tonality and frames from a small sample of annotated social media posts that can be compared to existing survey measures. In particular, the validity of several methods for extracting opinions from text so as to usefully complement survey findings is assessed. The second methodological work focuses on the detection of stance, which informs us about users' opinion about a given target (e.g., a policy issue or a person) or, in other words, whether users are opposing or supporting a given target. Better measuring stance from texts constitutes an indispensable task for being able to complement survey data with social media data. Indeed, while some measurements (e.g., the importance of a target and satisfaction with a target) can be made directly comparable through the use of social media proxies[1] (e.g., prevalence and tonality), this won't fit with the majority of survey questions that are interested in measuring respondents' stance towards a target. Stance detection is still a recent field of study in the realm of social media research with still few comparative works (Ng & Carley, 2022).

Chapter 4 focuses on what are typical and influential social media users and their interaction with public opinion. It addresses the following research question: What public are available on social media and how do they interact with public opinion? In contrast to data scientists, who have been quick to attempt to leverage social media data to generate novel understandings of society, social researchers have long worried that social media data describe nonprobability samples. An overarching concern has been the need to consider the demographic and ideological attributes of social media users (Barberá & Rivero, 2015; Mellon & Prosser, 2017). To this end, data and social scientists have developed methods to infer these characteristics (e.g., Mislove et al., 2011; Mancosu & Bobba, 2019), sometimes with a view to constructing weights that allow the collection of posts to reflect public opinion. Nevertheless, these corrections are sometimes difficult to apply given the lack of necessary sociodemographic information. However, any analysis of social media data seeking to draw valid conclusions about opinion, attitudes or behaviours should assess which groups of users are represented in the corpus of collected social media messages. The focus on social media groups improves our understanding of whether and how social media measurement complements more traditional survey data. Albeit social media users do not generally represent public opinion, they have the potential to lead and influence public debates largely and, thereby, to influence public opinion (Barberá & Steinert-Threlkeld, 2020). For instance, the COVID–19 pandemic has been accompanied by intensified public discourse on social media, including discussion about which policies and strategies are adequate for fighting the crisis (Gilardi et al., 2021). Furthermore, online discussions are often dominated by a minority of active users. Previous studies have shown that social media, especially Twitter, tends to over-represent the opinions of user groups that are very active online (e.g., political leaders, journalists, social movements and their followers, and influencers), which do not usually constitute the target samples of surveys. Moreover, political discourse is increasingly taking place in the online realm (Maia, 2017) and the effects of political events, such as

[1] Proxies are variables that are not in themselves directly relevant, but that serve in place of an unobservable or immeasurable variable.

social movements, on online discourse are growingly important (Segerberg, 2017). Consequently, the groups of users actively contributing to online discussions have concrete implications on the interpretability of findings derived from social media data. Chapter 4 proposes three empirical works. The first work assesses the characteristics of the general public that uses social media and what are the challenges in terms of media consumption practices. The second assesses who are influential actors on social media by focusing on the public with whom politicians interact and whether these interactions relate to their political success. The third work presents which groups of users are particularly involved in social media communication during a particular social movement. It also assesses how the claims of the movements that are voiced on social media mirror prevailing political views and public opinion on the subject.

Chapter 5 provides empirical examples of emblematic approaches in which social media research can complement survey research. It therefore aims to address the following research question: How can social media be used to provide a new lens through which to observe both well-established and under-investigated topics in social and political sciences by complementing survey data? Firstly, social media can be used to identify important dimensions of topics that are "under-investigated" by survey research, while proposing dimensions worth investigating in future surveys. Secondly, in combination with survey data social media can be used for addressing "old" research interests through a "new" lens. For instance, social media has created new avenues to examine well-studied phenomena, such as political agenda setting (Gilardi et al., 2021), issue congruence and responsiveness between politicians and the(ir) public (Stier et al., 2018; Ennser-Jedenastik et al., 2022), political polarisation on cultural and policy issues such as immigration or European questions (Gorodnichenko et al., 2017; Heidenreich et al., 2019). Based on our systematic literature review (chapter 2), the approach using social media to generate new insights can be described on a continuum of research designs going from an exploratory to an explanatory framework (see Figure 1.1). To improve the interpretability of the findings, the choice of the research design depends on the availability of survey data covering the topic of interest. Therefore, we propose three empirical works. The first work follows an exploratory design on a topic for which only few (or scarcely related) survey data are available. Thus, social media represents the main source of data for identifying relevant dimensions and opinions about the topic under investigation. The identification of user groups and innovations in data visualisation based on word embeddings are needed to render the research interpretable and useful for conducting future research. The second work uses a descriptive design to compare trends related to a common topic in survey data and on social media. This work aims to make a connection between political content from social media and the broader public attitudes. In particular, it shows how the communication of political authorities and experts is received by other users and how it reflects trends in public trust. The third work relies on an explanatory designs and suggests to conduct analyses by augmenting social media data with survey data about a shared topic to discover interrelations between online and offline trends, while also discussing the extent to which the findings are affected by different data collection strategies.

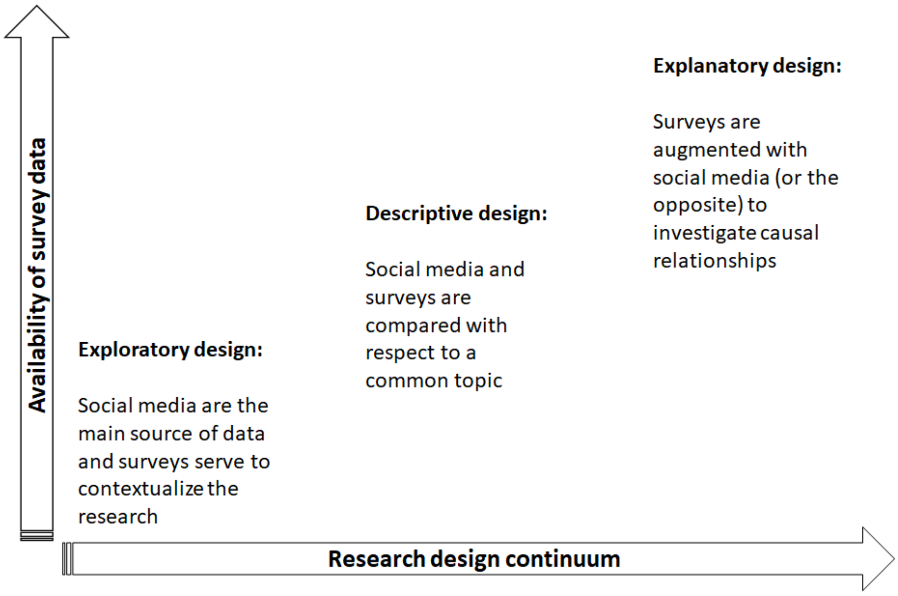


Figure 1.1: *Ways to efficiently use social media data for generating new insights in complement to surveys for the study of public opinion*

1.3. Definition of "public opinion" and implications for social research

The measurement of public opinion is a critical source of validation for social and political theories assessing the well-functioning of democratic systems (Welzien, 1995). However, there is little consensus about how to define "public opinion". From a historical perspective, the question of what exactly constitutes public opinion remains open with regard to how to define it, where to encounter it, and how to measure it. This necessarily produces visions and methodologies which are complementary (and sometimes contradictory) to each other.

Originally, public opinion represented the views of an elite class of people, with the opinion of the general population (still at a very local level) becoming of interest only later. In the 1850s, the objective of studies interested in public opinion revolved mainly around predicting electoral outcomes and assessing the approval (or legitimacy) of policy reforms. Concerning elections, mass media journals organised straw polls with the aim of predicting an overall voting trend by surveying only a sample of the readership. These polls were thus subject to a lack of representativity and could not achieve good generalisability beyond the journal readership. In 1936, Georges Gallup conducted a survey predicting Roosevelt's election and introduced a modern methodology that rested on a carefully chosen sample of only a few thousand people. This method was supported by the scientific debate on inferential statistics and the types of sample selection (e.g., random or quota).

From a societal perspective, opinion surveys were conceived as an indispensable instrument for democracy with concrete implications on the legitimation of the public order (Blondiaux, 1998). Indeed, modern survey techniques offered a direct access to citizens' opinions and views, and it was expected that, by revealing the general will of citizens, the levels of corruption within elected representative bodies and the power of lobbying groups' interests would be reduced.

In parallel to methodological discussions, scientific debates were asking what factors drive public opinion. In this respect, an important theory is certainly the *two-step flow of communication* by Katz and Lazasfeld (1955). According to this theory, individuals are indirectly influenced by the media through opinion leaders in specific issues (e.g., electoral choice). Thus, the people most interested in news and politics will be more receptive to this media content and will spread the media information to groups, friends, and relatives. The political and business consequences of such a theory are consequential, since it implies that is not necessary to reach the whole population, but only opinion leaders who will spread messages through little circles of influence. Today, social media act as an amplifier of such social persuasion by connecting people.

From the 1970s, important criticisms were raised concerningthe measurement of public opinion through opinion surveys (see Converse, 1970; Bourdieu, 1973). For instance, contrary to the belief that public opinion should be perfectly informed, and, thus, readily available, surveys ask questions about which individuals might have no opinion and no interest. Thus, public opinion solicited through surveys tends to reflect researchers' interests rather than the spontaneous expression of public interests. Furthermore, unlike the conception of public opinion as the product of collective processes (e.g., debates and discussions), surveys deduce public opinion by aggregating individual opinions. However, since society is permanently affected by polemics and confrontations, not all opinions possess the same value. These criticisms were seriously considered by survey methodologists and led to important developments in theoretical and empirical tools to improve survey measures and address survey biases. To date, survey researchers very often refer to and rely on the *Total Survey Error* (TSE) framework (Biemer, 2010).

Opinions surveys thus became a scientific and rigorous means of measuring public opinion which could be used by governing bodies for policy making, both as a means of control and as a method of anticipating societal trends and approval. In parallel, citizens could be informed of general trends and learn how to situate themselves with respect to a more general opinion concerning specific policies or decisions. However, there is still to this day no consensus about the definition of public opinion. Nonetheless, public opinion tends to be defined by the way it is measured through opinion surveys and typically relates to aggregated perceptions towards a given theme or policy. These perceptions are shared by (groups of) the population and are related to more general collective representations or understandings of the workings of society in a given contextual and temporal setting.

Another, less methodological, definition of public opinion was given by Key (1961) as "the opinions held by private persons which governments find it prudent to heed" (p.14). In this thesis, we borrow Key's definition as it is particularly useful for proposing practical guidelines about ways in which social media data can be a useful complement to survey data for the study of public opinion. Key's definition highlights that the conception of public opinion is highly political. Indeed, in his definition, public opinion is not merely conceived as collective social representations, but as positions that are shared by (groups of) citizens and that have a concrete political impact.

Key's definition enables us to better understand the role played by the tools to measure opinions on democracy. According to this definition, opinions become important for the workings of democratic systems when these opinions are considered important for decision-making purposes. It thus requires effective means by which publicly available opinions and attitudes can be accurately conveyed to political bodies. From this perspective, both opinion surveys and social media can bring insights about the state and dynamic of opinion debates by highlighting the multiple forms of expression and the competing viewpoints which compose the public debate. In this view, this thesis emphasizes these theoretical contributions on the nature of opinions measured through social media, as well as the empirical works validating and generating measures of opinions through social media texts.

1.4. Definition of "social media" and implications for social research

Turing to the term "social media", it is a collective term for websites and applications that focus on communication, community interaction, and collaborative content sharing. Kaplan and Haenlein (2010, p.61) proposed to define social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content". The term thus groups applications that can be distinguished into more specific categories.

In this thesis, the focus is on social networking sites, which have become an integral part of everyday communications for a large number of citizens, businesses, political actors and organisations. Such sites have deeply modified existing communication logics (van Dijck & Poell, 2013), among individuals (e.g., informal interactions with friends), institutional structure (e.g., activism), business applications (e.g., marketing of products) as well as political actors (e.g., governments engage with the people's concerns). Social networking sites group a variety of applications, which need to be distinguished. In particular, the goals of communication in social media (such as Facebook or Twitter) or networking sites (such as LinkedIn) are very different. While social media is a message-delivering channel to users, social networking involves two-way communication, and hence develops relationships.

Today, social media represent a new step to look at particular opinions and dynamics which could complement survey findings, notably because only specific groups are actively using social media with specific goals and audiences in mind. Opinion surveys thus remain an important standard to benchmark the state of public opinion in the general population, and it would be illusory to equate opinions available on social media as the true expression and reflection of public opinion. Furthermore, as opinions stemming from surveys are calibrated answers to pre-defined questions, there is a tendency to believe that opinions expressed on social media would be more spontaneous.

However, social media settings and conventions also impose structured ways to express opinions (e.g., number of characters and relational dynamics). The weight of external factors in formatting opinions is well-known from survey research, with such factors including the role of the question wording, the availability of answer modalities, and the role of the interviewer on the validity and sincerity of respondents' answers. Similarly, opinions on social media can also be artificially produced, namely through phenomena such as astroturfing and the action of bot armies (García-Orosa, 2021).

Key's definition also provides room for conceiving social media as a valid and complementary source for the study of opinions. For instance, social media can bring to the forefront views from groups that do not usually take part in surveys, notably participants of social movements. For instance, groups such as the *Gilets Jaunes* in France and Covid-19 anti-vaccination groups have largely communicated on social media platforms (particularly on Facebook and Twitter), notably to influence the larger public opinion by building awareness around specific causes. Social media can also complement surveys that assess the legitimacy of political actors (e.g., political candidates or leaders) by emphasising features that inspire credibility. For instance, a political leader's number of followers or mentions on Twitter is often associated with the idea of large public support (see review by Skoric, Liu, and Jaidka (2020)). This idea also applies to opinion surveys for election forecasting where measures of vote intention have an impact on the selection of candidates and on political campaigns (e.g., Rothschild & Wolfers, 2011). There are also cases where trends measured through surveys and social media can differ. For instance, people taking part in large street demonstrations and the silent (surveyed) majority can display different views (e.g., large anti-vaccination opinions on social media are not mirrored in survey trends). Social media trends can also offer a more polarised description of society than the trends measured in general surveys (e.g., online debates about immigration, discrimination, religion, etc). This divergence can certainly be explained by the fact that social media provide opportunities for the expression of extreme (e.g., populism) and negative (e.g., hate speech) voices.

The positive and negative consequences of social media on citizens' information and political expression, as well as on democratic processes, are largely documented (Persily & Tucker, 2020; Vaccari & Augusto, 2021). For instance, positive impacts include social media's potential to give a voice to (oppressed) groups of the population, notably in the form of (support to) social movements. Another positive impact relates to the development of more transparent political processes that mirror citizens' expressed opinions and preferences which are directly and rapidly accessible through social media. This echoes the hopes and expectations that characterized the implementation of modern survey techniques. However, negative impacts include, for instance, the potential for manipulating election outcomes and the polarization of societal debates. Although such processes existed before the rise of social media, the choice of algorithmic logics underpinning network communication and user affinities certainly enable the spread and intensification of these phenomena. It is nevertheless important to highlight that progress has been made by third-party bodies (e.g., the European commission) to regulate and moderate the content and privacy rules prevailing on social media, but also by social media companies themselves to modify their algorithms.

There are other societal trends that are important to consider for social research exploiting social media content to study opinions. Two of them are particularly consequential for research findings. The first relates to the availability of social media content. For instance, social media users are specific groups of the population, mainly it is the youngest people or people who do not read traditional news who are increasingly turning to social media to consume information, but also to actively post and share (non)political content. Political actors (e.g., political candidates and leaders) and other opinion makers (e.g., social movement figures and influencers) also constitute a prominent part of social media users and are responsible for a large share of social media content that is publicly available. The presence of these groups has important consequences on the interpretation of research findings based on social media messages, notably because each of these user groups has specific expectations and audiences in mind when using social media (Stier et al., 2018). For instance, political actors' social media communication cannot be completely disentangled from the fact that journalists are increasingly attentive to what happens on social media, and sometimes report messages in traditional news articles (McGregor, 2019).

The second important societal trend to consider relates to the way that the content and the user connexions are organised. Beyond the fact that the algorithms underpinning social media are not always made public by companies, there is a larger societal trend to organise the information from a demand perspective (Bivens, 2008). Indeed, where traditional newspapers organise information following an editorial line (thereby ordering the information based on the importance of societal and political events), today's social media tendency is rather to deliver a clustered content that matches peoples' views and preferences (thereby answering to "presentist" injunctions). There is thus a risk that the large-scale content available to conduct social media research be redundant and polarised instead of varied and balanced. The news media landscape also suffers from similar biases, notably due to the concentration of the media (e.g., media group acquisitions) and to the choice of fostering immediacy in the news.

1.5. Focus of the thesis: Twitter, "text-as-data" approach, Switzerland, and a descriptive approach of opinion

The theoretical framework and methods developed in this thesis can be applied to social media in general, but the empirical studies rely on Twitter as the main example. Twitter is a micro-blogging service that allows the sending out of short text-based posts[2]. Compared with other social media, its use for political purposes is more prevalent (Pew Research Center, 2021). It is also a platform that is much used by journalists and actors closed to politics (McGregor, 2019).It therefore represents a unique opportunity for researchers interested in the study of public opinion, especially in comparison with other platforms that have more restricted access to the data (e.g., Facebook). Although Twitter is only used by a subset of national populations (to a much lower extent than Facebook or Instagram), it remains one of the most popular platforms for academic research, notably because the content of the (non)political discussion that takes place on this platform can be easily accessed through an *Application Programming Interface* (API). Twitter offers other advantages for conducting social research. For instance, it is relatively easy to follow conversations and retrieve associated events or topics as publicly available tweets usually follow a hashtag logic and are referenced in google searches. After a period where it had become increasingly difficult for academics to access historical Twitter data, especially in the aftermath of the *Cambridge Analytica scandal*, the platform now offers a very large access to data and meta-information for academics (for instance, see the 2021 *Twitter API for Academic Research*). Twitter thus represents an important opportunity to study a range of views and topics from naturally occurring online settings to extreme circumstances, such as disease outbreaks or political protests.

[2] In the rest of the manuscript, the term "social media" is used to refer to social media applications such as Facebook and Twitter.

This thesis further takes a "text-as-data" approach (Grimmer & Stewart, 2013; Grimmer et al., 2021) as textual data forms the main source of analysis. If text is the focal point, it is also acknowledged that other multimedia components are becoming increasingly important, and sometimes occupy a central ground (e.g., Instagram), in social media communication (e.g., images, videos, etc.). In several empirical papers, there will also be emphasis on the importance of extracting information from the network of Twitter users (e.g., profiles of followers) for improving the interpretation of the findings from social media. Grimmer et al. (2021)'s framework timely addresses concerns held by social scientists and humanists in establishing natural language processing techniques to enrich many fields and applications of social inquiry from a descriptive, predictive, and causal perspective. Computational tools to extract information from the large amount of digital textual data (such as tweets) is unevenly adopted across social science disciplines. Indeed, several reasons are typically invoked by social scientists to opt against using social media data and computational text analysis methods in their research (Baden et al., 2021). Although this is often explained by reference to the rapid pace of development in computational tools and methods (Boumans & Trilling, 2016), as well as by the need for computational literacy (Domahidi et al., 2019), there are primarily important concerns related to social media ability to generate robust operationalisation and valid measurements of complex social constructs (Nicholls & Culpepper, 2020). Although the text-as-data approach educates researchers in the best uses of (non)supervised models for text classification, it does not emphasise the increased attention in other non-textual features (such as videos and images) or the recent developments of transformer models that dominate textual computation, such as applications with encoders for text understanding or decoders for text generation (Evans, 2022). These more recent developments are not covered in this thesis. In line with the pre-requisites aligned with survey methodology, the empirical chapters propose careful methods to constitute a corpus of social media messages, to pay specific attention to interpretive judgment for validation of findings, and to support multi-method analytical strategies.

Most of the empirical studies in this thesis focus on Switzerland as a case study. Albeit political and media users have become increasingly involved with social media in order to communicate with the public (Rauchfleisch & Metag, 2016), the broader public still live predominantly in the world of traditional mass media (Eisenegger, 2020). Furthermore, the consociational features of democracy combined with proportional representation reduce the impact that elections have on politics and policy outcomes compared to majoritarian political systems. In addition, the highly decentralised political landscape, with the cantonal and local parties playing an important role, renders national mobilisation less likely, especially on social media. However, the direct democracy component of the Swiss political system makes Switzerland a fruitful case study for investigating public discussion surrounding the multiple yearly referenda campaigns. From the perspective of the formation of public opinion, the Swiss case is also interesting because politicians and influential actors can gain a higher resonance on social media than they would in the framework of parties or organisations.

What is also absent from this thesis is extensive research about causality in textual analysis. Instead, this thesis provides empirical studies that provide descriptions and analyse dynamics of opinions. In general, social research can be differentiated between descriptive research which summarizes characteristics of a group or a theme, predictive research which aims to forecast future outcomes, and explanatory research which strives to understand underlying causal mechanisms (Shmueli, 2010). Each type of research requires different methods in terms of design, model building, and evaluation. Traditionally, the identification of causal relationships between phenomena has been the focus of social sciences, with prediction and description having a secondary role. For instance, description and prediction are valuable to provide the starting point for causal inference. Although the same classification can apply to social science research relying on social media data, description and prediction are prominent in social media based studies (Ledford, 2020). More specifically, computational social sciences tend to be dominated by descriptive models (Jungherr & Theocharis, 2017), especially in the form of data-driven approaches and simulation methods. The rising number of studies adopting a prediction approach based on social media data can thus provide valuable insights into the general goal of social sciences, which is to find general theories for understanding aspects of social and collective phenomena. A general problem for conducting causal research with social media data is that their inference from observational data is very challenging and usually requires the presence of a change (for instance, collecting data before and after an event or monitoring followers' networks) and the inclusion of important co-founders (for instance, media and algorithmic effects).

1.6. Public versus scientific usage of social media data for measuring opinion

Companies, political parties and organisations already look with great interest at what their followers are saying on their social media accounts. Moreover, journalists increasingly rely on social media content to convey public opinion (Dubois, Gruzd & Jacobson, 2018; McGregor, 2019). Many institutions have started to collect and analyse large volumes of data using a combination of computational and statistical tools (Lazer & Radford, 2017). For instance, health authorities were able to respond to public concerns in a timely manner based on the way users were disseminating terms like "H1N1" or "swine flu" during the 2009 pandemic (Chew & Eysenbach, 2010). Furthermore, social media enable us to assess who participates on various social media platforms and how they reflect or influence public opinion. This can help decision makers to develop better policies for tackling societal questions (Chen & Tomblin, 2021).

The rising role of social media within social movements and other social phenomena (e.g., political polarisation and fake news) has also encouraged social science researchers – driven by the wealth of information contained in online data – to investigate opinions and behaviours on a scale that was previously impossible due to the scarcity of data. For instance, social movement studies increasingly rely on social media data to analyse the emergence and development of collective identities and the usages of social media for collective actions (Segerberg & Bennett, 2011).

To date, it is still a very common practice to rely solely on surveys to study public opinions and attitudes towards important policy issues. However, given the proliferation of social media data and the increased importance they play in opinion formation, for anyone interested in studying public opinion "it would be foolish to ignore the information about public opinion revealed by social media data" (Klašnja et al., 2016, p.23). Furthermore, even if social media users are not representative of the general population, the opinions they express on social media might still be representative of important societal trends (Schober, 2016), some of which might not always be captured through opinion surveys.

In this view, comparing and combining survey findings with social media data is a more effective approach than relying on a single data source (Tufeki, 2014). For instance, instead of employing social media data as survey-like estimates of population parameters (Murphy, Craig & Dean, 2013), a fruitful approach is to use social media data for identifying nuances and intensities of viewpoints about a target or a topic of discussion.

In the meantime, social media data can best serve as a valid source of information for social research and contribute to the study of public opinion when they are benchmarked to or integrated with other data sources, such as survey data. Beyond the perspective of substituting surveys with social media data (Sajuria & Fábrega, 2016), the complementarity approach includes juxtaposing online and offline opinions about similar policy issues or integrating both data sources to explain the interactions between online and offline trends. In this way, survey research can be refined, since relying on both data sources enables us to develop a fuller picture of public opinion. Furthermore, carrying out social media analyses can be a more expedient use of resources for conducting preliminary and exploratory analyses than an expensive national representative survey. Social media can also offer a comprehensive understanding of an issue that is under-covered by survey research and provide frames that survey questions might miss.

Public opinion can be measured from descriptive, dynamic, and causal perspectives. Considering public opinion from these perspectives is central for the democratic debate as political representation needs to understand what opinions prevail at a given point in time, how opinions evolve, and what factors influence these opinions. For instance, political elites and journalists tend to consider Twitter as a barometer of public opinion (Jacobs & Spierings, 2016). Furthermore, social scientists have demonstrated an interest in applying social media data to understand public opinion, and even to replace traditional surveys (Gayo-Avello, 2013; Anstead & O'Loughlin, 2015), thereby leading to extensive research about the plea from Gayo-Avello et al. (2011) for a "model explaining the predictive power [or lack thereof] of social media" (p. 490).

Both opinion surveys and social media provide ways to study opinions from descriptive, dynamic, and causal perspectives. Furthermore, as social media are increasingly relied upon to impact political action (e.g., electoral campaigns, protest, e-participation, crowdsourcing), it is important to complement offline and online measures of opinions. Compared to surveys, social media certainly provide information that allows us to monitor public opinion almost continuously and instantly. However, to date, most studies relying on social media data aim no further than to describe and gauge the dynamics of opinions, the causal perspective still being dominated by survey research (Olteanu et al., 2016; Engel, 2022). From a descriptive perspective, it is important to take into account that, on various subjects, there are users that are actively voicing their opinions and others that are expressing their opinions very rarely, or not at all, by being passive observers (Vaccari & Valeriani, 2015). From a dynamic perspective, it is essential to account for the topics on which opinions are fixed and others on which opinions and views change. The causal perspective further highlights what event changes in opinions are taking place. In this view, complementing survey data and social media data can inform us about the effect of social media on citizens' opinions and attitudes.

1.7. Advantages and disadvantages of social media data for studying public opinion

Given the availability and richness of social media data, questions have been raised about whether traditional surveys are still required to observe public opinion. Indeed, social media have been perceived as cheaper than survey data, quicker to retrieve, larger in scale, and more diverse in content. However, social media pose substantial challenges for the study of public opinion.

During the last decade, several studies examined the use of social media, especially Twitter and Facebook, in measuring public opinion and social media's parallel with survey methods (Couper, 2013; Tufekci, 2014). The majority of these studies agree that social media data tend to be messier than survey data, unrepresentative of national populations, and require significant technical skills to derive valid measures of opinion (Klašnja et al., 2017). Furthermore, there are important differences in the practical and ethical considerations when using both data sources. Schober et al. (2016) provide detailed guidelines on these considerations, as well as a comparison between how survey respondents and social media users each understand their activity, namely responding *versus* posting. Other researchers have extended the existing framework for accounting for measurement and representation errors used in survey research – the TSE framework – to the realm of social media (Jungherr, 2016; Olteanu et al., 2016; Hsieh & Murphy, 2017; Sen et al., 2019). These studies represent fruitful attempts to arrive at a unified vocabulary across the main disciplines involved in the study of opinions and attitudes expressed online. Researchers interested in digging further into the respective differences of both data sources can rely on the above-mentioned literature.

The main pros and cons encompassed by each criterion are presented briefly. Table 2.1 summarises the main points of divergence that can be derived from the studies comparing social media data and survey data, in particularly: population coverage, topic coverage, geographical granularity, temporal granularity, data availability and authenticity of content.

Population coverage. In survey methodology, respondents form a representative random sample of a target population. In this scenario, representation errors arise when the responding sample systematically differs from the target population. The severeness of this error can be assessed using data available to describe the population (e.g., sociodemographics, location, language, etc). With respect to social media, the target population is typically identified via user accounts and is usually much larger in size than surveyed samples of the population (Sen et al., 2019). In practice, however, there is usually little information available to characterise the users included in a corpus of tweets. Furthermore, it is often the case that a minority of unique (self-selected) individuals dominates the discussion in terms of tweet and retweet volume, making oversampling of most active users very likely (Barberá & Rivero, 2014; Gruzd & Haythornthwaite, 2013; Mustafaraj et al., 2011). It is thus particularly important to account for who the users included in Twitter corpora are. In particular, there is a need to reflect on ways of including users that can be reasonably expected to represent individuals, groups or institutions. For instance, social media also has a growing availability of large volumes of relational data (e.g., networks of followers) that can be analysed to better understand social media trends (Holzinger, 2014). This is used especially for investigating the emergence and evolution of opinion-based communities, as well as information propagation in crisis situations. In the framework of the proposed thesis, we rely on the network information of social media users essentially for data collection purposes.

Topic coverage. According to Schober et al. (2016), understanding when topic coverage is achieved constitutes the central scientific problem for social media research. In survey methodology, topic coverage follows naturally from population coverage, as researchers control the constructs measured with survey questions. The scope of the investigated phenomenon is thus confined by the diversity of questions. For social media analyses, topic coverage need not be achieved with population coverage. However, the extraction and summarisation of relevant information and dimensions from social media messages often constitutes a major challenge. In this view, there is a need to focus on analytical strategies that can render social media data complementary to survey insights. When measuring opinions with social media data, results should be benchmarked against trusted ground truths (e.g., surveys, records, manual annotations, etc.).

Geographical granularity. Survey research enables us to conduct cross-national comparisons on similar concepts to assess cultural and geographical differences.

However, cross-national surveys face the particular challenge of trying to balance optimal survey quality within a country and comparability across countries. A major challenge is that understandings of concepts may vary greatly between countries or geographical areas (Dahlberg, Axelsson & Holmberg, 2020). Although survey data are delivered with country indications, they rarely enable the investigation of cross-national trends with the ease of social media. Indeed, these platforms offer an alternative way of collecting worldwide opinions on a common object (or policy issue) at relatively low cost, while also conserving the variety of views and understanding that prevail in different contexts. However, Hecht et al. (2011) found that up to a third of Twitter users do not provide any sort of valid geographic information. There is thus a need to propose ways to assess the geographical venue of users, notably by relying on lists of country-specific seed users (e.g., politicians, newspapers, etc.). It is also important to rely on analytical techniques which conserve the richness of social media information to enhance survey findings.

Temporal granularity. Survey data can efficiently measure opinion trends by submitting a similar questionnaire at several points in time. It can also draw causal inferences by conducting panel studies. However, conducting such survey research is generally expensive and involves considerable efforts to keep data quality high over time. The temporal granularity of social media is certainly one of its main advantages compared with surveys. For instance, social media data is very useful for observing the emergence and development of public opinion. Observing this does not depend on the analyst asking a preconceived question as in surveys. Thus, albeit social media data can be a barrier to measuring what the population is thinking on an issue, collecting large volumes of messages should make it possible to observe when new issues emerge or when opinions are changing on a given issue (e.g., change in sentiment, reactions towards events, etc.). However, the results obtained from social media data can change significantly when using different time windows (Morstatter et al., 2013; González-Bailón et al., 2014) or different collection techniques (e.g., Twitter's Firehose *versus* Twitter's streaming API). It is thus important to justify the technique of data collection chosen and to account for the data collection periods. Furthermore, unlike survey data, the collection of retrospective data is made easier with the history of social media platforms. Taking these differences into account is very important for understanding when surveys and social media insights can be complemented (and when not).

Data availability. In survey research, the data available to researchers are usually represented in a tabular mode with data points coming from individual respondents answering the same questions. Similarly, open-ended answers can be coded to categorised responses. There are well-established practices about how to report the methodological choices (e.g., weighting procedure, code book, questionnaire, etc.) and how to archive changes. However, survey data are usually released one or two years after data collection, which can decrease the actuality of a topic. On the contrary, social media platforms enable immediate access to the data (albeit depending on the data authorization access). Furthermore, the structure of the data available to researchers depends on the content of messages and on the analytical techniques to identify (e.g., select relevant content) and transform (e.g., transformed into survey-like format) the data.

Authenticity of the content. The authenticity of the content poses very different challenges depending on the data source. For instance, survey data can be affected by social desirability bias which suggests that respondents may answer in ways intended to make researchers evaluate them positively. However, the biggest challenge is probably linked to respondents' perceived burden of the survey, which either leads to abandonment or to renouncing participation. Social media content also suffers from social desirability bias in the sense that users try to manage the impression they make. In the extreme case, it might be that the sample of users who discuss the topic of interest does not exist. Indeed, it can be that individuals simply do not give their opinions on a sensitive topic on social media (Schober et al., 2016). It is also possible that individuals engage with the topic but temper their opinions when made publicly available on social media. Conversely, it is possible that individuals make their claims stronger online, thereby possibly leading to polarisation or exaggeration. Additionally, there is the possibility that individuals falsify their position, for example to be politically correct or to seek approval and reward from specific audiences. Other issues relate to the fact that users can create fake accounts or broadcast false information. So-called "trolls" and "adds" can post false information and produce political content for a specific goal (Tucker et al., 2018). There should always be reflection on whether (and how) these effects can impact the results of a study that measures opinion using social media data.

Table 2.1.: Advantages and disadvantages of survey and social media data according to essential methodological requirements

	survey data		social media data	
	advantages	**disadvantages**	**advantages**	**disadvantages**
Population coverage	Representative samples (inferences can be achieved)	Small samples (~1500-2500 individuals)	- Large-scale coverage of social media users - International span	- Unrepresentative of national population (e.g., non-human accounts) - Difficult to access personal information (e.g., location, gender, age, education, etc.)
Topic coverage	Well-defined concepts	Limited by the questionnaire length and the choice of survey items	- Data collection can be theoretical and data-driven - Topic coverage is not limited *a priori* by researchers	Topic coverage influenced by the method of data collection (e.g., choice of search queries, choice of seed accounts, etc.)
Geographical granularity	Cross-national studies enable geographical comparisons	Difficult to keep the concepts constant between countries and cultures	- High precision of geographical information (if available) - Conservation of the cultural richness of the data	Geographical information not always available (e.g., opt-in feature)
Temporal granularity	Longitudinal studies and the repetition of survey items enable temporal comparisons	Infrequent (e.g., round of every two years for cross-country surveys)	- Near real-time collection - Sensitivity to events - Frequent data collection - Monitoring of attitudes, opinion, and behaviours	- Results can vary greatly when using different time windows. - Difficult to measure the influence of social media discussions on public opinion (or the other way around)
Data availability	Data well-structured (e.g. tabular mode), transparent (e.g., weighting procedure, code book, questionnaire, etc.), and with changes actualised	- Release usually done one or two years after data collection - Retrospective data hard to collect (e.g., memory problem)	- Immediate access to the data - Retrospective data available (if not deleted)	No well-structured format (e.g., need extensive pre-processing)
Authenticity of the content	Respondents from carefully chosen samples	- True opinions: measures of opinions would not exist without researchers' intervention (e.g. questionnaire) - Social desirability bias	Data "as such" (not solicited by researchers)	- Perceived audiences or social norms online can influence the content of messages and the behavioural patterns are adopted (e.g., share, like, reply) - Fake news and potential manipulation

One of the most often mentioned concerns about the use of social media for conducting social science research is that social media users might not be representative of the general public. Most notably, social media users tend to be younger, more male, and more politically engaged than the general population (Mellon & Prosser, 2017). Furthermore, some users might be excluded (e.g., banning rules), missing (e.g., perceived audiences and elitist bias) or little visible (e.g., algorithmic logic and polarisation bias) because of the platform regulations and opportunities. Researchers must take these factors into consideration since they may affect the usefulness of Twitter data for social research but also the conclusion of longitudinal analyses. To correct for the potential non-representativeness of social media subsets, scholars from the field of computational social sciences have sought to extend post-stratification techniques from survey research to social media data (Gayo-Avello, 2013; Beauchamp, 2016). Other studies have proposed applying random sampling in the framework of social media (Morstatter et al., 2013). However, both techniques might become increasingly untenable given the proliferation of online modes of communication producing opinion data which are almost universally non-random. Furthermore, social media accounts also group many non-individual users, such as organizations, associations, media accounts and political parties, as well as non-human users, such as bots.

Despite the general lack of representativeness of social media users, there are essential arguments that make social media data relevant for the study of public opinion. For instance, many societal debates happen, and sometimes even start (e.g., social movements), on social media. Thus, simply knowing the opinion of social media users can be relevant. Indeed, even if the actual people who are tweeting are not a representative sample of the population, the content of online discussions can end up being representative of the concerns of the public (Farhadloo et al., 2018). At a time when survey response rates are declining and social media reliance is growing, it is no longer possible to ignore the opinions, attitudes or behaviours that can be found on social media (Klašnja et al., 2016). In this view, social media platforms, like Twitter, offer new possibilities of studying which actors are leading or strongly influencing the debate dynamics (Rauchfleisch, Vogler & Eisenegger, 2021). Additionally, by highlighting the opportunity provided by social media to find significant "profiles of thinking", one might adopt a perspective that goes beyond the notion of representativeness, making the quantification of the frequency of each profile a secondary concern.

There are additional reasons to consider social media data as valid social and political indicators of public opinion. For instance, social media content has the potential to impact what the broader public thinks on important societal matters. Furthermore, the questions covered by surveys are also limited by the scope of the questionnaire, thus making surveys less reactive to events and emerging societal questions. Social media data is more spontaneous (i.e., not research induced) and reactive to events, thereby enabling researchers to include an ethnographic dimension into public opinion research.

Now, the popularity of social media data for generating large datasets does not come without challenges for understanding public opinion. For instance, scandals such as that involving *Cambridge Analytica* where millions of Facebook users had their data collected for political advertising purposes pose serious concerns with respect to public opinion formation. A similar concern is raised by phenomena like the spread of fake news and the activity of bot accounts (i.e. accounts that operate without human involvement to post and interact with others on social media sites), the influence of which on public opinion trends constitutes a valuable line of study. Both of these phenomena have important implications when using social media to understand the formation and evolution of opinions (Venturini & Rogers, 2019).

When relying on social media data there is a need to better understand which data collection and analytical methods are best suited to answering particular research interests (Grimmer et al., 2021). There is, as yet, little consensus about how opinions should be measured via social media and how findings from social media data can reflect, impact, or lead public opinion. Early studies were very enthusiastic about the volume and reach of social media data for the purposes of conducting social science research. However, the fact that the amount of social media data is bigger (in the sense of the amount of available data) does not necessarily lead to better measurement of a phenomenon. Indeed, the "information-to-data ratio" from social media is comparatively lower than the information stemming from opinion surveys which rely on well-structured items (Groves, 2011; Lazer & Radford, 2017). Furthermore, Pasek et al. (2019) demonstrate that patterns observed in social media and survey data can measure very different things. Here, the choice of method and measures plays a crucial role. Indeed, they show that sentiment surrounding tweets about the president is no proxy for presidential approval, although attention to subgroups improves the extent to which survey and Twitter data yield similar conclusions.

To date, there seems to be an agreement about the use of social media data for social research: we should neither ignore the information about public opinion revealed by social media data, nor should we treat the measurement of social media data in the same way as we would well-designed and representative surveys (Klašnja et al., 2017). When investigating public opinion, representative surveys should still be the main go-to method. What social media can certainly offer are unique insights into the dynamics of (online) debates about given policy issues and events.

1.8. State-of the art "text-as-data" applications and approaches

This thesis proposes to view the concept of public opinion through the complementary lenses of social media data and survey data (see chapter 2). It does so in ways that are directly coupled with a critical assessment of the methods for identifying and reviewing opinions from social media data. Most notably, it provides a social science perspective on the most suitable computational methods for deriving information from social media data (see chapter 3). It also emphasises the importance of the study design for collecting the social media data, while underlining the necessity of accounting for the groups of users involved in social media discussions so as to better contextualise and understand social media insights (see chapter 4). It proposes ways to complement both data sources going from exploratory to explanatory research depending on the availability of survey data (chapter 5). This thesis thereby complements important efforts dedicated to the development of sophisticated methods, algorithms, and codebooks to render both data sources comparable. Before presenting how each chapter contributes to answer the research questions underpinning this thesis, this part provides a summary of the state-of-the-art methods to derive opinions from (social media) texts.

Textual features from social media data constitute the core complement of survey data in this thesis. Social sciences have considered text as a central research material since a very long time (e.g., news articles, transcripts of political speeches, radio broadcasts, and TV shows) because "text is arguably the most pervasive – and certainly the most persistent – artifact of political behaviour" (Monroe & Schrodt, 2008, p.351). However, social sciences have only recently relied on text-as-data approaches to quantitatively explore more text than it was previously feasible with manual annotations. To date, these approaches have become increasingly popular in the social sciences. This can be explained by the fact that the increased public reliance on (social) media platforms and the rapid development of computational (social) science tools have enabled researchers to answer new research questions and provide new perspectives on old research questions (Lazer & Radford, 2017; Evans & Aceves, 2016).

Unlike the well-established survey methodology measuring opinions and attitudes from representative samples of the population, research designs studying social phenomena based on social media data are not yet fully developed and there is still little relevant guidance in the existing methodological literature (Grimmer & Stewart, 2013). To date, text data, notably from social media, have a variety of applications. For instance, Gilardi and Wüest (2018) presents three specific kinds of applications in comparative policy analysis: concept identification, classification, and discovery. Most recently, Grimmer et al. (2021) classified the applications going from discovery to measurement and inference in social science.

Text-as-data approaches are about the compression of the high dimensionality of the textual data so to render the data interpretable and generalisable in ways that would be otherwise difficult or infeasible. Existing methods can be classified into two main approaches: unsupervised and (semi)supervised approaches (Grimmer & Stewart, 2013). Unsupervised methods of summarising text data eliminate human coding altogether but place the burden on the researcher to justify their chosen interpretation and validate the utility of the categories learned. Supervised methods help researchers to assign texts to predefined sets of categories but do not eliminate the necessity of manual coding. Semi-supervised between the supervised and unsupervised learning approaches as it uses both labelled and unlabelled data for training a classification task.

1.9. Choosing a method and assessing output quality

Approaches of text analysis are constantly evolving and there are few studies that organise existing methods and tools for helping social researchers in making informed methodological choices. Grimmer and Stewart's (2013) article is very helpful for getting an overview of the field. In particular, the authors underline the differences and respective advantages of supervised and unsupervised methods given the current state of knowledge and some research objectives. Table 3.1 summarises the aspect and task extraction, as well as the method, aggregation, and evaluation of the approaches used in this thesis. Following is a description of each approach.

Dictionary-based methods rely on a list of carefully chosen terms representing categories. Dictionaries are generally more difficult to validate, especially when a dictionary is developed outside the data under analysis (Grimmer & Stewart, 2013). However, dictionary-based approaches can enhance external generalisability as, in contrast to supervised models, dictionaries can cover cases that are not represented in the training set. This constitutes a major reason why several lexicons have been used extensively and on many different domains for classification tasks – such as sentiment detection (e.g., Bing, Afinn, or LexiCoder), emotion detection (e.g., LIWC) and policy issues classification (e.g., LexiCoder Topic, Media Frames Corpus).

Supervised machine learning is more appropriate when classification categories can be specified before analysis (for instance, based on theory). Therefore, the model performance in accomplishing classification can be assessed using accuracy metrics (e.g., F score, recall, precision). The validation is done using training data that already contains the categories (usually manually annotated). The distribution of the dictionary categories can be compared to this annotated training dataset. However, supervised learning that has been trained on a given corpus can suffer from decreased accuracy when a classifier predicts unknown documents from a different time period or a different domain. This underlines the necessity to have a representative sample of texts to train supervised models and to construct codebooks explaining manual annotations. Another issue with supervised document classification is that the categories are not necessarily clear-cut. The data might be too sparse to represent every category, which can lead semantic categories to overlap (e.g., climate and economy).

Unsupervised methods do not presume any predetermined categories as they rely on a data-driven approach to cluster documents into relevant categories. Therefore, these methods can offer a solution to the problems of supervised learning exploiting the semantic information. However, unsupervised methods impose more difficulties in interpreting the relevance and the accuracy of the clusters afterwards (Burscher et al., 2015). In this thesis, we relied on the following unsupervised techniques: topic modelling, distributional semantics, word embeddings, and cluster analysis.

- In topic modelling (Blei, 2002), the category predicted by the algorithm is given in the form of a probability for each category based on the features. Eventually, the category with the highest probability is selected to label to document. The outcome of topic modelling be tested for its validity. In a first stage, candidate models can be fitted with varying numbers of clusters or topics and examined within several parameters (e.g., coherence, exclusiveness, perplexity). In a second stage, human judgement is used to select a final model and to interpret the quality of the groupings or clusters. The human interpretation can include the interpretability (e.g., content coherence) and the evolution (e.g., time coherence) of topics.

- Distributional semantics can provide an answer to major drawbacks of topic modelling, namely: the fact that the topic categories are generally broad, and the fact that the reported top keywords are often either loosely related or overlapping. Distributional semantics is based on the hypothesis that "[y]ou shall know a word by its company" (Firth, 1957, p.11) in order to detect semantically similar words. Sahlgren (2006) shows that the size of the context plays a crucial role. By taking the document sequence into consideration (e.g., by ordering the documents according to time), a large observation windows can take the global ordering of the documents into account to gain an overall interpretation of the semantic space (e.g. conceptual maps based on Kernel Density Estimation). However, a serious challenge is that changes in parameters (e.g., window size) can lead to quite different results and it is sometimes difficult to assess which is better.
- Word embeddings are a modern incarnation of distributional semantics. Embeddings are nowadays routinely used in deep learning architectures where pretrained embeddings allow neural networks to capture semantic similarities among lexical features to conduct supervised tasks. Word embeddings enable us to provide detailed framings of a topic. To date, it has become popular to train word embeddings in an unsupervised way. Embeddings can be viewed as dense vectors that project words or short texts in a vector space. The trained embeddings can then be used as features in complement to a dictionary approach or used with machine learning algorithms.
- Cluster analysis is also a data-driven method to perform automated content analysis by exploring the content of documents. It relies on the co-occurrence of words to extract shared semantic regions. In particular, correspondence analysis enables to understand dimensions for categorical data (e.g., words) by projecting them on a two-dimensional factorial space where proximity of data indicates shared semantic meaning. To improve the interpretation of correspondence analysis, its results can be used in combination to cluster analysis to extract relevant regions (e.g., topics of discussion), which can then be analysed along other criteria (e.g., time, sentiment, actor groups).

Table 3.1: Summary of social media analytical steps from data collection to model evaluation of the method used in the thesis

Aspect extraction	**Task representation**	**Method**	**Aggregation**	**Evaluation (performance)**
Depends on the objectives of the study, especially the theoretical concepts:	Depends on pre-existing knowledge:	Depends on the multi- or single-labelled task:	Depends on the level of analysis:	Depends on the approach and specific method:
Classification into *a priori* defined categories (e.g., sentiment, policy issues, stance)	Known categories	- dictionary-based approach	- each tweet is labelled	- comparison with held-out sample of manually annotated tweets
	Known categories	- supervised approach	- each tweet is classified	- comparison with held-out sample of manually annotated tweets
Extract topics	Unknown categories	- topic modelling	- each word and tweet is assigned to multiple topics	- coherence and/or exclusivity of terms - interpretability (topics and time)
Extract clusters	Unknown categories	- clustering (e.g., correspondence analysis)	- group of tweets (or accounts or terms) are clustered	- interpretability
Derive semantic information	Unknown categories	- word embedding	- collections or word similarities are derived	- interpretability

1.10. Combining text classification approaches

The way in which a method is chosen, and findings interpreted, will depend on the research question. The focus should be placed on the research process and on the critical reliance on state-of-the-art methods to answer social science research questions. Importantly, there is no "one-size-fits-all" data source and method that can be applied to answer a given research interest. Rather, researchers often have to complement several data sources and try several methods to evaluate their performance and suitability to accomplish their research objective.

To date, an increasing number of research papers propose ways to combine these different approaches to augment their respective strengths. For instance, word embedding techniques can help to construct and expand the scope of existing dictionaries in ways appropriate to a given context (Amsler, 2020; Reveilhac & Morselli, 2022). Furthermore, the dictionary approach can be used to incentivise automatic classification into specific directions (Dobbrick et al., 2021).

In addition to the combination of approaches, it is also advisable to use a methodology that combines automation and human work. For instance, (un)supervised methods are certainly less laborious than manual coding, which is the main drive for using automated methods. However, these methods also need human coding as a check for their validation. Furthermore, triangulating the prediction from dictionary and (un)supervised models can compensate for the respective weaknesses of each approach.

Overall, computational social science need to forge a closer link between conceptual reflections (through theory-driven construct) and automated measurements (Baden et al., 2020). For instance, unsupervised methods are very useful to cluster texts into fewer interpretable categories. However, these categories are not necessarily of theoretical relevance given a research question. In this view, a promising approach in developing text-as-data applications is to combine different approaches of text classification. For example, the combination of off-the-shelf dictionaries with supervised machine learning has been conducted in different research areas, such as the study of populism (Gründl, 2021) and online debate quality (Dobbrick et al., 2021). Another possible approach consists in combining labelled documents, which serve as a baseline for the topical content of the textual data (e.g., open-ended survey answers), with unsupervised method to expand the domain of classification while also allowing for new topics to emerge (Stier et al., 2018).

Importantly, the combination of different approaches (e.g., dictionary, supervised and unsupervised learning) should enable researchers to improve the identification of the major source of weakness in single approaches and to propose alternative ways to classify texts reliably. For instance, dictionaries have the advantages of requiring little human input, they are easily interpretable and implementable. However, they suffer from major weaknesses related to the domain specificity for which they were elaborated. Dictionaries are therefore usually hard to transfer to different domains without adjusting the word lists to a new or unseen context (González-Bailón & Paltoglou, 2015). Furthermore, dictionaries focus on the word unity, thus, disregarding the grammatical structure of a text (Chan et al., 2021). Moreover, dictionaries usually follow an additivity assumption where all features are equally relevant to assess the presence of a construct (Young & Soroka, 2012). Machine learning can generally solve these problems and reach significantly better agreements with manually coded data (Van Atteveldt et al., 2021). However, we still have limited knowledge about the reasons why these approaches can perform better than dictionary-based methods (Dobbrick et al., 2021).

The methodological and empirical articles presented in this thesis (see chapters 3 to 5) generally follow an assumption-based framework that intends to locate major sources of weaknesses in the different text classification approaches. For instance, both methodological papers (see chapter 3) seek to capitalise on the information carried by dictionaries with pre-defined word lists while integrating (un)supervised tools to improve the classification task with a minimal amount of manually annotated data and to improve the interpretability of the classification results.

1.11. Validity and reliability of social media findings

As methods of text analysis are often developed as stand-alone tools, there is a risk that theoretical conceptualisation and concept validity may be downgraded in favour of measurability concerns and accuracy metrics (such as recall and precision) or post-hoc validity demonstrations (Baden et al., 2021). Reliability and validity both concern how well a research method measures some concept of interest. Reliability refers to the consistency of a measure, whether the results can be reproduced under the same conditions. In general terms, the notion of validity refers to the ability of measurements in a dataset to represent a theoretical construct. This construct validity is especially relevant when a concept is latent and has to be operationalized via some observed features. Other aspects of validity include internal validity, that is the extent to which the measurements correctly support the conclusions of the study, and external validity, that is the extent to which research findings can be generalized to other situations. External validity also suggests the ability to assess the extent to which results can be correlated to external factors (Trochim, 2006), thus also including ecological validity, that is the extent to which outputs properly reflect a broader real-world phenomenon (Ruths & Pfeffer, 2014). It can also refer to temporal validity which captures the extent to which an output (e.g., topic or concept) changes over time (Howison et al., 2011).

This section discusses the ways in which the validity and reliability of the findings from the methodological and empirical studies (see chapters 3 to 5) are affected by choices made regarding the methodological approaches, notably the data collection strategies (including the question of missing data), the reliance on manual coding (including a discussion about the efforts versus the gains of introducing the human-in-the-loop), and the choice of existing packages and libraries to conduct the analyses. It also briefly discusses the idea of "black-box" models and how they relate to validation concerns.

Choices made to collect the data must ensure that data faithfully represents the phenomenon being studied, most notably to ensure construct validity. When data collection is based on keywords, researchers should make sure that their list is inclusive enough and should discuss what content could be overlooked or ambiguous. The choice of search queries is very important as it can lead to the inclusion of data that are not informative (or loosely related to) the construct, but it might also miss the inclusion of important and relevant information (see discussion by Sen et al. (2019) on the notion of "trace selection error," which is similar to the notion of "measurement error" in survey methodology). This inclusion of relevant search strings can be particularly challenging for overwhelming theoretical concepts, such as democracy. For instance, a strategy proposed in the second methodological paper (see chapter 3) is to rely on an iterative process for developing a final list of keywords that starts from a reduced (theoretical) list of target queries (including hashtags and relevant synonyms) and uses word embeddings to generate new candidates that are manually checked before their iterative inclusion in the extended list of keywords. The extraction of new keyword candidates carried out by word embeddings is repeated until saturation is reached; that is, until no more new and relevant keywords appear in the list of candidates (see methodology developed by Amsler, 2020).

A similar strategy relying on word similarity and collocation can be applied to the collection of tweets on a topical basis, such as narrower theoretical concepts (e.g., immigration or vaccination) or specific events (e.g., women's strikes) as demonstrated in the empirical papers (see chapters 4 and 5). In general, the elaboration of a list of search queries requires researchers to make sure that the search queries are encompassing enough not to miss substantial aspects of the discussion about the topic, but precise enough to avoid including unrelated aspects. When the data collection is user-based (e.g., specific political accounts and/or their followers), the selection criteria is more straightforward and relies on the correct and complete identification of groups of users. When the collection is based on an external list, the over or underrepresentation of certain categories of users must be assessed against the "true" prevalence of the same groups in society (e.g., the share of elected representatives active on Twitter versus their share of seats in Parliament). However, when no external list exists and the user-based collection requires the identification of the users based on the presence of relevant profile description information (e.g., job activity or membership in association), it is necessary to apply a similar strategy to that of topic-based data collection (e.g., survey data can inform us about the share of the population that consider a topic to be of "the most important concern"). Nevertheless, the lack of profile information explicitly mentioning the search queries can result in the non-inclusion of relevant accounts (note that spelling mistakes and other grammatical specificities can also lead to missing data). Data collection based on geolocation information is another approach. However, this type of information is very often missing from user profile information. However, as certain research questions (see last empirical paper in chapter 5) require being able to gauge cultural contexts, it is essential to collect tweets which can be reasonably considered to stem from accounts nested in specific geographical contexts. Therefore, an additional data collection strategy used in this thesis relies on the network of target users (e.g., elected politicians or media) as a proxy to the geographical location of their (most active) followers. For samples of users that have a manageable size, a good practice is also to account for the share of users with valid geographical information (e.g., national versus foreign accounts). Finally, researchers must also decide whether to include single data items (e.g., original tweets) or data in multiple copies (e.g., retweets). Indeed, including identical (duplicates) or almost identical (near duplicates) content can sometimes distort results, yet redundancy may also be a sign of importance.

The reliance on manual coding can intervene at the data collection stage (see above discussion). It is also prevalent at the data analysis stage and has both advantages and disadvantages. Manual annotations are a useful form of enriching data by labelling tweets according to their categories, both for training machine learning applications and for interpreting the results of automated text classification models (see empirical chapters 4 and 5). Most of the time, human coding is essentially conducted to assess the accuracy of a (semi)supervised classifier by labelling a training dataset. In this view, human intervention is generally time intensive and constrained to small samples (unless crowdsourcing is carried out). Manual annotation is also useful for incentivising a model toward a specific direction by constructing classification rules (see second methodological paper in chapter 3). The main advantage of introducing the human-in-the-loop is certainly to improve the external validity of the results, which further complement the benchmarking of social media findings and the comparison of social media data against other data sources. This said, human intervention has also disadvantages for the validity and the reliability of the research process. For instance, manual annotations can include a subjective dimension which is poorly specified and too vague to be implemented automatically. Similarly, manual annotations can introduce noisiness when several annotators disagree (here, an inter-annotator agreement score is needed) and when some patterns are more easily recognisable than others (this can introduce an over-representation of certain categories). Furthermore, manual annotations are necessarily limited in the amount of text that can be labelled. With this respect, the paper by Barberá et al. (2021) offers recommendations about different annotation strategies (e.g., coding at the segment or sentence level, and the trade-off between the number of annotators and documents). The reliance on manual work to build rule-based models for text classification can also provide researchers with an additional control on the validity and interpretability of the results. For instance, even in cases where manual work may bring only little classification improvement compared to machine learning application, rule-based models allow researchers to assess how manual decisions about “closed” features (e.g., words that are chosen theoretically or words belonging to syntactic or cognitive categories) can lead to congruent (or different) results than those that would be obtained from “open” features (e.g., all words in the messages), which are typically used in machine learning models. Similarly, Jaidka (2022) tested the extent to which a closed-vocabulary approach can be more transferable across domains than a content-sensitive (or open-vocabulary) approach to detect different dimensions of deliberative quality.

Concerning the choice of existing packages and libraries to conduct social media research, there are two essential stages where the choice of packages intervenes: the data cleaning step and the data analysis step. Pre-processing generally refers to cleaning the text data for analysis so that further applications aiming at deriving value out of the textual data can do so with the uninformative and noisy features having been removed. Several well-known and up-to data packages are readily available. In this thesis, we mainly rely on the packages from the R programming language specific tasks, such as stop-word removal (see the tm package) and lemmatisation or stemming (see the udpipe package from Wijffels, Straka, and Straková (2018) or the TreeTagger software developed by Schmid (1994) and its available R wrapper from the koRpus package (Michalke et al., 2021)). Pre-processing may also involve transforming the text into other formats, such as a term-frequency or document-frequency matrix and n-grams (see the quanteda package (Benoit et al., 2018)). Other functions can rely on pattern matching (see the grep or grepl functions) to remove certain features (e.g., punctuation and URLs) and to normalize other features, such as separating concatenated expressions or words (e.g., ClimateChange becomes climate change). Pre-processing steps are important as the choices made here can lead to very different results. For instance, researchers could choose to keep specific terms (e.g., keywords from the list of search queries) or to add them to the stop-word list. When conducting pre-processing steps (other possible cleanings include lowercasing and spell corrections), researchers should be careful not to introduce errors and skew results during these steps (Denny & Spirling, 2016). Concerning the data analysis step, the choice of methods used for aspect of extraction depends on the research question being asked (e.g., measurement of the concepts) and the context of the research (e.g., pre-existing knowledge or exploratory analysis). In addition to the necessary justification of the choice of method, researchers should also be transparent about the model specifications, which are also typically linked to the choice of packages, software, and off-the-shelf lexicons. For instance, this thesis relies on several well-known and validated tools to conduct machine learning models (see R packages quanteda and H2O (LeDell et al., 2018)), topic modelling (see the Mallet software developed by McCallum (2002)), word embeddings (see the R package wordVectors), cluster analysis (see R package FactoMineR (Husson et al., 2020)), and lexicon-based analysis using off-the-shelf lexicons, such as LIWC (Pennebaker, Francis & Booth, 2001), Lexicoder (Young & Soroka, 2012), HuLiu (Zhao, Liu & Xu, 2016), AFINN (Nielsen, 2011), and NRC (Mohammad & Turney, 2013). The valid measurement of social or political concepts thus requires the concatenation of different text analysis methods (e.g., triangulating dictionary and machine learning approaches) or algorithms (e.g., applying the majority rule to label the relevant category with machine learning approach). The same social scientific text analysis frequently requires a multitude of different measurements, which may be easily combined in manual content analysis, but almost inevitably requires the concatenation of different measurement procedures (Schoonvelde et al., 2019). In this view, studies that compare different methods for measuring similar concepts, whether or not a gold-standard measure exists, are very useful to advance our knowledge of the suitability of given text analysis methods for a specific task (for instance, Van Atteveldt et al., 2021).

A frequent concern voiced by social scientists relates to the “black box” character of many text classification methods, especially machine learning models, and their sensitivity to model specifications. With the expression “black box”, researchers mainly refer to models containing complex mathematical functions (e.g., machine learning models such as support-vector machines and neuronal networks), as well as deep representational spaces (e.g., k-nearest neighbours). Internal and external validations are used for measuring the effectiveness of machine learning models. For internal validation, most studies rely on k-fold cross-validation[3]. For external validation, studies usually use two databases of the same nature that do not share any items, and where the first database is used for training the model (and used for internal validation) before being applied to the second database (as external validation). One of the main weaknesses of using these validation procedures is that the focus is on the performance (or accuracy) of the models rather than on the interpretability (or understandability) of the results. Against, this background, black-box models are conceived as more accurate than explainable models in some contexts, while explainable models (also called “white-box” or “glass-box” models) are seen as being more robust than black-box models because they can obtain comparable results and can explain results based on (linguistic) patterns and hypothesis

[3] K-fold cross-validation means that the original database is randomly partitioned into k equal-sized datasets, where a single dataset is retained as the testing dataset, and the remaining k − 1 datasets are used as the training dataset. After the cross-validation procedure is repeated k times, the k results are averaged to produce a single estimation.

testing (Loyola-González, 2019). Additional validity checks for machine learning models can include the analysis of the word features that have the biggest influence on the model output, and how many of these features go past a threshold (mean or median feature weight). This type of analysis is similar to the analysis of top words from topic modelling, which enables researchers to make sense of the topics extracted by the topic algorithm. For topic modelling, the difficulty lies in deciding on a meaningful number of topics to extract, which is often based on perplexity and/or coherence measures (see R package ldatuning by Nikita and Nikita (2016)). For both, machine learning and topic modelling (but also for lexicon-based analyses), additional validity verifications can include the comparison of the distribution of categories (or topics) across external factors (e.g., can peaks in the data be interpreted against "real-world" events? Does the prevalence of certain categories make sense based on the source of the data, such as the actor or the country?). For other types of analyses, such as word embeddings and cluster analysis, the reduction of the multidimensional space into two dimensions must yield interpretable results that correspond to theoretical expectations or that can be explained by the reliance on theory. Readers who would like to know more about the role of social theory for validating machine learning outputs can refer to the article by Radford and Joseph (2020), which emphasises the ways to test for the interpretation (what did the model learn, and how well?), explanation (why did the model learn this?), and theory-building (what does the model teach us about the world?) of models.

Before turning to the summary this thesis's content, it is worth noting that extensive theoretical work has been done to develop a quality assessment framework for social media data, most notably by adapting the TSE framework from survey methodology to encompass social media data. For instance, Olteanu et al. (2019) systematically point to errors and biases that affect the findings of studies based on social media data. To date, Sen et al. (2019) propose the most complete version of the extension of the TSE framework to social media data specifically, and digital trace data generally. Researchers interested in learning more about how to systematise a quality analysis of social media data can refer to these works.

1.12. The thesis in a nutshell

1.12.1. How social media data are used for the study of public opinion

Chapter 2 discusses how and whether social media data can complement traditional survey data to study public opinion. It draws from the scientific consensus that social media data are best suited to complementing rather than to replacing survey data. There are currently two broad applications of social media data for social research. Firstly, social media data are used as a way of knowing what the public is currently thinking about, notably in view of predicting salient current issues (e.g., election outcomes). Secondly, they are employed as a way of analysing opinions (or sentiment) on specific policy issues. Albeit these two categories comprehensively cover the current practices in social sciences, many more applications exist and are emerging.

Chapter 2 thus proposes an extensive review of the theoretical and empirical studies that use both data sources in the framework of public opinion research (187 papers are reviewed: 141 empirical and 46 theoretical). It focuses specifically on identifying and critically assessing the use of social media data for social research as a complement (and not as a replacement or substitute) to opinion surveys.

The review demonstrates five approaches complementing social media and survey data: i) comparison approach: this suggests comparing both data sources around a particular phenomenon (e.g., understanding of democracy or depiction of migrants); ii) enrichment approach: this generally relies on a data-linking strategy between existing surveys and social media, where social media data provide researchers with complementary information that is hard to collect through surveys; iii) "survey as proxy" approach: this uses social media data as the main source of analysis, while the survey data is used for contextualising or calibrating purposes; iv) recruitment approach: this recruits individuals (e.g., activists, social minorities or scientific experts) where few (or no) survey data exist on social media with a view to conducting a second survey phase; v) generating "new insights" approach: this aims to investigate "old" (e.g., policy issue ownership or responsiveness) or "under-investigated" (e.g., health technologies or interest in science) topics or theories using social media data with the aim of revealing unexplored dimensions of the public interest and perceptions.

The review highlights the implications of each approach and derives practical guidelines in line with different research purposes, namely validating survey findings, improving the sustainability of the research by diversifying the views on a phenomenon, improving the reliability of survey measures by specifying measurements, and improving the interpretability of social or political issues. An additional contribution of the review is to complement existing theoretical articles stressing the importance of developing a framework that accounts for the possible biases of social media data while remaining in, or mirroring, frameworks from survey research (e.g., TSE framework).

1.12.2. What reliable methods for extracting opinions from social media data are

Chapter 3 departs from the ideas that expanding the utility of social media data for the study of opinions suggests developing methods to make sense of this extremely rich data source that will create measures for social scientific inquiry. To do so, most studies rely on the toolboxes of computer scientists (Salganik, 2019). However, many social scientists raise concerns about text-as-data methods, especially those based on machine learning, which are perceived as "black box" models. The development of news methods and the critical assessment of existing ones can create opportunities for developing collaborations between social scientists and computational scientists. Indeed, while tools developed in computational science are indispensable for analysing textual data, the focus of social scientists on generalisability can help to improve the use of computational methods.

Chapter 3 proposes two methodological studies which combine existing methods and critically assess their ability to identify opinions (in terms of policy frames and tonality) and stance from a social science perspective. Opinions and stance are typically addressed within opinion surveys using (batteries of) items along pre-defined dimensions. When working with social media data, there is a need to examine the relevance and robustness of the derive measures. As the research process can be drastically improved by determining pivotal points for design decisions, chapter 3 also elaborate scientific workflows with iterative feedback between machine processes and human interventions – a procedure referred to as "human-in-the-loop" (Zerilli et al., 2019). The necessary role of human intervention at multiple stages of the research process is thus highlighted. The methodological articles specifically address two major challenges: i) the need to better understand the pros and cons of computational methods for extracting opinions from social media messages with a social science perspective; ii) the need to develop a reliable method for detecting stance (or an individual's position towards a target issue or person) from social media messages. Both articles point towards an interdisciplinary framework involving social sciences, computational sciences and computational linguistics.

The first challenge is critically investigated with reference to several methods in the first methodological article (see section 3.1). Social scientists are provided with concrete recommendations of where and how to use machine learning and dictionary-based classification approaches to relate the content of social media messages to traditional survey measures. The papers provides a case study which deals with a typical suboptimal research scenario characterized by skewed categories and a setting where researchers cannot afford large samples of annotated texts. Leveraging on the case of support for democracy, the paper guide researchers navigate decisions when producing measures of tonality and frames in view of comparing them with existing survey measures. The paper uses a mix of theory-driven and data-driven steps from the data collection stage to the analytical stage. For instance, the corpus of tweets related to democracy is based on a list of relevant terms referring to the workings of democracy that are salient in Swiss newspapers and that are grounded in democracy theory. After a few iterations of tweet collection, the initial list is extended to include relevant hashtags. The update list is used to collect the final sample of tweets, which are also filtered according to time and location (e.g., explicit Swiss geographical information and likeness to be Swiss). Regarding the classification tasks, the paper compares the results from the dictionary approach to a set of machine learning models which were chosen due to their conceptually different algorithmic approaches (Rocchio, SVM and BERT). Results show that supervised machine learning algorithms outperform dictionaries for tonality classification tasks. However, custom dictionaries are useful complements of these algorithms when identifying latent democracy dimensions in social media messages, especially as the method of elaborating these dictionaries is guided by word embedding techniques and human validation.

The second methodological paper focuses on the development of a new methodology for detecting stance from social media data using a computational linguistic perspective (see section 3.2). Most studies on stance detection have so far relied on sentiment (or emotion) detection as a potential synonym for approval or agreement measured through surveys. However, while sentiment can sometimes serve as a proxy for stance, sentiment does not necessarily equate to stance (Joseph et al., 2017). Therefore, sentiment information is only used in conjunction with other text features to classify the tweets. In addition to sentiment features, it is argued that it is particularly useful to rely on computational linguistic knowledge about the role of linguistic markers in stance detection. Linguistic markers (such as modal verbs) have been found to be a good means of determining a speaker's point of view and attitudes toward a given issue (Ehret & Taboada, 2021). The proposed methodology reads as a replicable recipe involving several steps. Most notably, a custom stance dictionary is built based on the manual selection and annotation of candidate hashtags and words identified through frequency analysis and tf–idf (term frequency–inverse document frequency) measure between the different targets. Then, use is made of the dependency information of pairs of words to derive relevant signals pointing to stance in a data-driven way. Following, the analysis of variable importance and the analysis of classification errors are conducted to decide about the weights that should be assigned to the combinations of each signal to carry out the classification of stance. While positive and negative evaluative words are the clearest markers of expression of stance, the results demonstrate the added value of linguistic markers to identify the direction of the stance more precisely. The classification model achieves an average classification accuracy of 75% (ranging from 67% to 89% across targets). This study concludes by discussing practical implications and outlooks for future research, while highlighting that each target poses specific challenges to stance detection.

1.12.3. What public are on social media and how they interact with public opinion

Chapter 4 investigates what are dominant social media users and their interaction with public opinion. Learning about who are active users on social media is a prerequisite for draw valid conclusions about opinion, attitudes or behaviours. Even if social media users are not representative of the general public, investigating who has an interest in talking about given subjects and who are influential groups of users on these subjects enables researchers to better grasp the extent to which what is said on social media can reflect or influence the views of the broader public which are assessed through opinion surveys.

In practice, the identification of user profiles can be relatively straightforward when the data collection strategies depart from specific accounts. For instance, studies can focus on "seed" profiles to tack co-evolution of different agendas and audiences' concerns. In this view, several audiences can be derived from network (e.g., follower-following) metadata, such as groups of "politically interested users" (Barberá et al., 2019; Gilardi et al., 2021). However, when social media data are collected based on a key-word approach, there are relatively few studies that systematically indicate the distribution of profiles when using social media data. This might be due to the fact that users usually provide little basic information – especially in terms of profession, age, gender and location – which makes it difficult to assess whether they are representative of (sub)populations.

To address the sparsity of user profile information, techniques have been developed for inferring basic user characteristics in cases where they are not actually available. The estimation of such characteristics can then help to build weighting procedure alike to surveys. However, these techniques are still not accurate for discovering basic sociodemographic information, such as age, gender or social class (Sloan et al., 2015). As a result, the best way to account for the type of users included in a corpus of social media texts remains to assign users – either manually or (semi-) automatically – in prevalent categories (e.g., media or journalists, politicians or parties, activists, etc.), communities (e.g., followers) or bots (Keller, 2020; Rauchfleisch et al., 2021). Therefore, when the main goal is to assess the relevance of social media data to depict trends in public opinion, treating social media data as exclusively traces from individual citizens and actors may lead to misjudgement (Duan et al., 2022).

Chapter 4 proposes three empirical studies where the identification of groups of Twitter users is central. It thus covers three contexts. Taken together the three contexts under study provide an exhaustive view of which groups of users are active in the Swiss Twittosphere, namely sub-groups of the general population, political actors and actors close to politics, as well as groups involved in online discussions surrounding social movements.

The first study (see section 4.1) investigates the profile of "average" non-political social media users, and whether this profile has evolved over time in a quickly changing digital environment. It relies on panel data to map the Swiss media consumption practices between 2013 and 2016. The paper aims to investigate whether the media consumption landscape can be defined by a digital-oriented versus a paper-oriented spaces. Then, it aims to explore what individual factors can explain the formation of this new media space. The study makes use of longitudinal data from the Swiss Household Panel (SHP) and relies on a dynamic version of multiple correspondence analysis for categorical variables. The panel nature of the SHP data allows to track changes in consumption patterns. The paper thus adopts a data-mining approach to investigate media consumption changes among

SHP respondents. It shows that, at the time of the surveys, Swiss social media users were likely to be citizens who were not particularly interested in politics and who relied heavily on free or lower-quality journalism sources. Based on these findings, we can expect to find multiple “user spheres” disconnected from each other, some of which can be more dominated by media consumers and socio-political elites, while others can lie some distance away from these major centres.

The second study (see section 4.2) allows a better grasp of which publics are particularly influential on Twitter and how this is linked to the evolution of the political scene. In Switzerland, the network of Twitter users engaging in political discussion can be best characterised as an elite network (Rauchfleisch & Metag, 2016). However, albeit not representative of the general population, investigating what user groups are politically active in politicians’ network and how these groups impact on political success can inform researchers about the potential of social media discourse on the formation of public opinion. The article presents a longitudinal study focusing on the active share of the Swiss Twittosphere and shows that Twitter offers an extremely good coverage of certain social groups, such as politicians, journalists, and other actors close to politics. After conducting a manual annotation of user profiles with whom politicians interact (in terms of replies and mentions), the article displays the evolving interactions with salient user groups (e.g., journalists and media, other politicians, consultants and experts, associations and lay citizens) thus pointing to possible strategic adaptations of politicians' communication. Then, the paper investigate the extent to which Twitter-based activity impacts politicians' political success, both in terms of political ranking and media coverage. It thus goes beyond the majority of studies predicting electoral success during political campaigns with social media indicators. The proposed analysis requires merging external information to the Twitter corpus, namely politicians’ ranking based on vote share and politicians’ media coverage. The explanatory variables are grouped into communication style variables and reactions’ to politicians’ tweets (e.g., proportions of retweeted and favourited politicians’ messages), and legislature dummies. Among the communication style variables, the proportion of replies emitted by politicians indicates their level of interaction with important user groups. Furthermore, the responsiveness of parliamentarians to citizen concerns is taken into account and requires to use machine learning to classify the tweets into relevant policy issue categories derived from electoral survey data. Politicians’ information dissemination practices (e.g., proportion of links) is an additional communication style variable. Results show that Twitter-based activity moderately impacts politicians' political success. However, this success strongly depends on the style of political communication and on the legislature under scrutiny. The strengths of the study are its historical perspective and the focus on data complementarity to measure political success.

The third study (see section 4.3) conducts an investigation into which actor groups are involved in gender equality discussions online, what the prominent and polarising ideologies are, and the main ways in which the debate is framed. Investigating what are engaged users in social media discussion surrounding a major social movement is indispensable to better understand the extent to which social media platforms constitute a tool for social movements to mobilise public opinion to promote social change. As social media algorithms tend to display information that people already like or are familiar with, new ideas might not always be visible to other users, thus impeding the dispersion of alternative viewpoints to a large community. To date, however, little is known about the extent to which activist and political claims formulated on social media echo what the general public thinks about gender equality. Twitter discussions surrounding the women’s strike that took place in Switzerland in June 2019 are analysed. Our data collection is based on relevant actor groups and keywords surrounding the women’s strike. In a first step, the article shows which actor groups are involved in gender equality discussions online. To do so, a comprehensive manual coding effort is done to classify these users into relevant actor categories. Then, the study looks at the association between the online salience of politically engaged users’ gender equality discourse and the opinions of citizens surveyed about gender equality while accounting for political positioning. These two first steps allow to investigate what are the prominent and polarising ideologies and how it reflects trends found in surveys. In the last step, correspondence analysis is used to display the argumentative features surrounding gender equality issues according to social media actors and to a representative sample of citizens concerned by gender equality. Findings indicate that organizational committees and their followers were the most active, followed by political actors. A polarisation effect on social media between left and right-wing oriented actors, which is more pronounced than trends drawn from opinion surveys was observed. Moreover, social media discussions were organised along a continuum, which ranges between calling for attention and discussing concrete policy measures. The data complementarity approach and the mapping of gender equality discourses surrounding the women's strike present a perspective on political polarisation by looking at the relationship between social media opinions and those expressed in surveys to assess gender equality related concerns.

1.12.4. How are social media used to provide a new lens on well-established and under-investigated topics in social and political sciences by complementing survey data

Chapter 5 provides empirical examples of emblematic approaches in which social media research can complement survey research. Three case studies are presented and each of them is emblematic of research contexts in which either only few or plenty of survey data are available about a given research topic. When few survey data are available, an exploratory approach is preferred. However, when survey data are available, a comparative or complementary approach can be envisaged. This will depend on the research goals.

In the first empirical study (see section 5.1), social media are used to identify important dimensions of topics that are "under-investigated" by survey research, while proposing dimensions worth investigating in future survey research. The study thus provides an example of an explanatory design. The main aims of this study are to highlight topics and the framings of health technologies (HT) diffused online by actors actively involved in these discussions on Twitter. The data collection was based first on the identification of accounts where the profile description contained relevant health technology-related terms, the relevance of the identified accounts then checked manually and their most recent tweets collected. This strategy for data collection has the advantage of focusing on accounts unambiguously linked to the topic of study, thereby enabling us to extract the most important frames and topics discussed in this still new health domain. However, it also has limitations, as it tends to focus on "expert" and "interested" users rather than lay citizens discussing the topic on social media. The proposed study also emphasise the utility of unsupervised methods, such as topic modelling and word embeddings, to classify tweets into relevant categories. The first descriptive step shows that the geographical distribution of important actors correlates with the citizens' reliance on social media to seek health information derived from survey findings. Then, results from topic modelling show that geographical factors and actor groups have an impact on the choice to promote particular HT topics. For instance, the United States focuses more on risk management and private funding, whereas Europe focuses more on health literacy, practitioners, and start-ups. Furthermore, institutions focus more on indirect, global, and strategic problematics, whereas specialists are more concerned with direct and concrete problems. The last descriptive step relies on creative visualisations displaying semantic relationships along important dimensions of HT and illustrates shifts in concerns related to privacy issues before and after the COVID pandemic. The need to combine these unsupervised methods with a manual annotation of the user profiles and with innovative visualisations to improve the interpretability of the findings is demonstrated. This study thus offers a larger theoretical hook for finding an explanation beyond what is observed and moving towards an understanding of why (Kar & Dwivedi, 2020). It further points to the necessity to complement social media findings about salient topics and frames with future research paths into what are potential further survey interests.

The second empirical research (see section 5.2) adopts a comparative design. Two years of tweets by policy makers and health experts about Covid-19 in Switzerland are analysed and compared with trend found in surveys. The corpus groups tweets emitted by major actors involved in COVID-19 online discussions in Switzerland, including scientists (experts), policymakers (government officials, cantonal executives, and other parties), and representatives of mass media. A first aim of this study is to explore correlation between tweet reception by other users and public trust during a time of societal and political crisis. Albeit surveys have shown a decrease in the levels of public trust towards political authorities during the crises, the study shows that the evaluation of lay citizens towards important actors (e.g., political authorities, the media, or health experts) is not necessarily reflected into patterns of engagement with social media messages from these targets (e.g., retweets, likes, replies). Indeed, the article finds little correlation between Twitter features (e.g., other users' engagement and negativity in other users' replies) and the level of public trust from representative opinion surveys. A second aim of this study is to extract salient episodes of the pandemic, notably by including important entities, specific periods of the public debate, particularly salient topics and target groups. To do so, topic modelling is used in combination with correspondence analysis, and includes additional variables for actor types and the period of the public debate to detect salient episodes related to the pandemic on social media. Results highlight that differing roles were played by the (health) experts and political authorities in terms of both topics and influence on the specific timing of the pandemic. Then, using hierarchical clustering, we are able to demonstrate the amount of other users' attention to each of the extracted clusters, thus providing information about how a given cluster of discussion on social media reflect other users' engagement and awareness of the same cluster (e.g., using interactive features such as likes and retweets, as well as negativity in other users' replies). From a survey perspective, these insights provide information about what prevalent online debates are likely to induce public attention offline. Results therefore help to derive conclusions for communication among political authorities, health experts, and the public.

The third empirical paper (see section 5.3) adopts an explanatory design by integrating survey data as explanatory factors of social media trends. This study starts by comparing migration discourses in traditional opinion surveys and social media in a cross-country perspective among five English-speaking countries (including the United States, Britain, Ireland, New Zealand, and Australia) to investigate the potential complementarity between social media data and survey findings. The data collection involves a two-step strategy (involving the retrieval of followers from seed accounts and the identification of tweets related to migration based on a list of search queries) and a comparison of different samples of users. The first sample includes randomly selected users following central media and party accounts for each country. The second sample is composed of politically interested users who follow central politicians' accounts. Methodologically, this paper thus demonstrates the necessity to reflect on the impact of different data collection strategies on the obtained findings. To investigate the salience of social media discussions across these different samples of users, the study integrates survey indicators (e.g., items about the impact of migration on cultural, social, and economic dimensions), as well as other contextual indicators (e.g., integration policy index and migrant acceptance index, and elite polarisation index), as explanatory factors for trends found in social media discussions. Augmenting social media data with public opinion trends from survey data also enables us to demonstrate that societal factors significantly impact the salience of migration online. To account for the effect of different general and specific framing of migration, a classifier to assign tweets among the following categories (e.g., civil rights, culture & identity, economy, foreign policy, law & order, and welfare) is built. These categories have been determined theoretically and inspired from survey research. Results show that, overall, there is a good correlation between salience of and sentiment toward migration, both in surveys and on social media. The study also demonstrates that societal factors significantly impact the salience of migration online albeit there are variations depending on the sample of users. There is thus a need to further demonstrate the different incentives that motivate users to engage with the migration topic online. To do so, a close-reading of a sample of tweets, as well as the interpretation of important words related to the generic frames are conducted. The major strength of this study is the cross-country approach combined with comparative assessment of survey data and public opinions expressed on Twitter. It also contributes to research on public attitudes toward migration, including a critical assessment of the relationship between sentiments on social media and sentiments derived from survey data. As social media messages also contain many dimensions of the immigration issue, a possible strategy is to compare averaged survey items covering several dimensions of the same topic with the mean sentiment derived from social media texts. Doing so, support is found for there being an overall good correlation between salience of and sentiment toward migration in surveys and on social media.

CHAPTER 2: HOW ARE SOCIAL MEDIA DATA USED FOR THE STUDY OF PUBLIC OPINION?

2.1 A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion[4]

Introduction

This paper provides a systematic literature review of how social media data (SMD) and traditional survey data have been used complementarily to study public opinion (PO) over the last decade. As social media users represent more than half of the world's population (see Adams-Cohen, 2020) and provide continuous reactions to daily socio-political events, it is not surprising that traditional survey research has been concerned about whether such data would make surveys obsolete or whether they could be used complementarily. Addressing these questions is particularly relevant in the area of PO. Social media plays a growing role in the formation of PO as user-generated content on these platforms is increasingly deployed as representations of PO (e.g. Dubois, Gruzd & Jacobson, 2018; McGregor, 2019). In addition, politicians increasingly consider social media, especially Twitter, to be a "barometer" of PO (Jacobs & Spierings, 2019).

Despite the extensive literature about the benefits and challenges of using SMD to answer social and political questions, as well as about SMD as a possible replacement for traditional surveys, a comprehensive overview of the complementarity of both data sources remains limited. The aim of this paper is to fill this gap by providing a systematic literature review focusing on how SMD and survey data can complement each other to study PO. Inspired by the influential study of Japec et al. (2015) which elaborated on the complementarity of survey data and "big data" (rather broadly defined), we want to concentrate, however, on one type of "big data", namely SMD. There are two main reasons for this choice. First, SMD are a specific type of "non-survey" data which possess specific arrangements (or conventions) and paradata that are different from other types of

[4] This chapter is a slightly adapted version of the article that has been published as Reveilhac, M. Steinmetz, S. and D. Morselli (2022): "A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion", Multimedia Tools and Applications, 81, 10107–10142.

administrative or "big data", especially when it come to the assessment of PO. Second, whereas there is substantial research on augmenting survey data with administrative (e.g. electricity or water consumption) or other type of "web data" (e.g. Google searches or citation metrics) to improve estimates of PO or official statistics, we still lack an overarching picture of the (new) developments and approaches of complementing SMD and surveys with each other.

Our analysis is based on an extensive survey of the literature capturing a representative sample of the best published theoretical and empirical scientific papers on the topic (N = 187). We have restricted the analytical period to the last decade (2010–2020) as the discussion on complementarity is still a young field of study (e.g. Moy & Murphy, 2016). On this basis, we have been able to identify six complementarity approaches which can be synthesised to four major purposes, namely predicting, substituting, comparing, and linking SMD and survey data.

In the next section, we situate our review within the existing literature by demonstrating how the scientific discussion surrounding the opportunities and challenges offered by SMD within survey research has evolved, especially by highlighting the complementary understanding of PO offered by both data sources. Then, we discuss more specifically which research approaches have emerged, and we classify them according to four main research purposes using both data sources complementarily. The analysis of the empirical studies aims to act as a guide for other researchers by identifying research gaps and highlighting the pros and cons of each approach. Furthermore, we underline areas for future improvements and point to technical and ethical considerations. We conclude by mentioning the main contributions and limitations of our review.

Background – The complementary understandings of PO

Surveys have long been the most predictive and accurate tools for collecting and measuring opinion. However, over the last decade, decreasing response rates have called into question the potential of using a random sample of individuals to represent an entire population (e.g. Groves, 2006; Keeter et al., 2007), thus posing important concerns about the sustainability of survey research. Even by adapting to new modes, such as push-to-web, to increase response rates, it remains unclear whether surveys will maintain this dominant role as communication habits continue to change (e.g. Schober et al., 2016).

Given the recent "survey crisis" (e.g. Brick & Williams, 2013; De Heer & De Leeuw, 2002), an increasingly rich source of PO data is commonly referred to as "big data". These "new" data take the form of extraordinarily large and complex datasets. There are three attributes that are generally agreed upon to describe this type of data (e.g. Dass et al, 2012), namely volume, velocity, and variety. Social media are a sub-type of big data where people express their thoughts and opinions with the purpose of sharing them with others (Couper, 2013). Due to their inherent properties, SMD have been seen as a promising complementary, and even alternative, source of data for exploring PO. However, researchers acknowledged early on that, almost universally, SMD are non-random, and thus discouraged using them as a means of making generalisable claims. This challenge is well highlighted by Schober et al. (2016), who claim that, while the social media researcher seeks to achieve topic coverage, the survey researcher emphasises population coverage as a central endeavour.

An entire strand of research thus focussed on how surveys and social media differ in several aspects. Table 2.1.1 attempts to classify the most prominent differences along which SMD and survey data are typically compared. We have identified several dimensions based on recurring criteria mentioned in the literature concerning the nature of and the relationship between both data sources. Often-cited criteria include the type of population and data signal, the unit of observation and analysis, and the available meta-data (for a thorough discussion of the differences see Couper, 2013, Schober et al., 2016, and Tefekci, 2014).

Table 2.1.1: Differences between SMD and survey data to study PO

	SM	Survey
Type of population	Selective	Representative
Data signals	Platform users and their signals (e.g. posts or tweets, #hashtags, @mentions, retweets, replies)	Opinion survey of individuals with a defined sampling frame
Types of data	Unstructured texts containing opinions and opinion strength or merely information (e.g. links)	Structured opinions measured by answers to pre-defined and pre-tested survey questions (scales)
Unit of observation or the level at which data are collected	Can be any of the following: users, search queries or keywords, #hashtags or @mentions, retweets, or replies, likes or emotional reactions, location	Individuals from a sampling frame representing a target population
Unit of analysis or level at which the data are analysed	Can be any of the following: users, location, texts from a specific topic or sentiment, overall texts, links, or other metadata	Individuals' responses to survey items or aggregated responses at the country, region or household level
Meta-data	Set of users' behavioural information (e.g. network, frequency of use, interactions) and contextual information (e.g. time and location)	Precise and quasi-complete socio-demographic information on individuals and auxiliary data (e.g. number of contact attempts, number of persons in the household)

To understand how to best use both data sources complementarily, it is also essential to reflect on how they construct PO differently. This is increasingly important, as what constitutes "the public" tends to be forged by the methods and data from which it is derived McGregor (2019). In survey research, PO is equivalent to the private opinion of a representative public, operationalised as a set of positions on a given topic. PO can thus be conceptualised as a reflection of a shared position among citizens on specific issues that are then amplified and reviewed by news media and political actors (Herbst, 1998). Survey measures of PO are constrained by the scope of the questionnaires, which usually provide little room for spontaneous expressions of opinion (except in open-ended survey questions). The diversity of opinions is thereby reduced into a set of discrete and aggregate data (e.g. Tourangeau & Galešić, 2008). Conversely, the reliance on social media for measuring PO expands the societal and collective components of opinions (Murphy et al., 2011) by conceptualising it in Habermas' (1991) terms as a complex system of representations. In this respect, SMD are better suited to capturing the conversational and relational nature of PO formation (Anstead & O'Loughlin, 2015). Hence, where survey data weigh precision and standardisation, SMD excel in multidimensionality and polyphony. In addition to their focus on solicited private opinions, surveys are also less reactive to opinion changes than SMD. In theory, opinion changes could be assessed by frequent short opinion surveys (e.g. every two months). However, the advantage of SMD is that they can cover opinion change more rapidly (and on an ad hoc basis), thus reacting faster to events, which is almost impossible for surveys (e.g. it takes more time to set up probability-based surveys for the study of COVID compared to what can be done with SMD).

Despite the advantages offered by social media for measuring more social and timelier opinions, the reliance on SMD raises important questions for empirical research on (automated) measurements of opinions and on the choice of the indicators employed to model opinions. Indeed, constructing measures of PO based on SMD can be very time consuming and can involve a lot of pre-processing effort before the data can be translated into meaningful measures of expressed opinions. Furthermore, it sometimes remains quite difficult to know what is driving the evolution of ideas and concerns found in online conversations. Consequently, a current strand of research seeks to better understand the issues of representativeness of social media communities and the validity of measured opinion, especially opinions stemming from sentiment analysis. While there is a rising interest in applying SMD to understand opinion, and even to replace traditional surveys (e.g. Anstead & O'Loughlin, 2015; Gayo-Avello et al., 2011), SMD alone are of limited use for social scientific research as they usually provide incomplete and imprecise information. However, the issues associated with SMD are not necessarily fatal to the proposition that they can be used to generate social insights, especially in complementing survey data. An efficient strategy to enhance research lies, therefore, in the analysis of how both data sources can complement each other in ways that maximise their strengths. In the next sections, we aim to show that there is a plethora of research practices in which both data sources complement each other for the study of PO. To date, however, there is still no consensus about the best way to use SMD for studying PO (Moy & Murphy, 2016). We are now at a point where we should reflect on what has been done so far, what lessons we can learn from it, and then specify suitable trends for social research. In this paper, we seek to fill this gap by reviewing research that uses both data sources complementarily for the purposes of measuring PO and by providing a critical evaluation of the identified research paths.

Method of analysis: Building a corpus of relevant articles

To build our corpus of scientific articles, we carried out several searches in bibliographic databases (focusing on Scopus and Google Scholar) using the software PublishOrPerish

(Harzing, 2007). We obtained an initial corpus of 3596 unique papers, which we reduced to papers that were relevant for the scope of our review. The initial corpus was deliberately based on a search-query that was broad enough to collect the relevant literature, while not missing important papers. We used the query "(social media OR twitter OR facebook OR instagram OR reddit) AND (survey OR surveys OR polls)" and specified that it should appear in the body of the text (using the keyword field) instead of appearing only in the title or abstract, which were found to be too restrictive to capture the literature of interest. The query was designed to restrict the focus of our review to SMD, thus ignoring other types of "big data" or "digital trace" data.
A first filter was applied to reduce the number of papers to journal articles, book chapters, and scientific reports (thus excluding books, theses, and conference papers) as we wanted to concentrate on high-valued scientific sources which have already been approved by the scientific community. In this respect, including conference papers would have drastically inflated the number of (duplicated) papers concerned with predictions and with replicating previous studies using alternative methods of analysis and algorithms. Among the remaining papers, we applied two eligibility criteria to disregard those that were not pertinent to the analysis as i) their focus was not on PO, ii) they were oriented towards a specific aspect of data treatment (e.g. estimating socio-demographics from texts or profile pictures) or an analytical strategy (e.g. elaborating algorithms). We also excluded articles mentioning survey findings without an explicit aim of supplementing, comparing, or combining those with SMD.

Results of the literature review on the uses of social media as a complement to surveys

Overall, the collection protocol left us with 187 papers - 141 of an empirical and 46 of a theoretical nature (these papers can be found in the Appendix). Most of these papers stem from political communication and computational social sciences journals. Although the sample of 187 papers may not cover the whole corpus of research on the subject, it is nonetheless sufficient to highlight the main research directions that have been endorsed on the topic of complementarity. Figure 2.1.1 provides an overview of the yearly repartition of the retrieved papers differentiating between those with a theoretical (N =

46) and an empirical (N = 141) focus. While the number of theoretical papers remains stable over the years, we can see a steady increase in empirical papers over time.

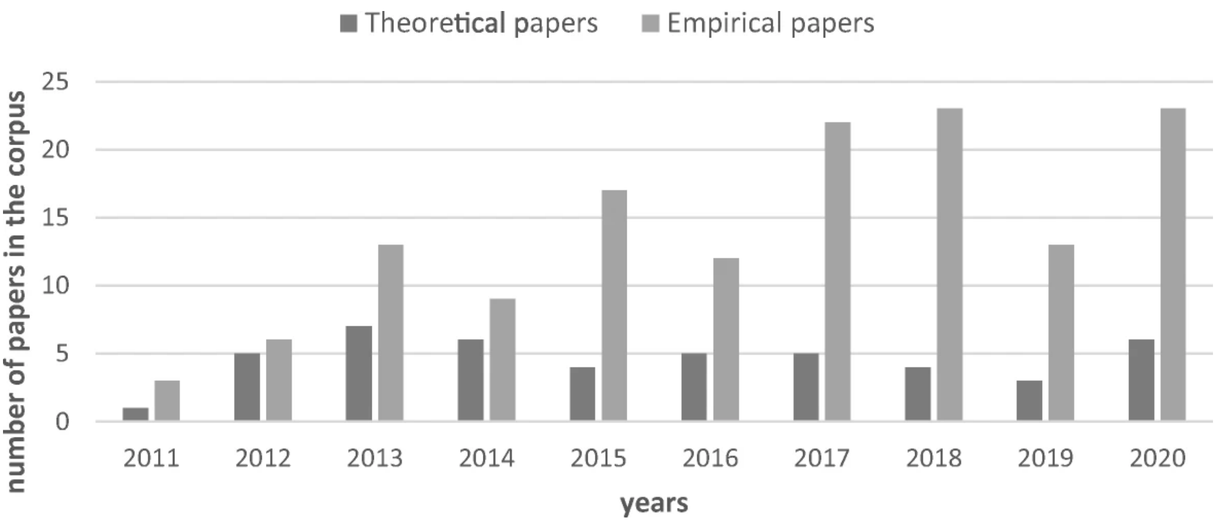


Figure 2.1.1: Number of empirical and theoretical articles according to our meta-review of the existing literature using surveys and SMD

Theoretical insights

Starting with the theoretical papers in our review (N = 46, see Table 2.1.2 in the Appendix), survey and social media researchers have explored ways in which social media and survey data can yield congruent conclusions (e.g. Schober et al., 2016). One part of these articles (n = 14) tries to establish a framework regarding the predictive power of SMD as a potential substitute for surveys. This line of research stems principally from the fields of election and economy forecasting (for recent reviews see Chauhan et al, 2020; Rousidis et al., 2020).

Another strand of theoretical articles (n = 14) focuses instead on the compliance of social media research with established reporting standards so as to guarantee transparency and replicability (e.g. Klašnja et al., 2018). Finding ways of integrating data obtained from different sources (n = 3) also constitutes a fertile path of research (Johnson & Smith, 2017). In this respect, Stier et al. (2019) provide the most advanced guide on how to systematically link survey data with information from external data sources, including SMD, at different level of analysis. The authors demonstrate that integrating traditional survey data and digital trace data is of growing interest, notably because of the limited reliability of self-reported behavioural measures and declining response rates. Additionally, enriching survey data with SMD could also help to reduce unit non-response and to control for the unrepresentativeness of SMD, as they are limited to those respondents having social media profiles and consenting to the linkage. Finally, a smaller share of research (n = 5) focuses on developing a quality assessment framework for SMD which is similar to the Total Survey Error (TSE) (Biemer & Christ, 2008; Groves & Lynerg, 2010). The TSE framework has been extended to encompass SMD and their inherent quality challenges (see the studies by Sen et al. (2019) on Twitter-based studies and Jungherr (2016) for a measurement theory to account for the pitfalls of digital traces). In a similar vein, Hsieh and Murphy (2017) analysed the potential benefits of evaluating estimates from surveys and SMD in common terms and arrived at a general error framework for Twitter opinion research. Olteanu et al. (2016) went a step further by pointing to the errors and biases that could potentially affect studies based on digital behavioural data, outlining them in an idealised study framework. The paper by Sen et al. (2019) provides the most advanced framework to date. It involves potential measurement and representation errors in a digital trace-based study lifecycle where they are classified according to their sources.

Other research (n = 5) tackles the ontology of SMD as compared to survey data. In these papers, prevalent discussions revolve around the conception of opinion as measured by both data sources, as well as debates related to the evolution of “new” research “paradigms” or “digital hermeneutics”. The remaining papers concentrate on behavioural research (n = 2), demographic research (n = 2), and small data analysis in political communication (n = 1).

Overall, the considered theoretical articles stress the importance of developing a framework that accounts for possible biases of SMD while remaining in, or mirroring, the TSE. Moreover, they also emphasize the need, in this debate, to focus on the complementarity rather than the replacing aspect, notably by developing clear and reliable linking strategies. These articles also encourage researchers to go beyond the dominant model for understanding PO from probability sample surveys to encompass other (“new”) expressions of opinions (e.g. Murphy et al. 2014) that can possibly supplement or even replace survey-based approaches.

Empirical insights

The empirical literature (N = 141) focuses on a rather narrow set of topics, such as elections, political issues, and approval ratings for the presidency (64%). Another important area of PO research using SMD complementarily with survey data is related to health (e.g. vaccination, drugs, etc.), equality issues, and climate or environment-related concerns. Most empirical studies in our review are based on Twitter data (73%), followed by Facebook (18%) and other social media (9%). This is related to the fact that not all social media platforms provide the same degree of data accessibility [8]. For instance, Facebook imposes severe limitations on the scope of retrievable data, whereas Twitter has less strong privacy settings, allowing researchers to get access to Twitter's historical data.

Overall, we derived six major approaches on how survey data and SMD can complement each other namely i) predicting social and political outcomes using SMD (n = 48), ii) comparing both data sources on a given phenomenon (n = 26), iii) using survey measures as a proxy in social media research (n = 18), iv) enriching surveys with SMD (n = 9), v) recruiting individuals on social media to conduct a second survey phase (n = 8), and vi) generating new insight on "old" or "under-investigated" topics or theories using SMD (n = 32). These approaches can be synthesised in four, partly overlapping, 'data complementing' research purposes: i) validating survey findings with SMD, ii) improving the sustainability of the research by diversifying the views on a phenomenon, iii) improving the reliability of survey measures by specifying measurements, and iv) improving the interpretability of social or political issues. Figure 2.1.2 summarises the relationship between the six approaches and the four research purposes. Furthermore, it shows that each purpose leads to a typical way of using both data sources complementarily. For instance, improving reliability by specifying a research question involves data linkage strategies, while generating new insights involves a sequential use of social media and survey stages.

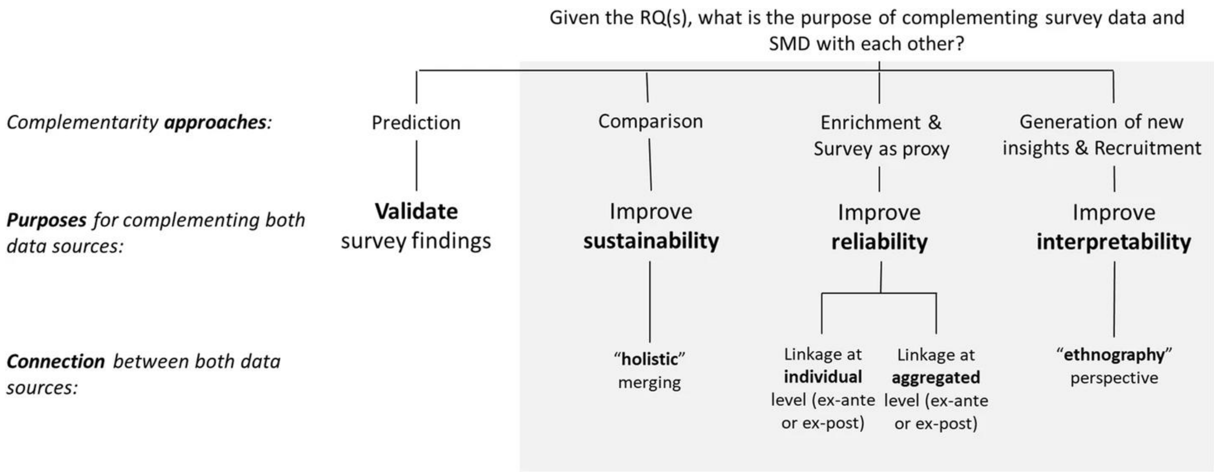


Figure 2.1.2: Complementary approaches using SMD and survey data for the study of PO

The analysis of our corpus suggests that the biggest part of research concentrates on whether SMD can potentially substitute survey data (n = 48, see Table 2.1.3 in the Appendix). This has mostly been done by trying to replicate survey findings by using SMD for forecasting (see recent review by Rousidis et al., 2020). The aim to predict real-world outcomes with SMD in the realm of PO has essentially been applied to elections. Most of these papers directly refer to the much-cited study of O'Connor et al. (2010) which purpose is to validate SMD against survey findings. While research in this area has tested a range of different methodologies, the results remain inconclusive, and only in some cases could elections be accurately predicted (e.g. Gayo-Avello, 2011; Jungherr, 2016). Recent literature reviews on the use of SMD for running electoral predictions (e.g. Chauhan et al., 2020) classify studies according to the employed methods of prediction, such as volume, sentiment, or network approaches. These reviews show considerable variance in the accuracy of predictions, which, on average, lag behind the established survey measurements. A common problem of the aforementioned studies lies in the decision about which approach can most accurately yield predictions (but also which social media platforms are better suited, and how that varies in different geographical or temporal contexts). This inference problem is quite complex as various elements are involved in skewing the samples in social media debates. To date, the inconclusive state of the research has led to a research agenda aiming to respond to the plea from Gayo-Avello et al. (2011) for a "model explaining the predictive power of social media" (p. 490). In this realm, for instance, the study of Pasek et al. (2019) assesses how patterns of approval among population subgroups compare to tweets about the president, while disentangling effects at the individual and group levels of analysis. On a more theoretical level, the study by Schober et al. (2020) seeks to elaborate when and under what conditions SMD can be used to make valid inferences. However, the inconclusive state of the research may also be linked to the fact that predictions are often done based on the content created by users and overlook the characteristics of the creating users. For instance, SMD can be biased towards a particular group (see Bakshy et al., 2015; Del Vivario et al, 2016). Moreover, interactions on social media platforms are not always the product of individuals, but also bots, organisations, political parties, etc. (Varol et al., 2017). Based on the evaluation of the body of articles falling under the 'substitution paradigm', a path for future research could be to better account for the characteristics of social media users, insofar as these characteristics can be useful for assessing how individual tweets can be converted into meaningful measures of expressed opinion. To do so, future studies could survey social media users identified using relevant key terms (e.g. hashtags or mentions) to gauge the relationship between social media measures of their sentiment and survey measures of their attitudes.

The second dominant approach in our review is related to how surveys can be enriched with SMD (n = 9, see Table 2.1.4 in the Appendix). Here, SMD are collected with the intention of improving the reliability of survey measures at the individual or aggregate level. Replication of survey-based opinions can be difficult, either because of improper interpretation of the findings or because insufficient information has been provided. Such issues undermine the credibility of survey research and make it difficult to evaluate the contributions of a given study. Research aiming to enrich surveys with SMD most often implies the adoption of a data-linking strategy. This can be done, for instance, either at the user level, public actor level, geographic level, or temporal level (see Stier et al., 2019). Enriching surveys with SMD can serve several goals. First, it can help to augment the explanatory potential of survey measures. For instance, De Sio & Weber (2020) adopted an innovative research design to explain election outcomes based on party strategy on social media with respect to policy issue salience. They did this by linking representative mass surveys from six European countries with Twitter analysis of campaign activity. Second, enrichment of survey data with SMD can also help to test research hypotheses by relying on "true" behavioural measures (instead of self-reported survey measures). For instance, Karlsen and Enjolras (2016) linked candidate survey data with Twitter data to study styles of social media campaigning. These differences in campaigning styles were then related to the extent to which candidates were successful on Twitter. Third, SMD also offer an opportunity to address issues of item non-response and calibration of novel measures. For instance, Shin (2020) studied the extent to which social media users selectively consumed like-minded news stories by linking survey responses from Twitter users with their media following and exposure to news via their friends. The study further showed some differences between self-reports and digital measures, such as more pronounced patterns of selective exposure in the SMD. Finally, linking social survey and SMD further provides an opportunity to explore the relationship between attitudes and beliefs reported through surveys and content (and behaviours) generated online. For instance, Cardenal et al. (2019) combined survey and Web-tracking data to analyse how Facebook-referred news consumption influenced social media users' agendas. They found that selective exposure increased with amplified news consumption. The core problem in these studies lies in gaining consent to carry out the data linkage. This constitutes a complex procedure in which issues of anonymity, security, and disclosure all come to the fore. An additional problem is that social media measurements provide only one partial view of opinions. For instance, while researchers can measure how many times a given message has been liked, shared, or retweeted, it is much harder to account for (or measure) how often a given message has been seen or has attracted attention. Moreover, our corpus shows that research relying on linking strategies tends to remain at the individual and public actor levels of analysis, which requires requesting consent to use the linked data. This may, in turn, introduce consent or selection bias. To mitigate such difficulties, future studies should also explore the potentials of linking both data sources at higher levels of analysis, such as country or according to topicality level.

A third purpose is to use surveys as a proxy in social media research. This approach therefore reverses the logic that SMD are always used as a complementary (side) element of the main survey-based analyses. In this kind of "survey proxy approach" (n = 18, see Table 2.1.5 in the Appendix), SMD are used as the main source of analysis, while the survey data are used for contextualising or calibrating SMD. A first strand of research relies on SMD to complement traditional research approaches in political communication and citizens' political engagement. For instance, the assessment of the importance of given public concerns in PO has been measured extensively with the "most important problem" survey item. Social media provide another way to measure this concern in an unintrusive way by (semi-)automatically classifying the content of social media texts, while also accounting for the extent to which different actors are responsive to these concerns.

Following this logic, the study conducted by Eberl et al. (2020) investigated the effects of sentiment and issue salience on emotionally labelled responses to posts written by political actors on Facebook. Another study, by Plescia et al. (2019), analysed the responsiveness of populist parties to the issue salience amongst the public. They did this by relying on survey data to measure public salience and tweets to assess salience issue for parties. A second strand of studies aims at facilitating cross-national comparisons. For instance, a possible application consists in using survey data for classifying parties and voters along important dimensions (e.g. see Ernst et al., 2017). Here, parties were placed on a left and right spectrum using the Chapel Hill Expert Survey (Bakker et al., 2021). Party score on the overall ideological stance was then used as an explanatory variable in subsequent analysis. Another example is the study by Park et al. (2017) which investigated the consumption of popular YouTube videos in countries that differ in cultural values, language, gross domestic product, and Internet penetration rate. A possible issue encountered by these studies is linked to spurious effects between survey and social media measurements (e.g. misleading or unexplained correlations). Furthermore, these studies tend to remain poorly equipped to explain actual motives behind social media users' expression of opinions or reactions. The "survey as proxy" approach requires a considerable dose of ingenuity and methodological innovation to mine social media for producing opinion estimates that can be merged with survey estimates. For instance, SMD corpora often deviate from a predefined (survey) coding scheme. Substantively, a future path of research should take advantage of the fact that a growing number of societal issues have become transnational, such as immigration, terrorism, women's rights, and climate change. Such research could involve the combination of word embeddings and survey opinion measures at the country level.

A fourth approach aims to compare SMD with survey responses that directly measure PO. Studies comparing SMD with survey data (n = 26, see Table 2.1.6 in Appendix) essentially aim at improving sustainability of the research, which consists in the ability to gauge PO consistently over time. Sustainability thus implies that we should develop designs that include opportunities for "holistic merging" of the data that will generate more inclusive and fine-grained research insights. There are several reasons that comparing both data sources is meaningful for social research. Firstly, comparing SMD and survey data can be very useful in times of protests and collective actions, notably due to the difficulty of generating survey data to properly assess these disruptive changes (see critique of survey data by Lee (2002)). The timing of an event might indeed not coincide with the timing of a survey, which is often done ex-post. For instance, Davis et al. (2017) examined the extent to which tweets about the affordable care act ("Obamacare") could be used to measure PO over time. Secondly, social media can be compared to surveys for research questions that require chronicity, on a weekly or daily basis, thus going beyond the few ongoing surveys that collect data monthly or yearly. For instance, Diaz et al. (2016) demonstrated how social media activity functions like an "opt-in panel" where users repeatedly discuss the same topics. This allows us to study, longitudinally, quite rapid shifts in individual opinions and behaviours, thus complementing survey panels which are prohibitively expensive. Another example is the study by Loureiro & Alló (2020), which aimed to complement surveys by providing up-to-date measurements about social concerns when debating mitigation and energy transition paths. Thirdly, survey questions are often designed to capture internal attitudes toward a specific object. However, the relevance of certain survey questions might vary over time and, in some cases, might no longer correspond to the issues discussed spontaneously online. For instance, at a geographical level, the study by Scarborough (2018) compared gender equality attitudes found in survey data to sentiments emanating from tweets. Fourth, SMD can produce quicker and less expensive statistics for enabling informed policy and program decisions. However, this requires gaining knowledge of where any possible disparities in attitude distributions between SMD and survey data may lie. In this respect, the study by Amaya et al. (2020) presented recent advancements. The authors compared attitude distributions between Reddit users and survey measures of political leaning, political interest, and policy issues. They showed that Reddit users tend to have more centrist and normally distributed scores than the survey data, skewing estimates toward the conservative end of the spectrum on all attitude measures. Another study, from Pasek et al. (2020), explained that SMD might be better conceived as providing insights about public attention rather than ("survey like") attitudes or opinions. To do so, the authors compared tweets mentioning the presidential candidates and open-ended survey questions about the candidates to assess whether spikes surrounding political events correlate between both data sources. Results display some support for the correlation between social media attention and survey data, but they also show systematic differences that need to be better understood to assess when SMD can best generate insights about select topics. The research comparing both data sources tends to remain focused on volume analysis and tonality assessment. This type of research also tends to pay little attention to the domain-specificity of the SMD collected as well as to ways of mitigating replicability and consistency issues (e.g. González-Bailónet al., 2014). For instance, the evolution of search queries around a given theme might lack precision and consistency over time. The connotation of hashtags can change or whole hashtags can even disappear. Better combining both data sources also requires elaborating more sophisticated measures of opinion and attitudes. One could think about pushing forward "stance detection" in complement to "sentiment detection", but also about advancing "narrative analysis" in complement to "topic or frame detection". These are avenues where computational social research would benefit from the expertise of applied computational linguistics.

A fifth approach implicates using SMD to generate new insights. This is especially useful when survey data are not available or when survey data are not recent enough (n = 32, see Table 2.1.7 in Appendix). Here, the main purpose is to improve the interpretability of the research by adopting an "ethnographic" methodology. By avoiding rigid research design plans, SMD can remain responsive to, and pursue, new paths of discovery as they emerge. Based on the papers collected, we found typical reasons for relying on SMD to generate new insights, such as capturing emergent opinions, expanding the scope of survey measures, validating survey measures, proposing novel approaches to get a more nuanced or dynamic perspective on PO, and making causal analyses (see column "Reason to complement" in Table 2.1.6 in Appendix). When used for capturing emergent opinions, SMD allow us to study the topical and normative climate around specific issues for which we have no theoretically grounded ideas yet. In this exploratory design, social media can provide survey researchers with a snapshot of important societal and political concerns worth surveying in future research. This is especially useful for emerging topics, such as nuclear power (e.g. Kin & Kim, 2014) or health-related policies (Robillard et al., 2013; Thompson et al., 2015). On these emerging issues, SMD can be used in an exploratory or ethnographic perspective to generate initial and qualitative insights into under-studied research objects in order to develop quantitative survey measurements. SMD can also be useful for expanding the scope of survey measures on topics that are difficult to survey. For instance, Hatipoğlu et al. (2019) used SMD to study international relationships with a case study on Turkish sentiments towards Syrian refugees using Twitter. Another study by Guan et al. (2020) relied on the social media platform Weibo to study Chinese views of the United States. SMD can also be useful for validating survey measures. For instance, the study by Dahlberg et al. (2020) investigated the meanings of democracy in a cross-country perspective to better understand differences in the usage of the term "democracy" across languages and countries. The authors' findings aimed to inform survey measurements about the different conceptualisations of democracy, notably by highlighting translations and language equivalence issues in survey items. Another reason is to propose novel approaches for achieving a more nuanced or dynamic perspective on PO. For instance, researchers can add new components and improve "old findings", which are difficult to measure with survey data. In this view, the study by Barberá et al. (2019) modelled policy issue responsiveness using Twitter data, thus going beyond the more static perspective on issue congruence offered by surveys. In another study, Clark et al. (2018) investigated organisational legitimacy in a case study about public reactions on social media to the Supreme Court's same-sex marriage cases. The authors argued that SMD can lessen some of the limitations of survey research in the field, notably by accessing not just policy positioning among individuals but also a variety of features of political discourse, such as opinion intensity and emotions like anger or happiness. SMD can also be used to make causal inferences in order to understand changes in opinion before and after an event, such as measuring the effect of a promulgated law on PO (Adams-Cohen, 2020). Here, SMD allow researchers to rely on spontaneous opinions expressed online rather than on retrospective survey questions, and this can help develop policy initiatives. For instance, Tavoschi et al. (2020) used Twitter as a "sentinel system" to assess the orientation of PO in relation to vaccination. Despite the advantages of SMD in providing new research insights, these studies tend to lack a rigorous contextualisation of the findings derived from SMD. In this respect, a reliance on SMD would benefit from implementing sequential designs, where social media help to identify specific populations or sub-topics, which could then lead to a second quantitative survey phase. Whenever possible, SMD would further benefit from a comparison with longitudinal surveys to assess the extent to which both data sources reveal similar dynamics of change. Future studies could further exploit SMD's ability to generate new insights for research in sensitive fields, such as war, racism, sexual orientation, and religious beliefs. These are often topics on which it remains difficult to collect survey data, notably because of the social desirability bias (e.g. Kreuter et al., 2008) and the like (e.g. extreme response style, moderacy bias, and acquiescence), but also because of the fear of being denounced or because the topic is controversial.

The last approach using SMD and survey data complementarily focuses on using social media to recruit survey respondents. However, in comparison with the previous approach, the studies collected here usually analyse SMD and survey data in sequential phases. As we only consider papers that are in some way also related to PO and are not solely about recruitment of survey respondents and their socio-demographic characteristics, the number of studies we were able to analyse is much smaller (n = 8, see Table 2.1.8 in the Appendix). Our review demonstrates that the papers essentially tackle the problem surveys have in recruiting specific politically involved sub-groups of the population. In particular, the research relies on social media to access representative samples of social media users, for instance, those who commented on their countries' elections (see Bekafigo & McBride, 2013; Bode & Dalrymple, 2016) or who posted at least one election-related tweet Vaccari et al., 2016). Furthermore, in these studies, ethical concerns (e.g. privacy, tracking, etc.), but also the technical affordability of the social media platform used, are discussed. The latter issue is important, as each social media platform has particular arrangements which are likely to influence the group of individuals that can be reached. Overall, future studies could think about extending the recruitment approach to enhance our knowledge of reactions to systematic events, topics, or other repetitive features (such as supporting an issue or taking part in actions), while eliminating recall errors. Furthermore, relying on SMD can help researchers pre-test their hypotheses for future surveys by uncovering relevant underlying discursive patterns or by making smaller-scale qualitative observations.

Summary and concluding remarks

The aim of this article was to provide a review of published papers on the complementarity of SMD and survey data for PO research. We started this review by situating our work within theoretical advances concerning the complementarity of both data source. There has been extensive work underlying the opportunities and (quality) challenges of SMD for answering social research questions. However, research attention has only recently turned to SMD as a source of expression of PO and of its measurement. Consequently, there is a need for more research to uncover the ways in which SMD can be best used for fostering the understanding of PO.

The main contribution of our review is to provide a complete picture of the empirical research on the topic while calling attention to the pros and cons of each approach and possible future paths of advancements. Though this review might not be exhaustive, it has enabled us to show six major complementarity approaches which were identified as responding to four different research purposes. Below we highlight the main research paths for each approach. Using both data sources complementarily for prediction purposes was by far the most prominent approach and it remains a research area which raises many questions about the potential generalisability of the findings, namely in terms of the representativeness and validity of social media measurements of PO. We believe that the most important difficulty lies perhaps in the manner in which these studies deduce political opinions or attitudes from SMD. Survey researchers readily admit that opinions are more difficult to measure than behaviour because they involve what people think and not just how they act. Thereby, the choice to rely on sentiment analysis or merely on volume metrics (such as the number of retweets or mentions) seems unclear, at least for the near future.

Approaches concerned with improving sustainability have a significant potential for advancing social research, as they allow researchers to combine the richness of SMD content with established survey measures. When SMD are used in similar contexts to survey data, we believe that a critical view should prevail, informed by current social science best practices and expertise. For instance, whereas surveys draw a sample of carefully worded and standardised questions, social media can cover many topics as well as different facets of the same topic, which are not necessarily defined a priori on a theoretical basis. This research avenue is most likely to be fruitful for studies aiming to augment surveys by mapping discussions that are topical on social media, while allowing variations at country or regional levels of analysis to be discerned (e.g. Bennett et al. (2021) on climate change opinions). Studies aiming to compare both data sources are certainly the most suitable to help improve our understanding about when and how both data sources can be validly combined. Survey methodologists can play a decisive role, notably by paying attention to the type of (open-ended) questions that can be more directly comparable with SMD. This direction can also inform the lack of consistent evidence for the first prediction approach.

Alternatively, studies aiming to improve reliability see research as mostly requiring control for the still severe limitations of using SMD appropriately in a PO context. In this respect, studies enriching survey data with SMD offer a solution to the fact that social media often lack relevant individual information, such as respondent's attributes (e.g. sociodemographic characteristics or personality traits) or key outcome variables (e.g. voting, social, or political attitudes). Additionally, the "survey as proxy" approach enables researchers to calibrate SMD according to standardized survey measures at the actor (e.g. political candidates or parties) or context levels by reversing the data linking strategy. Future paths for both approaches implicate opening up the analysis to non-individual levels.

Studies aiming to improve the interpretability of survey research by generating new insights or by recruiting respondents on social media for a second survey phase, and that use both data sources complementarily, offer additional fertile ways to consider for new analyses that would not be possible using survey data alone. In this view, SMD do not aim to replace opinion surveys, but aim to provide a broader context for interpreting opinion, which will then serve to improve the quality of survey questions. This research avenue is most likely to be useful for knowing more about hard-to-reach populations (e.g. the LGBTQI* or disabled persons communities) or topics that are difficult to survey (e.g. violence and racism), especially when conducting iterative phases of analysis. It is also useful to get "opinion climates" about topics which have long been under survey scrutiny (e.g. emerging concerns related to feminism or social inclusion) in order to develop "updated" survey measurements.

Bringing together the opportunities offered by these different approaches shows that samples of social media users do not necessarily have to be representative of the general public to be used meaningfully as a complement to surveys. Most importantly, we believe that SMD should supplement, but not replace, traditional methods and data sources in the study of PO. By keeping up with current developments, we believe that remaining in the framework of survey research when using both data sources complementarily is paramount for identifying potential non-survey data sources, accessing them, and assessing their quality and usefulness for the study of PO. Like mixed-method approaches combining qualitative and quantitative data (e.g. Greene et al., 1989), the primary motive for complementing survey and SMD with one another is to allow researchers to mix datasets in a meaningful way for developing an overall interpretation.

Technical and ethical note

Regarding sustainability, it is important to consider that the patterns of social media consumption are influenced not only by user preferences, but also by technological changes and the availability of the platforms. For instance, social media companies may not survive and whole platforms could disappear, thus impeding data access. With changes in consumption patterns, PO may be difficult to measure consistently over time. From a more technical perspective, it is also important to assess the extent to which databases composed of social media texts collected by different means (e.g. different search queries or different platform algorithms) might raise consistency and replicability issues (e.g. González-Bailón et al., 2014). As for reliability, several issues are worth considering. Even though SMD can provide complementary information to survey estimates though linkage, there are sometimes concerns about the veracity or honesty of the information collected. For instance, SMD may increase the potential for social stigmatisation, causing users to be more reluctant to share their true opinions (Pavalanathan & De Choudhury, 2015). However, the opposite may also be true: users could express more radical opinions to gain social approval (e.g. disinhibition effect). The identity of those who post can also raise veracity concerns (Lukoianova & Rubin, 2013), and it may be difficult to distinguish sarcastic content from texts that are straight-forwardly positive or negative (e.g. González-Ibánez et al., 2011). Another important issue is that we usually know how many people have liked a post, clicked on a link, or retweeted a message, but we rarely know how many people have seen the item and chosen not to take any action (Tufekci, 2014). Furthermore, due to algorithms that favour selective exposure and homophily of opinion (Barberá et al., 2015; Conover et al., 2011), it is important to assess the extent to which findings derived from online opinion generate more polarised opinions than the ones that would be obtained through the private setting of surveys.

When researchers aim to generate new insights, they should consider that each social media platform has particular arrangements. For instance, the orientation of the content (e.g. political, family-oriented, business-oriented) as well as the scope of the content (e.g. possible bias toward more visible events) can play a decisive role on what content is available and which user profiles are most likely to be active on the social media platform. Furthermore, the nature of the platform allows for different levels of engagement in debates (e.g. Twitter is mostly used for short text content, while YouTube and Instagram allow sharing and commenting on videos and pictures). Functional capabilities can not only influence the ways of recruiting respondents for a second survey phase (e.g. direct messages), but also the identifying of sub-groups of interest (e.g. differences between friend and follower networks, and the reciprocity of follower networks). In addition, social media platforms may give users control over the availability of the information (e.g. to suppress or filter unwanted comments), which will again impact what is available from whom and on what.

For each research purpose, we should also consider that there are important ethical factors that are likely to influence the possible paths of research relying on SMD. Each platform has its own rules which are subject to change at any time. For instance, anonymity settings also affect the content of SMD, with growing concerns about surveillance and the resulting loss of privacy (Ellison et al., 2011; Trepte & Reinecke, 2011; Turow et al., 2015), thus influencing what people are willing to post. There are also evolving rules about the banning of particular words and behaviours, as well as users, which may influence research findings (especially when conducting longitudinal research). SMD are private property of tech companies and can be arbitrarily erased or made inaccessible, compromising the replicability of research.

Outlook

Our review has several limitations. First, it focuses on social media but do not include other data sources that are frequently compared to survey data to model PO (e.g. Google trends, mainstream media, or administrative data). We thus encourage future research to extend the proposed complementary framework to additional data sources. This would allow the building of knowledge about the most suitable ways of combining these data for answering specific research purposes. Furthermore, our review entails a conceptual aim with less focus on the variety of methods used to either collect, clean, analyse, and aggregate the data to generate statistics. Discussing the pros and cons of methodologies employed by these papers could constitute the object of another review.

Notwithstanding these limitations, our study is not only of interest for social and political scientists concerned by the declining response rates and restrictive budgeting for survey research (Metzler et al., 2016). As social media have been established as multifunctional tools, and many companies and researchers implement strategies based on social media to collect opinions, make predictions, study behaviours, conduct experiments, or recruit hard-to-reach populations, this review is also of interest for practitioners.

Extracting PO from social media text can foster social sciences by moving it forward as an applied field, thus bridging gaps between computational models and interpretative research. We see this collaboration as particularly important for developing more advanced and reliable measures of opinion from social media texts. This also constitutes an opportunity to challenge the opposition of the so-called data-driven and theory-driven approaches, a simplistic dichotomy which further consolidates the misconception that social research can be conducted by relying solely on text-based data. We encourage researchers to acknowledge the different conceptualisation of opinion when measured by SMD and surveys, and we advise them to adopt a mixed-method strategy where the complementarity of both data is paramount.

CHAPTER 3. WHAT ARE RELIABLE METHODS TO EXTRACT OPINIONS FROM SOCIAL MEDIA DATA?

3.1. Dictionary-based and machine learning classification approaches: a comparison for tonality and frame detection on Twitter data[5]

Introduction: Textual analysis of social media

Political science and political psychology have looked with increased interest at social media analysis. On these platforms, in real time, people express opinions and evaluations of facts, the content of which can be used to measure peoples' thoughts and feelings as an alternative to traditional surveys (Schwartz & Ungar, 2015). This is typically done relying on text classification methods, which provide a way of estimating a category (or class) to which a given piece of text might belong. Grimmer and Stewart (2013) summarize different content analytical techniques lying on a continuum from dictionary-based, to (semi-)supervised (e.g. machine learning from a labelled training set), and fully data-driven methods (e.g. topic modelling). Because there is no guarantee that unsupervised methods return categories of theoretical interest, we suggest that it is important to develop a methodology that refine text classification combining dictionary and (semi-)supervised approaches.

Though these methods are under constant development, we argue that their application is sometimes limited for political and social scientists because of the narrowness of research questions and the consequent limited amount of data, compared with data science research. However, the baby shouldn't be thrown out with the bath water. In this article, we provide an example of application of some of these techniques to investigate tonality and frame detection (which are sub-tasks of text classification) in tweets related to democracy. More precisely, we address two main research questions: first, what are the advantages and disadvantages of each approach when relying on a small data set that is imbalanced with respect to the class distribution? And, are partial data-driven

[5] This chapter is a slightly adapted version of the article that has been published as M. Reveilhac and D. Morselli (2022): "Dictionary-based and machine learning classification approaches: a comparison for tonality and frame detection on Twitter data", Political Research Exchange, 4(1), 2029217.

techniques for custom dictionary inducement a worthy complement (or alternative) to other off-the-shelf lexical resources and supervised machine learning (ML) techniques in the realm of text classification?

The first research question draws from consideration of the difficulty supervised learning has in classifying the data on small data sets. Small data sets constitute a problem for the performance of machine learning classifiers, especially due to the limited number of examples. This difficulty is even amplified when the distribution of the data is skewed towards a few dominant categories in the data. Although small dataset and skewness of the data are separate issues for supervised learning, their combination can be very demanding for machine learning approaches. Here, dictionary-based approaches are especially beneficial for enhancing external generalizability (Amsler, 2020, p.135). Indeed, in contrast to supervised models that learn to classify the text into known categories based on a labelled training data set, dictionaries can cover cases that are not represented in the training set. For this reason, several lexicons have been used extensively and on many different domains for classification tasks – such as the General Inquirer from Stone, Dunphy, and Smith (1966), LIWC from Pennebaker et al. (2015) and Lexicoder from Young and Soroka (2012).

The second research question stems from recent advances in the methodology of deriving lexical resources where word embeddings are a component of the dictionary induction and enrichment (see Hamilton et al., 2016; Amsler, 2020). This strategy has the advantage of increasing the comparability to other data-driven approaches. We also insist on the fact that an iterative procedure between a data-driven step and a human validation step is paramount for allowing the inclusion of as much relevant information as possible in the dictionary.

Our empirical application is conducted on a corpus of tweets related to democracy. For the framing detection task (see also Amsler, Wüest & Schneider, 2016; Wüest, Amsler & Schneider, 2017), we aim to detect the presence (or absence) of democracy dimensions defined on the basis of previous survey research (Kriesi et al., 2013; Ferrín, 2018). Here, we define frames as 'schemata of interpretation' (Goffman, 1974, p.21) referring to the description (and the interest manifested) of the democratic decision-making process mentioned in tweets (e.g. political responsiveness or accountability). For the tonality detection task, we aim to detect the overarching sentiment conveyed in a given tweet, that is, its positivity or negativity. For both tasks, we compare the classification accuracies from a custom hand-curated dictionary, off-the-shelf lexicons, and supervised ML models. Our objective is not to build 'state-of-the-art' classifiers with optimal performance, but to understand how validly each method classifies the data under suboptimal conditions, that is utilizing a small and skewed training set. Thus, our goal is not to demonstrate that a classification method will always provide better classification accuracy than another. Instead, we aim to point how to test several methods and to decide which is the best for the given task and data. In this view, we also provide some guidelines to help researchers develop a custom dictionary and pre-process short texts before any tonality and frame detection can be produced.

Established methodologies in text classification

Dictionary-based and supervised learning approaches

When the volume of data is too large to be manually analysed, automated tools are needed to detect sentiment. Neuendorf (2016, 147) has argued that the fully automated approach is not a new procedure since it was introduced more than half a century ago with the General Inquirer (Stone, Dunphy & Smith, 1966), and has been widely used since then.

Dictionaries are perhaps the most intuitive way of classifying texts according to a priori defined classes. A dictionary uses the rate at which key words appear in a text to classify documents into categories (e.g. tweets, Facebook posts, news articles). It can do it either in a dichotomous manner or by using scores. Either way, the dictionary will correctly identify the categories only if the words contained closely align with how the language is used in the particular context under investigation. Indeed, the application of an 'off-the-shelf' dictionary to an area of research outside the substantive domain from which it has been developed can lead to classification errors, especially because similar terms can have different connotations in different contexts (see Loughran & McDonald, 2011). The writing conventions of different types of text (e.g. tweets often contain concatenated expressions, such as 'ClimateChange') further complicates the generalizability potential of dictionaries. In addition to the potential lack of domain generalizability, another possible pitfall of dictionaries refers to their restrictive domain coverage (or scope).Therefore, although a comprehensive set of keywords mapping unambiguously to a concept can produce highly reliable and efficient results, this method is sometimes criticized for its simplicity, and for its persistent difficulty in achieving completeness. To date, however, there exist methods of mitigating both of these difficulties – namely, the external generalizability and the vocabulary coverage – by making use of word embeddings to expand the dictionaries (e.g. Amsler, 2020).

Supervised learning methods provide an alternative for text classification into predetermined categories. They follow the following strategy. First, a sample of the corpus (the training dataset) is coded by humans for tone and frames. Then a classification method (ML algorithm) is selected, and the classifier is trained to predict the manually assigned labels within the training dataset. Here, multiple classification methods are generally applied and tested for minimum levels of accuracy. Finally, the chosen classifier is applied to the entire corpus to predict previously unseen texts (those not labelled by humans).

The supervised methods require it to be demonstrated that the classification from the ML algorithm replicates hand coding. A major advantage of this approach over the dictionary method is that it is necessarily domain specific (Grimmer & Stewart, 2013, p.275). However, it also has disadvantages, such as the necessity to label a substantial amount of text to train a reliable model (as the algorithm employed only learns from the features present in the training set). It also requires that the training data set not be skewed against a few classes (see Haixiang et al., 2017), otherwise the model may tend to overestimate larger categories at the expenses of smaller and sparser ones. For instance, in the case of tonality detection, most of the words carry no sentiment and the chosen algorithm can learn to wrongly associate neutral words with positive or negative sentiment (e.g. for the sentence 'e-voting sucks', the classifier will learn to associate the e-voting with the sentiment negative, and unlearning this will require training set instances with the word e-voting in them that are labelled positive). To optimize supervised models trained on small (or skewed) data sets, one prominent method for text classification is unsupervised pre-training. This approach has been widely adopted with the introduction of pre-trained word embeddings (Mikolov et al., 2013; Joulin et al., 2017), which rely on large data corpora. A recent strong performer in this line of research is BERT (Devlin et al., 2018). BERT models are unsupervised language representation, pre-trained using a plain text corpus. Thus, BERT is useful for generating context-specific embeddings, providing a pre-trained universal model. However, one of its major drawbacks is the computational resources needed to fine-tune and make inferences, especially on imbalanced classes. An additional difficulty with supervised models is that there is no easy way to fix classification errors (e.g. no control over induced learned parameters in the model).

In a nutshell and following Schwartz and Ungar (2015), no method is perfect, and each has advantages and pitfalls. Dictionary-based methods are accessible, theory-driven, abstract, and can be used with small samples, but they may overlook semantic context and misclassify text. Supervised learning can overcome some of these challenges for assigning documents to predetermined categories. Especially, it is necessarily domain specific and provide clear statistics summarizing model performance. However, it needs large data sets to 'learn' from the data to make accurate predictions.

We argue that the choice of one method over the other needs to be based on their performance, and that is necessarily context specific. As previously noted, some researchers warn against using 'off-the-shelf' dictionaries and emphasize the need to adapt them, or even create them from scratch for the task at hand (Grimmer & Stewart, 2013, p.275). However, the dictionary-based approach enables us to follow a strategy where transparency and sustainability are central pillars. This constitutes a major argument compared to the application of supervised learning approaches, which need a large amount of data to be trained and to balance possible class imbalance characterizing the data. However, the performance dictionary-based methods need to be established on a case-to-case basis.

Constrained but realistic research scenario: small and skewed manually labelled training set

A recent study from Barberá et al. (2021) provides guidance for researchers about the steps that need to be taken before any tonality assessment can be produced from newspaper coverage. The authors point out the decisions behind each analytical step, going from corpus selection to the choice of coding units, the trade-off between the number of annotators and the number of coded documents, as well as the comparison between supervised ML and a dictionary-based approach for tonality detection. They found that ML algorithms outperform dictionaries for tonality detection given a large enough training dataset. Their findings about the better performance of ML methods over dictionaries are congruent with the conclusions from Hartmann et al. (2019) in the marketing domain on social media.

In our study, we follow a similar endeavour in terms of assessing which classification method works best on a training data set. However, our study covers another research scenario that is characterized by the following aspects: we rely on tweets as textual data, we use a small sample of annotated data, our data are skewed toward few classes, we focus on tonality and frame detection, we build a custom dictionary, which means that we compare three classification methods (custom dictionary, off-the-shelf dictionary, and supervised learning models).

The research scenario that we propose is not unlikely. For instance, training data for generating novel prediction can be expensive and might not be possible in many real circumstances as supervised classifiers need to be trained on a quite large hand-labelled dataset to have enough predictive power. Furthermore, though social media are frequently used to discuss politically relevant topics, the popularity of some topics is often quite limited (e.g. fairness of the electoral system), meaning that the amount of data pertaining to one category can be underrepresented compared to other classes.

According to the existing literature aiming for classification accuracy, what is considered a ‘small’ training data set typically contains less than 2000 entries (Riekert, Riekert & Klein, 2021), which represent the usual size in social science. This, however, represents a suboptimal scenario when compared to the context of hundreds of thousands of labelled data that is used in computational linguistics and data science to build the ML models (e.g. Zhang, Zhao & LeCun, 2015; Joulin et al., 2017; Shen et al., 2018). The size of the training set is an important factor in determining which classification approach (dictionary or machine learning) or which classification algorithm is ultimately used. For instance, concerning ML, the models performing well on large training sets sizes tend to neglect the effect of the classes with very few examples on the performance (Cortes et al., 1994). Barberá et al. (2021) further discuss the trade-off between maximizing the number of coded documents by single coders to increase the size of the training set or having fewer annotated documents by multiple coders to increase the coding reliability. Their results showed that the ‘informational gains from increasing the number of documents coded are greater than from increasing the number of codings of a given document’ (p.30). Different research goals therefore imply different methodological strategies. Thus, researchers need to make the theoretically and practically appropriate choices in terms of the methods applied.

Sentiment analysis and frame detection as classifications tasks

Tonality and frame detection are two important research areas of text classification. They have become increasingly prevalent in research fields where texts are the primary data source. For instance, over 1200 papers applying automated text classification of sentiment have been published to date, spanning organizational science and marketing, psychology, medicine, and social science[6]. This line of research has doubled in size in the last three years, showing an expanding importance of these methods in different disciplines.

Tonality detection is often referred as sentiment analysis (or opinion mining) and is conceived as a flexible and powerful tool in some branches of political research, such as political psychology or electoral studies. Traditionally, sentiment analysis was conducted using dictionaries of word polarities (Young & Soroka, 2012). Although these approaches are being increasingly replaced by supervised classifications using supervised learning approaches, they continue to be actively developed. The long-standing popularity of sentiment analysis can be explained by the availability of ready-to-use methods and software, such as the Linguistic Inquiry and Word Count (LIWC, Pennebaker et al., 2015). These methods allow researchers either to rely on pre-compiled models (e.g. tonality dictionaries) or to easy tune research-specific dimension to be highlighted in text (Iliev, Dehghani & Sagi, 2015).

In parallel, a growing literature has aimed to develop tools for classifying texts in terms of issue or frame categories. Gilardi and Wüest (2018) summarize contributions relying on text applications for policy analysis into three broad research goals: extraction of specific information (concept identification), theory-driven allocation (classification), and inductive exploration of the underlying dimensionality (discovery). While fully automated (or inductive data-driven) methods of document classification such as topic modelling are a more common approach for frame detection (e.g. Blei, Ng & Jordan, 2003; DiMaggio, Nag & Blei, 2013; Roberts et al., 2014; Gilardi, Shipan & Wüest, 2021), there have also been works using dictionary-based approaches. Wüest, Amsler, and Schneider (2017) detected the presence (or claimed absence) of accountability of new forms of governance. In a similar vein, the Comparative Agendas Project aims to compare policies worldwide, thus investigating the trends in policymaking across time and between

[6] Data retrieved from Web of Science.

countries. In the frame of this project, the Lexicoder Topic Dictionaries were created to capture topics in news content, legislative debates, and policy documents (Albaugh et al. 2013)[7]. In addition to dictionary-based approaches, supervised algorithms have also helped researchers to identify dimensions from textual data. For instance, the Media Frames Corpus (Card et al., 2016) proposes a dataset of annotated news articles on 15 general purpose meta-frames (here called 'framing dimensions') and can enable researchers to develop and empirically test models of framing. Other examples include research by García-Marín and Calatrava (2018) to classify frames in the media about the refugee crisis and Gilardi and colleague's study (2021) about the co-evolution of different agendas (namely, the traditional media, parties' social media, and politicians' social media agendas) along specific policy issues.

The present study: tonality and frame detection in tweets related to democracy

The focus of this paper is on the evaluation of the tonality and frames of tweets related to democracy. We test the performance of a custom dictionary, off-the-shelf dictionaries, and unsupervised algorithms to achieve these classification tasks. After annotating tonality and theoretically relevant dimensions of democracy in a sample of randomly selected tweets from our corpus, we show how different dictionary-based and supervised classification approaches cope with the task of automatically detecting content and sentiment about democracy. We discuss the conditions under which the different approaches provides better classification results. Then, relying on the best classification models, we estimate tonality and democracy dimensions for previously unclassified tweets.

Unlike studies relying on large corpora, we are interested here in providing a procedure to obtain reliable results with small and skewed annotated data sets. We therefore draw from empirical works aiming to compare different research methods for text classification (e.g. Rooduijn & Pauwels, 2011; Hartmann et al., 2019; Barberá et al., 2021).

[7] Similar projects have used a dictionary-based approach to detect frames. We mention a sample. For instance, Maerz (2019) studied democratic systems relying on the analysis of the language of authoritarian leaders. Laver, Benoit, and Garry (2003) developed a dictionary of policy positions, where words have been selected semantically, based on how they relate to specific content categories as well as to a specific political party. Kraft (2018) explored whether and how individuals evoke moral considerations when discussing their political beliefs (e.g., harm, fairness, ingroup, authority, purity, and morality).

Using democracy as the study focus introduces several interesting aspects to our purposes. Democracy is, without doubt, one of the most complex concepts of contemporary political science. It can be framed in different way, with some frames (e.g. representativeness, responsiveness) being less spares than others (e.g. sovereignty; Fishman, 2016).

Democracy has also been the subject of multiple opinion surveys. To date, social media offer an alternative view on the working of democracy. On social media, discussions about democracy can arise from a variety of stimuli, such as attention to particular events or scandals, and personal motivation to post or not post a tweet. Even if unrepresentative of the wider public (see discussions of social media biases by Japec et al. (2015) and Schober et al. (2016)), social media discussions can thus serve as a complement to more stable surveyed attitudes. Therefore, the last section of our paper is dedicated to the correspondence between attitudes found in surveys and social media prevalence on similar democracy dimensions.

Methodology

Data collection: tweets related to democracy using a list of search-queries

We built our corpus by retrieving tweets related to democracy based on a list of relevant terms referring to the workings of democracy and extracted from in the main Swiss German[8] and French[9] language newspapers in 2018. Newspapers were retrieved from the Swissdox repository using 'democracy', 'populism', and 'Swiss people' as search entries for article titles. The lemmatized articles' text was used to extract the top 50 terms associated with the word 'democracy'. This extracted list was thus used as search keywords in the Twitter API to retrieve tweets. The restriction to three keywords could lead to a specific pre-selection and hence to a bias in the association scores, but the newspaper search only constituted the first step upon which the final list of search-queries is elaborated.

---

[8] Swiss German speaking quality newspaper (Neue Zürcher Zeitung), seven regional newspapers (Tagesanzeiger, Aargauer Zeitung, Basler Zeitung, Berner Zeitung, Neue Luzerner Zeitung, Die Südostschweiz, St. Galler Tagblatt), one free newspaper (20 Minuten), one tabloid press (Blick), and three Sunday press (Sonntagsblick, NZZ am Sonntag, Sonntagszeitung).

[9] One Swiss French speaking quality newspaper (Le Temps), seven regional newspapers (Tribune de Genève, 24 heures, La Liberté, Le Nouvelliste, Le Courrier, L'Express, L'Impartial), one free newspaper (20 Minutes), one tabloid press (Le Matin), and one Sunday press (Le Matin dimanche).

After retrieving the tweets, we extracted the hashtags from the corpus of collected tweets and added the ones related to democracy into the search list. We retrieved tweets with the extended list once again to make sure to include all relevant hashtags. Our final list of search-queries contained 51 French terms, 56 German terms, 50 Italian terms, and 71 hashtags related to general policymaking in Switzerland (see Annex 3.1.1: List of search-queries to collect the tweets).

The extracted tweets were also filtered by time and location. With respect to location, Twitter provides two classes of geographical metadata. The tweet location, which is available when users share location at time of tweeting, and the account location, which is based on the 'home' location provided by users in their public profile. However, very few users provide these meta-data, which is why we adopted another retrieval strategy. First, we concatenated the search-queries with the word 'Switzerland'. In this way, we obtain tweets that entail one of our search-queries and mention Switzerland. Second, we used the possibility of retrieving tweets from users situated in each region by specifying a geographical radius in addition to the search-queries. In this way, we obtained tweets that entail one of our search-queries and that are posted by users mainly in Switzerland or in regions very close to it. In a final step, we kept only the tweets emitted from January to December 2018, and collected the replies to those tweets to increase the size and variability of the corpus. Our final corpus of tweets is composed of 296,375 German, French, and Italian tweets.

Translation and pre-processing steps

To overcome the multilingual diversity in Switzerland, we decided to translate every tweet into English using Google Translate (see pre-processing section below). There are additional reasons for translating the tweets into English. Firstly, we should generate a sufficient amount of data to train ML models (see section 3.6). Secondly, we should have a sufficient amount of data to amplify our custom dictionary (see section 3.4). Thirdly, and more pragmatically, most of the off-the-shelf dictionaries are exclusively in English. Finally, Google Translate works best when translating into English, thus homogenizing the translation biases.

The translation process nonetheless entails two essential limitations. The first is linked to the interpretive sophistication. Indeed, words that pertain to complex socio-political phenomena can evoke different meanings regardless of the translation, a particularity with which the translator might not always cope. For instance, the word ‘nation’ can be translated to ‘state’ or ‘republic’ depending on the original language and on the context known by the translator. The second limitation is linked to the vocabulary coverage, which might be oriented toward most commonly used words instead of sophisticated expressions[10].

Before translation is applied, we conducted several pre-processing steps, namely links removal and split of concatenated words (e.g. ‘#ClimateChange’ becomes ‘Climate Change’). These pre-processing steps were important for maximizing the correctness of the translation. Further data cleaning steps were conducted after translation, namely removal of conventions (# and @), and lower-casing. We also removed stop-words, except negations, removal of which might harm the overall classification accuracy.

Tonality off-the-shelf dictionaries

We investigated the performance of five off-the-shelf dictionaries for sentiment classification. Typically, simple ratios (e.g. share of words with positive or negative emotions) or count scores (e.g. number of words) are computed.

First, we wanted to assess the performance of the dictionary included in the LIWC (version 2007), which is widely used in social science and psychology. Second, we analysed the performance of the NRC Emotion Lexicon (Mohammad and Turney 2013), a crowdsourced dictionary of sentiments which includes both unigrams and bigrams. Third, we included AFINN (Nielsen 2011), a dictionary dedicated to analysing tweets which emphasize acronyms and other expressions used in microblogging, such as emotional reactions (‘LOL’ and ‘WTF’). Fourth, we analysed Lexicoder, a dictionary designed specifically for political text (Young & Soroka, 2012). Lastly, we included Hu Liu’s dictionary (Hu & Liu, 2004), which was created on the basis of a set of seed adjectives (‘good’ and ‘bad’) and semantically expanded by applying the dictionary with synonymy and antonym relations provided by WordNet[11]. The choice of using these dictionaries was

[10] We also noted that depending on whether the grammatical and syntactic rules are very different from the resulting language, the translation can be far from the original word ordering, sometimes losing the meaning of the text. Furthermore, while some words can be suitable for direct translation into a single term, other words are complex to translate. In these cases, the automatic translator might be incentivised to translate with the most commonly used words.

[11] WordNet (Miller 1995) is a large lexical database of English. Nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. See more here: https://wordnet.princeton.edu/

based on the domain specificity of our corpus of tweets and on the language, as most lexical resources have been developed for English.

Custom dictionary for tonality and democracy dimension

The study on tonality about democracy (i.e. positive/neutral/negative connotations) was also conducted to extract several democracy dimensions. We focused on 11 dimensions theoretically based on the previous work of Kriesi et al. (2013) and Ferrín (2018) for the rounds six and ten of the European Social Survey (ESS).

These dimensions were adapted to the context of Twitter discussions. We grouped vertical and horizontal accountability, we differentiated voice and institutional participation, we added a category of sovereignty, and we did not differentiate between social and political equality. Compared to the survey items, our coding scheme differed in three further ways. First, we had to group some dimensions. The categories 'freedom' and 'rules' form two different survey dimensions, whereas they are grouped into one in our coding. Furthermore, the categories 'voice' and 'sovereignty' are absent from the survey items, but we introduced them in our coding scheme as they represented important dimensions in social media texts. Moreover, we did not differentiate between horizontal and vertical accountability when coding the tweets, as both dimensions were hard to disentangle in social media texts (see our final coding scheme in Table 3.1.1). Democracy could be framed along a number of other dimensions. We focus on these ones for the sake of our methodological study. However, the procedure outlined below can be applied to any dimension or concept, and is not necessarily limited to democracy.

A list of initial seed words associated with these dimensions were thus included in our democracy dimensions custom dictionary[12]. We used WordNet to look for other words that were obvious synonyms and antonyms of the words included in the dictionary[13]. Then, we followed the strategy elaborated by Amsler (2020), consisting in embedding the words from our dictionary into the overall corpus (not only a sample of manually annotated data) using word2vec (Mikolov et al., 2013) as implemented in the gensim

---

[12] If no previous concepts are available and the dictionary must be elaborated from scratch, software such as WordNet (Miller 1995) can be used to collect main concepts.

8. WordNet (Miller 1995) is not available in many languages in such comprehensive versions as for English.

[13] The traditional approach towards opinion mining or text classification in general (also called Bag-Of-Words approach) considers a sentence (or a document) as Bag containing words. It considers the words and their frequency of occurrence in the document, disregarding semantic relationship in the sentences (albeit the Bag can grow by searching for synonyms and antonyms).

library from Python. The general idea of word embedding is to learn word associations from a large corpus of text[14]. Once trained, such a model can detect synonymous words or suggest additional words for a partial sentence (e.g. man is to king as women is to X). We trained our own model based on the collected tweets and we used Amsler's LexExpander and LexEmbedder tools to further expand our dictionary.

For each entry in our dictionary, we determine whether they appeared in the sample of manually coded tweets and calculated a simple ratio of how many times they were coded as negative (wn) or positive (wp). This gives us the following equation: $\sum wp - \sum wn$. The entry is labelled as positive if the overall score is above 0, and negative if below 0. A neutral category was added for entries that do not appear in the sample of annotated tweets or if the ratio of negative and positive tweets is equal.

Our final dictionary was composed of 11 democracy dimensions and 1776 words or expressions distributed among our dimensions. Technically, the matching of dictionaries was implemented using software liwcalike() from the quanteda package for R (Benoit et al., 2018).

[14] A tweet was coded as neutral only if it contains an announcement of an event or a brief journalistic review. In total, we coded 50 tweets as neutral.

Table 3.1.1:. Coding scheme for the annotation of democracy dimensions.

Title	Coding label	Keyword examples	Description
Competition & fairness of electoral procedures	competition	*election, campaign, candidate*	Guarantees that elections are competitive, free, open, and fair (precondition for responsiveness). It also refers to the fact that the opposition is free to criticize the government.
Representation	representation	*elected representatives, governing coalition, congressmen*	Representation of citizens in the formal institutions, which translates in the fact that elected political actors act in citizens' interest. The notion also refers to the *crisis of democracy*, but also to the type of representation (e.g. majority vs. proportional).
Self-organized participation of citizens	voice	*social movement, strike, riot*	Citizen's self-organization into collective actions and reinventing political activism.
Participation of citizens through democratic procedures	institutional participation	*direct democracy, participatory budget, e-voting*	Citizens' views are voiced in the political dialogue to be heard by political authorities (e.g. parliament, government, court) through institutional channels (e.g. plebiscite, popular initiative, referendum, petition, e-voting, etc.).
Social fairness & social or political equality	fair	*redistribution, social justice, welfare*	Social equality is the elimination of social and economic differences (e.g. poverty, income inequality) that would stand in the way of political equality. Political equality ensures that all citizens have equal opportunities to participate politically and to access the law.
Efficiency	efficiency	*laundering, embezzlement, corruption*	Implies that the democratic system discourages political authorities' corruption and mismanagement of resources, but also guarantees parliament and government independence (e.g. from lobbies, interest groups, mafia, etc.).
Vertical & horizontal accountability	accountability	*transparency, blameworthiness, limited power*	Vertical accountability refers to mechanisms through which the people control their representatives. It also implies that elected politicians account, be responsible, answer for their political decisions and give justifications form their policy choices. Horizontal accountability refers to the *checks and balances* on government power via the courts.
Responsiveness to citizens	responsiveness to citizens	*answerability, solution oriented, public opinion*	Refers to the capacity to form and implement policies that citizens want (also accounts for democratic technologies such as e-democracy, e-governance, digital-government, etc.).
Responsiveness to other stakeholders	responsiveness to other stakeholders	*agreement, convention, freedom of movement*	Responsiveness can also be understood in terms of answerability to other governments, supernational bodies (e.g. EU, UN, etc.) or agreement (e.g. Schengen-Dublin).
Freedom & Rule of law	rules	*court, freedom of speech, impartiality*	Refers to the whole set of rights and liberties available to citizens (e.g. minority rights). The courts treat everyone the same, thus implying equality before the law. Gives information about what citizens think are essential characteristics of democracy (e.g. freedom of expression, freedom of the press, reliability of the press, etc.).
Sovereignism & nationalism	sovereignism	*self-determination, nationalization, autonomy*	Accounts for the competing views of *liberal democracy*, such as nationalism and sovereignism (defending a nation's political autonomy, preserving cultural identity, and shielding the domestic economy), which are often conflated with populism or Euroscepticism.

Manual coding of a sample of tweets

We coded 1426 randomly extracted tweets from our full dataset. The coding process was carried out by a single expert coder and was composed of two main tasks. First, each tweet was coded according to its tonality under three categories (positive, negative, and neutral[15]). Second, each tweet was assigned to its main democracy dimension based on the coding scheme shown in Table 3.1.1. Regarding tonality, the sample of annotated tweets entails slightly more negative tweets (n = 729) than positive tweets (n = 646). With respect to democracy dimensions, annotated tweets tend to be skewed towards three categories, namely 'institutional participation' (n = 263), 'representation' (n = 224), and 'rules' (n = 217), followed by 'accountability' concerns (n = 161), 'responsiveness to citizens' (n = 154), 'competition' (n = 131), 'voice' (n = 78), 'responsiveness to other stakeholders' (n = 76), 'fair' (n = 60), 'efficiency' (n = 37), and 'sovereignism' (n = 23).

Chosen ML classification models.

We tested a set of ML models due to their conceptually different algorithmic approaches. The ROCCHIO (Rocchio, 1971) classifier was implemented for both tonality and democracy dimensions tasks. ROCCHIO is a nearest centroid classifier applied to text classification. This method assigns to observations the label of the class of training samples which mean (centroid) is closest to the observation.

In addition, two different algorithms were used according to the specific task. Logistic regression, which is a classification algorithm used to solve binary classification problems, was used for tonality. For democracy dimensions we used SVM which are discriminative classifiers, fitting a margin-maximizing hyperplane between classes, and can be extended to non-linear problems of higher dimensionality using kernels that can accommodate any functional form.

In addition to supervised models, we also included an unsupervised BERT, which is a transformer for unsupervised language representation. Unlike word2vec, which generates a single word embedding representation for each word in the vocabulary, BERT considers the context for each occurrence of a given word. We pre-trained the BERT on the ensemble of collected tweets (n = 296,375). Therefore, unlike static resources, such as dictionaries, BERT calculates embeddings dynamically to represent the language data.

[15] A tweet was coded as neutral only if it contains an announcement of an event or a brief journalistic review. In total, we coded 50 tweets as neutral.

We then conduct multinomial regression on the top of the BERT representation for predicting tonality and democracy frames.

Preparing the training and testing data sets

The annotated data are split into training and testing sets by a ratio of 80% (n = 1140) and 20% (n = 286). The performance of both ML and dictionary models is evaluated using the training/test approach via accuracy, precision, recall, and the F-score. Accuracy is the ratio of correctly predicted observation to the total observations. Precision indicates how many among the predicted positive are positive (true positive). In other words, it is the ratio of true positives over all those classified as positive. Recall is the ratio between true positive and all the actual positive units. The F-score combines precision and recall, and scores one under perfect precision and recall.

Results

Tonality classification

Focusing on tonality, Table 3.1.2 shows that an established dictionary, such as Lexicoder, is a good fit for the task at hand. However, the custom dictionary also performs very well, suggesting that we were able to include relevant features in the dictionary and to score them in a relevant direction. Nonetheless, dictionaries necessarily limit the amount of information that can be learned from the text. This is shown in Table 3.1.2, where BERT and ROCCHIO models outperform the dictionaries. ML, based on logistic regression, comes in second place for accuracy.

In contrast, other existing dictionaries (HuLiu, AFINN, and NRC) perform less well on the sample of annotated data. Furthermore, LIWC does not exceed random chance to predict tonality on our sample. The lower performance of some dictionaries (HuLiu, AFINN, LIWC07, NRC) compared to ML is notably due to the context of study. LIWC contains negative and positive words based on reviews that follow a stringent logic, contain little noise, and carry emotion-laden words. However, the HuLiu dictionary, which is constructed around adjectives to detect the emotional orientation of customer reviews, performs better. This can be since the HuLiu includes processing properties that can cope with misspellings and morphological variations. The low performance of existing dictionaries can be explained since, in our real-world application of tweets about democracy, the tonality of a text tends to be conveyed in less obvious ways. However, this does not necessarily invalidate their usefulness for predicting tonality in other circumstances. It nevertheless points to the necessity of testing 'off-the-shelf' dictionaries before we can derive conclusions from them. The number of missing matches in our custom dictionary are essentially due to very unusual writing styles and the use of multiple languages in the same tweet, which can lead to poor translations.

Table 3.1.2: Classification results for predicting the tonality of tweets related to democracy.

Precision	Recall	F-measure	Accuracy	Model	Missing
0.92	0.87	0.89	0.90	*BERT*	0
0.86	0.89	0.88	0.88	*LR*	0
0.85	0.78	0.81	0.83	*ROCCHIO*	0
0.95	0.65	0.77	0.81	*Lexicoder*	91
0.83	0.74	0.78	0.79	*CustomDict*	35
0.84	0.61	0.70	0.74	*HuLiu*	199
0.83	0.58	0.69	0.72	*AFINN*	177
0.85	0.55	0.67	0.71	*LIWC07*	790
0.91	0.27	0.42	0.62	*NRC*	24

Classification of the democracy dimensions

Using a similar procedure to that for tonality, we compared dictionary and ML methods to classify tweets into the democracy dimensions. Because no off-the-shelf democracy dictionaries exist to detect our theoretical dimensions, for this task only our custom dictionary was used for the dictionary-based method.

Table 3.1.3 summarizes the accuracy of all democracy dimensions models. The average accuracy over all frame categories is as follows: 86% for the custom dictionary, 69% for BERT, 69% for ROCCHIO, and 66% for SVM. Our custom dictionary outperforms the other classification approaches for almost all democracy dimensions. BERT also performs generally better than ROCCHIO and SVM models.

Table 3.1.3: Classification results for predicting democracy dimensions of tweets related to democracy.

Category	**Precision**	**Recall**	**F-measure**	**Accuracy**	**Model**
Accountability	0.66	0.84	0.74	0.89	*CustomDict*
	0.38	0.56	0.45	0.73	*BERT*
	0.36	0.30	0.33	0.62	*ROCCHIO*
	0.23	0.26	0.25	0.59	*SVM*
Competition	0.69	0.83	0.76	0.90	*CustomDict*
	0.80	0.63	0.70	0.80	*BERT*
	0.73	0.50	0.59	0.74	*ROCCHIO*
	0.63	0.47	0.54	0.72	*SVM*
Efficiency	0.83	0.81	0.82	0.90	*CustomDict*
	0.50	0.17	0.25	0.58	*SVM*
	0.50	0.08	0.14	0.54	*ROCCHIO*
	–	0.00	–	0.50	*BERT*
Fair	0.78	0.81	0.80	0.90	*CustomDict*
	0.20	0.20	0.20	0.59	*SVM*
	0.33	0.10	0.15	0.55	*BERT*
	0.00	0.00	–	0.50	*ROCCHIO*
Institutional participation	0.67	0.81	0.73	0.86	*ROCCHIO*
	0.56	0.83	0.67	0.84	*BERT*
	0.63	0.72	0.67	0.81	*SVM*
	0.95	0.59	0.73	0.79	*CustomDict*
Representation	0.81	0.77	0.79	0.87	*CustomDict*
	0.40	0.61	0.49	0.72	*ROCCHIO*
	0.53	0.48	0.50	0.70	*BERT*
	0.42	0.45	0.43	0.67	*SVM*
Responsiveness to citizens	0.61	0.75	0.67	0.84	*CustomDict*
	0.50	0.32	0.39	0.64	*BERT*
	0.38	0.32	0.35	0.63	*ROCCHIO*
	0.29	0.29	0.29	0.60	*SVM*
Responsiveness to other stakeholders	0.69	0.80	0.74	0.89	*CustomDict*
	0.36	0.63	0.45	0.80	*BERT*
	0.40	0.50	0.44	0.74	*ROCCHIO*
	0.27	0.38	0.32	0.67	*SVM*
Rules	0.78	0.80	0.79	0.88	*CustomDict*
	0.67	0.73	0.70	0.83	*BERT*
	0.59	0.73	0.65	0.82	*ROCCHIO*
	0.50	0.56	0.53	0.73	*SVM*
Sovereignism	0.85	0.48	0.61	0.74	*CustomDict*
	1.00	0.14	0.25	0.57	*ROCCHIO*
	–	0.00	–	0.50	*BERT*
	0.00	0.00	–	0.50	*SVM*
Voice	0.75	0.70	0.72	0.84	*CustomDict*
	0.85	0.61	0.71	0.80	*ROCCHIO*
	0.63	0.56	0.59	0.77	*SVM*
	0.53	0.50	0.51	0.74	*BERT*

There are large differences in maximum accuracies between the different classification methods with respect to the 'efficiency' and 'sovereignty' categories, which entail very few annotated data and impede more sophisticated classification methods to 'learn' from the data. The dictionary-based classifiers do not need to learn custom classes automatically from training data. Instead, they require expert-crafted dictionaries for such purposes and are likely to perform well if the scope of the included features has a good coverage of the research dimensions.

Factors modelling the probability of a true value

The previous descriptive findings and statistical tests suggest certain plausible explanations for these differences. However, the different potential explanations for the observed accuracy differences cannot be disentangled. To further investigate which factors were possible drivers of performance, we ran a series of logistic regression models across the test dataset with accuracy of predicting human coding (correct vs. incorrect) as the dependent variable.

Text level factors, such as the number of words in the tweets, the number of hashtags (i.e. over or under five hashtags per tweet), the presence of mentions in the tweets, the original language of the tweets, the annotated sentiment, as well as the mention of any ambiguity encountered during the manual annotation process and whether the tweets contained an interrogation mark, were all inserted in the model as predictors.

With respect to the detection of tonality, we focus on the five best models (see columns 2–6 in Table 3.1.4). The regression results are displayed in Table 3.1.4 and show that tweets manually labelled as negative are more likely to be wrongly predicted by our custom dictionary and Lexicoder models in comparison to positive labelled tweets, although the reverse trend applies to ROCCHIO. Looking at the original language of tweets, French tweets were also more likely to be wrongly classified by our custom dictionary than German tweets. BERT and ROCCHIO also struggled to correctly assign tonality for tweets containing an interrogation mark.

With respect to the detection of democracy dimensions, we focus on the four best models (see columns 7–10 in Table 3.1.4). The regression results are displayed in Table 3.1.4 and show that the original language of tweets plays an essential role for the custom dictionary: French, and tweets in other languages, were more likely to be correctly classified by our custom dictionary than German tweets. Furthermore, the mention of an ambiguity during the manual annotation process leads tweets to be wrongly classified in every model. In addition, when looking at the tonality of the tweet, neutral tweets are more likely to be wrongly predicted by our custom dictionary and ROCCHIO in comparison to positive labelled tweets.

The results presented in Table 3.1.3 also point to the possible complementarity of the methods for text classification. Indeed, triangulating the prediction from the best dictionary and ML models could compensate the respective weaknesses of each approach. For instance, the methodology used in this article to induce and expand the custom dictionary can lead to the inclusion of terms that refer to several (albeit similar) concepts. However, ML models, especially BERT, are better at capturing single dense concepts that required more contextual understanding. Dictionaries could thus better identify the tweets containing rare (or complex) terms that can be matched with more than one translation, while ML models are better at integrating the contextual information.

Comparing surveyed attitudes with the prevalence of social media discussions

So far, we have applied multiple classification methods and tested for minimum levels of accuracy to determine the best classifier. In this section, we aim to compare surveyed attitudes from respondents to round six of the ESS on items covering the democracy dimensions that inspired the coding scheme for the extracted social media messages.

Concerning the measurement of democracy dimensions prevalence from Twitter messages, we select the best dictionary – our custom dictionary – and the best classifier – BERT –, and we apply them to the entire corpus to derive the prevalence of democracy dimensions from unclassified tweets (those not labelled). We selected only those tweets that were equally classified by both methods (43,859 tweets in total). The sovereignty dimension is not represented among these tweets as the two classification methods did not reach agreement on this dimension. This can be explained by the fact that the sovereignty dimension was hardly represented in the annotated corpus and could only be identified by our custom dictionary. Conversely, the dimension accounting for rules of law and freedom related concerns is more prevalent in the corpus than any other dimension.

Concerning the measurement of the prevalence of democracy dimensions among survey respondents, we first had to choose between two possible measurements. The ESS distinguishes between people's beliefs and expectations about what a democracy should be, asking whether different aspects are important for the workings of democracy, and people's evaluations of their own democracies with respect to these aspects. We chose to concentrate on respondents' rated importance of the democracy dimensions as it compares better conceptually to the estimated prevalence of democracy dimensions in social media. We organized the survey items related to democracy into the corresponding democracy dimensions following the template from Kriesi et al. (2013). However, we remove items that are negative formulations of positive items and items that are at odds with the functioning of democracy in Switzerland. The response scale to rate each aspect range from 1 'not important at all' to 11 'extremely important'. We selected only respondents who positioned themselves on all survey items (N = 686). To be able to compare the survey results with social media findings, we only considered answers above seven (excluded) when rating the importance of each dimension for the workings of democracy. We took the mean if more than one item represented the dimension.

Table 3.1.4: Logistic regression on method performance to predict tonality (left) and democracy dimensions (right) as a function text and coding process characteristics.

	Tonality					Democracy dimensions			
	CustomDict	**Lexicoder**	**BERT**	**ROCCHIO**	**LR**	**CustomDict**	**BERT**	**ROCCHIO**	**SVM**
(Intercept)	1.36 * (-0.65)	2.41 *** (-0.71)	1.31 (-0.87)	1.35 (-0.73)	0.62 (-0.81)	1.32 * (-0.63)	1.16 * (-0.54)	1.47 ** (-0.55)	0.93 (-0.54)
Number of words in tweet	−0.00 (-0.02)	0 (-0.02)	0.03 (-0.03)	0.01 (-0.02)	0.06 * (-0.02)	−0.01 (-0.02)	−0.02 (-0.02)	−0.02 (-0.02)	−0.03 (-0.02)
Tweet contains hashtags (ref: <5)									
Hashtag: >5	−0.16 (-0.41)	0.51 (-0.48)	0.35 (-0.65)	−0.38 (-0.43)	−0.36 (-0.5)	0.57 (-0.44)	−0.20 (-0.35)	−0.16 (-0.35)	−0.25 (-0.35)
Tweet contains mention (ref: no)									
Mention: yes	−0.06 (-0.3)	0.1 (-0.33)	0.84 (-0.48)	0.72 (-0.37)	−0.07 (-0.4)	−0.36 (-0.31)	0.04 (-0.27)	0.07 (-0.27)	0.12 (-0.27)
Original language (ref: German)									
French	0.69 * (-0.35)	−0.04 (-0.37)	−0.69 (-0.52)	−0.52 (-0.39)	−0.43 (-0.45)	0.75 * (-0.34)	0.16 (-0.3)	−0.15 (-0.3)	0.39 (-0.3)
Other	0.45 (-0.38)	−0.29 (-0.39)	−0.86 (-0.52)	−0.39 (-0.43)	0.15 (-0.53)	1.22 ** (-0.4)	−0.10 (-0.31)	−0.26 (-0.31)	0.2 (-0.31)
Annotated sentiment (ref: positive)									
Negative	−0.68 * (-0.29)	−2.21 *** (-0.37)	0.6 (-0.42)	0.73 * (-0.33)	−0.39 (-0.39)	−0.49 (-0.29)	−0.09 (-0.25)	−0.46 (-0.25)	−0.42 (-0.25)
Ambiguity in coding the tweet (ref: no)									
Ambiguity*: yes*	0.33 (-0.47)	0.09 (-0.5)	0.07 (-0.68)	−0.76 (-0.45)	0.52 (-0.78)	−1.35 ** (-0.42)	−1.11 ** (-0.42)	−1.28 ** (-0.43)	−1.06 * (-0.44)
Tweet contains a question mark (ref: no)									
Interrogation: yes	−0.46 (-0.34)	−0.18 (-0.37)	−0.97 * (-0.48)	−0.90 * (-0.39)	−0.36 (-0.48)	0.65 (-0.53)	0.2 (-0.4)	0.55 (-0.42)	0.45 (-0.4)
AIC	329.02	291.52	193.01	269.39	214.09	320.6	398.44	390.1	400.43
BIC	361.98	324.49	225.98	302.35	247.06	353.56	431.4	423.06	433.4
Log Likelihood	−155.51	−136.76	−87.51	−125.69	−98.05	−151.30	−190.22	−186.05	−191.21
Deviance	311.02	273.52	175.01	251.39	196.09	302.6	380.44	372.1	382.43
Num. obs.	288	288	288	288	288	288	288	288	288

Note: Significativity levels are coded as follows: ‘***’ p< .001, ‘**’ p < .01 ‘*’ p < .05.

To be able to directly compare survey and social media prevalence on the common democracy dimensions, we do a correlation analysis (see Figure 3.1.1). We computed Spearman rank correlation between the number of tweets classified in each dimension and the number of answers with a score higher than 7. The correlation coefficient is 0.58, confirming a positive correlation between the opinion surveyed in the ESS and those spontaneously expressed on Twitter. It also shows that a moderate proportion of variance differs between the scores extracted from the tweets and those in the ESS, supporting the argument for a complementarity of the two methods. Social media users and survey respondents emphasize the category related to the rule of law. Political accountability and competition are also perceived as important for the workings of democracy. On the opposite, responsiveness to other stakeholders (mainly the European Union) is considered least important.

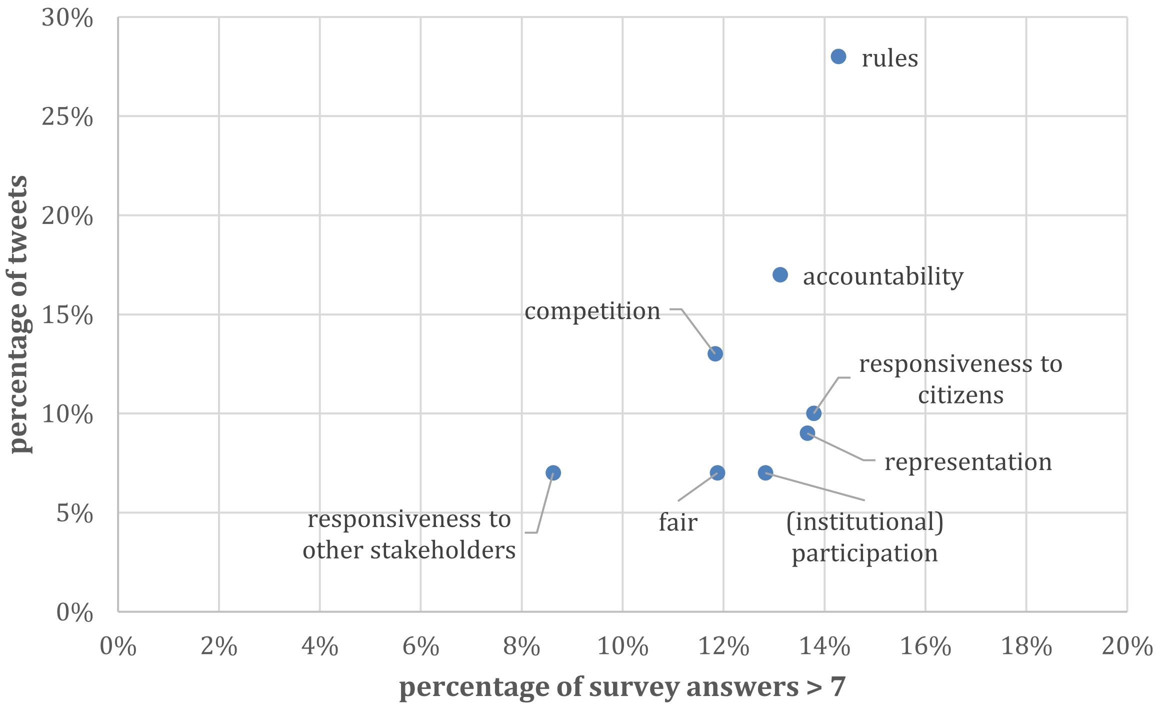


Figure 3.1.1: Relation between the importance of democracy dimensions for democracy in the ESS (x-axis) and the number of tweets related to each dimension (y-axis).

Discussion and concluding remarks

Where to use dictionaries and where to use ML?

The goal and main contribution of this manuscript is to provide easily implementable recommendations for increasing estimation accuracy under non-optimal conditions. With this goal in mind, we discuss how to run and compare different methods, in order to make informed and well-suited decisions. In this article, we develop an empirical application relating to tweets about democracy, from which we detect their tonality (positive versus negative sentiment) and frames (substantive democracy dimensions). We have relied on two main approaches for conducting both sentiment and framing analysis, namely supervised ML, and dictionary-based frameworks. We also employ a third approach which consists in using unsupervised classification methods relying on word embeddings to classify text by word vector similarity. These approaches allow researchers to code vast amounts of text that would not be possible with manual coding, and each presents unique advantages but also challenges.

Our main goal was to assess how well these approaches perform on a small and skewed sample of annotated data, and how off-the-shelf dictionaries compare to 'custom' dictionaries to accomplish these classification tasks. Our objective was not to build 'state-of-the-art' classifiers with optimal performance, but to provide a procedure with which to understand how methods utilizing minimal training data compare against off-the-shelf dictionaries.

In our study, the dictionary methods lagged behind BERT and ROCCHIO models for tonality classification, but performed better when the task was more complex, such as classifying tweets according to several democracy dimensions. This can be explained partially by the over-representation of negative tonality across all democracy dimensions (except for the institutional participation) in our sample of annotated tweets. Indeed, the fact that Twitter users tend to post more on aspects of democracy which are perceived could have 'nudged' classifiers in a negative direction, thereby enabling ML methods to classify negative tweets more correctly than dictionary-based approaches. This finding is confirmed by the regression analysis on the prediction accuracy, which shows that ML classify tonality more easily than dictionaries.

On the contrary, our custom dictionary performed well for detecting democracy dimensions as it contains words that contribute significantly to these dimensions.

The dictionary-based approach could thus serve as a useful complement to ML classifiers, which rely on every word feature contained in the training dataset fed in. However, if the reliance of ML models on the entire training dataset has the advantages of covering many more word features than the ones present in the dictionary, it may also be problematic when there is a bias toward most prevalent categories (skewness). This result should not be interpreted as indicating that dictionary methods are better than ML in absolute terms. On the contrary, we think that BERT models provide the most appropriate solution to date. In our analysis, BERT was indeed the most reliable among the ML models, though it performed less well than the custom dictionary method for very specific categories that were under-represented in our sample (e.g. 'sovereignism'). In cases of large training data and/or balanced categories, the ML methods might have outperformed against the dictionary ones. However, this is a context-dependent aspect which cannot be generalized, but needs to be investigated each time. Importantly, the fact that ML approaches tend to be error-prone for classes with fewer examples should not mean that lower performance is acceptable as long as the error is distributed in an unbalanced way. A possible solution would be to use the dictionary to over-sample the under-represented categories in the training set before ML models are employed.

What best practices can we derive from our case study?

Our results showed the usefulness of testing different approaches to carrying out a classification task (in our case, we took the best dictionary and the best unsupervised classifier), as it enables us to assess which classification method is more suited to the classification task at hand. Instead of blindly relying on preconstructed methods, a train-and-test framework can be easily set to empirically judge the accuracy and reliability of each method. Even the use of a small, annotated data set might be useful for understanding which among different methods of estimation or different dictionaries could be the most appropriate for the goal of the study. If larger training datasets are preferable, good proxies can lead to reliable results, even under poor research conditions. We also showed a procedure of building custom dictionary which complements embeddings techniques and human judgement. To construct the dictionary, we went through several steps by, first, departing from a few seed words derived from the literature and from mainstream media to which we added salient terms and hashtags from the sample of annotated tweets and, second, we trained a word vector model (see

Amsler's (2020) LexEmbbeder and LexExpander) based on the entire collection of tweets to add eligible word candidates to our dictionaries. This process involved a continuing translation between human validation and computational suggestions. We believe this method might be of particular interest for social and political scientists because it creates a loop between data-driven findings and theory-driven choices. On the one hand, it allows to build more exhaustive dictionaries, and on the other gives the researchers a control over the concepts they want to measure, relying less on black-box algorithms and estimating possible imbalances in the data.

The translation process of the tweets into English has a non-zero impact on the results, albeit it is difficult to really assess how. While we could have elaborated our custom dictionaries in the original languages of our corpus and translated the off-the-shelf dictionaries term-wise to apply these resources in the original languages of the tweets, there is an opportunity to seek a minimal transfer cost. Most notably, there is a non-negligible issue arising from the multitude of possible translations for single terms in each of the chosen languages (Vicente & Saralegi, 2016). Concerning the application of ML models, a possible solution would have been to align the text in each of the original languages contained in our data and to conduct separate models for single languages. Other solutions might include relying on transfer learning or cross-lingual embeddings, but this would require a non-negligible amount of training data.

Before we conclude, we would like to point to several aspects that are worth investigating in future case studies. The manual verification of the misclassified data by both methods enabled us to derive several common difficulties. For instance, both methods had trouble tackling specific linguistic patterns, such as double negations, thwarted expression, and negated expression. Future research should try to alleviate these difficulties using tagging and parsing. Furthermore, some tweets contained a dose of irony which complicated the classification task. Finally, we advise researchers to conserve as much as possible the language of the original tweets, as translations can introduce false synonyms and these translations may not always be of good quality.

3.2. Replicable semi-supervised approaches to a state-of-the-art detection of tweet stances[16]

Introduction: conducting stance detection in a replicable way

Our study contributes to the discussion surrounding reliable methods of stance detection, which, in addition to sentiment detection, also aims to determine the opinion (e.g. “favor,” “against,” and “none”) of a writer toward a specific target (Biber & Finegan, 1988; AlDayel & Magdy, 2021). Targets are, for instance, issues (e.g. legalization of abortion) or people (e.g. the current president). The application of stance detection is becoming a major endeavor in multiple research fields where the reliance on sentiment detection may be sub-optimal as it is not directly related to any entity or topic. However, stance detection is particularly relevant with respect to political, economic, and social questions where people can express differing opinions about a similar issue (e.g. voting object), event (e.g. mobilization) or person (e.g. president). The ability to detect stance from texts is particularly important from a social science and business perspective as it provides researchers and stakeholders with fundamental information about people’s opinions on a wide range of policy issues and entities. Furthermore, as texts are now an indispensable source of data for conducting social and economic research at an unprecedented reach, notably through social media data, it is increasingly important to strive for reliable methods to detect stance from texts. To date, opinion surveys are still the most reliable instrument for measuring opinion. However, decreasing response rates pressurize researchers to look for data sources that can be made comparable – serving as potential substitutes or complements – along important societal, political, and economic dimensions.

Recently, advanced machine learning approaches have been developed to offer new insights into stance detection (see review by Küçük & Can, 2020). Natural language approaches form the core of the methods in stance detection and strive to link language signals to better understand the positioning or polarization of speakers towards

[16] This chapter is a slightly adapted version of the article that is currently under review (minor revision and modification) as M. Reveilhac and G. Schneider: “Replicable semi-supervised approaches to a state-of-the-art detection of tweet stances”, Information Processing and Management.

particular issues (AlDayel & Magdy, 2021). These approaches have the advantage of offering reusable workflows (e.g. in the form of Python or R scripts) and reusable annotated datasets. However, these approaches also lack reliability (e.g. in providing no reasons for their findings) and replicability as their models tend to be very domain specific (Grimmer & Stewart, 2013: 268). These issues can introduce challenges to the reproducibility, replicability, and robustness of the research. Our methodological approach provides researchers with a workflow using existing tools that can be transposed to other domains with only little manual work, and with guidelines about which features are strong signals to stance.

Other studies rely on sentiment detection as a component to predict the stance. Some studies tend to display a confusion between sentiment and stance detection, either by considering sentiment as synonym or proxy of stance (Li & Caragea, 2019; Chauhan, Kumar & Ekbal, 2019). However, both measures convey different information about the textual data (Joseph et al., 2017). For instance, in AlDayel and Magdy (2021), only 35% of the “favor” stance tweets also had positive sentiment. Furthermore, Al-Ghadir, Azmi and Hussain (2020) showed that better classification results can be obtained if sentiment is not used. Ebrahimmi, Dou and Lowd (2016) propose a more nuanced view demonstrating that the best way to use sentiment to improve stance classification is through these multi-way interactions between input-sentiment-stance and input-target-stance. Sentiment strength and polarity are commonly detected using the lexica of words conveying negative and positive meanings and emotions (e.g. SentiWordNet, Bing Liu, and AFINN), which has become a common task of machine learning approaches (Verma & Thakur, 2018). Tonality lexicons have long been reliable tools for extracting the sentiment from texts, most notably for studying political communication from a linguistic perspective (Pennebaker, Francis & Booth, 2001; Taboada et al., 2011; Young & Soroka, 2012). Recent work by Li & Caragea (2019) built a multitask learning model that leveraged the sentiment of a tweet to detect its stance, which could provide a model’s overall F-score with a rating of 72.3%.

In this article, we propose to develop an alternative approach to stance detection that adopts a “rule-based” model (Patra, Das & Bandyopadhyay, 2016). We focus on the detection of stance in social media messages about several political issues while investigating the added value of linguistic markers compared to existing machine learning approaches. Our methodology involves the use of custom target dictionaries. To improve the scope of our dictionaries, we have also integrated "off-the-shelf" tonality lexicons in our workflow. This also includes custom lexicons of linguistic markers (notably, hedges, modals, and contrasts). Stance detection is conducted based on the composition of these signals at the tweet level, where the co-occurence of signals pointing to stance, sentiment, and linguistic patterns provide the final prediction of the stance.

Our objective is not to build a classifier with cutting-edge machine learning or the ability to set new performance records, but to offer easy access and feasibility to a state-of-the art performance of stance detection on new domains and tasks, and to understand how our rule-based approach compares to machine learning methods in terms of classification accuracy. We test our approach on the SemEval Task 6A dataset, which contains a relatively small number of manually annotated tweets about controversial topics. The small size of training datasets generally constitutes a problem for the performance of machine learning classifiers, which can be further complicated by the skewness of the data towards specific categories (Reveilhac & Morselli, 2022). Our methodology encompasses a "human-in-the-loop" component where theoretical and linguistic knowledge is brought by the researcher when computational approaches reach certain limits in recognizing fine-grained patterns. For instance, while words from the custom dictionary are extracted using a data-driven approach encompassing important contextual information, the choice to retain candidate words and the labelisation of the candidate words are decided manually and undergo an inter-coder reliability checking. However, the final decision about which features are strong indicators of stance relies on a data-driven approach based on the observed patterns from a training dataset, which enables us to assign weights from the different stance, tonality and linguistic signals in a data-driven manner. Our methodology further exploits the overall compositional aspect of a tweet by accounting for the distribution of these indicators and the morphological and syntactic pattern represented throughout the format of universal dependencies between pairs of words. In this view, our methodology explicitly models the semantic relatedness between a target and its context, which has been shown to be relevant for enhancing sentiment analysis (Xiang et al., 2022) as well as stance detection (Kyaw & Aungb, 2020).

More specifically, we address the following research questions: (1) What are important features of stance detection? (2) How can we build a framework that enables researchers to profit from integrating "off-the-shelf" lexicons into their workflows to conduct stance detection? (3) What are the advantages and disadvantages of rule-based methods as opposed to powerful machine learning methods in conducting stance detection?

To answer these research interests, our model makes use of a controlled dataset of tweets that are human-annotated for stance. More specifically, we rely on the SemEval-2016 task 6 dataset (Mohammad, Sobhani & Kiritchenko, 2017). It contains tweets on controversial subjects, some of which are very unbalanced in terms of stance (e.g. very few views opposing the idea that climate change is a real concern). We propose an experimental setting that entails several research steps: First, we assess the distribution of the linguistic features included in our model (namely, the linguistic markers and their combination with stance and tonality signals) across stance annotations. Second, we assess the importance of these features for stance detection and investigate which features are responsible for misclassification. Third, we derive a rule-based model for annotating a test dataset and compare the classification accuracy against machine learning methods.

Most prior works involving stance detection uses either machine learning or deep learning. In our article, we propose a rule-based model to extract stance from social media texts. Our models reads as a recipe with several steps involving both human intervention and interpretation (e.g. create custom dictionaries for each target and assign weights to each feature), as well as automatic classification (e.g. get feature importance and automatically tag the texts with the features for assigning a final stance). We replicate this recipe for each target separately and then assess the results of our classification model on each target using a test dataset. Although our approach has been elaborated on a specific domain for stance detection (namely, Twitter, the English language, and pre-defined targets), we were able to replicate our model on each target with acceptable levels of accuracy. To further improve the robustness and generalisability of our results, the proposed approach should be conducted on other datasets and on other types of data sources (for instance, newspaper comments or Reddit posts). Approaches to stance detection would benefit from more work in the direction of cross-dataset models. Future studies should therefore follow paths proposed by Schiller, Daxenberger and Gurevych (2021) and Ng and Carley (2022).

Theoretical Background

Approaches and challenges of stance detection

In social and political sciences, stance detection is a sub-task of opinion mining, which predominantly focuses on how social media users feel towards a certain topic or entity. A well-known example is the study of president approval (O'Connor et al., 2010; Pasek et al., 2019). This measure of approval aims to mirror the tradition in survey research of asking about the approval of political actors. However, surveys also aim at uncovering citizens' positions on a wide range of policy issues, so there is a need to develop reliable and replicable methods for stance detection that can be applied to multiple domains. This is particularly relevant for establishing the correspondence of social media measures with external data sources such as polls and opinion surveys on similar policy issues. It is also essential for assessing whether social media messages provide the same information as more traditional measures of opinions, or whether these messages represent another spectrum of public opinion.

In social computer sciences, stances are usually defined as text fragments explicitly or implicitly representing opinions or points of views with respect to a target (Rajendran, Bollegala, & Parsons, 2016). The text fragments can be social media messages, such as tweets or posts, but also political texts, such as parliamentary speeches. This textual data can thus serve to reveal a user's stance. Opinion mining typically consists of tonality detection and stance detection. Tonality detection mostly classifies the given text as generally positive, negative, or neutral. While tonality and stance may sometimes reflect similar patterns, this is not always the case, thus rendering sentiment analysis sub-optimal for social and political applications (Sen et al., 2020). Indeed, as Joseph et al. (2017) point out, while stance and sentiment are related, they are not the same: a negative sentiment of a text can be paired with a positive stance towards a particular target and vice versa. In addition to sentiment valence (positive, negative, or neutral) and intensity (low, medium, or high), approaches have also been developed to identify emotions (e.g. anger, pleasure, sadness) from texts. One of the most comprehensive and high-quality approaches is lexicon-based and is called LIWC, an acronym for Linguistic Inquiry and Word Count (Pennebaker, Francis & Booth, 2001).

The detection of stance, "usually considered as a subproblem of sentiment analysis and aims to identify the stance of the text author towards a target (an entity, concept, event, idea, opinion, claim, topic, etc.)" (Kücük & Can, 2020, pp. 12-2). As such, it can bring complementary information to sentiment analysis by considering not only the speaker's choice of message tonality but also the attitude taken towards an issue. Stance detection is complicated by the settings and conventions that apply on social media platforms. For instance, on Twitter, the stance of a user is usually expressed in very few words. This is less the case on Facebook, where users do not experience space restrictions. Furthermore, language on social media is heavily influenced by writing conventions (e.g. #, @, replies, etc.) and unconventional styles of communication (e.g. spelling mistakes, concatenated words, etc.).

Social media messages render the detection of stance even more complicated than more official texts (e.g. news articles or political speeches) given their brevity and unconventional writing style. The ever-increasing reliance on social media platforms to share opinions and interact with other users (e.g. politicians, media, activists, or engaged citizens) renders opinion mining a vital objective for enabling reliable measures of opinions (Sen, Flöck & Wagner, 2020; Kenneth et al., 2021). Interest in stance detection in social media messages has been growing since the SemEval 2016 competition (Al-Ghadir, Azmi & Hussain, 2021). According to the recent literature survey by Kücük & Can (2020) stance detection has mostly been investigated using supervised learning based on traditional features (such as n-grams, word embedding, and sentiment lexicons) and samples of human-annotated data. However, it is also not an easy task to generate enough training data to build efficient classification models. To address the scarcity of the annotated data needed to train accurate models, other techniques, such as transfer learning (Zarrella & March, 2016) and unsupervised learning (Xu et al., 2016), have been used to enrich the data with information related to the object of interest.

There are multiple approaches to stance detection with different types of targets, features used for detection, and various analytical approaches. Notwithstanding the merits of these studies, there are still major difficulties in distinguishing stance words corresponding to a "favor" or an "against" position. A major reason is that both positions tend to use similar vocabulary (e.g. the word "burqa" is used by opponents as well as by supporters of immigration or Islam, while referring to different argumentations and narratives). Another reason relates to the small size of manually labelled training data and the potential unbalanced distribution of stance towards a given policy issue (Reveilhac & Morselli, 2022).

Most approaches to stance detection are specifically limited to a narrow task and are costly to build, as "[a] separate stance classification model must be built for each target separately" (AlDayel & Magdy, 2021, p.7). A recent paper from Kannangara & Wobcke (2021), however, developed a new approach that needs only a few seed words for each issue and stance, as well as a sentiment lexicon to obtain a good level of accuracy. The authors show that using topic modelling in conjunction with sentiment detection can help to simultaneously identify target issues, sentiment, and stance on the issue from texts. By utilizing probabilistic topic modelling, their model outperforms supervised models of participating teams and the baselines of the SemEval 2016 competition and other deep learning models. Their model is particularly useful in cases where there is only one stance in the data (e.g. women's rights, animal rights), and in cases where it is difficult to find distinguishing stance words since favoring and opposing groups tend to use similar vocabulary. Our approach takes up this insight: the fact that certain entities are strong indicators of stance can be learned from a few instances, based on a small set of domain specific training material. For instance, the word "skeptics" in the context of climate change reliably points to positive stance on the need for environmental action, "prolifers" in the vicinity of evaluative language boosts the expressed stance. We have observed that many crucial patterns involve longer sequences of words, and have thus used syntactic dependencies. For instance, a positively connotated modal verb (can, could, must) with a negatively loaded verb dependent (e.g. die, refuse, abandon) leads to a strong feature indicating stance against an issue. We derive these patterns systematically by testing all possible combinations.

In addition to approaches in stance detection based on the content of a text, another strand of research has shown the merits of using social interactions and user network features for detecting stance (e.g. Bessi et al., 2016). However, the information about the user network is not always available. Overall, network features are superior to content (textual features) for most studies on stance detection, and supervised methods were found to be effective for different datasets (Aldayel & Magdy, 2019; Zhu, He & Zhou, 2020).

Stance detection from a linguistic perspective: rule-based method versus machine learning

In contrast to the notion of sentiment (or tonality) which has been extensively studied in stance detection (Kücük & Can, 2020), other markers of opinions, such as lexical and grammatical markers, remain understudied. However, these markers – such as stance adverbs, modal verbs, and hedges – have the potential to signal subjective and objective relations by serving as means through which a speaker expresses his point of view and attitudes toward a given issue (Ehret & Taboada 2021). We aim to operationalize linguistically motivated approaches to stance detection. First, modal verbs (Biber et al., 1999) are often used to express stance, particularly in objectively oriented registers like scientific writing (Hyland, 1996). Second, explicit personal statements (e.g. I, we) followed by normative verbs indicate statements where the speaker is personally involved, has no doubt and has an urge to tell the public what to do.

Our method is understood as “rule-based” insofar as it encompasses both human-coded rules and text classification methods that identify rules to elaborate our own model. The scope of our approach to stance detection is thus lexically-based and quantitative.

To the best of our knowledge, not many computational attempts have been performed using these linguistic markers for stance detection. However, the choice of these markers can be decisive in conveying the speaker’s opinion and attitude towards an issue or an entity, such as supporting environmental taxes or supporting a presidential candidate for (re)election. There is thus a need to understand the extent to which these different markers can serve the goal of detecting stance in a text, thereby complementing content approach to stance detection. These lexical features can bring a lot of information about stance towards a given target and complement other content (e.g. hashtags and target related keywords) and sentiment (e.g. tonality) features.

We hypothesize that a stance detection approach with a strong focus on well-studied linguistic features and patterns can play a decisive and complementary role in machine learning and dictionary approaches and can offer a simple way to adapt to new domains and tasks without much manual effort (no big training samples, nor extensive mathematical background). These complementary features can be especially useful for the detection of stance in short messages (e.g. tweets, Facebook, or blog posts), which are often written in an abbreviated or expressive style and where data sparseness is particularly acute. To elaborate efficient rule-based models that achieve a good level of accuracy, we need to have a good set of linguistic markers that can complement tonality and target related keywords. The approach proposed in our study, which includes the use of dependency information and the use of linguistic markers, has the potential to complement studies relying on machine learning methodologies and on the selection of stance features. Some of these studies demonstrate high classification performance, such as the study of Al-Ghadir, Azmi and Hussain (2020). The findings of our approach can also inform future studies which aim to assess the optimal number and type of features relevant to a given target, as proposed in the study of Vychegzhanin and Kotelnikov (2021).

Rule-based methods can be useful in addressing an important difficulty in methodologies that require the annotation of large datasets. A major advantage with this approach is that it enables us to conduct stance detection on multiple targets when there are few resources available. Indeed, even human annotation of social media content does not always capture stance as measured by public opinion polls. For instance, Joseph et al. (2021) compared individuals' self-reported stances to the stance inferred from their social media data and identified three factors leading to the disconnect between text and author stance: temporal inconsistencies, differences in how annotation tasks are constructed (Joseph et al. 2017), and measurement errors from both survey respondents and annotators.

Doing stance detection within a replicable framework

In those social science disciplines that rely on survey data, the terms of validity, reliability, bias, precision, and construct are prominent in evaluating the quality of measurements. In computer sciences, the prominent terms relate to correctness, accuracy (which usually includes precision and recall), efficiency, and reliability (Ladd et al., 2020). To repeat an analysis and to reach similar conclusions constitutes reproducible research. However, Baker (2016) conducted an influential survey with 1,500 scientists showing that reproducibility is the exception rather than the norm, because "more than 70% of researchers have tried and failed to reproduce another scientist's experiments" (Baker 2016: 452). In fact, we are faced with a reproducibility crisis. This crisis has been linked to several factors, most notably to incomplete method descriptions, unavailable raw data, and incomplete, undocumented and/or unavailable code. In contrast to reproducibility, replicability "refers to instances in which researchers collect new data to arrive at the same findings as a previous study" (National Academies of Sciences et al., 2019, p.43).

Reproducibility and replicability received increased attention as the use of computational tools expanded and several attempts were made to assess non-reproducibility or non-replicability. Instances of such attempts can be seen in social science studies (e.g. Camerer et al., 2018). The field of computational social sciences has expanded considerably since

the last decade, enabling researchers to improve their understanding of important societal phenomena to an unprecedented reach. However, this research is based on so-called "found" data which are inevitably subject to concerns about their internal and external validity, data availability, and privacy issues, thereby often impeding the reproducibility and replication of results (Lazer et al., 2020).
Our study aims to propose a methodology that promotes a replicable methodology for stance detection that can be easily replicated in other research subjects. In particular, our methodology insists on the transparency of the different analytical steps and relies on a long tradition of theoretical insights, most notably the use of sentiment dictionaries (e.g. Tausczik & Pennebaker, 2010) and the expression of sentiment through hedges or modal verbs (e.g. Ehret & Taboada, 2021). Furthermore, to maintain comparability, we exploited empirical data that have been employed in several studies. This enable us to further strengthen the trustworthiness of our methods and findings.

Materials and Methods

SemEval dataset

Training data for target detection can be expensive to generate. In addition to replicability and comparability, this is the second reason why we relied on a stance dataset that is annotated using a predefined set of labels (favor versus against and none). We use the SemEval-2016 (task 6) stance dataset (Mohammad, Sobhani & Kiritchenko, 2017), which covers a set of five different topics, namely: "climate change is a real concern," "atheism," "legalization of abortion," "feminist movement," and "Hillary Clinton." The inter-annotator agreement on this dataset is 81.85% and it is partitioned into training (70%) and test (30%) sets based on the timestamps of the tweets (Mohammad et al., 2016). To develop and test our rule-based model, we relied on these annotated training and held-out test datasets that are provided by the SemEval team. The training and testing dataset available do not include the possibility of integrating the users' network.

Workflow in several steps

Table 3.2.1 displays the steps from data preparation, data coding, elaboration of the rule-based model, and data analysis. The initial steps (steps 0 and 1) cover the detection of the tweet's target and the creation of a training and testing dataset. These steps were already conducted by the research team that made the data available. The subsequent steps form the core of our methodological approach. Our methodology acts as a "recipe" that can be reproduced step by step. Steps 2 to 3 prepare the necessary dictionaries with relevant target terms and linguistic features. Step 4 extracts the dependency information for the tweets. Step 5 extracts the tonality, stance, and entity information from the tweets using existing and custom dictionaries. It also extracts the linguistic features from the tweets. Steps 6 to 9 elaborate tests to detect key features and to establish decision rules. Step 10 applies the rule-based model to the test dataset.

Table 3.2.1: Successive steps of data preparation and analyses (The steps 0 to 9 are conducted on the training dataset, whereas the final accuracy of the model is computed on the test dataset).

Stages	**Explanation**	**Method**
Data preparation	~~step 0: detect the target of the tweet~~	(already conducted by SemEval)
Data preparation	~~step 1: annotate stance~~	(already conducted by SemEval)
Data preparation	step 2: build a target relevant custom dictionary	extract relevant words for main entities, entities, stance, and tonality (remove stop-words only for this step).
Data preparation	step 3: build custom lexicons with linguistic features	lexicons should entail hedges, modals, quantifiers, negators, passive, auxiliary verbs, etc.
Data coding	step 4: get dependency information	get the dependency information by applying the *Stanford Parser* to get word pairs in dependency relations
Data coding	step 5: extract the tonality, stance, entity, and linguistic features from the texts	annotate each text into a combination of 'signals' for tonality, stance, entity, and linguistic features
Rule-based information	step 6: detect important word features	retrieve distribution of features across stance & use conditional inference tree to assess each feature's weight
Rule-based information	step 7: detect which features are misleading	conduct logistic regression to detection features contributing to mis-classification
Rule-based information	step 8: establish decision rules	a) assign score (e.g. between -3 and +3) to each "signal" and calculate an overall score for each text b) assign a "none" stance if the tweet does not contain any words from the custom dictionary
Rule-based information	step 9: add corrections to incentivize the classification	correct the texts containing "very strong signals" (e.g. unambiguous hashtags)
Analyses	step 10: apply our model on the test dataset	compare classification matrices against ML models (e.g. naive bayes and random forest)

Step 2 involves building individual custom target vocabularies. To do so, we used tf-idf, which means term frequency–inverse document frequency, a classical keyword detection measure which is intended to reflect how important a word is to a document in a corpus. We also extracted words associated to social media conventions (e.g. # and @ appearing more than twice) to extract target specific words. We also considered common words that are used for different targets, as similar words can have different meanings depending on the context of discussions. We further extended the vocabulary by searching for synonyms and antonyms using WordNet (Fellbaum, 2005). We manually checked the target-specific vocabularies and removed unspecific words. Our final custom dictionary contains 2251 words (see Table 3.2.2 below).

Once the custom dictionary was ready, we manually assigned each word a score on several dimensions: "tonality," "stance," "entity," "main entity," and "no doubt." The "tonality" dimension can take the values 1 (positive), 0 (neutral), and -1 (negative). The "stance" dimension takes the values of 1 (favor), 0 (none), and -1 (against). The "entity" dimension indicates whether the word is a central aspect of the target specific debate (value of 1) or not (value of 0). The "main entity" dimension takes 1 if the word is an entity with a positive stance and takes -1 if the word is an entity with a negative stance. We thus differentiate between "entities", which are not specific to the target (e.g. Pope in climate change discussions), and "main entities," which have a direct relation to the target (e.g. Pope in atheism discussions). Finally, the "no doubt" dimension assigns a word with 1 if it unambiguously indicates a positive stance and -1 if it unambiguously indicates a negative stance. The inter-annotator agreement is above 0.9 for each target and dimension of our custom stance dictionary. Two annotators were involved in the coding process. A first annotator was taught how to code the different dimensions. Furthermore, an expert annotator who elaborated the coding scheme worked on the project. Both annotators independently coded all instances of the dictionary.

Table 3.2.2: Distribution of words for each dimension of the custom dictionary.

	#words	Stance			Tonality			"no doubt"		entity	"main entity"	
		favor	***none***	***against***	***favor***	***none***	***against***	***favor***	***against***		***favor***	***against***
Atheism	368	116	84	168	150	123	95	35	28	59	18	3
Climate Change is a Real Concern	289	169	93	27	64	139	86	36	14	87	4	0
Feminist Movement	593	143	264	186	178	132	283	44	66	63	13	1
Hillary Clinton	546	63	284	199	135	219	192	24	49	166	7	9
Legalization of Abortion	455	82	191	182	156	196	103	39	81	139	2	3

To get linguistic features (step 3) that can influence the stance of a tweet, we built lexicons that point to the following aspects: “non-specificity,” “personal view,” “vagueness,” “frequency adverbs,” “quantifiers,” “degree,” “introductory verbs,” “modal adverbs,” “modal nouns,” “evidences,” “modal adjectives for certainty,” “modal adjectives for uncertainty,” “positive modals,” “negative modals,” “contrast,” “negators,” “eventuality,” “neutral stance adverbs,” “negative stance adverbs,” and “positive stance adverbs.” Our list of features only includes 1-grams, but future works could also include expressions (e.g. “in my opinion”).

The custom dictionary and the custom lexicons with linguistic features were used along with the dependency information of the tweets (step 4). To get the dependency information, we used the Stanford Parser as implemented in the CoreNLP pipeline (Manning et al., 2014) and we relied on the wrapper in the R programming language by Arnold and Tilton (2016). The parser provided us with word pairs in dependency relations that consisted of two component words, one of which was present in either the custom dictionary, the custom lexicons, or the “off-the-shelf” tonality lexicons. Indeed, to increase the list of words that are relevant for tonality detection, we relied on existing lexicons, namely Lexicoder (Young & Soroka, 2012) and NRC (Mohammad & Turney, 2013). We only used words that were not already contained in our custom vocabulary. Each dependency of the tweet was therefore coded according to several “signals” (step 5). We also added some handcrafted rules. For instance, using tonality we get the following rules: Not + NEG = POS, Not + POS = NEG, POS + POS = POS, NEG + NEG = NEG. We also added several stance-tonality and lexical-tonality combinations. Indeed, our rule-based model does not rely on tonality as a feature per se but only in combination with stance and or entity. Annex 3.2.1 provides a glossary of the features.

To elaborate our rule-based model, we conducted two additional steps (steps 6 and 7). Firstly, we displayed the distribution of features according to stance and we combined conditional inference trees with the analysis of random forests to investigate the interplay between the various stance, entity, tonality, and linguistic features in stance recognition. To assess the importance of each individual predictor, we used the default settings for growing trees in the R language partykit package (Hothorn & Zeileis, 2015). Secondly, we identified the features responsible for misclassification of stance. Based on the findings from these steps, we assigned each “signal” a weight (from 3 to -3) to calculate an overall stance score (step 8). The choice of the value size (absolute numbers between 0 and 3) was based on data-driven observations (e.g. feature importance and misclassification analysis). The direction of the value (positive or negative sign) was based on theory-driven insights (e.g. "reject_neg" means that there is a rejection word followed by a negative, thus suggesting an overall positive language). Annex 3.2.2 provides the full list of features and their associated weights.

We then labelled the tweet with the associated feature weights and calculated an overall score (step 8). If the overall score was above 0, we assigned the tweet with a "FAVOR" stance. We assigned the tweet with an "AGAINST" stance if the overall score was below 0. We assigned a "none" stance if the overall score was 0 or if the tweet did not contain any words from the custom dictionary. We added correction rules to incentivize the classification toward very strong signals (step 9). The different correction rules follow a cascade order and include four steps. The first step gives a "none" stance to the tweets that do not contain any stance related signals (e.g. entities, main entities, stance). Then, the steps 2 to 4 follow a hierarchy of stance signals: the second step sums up the features of the main entity, the third step sums up the features for stance in the same way, and the fourth step gives priority to the unambiguous signals (namely, the "no doubt" feature). Annex 3.2.3 provides the full description of the correction rules.

Finally, we applied our model on the test dataset (step 10) and compared the classification accuracy against machine learning models. We compared the classification performance of our rule-based model against machine learning approaches using the F1 score.

Results

Features' distribution by stance and their conditional importance (training dataset)

A first interest is related to the proportion of these markers across the three stances available in our training dataset. Figure 3.2.1 shows that positive and negative evaluative words (namely, stance and tonality words from the custom and "off-the-shelf" dictionaries) are the clearest and strongest markers of expression of stance. Furthermore, we also note the usefulness of adding linguistic markers to make the direction of the stance more precise. For instance, the feature "main_entity_con_no_HoC_pos" (tweets mentioning a main entity opposing the target with a positive tonality without using hedges nor contrast) is clearly helpful to correctly identify an unfavorable stance. Another example states the merits of including linguistic features indicating "rejection_pos" (tweets entailing a rejection associated with a positive tonality) to identify supportive stance. Furthermore, we see that a favorable stance is expressed with evidentiality (evidences_pos, evidences), while stance against is less argumentative and is more based on personal preferences (personal_view). The fact that modal verbs appear in our lists of top features indicates that we catch the long tradition of stance description in corpus linguistics and pragmatics (eg. Biber et al. 1999), but also that features expressed without the slightest doubt (no_doubt_pos, evidences) are particularly reliable. Domain-specific entities suggest that lesser amounts of manual annotation allow us to significantly profit from contextual indicators.

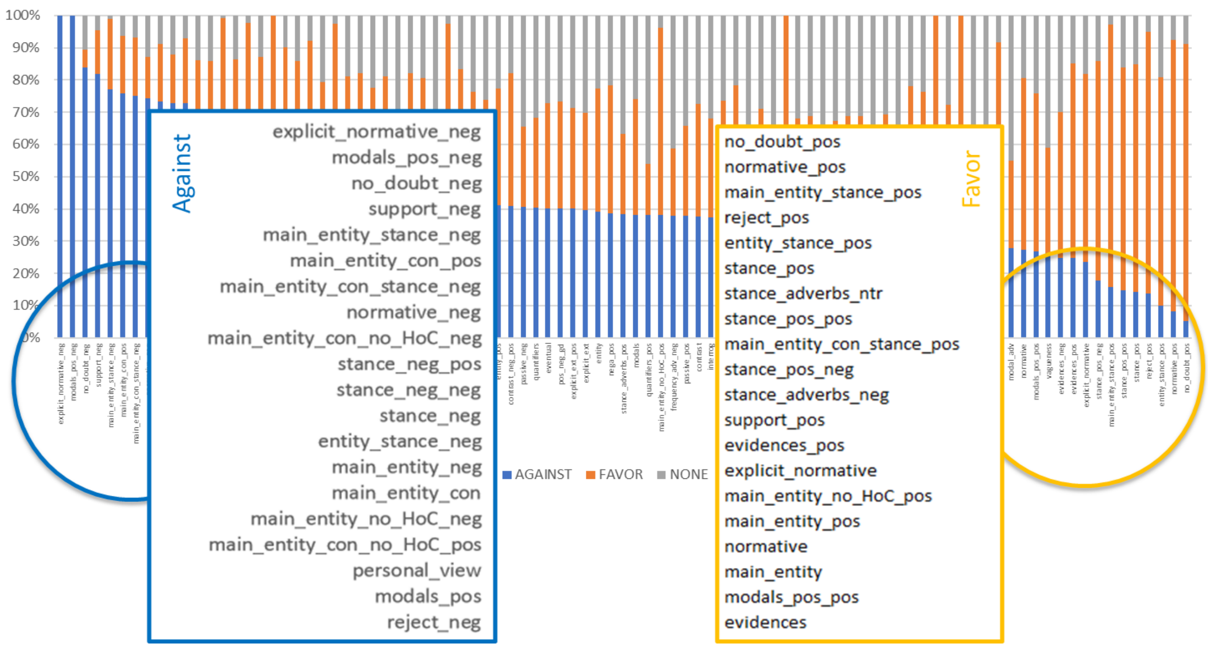


Figure 3.2.1: Distribution of features by stance (training set).

To analyze how the features from our rule-based model contribute to the prediction of stance, we utilized conditional inference trees, which is an ensemble method for classification based on an aggregate of single trees. This method is a powerful tool for analyzing complex interactions between many different predictor variables (in our case, features) and for assessing their importance. The dependent variable of our model is the stance category (three values: favor, against, and none). The predictor variables have similar scales of measurement (token frequencies). In this spirit, we constructed a feature matrix with the individual texts as rows and each of the variables (e.g. frequency counts of the various markers) as columns.

The ranking of variable importance presented in Figure 3.2.2 shows that (unambiguous) stance words from our custom dictionaries (e.g. "no_doubt_pos"), as well as the fact of explicitly mentioning a target entity (e.g. "entity," "entity_stance_neg"), are by far the most important predictors, followed by — at a considerable distance — linguistic markers, such as negators, normativity, vagueness, and modals. Typical bag-of-word classification for stance detection includes entities (as content words), but when it comes to linguistic markers they are partly excluded as stopwords. When they are included, their context, and particularly their function, is not included appropriately. The ranking of the importance of variables clearly shows that linguistic markers matter. For example, those pointing to the absence of hedges (like modal verbs or downtoners like maybe) or contrast ("_no_HoC_") matter in the classification of stance. Therefore, tweets with a direct positioning (without the weighting of arguments or pros and cons) should be easy to classify according to stance.

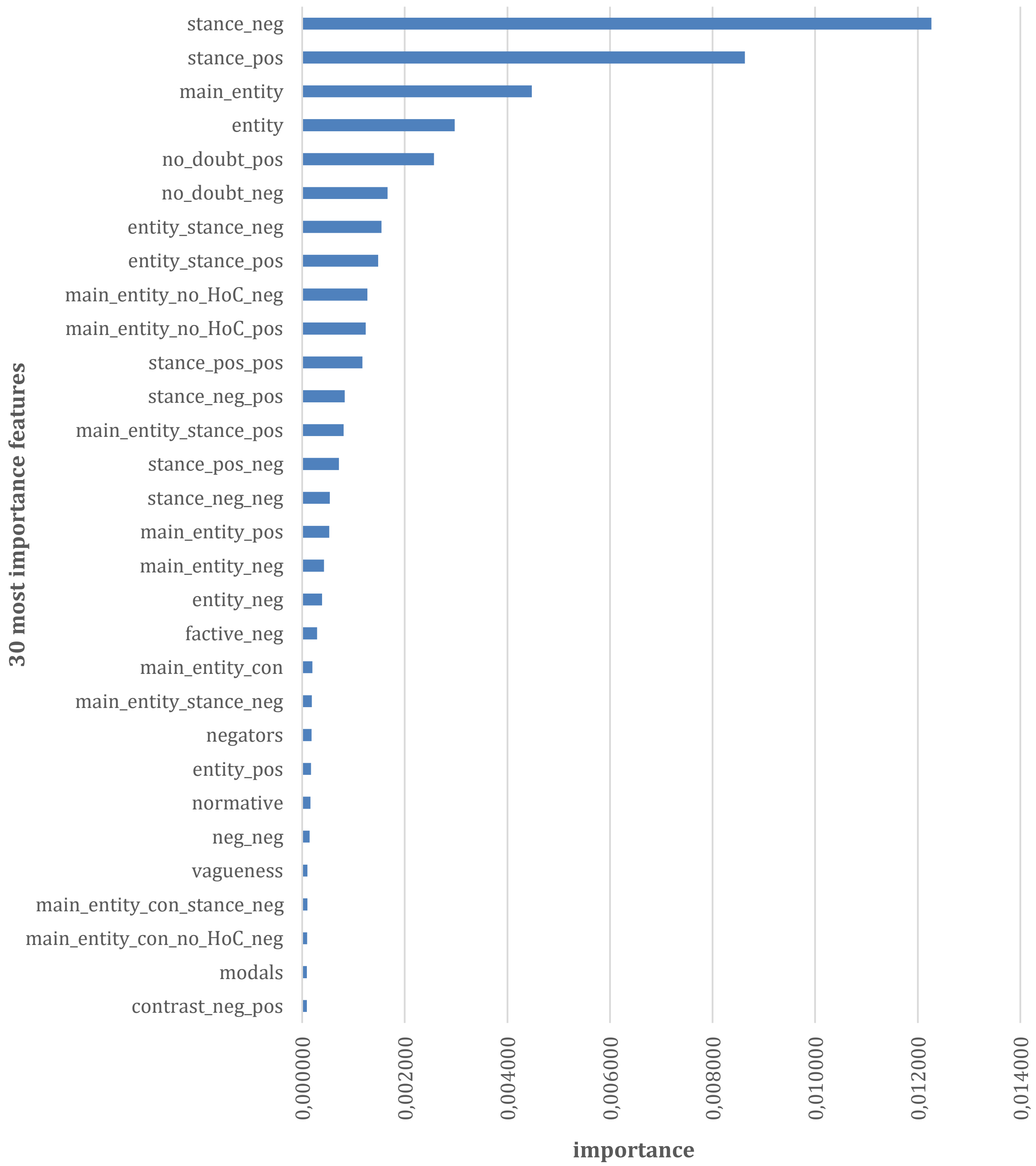


Figure 3.2.2: Conditional importance of the most important features (training set).

Table 3.2.3 shows a random sample of tweets coded according to our rule-based model. It involves several pieces of information: the target of the tweet (column 2), the original tweet (column 3), the coding derived from our rule-based model (column 4), the predicted stance from our model (column 5), and the true stance (column 6). For instance, the text 1406 (2278) about the presidential election was labeled as having no stance (NONE) by our model while it was manually assigned a negative (AGAINST) stance. Here, no word in the tweets clearly points to positive attitude, and the wordplay of hill as a nickname for Hillary involves so many inferences that an ML system can hardly pick up the signal. We observed an important level of confusion between tweets which do not mention the target at all (tweets with indirect favorable or unfavorable stances) and tweets that mention the target but do not express a stance towards them (considered as neutral tweets by our model, such as text 1406 (2278)). Another example from Table 3.2.3, one in which the model and the manual annotation agree, is the text 3 (1002). It contains several positive references associated with stance and entity words, making this tweet easily classifiable.

Table 3.2.3: Random sample of tweets coded according to our rule-based model and compared to the manually coded "true" stance.

Tweet id	**Target**	**Original tweet**	**Coding**	**"Predicted" stance**	**"True" stance**
3 (1002)	Climate Change Is a Real Concern	We need integrated #science with #indigenous knowledge to understand & adapt to #CFCC15	1:explicit normative explicit_normative 4:stance_pos normative normative_pos 6:pos_pos 7:pos_pos 9:normative introductory_verb 12:stance_pos entity entity_stance_pos	FAVOR	FAVOR
14 (1012)	Feminist Movement	Stupid Feminists, the civilization you take for granted was built with the labour, blood sweat and tears of men.	1:main_entity main_entity_con main_entity_neg main_entity_con_neg main_entity_no_HoC_pos main_entity_no_HoC_neg main_entity_con_no_HoC_pos 2:stance_neg stance_neg_pos main_entity main_entity_con main_entity_pos main_entity_stance_neg main_entity_con_pos main_entity_con_stance_neg 5:explicit_ext 17:pos_neg_gd 19:entity entity_neg	AGAINST	AGAINST
22 (102)	Atheism	Blessed are the peacemakers, for they shall be called children of God. Matthew 5:9 #scripture #peace	1:stance_neg 2:factive factive_pos 4:pos_pos 6:explicit_ext 7:modals 10:entity 12:stance_neg entity entity_stance_neg 16:stance_neg stance_neg_pos entity entity_pos entity_stance_neg 17:stance_neg	AGAINST	AGAINST
1406 (2278)	Hillary Clinton	over the river and through the woods, and UP WITH HILL WE GO! Yass #HillaryClinton #Hillary	12:explicit 14:main_entity 15:main_entity 16:main_entity	NONE	AGAINST
1711 (2555)	Legalization of Abortion	#ProLifeYouth know that life begins at conception.	1:stance_neg no_doubt_neg 4:entity 5:factive factive_pos 7:stance_pos	AGAINST	AGAINST

It is also important to assess which features are responsible for stance misclassification. Indeed, it is possible that features which are most important for detecting stance are also partly responsible for misclassification issues. To test this hypothesis, we relied on a logistic regression model where the dependent variable reads as follows: 1 for matching between the predicted stance and the true stance and 0 otherwise. Table 3.2.4 displays the result of the logistic regression only for coefficients which are statistically significant (all other features are included in the regression but are not displayed for readability reasons).

We see that one of the unambiguous features correctly detects the stance (e.g. “no_doubt_neg”) and that ambiguous features (e.g. “eventual” and “vagueness”) strongly contribute to misclassification. It is therefore important to identify such linguistic features when implementing stance detection. We also see that some features which were identified as most prevalent in detecting stance (according to the previous analyses in Figure 3.2.2) also significantly contribute to misclassification. For instance, the feature pointing to a main entity with a usual negative stance, followed by a positively loaded word without the inclusion of hedges nor contrast (“main_entity_con_no_HoC_pos”), has a high impact on misclassification. This might point to tweets using sarcasm and irony.

Table 3.2.4: Logistic regression analyzing which features are contributing to stance misclassification (minus sign).

	Correct classification (training dataset)
(Intercept)	0.71 (0.10) ***
stance_neg	0.18 (0.08) *
entity	-0.14 (0.05) *
no_doubt_neg	0.87 (0.17) ***
no_doubt_pos	0.86 (0.18) ***
normative_pos	2.20 (1.12) *
eventual	-0.37 (0.17) *
vagueness	-0.59 (0.17) ***
main_entity_no_HoC_pos	0.36 (0.17) *
main_entity_con_no_HoC_pos	-1.03 (0.44) *
main_entity_con_no_HoC_neg	1.02 (0.49) *
Pseudo R2	0.08
Num. obs.	2814

Classification by target and comparison with machine learning models (testing dataset)

The results of our rule-based model on the test dataset are detailed in Table 3.2.5. Table 3.2.5 provides the classification performance measures for the all the targets and for the individual targets. It also displays the evolution of the performance metrics of three models, showing how much improvement is achieved in the stance task for various targets with various features: stance and entity features only (Model 1); stance, entity, and tonality features (Model 2); stance, entity, tonality and linguistic features (Model 3). In general, we observe a pattern of improvement across the Models 1 to 3 with the addition of tonality and linguistic features (both in accuracy and F1 scores). The only exception is the target "Climate change is a real concern," which could be explained by the very skewed distribution of the stance categories towards the "favour" stance. The biggest improvement is found for the target "Feminist movement", which could be explained by the prevalence of ironic and sarcastic language that is difficult to capture with only stance and tonality features.

The overall accuracy of our model for all targets reaches an accuracy of 75% and a F1 score of 0.72. We obtained good accuracies (from 67% to 89%) and F1 scores (from 0.63 to 0.83) for every target separately. To provide a comparison baseline, we reviewed existing studies that have relied on ML to classify the stance of tweets from the same SemEval Task A dataset. Annex 3.2.4 displays the results of previous studies using the SemEval dataset and covering a wide range of machine learning algorithms[17]. Our approach reaches an acceptable performance compared to the average F1 score of the reviewed studies (see last column of the Table in Annex 3.2.4). For instance, the study of Vychegzhanin and Kotelnikov (2021) also relied on various stance and linguistic features and obtained a F1 score of 0.71, demonstrating that the most useful feature types for stance detection are n-grams and linguistic features (e.g. count of negations), while the most useless feature type is stylistic features (e.g. number of punctuation marks). There are, however, a few studies that demonstrate a higher level of performance than our proposed approach. For instance, the study of Al-Ghadir, Azmi and Hussain (2020) also proposed a methodology and involved an innovative feature selection (i.e. for the creation of dictionaries). They achieved a macro F1 score of 0.76, thus topping the current state-of-the-art performance by 2.01%, irrespective of whether the sentiment information was included or not.

[17] For a complete and updated review of stance detection studies, refer to Ghosh et al. (2019), who provide a survey of the performance of multiple supervised machine learning algorithms on the SemEval dataset. Comparative surveys for ML in stance detection were also published by Küçük and Can (2020) and AlDayel and Magdy (2021).

Table 3.2.5: Performance metrics (accuracy, average precision, average recall, average F1) for the stance classification results on the test dataset.

		Model 1: Stance + entity features	**Model 2: Stance + entity + tonality features**	**Model 3: Stance + entity + tonality + linguistic features**
Overall	Accuracy	0.71	0.73	0.75
	Precision	0.71	0.73	0.74
	Recall	0.68	0.69	0.71
	F1	0.68	0.70	0.72
Atheism	Accuracy	0.85	0.89	0.89
	Precision	0.78	0.82	0.82
	Recall	0.78	0.83	0.84
	F1	0.78	0.82	0.83
Climate Change	Accuracy	0.79	0.79	0.81
	Precision	0.67	0.67	0.68
	Recall	0.76	0.73	0.73
	F1	0.71	0.70	0.70
Feminist Movement	Accuracy	0.60	0.63	0.67
	Precision	0.62	0.65	0.67
	Recall	0.56	0.58	0.61
	F1	0.56	0.59	0.63
Hillary Clinton	Accuracy	0.65	0.68	0.71
	Precision	0.64	0.68	0.69
	Recall	0.62	0.64	0.66
	F1	0.62	0.65	0.67
Legalization of Abortion	Accuracy	0.72	0.73	0.74
	Precision	0.72	0.73	0.73
	Recall	0.67	0.68	0.68
	F1	0.68	0.69	0.70

Discussion of the Main Findings and Implications for Research

Summary of the Main Findings

In our article, we provide an experimental setting for uncovering the importance of various features for stance detection, including stance and tonality signals, as well as linguistic markers. To do so, we adopted a rule-based approach that can complement more sophisticated machine learning and network analyses. Our aim was not to construct a state-of-the-art classifier, but to raise awareness about complex linguistic relations that are rarely documented in stance detection (Ehret & Taboada, 2020). By doing so, we contributed to the identification of linguistic aspects that can hinder replicability and the high quality of research outputs in stance detection. This is especially important in the context of social media messages, such as tweets, which are noticeably short and generate sparse data problems, thereby justifying the study of the added value of linguistic features. Our major finding shows that our methodology performs better than the best pure machine learning approaches at SemEval 2016. Additional findings show that positive and negative evaluative words (namely, stance and tonality words from the custom and "off-the-shelf" dictionaries) are the clearest and strongest markers of expression of stance. Furthermore, we also note the usefulness of adding linguistic markers to precisely identify the direction of a stance. The ranking of variable importance clearly shows that linguistic markers point to the absence of hedges or contrast matter in the classification of stance. This could be particularly useful to address complex expressions of opinions, such as irony.

A close reading of several tweets misclassified by our rule-based model shows that each target highlights different challenges. For instance, climate change had very few "against" stances in the training set, which rendered the identification of different stances more challenging when relying on dictionaries. Furthermore, tweets related to feminism and Hillary Clinton contained a considerable proportion of ironic and sarcastic expressions. Moreover, tweets about the legalization of abortion posed challenges for identifying the neutral stance, because a neutral stance is rare in such a context. This could be improved by paying attention to events. While we could have added event as a variable, this is a double-edged sword: while improving performance, adapting closely to the event may lead to overfitting and incur the risk of jeopardizing our intended robustness.

Theoretical and Practical Implications

Theoretically, the proposed study has several implications. First, it proposes to assess the contribution of linguistic markers for conducting stance detection, notably by showing how these markers can complement prior work that have used external resources, such as sentiment lexicons. Second, we propose a rule-based model that provides an additional strategy to conduct stance detection compared to the most applied supervised ML approach. Our proposed framework has the potential to address one of the main drawbacks of ML approaches, which consists of the need for large amount of annotated data (Reveilhac & Morselli, 2022). Third, our approach proposes to support stance detection using a tonality lexicon, while considering tonality features as useful but insufficient to detect stance (Joseph et al., 2017). We therefore contribute to studies elucidating the interaction between stance and tonality (Sobhani, Mohammad & Kiritchenko, 2016; Li & Caragea, 2019). Fourth, our framework exploits the universal dependencies between pairs of words, thereby explicitly modelling the semantic relatedness between a target and its context and using it as relevant information for enhancing stance detection (Kyaw & Aungb, 2020).

Our results also have practical implications for research on stance detection. First, the proposed methodology is easily replicable across several targets and, thus, produces a replicable solution to stance detection. Most notably, the proposed model delivers a transparent recipe composed of the elaboration of custom dictionaries from textual extraction metrics and manual enrichment, the identification of relevant linguistic features through state-of-the-art methods (e.g. logistic regression), and the application of a rule-based method. The dialectic iteration, back and forth between manual annotations and computational methods, further enables researchers to get to know the corpus better while interacting with it. The obtained findings show the opportunities offered by the reliance on linguistic features to conduct stance detection, which can complement existing ML approaches. Second, our methodology integrates external resources that have a long tradition in tonality detection (Young & Soroka, 2012; Mohammad & Turney, 2013) and that can be adapted to the domain specificity and used as combined features. To the best of our knowledge, only one study relied on the creation of a custom stance lexicon as part of their workflow for conducting stance detection (Li & Caragea, 2019).

Concluding Remarks and Outlook

We have presented a replicable and transparent approach to stance detection, which is based on a linguistic motivation and has a state-of-the-art performance. Let us summarize our answers to the research questions that we have addressed:

(1) What are important features of stance detection?

Unambiguous stance words from our custom dictionaries (e.g. “no_doubt_pos”), as well as explicit mentions of a target entity, are the most important predictors. But linguistic markers, such as negators, normativity, vagueness, and modals, including their functions. For instance, the knowledge that modal verbs and downtoners add vagueness, which is specific in our approach, are significant factors in stance detection. For an overview, refer to Figure 3.2.2. We nonetheless show that the most important predictors can also be responsible for mis-classification (see Table 3.2.4), thus underlying the importance of using multiple predictors for improving stance classification. We have also learned that the linguistic features which we systematically include, such as modal verbs (Biber et al., 1999; Hyland, 1996) or personal involvement, are important indicators of stance.

(2) How can we build a framework that enables researchers to profit from integrating "off-the-shelf" lexicons into their workflows to conduct stance detection?

Stance and features combined with tonality words from the custom and "off-the-shelf" dictionaries are the clearest and strongest markers of expression of stance. Compared to a large body of studies on stance detection, we only considered tonality words in combination with other features and not as standalone parameters for classifying the tweets. This is because sentiment (or tonality) detection does not equate to stance analysis (Joseph et al., 2017). Adding linguistic markers helps us to make the direction of the stance more precise, see again Figure 3.2.2. For instance, the feature "main_entity_con_no_HoC_pos" (tweets mentioning a main entity opposing the target with a positive tonality without using hedges or contrast) is also among the most important features. Off-the-shelf dictionaries add a robust back-off for all words in which no customized entry has been made. Our methodology is easy to use and replicate in other domains. Another important advantage is that it relies on existing and validated tools (e.g. "off-the-shelf" lexicons) as well as established linguistic knowledge, while requiring only a limited amount of training data, little manual work, no development of complex algorithms, and a "human-in-the-loop" component.

(3) What are advantages and disadvantages of rule-based methods, as opposed to powerful machine learning methods, in conducting stance detection?

An advantage of our method is that it enables researchers to produce robust and replicable results in stance detection adapted to the domain, and, at the same time, offers transparent insights to the domain expert in the iterative development. As such, it has the potential to counteract the increasing loss of public trust in research from the Humanities and Social Sciences fields (Yong, 2018). However, our approach has the disadvantage that it needs to be adapted to perform best, even though the inclusion of off-the-shelf dictionaries offer a baseline. For instance, we could have placed the emphasis on additional tonality features such as emojis. We could also have focused on emotion indicators as developed in the LIWC dictionary. Furthermore, our approach still contains other limitations from a linguistic point of view. For instance, it does not allow us to identify sarcasm and irony, important aspects for expression of stance. Here, works from Potamias et al. (2020) can help to identify these forms of figurative language relying on advanced deep learning methodologies. Furthermore, the opinion, evaluation, or stance conveyed in a text might not always be the writer's own opinion (e.g. in cases where the user aims to report on an event without taking a position). Future work could also identify these specific cases, for instance, by looking at grammatical markers in the texts. Our study encourages the inclusion of other determinants of stance detection such as question marks. However, this specific feature can also be rhetorical or indicate a monologue from a unilateral communication style. The subtle interplay between stance, tonality, and irony makes it difficult to rely on a "one-size-fits-all" approach. The inclusion of and adaptation to particular events, and its effects on robustness, is a further avenue for research. Ideally, stance detection should also use a combination of approaches. For example, one can combine content and network approaches (Aldayel & Magdy, 2019; Lynn et al., 2019).

Going beyond the present application, we would like to suggest future paths for development. First, our methodology could be extended in the direction of active learning, which is a special case of machine learning in which a learning algorithm can interactively query a user to label new data points with the desired outputs. Second, an additional but related line of inquiry could go in the direction of argument mining, exploring more complex relations of stance propagation which is not only target-related stance, but also how arguments are backed up by other users (Zubiaga et al., 2018). In line with our application and with the proposed paths for future research, we believe that interdisciplinarity between computational sciences, insights from corpus linguistics, and social sciences approaches is indispensable to developing tools and guidelines that can address relevant research endeavors in the field of opinion mining.

Social media are increasingly relied on by a sizable world population. The focus of this study was to detect the stance of a specific microblogging social media platform, Twitter. To maintain comparability of the performance measures of our approach with previous studies, we relied on a single dataset, SemEval. However, this dataset is just one dataset and additional research is needed to assess the transferability of our approach to various societal and political domains. For future work, we plan to check the transferability of our approach to other data sources (e.g. newspaper comments, Wiki comments, Facebook posts, etc.) and to other targets (e.g. social movements)

CHAPTER 4. WHAT CONTENT AND PUBLIC ARE AVAILABLE ON SOCIAL MEDIA AND HOW DO THEY INTERACT WITH PUBLIC OPINION?

4.1 Digital Shift in Swiss Media Consumption Practices[18]

Introduction

In the spring of 2018, the oldest ad agency in Switzerland, Publicitas, collapsed within a span of a few weeks. It missed the digital shift, having been at the forefront of media until the turn of the year 2000. At the same time, the leading Swiss media company, Tamedia, deleted the printed version of one of its main daily newspapers, Le Matin, choosing to continue with only a digital version.

The news media are facing profound changes due to the development of digital technology, new competing media that are emerging, and the new individual media patterns that are developing (Willemin, 2018). Current media usage is marked by the arrival of new media companies on the Internet and by the prominence of social networks, and Switzerland is no exception. Both media technology and the very essence of news content and consumption are changing: while consumers are especially fond of free news and the dissemination of information is now immediate, online news is often obtained in a fast and superficial way (Flaxman et al., 2016; Eveland et al., 2004), and social media are becoming an important source of information, especially for young adults (Fög, 2017).

This study aims to map the media digital shift in Switzerland by drawing from individual media consumption practices, focusing particularly on the relationship between social media and other media usages. More precisely, we focus on two aspects. First, we are interested in detecting whether a digital shift has taken place in media use practices. In other words, can we account for a digital-oriented versus a paper-oriented media consumption space? Second, we aim to explore what individual factors can explain the formation of this new media space and discuss whether the generational divide can be supplemented by another divide related to news skills.

[18] This chapter is a slightly adapted version of the article that has been published as M. Reveilhac and D. Morselli (2020): "Digital Shift in Swiss Media Consumption Practices", Swiss Journal of Sociology, 46(3), 535-558.

The present study makes use of longitudinal data from the Swiss Household Panel (SHP) and relies on a dynamic version of multiple correspondence analysis (MCA) to answer our research questions. The panel nature of the SHP data offers insights that are otherwise masked when researchers rely on population aggregates because it allows us to track changes in consumption patterns. We thus adopt a data-mining approach to investigate media consumption changes among SHP respondents.

In the next section, we begin with an overview of the state of knowledge about media consumption in Switzerland. Then, we describe the main sources of data and present our method of analysis. In addition to providing substantive findings on the mapping of the Swiss media consumption patterns, this study suggests an innovative and dynamic use of MCA that has remained marginal in social sciences practices until now.

Changing media in a changing society

The societal role of mass media has been studied broadly, especially in relation to public opinion and political propaganda (Kaene, 1991; Hart, 1999; Street, 2005). In the last few decades, a constructionist approach to the study of public opinion (Bennett, 1993; Gamson & Modigliani, 1989; Kertzer, 2001; Neisser 1976; Price 1988) has moved away from considering media as a mere cause of opinion formation. These studies have moved in the direction of a more complex model in which people and media actors interact and co-create opinions and cultural and political value. According to this approach, the meaning of a particular phenomenon is constructed in the interaction between actors. In the case of media they should not be considered empty vessels to be filled with information; they are actors in interaction with another actor: the media (Ball-Rokeach & DeFleur, 1976). This approach to the analysis of media becomes even more relevant when looking at the rising importance of online media, in which actors are often in direct contact, instead of only cognitive exchange. For instance, social media allow people to exchange near-to-real-time information, transmitting news and user-generated content to each other. Similarly, journalists have increased their presence on social media to interact with readers, promote their articles (Hedman & Djerf-Pierre, 2013), and gather publishable information (Alejandro, 2010).

If the diffusion of new media has often been welcomed as a grassroots process for generating public opinion and making information more democratic, in reality, it goes hand in hand with several aspects that are worthy of concern. First, the cognitive processing of online news is fairly different for different categories of users. Some psychological research has shown that whereas experienced Internet users benefit from the online structure of news, expert users show a drop in attention and information processing (Opgenhaffen & d'Haenens, 2011; Tran, 2015). Web-based media might indeed facilitate a cognitive overload, which, in turn, is negatively related to information comprehension (Hou and Wang 2016). In other words, online media may make the reader process too much information, compromising the comprehension of its content and resulting in a superficial understanding of news.

Online media have been under the spotlight not only for comprehension but also for a series of side effects on the diffusion and formation of opinions. Some studies have shown that social media tend to increase the polarization of opinions by creating echo chambers, in which people are mostly exposed to opinions similar to their own via social media such as Facebook and Twitter (e. g., Quattrociocchi et al., 2016; Colleoni et al., 2014). This effect is amplified by the increasing use of personalized content by the major web corporations. Automated algorithms facilitate ideological segregation by offering only content from sources that fit the user's preferences and opinions and by creating filter bubbles in which users are rarely exposed to different opinions and viewpoints (Pariser, 2011).

Against the expectations of more horizontal and democratic communication, the effects of algorithms on public opinion might be particularly problematic for the correct functioning of a democratic system. Democracy is rooted in the assumption that heterogeneous worldviews and opinions can coexist and counterbalance each other towards a common goal and reach optimal solutions. Echo chambers and filter bubbles can instead create the illusion where a person believes that everyone believes what he/she believes and that there are no other opinions. Similarly, people might get the impression that only a limited number of events are occurring because news feeds are filtered by algorithms that prompt only preferred (i. e., the most browsed) types of news and topics.

Social media have had an increasing role in political communication, and it has been strategically used by political elites to shift votes and create opinions (Ratkiewicz et al., 2011; see also Woolley (2016) for a review). Thus, we might wonder about the role of online media on direct democracy systems, in which the population is called to vote on various issues several times a year and receiving information on the pros and cons of each issue is pivotal for the functioning of the system. In direct democracy systems, public opinion polarisation and the side effects of online media, such echo chambers and filter bubbles, may have profound political consequences. The long-lasting tradition of direct democracy makes Switzerland a case worth monitoring in the context of a changing landscape and a progressive shift toward online news and social media as information sources.

The media landscape in Switzerland

In recent decades, the media landscape in Switzerland has been characterized by multilingual complexity, strong public broadcasting of radio and television programming, and a general trend of media concentration and the downsizing of journalism. Echoing the situation in many other European countries, political pressure on public broadcasting has increased. In Switzerland, right-wing politicians launched a popular initiative in 2017, known as No Billag, which aimed at abolishing the provision of public funds to the public news service altogether. It was heavily rejected by more than 70% of the electorate. Swiss citizens thus remain loyal to the quality journalism offered by public broadcasting, despite the trends encouraging them to rely ever more heavily on the Internet and social media for information.

Concerning reliance on the Internet, the Swiss section of the World Internet Project – a comparative survey conducted every year since 2011 on a representative sample of the Swiss population by the Institute of Mass Communication and Media Research (MCMR) at the University of Zurich – showed a clear pattern towards a digital shift for news consumption (Latzer, 2017). The Internet was the most important source of media information in Switzerland in 2017, ahead of newspapers and television. It further showed that average Internet usage time has doubled since 2011 and currently stands at 25.5 hours a week, with young and poorly educated people spending the most time online. The number of non-users declined by half during the past six years, and the number of absolute non-users in Switzerland amounted to approximately 5% of the population.

The 2017 edition of the Annales survey showed that 41% of the Swiss population received information mainly from news sites or through social media (Annales, 2017). Concerning the reliance on social media as a source of journalism, the Digital News Report produced by Reuters in 2016 reported that 8% of Swiss news consumers said that social media had become their main source of journalistic information (Reuters, 2016), which is similar to trends founds in other European countries (Fög, 2018). In 2018, the same report showed that one-third of the Swiss population cited Facebook as a source of news, while trust in news content on social media also remained low, at 22% (Reuters, 2018). The 2016 report on media quality in Switzerland showed a correlation between reading only free or lower-quality journalism sources, including social media, and trust in the media system (Fög, 2016). Young people, however, remain over-represented among the users of social media platforms as a source of information (Fög, 2017).

The increasing reliance on the Internet and on social media as a source of journalism does not necessarily imply that readership is becoming less informed and less interested in keeping up-to-date about the current state of affairs. However, it raises questions about the necessary skills to process a wide amount of available news to distinguish between good and bad information. On May 10, 2017, the Swiss Federal Council published a status report on the legal basis for social media acknowledging that the increased influence of false information on political discourse is currently a source of lively debate, as social media play a central role in spreading fake news (Swiss Federal Council 2017). The growing global debate on fake news (Gorodnichenko, Pham & Talavera, 2018) is also reflected in the user behaviour captured in the abovementioned MCMR survey. In the information category, factchecking (78%) and searching for news (86%) have seen the most significant increases in the past few years. Until 2013, three-quarters of the Swiss population rated at least half of online content as trustworthy. This number dropped to 58% in 2017.

Two complementary hypotheses on media consumption patterns

These studies show that the Swiss media consumption patterns have constantly changed in the past few years and that readers are increasingly turning to online media for quick information. In this context, the quality of media must adapt to new consumer behaviours, and news agencies must face new economic constraints due to the declining revenue for traditional media.

The consequences of this digital shift concern the entire population, and the notion of what constitutes the news might also be affected. As suggested by Genner (2017), in addition to a generational divide that is expressed in the form of a heavier reliance on social media as a source of journalism among younger people, there is also a divide in the skills to treat the information available. The generational divide hypothesis may therefore also be complemented, rather than opposed, by a skills gap hypothesis. In the long run, there might even be a replacement of the former trend by the later as technological skills become diffused in society. Furthermore, as the online setting might also amplify opinion polarisation, it is also likely that political factors will play an increased role in the choice of media consumption patterns.

Given the on-going digitisation of news, our main aim is to understand how media information works in relation to online news and to social media while illustrating shifting media usage with individual sociodemographic and political factors. Determining media use, education, socioeconomic status, political interest, and political orientation can be complementary explanatory factors to age. Furthermore, the living context, such as the residential area, may also be important factors. For instance, the pools of potential customers of news media are below the national average in mountain regions and rural areas, where the public (and to a lesser extent private) broadcasting plays a central role, while free newspapers are less important than they are in towns and agglomerations (Hauptli, 2017).

Data and method of analysis

Data

The SHP offers unique panel data for Switzerland; it contains questions covering Swiss people's behaviours regarding social media and the broader use of the Internet. The SHP data provide information about individuals' political positioning and interests, as well as information on the occupational status and residential area, and they enable cohort analysis. We rely on wave 15 (2013) and wave 18 (2016) of the SHP questionnaire, which are the only available waves, including media consumption-related questions. We restricted our analysis to a subset of the sample that provided valid responses to a series of questions about individual media consumption in 2013. Our total sample consists of 1970 individuals.

The SHP questionnaire asked about the use of social media, and we recoded the data into three categories: Facebook and Twitter users, other social media users (including LinkedIn, Xing, MySpace and Google+), and people without any social media accounts. Although Facebook and Twitter may encompass different populations, we kept them in a single category to reach a sufficient number of people in each category. Then, we included variables accounting for three broad categories of Internet usage: frequency of chatting, frequency of reading news online and frequency of listening to the radio and watching TV on the Internet. The original scale was recoded into 1 = frequent use (which included 1 = every day and 2 = once to several times per week) and 0 = rare use (3 = once to several times per month, 4 = once to several times per year and 5 = never). Similarly, variables related to paper media consumption were retained and recoded using the same procedure: frequency of reading daily offline newspapers, free offline newspapers, and magazines.

To investigate political attitudes that could impact media consumption patterns, we used the SHP measures of political interest and self-positioning. The original scales ranged from 0 to 10, and to measure the changes in political interest and self-positioning, we subtracted the values from 2016 to 2013, which results in a scale ranging from -9 to +9. Concerning political interest, change was coded into seven categories: increased level of political interest (positive values from 2 to 9 from the subtraction); decreased level of political interest (positive values from -2 to -9 from the subtraction); no change (0 and ±1 from the subtraction) with medium levels (from 4 to 6 on the original scale); no change with very low levels (from 0 to 1 on the original scale); no change with low levels (from 2 to 3 on the original scale); no change with very high levels (from 9 to 10 on the original scale); no change with high levels (from 7 to 8 on the original scale).

A similar procedure was used to recode the change in political self-positioning into seven categories: change in political self-positioning to the right of the political spectrum (positive values from 5 to 9 from the subtraction); change to the left (negative values from -5 to -9 from the subtraction); no change (0 and ± 1 from the subtraction) with a centre political self-positioning (from 4 to 6 on the original scale); no change at the moderate right (from 7 to 8 on the original scale); no change at the extreme-right (from 9 to 10 on the original scale); no change at the moderate left (from 2 to 3 on the original scale); and no change at the extreme-left (from 0 to 1 on the original scale).

Several demographic variables were also included in the analysis. We assigned each person to one of six birth cohorts: <1942, 1943–1952, 1953–1962, 1963–1972, 1973–1982, and 1983–1999. Moreover, we included education level in 2016 that we recoded into four categories (tertiary, compulsory, general, and vocational). We also included occupational status in the year 2016 recoded as active (originally full-time paid work (at least 37 hours weekly), and work in protected atelier (for handicapped persons), part-time paid work (5–36 hours weekly and part-time paid work (1–4 hours weekly)), and inactive (originally retired people and other retired persons, other situations, further education, non-paid leave and work in the family company). Residential areas were coded into centres, urban areas (including suburban municipalities, mixed agricultural municipalities and peripheral urban municipalities), wealthy municipalities, tourist municipalities, industrial and tertiary sector municipalities, and rural areas (including rural commuter municipalities and peripheral agricultural municipalities). Last but not least, we also included the 7 large regions of Switzerland (Zurich, Central Switzerland, East Switzerland, Lake Geneva, Middleland, Northwest Switzerland, and Ticino), and gender was controlled.

Analytical strategy

To map the Swiss media landscape on the basis of individual media consumption measures, we used MCA, which can be understood as a multivariate factor analysis for categorical variables. MCA is usually visually represented in two plots: the graph of active modalities that determine the shape of the obtained map and the graph of supplementary modalities that serve to interpret the map. The spatial proximity between two modalities indicates in these graphs that these modalities are shared by a relatively large group of individuals. As people share more common patterns, they become more closely situated in the plan. Compared to other types of analysis, MCA has the advantage of accounting for complex (i. e., non-linear) relationships between variables and variable modalities. For this reason, MCA is a powerful analytical approach to analyse social categories and to investigate numerical variables once transformed into categories.

MCA allows us to model oppositions between variables in the logic of axes (or factors). The total information taken into account by each axis (inertia) is given by the variance rate explained (in the form of percentages) and by the eigenvalue of each axis. These indicators are thus used to determine the number of axes (factors) to retain for the analysis. For interpretation, we generally retain the number of axes that, when their respective rates are added, represent at least 80% of the cumulative variance rate. The first axis, which includes the variables with the highest inertia, represents the most important opposition; the second axis, the second most important, and so on. To interpret the difference between two modalities, it is customary to consider a difference of 0.5 as significant and a difference of 1 as very significant (Rossier, 2018).

As already mentioned, in MCA, variables can be analysed in two ways. First, a set of variables, named active variables, is used to define the axes and the distance between variables and between individuals. Once the space is formed by the oppositions between modalities of the active variables, it is possible to project additional (or illustrative) variables. Illustrative variables do not play any role in the formation of the axes.

Each active variable and each modality of an active variable contribute to a percentage of the inertia of each axis. By convention, active variables are considered as contributing to an axis when this percentage exceeds the average inertia contribution (100% divided by the total number of variables). Similarly, we consider a modality as contributing to an axis when it exceeds the average contribution of modalities (100% divided by the total number of modalities). To interpret the difference between two modalities, it is customary to consider a difference of 0.5 as significant and a difference of 1 as very significant (Rossier, 2018).

Our model had 7 active variables and 30 active modalities, for a total of 1970 respondents. Active variables are considered important when their contribution to an axis is larger than 14.3% (= 100/7). Similarly, the threshold of modalities is set at 3.3% (= 100/30). The details for our variables can be found in Appendix 1.

Faithful to a longitudinal perspective, MCA analyses the media use of the same individuals at two time points, namely, 2013 and 2016. We modelled the active variables by focusing on the respondents' positions between 2013 and 2016 (as described in the data section). It is thus possible to observe whether and how changes in individual positions over time affect media consumption patterns. This dynamic use of MCA has – to date – rarely been applied in the social sciences.

The different media consumption types in 2013 and 2016 were inserted as active variables in the model to examine the structure and evolution of media consumption in Switzerland. Then, we superimpose, as illustrative variables, the positions of the individuals on several variables that could explain the mapping of the media space. Illustrative variables include the different age cohorts, the variation in political interest, the variation in political self-positioning, as well as the level of education in 2016, the occupational status in 2016, the living area, the 7 big regions, and gender.

Results

Description of media usage

Between 2013 and 2016, the use of social media has evolved, with alternatives to Facebook and Twitter becoming more popular. Table 4.1.1 reports the descriptive statistics for our sample. The percentage of respondents without any social media account was the highest (56% in 2013 and 53% in 2016), followed by respondents with either Facebook or Twitter accounts (39% in 2013 and 40% in 2016). Respondents relying on other social media platforms had slightly increased (5% in 2013 and 7% in 2016).

If we focus on the three main reasons for consulting the Internet, online news consumption was higher than chatting and listening to the radio or watching TV online, and the proportions had increased between the survey years (from 66% to 72% for news consumption, from 14% to 17% for chatting and from 21% to 25% for listening to the radio or watching TV online). Concerning paper media, the consumption of daily news had decreased (from 83% to 80%), as had the consumption of magazines (from 54% to 51%). In contrast, the consumption remained the same for free news (54%). Overall, there was almost no change in the proportions of media consumption between the two survey years.

Table 4.1.1: Descriptive statistics for the active variables

	modalities	2013 count (%)	2016 count (%)	P
Social media (socmed)	*Facebook/Twitter*	775 (39%)	801 (40%)	0.450
	Other	97 (5%)	146 (7%)	0.021 *
	None	1133 (56%)	1058 (53%)	0.001**
Chating (chating)	*no*	1722 (86%)	1659 (83%)	0.006**
	yes	283 (14%)	346 (17%)	
Radio and TV (radio_TV)	*no*	1577 (79%)	1498 (75%)	0.003**
	yes	428 (21%)	507 (25%)	
Onlinenews (onlinenews)	*no*	690 (34%)	558 (28%)	<0.001***
	yes	1315 (66%)	1447 (72%)	
Newspaper (newspaper)	*no*	332 (17%)	410 (20%)	0.001**
	yes	1673 (83%)	1595 (80%)	
Freenews (freenews)	*no*	918 (46%)	923 (46%)	0.923
	yes	1087 (54%)	1082 (54%)	
Magazines (magazines)	*no*	924 (46%)	980 (49%)	0.079[a]
	yes	1081 (54%)	1025 (51%)	

Note: significance levels defined as **p < 0.01,*p < 0.05, a p < .08; N = 1970.

Table 4.1.2 reports the correlation among the study variables. Online behaviours (reading news, chatting and listening to the radio or watching TV on the Internet) were all positively correlated with each other for both survey years. Social media use was negatively correlated with reading traditional offline media, such as newspapers and magazines. It was instead positively correlated with free news and online news consumption, as well as with chatting and listening to the radio or watching TV on the Internet. These patterns hold for both survey years. There is thus a correlation between using social media and reading only free or lower-quality sources.

We further note that individuals who do not use social media also tend to avoid any online activities, such as reading news online, chatting, listening to radio or watching TV online. These persons also prefer reading newspapers instead of free news. These findings already indicate that a divide is taking place not only with respect to the use versus non-use of social media but also with regard to quality versus low-quality information consumption. Furthermore, the users of other types of social media, such as Facebook and Twitter, seem to follow a different logic because they tend to be actively involved in online practices but have no significant pattern related to offline news consumption.

Table 4.1.2: Correlations table

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. onlinenews_13													
2. chating_13	0.14 ***												
3. radio_TV_13	0.18 ***	0.11 ***											
4. onlinenews_16	0.55 ***	0.11 ***	0.17 ***										
5. chating_16	0.09 ***	0.30 ***	0.09 ***	0.10 ***									
6. radio_TV_16	0.21 ***	0.10 ***	0.37 ***	0.21 ***	0.13 ***								
7. dailynews_13	0	-0.10 ***	-0.03	0.02	-0.07**	-0.05*							
8. freenews_13	0.18 ***	0.07**	0.04*	0.16 ***	0.07**	0.06**	-0.02						
9. magazines_13	0.05*	0.02	0	0.02	0	-0.02	0.15 ***	0.01					
10. dailynews_16	-0.03	-0.09***	-0.03	-0.02	-0.06*	-0.04	0.55 ***	-0.08***	0.13 ***				
11. freenews_16	0.11 ***	0.06**	0.05*	0.12 ***	0.06**	0.05*	-0.02	0.48 ***	0.05*	-0.01			
12. magazines_16	0.01	-0.03	0.02	0.03	-0.06**	0.02	0.16 ***	0.03	0.43 ***	0.18 ***	0.07***		
13. socmed_13	-0.11 ***	-0.20 ***	-0.10 ***	-0.11 ***	-0.14 ***	-0.10 ***	0.15 ***	-0.08***	0.07**	0.12 ***	-0.06*	0.11 ***	
14. socmed_16	-0.09 ***	-0.17 ***	-0.06**	-0.09***	-0.14 ***	-0.04	0.12 ***	-0.09 ***	0.07**	0.10 ***	-0.06**	0.08***	0.65 ***

Note: Significance levels defined as *** p < 0.001, ** p <0.01, * p < 0.05; the endings _13 and _16 account for the survey year; the endings TW&FB account for Twitter and Facebook

Description of individual variables

Concerning the supplementary variables, the majority of respondents (84 %) did not express any change in their political self-positioning between 2013 and 2016: 47 % remained at the centre of the political spectrum, 21 % on the left, and 16 % on the right. Approximately one-tenth of respondents expressed a shift towards the left (9 %) or the right (7 %) of the political spectrum. With respect to political interest, approximately one-tenth of respondents expressed either an increased (13 %) or decreased (10 %) interest in politics. A total of 9 % of participants did not express any change in their level of political interest and remained uninterested in politics represent 9%, while 46 % remained interested in politics.

With respect to socio-demographic supplementary variables, the distribution of cohorts does not overrepresent the youngest age groups. Regarding occupational status, 67 % of the sample were active and 33 % were inactive. With respect to education, most of the included individuals have a tertiary education diploma (48%), followed by vocational school (38 %), general (10 %), and compulsory school (3%). Furthermore, women are more represented than men (57 % versus 43 %). Finally, respondents living in urban municipalities (41 %) were more represented than respondents living in centres (29 %), in rural municipalities (14 %), in tourist and wealthy municipalities (9 %) and in industrial and tertiary sector municipalities (7 %). Concerning the regions, people in our sample came mostly from the Middleland (25 %) and from the region of the Lake of Geneva (18 %), while very few live in Ticino (3 %).

Table 4.1.3: Descriptive statistics for the supplementary variables

	modalities	count (%)
Political self-positioning (politor)	*extreme-left stable (ext.-left)*	55 (3%)
	left stable (left)	365 (18%)
	center stable (center)	937 (47%)
	right stable (right)	288 (14%)
	extreme-right stable (ext.-right)	46 (2%)
	moderate change toward left (left+)	173 (9%)
	moderate change toward right (right+)	173 (7%)
Political interest (polint)	*very low stable (very low)*	54 (3%)
	low stable (low)	111 (6%)
	middle stable (middle)	447 (22%)
	high stable (high)	671 (33%)
	very high stable (very high)	250 (12%)
	moderate decrease (less-)	201 (10%)
	moderate increase (more+)	271 (14%)
Age cohorts (cohort)	*<1942*	206 (10%)
	1943-1952	367 (18%)
	1953-1962	541 (27%)
	1963-1972	460 (23%)
	1973-1982	254 (13%)
	1983-1999	177 (9%)
Language (PLINGU16)	*french*	510 (25%)
	german	1439 (72%)
	italian	56 (3%)
Occupation (OCCUPA13)	*fulltime work*	713 (36%)
	parttime work	683 (34%)
	at home	124 (6%)
	studying	41 (2%)
	retired	402 (20%)
	unemployed	24 (1%)
	other	18 (1%)
Gender (SEX13)	*man*	869 (43%)
	woman	1136 (57%)
Type of municipalities (COM2_13)	*tourist and wealthy towns*	176 (9%)
	centers	586 (29%)
	urban towns	815 (41%)
	Industrial and tertiary sector towns	140 (7%)
	rural towns	288 (14%)

Illustrating media consumption patterns with individuals and contextual factors

By retaining more than 80 % of the inertia, two axes contributed to structuring the map of the media consumption patterns. The variance rate of the first axis was larger than that of the second axis (67 % vs. 17 %), which indicates a particularly strong opposition in space along the first axis compared to the second.

Social media in 2013 and 2016 made above-average contributions to the first axis. On the negative side of the plot (left side), we find the absence of social media use and the absence of reading online news in 2013 and 2016. In contrast (right of the axis in the graph), we find the use of Facebook and Twitter in 2013 and 2016, frequent chatting and frequent listening to the radio or watching TV online in 2013 and 2016. We also find the absence of consulting newspapers and online news. Thus, we argue that the first axis refers to a main cleavage between the reliance on social media associated with frequent online behaviours, such as chatting and listening to the radio or watching TV online, and the absence of use of social media and the absence of online news consumption.

The second axis was defined mainly by the different categories of newspaper and magazine consumption (in 2013 and 2016). We find in the negative coordinates (bottom of the axis in the graph) the frequent reading of magazines, as well as the frequent consumption of newspapers. In the positive coordinates (top of the axis in the graph), we find the absence of reading online news, offline newspapers and magazines. Thus, this axis refers to the main cleavage between the consultation of news and the absence of reading offline or online media content.

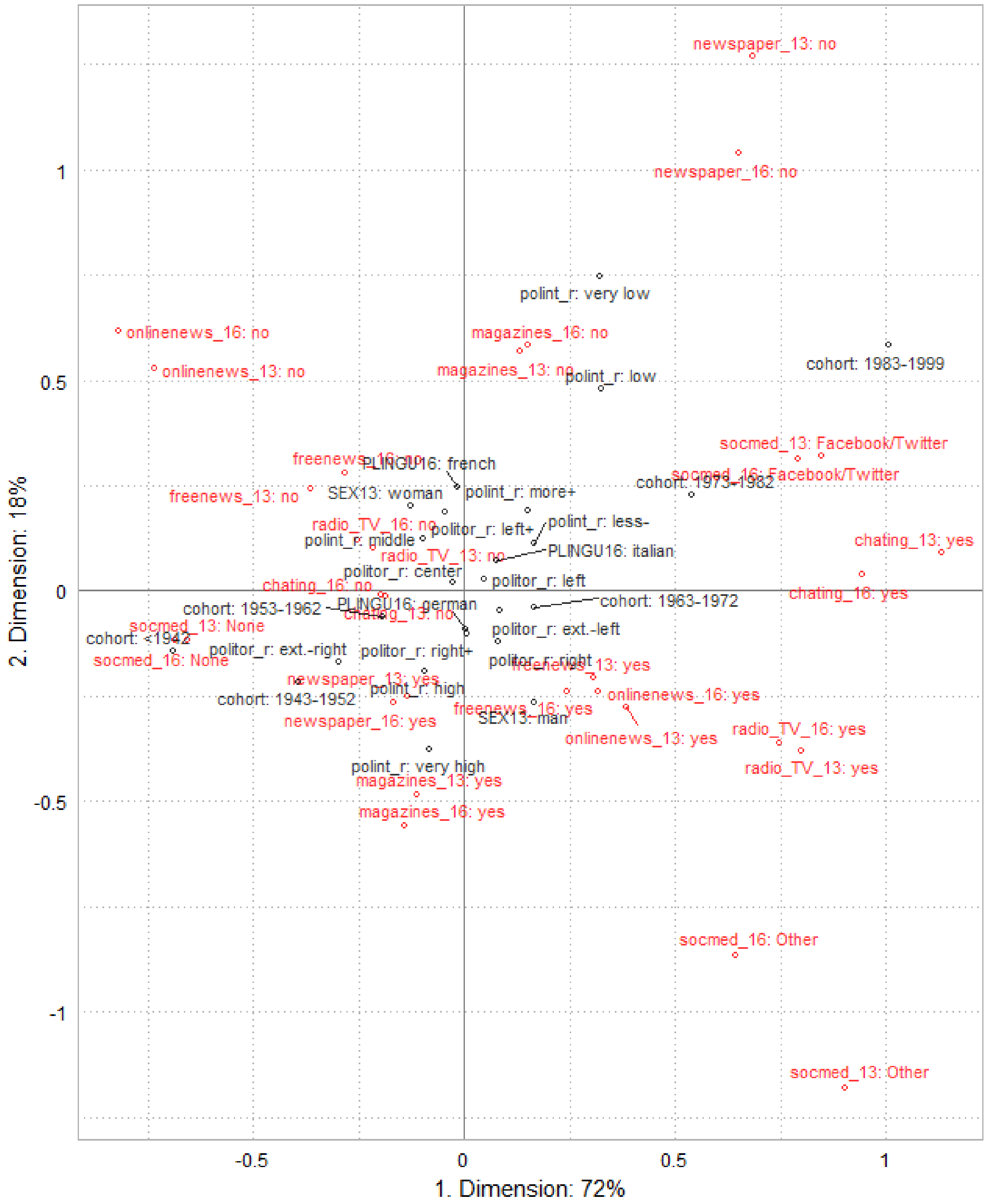


Figure 4.1.1: Map of active and supplementary modalities

Note: The endings _13 and _16 account for the survey year; active variables are in grey and supplementary variables in black; N = 1970

Some interesting trends can be highlighted by comparing the distances between media consumption measured in 2013 and 2016. First, whereas frequent consumption of offline and online news are located closer together on the map (lower-right quadrant), the absence of consumption of these media is situated further apart (upper-left quadrant). Second, the consumption of every media type included in the analysis has not changed much between 2013 and 2016 (as shown with the descriptive statistics in Table 4.1.1). Third, the reliance of social media such as Facebook and Twitter (situated in the upper-right quadrant) follows a different logic than reliance on other social media such as LinkedIn, Xing, MySpace and Google+ (situated in the lower-right quadrant).

To better understand possible age differences in the use of media, MCA was also performed separately by three groups of cohorts (1983–1999 for the younger group, 1973–1982 and 1963–1972 for the intermediate group, and 1953–1962, 1943–1952 and before 1942 for the older group). Figure 4.1.2 (a, b and c) shows that all of the variables had similar reciprocal relationships, and the axes of the overall model were replicated. The differences among the younger cohort were mostly prompted by not using social media and online news and by using the Internet for chatting or using social media other than Facebook and Twitter. However, these differences became less accentuated in 2016, showing a more homogenous consumption of the Internet and social media. A similar trend is observable for the other two cohorts. Among groups of cohorts, the use of offline media was also quite diverse, and more polarized consumption types were observed.

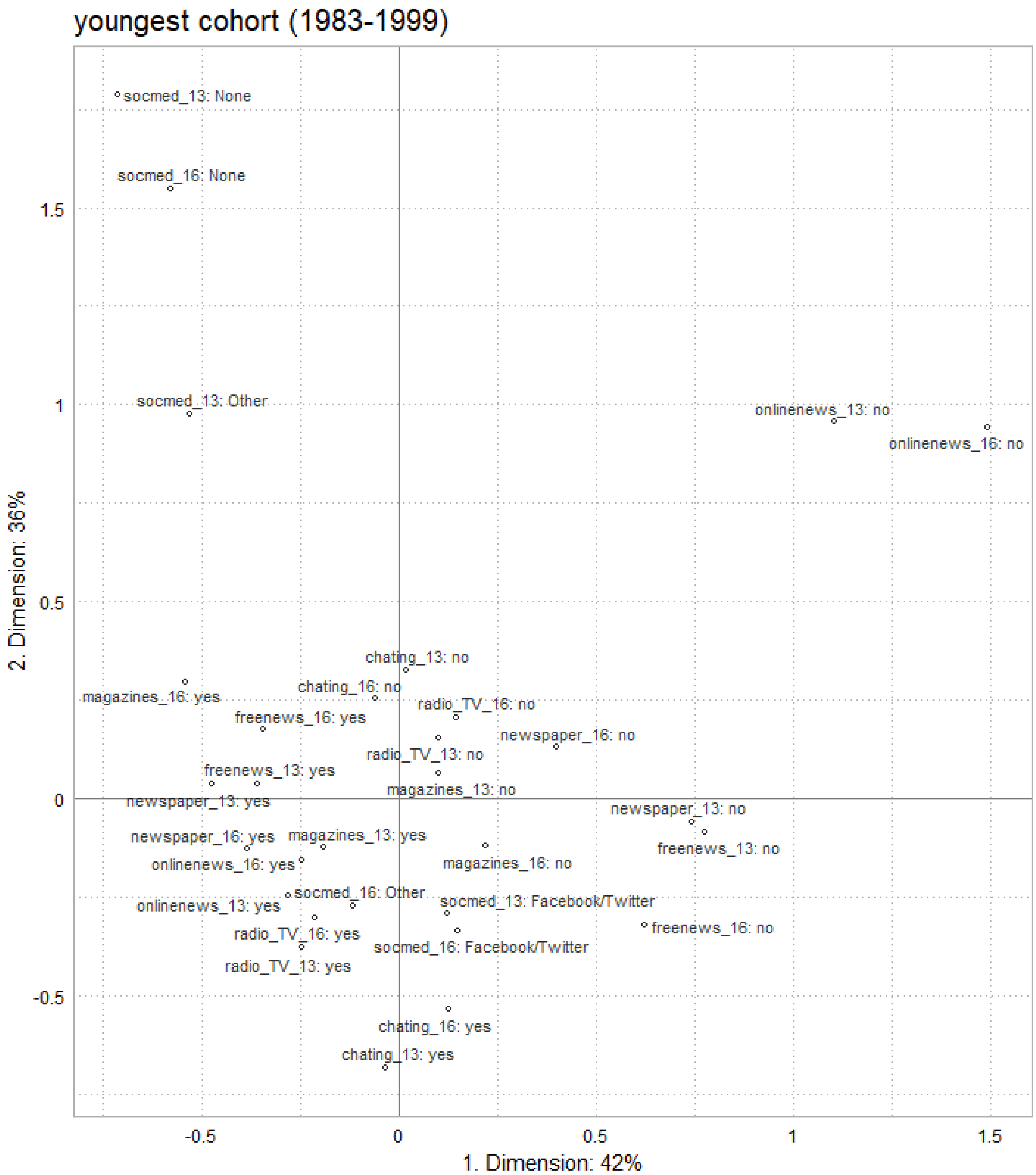


Figure 4.1.2 (a): Active modalities by age cohorts: youngest cohort

Note: Variable abbreviation are the same than Figure 4.1.1; the number of individuals equals 173 in the young cohort, 702 in the intermediate cohort, and 1095 in the older cohort.

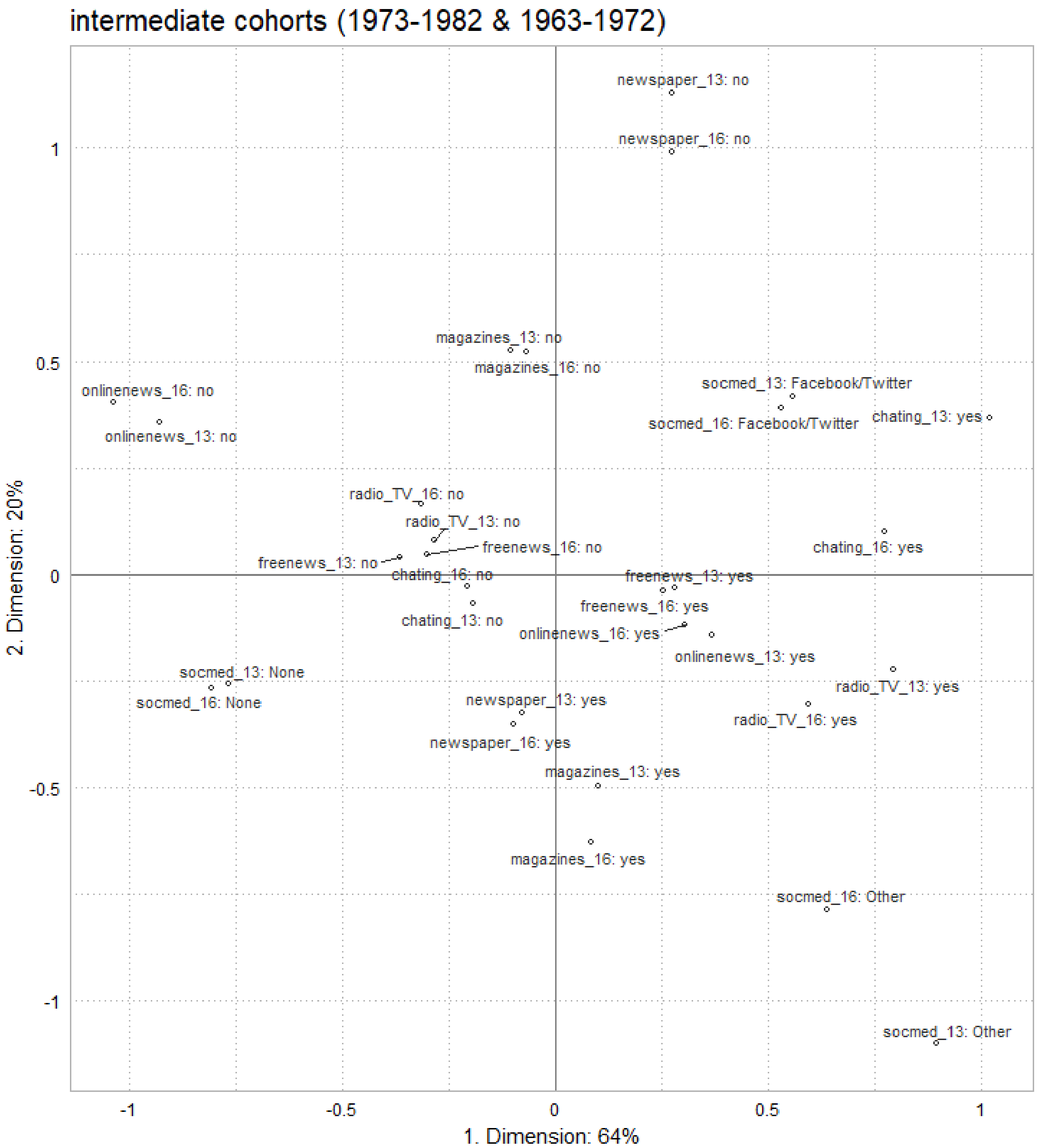


Figure 4.1.2 (b): Active modalities by age cohorts: intermediate cohorts

Note: Variable abbreviation are the same than Figure 4.1.1; the number of individuals equals 173 in the young cohort, 702 in the intermediate cohort, and 1095 in the older cohort.

Figure 4.1.2 (c): Active modalities by age cohorts: older cohorts

Note: Variable abbreviation are the same than Figure 4.1.1; the number of individuals equals 173 in the young cohort, 702 in the intermediate cohort, and 1095 in the older cohort.

Explaining different consumption profiles

Figure 4.1.1 shows the supplementary variables projected on the media consumption landscape (labels in black on the figure). Overall, there was little difference between political self-positioning and types of municipalities. The age cohorts were structured along the first axis, which summarizes the opposition between the reliance on or absence of reliance on social media. Younger respondents (see upper-right quadrant) were associated with reliance on social media, frequent chatting, and the absence of both online and offline news and magazine consultation, while older cohorts (see lower-left quadrant) were associated with the absence of social media usage, as well as the frequent consumption of daily news and magazines. With respect to the second axis, political interest and gender could explain the repartition of the active modalities. Respondents expressing no change in their level of political interest and who were highly interested in politics were situated in the upper-right quadrant. Gender also showed differences: women had positive values on the second axis, and men had negative values on the second axis.

Figure 4.1.1 also shows which of the supplementary categories are most important when interpreting the formation of the dimensions. The different age cohorts were structured mainly along the first dimension, whereas the changing level of political interest, as well as gender, were important variables when interpreting the second dimension.

To test the repartition of individuals, cluster analysis can be used to look for groups (or clusters) that bring together respondents sharing similar media consumption patterns. This procedure maximizes the homogeneity of clusters so that respondents in each cluster are most similar to one another and most of the differences are between clusters. Over several tested models, the best fit (i. e., best Duda-Hart, PseudoT2, and Beale indexes) was given by two clusters. The first cluster (named new media consumers) is characterized mainly by social media use, chatting, watching TV or listening to the radio online, and free news consumption, whereas the second cluster (news consumers) is defined mainly by information consumption practices through the use of traditional media outlets, such as offline newspapers and magazines, as well as online media outlets. In a final step, the two clusters were profiled using the supplementary variables not used to identify the clusters by means of a logistic regression (Table 4.1.4).

The results confirm that younger age cohorts were classified as new media consumers significantly more often than were older cohorts. Men were more likely to be associated with this group of consumers than women. Moreover, urban towns were more likely to be represented in the new media cluster than rural towns. People with high and middle levels of political interest are also more likely to be associated with news consumption than people with very low levels of political interest. Political orientation showed no significant results but tends to inform the polarization behaviour of those with rightist political inclinations, as we observed that people with extreme-right self-positioning tend to rely less on new media than left-leaning people. Finally, people living in East Switzerland are less likely to rely on new media than people living in the region of Zurich.

Table 4.1.4: Logistic regression on the new media consumers cluster (coded as 0) and the news consumers cluster (coded as 1)

		β (*SE*)	*p*
(Intercept)		-0.07 (0.57)	.89
Political self-positioning (politor)	*ext.-left*	0.08 (0.38)	.83
	left	0.14 (0.28)	.61
	center	0.01 (0.18)	.95
	ext.-right	0.91 (0.51)	.07*
	left+	0.14 (0.28)	.61
	right+	0.21 (0.30)	.48
Political interest (polint)	*low*	0.22 (0.52)	.67
	middle	0.82 (0.45)	.07*
	high	0.87 (0.45)	.05*
	very high	0.46 (0.47)	.32
	less-	0.15 (0.48)	.75
	more+	0.44 (0.47)	.34
Age cohorts (cohort)	*1943-1952*	-1.13 (0.27)	<.01***
	1953-1962	-1.63 (0.33)	<.01***
	1963-1972	-2.40 (0.34)	<.01***
	1973-1982	-3.07 (0.38)	<.01***
	1983-1999	-5.57 (0.80)	<.01***
Occupation (OCCUPA13)	*parttime work*	0.53 (0.18)	<.01***
	at home	0.56 (0.30)	.06*
	studying	-0.23 (1.11)	.83
	retired	0.28 (0.27)	.29
	unemployed	43525 (0.56)	.06*
	other	0.67 (0.69)	.33
Gender (SEX13)	*woman*	0.68 (0.16)	<.01***
Type of municipalities (COM2_13)	*tourist and wealthy towns*	0.35 (0.24)	.13
	urban towns	0.18 (0.16)	.27
	Industrial and tertiary sector towns	0.60 (0.27)	.02**
	rural towns	-0.08 (0.22)	.72
Language (PLINGU16)	*french*	-0.33 (0.15)	.02**
	italian	-0.45 (0.43)	.29

Note: Significance levels defined as **p < 0.01, * < 0.05, a p < 0.08; N = 1416.

Discussion and concluding remarks

The findings of the present study relied on an innovative way to exploit MCA by adding a temporal variation to the map. This use of MCA was first proposed by Mercklé (2017), Rossier (2018), and Rossier and Fillieule (2019), and some aspects of it still need to be deepened. However, it is a particularly promising technique and allowed us to highlight consumption patterns from a temporal perspective while seeking to explain them through individual and contextual factors. If, on the one hand, this approach has some limitations, such as the absence of clear-cut statistical tests, on the other hand, it allowed us to explore the multivariate and non-linear relationship between the study variables. Another limitation of this study was that it relied on a time window of only three years (2013 and 2016), which might be too narrow to observe radical changes in behaviour, consumption profiles and opinions. Our results showed that changes in both media usage and the covariate variables, particularly political interest and political positioning, have nevertheless occurred. Future studies relying on data from the SHP 2019 could test the hypothesis of opinion polarisation.

Our first research question asked whether it is possible to observe a digital shift in the analysis of media-use practices or, in other words, if it is possible to account for a digital-oriented versus a paper-oriented media consumption space. This distinction emerged clearly from the map of media consumption practices, showing a contrast between online and offline media consumption. However, the MCA found that more than one dimension explains the difference in consumption styles. This finding implies that the Swiss media space cannot be interpreted solely in a dichotomous offline-online view. Instead, it seems that social media, which offer the possibility for users to generate content and gain information, goes hand in hand with the consumption of other online news sources and offline media sources, mainly free news.

The relationship between using social media and reading only free or lower quality sources might be particularly problematic in the context of direct democracy because information plays a vital role in guaranteeing informed opinion and votes. Indeed, a narrow consumption of news, whether online or offline, was associated with low political interest. Our second research question, which addressed what individual factors best explain the formation of this media space, shed some more light in this direction. Our results show that the increase in online media is most common among younger cohorts.

The age cohorts were structured along the first axis, which summarized the opposition between the reliance or absence of reliance on social media: younger respondents more often relied on social media, reported frequent chatting, and reported the absence of both online and offline news and magazine consultation, while older respondents were characterized by an absence of social media usage, as well as frequent daily news and magazine consultation. Our findings therefore show that younger cohorts use social media in a homogenous way, giving higher preference not only to interactive media such as social media and chat services but also to online and free news. With respect to the second axis, which summarizes the opposition between the consultation of and absence of reading offline or online news content, political interest, gender, and, to some extent, political orientation, it illustrates the repartition of media consumption patterns.

The cluster analysis also showed that media consumption is cumulative: People with the habit of seeking information use different types of media at the same time. In line with research on interactive media usage (Opgenhaffen & d'Haenens, 2011; Tran, 2015), this audience might therefore develop expertise in using media, maximizing the benefits of online support and possibly contrasting the negative side effects of social media. This explanation complements the generational divide hypothesis by emphasizing the skill divide hypothesis. Indeed, as Genner (2017) suggests, the ways we consume information may well transcend questions of age, and the growing online news ecosystem is likely to increase the divide between individuals with versus those without the needed skills to sort through the available news and information.

To conclude, we would like to stress that the change in media usage habits is not necessarily negative for democracy. It can indeed demonopolize the news industry, creating grassroots and alternative sources of information. For instance, the use of social media in collective action and mass protest, such as the wave of protest that impacted Chile in 2019, has allowed activists to draw attention to (alleged) human rights violations committed by the police that had been neglected by official news. However, to maximize these benefits, a new culture of media consumption should be created. Users/readers should be encouraged to use multiple media sources and be aware of the way automated algorithms of news selection work in order to contrast the hidden risks of new media. An informed use of new media could thus represent a resource for, instead of a risk to, direct democracy.

4.2 The impact of social media use for elected parliamentarians: Evidence from politicians' use of Twitter during the last two Swiss legislatures[19]

Introduction: do social media make a difference for political success?

Campaigning on social media has become a core feature of political communication. Parties and politicians rely heavily on these platforms to promote their views, interact with citizens and actors close to politics, and generate traditional media attention (Spierings et al., 2018; Keller, 2020). By presenting themselves prominently on social media and by being responsive to public concerns, politicians can position themselves as candidates that voters can trust and build long-term reputability. In this article, we focus on the Swiss political environment and investigate whether the activity of politicians on Twitter is an effective strategy for gaining electoral success over a period of successive legislatures.

Despite extensive research on politicians' reliance on social media, two important research gaps remain. Firstly, the biggest share of the literature focuses on the use of social media by politicians during elections (Vaccari, 2017). However, due to permanent campaigning (Larsson & Kalsnes, 2014), politicians also frequently post messages between elections to increase their accountability and popularity. This is especially important for those elected politicians who rely on social media beyond intense campaigning periods. However, being active on social media requires an important additional investment from elected politicians that may not always be translated into (offline) political success, in terms of re-election or increased reputability. Secondly, there is already a wide body of research (e.g., Keller & Kleinen-von Königslöw, 2018) proposing that offline measures (e.g., vote shares) predict success online (e.g., followership). However, the reverse effect – accounting for the potential of social media communication to lead to offline success – is still understudied.

By investigating the effect of social media use on political success over successive legislatures, we aim to understand the extent to which social media presence can grant politicians the accountability to citizens and to the media that is important for a career in

[19] This chapter is a slightly adapted version of the article that has been published as M. Reveilhac and D. Morselli (2022): "The impact of social media use for elected parliamentarians: Evidence from politicians' use of Twitter during the last two Swiss legislatures", Swiss Political Science Review, 0(0), 1-24.

politics (e.g., through election, re-election, or higher position on the party list). Therefore, we measure offline success in two ways. First, we measure success in terms of the rank occupied by politicians based on the share of votes obtained in their constituency. Second, we measure success by looking at the amount of media attention generated by politicians during electoral periods. A throw-away brief remark can attract media attention and end up being cited in the traditional press the next day. This logic applies especially to Twitter as it offers a privileged means of contact with journalists (Keller, 2020).

Relying on longitudinal data enables us to assess how the effect of reliance on social media compares to other factors that contribute to politicians' success, such as parliamentary experience (e.g., incumbency) and responsiveness to citizen concerns. The effect of contextual factors, such as the legislature and the type of parliamentary chamber, can also be accounted for. To date, the majority of Swiss parliamentarians have a profile on social networks, but few of them are either really active (e.g., posting frequently) or adopt interactive behaviors (e.g. replying to citizens). However, not only social media usage per se, but also the style of online communication has been shown to affect politicians' (offline and online) popularity (e.g., Enjolras, 2014; Jungherr et al., 2017; Bright et al., 2020). In line with this area of research, our study also aims to assess whether and how different styles of online communication affect politicians' success.

To conduct our analyses, we rely on the online history of politicians who have had at least one parliamentary mandate during the most recent Swiss parliamentary legislatures (2011–15, 2015–19, and 2019–23). The longitudinal nature of the data allows us to assess which audiences are deemed important by politicians. We then assess the effect of social media use and style of communication over successive legislatures.

Theoretical background

The impact of politicians' use of social media on their electoral success

The literature on the patterns present in the adoption of Twitter by parties and politicians during campaigns is related to studies focusing on the adoption of other digital tools in the campaign repertoires of politicians. The findings are congruent across various countries and election cycles. According to Jungherr's (2016) review, parties and politicians in opposition are more likely to use Twitter than members of governing parties. However, politicians from well-established major parties, incumbents, and those with high campaign budgets are more likely to use Twitter than others. Furthermore, young politicians and politicians with urban constituencies appear to be more likely to use Twitter than others. Also, Twitter use in many cases seems to correspond with the intensity of electoral competition, former success with Twitter by members of the same party, and strong ideological positions.

The most visible motivation to explore politicians' reliance on social media in the electoral context is to infer attitudes towards politicians in view of predicting election results. However, this research endeavor has been criticized for its methodological inconsistencies (Gayo-Avello, 2012; Metaxas et al., 2011) and the arbitrariness of some of its choices (Jungherr, 2012). The current state of the literature delivers mixed results. Some studies have demonstrated that candidates' increased reliance on social media can influence the outcome of the election. For instance, researchers have found a positive relationship between social media use and increased success in the ballot (e.g., Kruikemeier, 2014; Bode & Epstein, 2015; Bene, 2018), while others do not (Vergeer et al., 2013). Therefore, no clear picture emerges about the connection between Twitter use and popularity or electoral chances. In the same vein, some studies have found links between the mentions that political candidates or parties received on Twitter and their election results (McKelvey et al., 2014), while other studies found no relationship between online popularity and vote share (e.g. Vaccari & Nielsen, 2013; Jungherr et al., 2017).

Thus, if there is a relationship between Twitter use and electoral success, it seems to be an indirect one, highly dependent on the respective electoral context. Indeed, there is little successful replication of the relationship between specific signals and specific metrics of support (Jungherr et al., 2017; Huberty, 2015). These arguments need no restating here. In a nutshell, there seems to be a consensus that signals contained in social media (e.g., number of tweets, likes, or mentions) should not be taken at face value for predicting the outcome of elections. However, these metrics certainly serve as a good proxy for politicians' reputation by demonstrating how actively other users react to their messages (Keller & Kleinen-von Königslöw, 2018).

To date, empirical studies have mostly investigated the social media contribution to politicians' success during election campaign periods, notably by studying the relationship between the reliance of political candidates on social media and their share of the vote (or vote outcome). Beside cross-sectional investigations about the extent to which efforts made on social media correlate with vote outcomes, few studies have observed change in outcomes over time. A notable exception is the study of Bright et al. (2020), which offers a stronger test of the impact of social media use on vote share outcomes. The authors covered two successive elections in the United Kingdom (in 2015 and 2017) conducting "pseudo-panel" analyses on a subset of political candidates who competed in both elections. They showed that the impact of Twitter use is small in absolute terms, but comparable with other factors, such as campaign spending.

In this study, we take a step back from approaches aiming to predict election outcomes. Instead, our main interest is to investigate the effect of social media use on political success over successive legislatures from two perspectives: i) the rank occupied by politicians when based on the share of votes obtained in their constituency; ii) the amount of media attention generated by politicians during electoral periods.

The impact of politicians' style of online communication on their electoral success

Political success also depends on the style of political communication. With social media, politicians can present themselves to their (potential) electorates more easily than ever before. For instance, they can generate attention by reporting on party activities and parliamentary work, but also by interacting with other users on social media, such as journalists, other politicians, actors close to politics, and citizens (Spierings et al., 2018; Keller, 2020). For new political candidates, this may facilitate a direct appeal to citizens to follow and, perhaps, support them during the next election. For political incumbents, it could consolidate their success and position them as leading figures in their party, thus boosting citizens' willingness to re-elect them. Studies have shown a certain variation between unilateral and interactive styles of political communication on social media (Enjolras, 2014). This is also linked to the platform's functional capabilities. For instance, Twitter provides several features for user interaction, such as retweeting, mentioning, and replying. Retweeting allows one user to repost another user's message that has triggered their interest, while specifying the original sender's username to provide a direct link to the initial source of information. Mentioning is employed to contact, reach out to, or acknowledge another user within a tweet. Replying occurs when one directly responds to another user's tweet. To date, studies have focused both on who has access to politicians' accounts and on who is drawing attention from politicians (Spierings et al., 2018; Keller, 2020). However, less is known about how and whether politicians' responsive behaviors (especially in terms of replies) affect their political success offline.

Replying and mentioning constitute emblematic features through which politicians can demonstrate their degree of responsiveness to public concerns and their interest in other audiences (see also the study by Tromble (2018) on politicians' reciprocal engagement with members of the public). Both features also enable politicians to actively demonstrate their interest in particular audiences. For instance, mentions allow politicians to invite and notify other users. Furthermore, replying is a task that requires that politicians get actively involved in discussions. From this view, replies demonstrate more involvement than retweets, which can sometimes be considered as an endorsement or a least as a sign that a tweet has been deemed interesting enough to be retweeted (Metaxas et al., 2015). However, the existing literature found little evidence that making use of interactive conversation strategies was impactful on vote share (Bright et al., 2020), thus challenging existing work which has criticized politicians for not engaging in social media in a more interactive fashion (e.g., Jungherr et al., 2017). The style of political communication, both in terms of interactivity and unilaterality, further varies greatly across party affiliation (e.g., Enjolras, 2014). For instance, Spierings and Jacobs (2018) showed that populist parties are less likely to interact with and respond to social media users than members of other parties.

These features complement other network measures (e.g., followers or friends), which tend to represent more passive (unidirectional or reciprocal) online behaviors, or informative behaviors (e.g., link sharing), that are not necessarily directed toward a particular audience (Bright et al., 2020). Most studies suggest that politicians generally use social media to unilaterally disseminate information instead of interacting with voters (Klinger & Svensson, 2015 for a review of this literature, see Jungherr, 2016). However, link sharing is also an important aspect of political communication on social media as it allows politicians to direct users to their official publications (e.g., manifesto, press release), to connect their messages to current events, or to diffuse information. This broadcasting behavior has been shown to be more successful in terms of vote share than more interactive styles of communication (Bright et al., 2020). This might be because it can generate increased media coverage by resonating with the news timelines and topicality (Broersma & Graham, 2012).

There is also evidence that the structure of a politician's network has an impact on their level of online popularity. For instance, Keller and Kleinen-von Königslöw (2018) found that, on Twitter, it is not the greater social media experience that ensures more reactions (in terms of favorites and retweets), but rather the greater professional follower networks politicians were able to build. Therefore, it is important to investigate with which publics politicians interact on social media and how this can, in turn, lead to success offline. Although there is little evidence of Twitter being an enabling device for dialogue between candidates and citizens, it remains unclear whether adopting a more interactive style of communication is likely to grant politicians more success offline. Yet, politicians need to be responsive to civil society to gain higher levels of legitimacy. Social media represent certainly one channel through which they can demonstrate their responsiveness to concerns from other users (Porten-Cheé et al., 2018), and also through which politicians can learn about public preoccupations to adapt their agenda (Ennser-Jedenastik et al., 2021). Therefore, in this study, we particularly focus on the effect of politicians' responsiveness on social media in terms of replies and policy issue responsiveness on their political success.

The impact of politicians' social media use to generate media attention

Our study also aims to assess the extent to which politicians' use of social media allows them to generate traditional media attention. In the article The political-media complex, Swanson (1992) denounces political communication in the US because of the particularly close relation between media and politics. This dominant media logic in political campaign coverage has been shown to drive individualization processes (Swanson & Mancini, 1996). In recent years, social media have also been identified as a channel with the potential to increase the focus on the personal side of politics (Karlsen & Enjolras, 2016). Furthermore, in line with Chadwick's (2013) analysis of media systems in Western democracies as hybrid media systems, contemporary mass media and social media have become intertwined. For instance, political journalists incorporate Twitter in their routines to keep up with campaign developments during elections (Wahl-Jorgensen, 2014).

Because traditional media remain among the main sources of information by which Swiss citizens forge political opinions (Eisenegger, 2020), it is important to gain an understanding of the possible relationship between politicians' social media usage and traditional media content. Furthermore, the share of citizens using social media as a main source of information is growing on a yearly basis, especially among young people (Eisenegger, 2020: 12). There is indeed a certain degree of inter-media agenda setting, where the traditional media are likely to report what politicians post online (e.g. Anstead & O'Loughlin, 2015).

To our knowledge, studies investigating the possible connections between politicians' online communication and their (traditional) media coverage are developing. We can identify three related strands of literature. The first two strands draw from early studies about the coevolution (e.g. Russell Neuman et al., 2014) and co-influence (e.g. Kruikemeier et al., 2018) of social media and traditional mass media.

The first strand of research on political communication investigates the relationship between politicians' reliance on social media and online news coverage with a view to assessing whether one agenda is leading the other (e.g. Barberá et al., 2019). Most recently, researchers extended this relationship by including other agendas. For instance, Gilardi et al. (2020) have assessed the coevolution of traditional media coverage and the agenda of parties and politicians on social media. They have shown that, overall, no one agenda leads the others any more than it is led by them. Overall, these studies found little evidence that a particular agenda decisively leads another.

The second strand of research investigates the major groups or individuals with whom politicians communicate, especially focusing on politicians' networks of followers and friends, as well as on their replies (e.g. Vaccari & Valeriani, 2015). For instance, Rauchfleisch and Metag (2016) focused on politicians' replies and found that parliamentarians received the greatest number of replies per actor, followed by local politicians, citizens, and journalists. Although there is this growing body of descriptive research about the configuration of politicians' social media networks (Keller, 2020), there is little knowledge of how elected politicians' online interactions have evolved over time. These studies generally endorse a descriptive approach that focuses little on the possible influence of politicians' strategies for communicating with specific audiences on their ultimate electoral success or on the traditional media attention they generate. Politicians who often interact with journalists on social media could likely generate more traditional media coverage.

More recently, a third strand of studies has investigated the resonance of social media content by showing that social media attention is often translated into news media attention (Schroeder, 2018) or is used to represent public opinion (McGregor, 2019). Studies have identified problematic practices. These related, for example, to the amplification of controversial content which would otherwise remain marginal (Donovan & Boyd, 2021). There is also the risk that news media uncritically report social media trends that might be the product of artificial content production (Kovic et al., 2018). Our study draws from this strand of the literature by investigating how politicians' reliance and communication styles on social media impact their mediatic success in terms of coverage.

Case study: The impact of Twitter on Swiss politicians' political success

Social media are not yet a primary source of information for citizens, traditional media still being central. For instance, the 2021 Reuters Digital News Report[20] demonstrates that only a minority of people rely on Facebook (27%) or Twitter (6%) for news consumption. Reveilhac and Morselli (2020) also demonstrated that the reliance on social media goes hand in hand with the consumption of other online news sources and offline media sources, mainly free news.

Despite this low public reliance, social media can still impact the formation of public opinion. For instance, an evaluation of online and offline debate on the environment and climate, equality, the EU, and migration and asylum during the 2019 election campaign showed that social media influenced traditional media coverage and vice versa (Gilardi et al., 2021). Particularly, political actors can rely on social media to influence the public debate in voting and election campaigns. These platforms offer them a channel for addressing specific target groups and for communicating about politics (Popa et al., 2020). In particular, Rauchfleisch and Metag (2016) point to Twitter as an important factor in the agenda-setting process because many journalists source news on Twitter, thus acting as multiplicators for the content emitted by politicians.

In comparison to many European and non-European countries, politicians from Switzerland are latecomers regarding their reliance on social media (Rauchfleisch & Metag, 2015). This rather late involvement can partly be explained by the lack of a prominent online leader. For instance, until very recently, only seven cabinet ministers (Federal Council) had a Twitter account. It can also be explained by the fact that in

[20] See the statistics from the 2021 Reuters Digital News Portal here: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021/interactive

Switzerland, election campaigns are geographically restricted, thus lowering the incentive for politicians to make a lot of effort in virtual spaces. The federal system incentivizes politicians to be popular in their home canton rather than striving for national reputation. Furthermore, the consensus-based system also suggests that election campaigns are less important than direct democracy campaigns, thus favoring communication about policy issues rather than political personalities (Klinger & Russmann, 2017). For instance, the popular hashtag #parlCH is often used by politicians to comment on ongoing political business.

A majority of Swiss elected parliamentarians now have a Twitter account, however, showing that social media have become more important for all major parties compared to previous elections (2011 and 2015). There are nevertheless significant differences in the use of social media according to party affiliation. In general, politicians from the left-leaning Socialist Party are more present than those from the right-leaning Swiss People's Party. However, in 2019, the Swiss People's Party had the most followers on Facebook. So far, none of the parties have used social media optimally as most of them are not highly active (Gilardi et al., 2020) and have not adopted intense interactive practices (Klinger & Svensson, 2015). The role of social media in Swiss politics is very subtle and it remains worthwhile to conduct the discussion more broadly, exploring how particular social media have an impact on politicians' success, both in terms of vote share and media coverage, to understand the dynamics between the political, public, and media arenas.

Data and method

Historical data from election politicians

Twitter data offer several advantages in conducting our study. For instance, Twitter enables researchers to access historical data (which is still difficult when using Facebook). Moreover, while it is less popular among the public than Facebook, Twitter is known to be primarily used to discuss and cover political issues (Popa et al., 2020: 329; Gilardi et al., 2020), differentiating this platform from other platforms such as Facebook, Instagram, TikTok, etc. Twitter is also characterized by specific capabilities and writing conventions. Even though tweets must be only 280 characters long (it was 140 characters until 2017), they offer ways of engaging (and interacting) with other users through such devices as replying, mentioning, and retweeting.

We rely on an original dataset of historical tweets emitted by Swiss politicians who have held at least one parliamentary mandate during the most recent legislatures (2015–19 and 2019–2023). We considered the careers of politicians since the 2011 federal election to assess their incumbency status in 2015. Therefore, for each politician included in our sample, we assess whether they are a candidate, an incumbent candidate already active in parliament, or a politician ending his parliamentary mandate. The tweets are collected based on the platforms' application programming interfaces (API) with the function get_all_tweets() from the R language package academictwitteR (Barrie & Ho, 2021). Figure 4.2.1 displays the tweeting frequency of politicians included in our sample over time. It shows the relative tweeting frequency (i.e., the number of tweets divided by the number of accounts) and compares it with the raw number of tweets on a monthly basis.

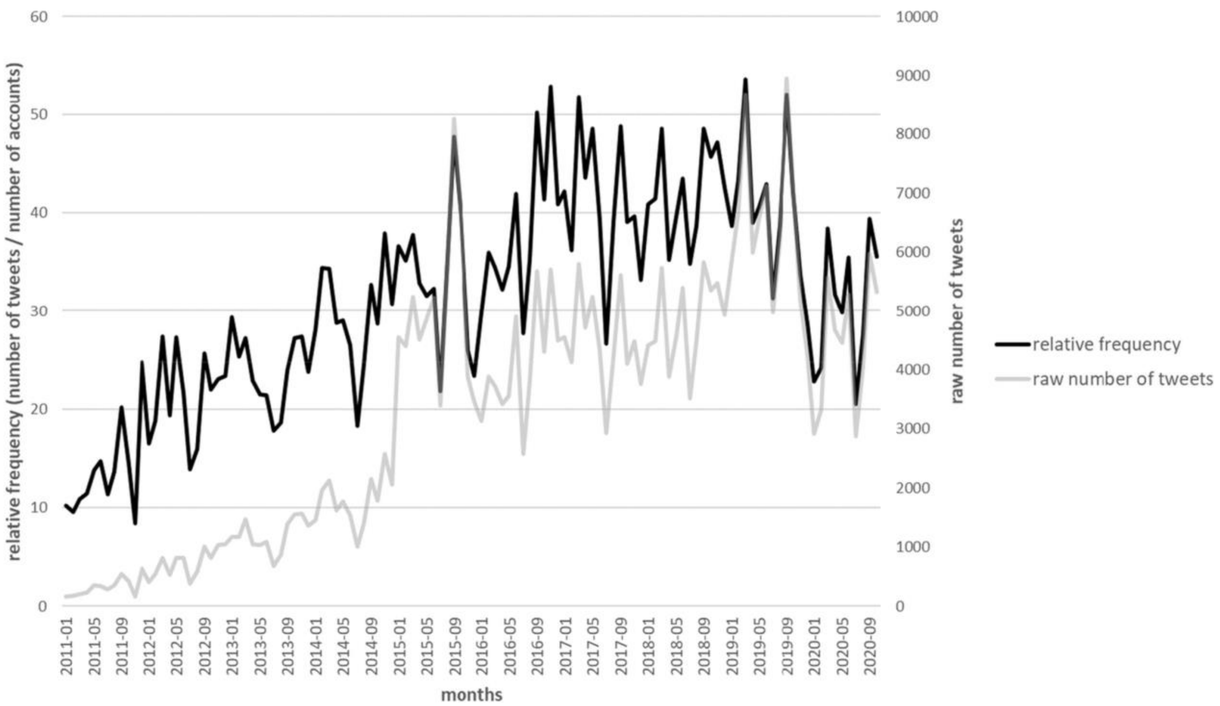


Figure 4.2.1: Distribution of the relative tweeting frequency (black line on the left y-axis) and the raw number of tweets (grey line on the right y-axis) by month.

In total, our dataset contains 227 unique politicians who were active in at least one legislature in either the National Council or the Council of States. The fact that we focus on politicians who have succeeded in building a political career means that our sample is biased toward politicians who are more successful. Table 4.2.1 below describes our sample of politicians with respect to their overall presence. For instance, we note that more than half of the politicians in our sample – taken from both legislatures – possess a

Twitter account (66% in 2015 and 70% in 2019). More than 70% of newly elected politicians in both legislatures are active on Twitter, thus pointing to the perceived increased importance of owning a Twitter account. We also specify the mean rank on the National Council based on each candidate's constituency (party and canton considered). We see that having a Twitter account was not associated with a higher ranking in 2015, but that it became an important asset in 2019 (differences are statistically significant at $p < 0.05$ for both legislatures). Understanding under what conditions politicians' presence on Twitter is advantageous for reaching top positions in the list will be at the center of our analyses. Table 4.2.1 also displays distributions with respect to incumbency and attrition (candidacy not renewed and no re-election) for both legislatures.

Table 4.2.1: Descriptive statistics for the sample comprising politicians who have occupied a seat in at least one of the two last legislatures (2015–19 and 2019–2023) in either the National Council or the Council of States

	With Twitter account		**Total (with and without Twitter account)**	**Difference**
Candidates in 2015	228	(66%)	343	115
Candidates in 2019	210	(70%)	299	89
Newly elected in 2015	52	(73%)	71	19
Elected and incumbent in 2015	109	(64%)	171	62
Not renewed in 2015	16	(30%)	53	37
Mean rank on NC list in 2015	4.04 versus 2.99 without*			
Newly elected in 2019	79	(81%)	97	18
Elected and incumbent in 2019	105	(69%)	153	48
Not renewed in 2019	38	(59%)	64	26
Mean rank on NC list in 2019	2.76 versus 3.33 without*			
Elected in 2015 and 2019	101	(74%)	137	36
Elected and incumbent in 2015 and 2019	65	(75%)	87	22

Note: The values in the table describe the two parliamentary chambers, except for the mean rank from vote share obtained in each election when running for the National Council (NC). * lower values indicate a higher ranking, while higher values indicate a lower ranking.

Analytical steps

Initially, deriving our information from Twitter meta-info and tweet content, we presented descriptive findings about the evolution of politicians' online interactions with other online audiences. To do so, we focused on politicians' replies, since these are emblematic of politicians' responsiveness to other users' concerns. In complement to this, we investigated the evolution of the audiences mentioned by politicians to account for the types of users to which politicians paid particular attention. Both measures are of interest for contextualizing the uses which politicians make out of Twitter. We considered the time span from 2011 to 2019, as some politicians included in our sample were already active in Parliament in 2011. Furthermore, the early 2010s coincided with the first rise in the use of social media by Swiss politicians (Rauchfleisch & Metag, 2015).

To code the audiences, we downloaded the Twitter profiles of users that had either replied to or been mentioned by politicians during the campaigning period and when they were active in Parliament. We identified the actors mentioned using regular expression to extract the username following the '@' sign. Usernames replied to by politicians were identified using the meta-information provided by the Twitter API. The full profile description was retrieved using the lookup_users() function from the R package rtweet (Kearney et al., 2020). We only coded the users that are replied to or mentioned more than 5 times (n = 9.412). Since our objective is to investigate the extent to which interactive behavior impacts politicians' success, we coded the users that sufficiently triggered politicians' attention so that they were replied to or mentioned in their tweets. Another possible strategy would have been to rely on politicians' follower networks to identify an "attentive public" or "party supporters" (Barberá et al., 2019). Albeit potentially skewed towards an elite sample, we avoided focusing only on partisan accounts, while also keeping relevant users from the perspective of politicians.

The labeling of users was done manually and based on their profile descriptions (as well as their locations and personal link fields provided by Twitter). The categories retained for labeling are inspired by previous works (e.g., Keller, 2020), but new categories found to be prominent (e.g., head of business/entrepreneur, consultant/communication manager, and experts) have been added because they were found to be important through the manual coding of users (Appendix 3 provides the description of the categories for the manual coding). When coding the users, if more than one category applied (e.g., local politicians can also be heads of business), we always selected the category that was of greater interest from the politicians' perspective. When this logic could not be applied easily, we selected the category that appeared first in the profile description. In the result section, we display only prevalent major audiences. The coding was done by a single coder, but a random sample of 100 tweets was verified by a second coder. The inter-coder reliability is very high (96%).

Then, we present regression analyses to assess the extent to which the level and style of Twitter activity in both legislatures impact politicians' success: most notably politicians' ranking and media coverage. Because we used longitudinal data about each Swiss politician who was elected at least once in 2015 and 2019, our dependent variables were built to fit models where observations are split by legislatures.

The different methodological steps to collect and prepare the data are summarized in Figure 4.2.2.

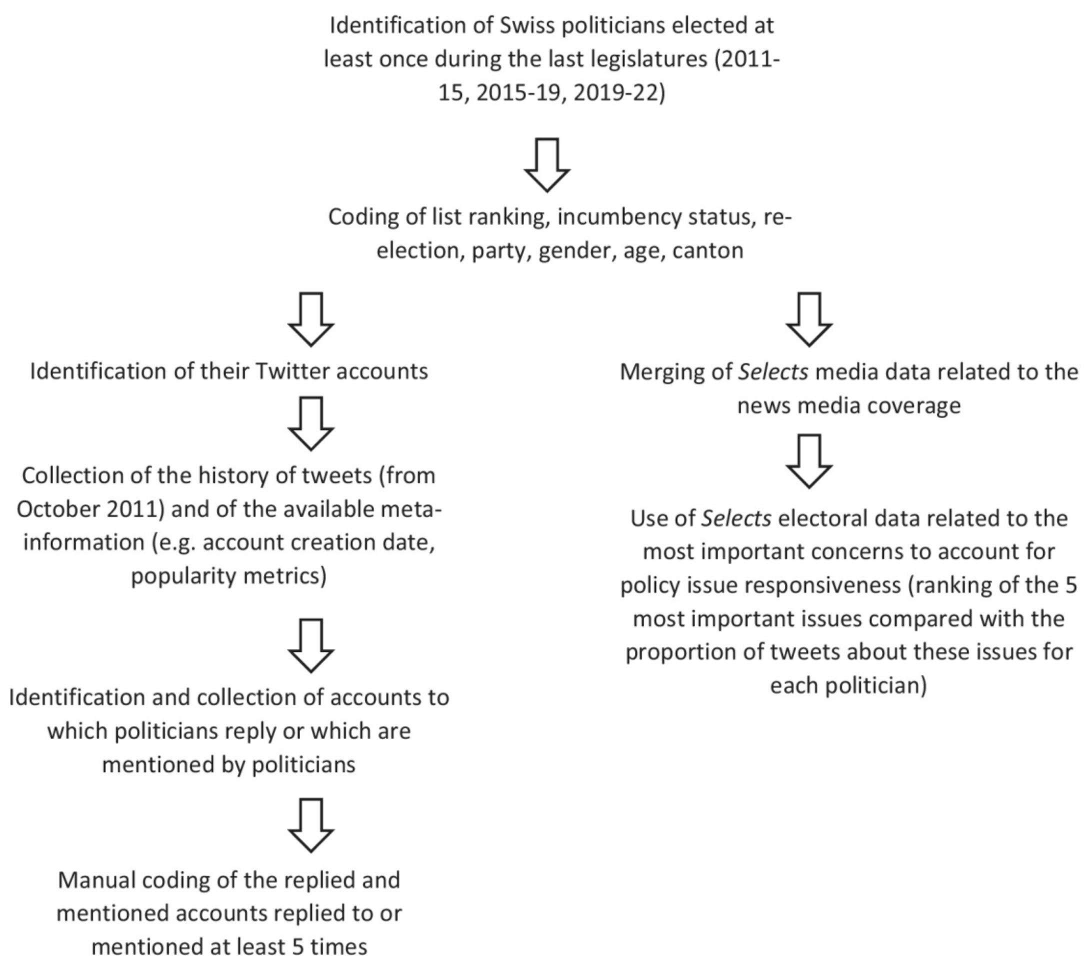


Figure 4.2.2: Methodological steps undertaken to collect and prepare the data.

Outcome variables: politicians' success and media coverage

Our outcome variables aim to model two important components of politicians' success, namely i) politicians' ranking, as well as ii) the level of media attention received by each politician during election campaigning periods[21].

Switzerland is a federal State with three levels of power (federal, cantonal, and communal). There are elections at each of these levels, each following its own rules. At the federal level, the vote in the lower house, the National Council, is proportional, while the election in the upper house, the Council of States, is by majority vote. Politicians' ranking is derived from the overall votes gained by each politician in the constituency in which he/she is competing. Compared to raw vote count, the rank has the advantage of improving comparability between election periods (i.e., it is not affected by changes in the population participating in elections).

Due to data availability restrictions, we will only look at politicians' media coverage during election periods. The data were collected and coded by the Selects team responsible for the media analysis[22], whose aim was to identify the most important actors (persons and parties) during the election campaign and to investigate which topics were covered by the media during the election campaign. The level of media coverage was measured by the number of articles published about the candidates during the election campaign periods (from January to November 2015 and 2019)[23]. We calculated z-scores (with a minimum at 0) for the media coverage in each election because the method for coding the articles changed between both years surveyed. The number of articles was matched based on the full name of politicians (all politicians included in our corpus were coded by the Selects team).

The regression models are built on Poisson regression because the distribution of our dependent variables is not normally distributed. The mean ranking is situated at the third place (with a standard deviation of 5) and the maximum ranking is at the 64th place. The

[21] We do not include media coverage for the whole legislature. Such investigation would require additional data collection and coding of traditional media article which goes beyond the scope of our paper.

[22] The technical report for the media analysis can be found here: https://www.swissubase.ch/en/catalogue/studies/13846/16968/datasets/1187/1877/files/document/18027/9151/physicalFile

[23] The number of articles has been collected and coded in the realm of the Swiss election studies by the Selects (Swiss Electoral Studies) survey team and the mandated research groups. The data are accessible on FORSbase under the project reference 13846 (for 2019) and 12447 (for 2015). For more information see: https://forsbase.unil.ch/

mean media coverage is situated at a z-score of 1 (with a standard deviation of 2) and the maximum is at a z-score of 12.

Furthermore, we present pooled-models, instead of panel-models, because the last election cycles were particular in view of the high replacement of the elite. To make the pooled-models reliable, we included a binary variable indicating the year of the legislature (2015 or 2019), as well as an interaction term between the legislature and the tweeting frequency[24]. To do so, we retrieved all tweets sent by politicians when they are active in Parliament or/and campaigning in an election. This allowed us to record the number of contributions politicians made to the platform during the legislatures and the campaigning period. Instead of using the raw count of emitted tweets, we measured the frequency of use of social media by taking into consideration the actual number of tweets each account made during the observation timespan divided by the number of days during the observation timespan.

Independent variables: political style of communication and reactions to politicians' messages

Concerning the communication style, we included the following variables:

First, we accounted for politicians' interactive practices on social media by including the proportion of replies emitted by a politician as an indication of their level of interaction with other users. We also specified with whom politicians are interacting by including the proportion of replies to prominent online audiences, namely journalists, national politicians, local politicians, and citizens.

Second, we also included a measure of political responsiveness for each legislature as a control variable. We modeled responsiveness as the difference between citizens' and politicians' emphasis on the five most important problems: ‘economy’, ‘environment & energy’, ‘EU, Europe’, ‘immigration & asylum’, ‘public health’, ‘social security/welfare

[24] The two last legislatures were noteworthy in terms of the renewal of the political elites. Indeed, in 2015, the number of parliamentarians not renewing their candidacy was one of the highest since 1987 (only 15% of members renewed their candidacy: 1/8 for the National Council and 1/4 for the Council of States). This constitutes a major reason why the regression analyses will focus on the two last legislatures and will be based on “pooled” models (instead of “pseudo-panel” models). In 2019, a record number of 4,652 candidates ran for the National Council. This increase can partly be explained by the greater investment of women candidates, but perhaps also by the increased number of younger candidates, notably in line with the rise of the “Green tide” (e.g., Bernhard, 2020).

state'. The public concerns were derived from electoral survey data[25], while machine learning was used to classify politicians' tweets according to similar policy issue categories (see Appendix 1 for more detail about the text classification and Appendix 2 for a description of the topic distribution).

Third, we accounted for politicians' information dissemination practices on social media by calculating the proportion of tweets containing links (proportion of links). To do so, we used regular expression matching for URLs (e.g., http[s]* or www*).

Concerning the reactions to politicians' Twitter messages, we included a measure capturing the size of politicians' active online audience by measuring the proportion of retweeted politicians' messages and the proportion of politicians' messages that were ' favorited.' We considered these measures a good proxy for the attention that politicians attract over time. We used these measures instead of the number of followers, which provides only the last updated statistic (this impedes us in modelling the increasing and decreasing number of followers over time). The popularity measures (likes and retweets) were updated by August 2021, as we collected the data historically at that time.

Control variables

We considered whether the politician already had experience as an elected parliamentarian (referred to as incumbent). We also accounted for whether the politician was active and/or running for the Council of States (upper House) or the National Council (lower House).

We also included control variables in our analysis, namely politicians' gender and a control for regional differences based on the language that each politician used most often on Twitter.

We further included politicians' left–right position. This last variable is based on their political affiliation (the scale ranges from 1 'left' to 8 'right'). We used the 2019 Chapel Hill expert survey[26] to rank political affiliations along the LRGEN ideological stance[27] (1 for

[25] Citizens' ranking of the five most important problems facing the country is derived from the survey items asking respondents "What is currently the most important issue facing the country?". We used the data from the last waves of the Panel Survey conducted in 2015 and 2019 by the Selects team (see project references above).

[26] For more information about the data and the codebook see: https://www.chesdata.eu/2019-chapel-hill-expert-survey

[27] The party acronyms read as follows: GPS/PES for the Green Party, SP/PS for the Social Democratic Party, GPL/PVL for the Green Liberals, CVP/PVC for the Christian Democratic People's Party, EVP/PEV for

GPS/PES, 2 for SP/PS, 3 for GPL/PVL, 4 for CVP/PVC, 5 for EVP/PEV, 6 for BDP/PBD, 7 for FDP/PLR, 8 for SVP/UDC).

Results

Changes in the audience politicians interact with

Of the 392 elected parliamentarians who have occupied a position in the National Council (lower chamber) or Council of the States (upper chamber) during at least one of the two last legislatures (2015–2019 and 2019–2023), 227 (58%) possessed a Twitter account. There are big discrepancies in politicians' Twitter activity. On average, politicians in our sample sent 714 tweets per year with a standard deviation of 1635 tweets between 2015–19, and 1231 tweets per year with a standard deviation of 3009 tweets between 2019–2023. Furthermore, 70% of politicians emitted less than one tweet per day.

the Evangelical People's Party, BDP/PBD for the Conservative Democratic Party, FDP/PLR for the Liberals, SVP/UDC for the Swiss People's Party.

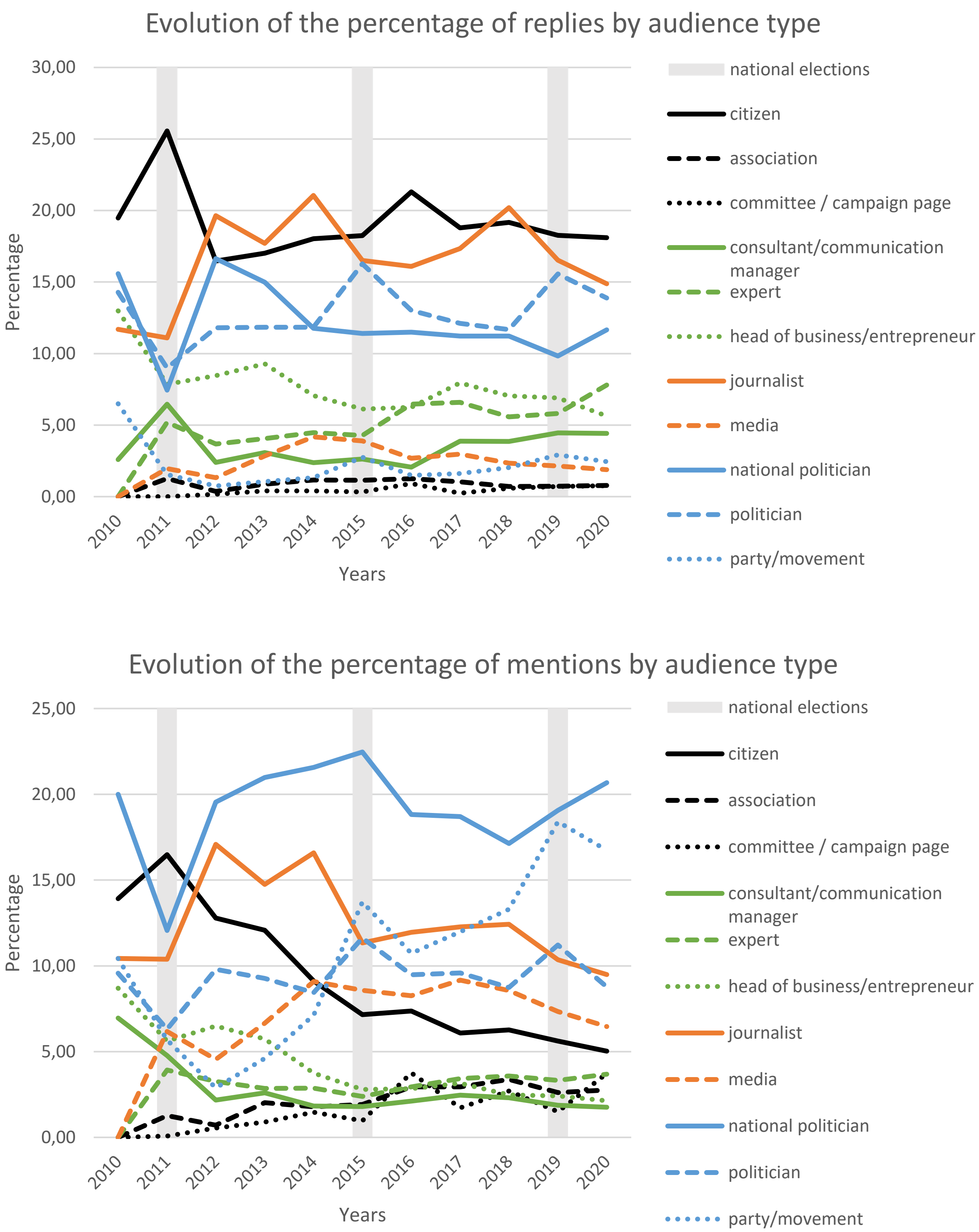


Figure 4.2.3: Evolution of the share of politicians' replies (upper pane) and of the share of politicians' mentions (lower pane) by audience type. Note that only original tweets (not retweets) are included.

For instance, politicians in our sample had an average share of replies of 15% with a standard deviation of 16% (the average share of replies slightly decreased between the

2015–19 and 2019– 2023 legislatures). Furthermore, on average, 59% of their tweets contained urls to external pages (with a standard deviation of 20%). Concerning measurements of popularity, politicians' tweets were liked on average 40% of the time (with a standard deviation of 21% and an increase between both legislatures) and retweeted 58% of the time (with a standard deviation of 24%).

Figure 4.2.3 displays the evolution of politicians' replies to and mentions of prominent audiences. In total, 9,412 unique accounts with which politicians interacted at least five times (through replies or mentions) were identified. There is a large overlap between the audiences to which politicians replied and the audiences mentioned by politicians. There are several trends worth mentioning.

Figure 4.2.3 shows that citizens, journalists, and national politicians represented the prevalent audiences to whom politicians replied (upper pane) and who politicians mentioned in their tweets (lower pane). We also observed that mentioning is less prevalent among politicians than replying. Furthermore, trends associated with replying and mentioning seem to be complementary. For instance, there was a steady decrease in the proportion of replies to (and mentions of) citizens over time, but a steady increase in the mentions of (and replies to) national politicians. Politicians also tended to reply to and mention other local politicians on Twitter, especially close to election periods (2015 and 2019). Mentions of parties/movements were more prevalent than replies to them. It also seems that users with a communication profile (mostly consultant/communication managers, but also experts) have recently become more prevalent in politicians' replies. However, the proportion of replies to and mentions of head of business/ entrepreneur seems to have vanished over time.

Additional observations can be made from Figure 4.2.3. For instance, mentions of associations (which also includes organizations and NGOs) and committees (especially citizen committees for voting purposes) are becoming more prevalent over time (notably since 2015), although politicians barely reply to them. Other trends are not displayed in the Figure for reasons of parsimony. For instance, the mention of foreign parties/movements and of political institutions/ embassies became more prevalent after 2017. Finally, we noted that replies to lawyers (or other law-related professionals, such as jurists and legal counsels) peaked around elections.

Effect of Twitter use on politicians' success over legislatures

Table 4.2.2 presents regression analyses to assess the extent to which the level and style of Twitter activity in both legislatures impacted politicians' success, namely politicians' ranking and media coverage. Overall, our models show quite different scenarios for explaining politicians' success according to their ranking (beyond the campaigning period and during the campaigning period) or media coverage (during the campaigning period). With respect to politicians' ranking, the coefficients in Table 4.2.2 suggest that positive values on the dependent variable mean that politicians had greater political success (more vote share), while negative values suggest that politicians had less success (less vote share). The ranking model shows that the tweeting frequency is only significant for the 2015–2019 legislature, but not for the 2019–2023 legislature. Higher tweeting frequency is associated with a lower ranking, thereby, a greater political success. The uses of Twitter impacted politicians' ranking in several ways. For instance, politicians who had a higher proportion of replies in relation to the media were likely to have more political success, thus, pointing to the long-term co-evolution between political and media agendas. Furthermore, higher levels of responsiveness to citizens' concerns were significantly associated with higher success, while link sharing was not associated with more success. Moreover, the number of favored politicians' messages was not statistically significant, contrary to the number of retweets of politicians' messages. As for the control variables, we did not detect any gender effect on the ranking position, nor did we observe an effect of the type of Council. However, we observed that incumbents had more political success. Finally, left-leaning politicians benefited more from their reliance on Twitter with respect to their political success. Although this effect can be explained by the higher rate of Twitter reliance from left-leaning politicians (e.g., the Swiss Peoples' Party is known to be more active on Facebook than on Twitter), the tweeting frequency variable controls for this imbalance.

Table 4.2.2: Poisson regression models of politicians' ranking (whole period covering the legislatures and election campaign) and politicians' media coverage on the study variables (focus on the election campaigning period)

	Politicians' ranking (whole)			Politicians' media coverage (election)			Politicians' ranking (election)		
	Std. Coef. (Std. Error)	***p*-value**		**Std. Coef. (Std. Error)**	***p*-value**		**Std. Coef. (Std. Error)**	***p*-value**	
Constant	0.669 (0.25)	0.008	**	−1.181 (0.746)	0.113		0.247 (0.338)	0.466	
Communication style:									
Proportion of replies	−0.005 (0.003)	0.089		−0.003 (0.006)	0.624		0.006 (0.003)	0.068	
Proportion of replies to journalists	−0.001 (0.003)	0.627		−0.005 (0.005)	0.352		−0.004 (0.003)	0.191	
Proportion of replies to media	−0.011 (0.004)	0.008	**	−0.013 (0.011)	0.215		−0.001 (0.003)	0.776	
Proportion of replies to national politicians	−0.003 (0.002)	0.118		0.006 (0.003)	0.044	*	−0.003 (0.002)	0.128	
Proportion of replies to local politicians	0.003 (0.003)	0.22		−0.003 (0.004)	0.377		−0.003 (0.002)	0.216	
Proportion of replies to parties	−0.009 (0.005)	0.076		−0.011 (0.008)	0.157		−0.003 (0.003)	0.364	
Proportion of replies to citizens	−0.001 (0.002)	0.635		−0.008 (0.006)	0.136		−0.008 (0.003)	0.004	**
Responsiveness to public concerns	0.112 (0.045)	0.012	*	−0.018 (0.083)	0.828		0.151 (0.044)	<0.001	***
Proportion of links	−0.002 (0.002)	0.203		−0.005 (0.004)	0.247		−0.004 (0.002)	0.019	*
Reactions to politicians' tweets:									
Proportion of retweeted politicians' messages	0.003 (0.002)	0.041	*	0.009 (0.004)	0.023	*	0.004 (0.002)	0.017	*
Proportion of favourited politicians' messages	−0.001 (0.002)	0.601		0.008 (0.003)	0.013	*	−0.001 (0.002)	0.34	
Legislature dummy:									
Tweeting frequency	0.008 (0.031)	0.804		−0.069 (0.109)	0.523		0.052 (0.037)	0.156	
Legislature dummy: 2019–22 (ref. 2015–19)	−0.336 (0.095)	<0.001	***	−0.164 (0.477)	0.731		−1.345 (0.191)	<0.001	***
Tweeting frequency x legislature dummy	0.17 (0.034)	<0.001	***	0.357 (0.103)	<0.001	***	0.163 (0.037)	<0.001	***
Control variables:									
Gender: woman (ref. man)	0.076 (0.061)	0.215		0.152 (0.123)	0.217		−0.017 (0.064)	0.792	
Regions: Latin (ref. German-speaking)	−0.086 (0.083)	0.297		−0.433 (0.18)	0.016	*	−0.176 (0.087)	0.043	*
Left–right position	0.082 (0.013)	<0.001	***	0.012 (0.024)	0.635		0.098 (0.013)	<0.001	***
Incumbent: yes (ref. no)	1.033 (0.07)	<0.001	***	−0.559 (0.162)	<0.001	***	1.017 (0.069)	<0.001	***
National Council (ref. Council of States)	−0.071 (0.136)	0.601		0.422 (0.17)	0.013	*	−0.12 (0.14)	0.391	
Adjusted R2:	0.28 (28%)			0.55 (55%)			0.28 (28%)		
Number of observations:	339 observations			321 observations			321 observations		

Note: significance levels read as '***' for p < 0.001; '**' for p < 0.01; '*' for p < 0.05.

The model for politicians' ranking is also based on politicians' tweets from the period covering the election campaigns (last model in Table 4.2.2). Indeed, politicians' use of social media likely differs across electoral periods. The focus on the electoral periods also shows that the proportion of replies to citizens negatively impacted on politicians' ranking. However, in line with the first model for the whole legislature period, we observed a positive relationship between higher levels of political responsiveness to citizens' concerns and political success. Furthermore, link sharing also helped gain political success, possibly because links shared by politicians also enabled them to publicize their political agenda. These findings could be explained by the fact that replying to lay citizens can be detrimental to politicians' campaigning image as it shows points of contention with given sub-groups of Twitter users. However, showing responsiveness to public concerns and sharing information (e.g., links to party program or major events) grant politicians more political success. Table 4.2.2 also shows that the impact of the proportion of politicians' retweeted messages matters for explaining a higher political success. The control variables display similar trends as those found in the model covering the whole period. In addition, the last model shows that politicians from the German-speaking regions benefitted more from their involvement on Twitter than politicians from French and Italian speaking regions.

Considering politicians' media coverage, the coefficients in Table 4.2.2 suggest that positive values on the dependent variable mean greater press coverage for politicians during the campaigning period, while negative values imply little press coverage. The model for media coverage shows that the proportion of replies to national politicians had the most significant impact on politicians' ranking. Again, this could indicate an interdependence between political and media agendas. However, the proportion of replies to journalists did not have a significant impact on the levels of media coverage. Contrary to the models explaining political success, we observed no relationship between higher levels of political responsiveness to citizens' concerns and media coverage. Table 4.2.2 also shows that the proportion of retweeted and favored politicians' messages is statistically significant and positively associated with higher media coverage. Furthermore, the interaction variable between tweeting frequency and the legislature informs us that the different trends were prevalent during the two legislatures, as the tweeting frequency had more impact in the 2019–2023 legislature than in the previous legislature. Interestingly, incumbents did not necessarily benefit from a higher media coverage compared to new candidates. This can be explained by the high turnover in the elite staff during the two last elections (e.g., many politicians did not represent as candidates), but also by the fact that new topics (e.g., climate change and gender issues) were at the forefront of public debates. Finally, unlike the previous models, there was no effect of political leaning on media coverage, thus, pointing to the neutrality of the press concerning partisan positioning.

In order to control for the potential effect of the variable related to politicians' responsiveness to public concerns, we also provided the regression models without the inclusion of this variable (see Appendix 4). No notable change is observed with respect to the direction of the coefficients for the first model (politicians' ranking over the whole period). We only note that removing the responsiveness variable reinforces the significance level of the variable related to the effect of the proportion of replies to parties, thus consolidating its negative impact on politicians' ranking. Regarding the second model (politicians' media coverage), we do not observe any change, neither in the direction nor in the statistical significance of the coefficients. Concerning the third model (politicians' ranking during elections), we noted that the removal of the responsiveness variable reinforces the statistical significance of the variable related to the proportion of replies, as well as the significance of the variable related to the proportion of retweeted politicians' messages. Furthermore, the removal of the responsiveness variable introduces a statistical significance for the negative effect related to the proportion of favourited politicians' messages. These findings suggest that removing the responsiveness variable translates into an increased statistical importance of the impact of variables linked to the public reactions to politicians' tweets (in terms of retweets and favourites). Moreover, the removal of the responsiveness variable cancels the statistical significance related to the proportion of link sharing and to regional differences. This suggests that the inclusion of the responsiveness variable enables us to show the positive impact of information sharing and to highlight regional differences.

Discussion of the main findings

According to our findings, it is very unlikely that the results of elections are determined by what politicians do (or fail to do) on social media. However, we might also expect that politicians' success does not solely depend on the frequent use of social media, but rather on the communication strategy they adopt. By adopting certain interactive communication styles, politicians can become more visible and consolidate their public image.

In terms of the changes in the audiences that politicians interacted with across successive legislatures and elections, the changes in the proportions of particular audiences that politicians replied to and mentioned underlines the fact that politicians are still adapting their political communication styles. The patterns of interactions of politicians with other users will likely continue to evolve and be influenced by societal trends, particularly because the share of politicians relying on social media is still growing.

Concerning political success measured in terms of political ranking, the multivariate analyses have shown that tweeting frequency per se does not decisively impact politicians' success. However, we also show that the effect of Twitter depends on the usages related to differentiated styles of political communication. For instance, politicians are likely to achieve higher levels of political success in a long-term perspective if they reply to the media. Correlatively, politicians are more likely to benefit from a higher media coverage if they engage in reply behavior with other national politicians. Furthermore, politicians have higher level of political success during electoral periods if they abstain from replying to other non-political and non-media users. However, responsiveness to citizens' concerns is also an adequate communication strategy to reach higher levels of political success.

Overall, replying to other audiences had a positive impact on politicians' success and media coverage, which could be explained by the fact that replies on Twitter also tend to contribute to a politician's image of being responsive to other audiences' concerns. For instance, a higher share of replies from a media account is strongly related to increased political success for the politician. Interactions with other users close to politics can help build a strong network of mutual support, thus promoting online political debate that can be of interest to the general public (Keller, 2020). Furthermore, replying to citizens is especially important for increasing political success, especially when focusing on election campaign periods (Tromble, 2018). Hence, avoiding answering people on social media can have a negative long-term impact on reputation.

Concerning media coverage, it is interesting to note that political incumbents have benefited less from higher levels of media coverage during the last elections. We think that these trends indicate that the last elections favored the overhaul of the political elite.

This reflects recent political trends affecting the composition of the parliamentary elite in Parliament (Bernhard, 2020), most notably the trends favoring an overhaul of the political elite and encouraging more progressive politics. The new political staff might also be more likely to generate public attention and, thus, to be retweeted. This could explain the stronger positive relationship found between higher levels of political responsiveness and political success when focusing on the campaigning period. This would suggest that politicians with lower rankings are merely challengers who need to emphasize the congruence of their agenda with public concerns to increase their chances of being elected. Additionally, the positive impact of the proportion of politicians' messages that are favorited is in line with recent literature observing that more visible political tweets are more likely to be spread (or taken on) by journalists (e.g., Metag & Rauchfleisch, 2017). The negative relationship between a higher proportion of replies to citizens and media coverage is hard to explain. We think it can result from politicians' incentives for strategic communication online that pushes them to choose between being pro-active with audiences closer to politics (e.g., other politicians, journalists, or media) versus being responsive to citizen demands and concerns. As such, politicians who dedicate more time to interacting with citizens tend to get less media attention.

A somewhat surprising finding is that the proportion of replies to journalists had no effect on political success, in terms of either ranking or media coverage. However, the literature depicts journalists as important actors with whom politicians interact on social media (Spierings et al., 2018; Keller, 2020). A potential explanation is that politicians' short-term engagement with journalists (e.g. during the campaign) is not enough to build a longstanding reputation with media actors. Another possible explanation echoes politicians' perception of the low media responsiveness to their social media efforts. Indeed, current trends found in survey data of political candidates[28] show that social media are perceived to be useful for convincing voters (above 70% agree or strongly agree), attracting attention to salient policy issues (80%), and communicating personal views on politics (87%). Social media are nevertheless considered as less well-suited for learning what citizens are concerned about (60%) and sharing opinions and activities to be picked up by traditional media (48%).

[28] Please refer to the 2019 Swiss wave of the Comparative Candidate Survey in which political candidates are asked to position themselves on several propositions regarding the usefulness of social media.

Concluding remarks and outlook

Our study has several limitations. First, our analyses focus on a single social media platform and on a single country. Second, we only focus on a specific measure of political success based on the share of electoral votes. Third, other independent variables not included in our analyses could impact the success of politicians (e.g., scandals can go viral online and affect politicians' reputability and career). Future studies could address these limitations. For instance, a cross-national perspective could help assess the effect of social media on politicians' success by taking the political and institutional contexts into account. Furthermore, future studies could envisage building indexes to account for a variety of indicators that reflect politicians' success. For instance, greater success could be indicated by a number of things: a politician's higher position on the party list, re-election, or promotion to the higher chamber of Parliament. This "success index" could thus account for the evolution of politicians' status over various legislatures. It would also enable us to conduct a more robust causal (or "pseudo-panel" model) where observations are compared to the first legislature. Similarly, future studies could also envisage building a "media coverage index," taking into account whether politicians have gained media coverage over the course of a legislative term.

Finally, we cannot entirely exclude the possibility that the overlap between the policy issues that are salient in politicians' communication and in the public opinion results from the dynamics of politicians' communication. In other words, the mere overlap could be the result of the audience's responsivity to topics politicians try to promote online. Given the nature of our data, we cannot directly test the "true" direction of the causality: do politicians react to citizens' concerns or do politicians make these concerns salient by communicating about them? In our study, the responsiveness of politicians corresponds to an overlap between politicians' communication and citizens' concerns. As politicians are essentially involved in a unidirectional style of political communication (Graham et al., 2013), as demonstrated by the rather low proportion of replies to other users, we can be confident that politicians in Switzerland merely rely on social media to share opinions about policy issues that are important to them or that are salient on the party agenda. In line with the idea of a feedback loop between politicians and the public emphasizing the circumstances in which public opinion may facilitate political discursive elements (Reveilhac & Morselli, 2022), we assume that an increased overlap between politicians' communication and citizens' concerns in terms of policy issues is likely to attract more attention from the public (citizens or media actors).

Despite these limitations, we believe that our study makes two important contributions. Firstly, by using historical data we can provide an exhaustive picture of political communication trends over time. Studies covering such a large period of Twitter uses by elected politicians are still rare. This endeavor is important as it enables us to assess which communication patterns are perceived as well-suited by politicians, thus complementing political candidates' surveyed perceptions of the usefulness of social media. Secondly, the focus on political success enables us to provide a complementary picture to the extensive research already undertaken on political candidates' success during election campaigns. Permanent campaigning has become an important feature of politics, and the efforts politicians put into social media beyond the heated election periods demonstrate their aim of seeking greater accountability and (offline) popularity. In this view, our study is among the few which enable researchers to grasp how and whether social media communication can lead to success offline.

4.3 Political Polarisation on Gender Equality: The Case of the Swiss Women's Strike on Twitter[29]

Introduction: Gender Equality Discourses on Social Media and in the General Public

Social media are widely used in social movements for fast information diffusion and for raising attention to specific social and political claims (Sini, 2017). For instance, the use of social media was an essential tool for activism during the Arab Spring and for the Black Lives Matter movement. In the gender equality context, one of the most famous examples of the centrality of social media is the emergence and consolidation of the MeToo movement, which aims to raise awareness around sexual harassment and assault. While these examples show that social media give a voice to social movements, the extent to which they represent public opinions remains less clear.

Indeed, most studies investigating social media discussions around social movements rarely compare online (e.g., opinions expressed on social media) to offline (e.g., opinion surveys) trends. Yet, comparing opinions expressed online and offline is especially important given that the two may not often converge. While social media merely serve to connect like-minded users who already share and support similar ideas and concerns (Cinelli et al., 2021), opinion surveys focus on representative samples (of sub-groups) of the population. Further, most social media content about social movements' agendas tends to be produced and discussed by a minority of politically engaged users (Huges & Wojcik, 2019) who do not necessarily represent the broader public (Tucker et al., 2018). In addition, social debates are not only the product of individuals (i.e., as measured in surveys), but also of institutionalised groups, political actors, media, journalists, organisations, and other particular sets of actors (Tucker et al., 2018) which are also active on social media. It is thus important to assess the extent to which social media discussions stemming from different actors reflect the prevalent concerns raised by the broader public. This is particularly important as the reliance on social media by these different actors to express their views can lead to an increased polarisation of societal debates (Quattrociocchi, Scala & Sunstein, 2016).

[29] This chapter is a slightly adapted version of the article that has been published as M. Reveilhac and L. Eisner (2022): "Political Polarisation on Gender Equality: The Case of the Swiss Women's Strike on Twitter", Statistics, Politics and Policy.

In the present study, we focused on social media content centred around a social movement for more gender equality. In a first descriptive step, we examined which user groups mobilise on social media surrounding a social movement. We then focused on three more substantial research interests. Namely, we looked at the extent to which the involvement of politically active users reflects trends from public opinion. We then investigated the extent to which political polarisation about gender equality is increased on social media. Finally, we investigated what the content of social media data tells us about the potential of social media to promote the claims of a social movement. To do so, we examined the discursive content of social media actor groups, upon which we also mapped claims from a representative sample of citizens concerned about the gender issue. Juxtaposing social media data and opinion survey data enabled us to assess the link between the expressed digital rhetoric and the validity of surveyed opinions.

To achieve our goals, we focused on Twitter content which centred around the national women's strike for gender equality that took place in Switzerland in June 2019. The data collection is based on relevant actor groups (e.g., strike organisation committee and its followers, but also politicians and other users tweeting about the strike) and keywords (e.g., hashtags pointing to the strike). The collected data covers the period from January 1st to December 31st, 2019. To examine our research questions, we conducted several coding steps. First, we identified influential users participating in the debate based on the tweeting frequency and we undertook a comprehensive manual coding effort to classify these users into relevant actor categories. We then looked at the association between the online salience of politically engaged users' gender equality discourse and the opinions of citizens surveyed about gender equality while accounting for political positioning. Finally, we relied on factor analysis to display the argumentative features surrounding gender equality issues according to social media actors and to a representative sample of citizens concerned by gender equality.

Background

The Use of Social Media by Social Movements Promoting Gender Equality

There is currently a consensus that social media are widely used in social movements as they serve for fast information diffusion and for raising attention to specific social and political claims (Soares & Joia, 2015). Consequently, social media have emerged as a key venue for political debates. Social media, especially Twitter, serve as a place to engage in civic-related activities, notably by using viral capabilities such as #hashtags, which are often referred to with the concept of hashtag activism (Xiong, Cho & Boatwright, 2019). In the case of feminist movements, the topic of the present study, Dixon (2014) traced the different ways that hashtag feminism has been enacted, such as through the sharing of personal experiences and the challenging of dominant discourses. In our study, we used the case of the women's strike movement in Switzerland, which also developed specific hashtags to promote fast information diffusion and raise attention to the social and political claims of the movement (e.g., #Frauen*streik in German and #Grevedesfemmes in French).

Social media platforms constitute an indispensable tool for social movements to organise their actions and mobilise public opinion around particular claims to promote social change (Eltantawy & Wiest, 2007; Poell et al., 2015). While social media have been a major tool for spreading equality claims and actions since the MeToo movement in 2017 (Modrek & Chakalov, 2019), a recent study also showed that levels of modern sexism among the American mass public did not respond to the rise of MeToo (Archer & Kam, 2020). One reason to this might be that gender equality related issues are typically a topic on which political polarisation is strong, especially on social media where the progress achieved in women's rights and gender equality has become the target of a backlash driven by anti-gender users and right-wing populists (Wallaschek et al., 2022). Despite this, we still have limited knowledge about how the scope of ideologies and content from social media discussions can influence the reach of social movements' claims in public opinion.

How Social Media Content Surrounding Social Movements Connects to Public Opinion

The details of social movements engagement on social media with the public discourse in society are of utmost importance as the goal of a social movement is to bring concerns to the forefront of the political agenda. In this view, the fact that social media users are typically unrepresentative of the general public (Mellon & Prosser, 2017) does not mean that social media content is unrelated to what the public thinks, notably because influential user groups also have the potential to influence public opinion (Weeks, Ardèvol-Abreu & de Zúñiga, 2017).

Figure 4.3.1 illustrates the conceptual framework underpinning our study. It is inspired from the study of Gordon (2015). In Gordon's model, social movements rise from the development of a community of interest in response to a set of underlying grievances stemming from the elaboration of an agenda. Until a call to action takes place, following an event raising the claims of the social movement, the community is likely to either lie silent or even fade away. Likewise, a call to action that is not supported by a community consists of an individual cause, not a social movement. The outcome of the movement is likely to depend on the success of the mobilisation and can consist of short-term, as well as long-term, changes. Although the model appears quite linear, there are many moments where the development of the social movement can be stopped, re-oriented or cancelled either by the community of interest itself or by external factors.

This model enables us to better situate our study in regard to the different stages of a social movement. In the present paper, we only focus on the elements referring to the mobilisation on social media and the type of actors involved in the online conversations. However, in addition to Gordon's model, we consider that the mobilisation and call for action surrounding a social movement (in our study, the women's strike) are embedded in a broader context. In this context, public opinion and political polarisation towards gender equality questions prevail and are generally measured with opinion surveys. While social mobilisation around the women's strike happened on the streets, in newspapers, and in official communication, it also took place on social media, which is the focus of the present study.

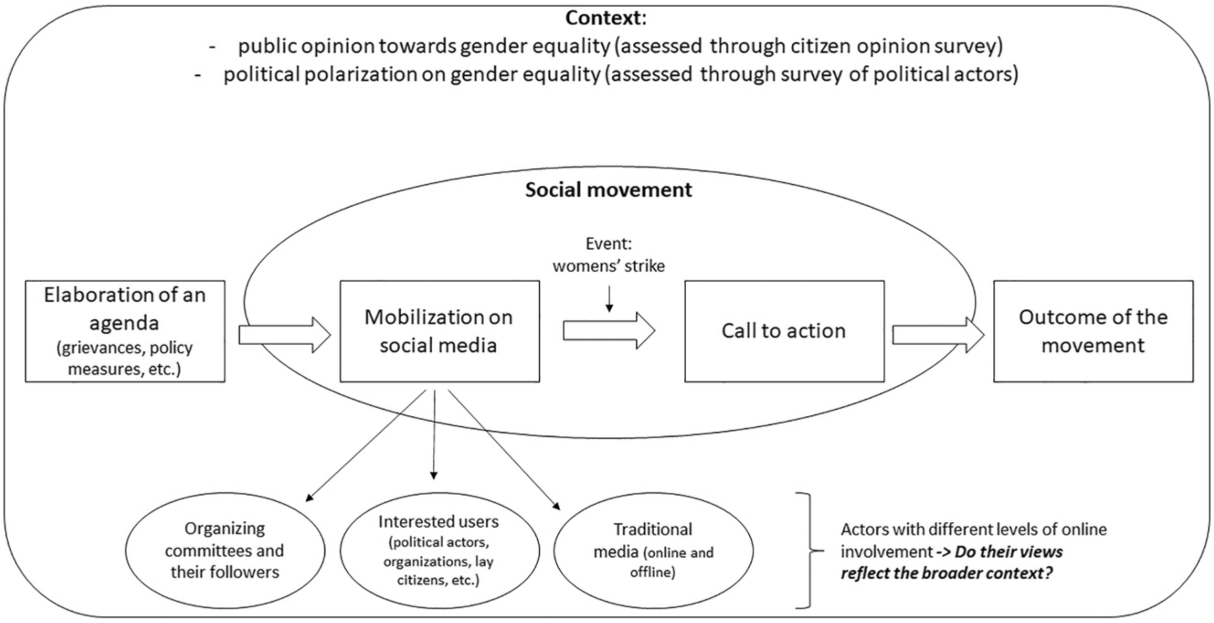


Figure 4.3.1: Conceptual framework underpinning the study (inspired from Gordon 2015).

To understand the interrelation between offline opinion and online discourse, it is important to identify the actors who mobilise on social media, both supporting and opposing the strike. Indeed, if social media are now a widely used communication tool to mobilise supporters around common grievances and to implement calls to action, opponents of social movements also use social media for developing counter-protests and for spreading their counter-arguments or programs (Gordon, 2015, p.19). Hence, in a first step, we aimed to identify the user groups who mobilised on Twitter surrounding the women's strike.

At the same time, studies have shown that the reliance on social media by these different actors to express their views can lead to increased polarisation (Quattrociocchi, Scala & Sunstein, 2016). Therefore, we aimed to assess the extent to which the views expressed by the politically engaged users on social media reflected trends prevailing in the broader context. We further looked at the extent to which social media increased polarisation trends observed in public opinion surveys and in traditional political surveys. Finally, to better grasp if social media can be used to advance social movements to promote its messages and to achieve its goals, we took advantage of the textual nature of social media data to explore the views and arguments of each actor surrounding the women's strike. This online content can also be compared to open-ended survey responses from citizens concerned about gender equality related issues. In the next sections, we will develop our research questions and the respective hypotheses in further detail.

Assessing the Congruence Between Social Media Discussion and Public Opinion Regarding Gender Equality

Not only has the topic of gender equality become central to scientific literature, but it has also been prevalent in political agendas and parliamentary debates across the world (Hooks 2000). As activists, politicians and other groups seek to find an agreement on gender equality policy measures, gaining knowledge about the overlap between online and offline opinions around gender equality is key to better grasping the opportunities for social change.

The study of Scarborough (2018) takes a step in this direction by demonstrating to what extent social media data, specifically tweets, can be used to account for gender equality attitudes. The main premise of the study states that if tweets about feminism deal with issues that are central to gender relations, then they should capture the same underlying dimensions as those opinions measured through gender attitude surveys. In a similar vein, other studies have pointed to difficulties in the identification of opinions expressed online, notably the supportive and opposing views about gender equality. For instance, Kirkwood et al. (2018) were unable to differentiate between very polarized pro-, neutral, and anti-feminist views in discussions on Twitter, especially because of content-specific challenges pointing to the difficulty of extracting relevant tweets and of precisely classifying the diversity of sub-topics. Both studies, however, place little emphasis on the actor groups involved in gender equality discussions. Furthermore, by considering sentiment (or tonality) as the main content feature to be correlated with surveyed opinions, they say little about how social media discursive content reflects what the broader public thinks. In the framework of public opinion surveys, Baldassarri and Park (2020) found that the U.S. population is moving towards more progressive views on a host of issues – from LGBTIQ+ rights to gender roles and sexual behaviours (see also Eisner, Spini, and Sommet 2020 in the Swiss case). However, the authors warn that, contrary to public opinion dynamics on economic and civil rights, the above-mentioned issues can less clearly be described in terms of increased issue partisanship. Wallaschek et al. (2022) found a similar trend towards the support of gender equality on social media. The authors investigated the users' engagement and the content of debates about gender equality in tweets about the 2021 International Women's Day in Germany, Italy, and Poland. They showed that social media users and discussions were predominantly supportive of gender equality, as users engaged with the value of gender equality mainly in an acclamatory fashion. They also showed that political and societal actors exhibit high levels of online engagement. These studies demonstrate that actors and content dominant in social media discussions are also mainly in favour of gender equality measures.

In line with the findings outlined above, and to answer our first research question (i.e., to what extent does the involvement of politically active users reflect trends from public opinion?), we expect to find a high congruence between the online involvement and the offline support for gender equality in terms of political ideology (Hypothesis 1).

Assessing the Extent to Which Social Media Leads to More Polarisation on Gender Equality than Trends Prevailing Offline

Notwithstanding the closing of the gap on gender equality measures between progressive and conservative positioning, there remain important divergences between political elites on gender policy. This can lead to important polarisation between politically involved actors and the wider public, which could be further heightened on social media. Therefore, our second interest is to investigate the extent to which political polarisation surrounding gender equality is increased on social media.

With respect to gender equality, polarisation can be conceived in terms of positional dynamics relying on rhetorical moral arguments (De Wilde & Zürn, 2014; Roggeband, 2018). For instance, Kantola and Lombardo (2020) conducted a qualitative analysis of populist interventions in EP plenary debates on gender equality in the European Parliament and found a variety of radical right opposition strategies to gender equality, mainly drawing on old and traditional gender imaginaries packaged in novel populist ways. Their findings reflect previous studies on political elites' opinion polarisation displaying similarly extreme opinions about connected topics, such as sexual minority rights (Wojcieszak, 2010) and homosexuality (Munro and Ditto 1997). Furthermore, social media play an important role in the polarisation of the political debate on gender equality. For instance, Russell et al. (2020) showed that hyper-partisanship in Parliament extends from the legislative process into politicians' social media strategic communications. Social media are thus useful as they can cover a large spectrum of political positions that underlie the topic of gender equality, thus also providing a platform for the backlash against the ideas and goals of feminism (Lawrence & Ringrose, 2018).

However, notwithstanding the increased opportunities to express political opinions online, social media can also amplify the phenomenon called the “spiral of silence” (van Aelst et al., 2017), which consists of self-censorship behaviours on the part of politicians or citizens who do not express their opposing views about a topic (Noelle-Neumann, 1984). This silence can be due to the fact that individuals perceive a majoritarian public consensus (Sunstein, 2017) or simply because the topic does not trigger enough of their attention (Lasorsa, 1991). On social media, this self-censorship behaviour can lead to an increased polarisation, notably as certain views will become inflated at the expense of other opinions (Dubois & Szwarc, 2018). Additionally, the effect of filter-bubbles (or echo-chambers) is likely to lead to a fragmentation of (more extreme) opinions towards political issues (Zuiderveen Borgesius et al., 2016).

These studies thus point to the fact that social media serve as a tool for conservative politicians to voice their opposition to gender equality measures. These studies also suggest that there are dominant ideologies surrounding the gender equality debate, thus leading to political polarisation on gender related issues. In line with these findings, and to answer our second research question (i.e., to what extent is political polarisation surrounding gender equality increased on social media?), we expect to observe increased levels of polarisation of the online debate on gender equality compared to trends observed through the lens of opinion surveys (Hypothesis 2).

Identification of Gender Equality Opinions in Social Media Content

In addition to the focus of political polarisation on gender equality issues, we also aim to investigate the extent to which social media claims have the potential to echo what the broader public thinks. This is paramount to understanding the extent to which social media mobilisation can help social movements build long standing support in public opinion.

In line with the idea that social movements have the ability to connect with public opinion, Mirbabaie et al. (2021) investigated how specific user groups participated over the course of the MeToo debate in 2017 and 2019. Drawing from the theory of connective action, they found that the framing of and the attention to the movement were spread in different ways according to actor groups – namely, the starters and the maintainers. Overall, the authors found little variety in the content of online discussions, although they pointed to different underlying motives, ranging from self-serving and branding intentions to calls for attention and action.

Their findings echo the results from Baik, Nyein, and Modrek (2021) by showing how difficult it is for social movements to garner new adherents. The authors concluded that, although the movement they analysed did increase awareness and participation among those already sympathetic to the movement, it might not have enlisted new supporters. This is mainly because social media are used by user groups to promote their ideas through different communication strategies. For instance, direct supporters have an interest in promoting the event, whereas politicians point to fewer collective goals and highlight specific aspects of the debate which resonate with their own agenda.

The studies outlined above remain, however, in the framework of social media research without mobilising other data sources, such as opinion surveys, to assess the potential of social media messages to impact public opinion on gender equality. The study of Adams-Cohen (2020) proposes to address the question of causality in the domain of same-sex marriage. More precisely, it uses Twitter data and machine-learning methods to analyse the causal impact of the Supreme Court's legalization of same-sex marriage at the federal level in the United States on political sentiment and discourse towards gay rights. Results showed that there was a relatively stronger negative reaction in public opinion towards same-sex marriage in states where the Court's ruling produced a policy change as compared to that of other states. Nonetheless, this study is also not able to rely on survey data to benchmark its findings.

In our study, we propose to take advantage of the textual nature of social media texts and of open-ended survey questions to look at the extent to which ideas and views expressed online by users involved in discussions surrounding the strike are similar to open-ended answers of citizens concerned with gender equality. Based on the literature, and to answer our third research question (i.e., what does the content of social media data tells us about the potential of social media to promote the claims of the social movement?), we expect to observe a continuum between calling for attention, on the side of the strike organisation committee and of its followers, and discussing concrete policy measures, on the side of politicians and citizens (Hypothesis 3).

Data and Method

Data Collection: Original Tweets from Different Actor Groups

In the present study, we focus on the 2019 women's strike in Switzerland. The strike consisted of demonstrations in the country's major municipalities and revolved around the issues of equal pay, recognition of unpaid care work, and governmental representation. It followed the first 1991 Swiss women's strike, which was organised 10 years after the Swiss population's acceptance of the constitutional article on the equality between women and men.

To construct our corpus of tweets about the women's strike covering the period from January to December 2019, we retrieved social media texts from the social media platform Twitter using the Application Programming Interface. Our data collection strategy was based on both seed user profiles and search queries.

Concerning the user profiles, we extracted all tweets emitted from the seed user profiles. This included the main organisation committees of the women's strike that had a Twitter account. We also collected all tweets from the committees' followers (as it was in June 2019), as well as tweets from political accounts (candidates of the October 2019 federal elections, elected politicians, and national political parties) referring to the women's strike between January 2019 to June 2019. For the followers, political accounts, and trade unions, we then filtered out tweets that did not explicitly refer to the women's strike by using a list of search queries. The list of search queries read as follows: '.*womenstrike.* | .*frauen.*streik.* | .*feministi.*streik.* | .*femstreik.* | .*frauen.*strassen.* | .*frauen.*mobilisier.* | .*grève.*femmes.* | .*greve.*femmes.* | .*grève.*féministe.* | .*femmes.*grève.* | .*femmes.*greve.* | .*femmes.*rues.* | .*femmes.*mobilis.* | .*femmes.*manif.* | .*14.*juin.* | .*14.*juni.*'.

We also included tweets that contained specific search queries stemming from the most common hashtags found in the collected data. This strategy enabled us to make sure that the queries were precise enough to collect tweets related to the specific event of interest, but not too large as to include unrelated tweets. The search queries read as follows: 'frauenstreik | 14juni | femstreik | feministischerstreik | fstreik | femstreik | frauendemo | frauenbewegung | grevedesfemmes | 14juin | grevefeministe | femmesengreve | grevefeministe'. The hashtag #womenstrike was not used as it returned tweets mostly unrelated to the event of interest. We kept only tweets not emitted from the above-

mentioned groups (organisation committees, followers, politicians, candidates, and trade unions).
We did not include retweets in our sample because we are interested in the content of original tweets and to what extent different actor groups contribute to the elaboration of the content of social media discussions. Our final corpus contained 41′062 tweets and 15′919 unique Twitter accounts.

Manual Annotation of Tweets

To address our first research hypothesis, we conducted a manual annotation of the entire list of Twitter profiles that contained a description. We coded for the type of accounts that were not already included in the list of seed accounts. We thus added the following additional categories: foreign politicians, media, and journalists (national and foreign), activists (national and foreign), other political users (e.g., national and foreign ambassies, governmental departments), organisations with feminist-related aims (national and foreign), other users tweeting actively (more than five tweets) about the strike, and other unlabelled users. These categories were heuristically found to be encompassing enough to describe our sample of top users. The category for followers only indicates users which are not labelled in any other user categories. Finally, the category for trade unions does not differentiate between national and foreign organisations.
We also labelled the profiles to specify whether they are Swiss, foreign, or unknown accounts. The Annex 4.3.2 provides more information about the distribution of accounts according to geolocation. This annotation was also done for every profile and shows that most users included in our sample stem from Switzerland (27%), France (16%), Germany (14%) and other European countries (11%).

Correlation Between Social Media and Survey Data According to Political Affiliation

To address our second research hypothesis, we aimed to identify the party affiliation of Swiss politicians on Twitter. To do so, we manually specified the party affiliation of the Swiss political accounts. This enabled us to correlate the online attention with the offline support to gender equality along partisan leaning. In order to achieve this, we relied on survey data to measure politicians' and citizens' opinions about women's rights. The obtained survey scores are reflected along a left-right political continuum (e.g., the party affiliation of politicians and of citizens who declare a party affiliation). These scores are also correlated to the prevalence of Twitter discussions about the women's strike for each political affiliation.

Concerning politicians, three items from the Swiss part of the 2019 Comparative Candidate Survey ask politicians about women's rights. We used the following items measured on a 5-points Likert scale: 'Women should be given preferential treatment when applying for jobs and promotions,' 'The government should take measures to reduce differences in income levels,' and 'Women should be free to decide on matters of abortion'. From these items, we built a mean score for gender equality support by political affiliation. The mean of the score is 3.4 with a standard deviation of 0.8. Concerning citizens, two similar items were selected from the wave 22 of the Swiss Household Panel survey. The items were measured on a scale from1 to 10 and read as follow: 'Gender: Women in general penalized' and 'Gender: In favour of measures.' We also built a mean score for gender equality support by political affiliation. The mean of the score is 5.8 with a standard deviation of 1.1.

Pre-processing Steps for Data Cleaning

Switzerland is a multilingual country with German and French being the most represented languages (Italian and Romansh are the other two national languages). To preserve the most authentic content of discussion, we did not translate the tweets into a single language. We nevertheless conducted several pre-processing steps. For instance, we filtered out URLs and characters that are not natural language texts. We also filtered out stop words, which are words that provide no information towards the analysis. We further split concatenated words (e.g., WomenStrike becomes women strike) and we lowercased the text. All further typos, misspellings, and slang terms remained intact. We then lemmatised the text using the library udpipe (Wijffels, Straka & Straková, 2018) for the programming language R.

Unsupervised Text Representation

Correspondence and cluster analysis of our corpus of tweets were used to investigate the theme and opinions surrounding the strike to address our third research hypothesis. To explore the content of the tweets and how opinions relate to the different actor categories, we analysed the co-occurrence of words in tweets, extracting shared semantic regions via correspondence analysis. To do so, we used the library FactoMineR (Husson et al., 2013) from the R language. Correspondence analysis can be understood as principal component analysis for categorical data. It is used to discover structure in textual data (D'Enza & Greenacre, 2012; Morselli, Passini & McGarty, 2021). Correspondence analysis works as an unsupervised bag-of-words approach where the words are projected on a factorial space such that the proximity between words indicates a higher association (or a shared semantic meaning). Correspondence analysis calculates the contributions of each word to the inertia of a factorial axis, showing how each word contributes to identifying the axis. Hence, words that are projected further from the centre of the axis provide a higher contribution.

For our analyses, the data-term matrix was aggregated by the user categories. Using a graphical representation on a two-dimensional space, we projected the Twitter vocabulary on the correspondence analysis space to visualise ideas and opinions associated with each user category. As such, user categories that appear closer on the graph share a similar vocabulary and set of ideas. To account for the polarisation between left and right-oriented accounts, we created one category for each political orientation. Furthermore, to look at the extent to which what was said on Twitter is representative of what the lay audience thinks, we added open-ended responses from the Selects survey respondents to the items asking about the first and second 'most important issue facing Switzerland'. The pre-processed (same steps as the Twitter data) respondents' vocabulary was used as supplementary rows and was not used for the definition of the principal dimensions. Their coordinates were predicted using only the information provided by the performed correspondence analysis on the active vocabulary from the Twitter accounts.

Results

Description of the Corpus

Table 4.3.1 displays the involvement of the different actor groups in our corpus of tweets. It provides the number of accounts, the number of tweets, and the tweeting frequency of the different user groups involved in online discussions about the women's strike. In total, almost 16′000 unique accounts took part in the online discussions about the strike with an overall tweeting frequency of three tweets. Table 4.3.1 shows that the organisation committees were the most active users when considering the tweeting frequency. They are followed by their followers, Swiss political parties, and other involved users that tweeted more than five times. Table 4.3.1 also shows that Swiss political actors (parties, elected politicians, political candidates) emitted 11% of all collected tweets. When looking at the distribution of these political accounts according to political ideology (see Annex 4.3.1), we see a domination of the left in terms of number of accounts (55% of the Swiss political accounts) and tweets (75% of the Swiss political tweets). However, political accounts from the right were not absent from the online debate.

Table 4.3.1: Description of the Twitter user groups in our sample among seed users and additional users (in bold, first column) with total number of accounts and tweets (in bold, bottom line).

User groups	**Number of accounts**	**Number of tweets**	**Tweeting frequency**
Seed users			
Strike committees (Swiss)	8	1715	214
Unclassified followers (status in 2019)	165	3536	21
Elected politicians (Swiss)	194	1242	6
Political candidates (Swiss)	298	1896	6
Political parties (Swiss)	124	1203	10
Additional users (national and foreign)			
Trade unions	125	666	5
Organisations with feminist/gender aims	398	2801	7
Proclaimed activist	409	1347	3
Media/journalists	2067	7134	3
Other political users	290	763	3
Other users with tweeting frequency >= 5	371	3782	10
Other users with tweeting frequency < 5	11,468	14,909	1
Total	**15,917**	**40,994**	

In complement to Table 4.3.1, Figure 4.3.2 displays the number of original tweets about the women's strike emitted by actor groups over time. We see that organisation committees' followers, other top users, and the media or journalists were essential actors that generated original content about the strike. Political candidates formed the third most prolific group. We also notice two peaks in the collected data pointing to two major events, namely International Women's Day in March and the Women's Strike in June.

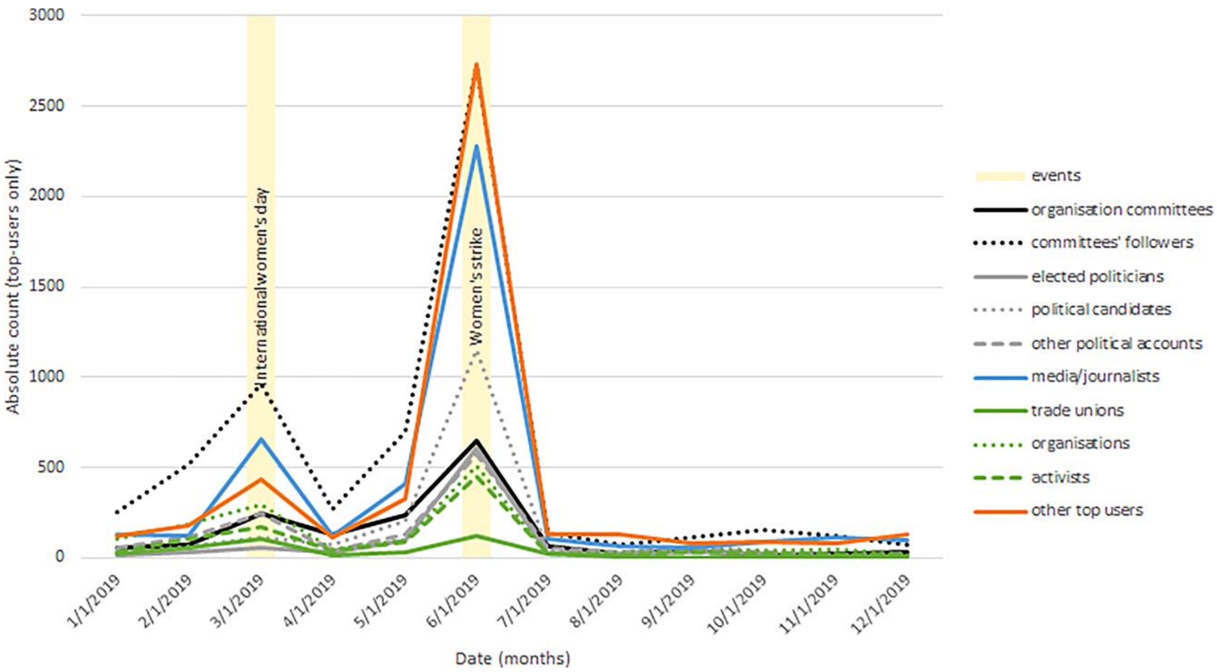


Figure 4.3.2: Prevalence of tweets by important actor groups over time.

Congruence Between the Online Involvement and the Offline Support for Gender Equality in Terms of Political Ideology

Figure 4.3.3 provides four panes displaying the relationship between the political leaning (x-axis) and the support for gender equality measures (y-axis). The panes are organised so that the offline and online patterns are compared horizontally, and so that the citizen and political patterns are compared vertically. The upper left pane describes the relation between citizens' left-right positioning (x-axis) and citizens' support for gender equality (y-axis). The lower left pane displays the same relation for politicians. Both of these left panes are solely based on survey data. The upper right pane includes the relative tweeting frequency of Swiss political actors active on Twitter (x-axis) in relation to citizen support for gender equality (y-axis). The lower right pane includes the relative tweeting frequency of Swiss political actors active on Twitter (x-axis) in relation to politicians' mean support for gender equality (y-axis). Both of these right panes combine the salience of Twitter discussions about gender equality on social media (x-axes) and the opinions towards gender equality measured in survey data (y-axes).

Figure 4.3.3 allows us to address our first research hypothesis about the distribution of political ideologies in relation to the support of gender equality measures. More concretely, we assess whether there is a congruence between the online involvement and the offline support for gender equality in terms of political ideology. The upper left pane shows that citizens with a left-leaning orientation have a more positive attitude towards gender equality measures compared to citizens with a right-leaning orientation. The lower left pane displays a similar pattern for politicians that responded to the survey. With respect to Twitter conversations, the upper right pane shows that politicians with a left-leaning position were more involved than politicians with a right-leaning orientation. Overall, social media discourses from politicians reflect the pattern survey data from their potential electorate. Here, citizens and politicians with a leftist orientation are clearly the most favourable towards gender equality measures and the most involved on social media.

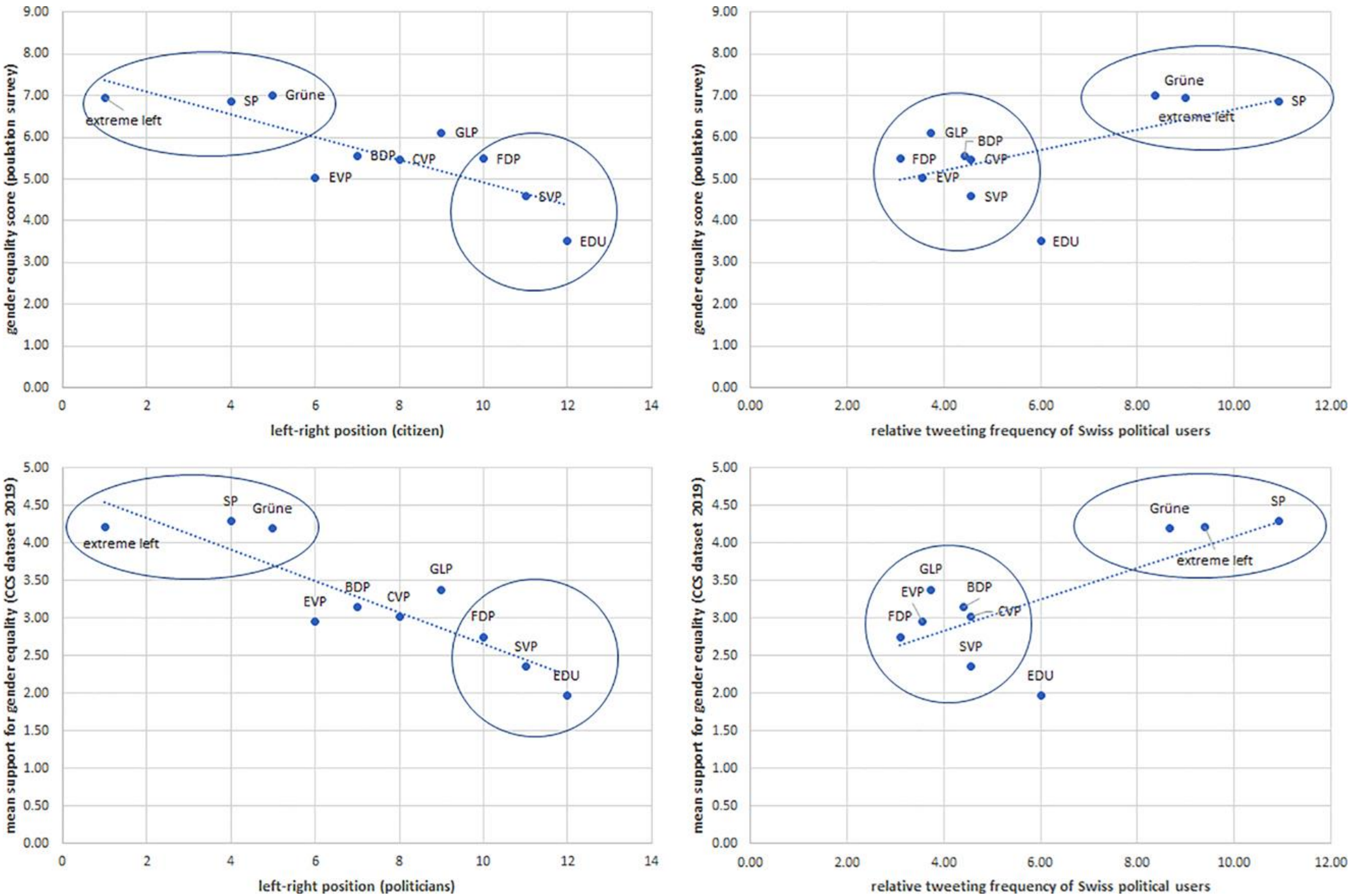


Figure 4.3.3: Relationship between online and offline gender equality opinions.

Figure 4.3.3 also enables us to test our second hypothesis according to which we should observe increased levels of polarisation of the online debate on gender equality compared to trends observed through the lens of opinion surveys. This polarisation can be observed on both the upper and lower right panes of Figure 4.3.3, where there is a clear segregation between politicians with a left-leaning compared to the remaining politicians with either a centrist or a right-leaning orientation. The EDU (Eidgenössisch-Demokratische Union in German or Federal Democratic Union in English) is an outlier in comparison to the other rightist parties as it demonstrates a high relative tweeting frequency. This can be explained by the fact that the EDU made several actions to voice its opposition towards the legitimacy and the usefulness of the strike (see the "a rose for you" leafleting campaign, where the EDU aimed at thanking the women who would not go on strike). Overall, we observe that social media increases the polarisation between left and right-leaning politically involved actors compared to opinion surveys.

We rely on the correspondence analysis displayed in Figure 4.3.4 to test our third research hypothesis according to which we expect to observe a continuum between calling for attention, on the side of the strike organisation committee and of its followers, and discussing concrete policy measures, on the side of politicians and citizens. Figure 4.3.4 thus enables us to observe if online discussions are along a continuum between calling issues to attention and discussing concrete policy measures. The obtained two-dimensional space reveals the structure from the vocabulary employed by the actor groups. The final data-term matrix is based on 22′341 German tweets and 550 open-ended survey answers. The matrix includes 2′ 723 lemmatized terms which are either nouns or adjectives. Figure 4.3.4 shows the projection of the terms on a factorial space with the active user categories (in red) and the passive user category from survey respondents (in brown). The terms were automatically translated in English using deepL and the translation is given after the "_" on Figure 4.3.4.

The first dimension explains 18.3% of the variance. On the negative side of the axis, it includes terms such as "justice", "assembly", "consent" (lower left quadrant), and "principle" (upper left quadrant) in relation to terms such as "patriarchy" and supportive actions or movements (e.g., "collective", "flyer"). On the positive side of the axis, it includes terms such as "council of states" and "understanding" (lower right quadrant), and "regulation" (upper right quadrant) in relation to policy issues about taxpayers, childcare, and wage discriminations. Therefore, this axis seems to refer to a "normative-representative" continuum going from the defence of women's rights through norms and actions to the implementation of policymaking in the political arena.

The second dimension explains 15.1% of the variance. On the negative side of the axis, the figure includes terms such as "chinderchübu" and "gängeviertel" referring to public and open projects (lower left quadrant), and "albanieen" or "saxon" referring to debates about foreign aspects on Twitter (lower right quadrant). On the positive side of the axis, the figure includes terms such as "spfrau" and "parental leave initiative" referring to Swiss political initiatives in the framework of gender equality (upper right quadrant), and "strass" and "industry group" or "demonstrators" referring to important supporters and stakeholders actively taking part in the Swiss women's strike (upper left quadrant). Therefore, this axis seems to refer to a "foreign-national" continuum going from the reference to foreign projects to the concrete Swiss mobilisation for the defence of women's rights.

The first axis differentiates between the strike organisation committees, their followers, and organisations with gender-related aims on the left side, and political actors on the right side. Accounts from trade unions, media, activists, and other users are grouped in the centre. The second axis differentiates between the different political leanings of political accounts. The survey respondents are located close to the Swiss political accounts, at an equal distance from left and right-oriented accounts.

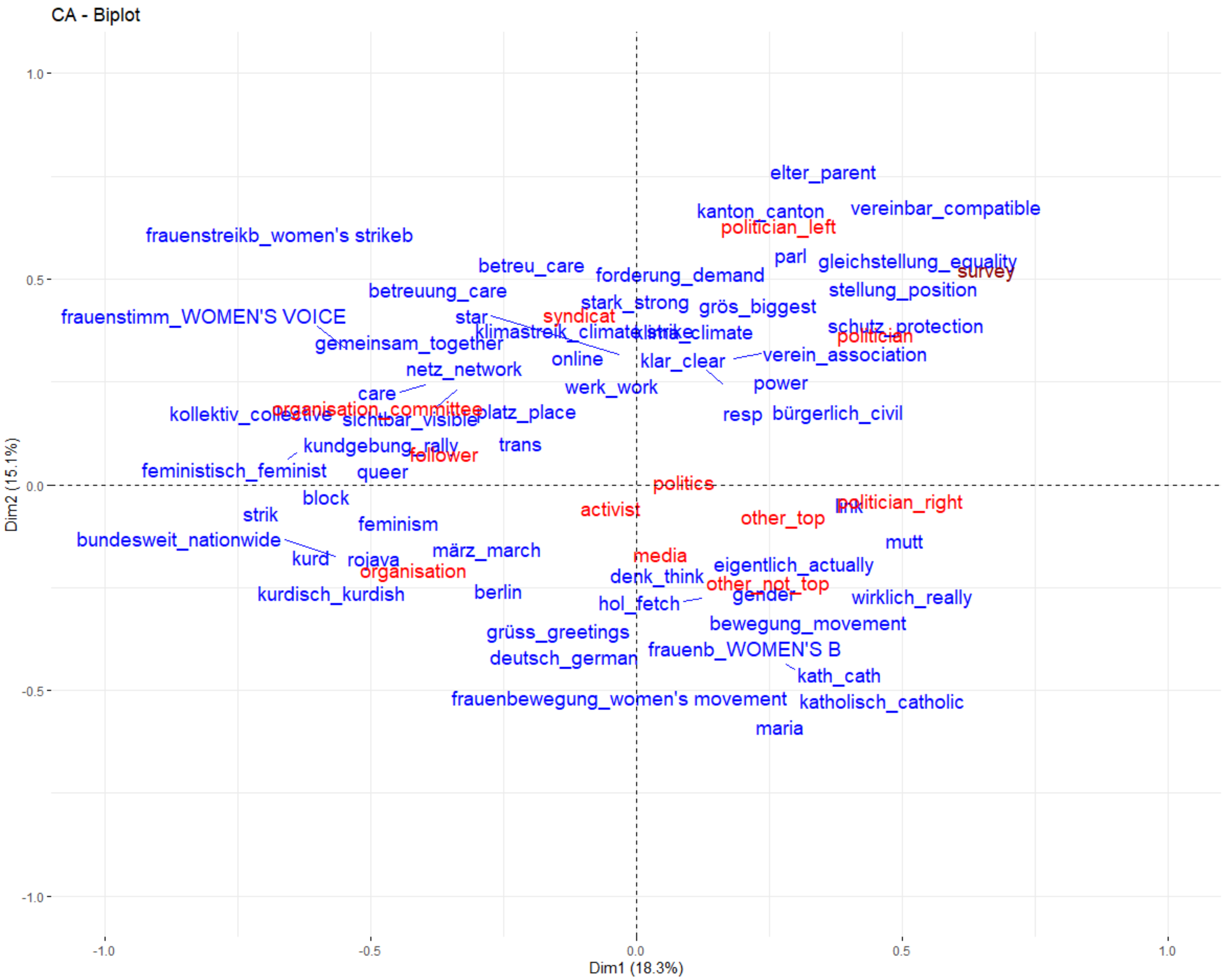


Figure 4.3.4: Graphical representation of the correspondence analysis.

Discussion of the Main Findings

In this article, we described which actors are particularly involved in social media discussions surrounding a social movement promoting gender equality. Then, more substantially, we also aimed to investigate how these online discussions echoed what the broader public thinks about gender equality. To do so, we investigated the relationship between online and offline gender equality support along the spectrum of political leaning. We also examined the correspondence between the discursive content of social media actor groups, upon which we also mapped claims from a representative sample of citizens concerned about gender issues. We examined our research interests by analysing the Twitter accounts involved in social media discussions about the women's strike that took place in Switzerland in June 2019. The data collection strategy should be representative of the range of gender related discourse on social media in Switzerland given that it comprises of a variety of user categories and that it was possible to code for the geolocation of the most active accounts in most cases.

In a first descriptive stage, the distribution of Twitter user profiles shows that organisational committees and their direct followers were the most active contributors to the online content. Swiss political accounts also actively participated in online discussions (11%). The high involvement of Swiss political accounts can be explained by the fact that the year 2019 was also a federal election year in Switzerland, thus providing politicians with an increased incentive to voice their positions online. This finding echoes the literature showing that most social media content about social movements' agendas tends to be produced and discussed by a minority of users (Huges & Wojcik, 2019).

We were able to confirm our first hypothesis, according to which we expected to find a high congruence between the online involvement and the offline support for gender equality in terms of political ideology. In particular, Twitter discussions were dominated by left-oriented political accounts (see also Annex 4.3.1), which also reflects the more positive attitudes towards gender equality from citizens and politicians with a left-leaning orientation as measured in surveys. However, our findings also show that politicians from the extreme-right (SVP and EDU) also engaged in intensive tweeting to voice their opposing views. Therefore, although right-leaning parties and politicians talked less (in terms of prevalence) about gender related issues on Twitter, they may have been talking more negatively about gender equality than left-leaning actors. Overall, we find atypical behaviours from political actors from the extreme left and the extreme right as they tend to address gender equality more frequently on social media than offline.

We were also able to confirm our second hypothesis, which suggested that political polarisation on gender issues is more pronounced on social media. Indeed, the results emphasise that social media have a clear polarising effect that segregates between the left and the rest of the political spectrum. The added value of our findings relates to the fact that social media tends to increase the polarisation between the left and the right of the political spectrum in comparison to that observed in survey trends. This finding echoes the literature on the contribution of social media to political polarisation (Tucker et al., 2018).

The results from the correspondence analysis confirm our third hypothesis. Thus, our expectation to observe a continuum between calling issues to attention and discussing concrete policy measures was substantiated. Indeed, in line with previous research (Mirbabaie et al., 2021), we find that organisation committees and organisations use social media to call for attention and action. Contrastingly, political accounts are engaged on social media to make visible policy measures addressing gender equality and to link to other possibly related policy issues, such as child or family policy (especially in the case of right-oriented accounts) or climate change policy (namely in the case of left-oriented accounts). We thus suggest that social media discussions surrounding the women's strike provided politicians with an opportunity to promote their own policy agenda.

Conclusion and Outlook

The main purpose of this study is to contribute to the mapping of gender equality discourses by investigating Twitter discussions surrounding the women's strike that took place in Switzerland in June 2019. This article presents a perspective on political polarisation by looking at the relationship between social media opinions and those expressed in surveys to assess gender equality related concerns. The use of data collection for the assessment of the dynamics of gender communication is the strength of the work. The juxtaposition of several data sources on attitudes about gender equality is particularly relevant from a practical perspective. The main reason for this is that social movements use social media to develop their actions and to build long-standing support around particular claims to achieve social change while being confronted with (pre)existing public attitudes (see also Eisner et al., 2021).

Relying on an extensive manual annotation of the identified Twitter accounts and using unsupervised content analyses, we showed that Twitter triggers a stronger pattern of political polarisation on the topic than that observed in survey data. For instance, political accounts from the extremes of the political spectrum gave more prominence to the topic on Twitter than they did offline. Furthermore, a reinforced polarisation of the left and right-wing positions along the political spectrum has an effect on online discourse. This suggests that the possible impacts of polarisation in society or on social media lead to heightened attention to the topic, especially as the year 2019 was also an election year in Switzerland and gave rise to a surge in women's political representation (Giger et al., 2021). During our observation period, other strikes (e.g., climate change mobilisations) and popular votes (e.g., preparation of the campaign for the popular vote on paternity leave in September 2020) may also have impacted the content of online discussions. For example, it may be that these themes were put forward as other possible issues on which the political realm was expected to provide policy solutions.

This study set out to complement research on one of the most relevant topics in European countries; namely, public discourse about gender equality. Applying a comparative approach between social media content and survey data enabled us to compare the discourse on gender equality from a variety of political actors and to shed light on the relationship between left-right ideology and gender equality discourse. This study also improves our understanding of gender-related communication strategies of political actors on social media.

Our study has its limitations, which future research may want to address. Our sample provides a good sample of users engaged in social media discussions about gender equality, but it does not offer a comparison between countries. Moreover, by limiting our analysis to non-retweets, we might have lost some information about the most discussed topics related to gender equality. Other limitations relate to our choice to not translate the tweets into a common language and to use only German tweets for correspondence analysis. Indeed, other topics and trends could prevail in the remaining regions of Switzerland. In the future, it may be possible to use more resource-intensive approaches. Finally, although Facebook has shut down parts of its API, we would encourage scholars to pursue research on Facebook posts, given that this social media has a higher reach than

Twitter in most countries. However, we were interested in retrieving political content and, in this view, Twitter is better suited (van Dijck & Poell, 2013).

Based on our findings, we suggest future studies try measuring the influence of social media discourse on public opinion in the field of gender equality. Such a research design would require that the online discourse can be integrated with longitudinal survey data. However, this type of survey data is rarely available to researchers. Indeed, while survey data are by far the most popular source of data in studying public opinion, collecting data around social movements is time consuming as one would need to run comparable polls immediately before and after the event (see similar discussion in Brickman and Peterson (2006)). As a final outlook, we would like to suggest that researchers need to make a greater effort to understand the gender equality topics to which citizens are exposed by investigating non-textual content (e.g., pictures or videos) and by expanding the already existing qualitative analyses to a quantitative perspective by developing appropriate technical tools.

To date, surveys have been the way to assess this congruence between the public and politicians' positioning on similar issues (Reveilhac, Steinmetz & Morselli, 2022).

However, social media can serve as a complementary picture by providing online dynamics. Beyond the social movement literature, the idea of interrelated offline and online agendas represents a major topic in the field of political communication (Gilardi et al., 2021; Posegga & Jungherr, 2019). For instance, politicians' involvement online may not only depend on their ideology, but also depend on how they anticipate their audience to share their same political ideology. Social media enables us to investigate how conflicts that take place offline are reflected in the digital and social media spheres, thus illustrating the new mediatized logic of value contestation.

CHAPTER 5. HOW CAN SOCIAL MEDIA BE USED TO PROVIDE A NEW LENS TO WELL-ESTABLISHED AND UNDER-INVESTIGATED TOPICS IN SOCIAL AND POLITICAL SCIENCES BY COMPLEMENTING SURVEY DATA?

5.1 The framing of health technologies on social media by major actors: Prominent health issues and COVID-related public concerns[30]

Introduction: studying health care issues using social media

Platforms such as Facebook and Twitter have recently attracted the attention of enterprises and public institutions working in the field of health technology (hereafter HT) as potential communication channels for promoting their policies and products (Lupton, 2012). In our study, we aim to gain a better understanding of the major actors leading the online HT debate and, thereby, the prevalent topics and discursive frames they emphasise on social media. Providing answers to these questions is paramount as the internet in general – and social media in particular – is increasingly important for citizens who want to inform themselves about health-related issues (OECD, 2020).

We adopt a sociological approach which focuses on the role of prominent actors in the depictions of HT in publicly accessible discourses. Here, the reliance on social media by important actors in the field of HT is likely to provide a fertile source of information about the current public debate concerning HT. Indeed, social media have become important platforms through which HT companies and professionals can position themselves and get in touch with a public audience (Lupton, 2012). Meanwhile, social media are also becoming an important source of information from which citizens get health-related information (see (European Commission, Brussels 2015); Weber Shandwick, 2018).

Our study thus contributes to shedding a complementary light onto studies focusing on health-related practices on social media (Lupton, 2012) and on the potential of social media applications for health behaviour and information (Koteyko et al., 2015). It raises two main research questions: Who are the important actors in the HT field that are active

[30] This chapter is a slightly adapted version of the article that has been published as M. Reveilhac and A. Blanchard (2022): "The framing of health technologies on social media by major actors: Prominent health issues and COVID-related public concerns", International Journal of Information Management Data Insights, 2(1), 100068.

on Twitter? What important topics and framings of HT are promoted by these actors online?

There are three essential motivations for undertaking this study. First, there is still little empirical knowledge about who is involved in social media conversations concerning HT and how these conversations relate to business and to raising public awareness of these technologies Kushwaha et al. (2021). study emerging management areas that are supported by big data (including social media). They show that one of the most significant areas of development relates to healthcare management and two aspects in particular are studied: research about the usage of sensor-generated data to help in addressing diseases and the use of health data to manage patients. To complement this managerial approach, we argue that it is also important to have greater awareness of which common understandings and practices are promoted on social media to raise the public medical awareness and, thereby, to generate trust in the business overall.

Second, Grover et al. (2018) show that discussions about HT on social media tend to be skewed towards computing algorithms, while they show no differences in discussions of acute and chronic diseases, nor in discussions of communicable and non-communicable diseases. In addition to this social technical perspective, we argue that there is a need to better segregate HT related tweets with respect to the professions (or profiles) of the authors. Indeed, literature indicates that health care promotions fulfil different aims according to their target audiences. For instance, healthcare firms aim to improve patient trust and satisfaction (Jiang, 2019), whereas professionals aim to provide guidance to physicians (Peluchette, Karl & Coustasse, 2016). In addition, it is conceivable that the aim of companies is to put forward their latest technology for better patient outcomes, while influencers aim to give support to particular groups and products.

Third, another major motivation underlying the proposed study is to further investigate which aspects and concerns of the public debate could lead to the development of public opinion survey items. Indeed, relying on online data from major actors has the potential to complement existing analyses. To date, most large-scale quantitative research on HT has been conducted in the form of surveys conducted by government or national institutes (WHO, 2015) or eHealth professionals from multiple European countries (HIMSS Analytics, 2019). In addition to HT experts' answers to well-defined survey concepts, it is worth considering what major actors in the field consider important to share with the wider public on social media platforms.

To conduct our empirical analyses, we rely on messages from more than 4,000 identified actors active in HT discussions on Twitter. In a first descriptive step, we identify important actor groups active on social media to promote HT and investigate whether the geographical distribution of these actors correlates with the general public reliance on social media to seek health information. Our data suggest that institutions (e.g., governmental agencies or private enterprises) and specialists (e.g., physicians or experts) are the two major groups involved in the online HT discussions. Our data further show a correlation between the retweet share of major actors' messages and the general public's reliance on social media to seek health related information.

The key findings of our research demonstrate a positive correlation between the share of retweets of major actors involved in HT and the public share relying on social media to seek health information. It also identifies prevalent topics about HT found in tweets addressing technological priorities, professional skills, and privacy issues. Word embedding enables us to demonstrate that current challenges lie in the relationship between patients and professionals, notably patients' empowerment and access to health data. It further suggests that the COVID pandemic led to a shift away from concerns related to (cyber)security towards a focus on data storage and computing.

Another contribution of our study is to promote a computational approach to disclose topics and frames in the field of HT. Therefore, in a second research step, we rely on 'state-of-the-art' computational social science methods and creative visualisations. These methods are already used widely in the fields of linguistics and digital humanities. However, they remain underused in the field of sociology. Our article thus contributes to the promotion of these methods within the field and also provides a detailed explanation of how they can be implemented in practice to address other research questions. In our study, we investigate which salient topics are discussed online and how their prevalence differs in terms of geographical coverage and actor type. Additionally, we provide a more fine-grained view of the framing of specific aspects of HT in relation to important dimensions and relationships. For instance, we look at the framing of HT in terms of challenges and opportunities, technological advances, as well as privacy concerns. We differentiate these framings by actor group and by period (e.g., pre- and post-COVID pandemic).

Study background: the study of HT perceptions through quantitative and qualitative methods

Public opinion about health technologies

HT can be defined as healthcare innovations relying on continuous data collection and algorithmic evaluation. In recent years, social media apps and other mobile devices have increasingly been adopted by health professionals to 'personalise' health treatment by sending people tailored messages in relation to their individual health concerns and conditions (Fagerlund et al., 2019). Qualitative studies have thus explored the experiences of organisations in the development of disruptive health services. For instance, the study of Sterling and LeRouge (2019) investigates the integration of telemedicine services. While qualitative studies have the advantage of advancing our understanding of how to encourage people to voluntarily share health information with the authorities (e.g. governmental agencies, organisations, practitioners, or experts) and of the new deployment of business models and strategies, they generally lack generalisability in terms of concerns of the general population towards health technologies.

In the meantime, privacy issues have been raised in discussions about the use of personalised computerised technology (e.g., Lyon, 2010), thus underlying the variety of concerns and expectations of different actors and stakeholders. Against this background, one strand of research focuses on the psychological mechanisms underlying the intention to use personal health devices. For instance, Tsai et al. (2019) aim to explain why people accept or reject telehealth usage. Their study suggests that technology anxiety takes on a critical role. More recently, the experimental study from Ross (2021) about COVID-19 contact-tracing apps showed that the intention in using health apps was positively related to chronic prevention focus and that this relationship was mediated by privacy and information security concerns.

The role of social media for assessing health information

To investigate the publicly accessible discourse about HT, studies have relied on social media data to study citizens' interest in, and their responses to HT. For instance, a study by Grover et al. (2018) investigated Twitter discussions on 'technology-enabled health' to identify top technologies and their relationship with specific diseases. The authors could confirm the role of technologies for treating, identifying, and healing various diseases, while being skewed towards computing algorithms. Another study by Lee et al. (2019) analysed health technology trends and sentiments related to health information technologies in tweets so as to examine the opinions of members of the public and identify their needs. Relying on an ontology and sentiment dictionary, they showed that social media constitute a useful tool for studying the public's responses to new HT. Their study makes a strong contribution to assess public concerns towards HT, notably because of the lack of survey data of the topic.

Social media platforms do not only play an important role in citizen information and expression of opinion, but they are also a means used by institutional actors and specialists to maintain public relations, promote products, and construct social events around specific interests (Lupton, 2012). In view of investigating professionals' perceptions and uses of HT, qualitative studies have focused on the perception and use of these technologies by practitioners and physicians (e.g., Brandt et al., 2018; Johansen, Holm, & Zanaboni, 2019). These studies relied on semi-structured interviews with convenient samples of general practitioners to uncover perceptions, as well as on digital health records and electronic health consultations. Other quantitative studies relied on survey data from health professionals (e.g., IT staff, administrative staff, clinicians, CIOs, CEOs, physicians, nurses, professionals from consulting companies and from eHealth related sectors). For instance, the Annual European eHealth Survey is conducted two to four times a year to provide insights into specialists' current and expected developments within eHealth in Europe (HIMSS Analytics, 2019).

According to the literature review of Zhang et al. (2020), social media act as a research context for public health research when it is 'mere reference', used to recruit participants and for data collection. The authors also note that, while qualitative and quantitative methods are frequently used, 'state-of-the-art' computational methods play a marginal role. Furthermore, their review shows that discourse (as well as behavioural) data on social media (e.g., Twitter) have essentially been used by professionals and organisations for public health management, such as disease surveillance, assessment, and control. Concerning HT (eHealth specifically), the authors underline that social media have substantially altered how individuals seek and share health information, discuss health issues, and engage in health behaviours. This constitutes a primary motivation for further investigating the discursive content of online messages posted by actors actively taking part in the promotion and discussion of HT. For instance, social media can be used for promoting open innovation in digital health through hashtag-based campaigning. Kletecka-Pulker et al. (2021) investigated the impacts of the biomedical hashtag #DHPSP to promote visibility of patient safety and personalised medicine. The authors found that the campaign achieved high visibility with a large body of Twitter users participating in the online debate. Moreover, the campaign resulted in an increase of member enrolments and website visitors.

The current state of the literature shows that, despite social media's important role in spreading information and opinions about health applications and technologies, the role of social media as tools to spread HT awareness by actors actively involved in online discussions about HT has been little researched. At the same time, data availability and accessibility to various platforms are changing the nature of information systems studies. Particularly, Kar and Dwivedi (2020) underscore the need to explain beyond what is observed by moving towards why the observations happen. In our study, we seek to examine who the major actors producing HT content online are and what topics and framing of HT are prevalent in their online messages. Furthermore, we study changes in content before and after the COVID-19 pandemic which enables us to overcome the limitation that cross-sectional data can only be used to observe the relationship at a certain time.

Text classification methods for retrieving textual information

The large amount of data obtained from social media platforms makes it challenging to summarize the information in an interpretable way. This issue is especially salient in explorative research when content categories or semantic groups are not defined a priori by researchers. To address this difficulty, there is a need to apply unsupervised natural language processing techniques. Our study relies on two of them, namely topic modelling (hereafter TM) and word embeddings.

TM is widely used for producing data insights (Garg et al., 2021a). In fact, topic modelling consists of grouping together a collection of words in a way where each group represents a topic in a document. TM is beneficial for analysing the content of a corpus of documents with a knowledge discovery perspective (Bundschus, Tresp & Kriegel, 2009). However, one big issue with TM is determining the adequate number of topics to consider or opt for. Recently, several studies adopted TM analyses on tweets to identify public concerns. This trend has increased significantly with the COVID-19 pandemic with the need to rapidly identify important themes of discussion and public concerns. For instance, Abd-Alrazaq et al. (2020) examined the tweets posted in English related to COVID-19 from February to March 2020 by adopting Latent Dirichlet Allocation. Furthermore, Cinelli et al. (2020) collected data related to COVID-19 on Twitter, Instagram, YouTube, Reddit, and Gab to examine public engagement on the topic of COVID-19. They extracted all of the topics related to COVID-19 by generating word embedding and then analysed the topics. Moreover, Mahdikhani (2021) introduced a novel approach to extracting the features from tweets and to predicting their retweetability using supervised machine learning algorithms. In our study, we pay particular attention to how the extracted topics are distributed among countries and actor groups on social media to enhance the validity of our findings.

Word embeddings enable us to achieve dimensionality reduction using an unsupervised learning algorithm for obtaining vector representations for words. However, it is also used to achieve accurate text classifications (Singh et al., 2022). Recently, the popularity of word embedding techniques – such as Word2Vec (Mikolov et al., 2013) – have been increasing in various applications because of its capturing of word semantics and syntactics. For instance, cosine similarity measure is used to compare the found lists of resources and expand the queries (Garg et al., 2021b). Since Word2Vec treats each word equally in a corpus (or a document), it cannot distinguish the importance of each word. Therefore, it is useful to combine it with a weighting scheme to improve a given information retrieval task. In our article, we combine it with relative frequency. Furthermore, the word embedding approach evaluates the similarity score between words, but it does not answer why as to a similarity occurs. In our article, we propose to several visualisations that enable us to support similarity justifications between words.

Methods and data

Identification of major actors involved in the public HT debate on Twitter

The R library rtweet was used for data crawling and for natural language processing. Using the rtweet library, we extracted users whose profile description on their Twitter accounts contained specific keywords. The list of keywords was built upon the selection of relevant hashtags and words using tf-idf as a method of keyword extraction from Twitter conversations. In information retrieval, tf-idf means term frequency-inverse document frequency and serves as a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. The resulting list contains the following queries: 'healthtech', 'health AND (technology OR technologies)', 'digitalhealth', 'digital AND health', 'mhealth', 'medtech', and 'ehealth'. The term 'ehealth' refers to healthcare practices supported by electronic processes and communication and includes the networking of oIT staff, administrative staff and clinicians from health facilities, professionals from health-IT related software and consulting companies. Dating back to at least 1999, the usage of the term covers not just Internet medicine but also virtually everything related to computers and medicine. The term 'mhealth' is an abbreviation for mobile health and encompasses the practice of medicine and public health supported by mobile devices (e.g., mobile phones, personal digital assistants, wearables). The user accounts were retrieved using the search queries. Then, every account retrieved was manually checked for its relevance, coded according to an actor group category by two coders. The actor categories are the following: 'institution', 'specialist', or 'advocate'. The coders further assigned the country of emission when the location field allowed them to do so. The actors identified as relevant for our study are included in our sample (N=4,120).

Selection of relevant tweets and pre-processing steps

From each of these Twitter users, we then collected up to the most recent 3,200 tweets (which corresponds to the rate limit authorised by Twitter API) which left us with more than 7.5 million tweets in total. To keep only the most relevant tweets about HT, we applied the following search query: '.∗ health.∗ | .∗ medicine.∗ | .∗medical.∗ | .∗patient.∗ | .∗technolog.∗ | .∗ medtech.∗ '. We also only selected tweets from our corpus that had been posted since January 2019. We applied several pre-processing steps including the removal of stop-words (e.g., 'the', 'our', 'of', 'at'), of special characters and symbols (e.g., '#', '@', emojis, emoticons), of punctuation, and of links (e.g., 'http(s)', 'www'), as well as the splitting of concatenated expressions (e.g., 'HealthTech' becomes 'health tech') and the lowercasing of the text.

Identification of salient topics surrounding HT

We conducted TM to provide more prompt and accurate insights into trends related to HT. TM enables us to extract dominant or salient topics in the tweets collected for the study. For instance, it can automatically identify important health topics related to HT and other important themes for the actors whose tweets were retrieved. A 'topic' consists of a cluster of words that frequently occur together. The logic behind TM uses contextual clues to connect words with similar meanings and to distinguish between the uses of words with multiple meanings (Blei, 2012). TM thus aims to reduce the complexity of the tweets to 'core' meanings so that we can identify what a given tweet is about. Topic models maximise the equation p(topic|document) x p(word|topic) for all given tweets in our corpus. It thus combines document classification (p(topic|document)) and keyword generation (p(word|topic)). Documents and words are given, topics are fitted iteratively starting from a random configuration. We used the popular implementation algorithm of Latent Dirichlet Allocation as implemented in the Mallet software to conduct TM (McCallum, 2002). We set the number of topics to be extracted to 150, which appeared to be the most relevant number of topics after several attempts. The extracted number of topics demonstrates a good internal and external coherence, which are two criteria proposed by Grimmer and Stewart (2013) to assess the reliability of the topic extraction. Each topic is represented by a list of top related keywords, which then need to be manually labelled with a view to proposing a possible interpretation. The distribution of topics can be assessed for external parameters, such as actor group and location.

Identification of framings of HT along important dimensions

We also aim to uncover framings of HT along important dimensions of the debate. To do so, we apply word embedding (hereafter WE) analyses which enable us to better understand the relationships between words. Therefore, instead of extracting a fixed number of topics as in TM, WE lets us choose how expansive the explored space should be as it provides a low-dimensional representation of the meaning of words (Sahlgren & Lenci, 2016). The underlying logic of WE implies that the model 'learns' scores for each word in the text for some arbitrary number of characteristics (also called dimensions). The WE method represents words as vectors, where each word gets a series of scores that position it in a multi-dimensional space. WE is thus useful for retrieving important synonyms and associations surrounding important dimensions of HT. It is also well-suited to build information retrieval contexts while letting us choose how wide the discursive space should be. We relied on the R library wordVectors (Schmidt, 2017) to train WE models (the models that we employed uses the function train_word2vec). This library enables us to achieve matrix operations that are useful in exploring embeddings, including cosine similarity, nearest neighbour, and vector projection with some caching that makes them much faster than the simplest implementations. The input must be in a single file and pre-tokenised, and the algorithm relies on the existing word2vec code implemented by Google in the C language (Mikolov et al., 2013). The algorithm produces a vector space, typically of several hundred dimensions, with each unique word in the corpus being assigned a corresponding vector in the space.

We followed some advice on the optimal set of parameters to use for training as defined by Mikolov et al. We used skip-gram as argument type which is better for infrequent words. We used hierarchical softmax as training algorithm. We produced 100 dimensions of the word vectors and used the argument window of 10, which is appropriate for skip-gram. More vectors usually mean more precision, but also more random error, higher memory usage, and slower operations. We used 3 threads to run the training process on. Furthermore, we did not use any minimal word frequency and we made no use of the epoch (or iter) parameter which provides passes to make over the corpus in training.

We can use visualisations to obtain a concept map plotting similar words close to each other. Words that are found in most discourses appear near the centre of the map, those which are restricted to very few documents appear on the fringes of the axes. We built models for the whole dataset, but also for each actor group (‘specialists’, ‘institutions’, and ‘advocates’) in view of generating an additional interpretative dimension related to the actors. We also applied stemming using the textstem R package (Rinker, 2018).

The proposed approach for conducting our research is summarised in Figure 5.1.1. Three main stages have been followed. Stage one captures the profiles and the relevant tweets using a list of search-queries. Stage two delivers insights from the tweets through various techniques, namely TM and WE. Stage three presents the findings in form of graphical representations and innovative visualisations.

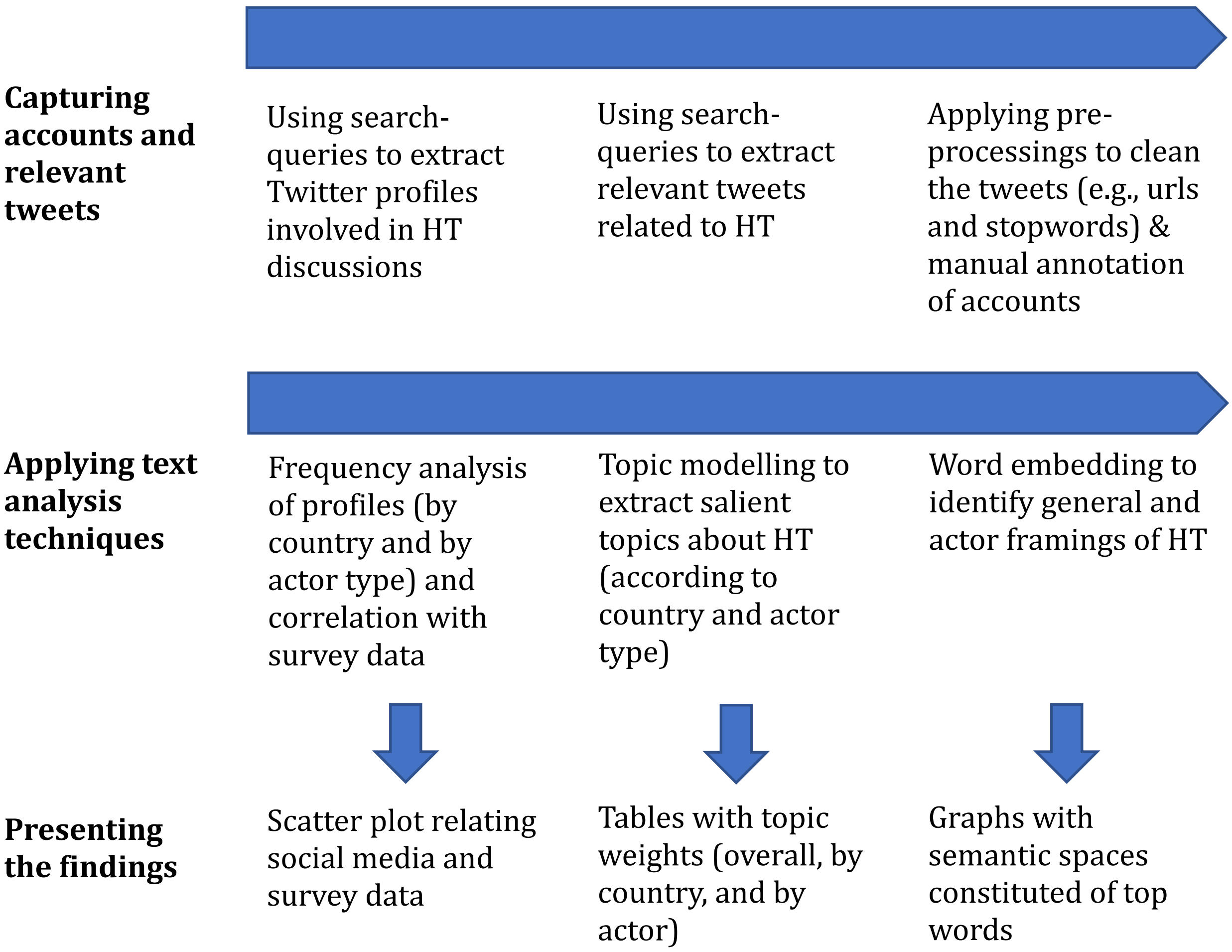


Figure 5.1.1: Proposed approach for conducting the research.

Results

Identifying the main actors involved in health technologies who are active on Twitter

The common population of users on social media platforms consists of non-affiliated users, users self-identifying with an organisation in their profile, official organisational accounts, influencers, fake accounts, and bots. Our sample of social media users active in the field of HT is divided between 40% institutions (public and private), 40% specialists (or practitioners), and 20% advocates. To be included in our sample, advocates must refer explicitly and primarily to HT in their profile description. For instance, journalists who cite HT as one of their minor interests are not included in our sample. Neither do we include users with either an irrelevant profile description or a very minor interest for HT. The profile descriptions allow us to derive shared characteristics among the different groups of users. Among organisations, there are as many public as private actors (including: universities, research institutes, hospital services, health authorities, private organisations, or corporations). Institutions use Twitter to promote their services (e.g., technological advancements) or policies (or regulations). With respect to specialists, they are essentially CEOs, CIOs, practitioners, research fellows working in universities, or private entrepreneurs. Specialists, in particular, rely on Twitter to publicise their research, their research agenda (e.g., events, conferences, webinars, etc.), and new challenges associated with their practice. Our sample of Twitter users thus reflects similar specialist positions as the respondents covered by expert surveys (e.g., HIMSS Analytics, 2019).

Our corpus is mainly composed of social media users from the United States (50% of users are from the United States), followed by users from Europe (40%), and a residual share from other countries (10%), including Canada, New Zealand, India and African countries. According to a spring 2019 Pew Research Center survey (Schumacher & Kent, 2020), the social media penetration rate is more pronounced in the United States than in Europe. In European countries, the use of social media varies significantly between countries Figure 5.1.2 below illustrates the distribution of the share of retweets in our corpus in relation to the national share of respondents from representative samples of national populations seeking health information on social media (we used the survey data from the 2014 Eurobarometer[31] to plot European countries (European Commission, Brussels 2015), and survey data from the 2013 Great American Search for Healthcare Information[32] for the United States). Figure 5.1.2 shows a positive correlation between the share of retweets about HT and the share of national social media users relying on social media to seek health information (Pearson correlation of 0.7 only for European countries and of 0.65 with the United States included).

[31] For more information on the survey report, see: https://ec.europa.eu/commfrontoffice/publicopinion/flash/fl_404_sum_en.pdf

[32] For more information on the survey report, see: https://www.webershandwick.com/wp-content/uploads/2018/11/Healthcare-Info-Search-Report.pdf

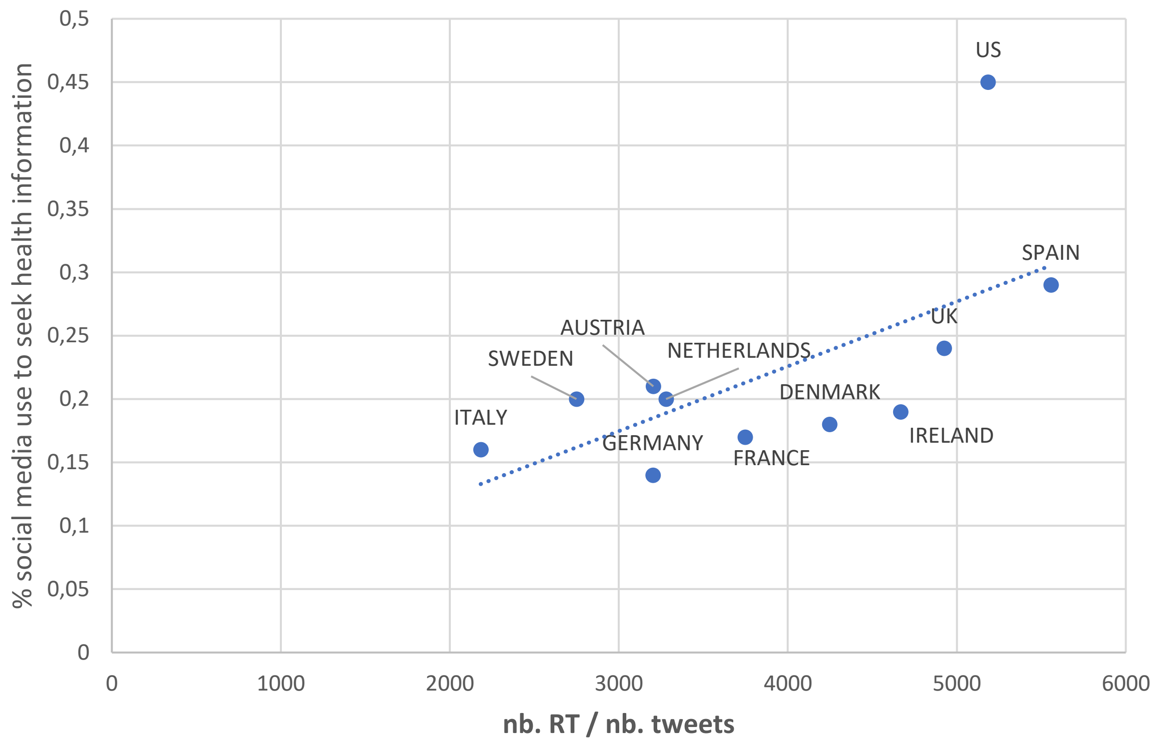


Figure 5.1.2: Relationship between the share of retweets from health technology actors (x-axis) and the reliance on social media to seek health information (y-axis).

Extracting salient topics surrounding health technology

In this section, we assess more precisely what the important topics addressed on Twitter by the users included in our sample are. The topics extracted are multi-fold and range from 'trendy' topics gathering interest on Twitter to recent advances in the field of HT and policy regulations. The vast majority of topics are relevant for our analysis. More than 80% of the topics extracted have a clear interpretation and are mutually exclusive. The remaining 20% are related to news, or to summaries of events, and are difficult to differentiate (the full manual coding and the mean topic weights that refer to the text mass that this topic covers, presented in percentage terms by regions – 'United States' and 'Europe' – and actor type –'specialist', 'institution', and 'advocate' – can be found in Annex 5.1.1). The findings from TM shows salient themes addressing the patient–doctor relationship, patient-centred initiatives and needs, healthcare systems, innovative solutions, big data challenges, market opportunities, and customer experience (see Annex 5.1.1).

Because there are major differences in the health systems prevailing in the United States and in Europe, we also assess whether these differences are reflected in the prevalence of topics in a cross-cultural perspective. For instance, OECD data (2019) show that the amount of money Americans spend on healthcare services is higher than in any of the other developed countries in the world. At the same time, only 23% of Americans think that they get the best care possible, compared to an average 70% of EU citizens who are satisfied with the quality of health-care. We therefore expect to find differences between cultural contexts, especially in terms of which topics are emphasised to meet the expectations of patients, citizens, and communities. To test this hypothesis, we assess the difference between the United States and Europe in the prevalence of the topics extracted (see Annex 5.1.1 for the topic weights for the two regions 'United States' and 'Europe'). The topic weights show differences in topic prevalence between European countries and the United States. For instance, the latter places greater emphasis on risk management and private funding, whereas European countries focus more on health literacy, practitioners (as opposed to scholars), and start-ups.

We also expected to find different topic salience across actor types, which we test using the topic weights (see Annex 5.1.1 for the topic weights for the actor types 'specialist', 'institution', and 'advocate'). Specialists tend to focus more on concrete and direct challenges and topics. For instance, they focus on subjects such as patient happiness and patient monitoring, as well as on the latest technological developments and the COVID pandemic response. Furthermore, specialists have a direct interest in learning/training/teamwork, which are additional direct concerns in their daily practice. In contrast, institutions focus more on indirect problematics, such as corporate policies, projects, and finances (e.g., funding, market growth, profits margins, and marketing), as well as on more strategic or global topics such as general policies and health concerns (e.g., smoking, home care, and pregnancy). Whereas institutions and specialists have a scientific and economics-oriented discourse about HT, advocates spread content mostly related to highlights, wellness, well-being, and wearables.

General framing of health technology on social media

In this section, we apply WE using different strategies to extract relevant framing related to HT. Compared to TM, which provides one particular idea of a given theme, WE models enable us to search for relationships embedded in words. They can thus provide us with an overview of families of related terms, i.e. words that are found in similar contexts. In this respect, WE is a good strategy to reveal word relationships. It separates and clusters words that are semantically similar. A way to make sense of the WE is to build a text network to derive the similarities between each pair of words. Based on this network, we can build a visualisation of word relationships. This visulalisation is also referred to as a 'conceptual map'. The Figure 5.1.6 in the Annex 5.1.2 displays such a map based on top terms of our corpus of tweets. It shows pairings where words with similar meanings are nearby[33]. For instance, 'io' and 'robotics' (see upper middle pane) clearly have something in common and are plotted next to each other. Terms that appear together (e.g., 'interoperability' and 'telemedicine') cluster together on the chart (see lower middle pane).

In the following analyses, we rely on WE to obtain ways of interacting with the vector space beyond word pairings in order to build information retrieval contexts. For instance, we can thus assess the distance between two words, or between one word and several related words. In our application, we used a list of diseases and health issues for which we calculated the distances to HT related terms (notably, 'healthtech' and 'medtech'). This enables us to demonstrate that certain health issues – such as obesity, addiction, heart disease and COVID – are perceived as more 'well-suited' in terms of HT (see Table 5.1.1 containing cosine distances between our list of health issues and HT related terms). It will be important to take this result into account when interpreting the next analyses as Table 5.1.1 indicates what health 'domains' are likely to be prevalent in our corpus of HT-related tweets.

WE can also be used to highlight connections between concepts in terms of word-vector relationships. This lets us plot a number of terms in a given discursive space. However, instead of specifying vocabulary items, we can also create text visualisations corresponding to word relationships. For instance, just as 'patient' and 'professional' are individual vectors, 'patient – professional' can also be represented in a semantic space. We can simply indicate this by comparing our words to a new vector defined as the difference between the two words ('customer' and 'industry') within the same vector space. This enables us to score any words based on their relationships in order to create word representations specific to any desired word relationship.

[33] We relied on the python library texplot and on the Gephi software (see implementation by McClure, 2015).

Table 5.1.1: Cosine distances between health issues and HT related terms ('healthtech' and 'medtech').

Disease or health issue category ('term used')	cosine	cosine rank
Weight disorders ('obesity')	0.061	1
Addiction disorders ('addiction')	0.060	2
Cardiovascular diseases ('heart')	0.050	3
COVID ('covid')	0.047	4
Mental disorders ('mental')	0.041	5
Diabetes I & II ('diabetes')	0.039	6
Liver problems ('liver')	0.038	7
Hypertension ('hypertension')	0.035	8
Vascular diseases ('vascular')	0.028	9
Gerontology ('gerontology')	0.020	10
Oncology ('oncology')	0.019	11
Neurological pathologies ('brain')	0.017	12
Alzheimer's ('alzheimer')	0.017	13
Lung diseases ('lung')	0.016	14
Blood diseases ('blood')	0.010	15
Neurology ('neurology')	0.004	16
Sexually transmitted diseases ('aids', 'hiv')	0.004	17

Figure 5.1.3 displays a semantic space composed of two-word relationships: 'patient – professional' and 'challenge – opportunity'. The relationship 'challenge – opportunity' aims to illustrate an important opposition in health data usages. The increased availability of HT offers opportunities to improve important aspects relating to diseases and injuries, but HT can also be framed with respect to emerging challenges and concerns, either from the patients' or the professionals' perspective Figure 5.1.3 captures distinctions between these two continuums.

We will now explain the methodology applied to extract the words plotted in Figure 5.1.3. A similar methodology will be used for the subsequent figures (also refer to Schmidt (2017) who presented the method on which we elaborated to build our own analyses). Regarding Figure 5.1.3, we first extracted top words mostly associated to HT using the following query: 'healthtech | medtech | digitalhealth | ehealth | digihealth'. Then, we extracted the top words closed to opportunities (query: 'opportun | solute | advanc') from which we subtracted the top words closed to challenges (query: 'challeng | difficulti'). This forms the 'opportunity vector'. We also extracted the top words closed to patients (query: 'patient'), from which we subtracted the top words closed to professionals (query:

‘profession’). This forms the ‘patient vector’. On this basis, we calculated the cosine similarities between the ‘HT vector’ and the ‘patient vector’, as well as between the ‘HT vector’ and the ‘opportunity vector’. Because of the big differences in the frequency of individual words, we weighted the cosine scores by the relative frequency of each word. For readability purposes, we only plotted the top 130 words.

The shape of the word distribution on Figure 5.1.3 shows that HT tend to be framed as opportunities on the patients’ side and as challenges on the professionals’ side. Words on the upper left pane (such as ‘digitalmental-health’, ‘clearhead’, ‘behaviourchang’) are related to the patient and the opportunity space. Words on the lower right pane (such as ‘patientcentr’, ‘patientexperi’, ‘clinicaltri’) are related to the professional and challenge space. These words indicate areas in which there is a need to improve the application of HT, with a focus on digital and virtual HT (e.g., ‘tele-health’ and ‘virtualcar’) Figure 5.1.3. enables us to assess further salient trends. First, on the patient side, there are words related to concerns about data safety and the guarantee of their privacy (e.g., ‘dataprivaci’), as well as a call for more ethics (e.g., ‘techforgood’) and medical knowledge (e.g., ‘digitalhealthliteraci’ and ‘mindblow’). Second, this trend is shared by the professionals who emphasise patient empowerment (e.g., ‘patientdrivenhealthcar’).

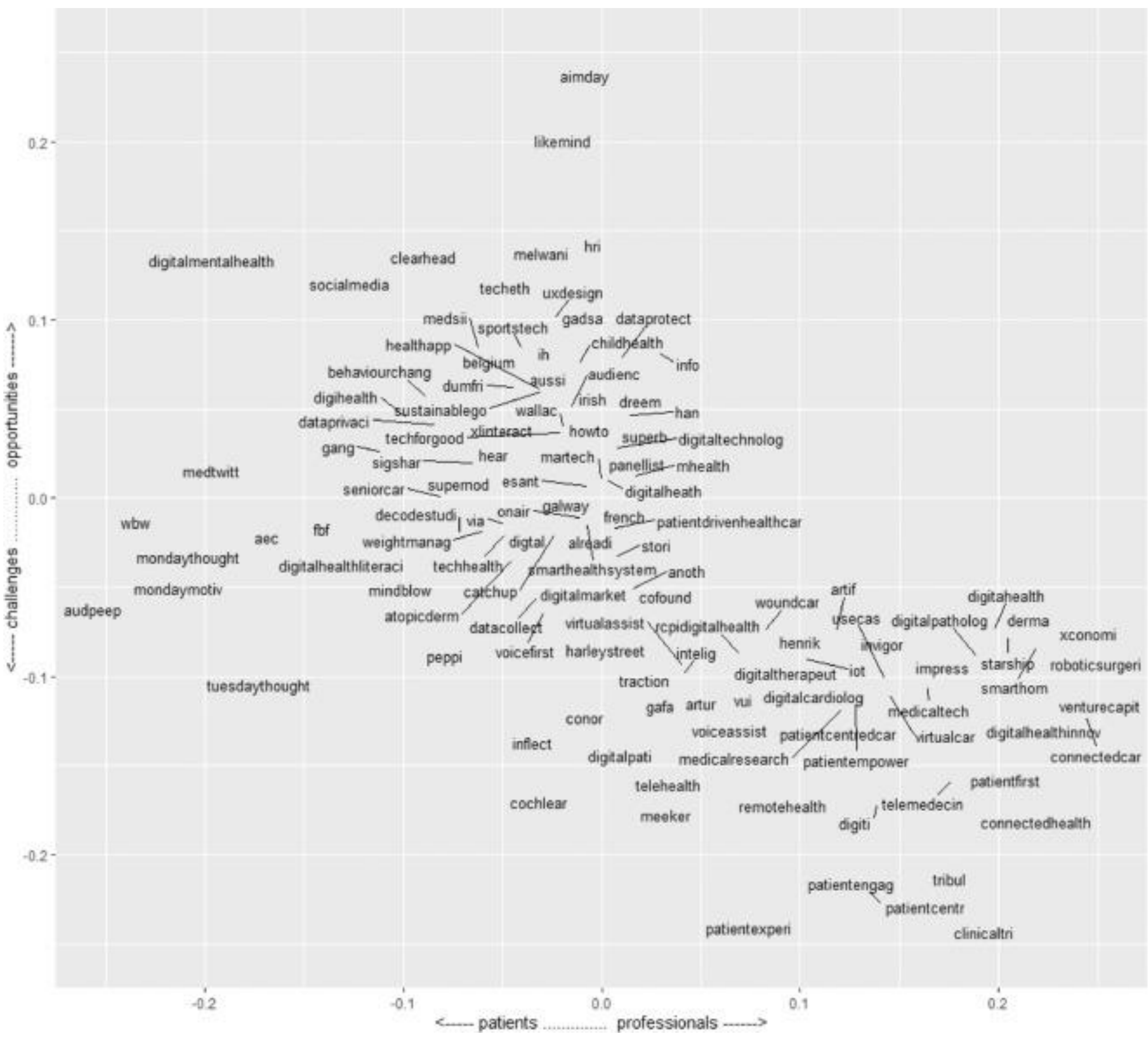


Figure 5.1.3: Semantic space composed of the two-word relationships 'patient – professional' (x-axis) and 'challenge – opportunity' (y-axis).

Overall, our findings tend to indicate a positive 'tonality' (or connotation) at the word level. This can result from the fact that the actors included in our sample are more likely to be favourable than critical toward HT. To assess this possible bias, we conducted a sentiment analysis of the tweets from the three groups of actors using the R package sentimentR (Rinker, 2019). We found that there is a general pattern toward positive language about HT (see Figure 5.1.7 in the Annex 5.1.3). However, the three groups significantly differ in their mean sentiment, with institutions and specialists relying on a more positive language than advocates (significance level of Student-test for p-value <0.05; see also Annex 5.1.4 to see the distribution of sentiment by actor group). This means that our analyses are more representative of the perspective of actors who are rather supportive of HT, thus under-representing views from other Twitter users who are critical (or sceptical) about the benefits of HT.

Actor framing of health technology on social media

In this section, we apply WE to extract relevant actors' framing of opportunities and challenges associated with HT. To do so, we can also retrieve similarity scores while keeping the information about the actor type. To maintain discursive distinctiveness between actors, we trained word vectors separately for tweets from each actor ('specialist', 'institution', and 'advocate'). Merging the scores for each actor enabled us to identify terms that are shared among all actors ('shared' words) and terms that are more salient for a given actor compared to the other actors ('specialist', 'institution', and 'advocate') Figures 5.1.4 to 5.1.6 are based on this logic and display the discursive differences for each group of actors.

Figures 5.1.4 and 5.1.5 focuses on the similarity scores associated with new technologies and with privacy in relation to the terms 'professional' (y-axis) and 'patient' (x-axis). Top words are plotted in this discursive space and coloured according to the actor type. In Figure 5.1.4, specialists' framings especially emphasise concerns related to their daily practices and research (e.g., 'medtechinnov', 'showcas' and 'futurofhealth'). In contrast, institutions' framings mainly emphasise business opportunities (e.g., 'charitesummit', 'investor' and 'standout'), but they also focus on the opportunities offered by the collection and analysis of health data (e.g., 'healthanalyt' and 'healthinfo'). The advocates mainly emphasise the concrete applications (e.g., 'videoconferenc', 'healthit' and 'voitech') and domains of HT (e.g., 'biotech', 'prosthes' and 'ophtalmolog'). The shared discursive space provided by Figure 5.1.4 is in favour of more predictive medicine and new research skills, notably with the reliance on artificial intelligence and big-data analytics.

In Figure 5.1.5, we show the top words associated with the term 'privacy'. There is a trend to associate 'privacy' concerns to the 'patient' side (x-axis) rather than the 'professional' side (y-axis). Furthermore, shared words demonstrate that data-driven technologies raise data privacy discussions associated with professionals' obligations (e.g., 'compli', 'transpar', and 'ethic'). Specialists also emphasise data access and algorithms to analyse these data. On their part, institutions are more concerned with security issues (e.g., 'hitsecur' and 'cyber'), as well as with data sharing and authenticating strategies (e.g., 'patientaccess' and 'authent'). Advocates emphasise the need for accountability (e.g., ‘inform’), confidentiality, interoperability and security (e.g., ‘cyberattack’, ‘protect’) concerning HT.

HT based on continuous data collection and algorithmic evaluation have gained importance during the COVID pandemic (Scott et al., 2020). The growing interest in continuous data collection and the algorithmic evaluation of personal health data exacerbates concerns about data privacy. To highlight recent important trends, we use similarity scores associated with privacy concerns based on their distance from HT before and after the COVID pandemic.

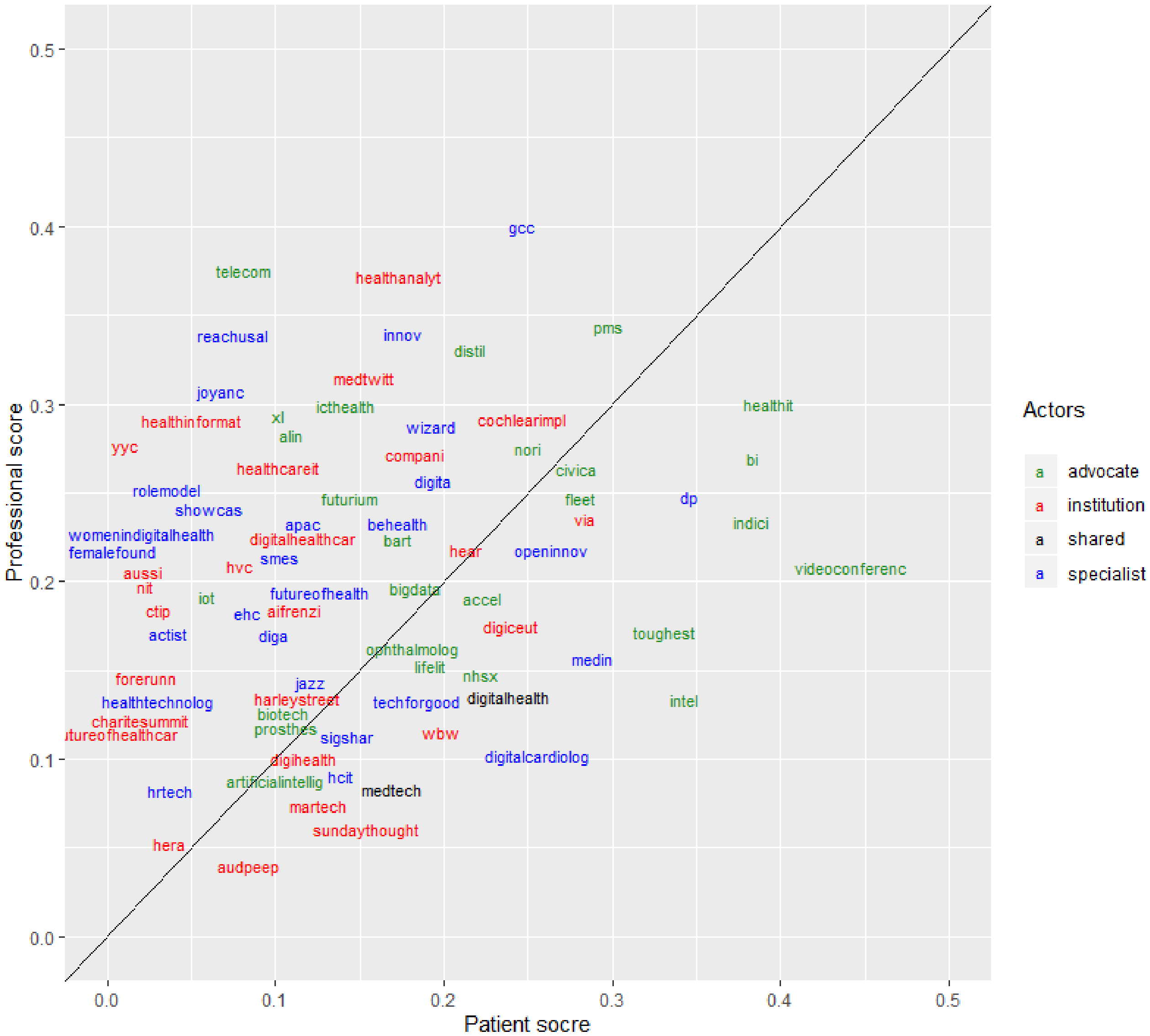


Figure 5.1.4: Top words associated with technology by similarity to patient (x-axis) and professional (y-axis) by actor type.

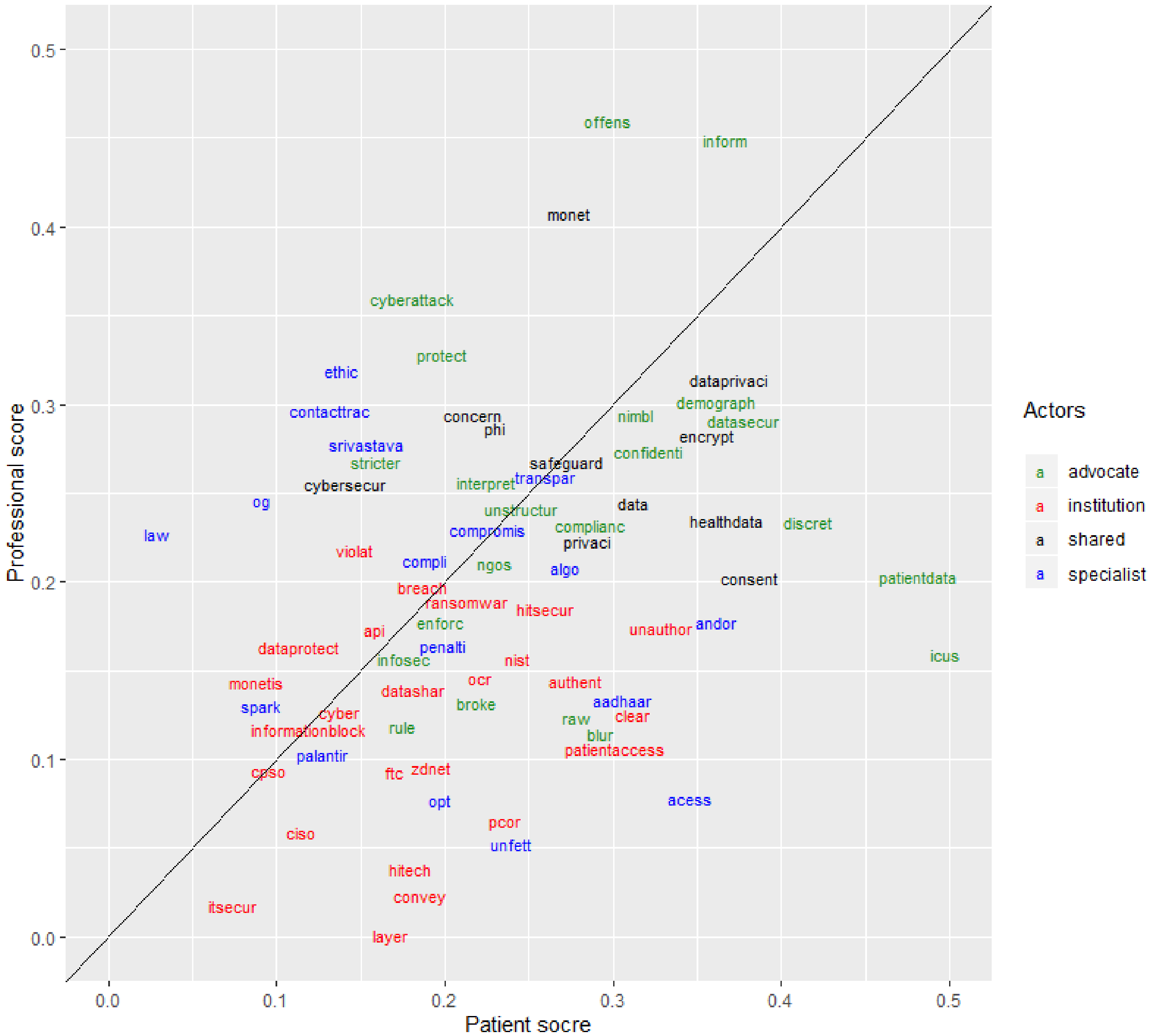


Figure 5.1.5: Top words associated with privacy by similarity to patient (x-axis) and professional (y-axis) by actor type.

Figure 5.1.6 shows a discursive shift between before and after the COVID pandemic, with focus moving from the professionals' to the patients' side. Furthermore, there is also an evolution from concerns related to (cyber)security to data storage and computing between the 'pre-covid' and the 'post-covid' periods. The 'pre-covid' period also rassembles more words associated with ethical considerations (e.g., 'liberti', 'imbal', and 'dilig'). In a similar vein, the 'pre-covid' period also emphasises non-discrimination issues. The legal orientation of HT discussion is present both before and after the pandemic (see shared terms in black: 'law', 'regulation', 'compliant' and 'rule'), although it seems to have taken a more punitive orientation in the 'post-covid' period (e.g., 'judgment', 'sanction'). This can be explained by the fact that the 'post-covid' period seems to be characterised by terms related to emergency (e.g., 'lifesav').

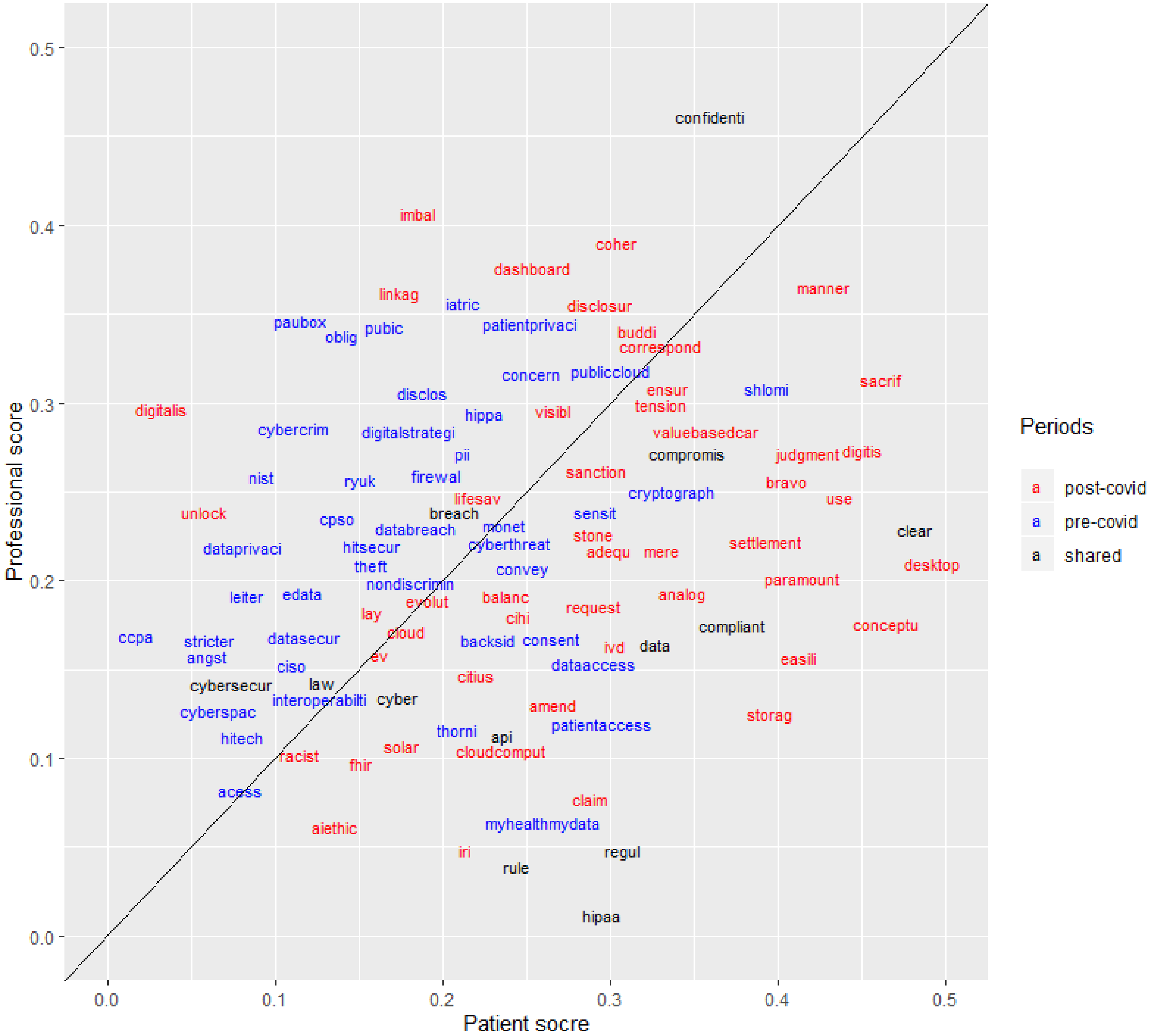


Figure 5.165: Top words associated with privacy close to patient (x-axis) and professional (y-axis) before and after the COVID pandemic.

Discussion of the main findings

In the first research step, we identified the major actors leading the public debate on HT on Twitter. We showed that the most represented actors are institutions and specialists (80% of our corpus) who are mainly located in the United States. We also found a positive correlation between the share of retweets from major actors' tweets and the share of the public relying on social media to seek health information. The lesser representation of HT advocates provides a partial explanation of the low proportion of topics related to news or lighter topics

In a second step, we relied on TM to extract topics and concerns underlined by major actors involved in the field of HT. We assessed important differences across cultural contexts (i.e. Unites States versus European countries) and actor types (namely institutions, specialists and advocates). Using TM, we showed that there are important differences between the United States and Europe in the prevalence of topics related to HT. For instance, the United States focuses more on risk management and private funding, whereas Europe focuses more on health literacy, practitioners, and start-ups. The topics extracted also showed different focuses among the actors. Institutions focus more on indirect, global, and strategic problematics, whereas specialists are more concerned with direct and concrete problems. Our dataset shows no particular pattern for advocates. Advocates are also active actors in the HT field, but they focus on less substantive themes, such as wearables, well-being, and healthy lifestyles.

In a further step, we relied on WE to gather general and actor-specific understandings of HT along important dimensions (see Figure 5.1.3). The semantic space crossing two relationships, namely 'patient–professional' and 'opportunity–challenge', shows that current challenges lay particularly in the relationship between patients and professionals, both in terms of patients' empowerment and in access to health data and information. There is also an emphasis on new development opportunities (e.g., equipment and wearables). Furthermore, professionals focus on what could be well-suited domains (e.g., imaging and videos for diagnostics), whereas patients are concerned with data protection issues (e.g., in terms of artificial intelligence, demystification, and customer experience).

The discursive spaces along the 'patients' and 'professionals' dimensions display important (and perhaps opposing) challenges between these two actors in terms of technological innovation and privacy concerns. For instance, specialists and institutions focus on adapting to HT by learning and developing new applications, whereas advocates are concerned with data privacy and also insist on the importance of data protection (see Figure 5.1.4).

The new challenges regarding privacy imply that practitioners will tend to focus on their responsibilities and obligations (or liabilities) by focusing on legal, ethical, and IT security concerns (see Figure 5.1.5). There is also a clear patient demand for more control of health data (e.g., in terms of transparency, access, and interoperability). There is, thus, patient demand for a more horizontal relationship with practitioners.

We concluded by analysing a possible shift in concerns related to privacy issues before and after the COVID pandemic. We note that word scores linked to privacy have generally become more prevalent in relation to the patients' side since the beginning of the pandemic. Furthermore, we discern two broader categories of terms related to either 'legality' or 'ethics'.

Theoretical contributions

In line with the literature suggesting that social media serve efficiently for health care discussions (Jiang, 2019), our findings demonstrate the usefulness of investigating HT-related discourses online. In particular, the proposed study discusses some of the opportunities and concerns expressed by users posting about HT on social media, while also discriminating the groups of users. The different groups of users that we investigated strategically use social media according to their characteristics (e.g., public or private entities, practitioners or business managers, and influencers) and according to the purpose of the information delivered (e.g., raising public awareness, selling products and services, and raising concerns). In this study, we have conceptually identified the most salient topics and framing of HT based on the words that are relevant for these groups of users.

Another theoretical contribution relates to the methods used for investigating HT discourse in terms of topicality and framings. First, we adopted fully automated methods to collect and analyze the collected tweets, which enabled us to have significantly more variety in the topics and frames analysis than would be feasible with manual annotation. Furthermore, the use of word embedding combined with innovative visualisations along important dimensions can be applied in other fields of technologies and information management to complement managerial and social technical perspectives. This methodology will hopefully guide future researchers to perform in-depth analysis in individual HT subdomains (e.g., privacy concerns, business opportunities, crisis management).

Implications for practice

Our findings reveal that social media are not only a useful source of information about the current state of HT (e.g., business opportunities), but also about which concerns surround HT policy and the role of HT in crisis management. The practical implications of our study can thus be segregated into several audiences, namely the users, the (public or private) compagnies, and the (private) practitioners or business managers. On the basis of these findings, the different user groups can decide what aspects should be prioritised and how to frame them so as to address salient concerns. Concerning the users, our analyses reveal that they are mostly concerned about privacy and security when discussing HT. Therefore, it is important that social media platforms provide users with authentic and balanced information about HT so that users can make informed choices and find answers to their concerns. Concerning companies, we show that social media can be used as useful channels to raise public awareness by promoting specific campaigns and to monitor trends in disease conditions (e.g., COVID-19 crisis management). Concerning the practitioners, they can usefully rely on social media to provide innovative solutions to diseases while putting forward their own business.

From a methodological perspective, our findings also have practical implications for the research community. Studying topics and frames stemming from social media accounts of specific users enables us to derive the most salient dimensions of the debate about HT. However, we still know little about whether the concerns and opportunities expressed are representative of those of the general population. It would therefore be useful to complement the proposed methodology by using additional methods, such as opinion surveys. In this case, the findings of our study could serve as a basis for identifying the HT areas and aspects worth surveying at national levels while considering possible country-effects on health care systems and health promotion.

Concluding remarks and outlook

Our study makes two important contributions to the research on HT. First, it provides an exhaustive picture of the major actors in the HT field actively posting on social media and of what topics and framings they share with the wider public on Twitter. In this view, our study represents an important step towards a better understanding of how and why social media can impact citizens' health attitudes and behaviours. The second contribution of our study is to provide an innovative methodology for investigating important HT framings using creative visualisations.

Our study nonetheless entails several limitations that would be worth addressing in future research. First, Twitter is only one possible social media platform, with specific rules and conventions. It is less used than Facebook and allows less extended user contributions. However, Twitter data are submitted to fewer access restrictions and also cover an international population. These characteristics make Twitter data suitable for the purpose of our analysis. Nevertheless, other professional platforms, such as LinkedIn, could offer an alternative source of data for studying in greater depth how institutions and specialists in the field of HT portray themselves and recruit specific profiles.

Second, future studies could also examine the evolution of HT discussions online by accessing historical data. Our study is limited to the most recent tweets and, thus, does not allow for the study of the evolution of HT themes or concerns over time. Our corpus of actors testifies that a historical study is feasible, as the majority of Twitter accounts were created several years ago (the majority were created from 2011 onward). HT are characterised by rapid changes in the health and social care sector, and the development and impact of these changes are hard to predict. Our data already account for the current shifts in information technology and big data, automation, and artificial intelligence. This shift was brought to light in a recent study by the OECD, which identified a new demand for skills and specialisations among health and social care workers, while reducing the importance of other professional roles (OECD, 2019).

Third, we restricted our analysis to major actors, which, possibly, does not give voice to more negative or concerned opinions about the use of HT (see end of section 4.2). Therefore, future studies might include the network of Twitter followers to seek a more global view of HT as perceived by the public. In a similar vein, we encourage the development of surveys covering public reliance and concerns about HT that can complement existing surveys conducted with official health actors. Fourth, we focused on discourses surrounding HT from the perspective of topicality and framing. However, another important discursive component relates to tonality and emotion, also referred to as opinion mining. For instance, Ridhwan and Hargreaves (2021) relied on opinion mining to investigate public sentiment about the COVID-19 outbreak in Singapore. They showed how policy measures triggered different emotions, drawing from previous studies using social media to monitor public health-related issues expressed online

(García-Díaz et al., 2018). We should nonetheless note that opinion mining does not always reflect stance (e.g., favouring or rejecting a policy issue). Other metainformation, such as retweets or likes, could also be useful in measuring support for – or the contestation of – given HT aspects. This would be useful for understanding how the broader public reacts to the tweets posted by each user group.

To date, most surveys about HT have been conducted with specific groups (such as health professionals and institutions), but there are few indicators of the perception and usage of HT by the general public (or representative samples of national populations). A space has thus been incentivised for research that identifies people's experiences when taking up or resisting new digital HT. Our study provides insights about what could also be potential survey interests. Developing survey items about HT would allow for a direct comparison between spontaneous online discussions and structured survey opinions.

Despite these limitations, we are confident that the findings from our study can help major health actors (such as HT companies and practitioners) to better target their campaigns while considering the concerns expressed by the different online audiences. This is in line with findings from Obembe et al. (2021) who studied tourist public responses on social media to crisis communications during the early stages of COVID-19. Indeed, the authors have shown that online publics played a key role in shaping the narratives of the crisis, thereby facilitating public engagement. However, a combination of analytical strategies and data sources is needed to take the next steps beyond the 'what has happened' to the 'why it happens' (Kar & Dwivedi, 2020).

5.2 The deployment of social media by political authorities and health experts to enhance public information during the COVID-19 pandemic[34]

Introduction: Political authorities' and experts' use of social media to inform the public about health issues

Social media have increasingly been used by officials – such as political authorities and health experts – to disseminate health information to the public (Gough et al., 2017). This trend reached unprecedented levels during the COVID-19 pandemic in efforts to enhance trust in scientific expertise (van Dijck & Alinejad, 2020). In parallel, there was an increase in citizens' reliance on social media to obtain news and COVID-19-related information (Nielsen et al., 2020). Social media platforms thus played an important role in political authorities' communication with citizens as the pandemic led to a narrowing of the topic agenda on these platforms, with an increased level of Twitter activity by political and health experts (Rauchfleisch, Vogler & Eisenegger, 2021). Specific health and political actors' communications are also likely to be directed to (or at least mention) target population groups, whose acceptance of the measures is essential to achieve the desired policy outcome (e.g., Martin et al., 2020).

Nevertheless, the roles of social media and public trust during the COVID-19 pandemic are relatively new topics that deserve more attention from the research community. Researchers still possess little knowledge of how public trust in health experts and political authorities is comparable with social media trends during an epidemic. Our study contributes to understanding these processes by drawing on the Swiss context. We are interested in answering the following research questions: Is the level of public trust measured through opinion surveys also reflected in social media users' engagement with messages from figures of authority? What were salient clusters of the COVID-19-related online discussions? How were these clusters received by the broader audience? The investigation of levels of public trust in (political and health) authorities in the context of (health) crises is not new. Most such studies have relied on survey data. For instance,

[34] This chapter is a slightly adapted version of the article that has been published as M. Reveilhac (2022): "The deployment of social media by political authorities and health experts to enhance public information during the COVID-19 pandemic", SSM - Population Health, 19, 101165.

psychological variables, such as trust and worldview, strongly influence people's risk perceptions and acceptance of health measures (Siegrist & Bearth, 2021). Sustaining trust has thus been a challenge over the course of the COVID-19 pandemic. Previous studies have found that trust in health authorities and government institutions is importantly correlated with citizens' compliance with public health policies and guidelines (Blair et al., 2017; Vinck et al., 2019). Public trust in political authorities can also be influenced by social media. Indeed, previous studies have identified criticisms of political authorities and institutions, thus suggesting an expression of distrust in political authorities (Linde-Arias et al., 2020; Roy et al., 2020).

In addition to survey data, social media have offered important advantages in delivering interactive communications between political authorities and citizens during the COVID-19 pandemic (Chen et al., 2020). Furthermore, social media are a useful source of data for rapid and exhaustive data collection to support evidence-based decisions based on public reactions (Huerta et al., 2021) when traditional face-to-face approaches are deemed difficult (Grow et al., 2020). Nonetheless, this focus on social media communication does not come without challenges, as these platforms also constitute a means for citizens to bypass traditional media outlets, provide channels for carrying out verbal attacks on political authorities and facilitate the spread of mis/disinformation (Brennen et al., 2020). Empirical evidence on how trust in health experts and political authorities has evolved in the different phases of the COVID-19 pandemic, acquired by comparing offline and online trends, is needed. Furthermore, van Dijck and Alinejad (2020) found evidence that the current phase of the public debate is an important factor that affects opportunities for actors to communicate with the public. In this regard, our study shows how the reception of the online communications of political authorities during the pandemic can provide us with complementary insights to better grasp the levels of trust measured in opinion surveys. In particular, we draw on two complementary measures of public engagement related to major actors' handling of the COVID-19 pandemic in Switzerland. These actors included not only health officials (e.g., Federal Office of Public Health (FOPH) and Taskforce experts) and political authorities (e.g., Federal Council, cantonal executives, members of Parliament, representatives of national parties, and elected politicians) but also the media, important economic actors, and universities. Including social media and survey data within the same study enables us to investigate the real-time dynamics of public engagement and trust in scientists and public health authorities.

We conducted several research steps:

First, we benchmarked how these actors' social media messages were received by the broader online audience compared to survey measures of public trust in similar actors. As a reception measure on social media, we relied on negativity in replies and on the sum of interactive features (e.g., likes and retweets) attached to messages from health and political authorities. Liking and retweeting are low-effort interactions that likely indicate endorsement and support of the original message. For this reason, they convey a measure of the popularity of a message (Guerrero-Solé, 2016), and summing these interactions provides us with a reasonable measure of engagement. In contrast to retweets and likes, interactions with replies require writing and do not necessarily reflect users' support. Therefore, we also relied on the overall negativity of user replies to messages from health and political authorities, which was assessed through dictionary-based sentiment detection.

Second, in line with previous studies demonstrating the usefulness of including the main topics of discussion among the active variables used in correspondence analysis (Gesualdo et al., 2022; Zengul et al., 2021), we identified major associations in online discussions of the COVID-19 pandemic that were characterized by different political and health authorities, by the phases of the COVID-19 pandemic, by the topics extracted through automatic text classification, and by the mention of target populations. The main clusters of discussions were then approached with hierarchical clustering to consider how they were characterized in terms of interactive features (e.g., likes and retweets) and of the negativity in other users' replies.

Study background

The role of social media in crisis communication

Despite the existence of studies underlining the role of social media in informing the dynamics of the public debate about the pandemic, few studies have relied on social media to examine political authorities' communication behaviour when attempting to raise public awareness (Zhao et al., 2020). Research on political authorities' use of social media to increase public attention to epidemics has been conducted in the contexts of the Ebola outbreak (Strekalova, 2017) and the H1N1 pandemic (Liu & Kim, 2011).

Shortly after its emergence, COVID-19 became one of the major concerns for policymakers and the public worldwide. Meanwhile, social media platforms have increasingly become primary sources of news and information (Mitchell, 2016). In this context, political authorities and health experts have strong incentives to maximize their social media efforts, especially during crises (Graham et al., 2015; Tagliacozzo & Magni, 2018). During the COVID-19 pandemic, social media have played a particularly important role in many countries (Sha et al., 2020; Thelwall & Thelwall, 2020), but they have also been linked to the spread of misinformation (Bauer, Freitag, & Sciarini, 2013).

Studies using social media to measure these attitudes and behaviours in relation to actors' communication have found that actor expertise has an important impact, as users' exposure to scientific social media messages leads to improved public knowledge (Vraga & Bode, 2017). Furthermore, evidence has suggested that mass publics are receptive to important information from governments (Goldberg et al., 2020). Both political communication and scientific expertise converge in their intention to incentivize specific behaviour and attitude changes. Indeed, government communication during crises is most effective when it translates scientific and technical information (Herovic et al., 2020). Gilardi, Gessler, Kubli, and Müller (2021) further found that social media challenge the capacity of party and media elites to craft a consensus regarding the appropriateness of different measures as responses to COVID-19.

However, social media also challenge political authorities' communication, as they allow multiple stakeholders and groups to shape social and political agendas while bypassing traditional gatekeepers such as news media (Jungherr & Gayo-Avello, 2020). For instance, Gilardi et al. (2021) investigated policy responses to COVID-19 promoted by Swiss political and health authorities, with a special focus on policy solutions, namely, face mask rules and contact tracing apps. The authors analysed the salience of these policy solutions to the COVID-19 problem. They found that the debate on face masks was led by the attentive public (a group of users who follow the accounts of at least five Swiss news outlets) and by politicians, followed by parties and newspapers. Social media thus challenge the capacity of party and media elites to elaborate appropriate measures as responses to a major health crisis.

In addition to the type of actors engaged in this debate, the stage that the public debate has reached is an important factor influencing how actors' messages are received by the public. The development of crisis management strategies by actors can indeed highly impact perceptions of their institutional expertise for handling the pandemic. For instance, van Dijck and Alinejad (2020) showed that the phases in the public debate were an important factor affecting how institutional actors engaged in communication with the public on social media. Social media are deployed to both undermine and enhance public trust in scientific expertise, and actors need to adapt their communication at the various stages of a public debate.

Measuring the success of actors' messages on social media

In public health crises, the communication of actors, such as political elites and health experts, is crucial for compliance with policy measures. When communicating with the public, these actors aim to reach target audiences with a view of inducing behavioural and attitudinal changes (Vinck et al., 2019). Measuring these changes is extremely difficult, but social media provide us with strong signals of how actors' messages are received by the public. The main way of interpreting these data consists of noting reactions (Cho et al., 2014) – such as expressions of satisfaction (e.g., likes), responses to messages (e.g., replies), and the propagation of messages (e.g., retweets) – which are widely used metrics in social media research (Stone & Can, 2020).

Against this background, it is important to measure the success of actors' messages, in particular by assessing the dynamics of the debate between actors and citizens but also by measuring the extent to which citizens spread these messages and interact with them. However, users choosing to engage in online debate about the COVID-19 pandemic may not be representative of the average person in terms of behavioural characteristics. For instance, social media discussions can be ideologically polarized and organized in echo chambers, as has been demonstrated in the case of the vaccination debate in Italy (Cossard et al., 2020) or of COVID-19 discussions in the United States (Jiang et al., 2020). Even though social media users are generally unrepresentative of the general public (Mellon & Prosser, 2017), studying social media reactions towards political authorities can still validly provide us with complementary information about the public attitudinal dynamics towards political authorities and health experts.

Several studies have investigated how political authorities' communication is received by a broader public of social media users. For instance, Raamkumar, Tan, and Wee (2020) examined the COVID-19-related outreach efforts of public health institutions in Singapore, the United States, and England and the corresponding public responses to these outreach efforts on Facebook. Using sentiment analysis, the authors found cross-country variations between overall sentiments towards public health authorities. Moreover, Mahdikhani (2021) studied public opinion and emotions at different stages of the COVID-19 pandemic, from the outbreak of the disease to the distribution of vaccines, and found that tweets with higher emotional intensity were more popular than tweets containing information about the COVID-19 pandemic. Furthermore, Teichmann et al. (2020) showed that different posting strategies on Twitter and Facebook were effective in drawing public attention to political authorities' health messages. For instance, the poster could often be more important than what was posted, and concise messages with clearly formulated health directives tended to receive widespread engagement. However, evidence suggested that health authorities faced low engagement with their social media posts related to the pandemic (Berg et al., 2021).

Comparing survey measures of public trust with social media reactions

To date, the most efficient way of accounting for public trust in political authorities and for public approval of policy measures is to rely on opinion surveys. In the context of COVID-19, national and international surveys have been developed to understand how public attitudes and behaviours are evolving in relation to the pandemic. Contrary to surveys that pose questions on well-defined concepts but usually require intensive resources to collect data, social media provide signals of opinions that can be accessed in a timely manner without the intervention of researchers (Diaz et al., 2016).

However, social media data are often messy, thus requiring specific pretreatments and cleanings before they can be meaningfully analysed (Klašnja et al., 2015). They are also usually not representative of national populations and lack information about personal attributes (e.g., Barberá & Rivero, 2014). Indeed, social media tend to be used predominantly by active user groups – such as politicians, influencers, journalists, and bots – who are influential in terms of public opinion (e. g., Hargittai, 2018), while survey data aim to be representative of what the wider public thinks.

Due to their respective characteristics, comparing the two data sources can provide meaningful insights into the congruence between offline and online support of political authorities and the policy measures that they propose to fight the COVID-19 pandemic. We view this comparison as useful for informing future actors' reliance on social media as a means of communicating with the public on health issues. More specifically, it enables us to assess whether their perceived trustworthiness within the population is congruent with how their communications are received online.

Case study: Public trust and the stages of the COVID-19 public debate in Switzerland

People in Switzerland have demonstrated an exceptionally high level of confidence in their government in recent decades (Mabillard & Pasquier, 2015). This can be explained by the sense of participation in political decision-making due to (semi)direct democracy and by the trust in political authorities' communications during critical situations (Freitag & Ackermann, 2016). Furthermore, Swiss citizens are highly satisfied with their healthcare system (OECD, 2019). The Swiss population continues to rely more heavily on traditional news media than on social media for information (Reveilhac and Morselli, 2020). Humprecht et al. (2020) further found that Swiss people are more reluctant than citizens of other countries to share false information about COVID-19 on social media.

The COVID-19 pandemic has been spreading in Switzerland since February 25, 2020, the date of the first confirmed case. Each phase of the public debate has been marked by political decisions leading to protective measures being taken against the pandemic (for a chronology of the pandemic in Switzerland, see Annex 5.2.1). While a large majority of the Swiss population had confidence in the federal authorities in the first wave of the pandemic, the study conducted by Hermann et al. (2022) demonstrated a drop in confidence in the actions of governmental bodies around January 2021. This decline in the level of trust was renewed during autumn 2021, which points to the enduring polarization of society between sceptics – who show little support for measures promoted by federal institutions and generally distrust mainstream information – and people supporting the actions of governmental bodies and showing confidence in the information coming from these official bodies. The authors also found that 9% of adults in Switzerland had taken part in at least one demonstration against COVID-19 measures in the past two years of the pandemic. This number is still significantly lower than the 40% of people who voted against the two COVID-19 proposals (in June 2021 and November 2021). This climate of distrust is particularly problematic in conjunction with direct democracy systems, as it can open the door to easy populist loops (Reveilhac & Morselli, 2022).

Data and analytical strategy

Assessing the relationship between public trust and social media reactions

To answer our first research question about whether the level of public trust measured through opinion surveys is reflected in social media users' engagement with messages from authority figures, we relied on two data sources. Figure 5.2.1 summarizes the architecture of the study to visualise the steps related to the method and data collection: We relied on tweets emitted by actors – namely, Swiss political authorities (the Federal Council and the Parliament as well as national parties and politicians), health experts (the FOPH and the national group of COVID-19 experts), news media (main daily or weekly newspapers), Swiss universities, and major business actors (business federations or trade unions) – from December 2019 to December 2021 (see Annex 5.2.2 for the list of Twitter accounts). We kept only original tweets and identified those related to the COVID-19 pandemic by relying on a dictionary approach using a list of COVID-19-specific search queries (see Annex 5.2.3). For all the tweets that were replied to by other users, we also retrieved these replies. Annex 5.2.4 displays the number of tweets that were collected for each group of actors and the number of selected tweets based on the list of search queries. The final sample contained 115,600 original tweets.

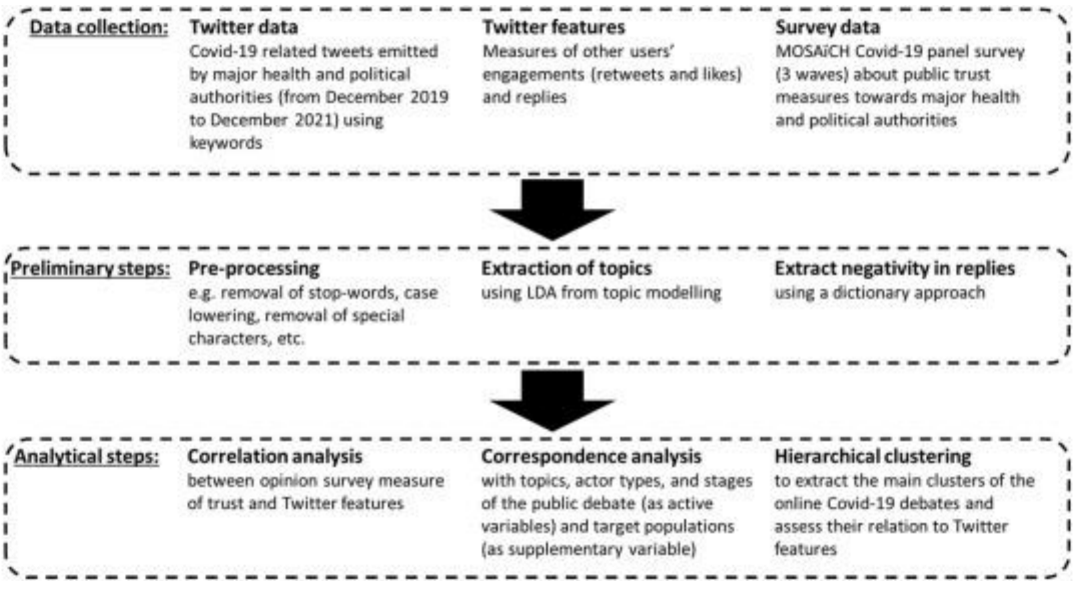


Figure 5.2.1: Analytical framework of the study.

Figure 5.2.2 (upper right panel) shows the relative proportion of other users' reactions (likes and retweets) to each actor's tweets on a monthly basis. In general, the news media and the political elite (encompassing the Twitter accounts of the Parliament, of the elected politicians, and of the parties) trigger the largest share of other users' engagement. Furthermore, the highest share of reactions for health experts was reached before the second wave.

Figure 5.2.2 (lower left panel) displays the relative share of replies from other users triggered by each actor group. The was a notable increase in the relative share of replies to the experts' tweets (including the Twitter accounts of the Taskforce and its members as well as the account of the FOPH). The relative share of replies to tweets from the political elite varies over time, with peaks around April 2020, March 2021, and October 2021.

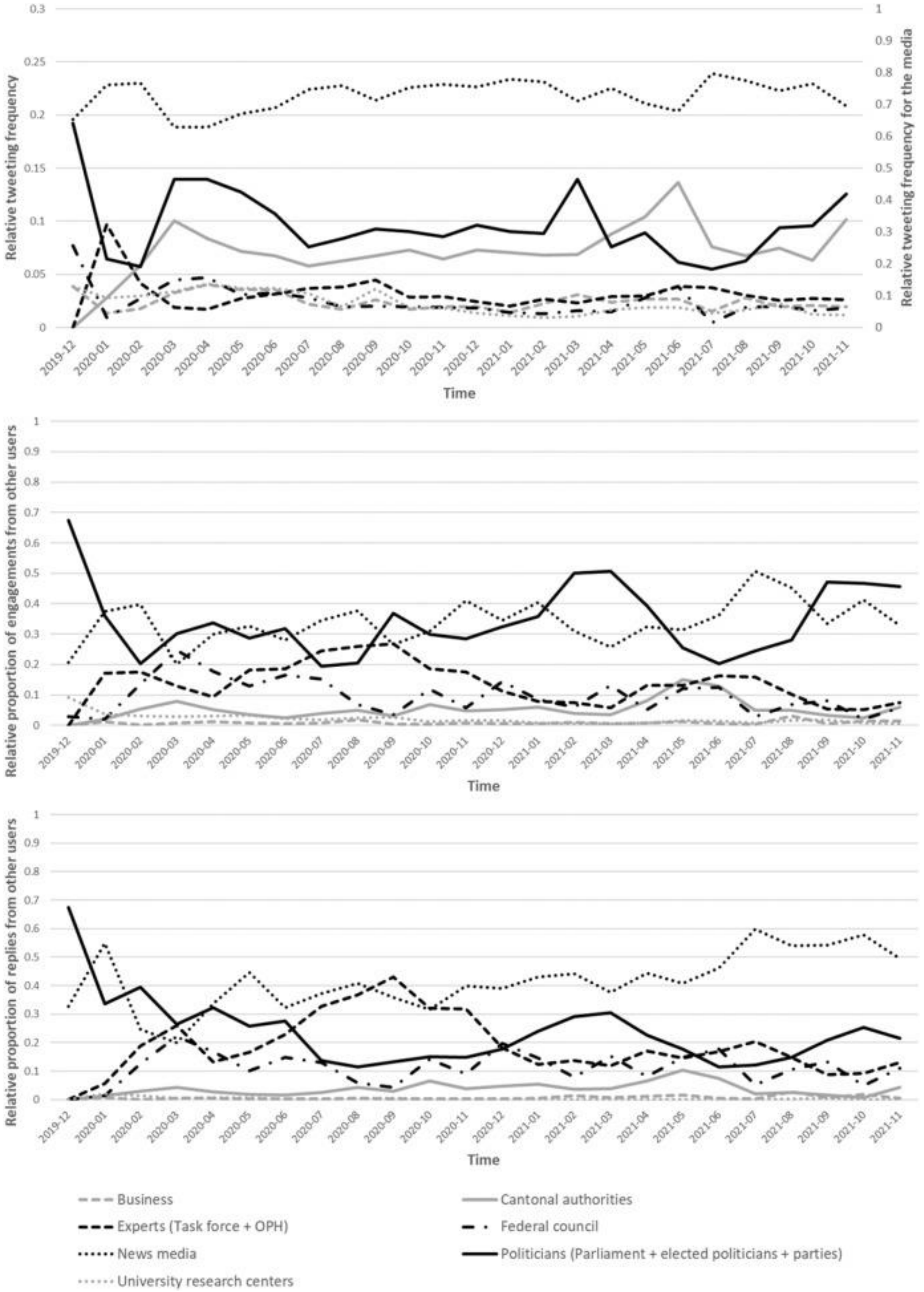


Figure 5.2.2: Relative share of actors' Covid-19 related tweets over time (upper left pane); Relative share of other users' reactions in terms of aggregated likes and retweets (upper right pane); Relative share of other users' replies over time (lower left pane).

Correlation between opinion survey measure of trust and social media features

To investigate whether the level of public trust measured through opinion surveys was also reflected in social media users' engagement with messages from figures of authority, we relied on public opinion data from the MOSAiCH-ISSP COVID-19 survey[35] to access measures of trust in similar official groups. The survey item used to measure trust was graded on a scale from 0 to 10 and read as follows: "How much trust do you personally have in the following institutions?" The survey spanned several waves, thereby allowing for responses to the same question to be obtained at different phases of the pandemic. We calculated the difference in public opinion between the first (from April 30th to July 13th, 2020, thus beginning at the end of the first wave of the pandemic) and the third (from March 19th to April 18th, 2021, thus corresponding to the third wave of the pandemic) survey waves for each institution (the second wave took place between October 2nd and November 2nd, 2020).

To assess the relationship between offline and online trends, we compared the difference in surveyed public trust to reactions from other social media users. We relied on the difference in engagement metrics (including likes and retweets) and negativity in replies from other users between the periods corresponding to the first and third survey waves. The negativity in replies was assessed for each language separately (only for German and French replies) using a dictionary-based approach. More specifically, we triangulated the LIWC (Pennebaker et al., 2015) and NRC dictionaries (Mohammad & Turney, 2013) to detect the negativity in replies from other users. Both dictionaries have been carefully translated by psychology and computational research teams in multiple languages. We label a tweet as negative if it was matched to words with a negative tonality in either dictionary.

This dictionary-based approach was chosen for this study rather than a machine learning approach. With a machine learning approach, the results would have been dependent on the quality of the training data, which, in our study, would have been difficult to obtain given that the data are in multiple languages and that human evaluators may not always be able to label them consistently according to the sentiment categories. However, to ensure that the results could identify negativity in replies with acceptable accuracy, we manually annotated a random sample of 100 replies in German and 100 replies in French.

[35] For more information, please see: https://www.swissubase.ch/en/atalogue/studies/13871/16853/overview

Both languages showed acceptable levels of accuracy for detecting negativity (89% in French and 92% in German). Cases where the dictionaries could not correctly identify negativity mostly contained colloquial slang or profanities. We used the negativity proportion for each actor to compute the correlation between survey and social media trends.

Salient topics in social media discussions

To answer our second research question about the salient clusters of the online COVID-19-related discussions and what reactions they triggered from the broader audience, we relied on two classification methods: topic modelling and correspondence analysis.

We used an unsupervised text classification method to extract relevant topics from political authorities' tweets. A "topic" consists of a cluster of words that frequently occur together and thus have similar meanings (Bauer, Freitag, & Sciarini, 2013). Documents and words were given, and topics were fitted iteratively starting from a random configuration. We used the popular implementation algorithm latent Dirichlet allocation (LDA), as implemented in Mallet software (McCallum, 2002), using a hyperparameter optimization of every 10 iterations. We pre-processed the tweets in the following sequence: i) removal of French and German stop words, ii) removal of URLs, iii) removal of special characters (e.g., #, @) and punctuations, iv) division of concatenated expressions (e.g., StayAtHome become stay at home), v) removal of words shorter than 2 characters, and vi) removal of words with a direct reference to COVID-19 (e.g., covid, covid19, cov19, corona, coronavirus, pandemic).

Topic modelling combines document classification with the strong semantic unity of the discourse of a topic and of a document by optimizing the following equation: p(topic|document) *p(word|topic), for all given documents in a collection. Document classification can be expressed by the first part of the equation: p(topic|document). The task of document classification is to find the most likely class given the document (or the tweet). The second part of the above formula is the keyword generation probability: p(word|topic). It expresses that for a given topic, certain keywords are particularly likely. Documents and words are given, and topics are fitted iteratively. The user must set the number of topics that the algorithm will use. This fitting process ensures that the overall probability of the given documents and words is as high as possible. We calculated the best number of topics to extract (see Annexes 5.2.5 and 5.2.6) using the function

FindTopicsNumber() from the R package ldatuning (Nikita & Nikita, 2016). The four metrics composing the function were computed by training several LDA models with the number of topics ranging from 5 to 200. The results suggest that the optimal number of topics with respect to these metrics is 60 for French and 70 for German tweets. The number of topics was also determined via manual inspection of a variety of topic sets trained using several different numbers of topics.

Each topic was represented by a list of top related keywords, which then needed to be manually labelled to propose a possible interpretation. We reduced the possible topic labels to 12 categories that were found to encompass enough to summarize the content of the tweets. For instance, many different topics referred to vaccination – such as laboratories, number of vaccinated people, patents, etc., and can be summarized under a single generic "vaccine" label. We assigned each tweet the topic with the highest prevalence (highest gamma value).

Salient clusters of social media discussions and other users' reactions towards these clusters

We used correspondence analysis to reveal the associations among the identified topics found in Twitter discussions, the actor type, the different stages of the public debate, and the target population, relying on the FactoMineR package for R (Husson et al., 2020). Hierarchical clustering was subsequently applied to extract the main clusters of discussions from the correspondence analysis results and investigate what were other users' reactions towards salient clusters of discussions.

Correspondence analysis can be understood as principal component analysis for categorical data (Reveilhac & Morselli, 2020) and is also used to discover structure in textual data (D'Enza & Greenacre, 2012). The variables are projected on a factorial space such that the proximity between variables indicates a higher association. To this aim, we relied on a two-dimensional space to plot the variables to assess how closely they related to each other. Correspondence analysis calculates the contributions of each variable to the inertia of each factorial axis.

The projected variables are distinguished between active and supplementary variables. The active variables are used for the determination of the two-dimensional space, while the coordinates of the supplementary variables are predicted using only the information provided by the performed multiple correspondence analysis on active variables. As active variables, we explored how the content of tweets related to the different actor categories while also taking the topic and the temporal dimension into account. As a supplementary variable, we considered which target population (including children, women, adults, elderly individuals, and patients) was mentioned in the tweets. The identification was based on lists of search queries (see Annex 5.2.7).

With respect to the different stages of the public debate, we differentiate between the COVID-19 waves (W0: first international increase in the number of cases from January to March 2020; W1: first increase in Swiss cases from March to April 2020; W2: second COVID-19 wave in Switzerland from October to December 2020; W3: third COVID-19 wave in Switzerland from January to May 2021; W4: fourth COVID-19 wave from September to mid-November 2021; W5: fifth wave from mid-November to mid-December 2021) and the normalization periods (N1: decrease in the number of cases and relaxation period from the end of April to mid-June 2020; N2: end of the state of emergency; N4: public spaces partially reopen from mid-May to August 2021; N5: establishment of national “2G” (vaccinated or cured) rules since mi-December 2021). Annex 5.2.1 provides a detailed description for each stage (the modalities starting with a W indicate the different COVID-19 waves, and the modalities starting with a N indicate the normalization periods).

The results of the correspondence analysis were then used to perform a hierarchical cluster analysis with the Ward method to classify the tweets of the corpus into salient momentums and to investigate the reactions of other users. The best number of clusters was determined visually (see Annex 5.2.8). Each cluster is analysed according to Twitter features (including engagement in terms of likes and retweets, as well as negativity in replies).

Results

The original COVID-19-related tweets that were collected using a list of curated search-queries representing 16% of the total tweets emitted by major actors from December 2019. The government posted significantly more about non-COVID-19-related issues than health experts did. With respect to measures of popularity, experts’ tweets had the highest mean retweet rate, while the government triggered the highest mean like rate. Therefore, we observed that the main entities responsible for producing recommendations for handling the pandemic (FOPH and Taskforce) had the highest share of COVID-19-related tweets (more than 60%) compared to that of other actors. It is also interesting to note that, on average, the cantonal authorities tweeted more about COVID-19 than federal institutions. The cantons also triggered the most positive replies, thus suggesting that they were supported by the online audience. In general, the mean number of reactions (replies, retweets, and likes) was the highest for the main entities in charge of the COVID-19 communication, namely, the Federal Council and the FOPH, followed by political parties and politicians. The mean likes and retweets were also high for Taskforce actors.

To investigate how public trust in health experts and political authorities is comparable with social media trends during an epidemic, Table 5.2.1 displays the relationship between the average difference in public trust in political authorities measured between the first and the third waves of the MOSAïCH survey (y-axis) and the average differences in social media features covering the same period (x-axis). Based on Table 5.2.1, we can see that the majority of actors (except business industries, media, and research centres) suffered from a decline in public trust. Experts from the FOPH were especially affected by declining public confidence. When juxtaposed to the interactions on social media, we can see that there was an increased negativity in replies for experts and parliamentarians. However, there was a decline in the negativity in replies to the government and to the media, thus suggesting an opposing trend between survey trust and negativity in replies for these actors. Furthermore, the cantonal authorities also benefited from an increased positivity in the replies to their tweets. Table 5.2.1 also shows that the average number of engagements (in terms of likes and retweets) decreased for most actors, but especially for the government. However, there was an increased number of engagements towards parliamentarians and the media. In sum, the correlation between the survey and social media trends is generally low regarding the change in the negativity of other users' replies and with respect to engagement between the two waves. However, a closer look at each wave demonstrates a significant negative correlation between trust and negativity during the third wave.

Table 5.2.1: Relationship between the levels of public trust measured in surveys and the negativity in other users' replies and engagements (including likes and retweets).

	Wave 1: (April 30th to July 13th, 2020)			Wave 3: (March 19th to April 18th, 2021)			Difference between wave 3 and wave 1		
	Twitter		Survey	Twitter		Survey	Twitter		Survey
	negativity	engagement	trust	negativity	engagement	trust	negativity	engagement	trust
Business and industry	0.32	2.4	4.2	0.43	2.23	4.7	0.11	−0.17	0.5
Cantonal authorities	0.35	5.01	6.6	0.3	4.43	6	−0.05	−0.57	−0.6
Federal Office of Public Health	0.35	2.83	7.3	0.36	2.49	6.3	0.01	−0.35	−1
Federal Council	0.36	36.55	7.2	0.34	17.65	6.7	−0.02	−18.9	−0.5
News media	0.39	11.2	4.9	0.39	14.67	4.9	0	3.47	0
Parliament & elected politicians	0.35	10.73	6.3	0.37	13.88	5.9	0.02	3.15	−0.4
University research centres	0.38	6.5	7.6	0.25	5.51	7.6	−0.13	−0.99	0
Pearson correlations of Twitter features with surveyed trust (p-value)	**0.27 (0.26)**	**0.33 (0.56)**		**−0.87 (0.01)**	**0.09 (0.97)**		**0.28 (0.54)**	**0.21 (0.65)**	

Figure 5.2.3 displays the results from the correspondence analysis along a two-dimensional space. The vertical axis is essentially useful to account for the different phases of the public debate. For instance, the first stages are represented by groups in the upper quadrants, whereas the later stages are represented by groups in the lower quadrants. The horizontal axis groups the topical content and the actor groups. Tweets from the news media were not included in the correspondence map because of their essential broadcasting behaviour and because the news media do not represent public opinion.

We can thus investigate the major accents placed by political authorities on COVID-19 by showing what topics were prioritized in their communication on social media (see also Annex 5.2.9 for the topic distribution by actor). For instance, the upper right and left quadrants show the special role of the health experts and the research centres during the pandemic. The health experts heavily promoted policy solutions (e.g., testing), while research institutes focused on technologies and tracing applications.

Furthermore, the cantons are situated apart (see lower right quadrant) from the remaining political and business actors (see centre left position on the map) and are situated close to the vaccination and certificate policies. We also see that the cantons had an increasingly important role in collecting and sharing information about new cases over the pandemic. In contrast, the government was especially active at the beginning of the pandemic and tried to inspire more confidence in COVID-19 policy measures (e.g., masks and quarantine) by accounting for the national policy while also pointing to economic difficulties and responses. The business and industry branches especially focused on the economic and international aspects of the pandemic.

Figure 5.2.3 also enables us to assess how the major actors' communication about COVID-19 is associated with specific target populations. For instance, the cantons had a clear focus on elderly individuals and patients. Furthermore, the political and business actors focused more on women and children. References to children and to the medical staff were not significantly associated with communication.

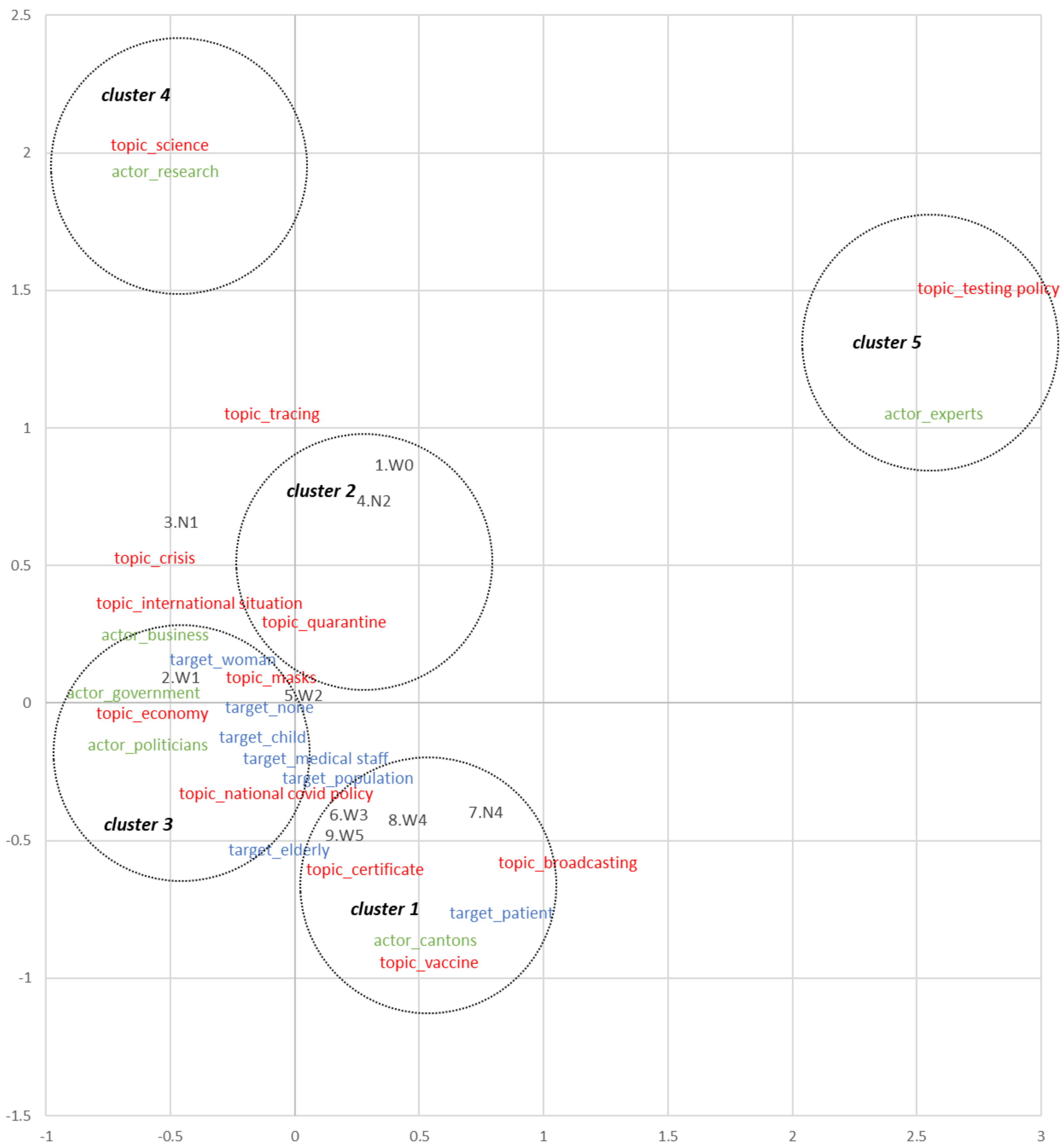


Figure 5.2.3: Correspondence analysis including debate stage (W indicate Covid-19 waves and N normalization periods in black), topical content (in red), actor groups (in green), and target population (in blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 5.2.2 provides the results from the hierarchical clustering based on the previous correspondence analysis. Each cluster is characterized by tweets related to the topic content, actor groups, stages of the public debate, and target populations. In comparison to Figure 5.2.3, additional important information provided in Annex 5.2.9 relates to the measures of Twitter features (including other users' engagement and proportion of negativity in replies) that relate to each cluster. Table 5.2.2 describes each of the seven identified clusters.

Table 5.2.2 shows that cluster 1 is characteristic of the third wave of the pandemic. The cantonal authorities and politicians were particularly engaged in discussions about the national COVID-19 policy, where face masks remain an essential complementary measure to the vaccination. The level of other users' engagement is one of the lowest across the clusters (both in terms of mean and median), and the proportion of negative replies is among the lowest. However, this cluster also encompassed tweets against the national policy. For instance, the tweet that triggered the highest number of engagements was writer by a politician voicing against the enlargement of Covid-19 certificate.

Cluster 2 overlaps the third wave of the pandemic and the following normalization period. The communication of health experts (FOPH and Taskforce) is especially salient in this cluster, with a focus on testing policy and COVID-19 case broadcasting. The median number of engagements is the second highest, thus showing a high level of other users' interest and potential for supportive replies. For instance, the tweet from the group of health experts with the highest number of engagements was written by the FOPH and thanks the former head of the federal office for his devotion and years of services. However, this cluster has one of the highest proportions of negative replies, which suggests that other users were in general less supportive of the experts' communication. For instance, the tweet with the highest number of replies and the highest negativity provides an update on the quarantine obligation and the list of countries with worrisome virus variants.

Cluster 3 centres on the first normalization period of the pandemic as well as the subsequent COVID-19 waves. It especially underlines the implication of the politicians who discuss economic issues and the national COVID-19 policy. This cluster is characterized by a high share of negativity in replies and has the highest standard deviation of engagement. The tweet with the highest number of engagements was written by a politician contesting the efficiency of lockdown measures.

Cluster 4 characterizes the first wave of the pandemic and the following normalization period where researchers focus on scientific solutions to the crisis. The level of other users' engagement is the lowest across the clusters (both in terms of mean and median), and the proportion of negative replies is among the lowest. There is thus less interest from the public in this cluster than in the other clusters. For instance, the tweet written by a research centre that triggered the highest number of engagements promotes the elaboration of a tool to track COVID-19 cases.

In Cluster 5, the communication of health experts is the most salient and focuses on the testing policy. Tweets from this cluster essentially relate to the second normalization period as well as the third COVID-19 wave. This cluster triggers the highest median number of engagements from other users. The tweet triggering the highest number of engagements was written by the FOPH and announces the lifting of certain restrictions from July 2020. This tweet receives 111 replies, of which 35% are negatively loaded.

With respect to the target populations, which serve as a supplementary variable, we observe that the mentions of patients and elderly individuals are associated with the vaccination and certificate topics (see cluster 1). This is in line with the fact that the vaccination campaign started with older and more vulnerable sections of society. However, the communication of political and health authorities on Twitter is little focused on target populations, as the majority of tweets do not mention any target population.

Table 5.2.2: Description of the clusters according to actor groups, debate stage, topical content, and target populations.

	cluster1	**cluster2**	**cluster3**	**cluster4**	**cluster5**
engagements - median	6.00	26.00	8.00	6.00	58.00
engagements - mean (sd)	27.35 (90.09)	53.63 (119.11)	37.47 (139.38)	13.8 (53.4)	71.29 (50.27)
engagements - [min. - max.]	0–2465	0–2564	0–3740	0–1695	7–509
% negativity in replies	0.32	0.36	0.36	0.32	0.35
replies – mean (sd)	5.57 (20.41)	12.00 (22.49)	4.90 (18.33)	0.70 (3.37)	30.16 (29.44)
Stages of public debate					
1.W0	0.00	0.04	0.01	0.02	0.00
2.W1	0.10	0.08	0.17	**0.21**	0.04
3.N1	0.02	0.06	**0.21**	**0.28**	0.12
4.N2	0.02	0.11	0.16	0.19	**0.22**
5.W2	0.13	0.15	0.15	0.11	0.13
6.W3	**0.34**	**0.26**	0.16	0.09	0.19
7.N4	0.19	0.16	0.06	0.05	0.16
8.W4	0.16	0.11	0.05	0.04	0.11
9.W5	0.04	0.02	0.01	0.01	0.02
Actor groups					
business	0.04	0.03	0.16	0.11	0.00
cantonal authorities	**0.59**	0.05	0.11	0.00	0.00
experts	0.00	**0.76**	0.00	0.00	**1.00**
government	0.06	0.04	0.14	0.01	0.00
politicians	**0.30**	0.11	**0.55**	0.07	0.00
research	0.00	0.01	0.05	**0.81**	0.00
Topical content					
broadcasting	**0.24**	**0.31**	0.03	0.00	0.00
certificate	0.04	0.02	0.01	0.00	0.00
crisis	0.01	0.02	0.13	0.14	0.00
economy	0.04	0.01	**0.28**	0.03	0.00
international situation	0.00	0.01	0.04	0.02	0.00
masks policy	0.03	0.02	0.09	0.06	0.00
national covid policy	**0.29**	0.05	**0.26**	0.07	0.00
quarantine policy	0.00	0.04	0.05	0.01	0.00
science	0.00	0.06	0.05	**0.65**	0.00
testing policy	0.00	**0.24**	0.00	0.00	**1.00**
tracing	0.00	0.03	0.03	0.03	0.00
vaccine	**0.35**	0.19	0.04	0.00	0.00
Target population					
child	0.04	0.04	0.04	0.06	0.00
elderly	0.02	0.01	0.01	0.00	0.00
medical staff	0.02	0.01	0.01	0.02	0.00
patient	0.05	0.02	0.01	0.01	0.00
population	0.03	0.02	0.02	0.02	0.00
woman	0.01	0.01	0.01	0.01	0.00
none	**0.84**	**0.89**	**0.91**	**0.88**	**1.00**
Number of tweets	8898	1687	9407	1559	989

Discussion and implications

The objective of this study was to compare offline and online trends to provide empirical evidence regarding how trust in health experts and political authorities has evolved throughout the different phases of the COVID-19 pandemic. This study contributes to understanding these processes by drawing on the Swiss context.

Our data collection shows interesting patterns. For instance, COVID-19 tweets emitted by actors replicated the COVID-19 case curve, including a lower level of Twitter activity between waves of the pandemic. This finding is in line with the results obtained by Pang et al. (2021), which showed that government engagement on social media was relatively low at the beginning of the pandemic and then surged in the acute stages, with a trend towards a decrease in engagement during the chronic stages. Federal institutions tweeted less about COVID-19 than health experts did due to the much broader range of topics that political authorities have to deal with. As such, political authorities tweeted significantly more about non-COVID-19 topics than health experts did. Overall, the COVID-19 tweets of experts triggered more engagement and replies from other users than those of authority figures (e.g., Federal Council executives). This is in line with findings that experts are more likely to be listened to than political authorities (Drescher et al., 2021).

Our first research question asked whether the level of public trust measured through opinion surveys is reflected in other users' engagement and negativity in replies. Comparing negativity in replies with trends in public trust from opinion surveys shows a similar decrease in public support for experts. Furthermore, the general decline in the mean rate of engagement from other users suggests that while the initial COVID-19 outbreak was characterized by increased trust in scientists and health authority experts, there was weakened trust in public health authorities as exposure to the epidemic became prolonged (Battiston et al., 2021). Overall, we found little congruence between the survey measure of trust and social media trends in terms of engagement and replies. This lack of congruence might suggest that people express more dissent on the internet than in surveys because either surveys are biased by desirability effects or the reply feature in Twitter is used mostly by discontented people, as tweets are unsolicited reactions. Either way, the findings reveal a complementarity need between the two data sources. Our second research question focused on the salient associations between the topics, the actors, and the different stages of the online COVID-19-related discussions. We found that health experts, research centres, and cantonal authorities contributed greatly to the formation of the correspondence space. Furthermore, these actors focused on distinct topics and target audiences. For instance, the cantonal authorities spent an essential part of their communication on broadcasting the COVID-19 case and promoting the vaccination (as well as the certificate). Thus, patients and elderly individuals constituted their target audiences. Health experts were oriented towards the promotion of the testing policy, while the research centres focused on technological innovations and tracing applications.

Our third research question asked how the extracted clusters were received by a broader audience. We confirmed that the cantons played an important role in the management of the pandemic, notably due to federalist and subsidiarity principles (e.g., by being in charge of broadcasting and the application of COVID-19 policy measures at the local level). The fact that the cluster characterizing cantonal communication triggered one of the lowest levels of negativity suggests that cantonal authorities were able to build a good online followership and reputation for managing the crisis. The close reading of emblematic tweets from the clusters in terms of engagement and negativity also enabled us to highlight the variety in other user supportive and contesting behaviours towards the major actors during the pandemic. As in other European countries and the rest of the world, we found evidence that Switzerland is experiencing a mobilization against COVID-19 policy measures.

Theoretical contributions

From a theoretical perspective, the results of this study provide helpful conclusions about the communication between government authorities and the population in (health) crisis situations. Most notably, it reveals that the type of actors, the type of content, and the stages of public debate have an impact on other users' reactions to actors' messages. This is particularly important because actors' tweets can reach a large audience, potentially helping to raise public awareness and public acceptance of policy measures whose reception has been, to date, measured mostly through surveys. Relying on social media data enables us to access unsolicited behavioural and rhetorical measures that can be compared to survey results. Our findings resonate with the study of Gilardi et al. (2021), which states that social media challenge the capacity of political and media elites to craft a consensus regarding the appropriateness of different measures as responses to a major crisis. Indeed, actors seem to have a limited capacity to influence broader audience opinion. In this study, we show a low correspondence between social media and survey data sources, but it may well be that in other countries, we could have observed more contesting behaviours on social media (e.g., Haupt et al., 2021). The methodology used in this paper could be applied from a cross-country perspective. We found some evidence of public fatigue in Twitter features. We therefore encourage future studies to link online reactions to offline attitudes to account for the role played by social media in the public debate and to assess whether online and offline opinions are congruent.

Implications for practice

The results of this study have implications for governments, health organizations and experts, the media, and researchers in selecting suitable communication strategies that may foster the active liking and retweeting of messages on social media. For instance, we found a general decrease in the number of engagements from other users to tweets from health and political experts. This might be due to the fatigue effect in the public, which, in turn, might increase public concern about the legitimacy of COVID-19 policy measures. A better understanding of the communication and content dynamics among authorities and (online) public debate is thus pivotal to ensure the well-functioning of democratic institutions. Furthermore, we show that tweets that are clearly linked to a policy issue tend to trigger more engagement (in terms of likes and replies) from other users. This suggests that actors should adopt a communication strategy that promotes and discusses clear policy recommendations and measures instead of adopting a broadcasting behaviour (e.g., tweeting about the number of COVID-19 cases). Moreover, there is an incentive for actors to make use of hashtags (instead of mentions or links) to generate public attention and approval.

Conclusion and outlook

The findings obtained in this study enable us to improve our understanding of how the types of actors emitting messages, the variety of COVID-19-related topics, and the stages of the public debate all affect the reception of the messages. Social media play an important role because they allow actors to bypass institutional gatekeepers – such as political parties and newspapers – with a view to achieving public compliance with the promoted policy measures and motivating citizens to adopt preventative measures.

Despite signs of rising fatigue characterizing the later stages of public debate, our results indicate that actors' efforts to communicate on social media are generally well received by the online audience. This overall positive picture reflects public support for governmental authorities, as demonstrated during the two popular votes about the modification of COVID-19 measures. The first law was supported by 60% of voters on June 13, and the second saw an increased share of public support, with 62% of voters on November 28.

Our study contributes to political communication research in times of crisis by investigating how actors' online messages resonate with and are received by the wider audience. To the extent that elite communication is crucial for compliance with policy measures, the findings suggest that social media may hamper success in achieving COVID-19 responses in the later stages of public debate, as there seems to be increasing fatigue among the public. It is difficult to provide officials with a clear pathway to communicate their crisis response through social media. However, we can formulate the following recommendations: messages should be oriented towards specific policy issue measures instead of merely broadcasting statistics, messages should make use of content features such as hashtags, and cantonal authorities should continue to play a decisive role in crisis communication.

Building on our methodology, future research could adopt a cross-country perspective to assess the extent to which our conclusions are generalizable to the context of other countries. Furthermore, other popularity measures could be used to assess actors' reputation, such as the evolution of a network of followers. Finally, Twitter is a particular social media platform that encompasses a specific audience (and is perhaps more elitist than Facebook), and it is possible that this could have impacted our results, as the actors may have been depicted more positively on Twitter than on other platforms.

5.3 Assessing how Attitudes to Migration in Social Media Complement Public Attitudes Found in Opinion Surveys[36]

Introduction

Migration has been dominating media and political discourse worldwide, especially with respect to the European refugee crisis since 2011 and Trump's 'build the wall' campaign in 2016. Previous studies have mapped migration discourses in traditional media (Vliegenthart & Boomgaarden, 2007), conventional channels of party communication (Charteris-Black, 2006), and politicians' social media accounts (Heidenreich et al., 2019; Combei et al., 2020). However, studies comparing different countries and different data sources on migration remain scarce (Eberl et al., 2018).

The proposed study investigates the extent to which migration discourses on social media can provide a complementary understanding of attitudes to migration in traditional opinion surveys. It also investigates what factors explain the prevalence of migration discussions on social media, especially in relation to societal factors (for example, migration integration policies and elite polarisation on the topic), public attitudes (such as acceptance of migration and migrants), and framings of migrants (for instance, generic framing of policy issues and specific depiction of migrants). Regarding framing, we draw from the definition of Entman (2007), who suggests that framing is inherently part of communication and implies choosing "a few elements of perceived reality and assembling a narrative that highlights connections among them to promote a particular interpretation" (p. 164), as well as from Matthes and Kohring (2008) in our choice of a quantitative approach. We distinguish between generic frames, which offer a systematic platform for comparison across frames, and issue-specific frames, which allow for "great specificity and detail" (de Vreese, Peter & Semetko, 2001, p.108). In other words, whereas issue-specific frames emphasise unique ways to contextualise a topic (for example, migrants as victims or criminals), generic frames promote a particular discourse (for

[36] This chapter is a slightly adapted version of the book chapter as M. Reveilhac and G. Schneider (2022): "Assessing how Attitudes to Migration in Social Media Complement Public Attitudes Found in Opinion Surveys", Swiss Papers in English Language and Literature, 41(41), 119-153.

instance economic or cultural) that has obvious relevance to a bigger set of topics within which the unique topic (e.g. migration) is located (Brüggemann & D'Angelo, 2018).

To date, opinion surveys remain the main way to assess public attention and attitudes towards migration and its different societal dimensions. However, social media also offer opportunities for spontaneous discussions of these topics, without the intervention of pre-defined survey interests. Although Twitter users are not representative of national populations (Ceron, 2017), they form a politically interested audience whose voices about migration are likely to interact with the current public debate . Given the important impact of social media on political views and outcomes (Zhuravskaya, Petrova & Enikolopov, 2020, p.429), it is critical to examine whether social media serve as a reflection of or a substitute to broader public attitudes. More specifically, previous studies have demonstrated the potential influence of anti-immigration social media groups in shaping broader audiences' migration-related attitudes (Törnberg & Wahlström, 2019). However, beside these most active user groups, we still know little about what influences the general salience and sentiment found in online discussions about migration.

Our study analyses migration discourses on the Twitter accounts of followers of major political accounts across 5 English-speaking countries (United States, Britain, Ireland, New Zealand and Australia) and compares the obtained results to attitudes found in opinion surveys. In a first step, overall comparisons are conducted, both in terms of issue salience and tonality. For instance, we display insights into the extent to which the salience of online migration discussions on social media is congruent with the perceived importance of migration measured in surveys in the selected countries. Furthermore, we correlate the online sentiment about migration with a sample of relevant attitudinal dimensions inspired from survey research. In a second step, we strive to disentangle the impacts of societal factors and public opinion on the salience of online messages about migration. To address these research interests, we complement several data sources to extract migration-related opinions, most notably social media messages and opinion survey responses. It is important to note that, compared to other social media, such as Facebook, Twitter is known to be dominated by elite actors (e.g. politicians, journalists, opinion leaders, etc.) who have strategic goals in terms of persuasion or opinion making (Freelon and Karpf, 2015; Rauchfleisch, Vogler & Eisenegger, 2021). This is important from the perspective of public opinion, especially because these actors have a prominent position on Twitter (e.g. more followers, more reach, circulation, etc.) and, thereby, a high potential to influence public opinion formation (Barberá & Steinert-Threlkeld, 2020). Another contribution of the present study is to build a bridge between the fields of linguistics – using computational linguistic methods – and cultural studies. Michel et al. (2010) argue that their approach of analysing correlations between lexical frequency and time (e.g. frequency peaks), opens up an entirely new field of research, which they call culturomics.

"[T]his approach can provide insights about fields as diverse as lexicography, the evolution of grammar, collective memory, the adoption of technology, the pursuit of fame, censorship, and historical epidemiology. Culturomics extends the boundaries of rigorous quantitative inquiry to a wide array of new phenomena spanning the social sciences and the humanities". (p. 1)

We use more advanced computational approaches (supervised classification), which allow us to focus on relevant (generic and specific) framings of migration in a cross-country perspective. Importantly, we also assess how different tweet collection strategies impact similarities between social media and survey attitudinal distributions. The proposed methodology – based on automated content analyses and the linking between social media and surveys – can be extended to other countries and to other research arenas where comparison between data sources is valuable to provide a more nuanced view of a phenomenon.

Study Background

Salience of Migration and Attitudes Towards It

There are notable global surveys that include questions on immigration and immigrants, such as the Gallup World Poll, the International Social Survey Program, the World Values Survey, and the Ipsos Global Trends Survey. All cover a large cross-section of countries and contain multiple waves in which the same general questions are asked to respondents. Specific question items also serve to build global trend indicators, such as Gallup's Migrant Acceptance Index. Surveys of public attitudes toward immigration have shown that the salience of immigration as an issue has varied wildly over time (Dempster, Leach & Hargrave, 2020, p.25). In particular, the salience of immigration has risen in Europe over the last decade. Regardless of the salience, it is also notable that attitudes toward immigration actually improved in most European countries (Gonzalez-Barrera & Connor, 2019).

Beyond survey research, other studies found that public attitudes on immigration are increasingly expressed online, especially in the discursive construction of immigrants and refugees (Ekman, 2019, p.606). Yet, compared to nationally representative samples of respondents, social media users are usually unrepresentative of national populations (Ceron, 2017). Furthermore, social media platforms are likely to be polarising spaces (Krasodomski-Jones, 2016), thereby, starkly contrasting with the calibrated setting of opinion surveys. For these reasons, the online debate on immigration does not necessarily reflect public opinion but rather creates a space which amplifies the strongest views (Rutter & Carter, 2008, p.35). For instance, posts that are no longer socially acceptable in a face-to-face conversation and that contain prejudiced and hateful comments on immigration can reach a wide audience through social media (Rutter & Carter, 2018, p.165). As a result, the connection between social media messages and public opinion measures on the migration debate remains generally hard to disentangle.

On the one hand, it is complicated to evaluate the impact of social media coverage of immigration on how the broader public views immigrants and immigration. This is notably due to the fact that it is difficult to discern whether people learn their political views from social media pages (or threads), or whether they choose to consult social media pages that reflect their existing political views.

On the other hand, it is also unclear how public opinion and contextual factors affect the salience of immigration debates on social media. This is because it generally remains unclear whether social media serve to amplify or substitute public opinion (see similar discussion about elite communication by Castanho Silva & Proksch, 2021) and these platforms may have a similar amplification effect as news media (Gilardi et al., 2022, p.42). According to the substitution logic, social media would just serve as another channel for people to express similar attitudes as during face-to-face interactions. Aggregated patterns of social media discussions should thus reflect similar trends found in surveys, despite the non-representativity of social media users. With respect to the amplifier logic, social media present tools for more personalised, and perhaps also more polarised, messages which may not be expressed in other arenas, thus circumventing the mainstream debate. Therefore, aggregated patterns of social media discussions should display quite a different distribution than opinion surveys. For instance, we could expect that, in addition to be highly polarised, social media messages would be slanted towards negative sentiments. However, studies show mixed results. For instance, Heidenreich et al. (2019) did not find support for the assumption that right-leaning parties talk more, and more negatively, about migration. However, Hameleers (2019) demonstrates that social media platforms offer opportunities for ordinary citizens to generate populist discourse to predominately target elites and marginalized groups in society.

Despite their inherent unrepresentativeness, social media data can provide statistics to make informed policy and programme decisions (Japec et al., 2015, p.846). The topic of migration is no exception here. Drawing from these premises, research has been undertaken to better understand whether social media data can produce distributions of attitudes and salience similar to those from survey data. Concerning attitudes, Amaya et al. (2020, p.173) take a critical view: with respect to the salience of discussions, the broad correlation between frequency and opinion is generally accepted. Roberts and Wanta (2012), for instance, investigate the correlation between media coverage and private electronic conversations. Ghanem (1997) states that a strong correlation has been recognised, and that salience may be the best predictor:

“Agenda-setting studies have focused on how frequently an issue is mentioned in the media. The frequency with which a topic is mentioned probably has a more powerful influence than any particular framing mechanism” (p. 12)

In this article, we aim to better understand the congruence between surveys and social media messages on the sentiment and salience of immigration. We therefore raise two overarching research hypotheses. First, we hypothesise that the salience of migration online correlates with the extent to which migration is perceived as an important concern in representative opinion surveys. Second, we hypothesise that the tonality related to migration online correlates with the overall satisfaction toward migration found in representative opinion surveys. We answer these two hypotheses relying on correlations comparing salience and support towards migration between different groups of Twitter users and responses from survey respondents.

Impact of Contextual and Political Factors on the Salience of Migration-Related Tweets

In connection to real-world events, several factors can explain variation in the salience of social media messages referring to migration. For instance, salience can be influenced by contextual factors, such as the type of institutional response to migration related issues (e.g. integration policies). It can also be linked to political factors, such as the degree of party polarisation with respect to the topic of migration.

Overall, surveys demonstrate that salience increases when immigration is perceived as problematic and decreases when it is perceived as being under control (see Blinder & Richards, 2020). As such, the institutional capacities to deal with migration related issues can decisively impact the salience of migration debates. For instance, national and local governments are responsible for integration policies which help facilitate immigrants becoming part of the host country (such as through schools, workplaces, and communities). The Migrant Integration Policy Index (https://www.mipex.eu/) is a tool dedicated to account for policies undertaken to integrate migrants in host countries.

Demonstrating the connection between political rhetoric and public attitudes to migration is a more complicated task. However, whereas the ability of politicians to directly influence attitudes through their rhetoric is unclear, the political rhetoric has a clearer influence over the salience of an issue (Hatton, 2017, p.19). For instance, the anti-immigration rhetoric has the potential to make attitudes towards immigration more consequential for voting behaviour (Rooduijn, 2020).

In this article, we aim to better understand how institutional settings (namely, the institutional responsiveness to migration) and the degree of elite polarisation impact the salience of social media messages about migration. We therefore add two further overarching research hypotheses. Our third hypothesis states that the salience of tweets related to migration is more pronounced when societal and political factors (migrant integration policy and elite polarisation) are unfavourable to migrants and immigration. Our fourth hypothesis suggests that the salience of tweets related to migration is correlated with lower levels of public acceptance of migration. We aim to address these hypotheses relying on multivariate regression.

Impact of Specific and Generic Frames on the Salience of Migration-Related Tweets

Immigration debates have been increasingly marked by a rhetoric of emergency and threat (for example, calls for stricter policing of borders and the limitation of mobility). From a survey perspective, Dempster, Leach and Hargrave (2020) have noticed important implications for the interpretation of opinion data on migration. In particular, the authors note inherent contradictions in the messages received from these surveys as people can seemingly hold two opposite views. For instance, within the same survey, people can hold the opinion that immigrants both take jobs and create jobs (Dempster, Leach & Hargrave 14). Duffy (2019, p.207) suggests this inherent contradiction may be due to the framing of the question, or to the level at which respondents prioritise the impacts of migration (namely, locally, nationally, or internationally).

Despite being 'gold-standard' for measuring public opinion, surveys can also test a limited and pre-defined set of dimensions, and are vulnerable to changes in methodologies and timing (Crawley, 2005). Compared to social media messages which explicitly refer to the perceived important (or problematic) aspect of migration, it is often difficult to know what respondents are thinking about when they answer a survey question. In surveys, the wording of a question is of utmost importance as it should be unambiguous and unequivocal.

That said, survey data are a valuable barometer of public attitudes, especially when consistent over time and between waves. Survey data are also useful for calibration purposes with other types of opinion data, such as social media messages, especially when comparing different framings of migration. For instance, surveys have particularly focused on the impacts of migrants and migration, typically assessed in terms of economic, social, and cultural burdens for the country. Yet, it is unknown how these more or less positive assessments of migrants impact the salience of social media discussions about migration.

This generic framing of migration is usually complemented by more specific narratives about migrants in public discussions. For instance, there is some evidence that people adopt elite rhetoric to a certain degree, either negatively (Doherty, 2015, p.57) or positively (Crawley & McMahon, 2016, p.13). For instance, the anti-immigration rhetoric is at the core of far-right populism (Schwartz et al., 2020), immigrant movements being described as invasions and narratives drawing on the concerns that people may perceive refugees and migrants as a challenge to values and culture, a source of terror and crime, and a threat to living standards, jobs, and public services (ODI & Chatham House, 2017, p.1), which form main frames against which the impact of migration are assessed. Social media have been shown to play a decisive role in the spread of the populist rhetoric, notably anti-immigration (Ernst et al., 2019, p.18), and also in the depiction of migrants as a threat (Lorenzetti, 2020, p.87).

Several studies have investigated the specific depiction of migrants and asylum seekers (see Milioni & Spyridou, 2015; Van Gorp, 2005). More recently, O'Regan and Riordan (2018) relied on a combination of methods in corpus linguistics and critical discourse analysis to explore the representation of refugees, asylum seekers, immigrants and migrants. This research has been essentially applied to the study of news articles, and more rarely to social media messages (de Rosa et al., 2021). A common finding of these studies is that migrants and asylum seekers are mainly described as 'innocent victims' or 'intruders.' Furthermore, while asylum seekers can generate empathy, this is less the case for migrants who are also perceived as 'profiteers.' The depiction of migrants in negative or positive terms has implications on countries' choices to develop exclusion or inclusion policies.

In this article, we aim to better understand what framings of migrants and migration impact the salience of social media messages. We therefore add one last overarching research hypothesis that the salience of tweets related to migration is positively associated with discussions about migrants and migration using a threat related rhetoric. To test this hypothesis we conduct multivariate regression, but we also rely on close-reading of a sample of tweets, as well as on the interpretation of important words related to the generic frames.

Data and Methods

Twitter Data Collection

We collected two samples of Twitter users relying on a similar data collection strategy. These two samples differ in the choice of 'seed accounts' from which followers are extracted. For the selected followers, we then collected the last 3'200 tweets (which corresponds to the authorised limit by the Twitter API). We identified tweets related to migration based on the following list of search queries: ".*migration.* | migrant.* | immigrant.* | emigrant.* | foreigner.* | asyl.* | refugee.* | undocumented worker* | guest worker* | foreign worker* | freedom of movement | free movement". We then retrieved the followers of these seed accounts (max. 75'000 followers for each of the seed accounts authorised by the Twitter API) and applied filters to keep the most relevant Twitter accounts[37]. For each sample of Twitter followers, we decided to take random samples of 100'000 followers to keep the tweet collection stage reasonable in time and size. We also decided to include only tweets emitted after January 2019 in our final dataset of tweets about migration. The main reason for this is that we wanted to equilibrate the tweets of users with different dates of account creation and tweeting frequency as much as possible. Concerning the first sample, we identified central media and party accounts for each country of interest (United States, United Kingdom, Ireland, Australia, and New Zealand).). Because we rely on these seed accounts, we expect that a majority of users in the "random sample of twitter users" and also in the "interested sample of users" come from the same countries as the seed politicians and media. This data collection strategy enables us to tackle the fact that only a minority of users provide geolocating information. The distribution of this random sample of followers is given by country in Table 5.3.1. The size of the final dataset contains 310,247 tweets from 27,649 unique users. Overall, between 25% to 39% of users tweeted about migration. The overall tweeting frequency about migration in our sample has a mean of 11 and a standard deviation of 34 (with a maximum of 1 and a maximum of 2382).

Concerning the second sample, we identified central politicians' accounts for each country. The distribution of this politically interested sample of followers is given by country in Table 5.3.1. To identify the relevant politicians' accounts, we relied on the *Twitter Parliamentarian Database* (van Vliet et al., 2020)[38]. We selected the politicians who were active in parliament from the year 2019 onward. The size of the final dataset contains 347,003 tweets from 28,966 unique users. Overall, between 18% and 37% of users tweeted about migration. The overall tweeting frequency about migration in our sample has a mean of 14 and a standard deviation of 43 (with a minimum of 1 and a maximum of 2941).

---

[37] We apply some filters to keep only relevant users in our sample of followers. Notably, we apply the following filters: the user has to provide a minimal description in the user profile field, the user account must have been created before 2020-01-01, the number of emitted tweets must be 'reasonable' (above 5 per year and below 10'000 per year), and the main language of the account must be English.

[38] The data can be found here: https://figshare.com/articles/dataset/The_Twitter_Parliamentarian_Database/10120685

We also give a description of the Twitter sample of the 1,951 identified politicians in Table 5.3.1. Among the entire sample of politicians, 966 (50%) tweeted about migration. This left us with a total of 40,455 emitted tweets. The overall tweeting frequency about migration in our sample has a mean of 28 and a standard deviation of 50 (with a minimum of 1 and a maximum of 457).

Table 5.3.1: Description of the samples of Twitter followers and politicians

Random sample of Twitter users	Selected followers	Sample of 100'000 followers	Followers tweeting about migration	Number of tweets
US	98363	14324	5303 (37%)	51219 (ratio: 9.7)
UK	161458	23508	9364 (39%)	106110 (ratio: 11.3)
Ireland	96406	14189	4263 (30%)	45707 (ratio: 10.8)
Australia	174618	25343	6647 (26%)	85556 (ratio: 12.9)
New Zealand	57740	8179	2072 (25%)	21655 (ratio: 10.4)
	588585	**85543**	**27649**	**310247**

Interested Twitter users	Selected followers	Sample of 100'000 followers	Followers tweeting about migration	Number of tweets
US	12058554	20000	7481 (37%)	88976 (ratio: 14.4)
UK	8515174	20000	7361 (37%)	94535 (ratio: 14.8)
Ireland	1002356	20000	4943 (25%)	53038 (ratio: 12.0)
Australia	1125590	20000	5614 (28%)	83312 (ratio: 16.9)
New Zealand	520743	20000	3567 (18%)	27142 (ratio: 8.4)
	23222417	**100000**	**28966**	**347003**

Politicians	Selected politicians	-	Politicians tweeting about migration	Number of tweets
US	873		311 (36%)	30572 (ratio: 98.3)
UK	590		454 (77%)	7291 (ratio: 16.1)
Ireland	150		87 (58%)	825 (ratio: 9.5)
Australia	134		77 (57%)	1057 (ratio: 13.8)
New Zealand	204		37 (18%)	710 (ratio:19.2)
	1951		**966**	**40455**

Survey Data from Representative National Samples of the Population

We test our hypotheses 1 and 2 by relying on the comparison between the collected tweets and measurements from opinion surveys. We present the comparisons between Twitter and survey data using visualisations in the form of scatter plots.

Hypothesis 1 centres on the salience of migration. On Twitter, we measure salience as the proportion of sent tweets related to migration by country. In surveys, we rely on the 'most important concern' question item, which asks respondents to mention what they perceive as the most important policy issue facing the country. We use data from the 2019 Eurobarometer for the United Kingdom and Ireland, from the 2021 Survey of US adults for the United States, and from the 2019 Roy Morgan survey for Australia and New Zealand. To measure salience, we rely on the proportion of respondents mentioning migration as the most important concern.

Hypothesis 2 focuses on the sentiment towards migration. On Twitter, we measure sentiment using the sentimentr R package (Rinker, 2021) which calculates text polarity sentiment in the English language at the sentence level. The number can take positive or negative values and expresses the polarity of the sentiment. We recode the scale into positive (values above 0), negative (values below 0), and neutral (value of 0). In surveys, we rely on the combination of question items asking respondents to assess the impacts of migration on cultural, social and economic dimensions. Data from the 2019 European Social Survey[39] are used for the United Kingdom and Ireland, while data from the 2019 World Values Survey[40] are used for Australia, New Zealand and the United States. To make the survey items most comparable and to account for the degree of positivity toward migration, we recode the variable scales into three categories ("agree", "disagree", and "neutral") and we sum up the proportion of respondents rating the impact of migration positively on the three mentioned dimensions (cultural, social, and economical). When comparing Twitter sentiments with a set of variables from surveys, possible

[39] The question items are: 'Country's cultural life undermined or enriched by immigrants' (answer scale from 0 to 10, where 0 states that the cultural life is undermined and 10 states that the cultural life is enriched), 'Immigrants make country worse or better place to live' (answer scale from 0 to 10, where 0 states 'worse place to live' and 10 states 'better place to live'), 'Immigration bad or good for country's economy' (answer scale from 0 to 10, where 0 states 'bad for the economy' and 10 states 'good for the economy').

[40] The question items are: 'Immigration in your country: Strengthens cultural diversity' (answer scale from 1 to 3, where 1 states 'disagree' and 3 states 'agree'), 'Immigration in your country: Leads to social conflict' (answer scale from 1 to 3, where 1 states 'disagree' and 3 states 'agree'), 'Impact of immigrants on the development of the country' (answer scale from 1 to 5, where 1 states 'rather bad' and 5 states 'very good').

methodological problems can arise in terms of data comparability. For instance, questions like "[immigration] strengthens cultural diversity" and "[immigration] leads to social conflict" measure different aspects that do not necessary link to sentiment. Namely, one can be generally positive (sentiment) to immigration (for humanitarian reasons) but still agree that it "leads to social conflict". As social media messages also contain many dimensions about the immigration issue, a possible strategy is to compare averaged survey items covering several dimensions about the same topic with the mean sentiment derived from social media texts. Doing so, we test the correlation between sentiment toward migration in surveys and on social media.

Statistical Model Specifications

We test our hypotheses 3 to 5 using linear regression modelling. The dependent variable is the logged number of tweets mentioning migration for each Twitter follower. According to our hypotheses, this salience of migration at the user level can be explained by several independent variables.

To test hypothesis 3, we include contextual factors, namely an integration policy index (MIPEX: https://www.mipex.eu/) and a measure of elite polarisation. The MIPEX summarises policy indicators to create a multi-dimensional picture of migrants' opportunities to participate in society. Lower values indicate more restrictive policies whereas higher values indicate more integrative policies. The measure of political polarisation is based on the expert coding of the positiveness toward migration for the political parties within each of our selected countries. The coding is done by the experts from Manifesto Project (https://manifesto-project.wzb.eu/). For each country, we calculated the level of polarisation by taking the absolute difference between the higher and the lower party value for viewing immigration as positively impacting the national way of life[41].

To test our hypothesis 4, we include a public opinion measure of migration acceptance, the Gallup's Migrant Acceptance Index. The index is based on three questions that Gallup asked in 138 countries in 2016 and in the U.S. and Canada in 2017. The index is a sum of

[41] For more information see the codebook: https://manifesto-project.wzb.eu/down/data/2021a/codebooks/codebook_MPDataset_MPDS2021a.pdf

the points across three questions: whether people think migrants living in their country, becoming their neighbours and marrying into their families are good things or bad things. It has a maximum possible score of 9.0 (all three are good things) and a minimum possible score of zero (all three are bad things).

To test our hypothesis 5, we include general and specific framings of migration on Twitter. To classify the tweets along general policy issues, we build a classifier to assign tweets among the following categories: civil rights, culture & identity, economy, foreign policy, law & order, and welfare. These categories have been determined theoretically and inspired from survey research. To extract a sample of emblematic tweets corresponding to these categories in view of training the classification model, we annotated the tweets using the policy issue Lexicoder dictionary. After preprocessing (most notably, removal of stop-words, removal of punctuation, lemmatisation, and generation of bigrams), we trained an ensemble model based on Random Forest and Gradient Boosting Machine using the R package h2o (LeDell et al., 2018). The accuracy of the classifier is shown in Table 5.3.2.

Table 5.3.2: Accuracy of the classifier for the generic frames in both samples of Twitter users

		Twitter sample of politically interested users				Twitter sample of random users			
Generic frames	**Original Lexicoder categories**	**precision**	**recall**	**F1**	**accuracy**	**precision**	**recall**	**F1**	**accuracy**
civil rights	civil rights	0.78	0.79	0.79	0.87	0.78	0.67	0.72	0.82
culture & identity	culture, education, religion	0.72	0.67	0.69	0.81	0.66	0.64	0.65	0.80
economics	labour, macroeconomics	0.72	0.70	0.71	0.83	0.72	0.66	0.69	0.81
foreign policy	international affairs, defence	0.76	0.75	0.75	0.86	0.68	0.72	0.70	0.83
law order	crime	0.80	0.84	0.82	0.88	0.77	0.83	0.80	0.87
welfare	healthcare, housing, social welfare	0.79	0.78	0.78	0.86	0.67	0.70	0.68	0.81

We also consider specific frames of migrants in terms of ‘victims’ and ‘criminals’ using lists of search queries. The list for ‘victim’ which we use is “.*victim.* | .*scapegoat.*”. The list for ‘criminal’ reads as “.*criminal.* | .*rapist.* | .*rape.* | .*murder.* | .*illegal.* | .*intruder.* | .*alien.*”.

Finally, we also include a number of control variables in our regressions. For instance, we control for users’ tweeting frequency because this can be a strong predictor of the number of migration related tweets, since it accounts for users’ general level of online activity. We also include user’s mean sentiment on immigration.

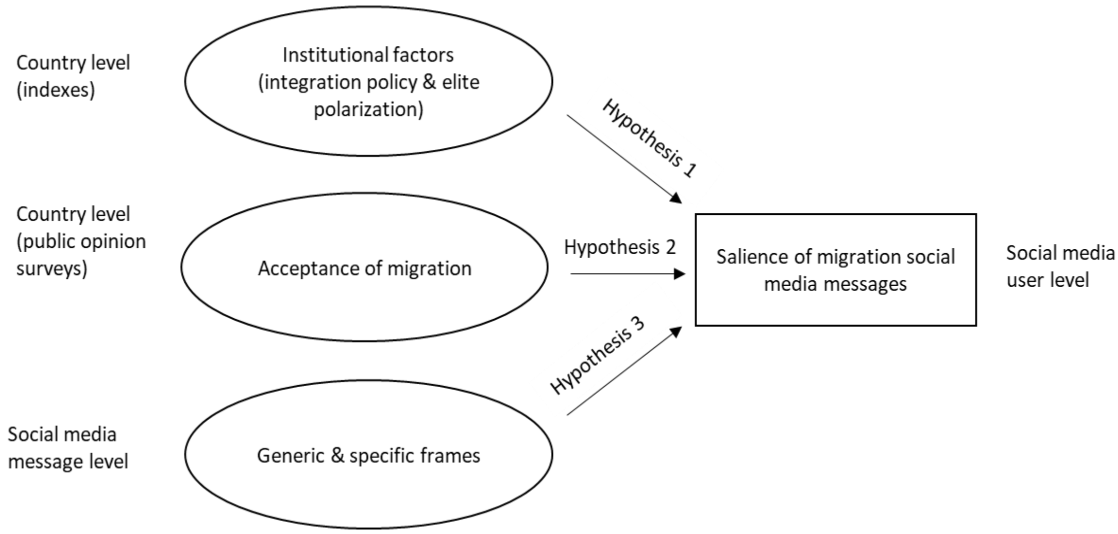


Figure 5.3.1: Conceptual framework summarising the explanatory factors of the salience of migration on Twitter

Results

Comparing the Salience of and Tonality toward Migration Online and Offline

The salience of migration as a topic of social media discussion relates to its visibility and can be compared to survey respondents’ perceived importance of the topic. Furthermore, the sentiment (or tonality) of social media discussions about migration is important to understand the evaluations of online users as compared to representative samples of the population.

Figure 5.3.2 displays the salience of migration related discussion on social media for our different samples of Twitter users (random users, interested users, and politicians) and compares it to the survey distribution related to respondents’ perceived importance of migration as a policy concern. The salience is given as a percentage of the number of tweets mentioning migration of the total of the collected tweets of users from each sample. We observe that in a majority of countries (Australia, New Zealand and the United Kingdom) politicians pay more attention to migration than Twitter users and survey respondents, which is probably related to Brexit in the case of the UK and possibly Ireland, and the contested detention policy in Australia. Furthermore, when taking different samples of Twitter users (random and interested users), we end up with similar distributions in most countries. Pearson correlation between the salience of migration in surveys compared to the different Twitter samples indicate that the correlation is the highest with the random sample of Twitter users (0.81), followed by the sample of interested users (0.67), and politicians (0.16).

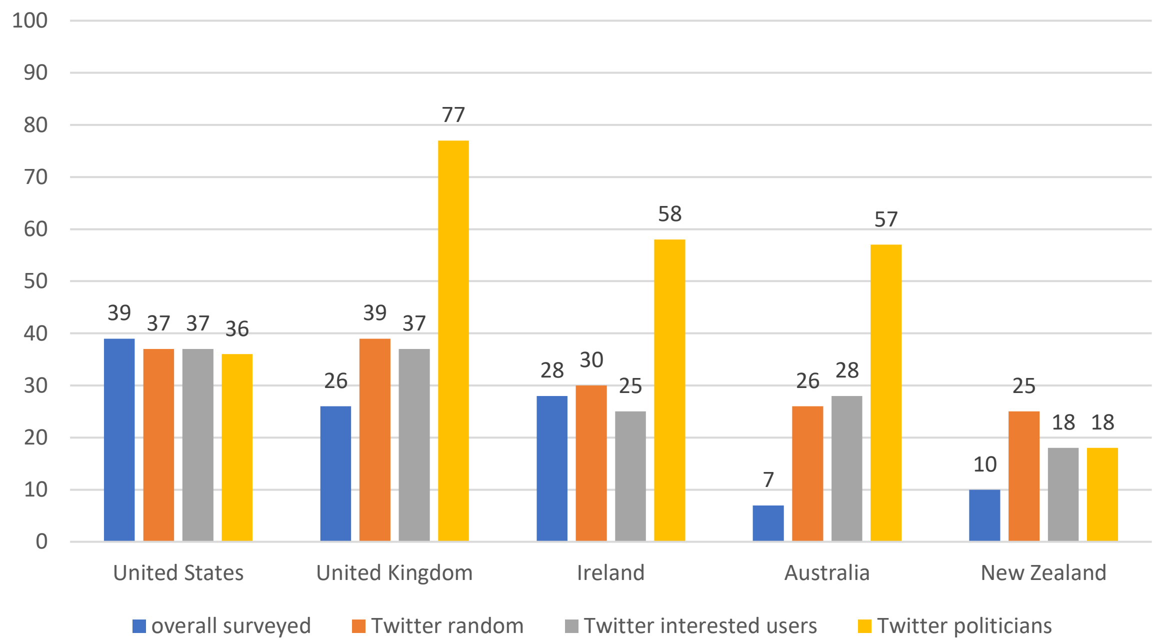


Figure 5.3.2: Salience of migration on Twitter for our different samples of users and in public opinion surveys by country

Figure 5.3.3 displays the correlation between sentiment toward migration on Twitter (for both samples of Twitter users and for politicians) and the percentage of public support for migration in public opinion surveys. The Pearson correlation between the sentiment of migration in surveys compared to the different Twitter samples indicate that the correlation is the highest with the interested sample of Twitter users (0.94), followed by the sample of random users (0.80), and politicians (0.38). The ascent on Figure 5.3.3 is much steeper for interested Twitter users than for random Twitter users, thus indicating polarisation and the fact that this is a more indicative user group. The fact that the correlation is much lower from the plot including politicians is particularly due to the outlier behaviour of the politicians from the United Kingdom on Twitter. This can be explained by the fact that the political discourse is much more polarised than in other countries due to discussions surrounding Brexit.

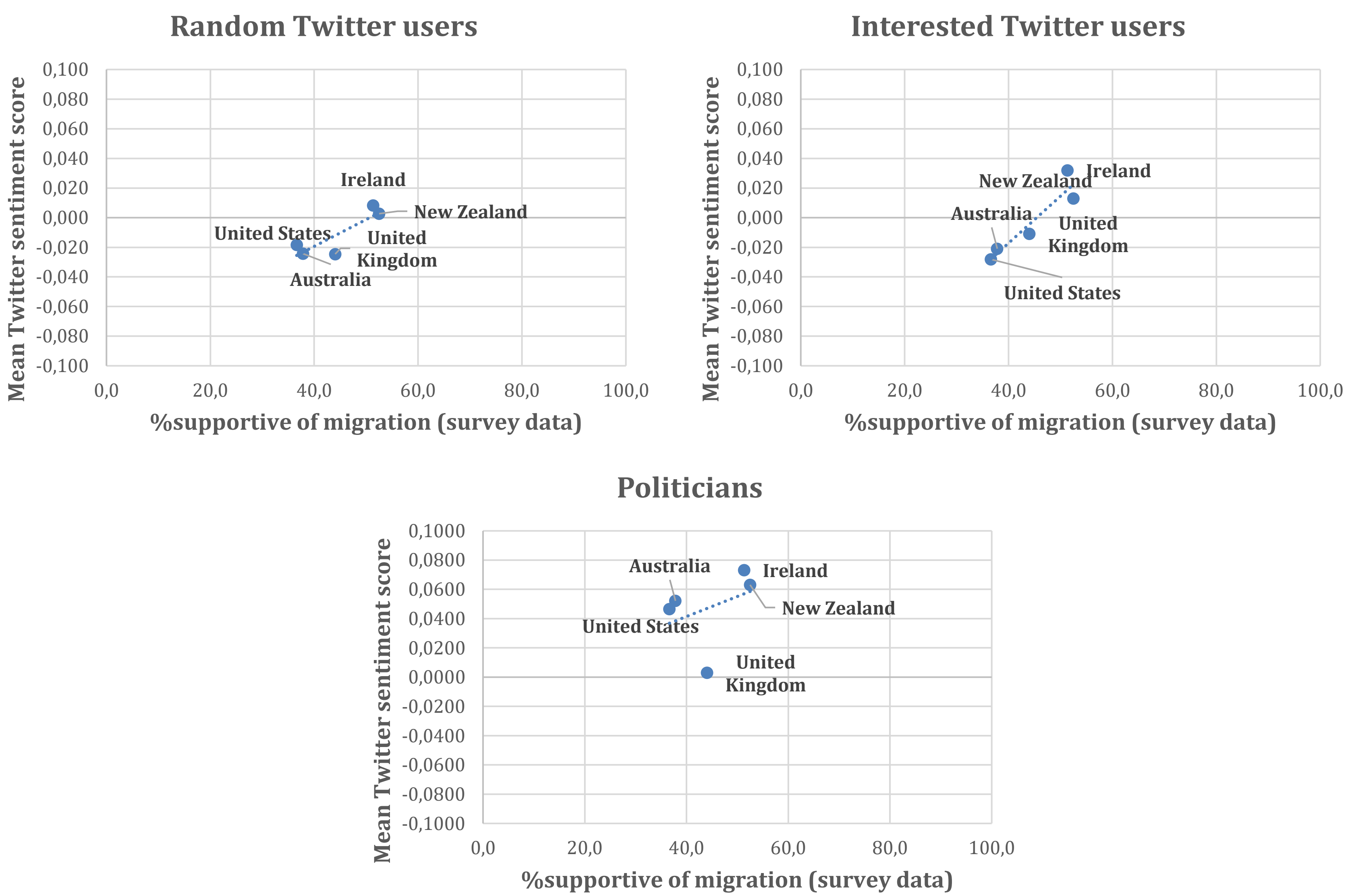


Figure 5.3.3: Correlation between sentiment toward migration on Twitter (for both samples of Twitter users and for politicians) and the percentage of public support to migration in public opinion surveys, split by country

The differences in sentiment distribution for the samples of Twitter users are displayed in Figure 5.3.4. Overall, we can see that the sentiment score is much lower for politicians than in the other Twitter samples. Random users have a mean of 0.37, compared to the mean sentiment of interested users of 0.53. The difference is highly significant at $p<0.001$ (t-test). Random users compared to politicians, who have on average a mean sentiment of 0.05 $p<0.001$ (t-test), also deliver a highly significant difference at $p<0.001$ (t-test). Finally, interested users versus politicians is also significant at $p<0.01$ (t-test).

Furthermore, the sample of politicians shows a more polarised distribution of sentiment than the sample of users. The standard deviation of both random users and interested users is 0.05, while the standard deviation of politicians is 0.13. We also observe that some of the most negative tweets come from the politicians, possibly aiming to incite their followers. The most negative tweets in the sample of politicians forcefully reject right-wing immigration policies, but there are also negative statements about immigrants. The most positive tweets are related to the advantages of highly skilled immigrants for the receiving country. We first thought that the positivity of interested users may sometimes be due to them applauding the politicians they follow rather than the topic of migration, but the data shows very few such instances.

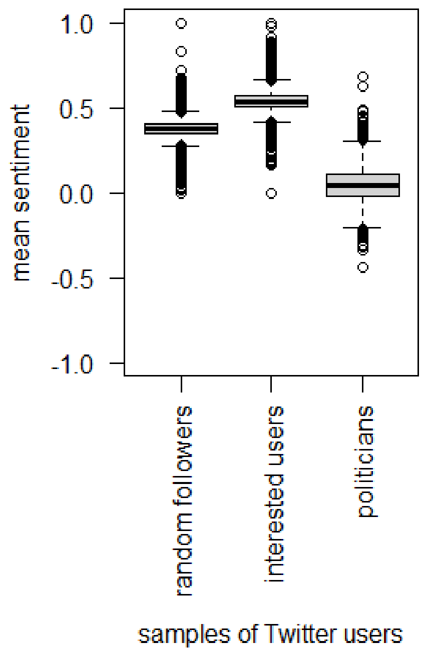


Figure 5.3.4: Distribution of sentiment in tweets for both samples of Twitter users and for politicians

Explaining the Salience of Tweets about Migration

In this section, we discuss the visibility of migration on social media and focus on the impact of the salience on migration in online discussions at the user level. We especially focus on how several factors impact the salience of online discussions about migration at the user level. To do so, we link social media messages, survey data, and societal indicators by applying linear regression models as explained in Section 3. We predict the salience of random twitter users in Model 1, and the salience of interested twitter users in Model 2. As predicting factors, we on the one hand use societal variables like Gallup's Migration Acceptance Index, elite polarisation, the Migration indexation Index (MIPEX; Section 2). On the other hand, we add linguistic factors, namely sentiment and specific content features which are indicative of generic frames, specific frames. The result is given in Table 5.3.3.

With respect to societal factors, we note several interesting effects. First, the level of elite polarisation slightly, but significantly, impacts the salience of migration discussions. This impact is negative, which means that higher levels of elite polarisation tend to be associated with a decreased salience of migration discussions online, thereby suggesting migration is discussed less prominently. Second, the effect of the index about migration integration policies is also negatively associated with the salience of migration on social media. This suggests that the salience of migration discussions on social media is higher when there are fewer institutional responses dedicated to the integration of migrants.

With respect to the connection between social media and public opinion, Table 5.3.3 shows that migration acceptance significantly impacts the level of salience of migration for both of our Twitter samples, random and interested users. However, while the effect is negative for the former, it is positive for the latter. This suggests that the more polarised sample of Twitter users is more likely to tweet about migration when the level of public acceptance is high at the country level, which indicates that this sample of users may be more likely to gather dissenting voices on migration.

Regarding the generic policy issue frames of migration, there are notable differences between both samples of Twitter users. In a nutshell, economy, foreign policy, and law & order framings have the effect of opposing direction between our samples of Twitter users. A possible explanation is that both groups of users pay attention to different narratives feeding into similar generic and specific framings. For the example of economy, the development of the economy is of more direct concern to the general population (see James Carville's famous quote from Clinton's campaign in 1992 "it's the economy, stupid") than to interested users, who may be willing to sacrifice economic success to the benefit of political or ideological views. Examples of tweets from random users supporting this interpretation are:

(1) @JoshVanVeen @philipsophy But why don't they go down the economic populist route? That's where the open lane is. I think they'll fail with rw populism: anti-immigration & culture wars. First one is irrelevant with borders closed & who's concerned with culture war issues with an economic crisis coming?

(2) Immigration Bill before parliament today. A Bill that would block entry to all care workers, cleaners, shop workers, delivery drivers & other low paid key workers who we clap for every week. Our @JCWI UK polling shows people do not want this. #r4today https://t.co/MJVxdVjp5K

The following is an example of a tweet from an interested user explicitly giving low precedence to economic issues:

(3) @SenatorLeahy @DHSOIG What a Joke, support your country, we're being overrun by illegal migrants. Do your damn job fraudulent hacking hypocrite

Similar arguments can be adduced in the discussion of the welfare frame. Namely, migrants are more likely to be constructed as being given unfair access to benefits and threatening the welfare State. For instance, an interested user writes,

(4) Eighty. Six. Million. Dollars.For hotel rooms. To house illegal immigrants. "Scoop: ICE securing hotel rooms to hold growing number of migrant families" https://t.co/xZyoFmh7IU

The correlation to welfare is high among interested users because many of them ask for support to migrants, particularly in difficult situations which coincide with immigration waves. (5) is from an interested user, (6) from a random user:

(5) I've been moved by the plight of refugees risking their lives in unimaginable ways to get to a safe place. I'm not much of a runner but I'm pledging to run 22 miles in September. Please sponsor me! <U+0001F64F> Thank you. <U+0001F496> @everydayherouk #everythingcounts https://t.co/KXRE5hEhZ1

(6) @lilibellmia @BlueSea1964 "The true measure of a man is how he treats someone who can do him absolutely no good." - Samuel Johnson. After the Golden Age Illegal Immigrants & the WALL = MOOT. Nobody should have to LIVE in FEAR. Put yourselves in their SHOES. https://t.co/81zdWVspxE

Regarding the issue-specific framing of migrants, both the samples of interested and random users put emphasis on frames depicting migrants as 'criminals.' An example from random users is given in (7).

(7) Who voted for mass open door immigration and who wanted to see the sort of aggressive scenes on the streets of Britain we are witnessing in Batley? Multi culturalism has never in history ever worked. It can't. The left have caused this. I was called a racist.

Compared to the sample of interested users, where the 'victim' frame is frequent, such as in (8), it is under-represented in tweets from random users but can also be found, for instance (9).

(8) @DavidFrankal She is trying to copy the Australian asylum system by sticking them in unspeakable camps like in Nauru and PNG. You only have to watch @4corners docus to see how bad they are.

(9) Trump shared a video that begins "the only good Democrat is a dead Democrat." Remember: when he referred to immigrants & asylum seekers as an "invasion", an "infestation", as "animals" they - and anyone perceived to be an immigrant - became targets. 23 people murdered in El Paso.

In view of investigating the variations in the narrative about migration, we do not only measure the direct impact of generic frames, but also to their effect in conjunction with tonality. To do so, we include interaction terms between generic frames and sentiment in tweets. Figures 5.3.5 and 5.3.6 display the results for the mean (and +/- 1 standard deviation) of sentiment. For the interested Twitter users, economy and sentiment are particularly strongly correlated (pane C): a more positive sentiment leads to more tweets on economy. But also law & order is strongly correlated to sentiment (pane E) with interested users: here, a more negative sentiment leads to more tweets. Indeed, more negative opinions about economy and law & order appear more often within the sample of interested users compared to the random user sample. Examples with strong negative sentiment from interested users are given in (10), which is from law & order, and (11), from economy.

(10) This is evil. Days after immigrants were gunned down in El Paso, Trump is continuing the attacks on immigrant families. Our job is to reject Trump's racist agenda,

end the terror inflicted on immigrant communities and bring families together, not tear them apart. https://t.co/NAjZes02Aw

(11) And by the way, this India immigration bill (HR.1044 & S.386) is a disaster. It's a big-tech subsidy. India would dominate all employment green cards for the next decade. Is this what they call diversity? Shame on @MikeLeeforUtah. https://t.co/hKBgpWB599

The most strongly correlated factors for the random users are foreign policy and sentiment. Examples in this class are given in (12) and (13).

(12) @Nigel Farage @BorisJohnson No. Boris is wet, weak & woke and will happily accept mass immigration on an even larger scale than Blair.

(13) Our PM is shocked at alleged war crimes by our #SAS in Afghanistan while being part of a political party that for past 20 years has been demonising & dehumanising Muslim refugees from the Middle East. What message do you think this sent and what culture did this foster? #auspol

The strong correlation indicates that foreign policy kindles the strongest feelings, and that sentiment among the general public is typically higher. Non-experts tend to associate migration first and foremost with foreign policy.

Concerning the control variables, there is a small, but significant and negative effect of the users' sentiment, suggesting that online discussions about migration are generally unfavourable towards migration. Furthermore, the tweeting frequency has a significant and positive effect, suggesting that users who rely more heavily on Twitter are also more likely to address the topic of migration. Here it would be tempting to compare the effect size of the tweeting frequency between the two groups, which is higher for interested users. But this is statistically not permissible, as the sample sizes, which affect absolute frequency weights, are different.

Table 5.3.3: Linear regression model explaining salience of tweets (log transformed)

	Model 1 **Random users**	**Model 2** **Interested users**
(Intercept)	-0.03 (0.01) ***	0.06 (0.01) ***
Gallup's Migration Acceptance Index	-0.03 (0.01) ***	0.06 (0.01) ***
Elite polarization	-0.03 (0.00) ***	-0.04 (0.00) ***
Migration integration index (MIPEX)	-0.10 (0.01) ***	-0.07 (0.00) ***
Generic frames		
civil rights	7.37 (1.07) ***	17.21 (2.13) ***
culture & identity	33.52 (1.19) ***	12.56 (0.98) ***
economy	12.12 (0.71) ***	-52.50 (2.69) ***
foreign policy	-20.36 (1.60) ***	0.55 (0.37)
law & order	-2.89 (0.93) **	23.52 (1.27) ***
welfare	19.68 (0.84) ***	6.04 (0.94) ***
Generic frames x sentiment		
civil rights x sentiment	-16.41 (2.59) ***	-27.94 (3.72) ***
culture & identity x sentiment	-74.07 (2.79) ***	-18.77 (1.63) ***
economy x sentiment	-27.56 (1.86) ***	96.99 (4.89) ***
foreign policy x sentiment	52.43 (4.24) ***	1.27 (0.68)
law & order x sentiment	10.08 (2.51) ***	-42.06 (2.34) ***
welfare x sentiment	-45.63 (2.13) ***	-9.69 (1.66) ***
Specific frames		
criminal	0.21 (0.02) ***	0.77 (0.07) ***
victim	-1.97 (0.10) ***	0.04 (0.04)
Sentiment on Twitter	-0.03 (0.00) ***	-0.04 (0.00) ***
Tweeting frequency	0.26 (0.00) ***	0.34 (0.01) ***
R^2	0.46	0.46
Num. obs.	29498	25172

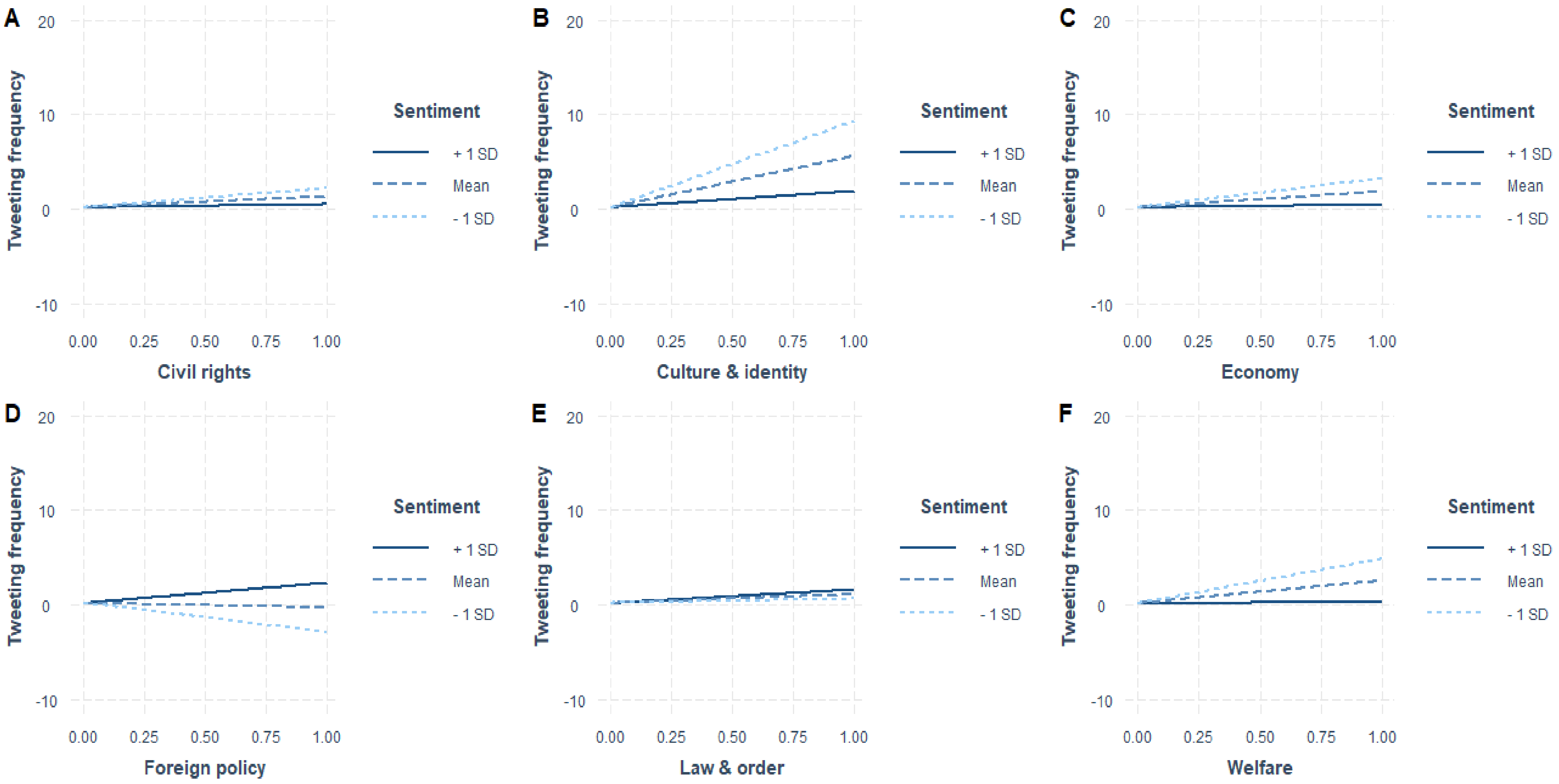


Figure 5.3.5: Interactions between generic frames and sentiment in tweets for the sample of random Twitter users

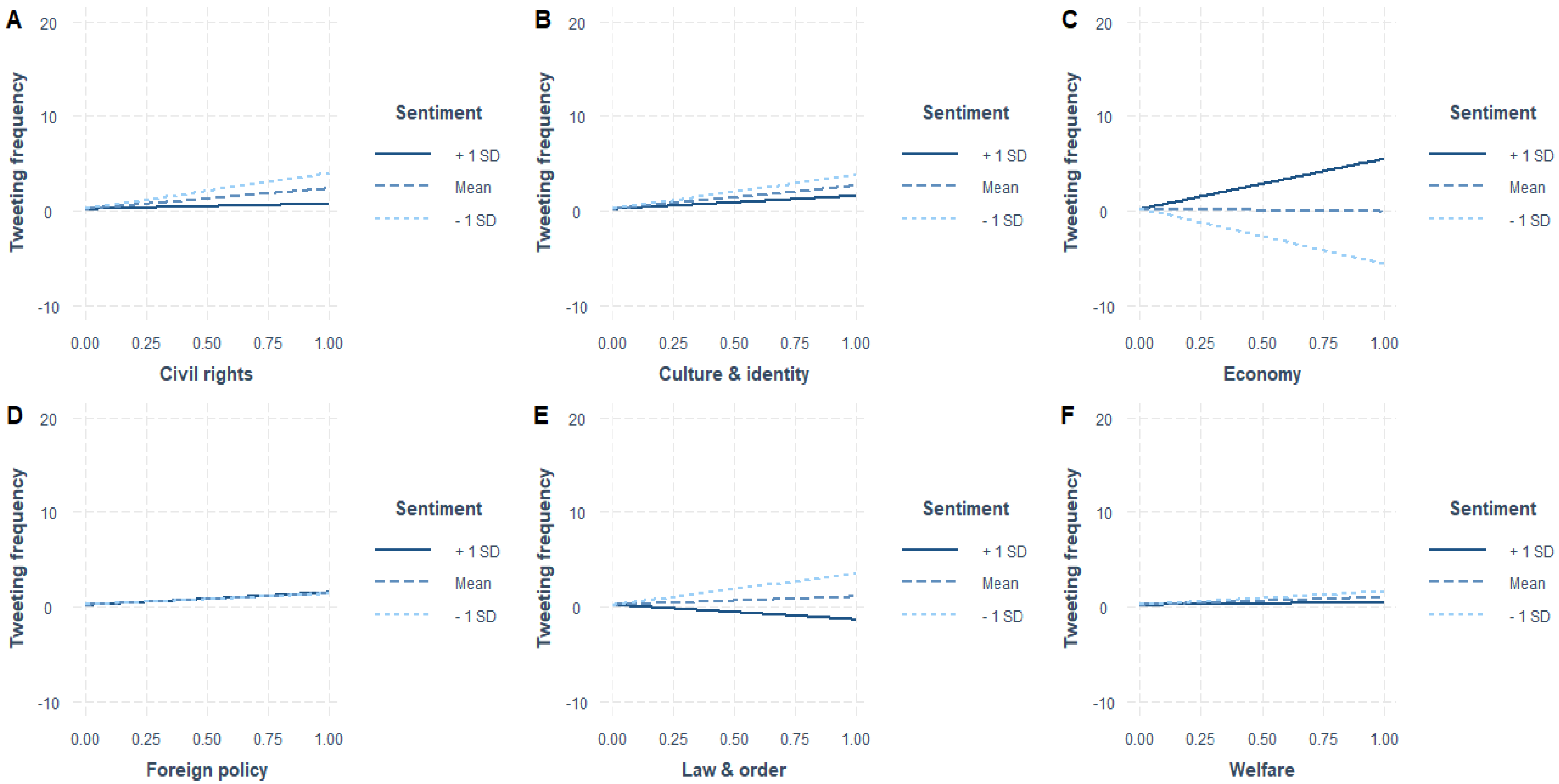


Figure 5.3.6: Interactions between generic frames and sentiment in tweets for the sample of interested Twitter users

Interested users are much more specific and engaged. This is shown in Figure 5.3.7 which extracts the most discriminatory words for each generic frame using the tf-idf measure. This is a classic measure to detect keywords, that is, to reflect how important a word is to a document in a collection or corpus.

Major differences can be observed in foreign policy where the two Twitter samples tend to adopt different behaviours. For instance, the random samples rather engage in a broadcasting style of communication by citing 'hot' events (e.g. Brexit, flee), typical entities (e.g. Macron, Erdogan), and agreements (e.g. pact) of the public debate. The interested users tend to be more engaged in the migration debate by using more specific terms that link migration to direct political events (e.g. election, council) and concrete policy making (e.g. fairness, dialogue).

A similar logic applies to the economy, where the random users cite numbers and figures (e.g. billions, yearly), whereas the interested users are more engaged with concrete policy measures (e.g. wage for all, stimulus) and refer to ways of life (e.g. dreams, growth).

With respect to civil rights, the random users are, again, rather nonspecific and call to overarching principles (e.g. constitution, equity), whereas interested users refer to specific social movements and events (e.g. migrants' lives matter, migrants stuck offshore).

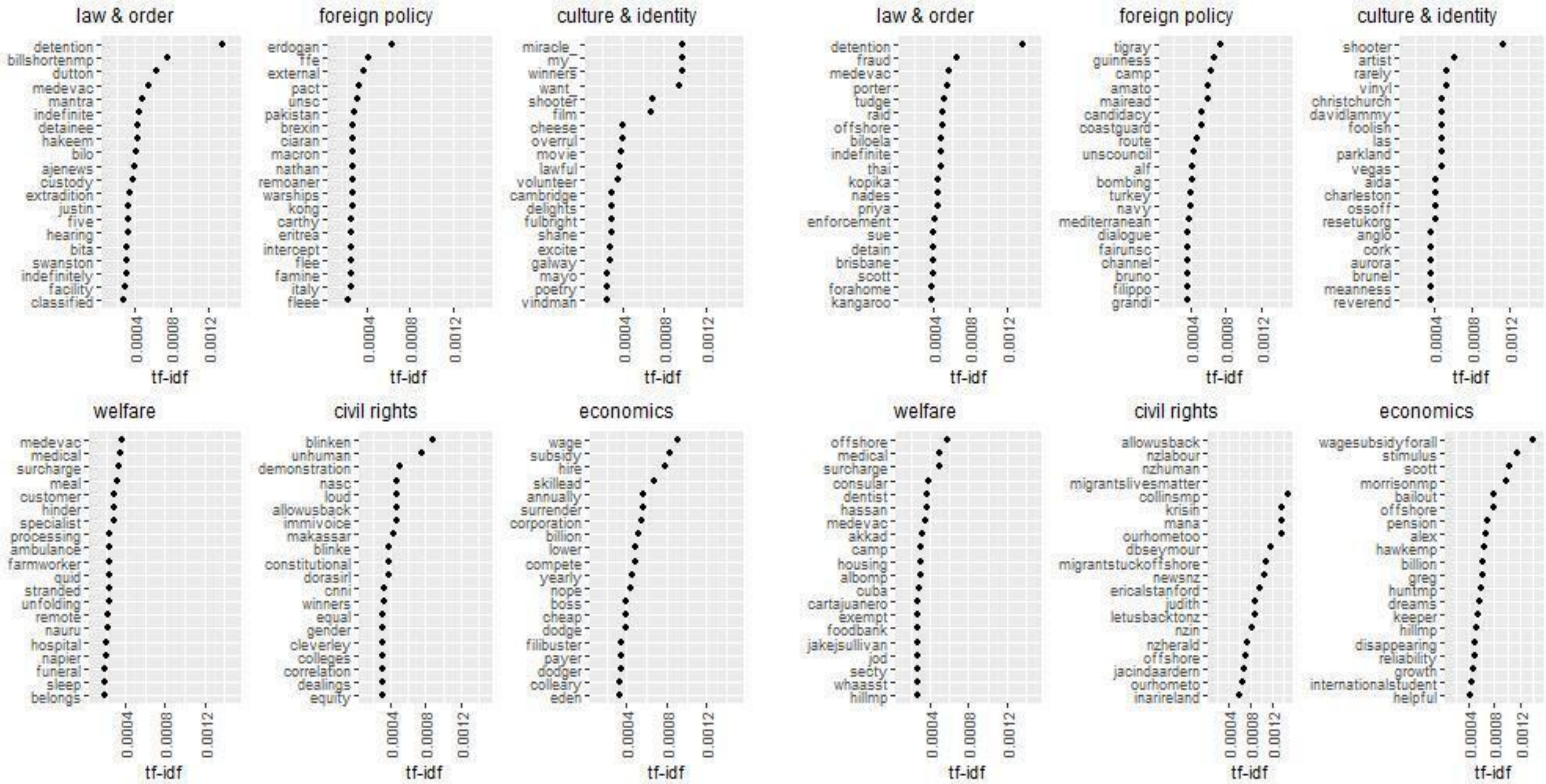


Figure 5.3.7: Top words by generic frame for the sample of random users (left pane) and interested users (right pane)

Discussion of the Main Findings and Concluding Remarks

In our study, we proposed to compare migration discourses in traditional opinion surveys and social media in a cross-country perspective among five English-speaking countries. Let us discuss our main findings with reference to our initial hypotheses.

Main Findings

Hypothesis 1: the salience of migration online correlates with the extent to which migration is perceived as an important concern in representative opinion surveys.

We can answer this hypothesis positively. Pearson correlation between the salience of migration in surveys and on Twitter shows a high correlation of 0.81 with the random sample of Twitter users. This indicates that salience can be used as a good approximation to surveys. Correlation to interested users and politicians is lower, which is also expected as polls aim to capture the stance of the general population.

Hypothesis 2: the tonality related to migration online correlates with the overall satisfaction toward migration found in representative opinion surveys.

We can answer this hypothesis positively. We observe a very high Pearson correlation of 0.94 between the sentiment on migration in surveys compared to the interested Twitter users, and a high correlation of 0.80 to the random users. While we find a good match between social media and surveys, we do not claim that Twitter users are representative of the national populations. Rather, it merely suggests that there is a shared public mood at the national level when looking at aggregated measures between survey respondents and social media users. The merit of relying on two different samples of users enabled us to show how it might influence sentiment distributions, particularly with respect to interested Twitter users, which overrepresent engaged, politically active or strongly opinionated users. Our study thus nuances earlier findings that sceptically concluded that social media users are not representative of the population by showing that users produce significantly different averages in sentiment compared to survey respondents, especially by being less supportive of migration (see Amaya et al. (2020)'s study on Reddit). Our aim was not to construct equivalent distributions of sentiment toward migration as found in opinion surveys, but to use both data sources in tandem to better understand how public attitudes toward migration interplay.

The fact that the correlation of sentiment is higher than the one of salience indicates, on the one hand, that it is worth adding a linguistic analysis, albeit simple and not adapted to the domain. On the other hand, it may also suggest a revision to Ghanem (1997)'s statement that "[t]he frequency with which a topic is mentioned probably has a more powerful influence than any particular framing mechanism" (p.12). It is also worth observing that politicians only correlate with 0.38.

Hypothesis 3: the salience of tweets related to migration is more pronounced when societal and political factors (migrant integration policy and elite polarisation) are unfavourable to migrants and immigration.

We found no support for this hypothesis. Indeed, higher levels of elite polarisation tend to be associated with a decreased salience of migration discussions online, which contradicts the direction of our hypothesis. It could be that if the elite is devised on a topic, then it may mostly affect citizens' positions on an issue but not necessarily their perceived importance of the issues. Furthermore, the MIPEX is also in the opposite direction from what we hypothesised. This could be explained by the fact that social media discussions on migration are most likely to take place in countries where integration mechanisms are scarce.

Hypothesis 4: the salience of tweets related to migration is correlated with lower levels of public acceptance of migration.

Gallup's Migrant Acceptance Index is a significant factor in the regression analysis. However, its effect is negative, as expected, only for the random sample, while it is positive for the sample of interested users. We suggest that this could be linked to the fact that, in contexts where there is a high public acceptance of migration, interested users are likely to voice their positions, perhaps in a dissenting direction and as a counter-reaction movement to the general acceptance of migrants. On the reverse, the random sample of users tend to be less involved in migration discussions on social media when the acceptance of migration is high and, thereby, presumably perceived as under control.

Hypothesis 5: the salience of tweets related to migration is positively associated with discussions about migrants and migration using a threat related rhetoric.

We could confirm this hypothesis in several respects. For instance, we showed that the law & order frame is especially prevalent for interested users, especially from a standpoint on migration. We also noticed that there are different affordances according to our samples of users to pay attention to generic frames. For instance, the development of the economy is of more direct concern to the random sample than to interested users, who balance economic concerns with the benefit of political or ideological views. When looking at the specific depictions of migrants, we noticed that they are more prominently characterised with the criminal rather than the victim frame.

Migrants and migration are thus generally associated as being a threat to the country of arrival. However, a closer look at the top words used in each generic frame also allows us to derive more positive attitudes towards migrants, namely through concerns related to the threat to life for immigrants on their journey.

Study Limitations

Our sample provided a cross-country analysis including only English-speaking countries. However, future studies would benefit from including other regions of the world and additional countries. For instance, it would be interesting to compare multiple receiving European countries. However, this poses additional challenges due to the language variety. Furthermore, the countries included in our sample are essentially receiving countries. Other studies could also envisage conducting temporal analyses, such as the study of Yantseva (2020) comparing multiple media sources.

Moreover, although we implemented different tweet collection strategies, it may well be that limiting our analysis to the followers of political accounts excludes groups of users with different ideas about migration. However, we are confident that we could sample users with enough variation in the countries and ideological orientations. We recommend that similar and other collection strategies be made for the sake of comparability between countries and years of analysis, but also for different social media platforms (e.g. Facebook, Instagram, TikTok).

Another limitation lies in the use of a dictionary-based approach for sentiment analysis. In the future, it may be possible to use more machine-learning algorithms with domain specific validation. Furthermore, we should note that sentiment contained in tweets does not necessarily equate the stance of a speaker toward migration. Future improvements could also be made in this direction to render social media data more comparable to surveyed attitudes.

Additionally, Twitter represents an important source of social media discussions. However, we encourage future research to additionally use other platforms, such as

Reddit or Youtube (see Lee & Nerghes, 2018), but also to use other types of content, such as pictures or videos, to study attitudes towards migration.

Finally, the perceived importance and specific framings of migration may alter citizens' perceptions of and attitudes toward migration or migrants (Vliegenthart & Boomgaarden, 2009, p.309). To analyse such relationships, most studies have relied on public opinion data from (panel-)surveys that link to aggregate analyses of relevant media coverage (Eberl et al., 2018, p.210). However, future research needs to test the influence of migration discourse on public opinion by integrating broader media samples, including and social media.

The salience of migration and the tonality with which it is publicly discussed are important as they may influence broader public opinions on migration and migrants. Albeit this relationship has been tested by combining (panel) surveys with media analysis (see review by Eberl et al., 2020), it is so far understudied with respect to social media discourses. A notable exception is the study of Heidenreich et al. (2020), who focused on party communication on social media. Our study contributes to this line of inquiry and provides an approach that can be usefully extended to other countries and frames of migration.

CHAPTER 6. CONCLUSION

Each part of this thesis has addressed a specific problem and made an independent contribution to the study of public opinion using social media data in complement with survey data. In this concluding chapter, the lessons learned from the chapters are discussed. The limitations of this thesis are also explained and there is reflection on future research paths.

6.1. Lessons learned

6.1.1. Relevant approaches complementing social media data and survey data for studying public opinion

The first research question asked how social media data are used for the study of public opinion. To answer this question, chapter 2 entailed a theoretical review proposing a historical perspective that is representative of scientific discussions covering the last decade. It thus provided a faithful picture of the reflections, challenges and consensus that have been conducive to the establishment of current best practices. It offers several contributions.

The first contribution is to demonstrate that the main approach relying on both data sources to date has been from the perspective of replacing opinion surveys with social media data (e.g., election outcomes). However, albeit successful examples with both data align exist, the mechanisms underpinning the congruence between both data remain largely unclear. For instance, there are doubts about whether (aggregated) features from social media and survey answers can measure the same phenomenon (e.g., sentiment *vs* approval). Based on these concerns, the proposed review suggests to consider other ways to complement both data sources, thus going beyond the replacement perspective.

Therefore, the second contribution of the review is to highlight other complementary approaches which have different research purposes: validating survey findings, improving the sustainability of the research by diversifying the views on a phenomenon, improving the reliability of survey measures by specifying measurements and improving the interpretability of social or political issues. Highlighting these different purposes offers researchers a framework with which to guide research designs that complement both data sources for the study of public opinion.

The third contribution of the review is to discuss how future directions about the proposed framework can be extended to encompass other types of textual data, such as transcripts of parliamentary debates or news articles.

6.1.2. Reliable methods for extracting opinion and stance from social media text

The second research question asked what reliable methods there are for extracting opinions from social media data. Chapter 3 dealing with this question provided two methodological papers which aimed at discussing the pros and cons of existing text-as-data methods for answering questions of relevance for social sciences. In doing so, a critical perspective on computational tools was adopted, while adjustments were proposed that were relevant in research scenarios typical for conducting social research (e.g., skewed data, small available annotated data, few pre-existing domain-specific dictionaries).

In the first methodological paper, a social science perspective on tools from computational social science was proposed: in particular, an investigation of the pros and cons of different approaches to classifying democracy-related tweets in a sub-optimal research framework with texts skewed toward categories and researchers unable to afford large samples of annotated texts. In the second methodological paper, the importance of developing a reliable method for detecting stance from social media data was highlighted, specifically a method that can be easily adapted to a variety of topics of discussion. Both methodological papers converge with respect to several conclusions.

Firstly, both papers demonstrated the merits of combining multiple approaches for conducting text classification tasks (e.g., identifying relevant democracy dimensions and identifying stance). For instance, the papers showed the usefulness of a dictionary-based approach when complemented with data-driven insights (e.g., from word embeddings or tf-idf scores) to improve the scope and the domain-specificity of custom dictionaries.

Secondly, both papers provided evidence for introducing the "human-in-the-loop" during intermediate verification steps (e.g., reviewing candidate words for dictionary) or for incentivising the classification model in specific direction (e.g., strong features in stance detection). Practically, this enabled us to enhance transparent and replicable research by showing the decisions made in a stepwise manner.

Thirdly, both papers demonstrated the usefulness of an interdisciplinary setting including social science data management knowledge, computational science technologies, and computational linguistic foci on semantic information. For instance, the paper on stance detection proposed to take a social science perspective on computational methods that improves the reliability and robustness of classical classification tasks, including sentiment and frames detection tools from computation science, while also adding knowledge about linguistic features.

6.1.3. Influential social media user groups and how they related to public opinion

Our third research question asked what publics are available on social media and how they interact with public opinion. Chapter 4 covered three contexts of social media uses through three empirical papers. The departure point was the idea that, although not representative of the general public, social media users can convey opinions which are interesting to study in themselves, especially because influential users have the potential to represent (or influence) what the broader public thinks. Three empirical studies focusing on the evolution of the social media users were conducted, and the importance of systematically reporting which audiences are influential on a given research topic underlined.

In the first empirical paper focused on profiles of non-political citizens (using survey panel data), thus pointing to potential evolutions related to the Swiss media consumption landscape. Notably, it suggested that there is no simplistic opposition between the consumption of online and offline news, but rather the emergence of a requirement of citizens to possess the necessary skills to process a wide amount of available news to distinguish between good and bad information. This finding is important from a societal perspective, notably with respect to phenomena such as fake news and misinformation, both of which have the potential to influence (streams of) public opinion.

The second empirical paper examined the evolution of politically involved users by relying on the history of users with whom elected politicians interact. It aimed to assess whether politicians' interaction with these (non-)political audiences enables them to gain political success online and offline. The main theoretical contribution is to show that the interactions between political actors and their audiences is still evolving, thus pointing to possible adaptation in politicians' communication strategy. This findings can have important implications when interpreting the content posted by politicians as it requires considering which audiences politicians have in mind. To date, however, Twitter-based activity is moderately impacting politicians' political success, both in terms of political ranking and media coverage. This findings may change in the near future as social media use is increasingly the norm in political communication and can, thereby, be increasingly relevant for shaping public opinion.

The third empirical paper was motivated by the fact that there are still relatively few studies that systematically indicate the distribution of profiles included in their corpora of social media data. However, who speaks about a given topic might strongly influence the interpretation and the generalisability of the results. User groups that were particularly involved over the course of a social mobilisation, namely the women's strike that took place in June 2019 in Switzerland, were identified. Focusing on politically involved users declaring a political affiliation, it was shown that that the polarisation effect of the topic of gender equality is amplified online compared with trends found in survey data. The fact that the political extremes (from the left and the right) voice their positions more strongly than what is measured through surveys may also indicate that social media serve as channels of opportunity for these parties, which generally possess less resources (in terms of members and finances) than mainstream parties. Furthermore, taking advantage of the textual nature of social media text enables us to highlight different framings of gender equality online and how this could reflect public opinion trends was discussed.

6.1.4. Social media data are useful from exploratory, comparative and explanatory research depending on survey data availability

The fourth research question asked how social media can be used to provide a new lens through which to view topics that are well-established or under-investigated in social and political sciences by complementing survey data. Social media-based research has vast potential as it can access very large groups of individuals who are publicly voicing their thoughts and opinions on a variety of social and political topics. In this view, it is important to assess whether social media users talking about a topic (e.g., amount of attention to the topic, framing of the topic, and satisfaction towards the topic) informs public opinion research.

In chapter 5, three empirical researches were proposed displaying three emblematic ways to complement social media and survey data along a continuum going from exploratory (i.e., how social media insights can inform future survey research), descriptive (i.e., whether social media insights are congruent with public opinion trends), and explanatory (i.e., how social media and survey data can be integrated with each other to provide a more complete picture of societal trends). These studies demonstrate the usefulness of using social media data to identify important dimensions of under-investigated topics in survey research by offering a windows into actual and past opinions with unprecedented reach and content details. This is especially useful when the aim is to conduct exploratory research. The presented studies further show that social media can be relied upon to provide new approaches to well-established concepts, especially when conducting descriptive and explanatory research. Importantly, every approach must place the complementarity of both data sources at the centre of their research design.

6.1.5. Thoughts about the text-as-data approaches and output visualisation

In this thesis, multiple supervised and unsupervised approaches were used to make sense of social media data. Chapter 1 introduced these text-as-data approaches according to the framework proposed by Grimmer and Stewart (2013). The experience gained along this thesis suggests that researchers interested in using unsupervised approaches might increase their understanding of social and political phenomena by pushing further survey insights into the direction of "what" and "how" questions. For instance, topic modelling and word embeddings are suitable to indicate "what" are salient topics (e.g., what politicians are most likely to talk about) and narratives. Such techniques are also useful to understand "how" a given topic (or theme) is talked about (e.g., what are salient words or aspects associated with a topic) and "how" it is accompanied by semantic or linguistic patterns. Furthermore, clustering techniques, such as correspondence analysis, can be useful to answer both the "what" and the "how" and can be combined with meta-information that enables us to structure or improve the readability of the conceptual maps. An important remaining concern with unsupervised approaches is the need to validate the output internally (e.g., semantic coherence) and externally (reflection of "true" patterns in the economy).

The highly heterogeneous techniques of text-as-data analysis can produce divergent outcomes which contribute to the discussion about the choice of the relevant method(s) to address a specific research question. In general, there are two approaches within quantitative content analysis (Hogenaraad et al., 2003): the correlation and the substitution approaches. The first approach heavily emphasizes the co-occurrences of words in view of classifying them into categories. The second approach classifies words based on ad-hoc dictionaries containing themes and built by researchers to test particular hypotheses and theories. In this view, this thesis suggests that triangulating text-as-data approaches might be the most useful for finding a middle ground between these approaches and for improving confidence in the results, rather than focusing on achieving the highest possible model accuracy.

An additional observation worth mentioning is that, although unsupervised methods are widely used in computational science, more work is needed to improve the visualisation of the output (Schneider et al., 2017). However, visualisation techniques for displaying many of the salient features are equally essential for validity purposes. For instance, the chapter 5 offers several examples of how the output from one unsupervised model can feed another method for displaying the results. These combinations can complement existing visualisation techniques (e.g. tethne, pyLDAvis, textplot) and previous studies showing the utility of adding meta-information to conceptual maps for more qualitative interpretation (Schneider 2022; Reveilhac & Schneider, 2022). Furthermore, it is interesting that correspondence analysis, which has been developed in sociology, can be expanded to include new data sources, such as social media. To do so, visualisation tools were developed to facilitate text analysis for social scientists (e.g., FactoMineR or R.TeMis). To date, these tools are developed mostly to as stand-alone visualisation techniques. However, the frequent combination of text-as-data approaches in the framework of a single project suggests that increased transferability between the tools and methods is desirable.

The visualization tool also has to be adapted to the research objective. For instance, correspondence analysis is very useful when creating word maps on which external (or passive) features (e.g., group or individual characteristics) can be introduced as supplementary variables (i.e., variables that do not affect the shape of the word space) to improve our interpretation of the map (see article about the women's strike in chapter 4). However, conceptual maps built on kernel density estimations are useful to grasp the relationship between the word features, such as synonyms and collocates (see article about health technologies in chapter 5). Although external features can also be added to conceptual maps, the inclusion of these features will necessarily impact the shape of the obtained maps, as well as the ordering of the documents.

6.2. Critical reflections about the thesis

This thesis entails several limitations worth addressing in future studies.

First, this thesis focusses on **Twitter**, which represents only one possible social media with specific technological affordances for research (e.g., data collection API, available meta-information, dominance of specific user groups). For instance, compared to other social media platforms, Twitter is especially popular among political actors and journalists, as well as among other groups interested in politics (e.g., activists, NGOs, companies, experts, etc). While the composition of the Twittersphere is very interesting for collecting opinions about political topics (e.g., elections and policies) and for understanding how these opinions reflect and/or influence offline opinions, it would be illusory to equate opinions expressed on Twitter with the public opinion measured through surveys. However, social media in general, and Twitter in particular, offer opportunities for identifying opinions that are under-represented in or concurrent to opinions measured through surveys. Importantly, online opinions can sometimes also be precursors of opinions that will occupy public debates only later on (e.g., social movement claims can first appear on social media). Therefore, Twitter data cannot be overlooked when studying questions related to public opinion. Beyond the issue of the non-representativity of social media users, social media data are characterized by a low information to data ratio. This is partly due to the fact that data collection is often opportunistic, meaning that it is not specifically designed to answer a specific question. Researchers have to apply search queries and other cleanings to remove irrelevant or unnecessary data before analysing the data. As design choices and the social media platform algorithms can introduce bias into the data, it is important that researchers clearly describe how they collect the data (e.g., keyword-based, actor-based, location-based, etc.) and what actors or groups are prominent in the final database. In some cases, it is possible that initially large datasets turn into relatively small datasets once filtered and cleaned so as to only keep relevant data. From the perspective of the study of public opinion, this suggests that more social media data does not necessarily translate into better data (Boyd & Crawford, 2012; Hargittai, 2015). Cross-platform analyses, as well as the reliance on alternative data sources (e.g., register data, health data, geo data, transcripts of political debates), which are also increasingly used in many disciplines of social science, constitute necessary steps to better assess the validity and the generalisability of the findings from Twitter studies. Following the path proposed by this thesis, the main question would be to assess how much these data can contribute to the study of public opinion.

Second, this thesis mainly draws conclusions from the **Swiss case**. Albeit this is not a limitation *per se*, expanding the analyses to other countries would enable cross-national comparisons. Indeed, the generalisability of the findings is necessarily limited by the specificities of the use of a given social media platform in a country. This is important as the use of social media might be very different from one country to another. For instance, in Switzerland, the share of the public using Twitter is rather low in comparison to that of other European countries. Furthermore, while it was still limited a few years ago, the share of political actors today relying on social media, and especially on Twitter, is still increasing. What must be noted, however, is that politicians' increased reliance on social media has little probability of drastically impacting the democracy and the process of decision making, at least in the near future. Although being very useful for grasping different views and for understanding common issues (e.g., immigration, climate change, poverty, etc.), cross-country comparisons can nevertheless be challenging, most notably due to the general lack of available geographic information about users. Even though there exist ways to proxy the geographical location of users (e.g., by keeping users who follow a minimum of nationally bound seed users, such as political and media accounts), we do not have reliable access to the geographic information of most of the data available online.

Third, this thesis relied on **text-as-data** approaches and makes thus little use of additional features. For instance, it makes use of network information mostly at the data collection stage, but there are other opportunities to use network information, for example as an additional information to detect ideologically similar groups or groups that share a similar stance about a target issue. Content analysis is worth extending to other types of data, such as images. Although little is yet known about the reliability of such data to understand opinions, research should put under scrutiny whether "a picture is worth a thousand words". However, it is much more difficult to determine with features should be extracted from image and video than from text. Text implies a form of translation of ideas into something understandable by other people, while an image or a video does not necessarily need this translation step. Furthermore, audio files are also a promising source of data for research. Indeed, automated speech recognition or a "speech-to-text" approach could supplement human annotation and lead to the generation of rich and new material for social scientists (see Pentland et al., 2021). Unlike standardized survey items that can be submitted to respondents from different countries to assess how public opinion differs or is similar across cultures (assuming that the question wording is unequivocal across countries), social researchers might be sceptical about conducting similar analyses using text data. In the meantime, textual data are unavoidable as it is essentially through language that humans express their views and understandings about society. When relying on text data, it is important to acknowledge that there might be fundamental ambiguities in language, as language is by its nature context-dependent (e.g., similar words can mean very different things according to the context). To address this issue, studies relying on text classification methods typically intend to maximize intercoder agreement. Writing careful (and theory-based) coding rules and thinking ahead about potential ambiguities and misinterpretations are also part of the potential solutions.

Fourth, this thesis focusses on the **production of descriptive knowledge** (examples of research questions are: What happens? Why it happens? How important is it? What is the context?). This emphasis on the utility of quantitative description goes somewhat against the trend in quantitative social sciences in favour of causal inferences (Gerring, 2012). However, providing accurate descriptions can also avoiding spurious causal research. In the near future, more research will also be needed to examine causal relationships between social media exposure, public attitudes and behaviour (Schuck, Vliegenthart, & De Vreese, 2016). This would enable the field to move beyond exploratory findings. This said, the production of descriptive knowledge and of ruled-based models remains essential to improve our understanding of outputs produced in the framework of computational social science, as the field is prominently marked by the use of prediction models (e.g., elections forecasting) and concerns about the "black box" nature of these models. Even though the richness and scope of social media data can give the impression that the information available can push the boundaries of social sciences, more research is needed to compare the performance and interpretability of different methods to address a specific task. This can be done by introducing variations in the research setting (e.g., by adapting the size of the databases and/or the skewness of the division of categories) and in the research design (e.g., by varying the data collection strategy). More replicability and generalisability studies are also needed (e.g., by transferring the models to different, but domain specific. databases).

Fifth, we did not include a case study directly **integrating social media data into a given survey**. This alternative strategy involves contacting social media users directly via surveys (Vaccari et al., 2013) or obtaining consent from survey respondents to track their online media consumption behaviour (Guess, 2021). Integrating social media data and panel survey data (e.g., by merging them on a topical or temporal basis) has the potential to highlight the causal mechanisms at play in public opinion formation, which would be difficult or impossible to investigate using only one source of data. It might also enable us to shed new light on the direction of certain effects. For instance, while certain effects seem more plausible, such as ideology influencing vote choice, the direction of other effects is less straightforward, such as the relationship between national symbolic frames and vote choice. Nevertheless, integrating both data sources to address substantive research questions can be challenging as respondents who agree to share their social media information may not post about topics that are the focus of a survey. To ensure sufficient social media data, surveys would need to include a very large number of respondents. However, even increasing the survey sample size is no guarantee that the topic will be covered on social media and in a sufficient variety of aspects to address topical social science questions. Integrating both data sources can nevertheless be of interest to cover behavioural questions, thereby bypassing survey recall error or desirability bias (e.g., online news consumption practices and the relationship between user networks and attitudes towards given policy issues).

Sixth, this thesis only used a **sample of text analysis methods**. Other methods are also valuable for conducting opinion research, such as text scaling (e.g., wordscore or wordfish), and for measuring proportion when categories are known (e.g., ReadMe). Furthermore, the thesis did not rely on advanced techniques pertaining to deep learning, which are claimed to be "featureless" models as they bypass feature extraction. Furthermore, no comparisons between "zero-shot" (when a machine is taught how to learn from data without accessing the data itself) and "few-shot" (when a machine is taught how to use data to learn from a specific point of view) models are performed, nor does the thesis employ transfer learning, which stores knowledge gained while solving one problem and applies it to a different but related problem. Moreover, the thesis does not rely on classification strategies based on a human-in-the-loop machine learning paradigm, in which one or many human annotators and an automatic classifier label incoming instances. Such applications are very useful to contribute to the creation of accurate classifiers. For instance, Pandey et al. (2022) propose an error-avoidance approach to the active learning paradigm that is robust to key human errors during annotations (see also Reason (2000) for a detailed description of human errors in terms of mistakes and slips). Finally, this thesis takes social media text data as the primary source of analysis but does not investigate the potential of automated text generation for social science research.

6.3. Future theoretical research paths

The theoretical and empirical studies presented in chapters 2 to 5 can open up to future theoretical research using social media with a public opinion perspective.

Chapter 2 provides a theoretical framework which could be extended to a more cohesive framework that clearly identifies best practices in the selection and coupling of appropriate methods and technologies for social media research. The proposed framework could encompass other data sources (e.g., news articles, registries, archives, meta-data, sensor-data), and could be made suitable for including qualitative or mixed methods research. For instance, it can include sequential quantitative and qualitative phases, as well as qualitative (close-) reading of social media messages to better grasp the context of public discussions.

Chapter 2 also invites future studies to move beyond the idea that the representativity issue constitutes a sufficient reason for social media data being unable to contribute to the study of public opinion and inform social research. From a survey perspective, a major drawback of social media lies in the lack of generalisability of the findings to some overall population. This constitutes an issue that is unlikely to be resolved, at least in the near future, given that citizens around the world do not use social media with a similar intensity and that many accounts are either non-individual or non-human. Researchers have made substantial advances in understanding and adjusting for representation errors (Pasek et al., 2019; Barberá, 2016), notably by using external data such as polling results to (in)validate the validity of aggregate measures from social media (e.g., electoral outcome, presidential approval, economic satisfaction). Conversely, it may also be possible to reverse the validation process by training machine-learning models to find features that directly correlate with variations in the quantities of interest, such as past polling measures of vote intention (Beauchamp, 2016).

Based on the experience from chapter 3, there is a need to identify minimal criteria that make social media data a useful source and a credible complement to survey data for studying opinions. In addition to sentiment or emotions detection, further methods that can reliably assess users' stances need to be developed. Indeed, stance measures are most suited to comparison with survey items formulated in terms of agreements or support. In particular, stance does not necessarily equate to sentiment in social media messages, and satisfaction does not necessarily equate approval in survey answers. It is therefore essential to dedicate increased effort in future to assessing the validity, reliability and robustness of stance detection methods applied to social media data, notably by detecting features of importance for conducting transferable text classification tasks. Researchers are thus encouraged to adopt approaches that triangulate different methods of classifying social media text into pre-established categories. The combination of multiple methods to reinforce the accuracy of classification tasks is suggested as a fruitful endeavour. In addition, researchers are also encouraged to introduce "human-in-the-loop" validation steps to create custom dictionaries and evaluate the importance of textual features that affect the classification tasks.

Based on the findings from chapter 4, future research should systematically account for the actors included in corpora of social media messages to understand how online trends interact with or reflect public opinion. Indeed, it is important to consider that a large proportion of users are passive observers rather than active posters of social media content. Consequently, most users are either not active at all or choose not to engage in social media discussions. Here, it should be remembered that previous survey methods have also suffered from similar problems. For example, telephone surveys struggle to contact people who are ex-directory, whilst face-to-face interviews are biased towards those who are at home during the day. However, as opposed to opinion surveys, social media data rarely come with demographic information, thus making it difficult to identify which groups of social media users are over- or under-represented. Not only can social media data miss some specific opinions, but it can also hardly be assumed that peoples' social media activity perfectly reflects their personal (political) views.

Based on the findings from chapter 4 (and to some extent, from chapter 5), there are opportunities to use the information stemming from the manual annotation of user profiles to develop automatic classification of user categories based on the profile description fields or user network information (Rauchfleisch et al., 2021). Manual annotations, albeit time-consuming, was found to be an effective way of identifying salient and influential user profiles for which it would have been difficult to define *a priori* user categories. The (manual) classification of social media users into groups is of utmost importance to improve the interpretability of social media findings. This is important, as social media platforms are likely to evolve (e.g., change of rules government data access, and change of focus or uses), and this will induce the migration of social media users to other platforms (e.g., Facebook to Telegram for the anti covid-measures users). Knowing who are the involved users also enables us to conceive which audiences they potentially have in mind when posting on social media. This matters a great deal for understanding the impact of social media on public opinion, as Twitter is geared more towards elite actors who have strategic goals in terms of persuasion or opinion making (e.g., politicians, journalists, opinion makers, etc.). This fact is rarely problematized in published papers, especially because these actors have a prominent position on Twitter (e.g., more followers, more reach, circulation, etc.) and, thereby, a high potential to influence public opinion formation.

Chapter 5 invites to think about the usefulness of social media data as a complement to survey data not just on whether it can answer research questions, but whether it can improve on existing methods and insights in some respect. Future studies should try to disentangle cases where social media either reflect or influence public opinion on important social matters. Indeed, trends occurring on social media can no longer be understood in the way that media communication was understood previously – messages merely passing through press organisations and affecting the opinions of citizens. Instead, social media users are active producers of information and have precise audiences in mind when producing social media content. Albeit Twitter is dominated by particular user groups, it is important to disentangle cases where social media are leading public opinion formation and when social media are mirroring the public opinion. However, such research designs are often complicated by lack of suitable datasets as they mostly require panel survey data (or very frequent cross-sectional surveys) that can be augmented with social media data.

6.4. Future practical research paths

A practical challenge for future research on public opinion is finding survey measures against which social media measures (e.g., sentiment) can be compared, both in terms of values (e.g., self-produced content *versus* assessments on pre-given scales), but also more generally in terms of formulation (e.g., messages on Twitter *versus* attitudes in standardised survey formulations). In this view, survey methodologists could propose (new) ways better allow for other data integration when setting up the survey questionnaires, for instance, by allowing for more open-ended survey questions.

Beyond survey methodology, increased collaboration between social scientists and computation social science researchers is needed to move towards the elaboration of more sophisticated and valid measures of theory-driven concepts (Baden et al., 2020). To do so, future collaborations could investigate which data-error assumption brings improvement in agreement between the text-as-data analysis and the hand-coded gold standard. For instance, chapter 3 proposed a data-driven extraction of relevant signals (e.g., pointing to frames and stance) with a manual coding of the identified patterns. However, more collaborative work is needed for assessing which underlying parameters (from both supervised and unsupervised models) impede reliable classification tasks. In this view, multi-step error analysis frameworks appear as a fruitful research avenue (Ng, 2019; Dobbrick et al., 2021).

In addition to the collaboration between survey methodologists and social media researchers, the inclusion of ongoing efforts in computational linguistics and computational tool developers offer a fertile way to promote more transparent and robust research infrastructure explicitly addressing how different parameters (such as linguistic factors and pre-processing steps) impact the performance of computational tools (Chan et al., 2020). As new channel of communication may replace social media in the future, there is a need to develop methodologies that can be transferred to other data sources which are used across disciplines.

The collaborations between these different disciplines also offers the potential for providing solutions to the “replication crisis” which concerns empirical branches of (social) sciences since the early 2000s (Sönning & Werner, 2019). Despite the widespread attention to the issue, there is still a lack of agreement about what, precisely, the crisis consists of (e.g., “reproducibility crisis”, “theory crisis”, “generalization crisis” among others). A partial answer could be found in the formulation of concepts that can be concretely tested and implemented among different data sources and (social) media platforms. This suggests bridging the gap between theory-driven and data-driven approaches, for example by adopting hybrid approaches based on the deductive analysis of social media data and the extraction of theoretical concepts from the textual data. Here the expertise of (digital) communication scientists to manage large-scaled discursive data can help to solve dilemma related to trade-off between efficient computational strategies and theoretical sensitivity.

There are also practical questions with respect to data collection. Social media platforms have granted access to researchers to collect data posted by users with different access rates (see recent review of Twitter as a research data by Chen et al., 2022). For instance, Twitter is very open to data collection, storage, and analysis (even the retrieval from historical data is available), while Facebook is much more restrictive albeit it is by far the most popular social media. A major difficulty to better contextualise and explain public opinion with social media data is the fact that, while there is generally some meta-information about how much attention is triggered by a social media message (e.g., in terms of likes and shares), normally there is little idea about how many people saw a post and how (or whether) it affected their opinion on a given topic. Researchers should use this information whenever it become available with new technological arrangements of social media platforms.

6.5. Ethical note and thoughts about the societal impacts of social media research

Before concluding, this ethical note addresses some of the core issues applying to publicly available Twitter data, but not to non-public data (e.g., private profile information or direct messages). It describes a range of factors that were central to each empirical study presented in this thesis, namely legal compliance, user consent, user privacy and anonymity, data sharing, and publication. The information given below is a reflexion about the ethical concerns that took place during the retrieval and analysis of the Twitter data, as well as during the publishing of the results. Researchers may also be interested to know more about Twitter's terms of service[42] and Twitter's developer agreement and policy[43]. It is also recommended that researchers seek for the relevant aspects of the GDPR (General Data Protection Regulation) that apply to their particular research. There might also be important and influential national or local regulation considerations.

A key ethical principle in relation to Twitter data retrieval and analysis is the consideration of autonomy, and most notably the concepts of informed consent and participant expectations. From a legal standpoint, Twitter users agree to Twitter's terms of service when signing up to get a (personal, corporate or parodic) account, or simply by

[42] See: https://twitter.com/fr/tos
[43] See: https://developer.twitter.com/en/developer-terms/agreement-and-policy

browsing its content via its website or its apps. As tweets are mainly posted using official or third-party apps on mobile or fixed devices, they are often accompanied by, or contain, metadata (e.g., geolocation or posting time). In practice, a Twitter dataset is dynamic as its content changes regularly. There may thus be changes that users make to the availability of their content (e.g., user-driven addition and deletion of tweets, as well as changes to the status of available information). Therefore, it is essential that a research design regularly considers ethical questions during retrieval, retention, sharing[44], and publication. Importantly, appropriate considerations must be given as to whether the benefits of the research outweigh that aspect of ethical risk, especially by mitigating informed consent (and autonomy), as well as privacy and anonymity concerns.

Regarding consent, Twitter informs its users about the fact that their data may be used for research and provides documentation explaining the public visibility settings. Pragmatically, nevertheless, it is important to acknowledge the barriers that can impede users being informed of their rights (including consent and privacy). For instance, it can be difficult for users to stay up to date with terms and conditions. Documents about the Developer API further inform researchers about what uses can be made of Twitter data. For instance, in January 2021, Twitter allowed a preferential access for academic research (Tornes & Trujillo, 2021). A frequent practice in research using Twitter data is to argue that the requirement of informed consent could be nuanced by the fact that research uses publicly visible data, thus considering consent as implicit (Gold, 2020). Indeed, for most research projects, asking for an active informed consent from users might be impractical. In addition, Twitter content is typically dynamic as user-driven changes affect the availability of public information. For example, when a user deletes a tweet, it might be interpreted as withdrawal of consent (Kamocki et al., 2021). Compared to social media research, the process of information update in survey research is more explicit, as respondents can ask for the removal of their answers from the database at any time.

Protection with respect to privacy and anonymity is similar to any other research handling personally identifying data. However, social media data is almost impossible to anonymise. For instance, Internet search engines can trace the content back to its source,

[44] Concerning data sharing, Twitter permits the sharing of tweet IDs and user IDs in a dataset for others to use, thus suggesting that other users must re-hydrate IDs into tweets using Twitter's API. This ensures that anything that has been deleted on the platform is not shared because the ID will not resolve to anything.

thus identifying the user. Researchers must thus apply appropriate data protection management and conduct risk assessment to address privacy and anonymity concerns (Williams et al., 2017), while also considering that Twitter places restrictions on types of derivation and inference for certain analyses (e.g., ascribing an individual characteristic irrespective of whether a user had stated this information directly). In this respect, data matching (e.g., associating Twitter user data with external private or geographical data), individual profiling, facial recognition, and monitoring of sensitive events and groups are considered sensitive analyses[45], which may harm and stigmatise users (Williams, Burnap & Sloan, 2017). Aggregating Twitter information may constitute an acceptable form of data protection management (e.g., Mahoney et al., 2022), especially when aggregate analyses also do not require the storage of personal data and individual identifiers.

With respect to publishing, Twitter requires certain items of data to be retained in the published form, which may, in turn, pose a challenge to privacy (e.g., the anonymity of the users) and to the synchronisation requirements (e.g., a paper needs to be modified if a user deletes or protects a tweet which was quoted in that paper). Furthermore, to protect the identity (and the reputation) of users, researchers may need to take the potential protection steps of using indirect citations or of paraphrasing instead of using verbatim words (or using quotes without mentioning the usernames). However, this could raise concerns of users' tweets being taken out of context or used in a distorted manner to promote views and ideas that were not intended (Golder et al., 2017, p.9).

That said, researchers conducting studies based on social media data must also consider the potential societal impact of their findings. For instance, particular attention should be paid to a study's potential for controversy, its potential for public interest (e.g., press, political authorities, etc.), and its potential effects on public opinion at large, on specific groups that may be concerned by the study's implications (e.g., NGOs, activists, followers of a party, social minorities, vulnerable groups, etc.), or on any reader of the findings.

When expressing views and judgements about their results and commenting about the theoretical and practical contributions of their study, researchers must reflect on the public reception of their study. Beninger et al. (2014) have shown that users' views about social media research can fall within the realms of scepticism, acceptance and ambiguity

---

[45] A detailed description of these restrictions are fully documented in Twitter's various agreements and API uses.

depending on the research context and on users' awareness of social media platforms. For instance, feelings of scepticism were found to be related to uncertainty about the validity of the data compared to traditional methods. Furthermore, Hemphill et al. (2022) showed that survey participants think of their social media data as moderately sensitive compared to other private or sensitive information (e.g., driver's licence number or health record). Generally, respondents were supportive of social media research, but also insisted on the necessity for transparent information about how the data would be used in the research (e.g., who will use it and for what purpose). The most enthusiastic respondents even perceived social media research as being a valuable and quick means of accessing diverse and timely information that could mitigate the effect of false information or extreme views and improve research accuracy. However, they also perceived such research as a tool to access views that would be more spontaneous than those expressed when having to answer questions in the presence of others, such as in a survey.

From the perspective of research findings readers, it is also important for researchers to not only reflect on how the results can be used to promote certain policy decisions and used to influence public discourse (e.g., the depiction of migrants through identified narratives), but also to reflect on how the research can contribute to maintaining and improving public trust in the value of such research (Golder et al., 2017). Therefore, when making the study results publicly available, researchers must describe the observed indicators and trends accurately, interpret the results by referring to theoretical expectations, and acknowledge the limitations (e.g., in terms of representativeness and validity) of the study. It is therefore essential that researchers give sufficient information about how to replicate the study, how valid and reliable the measures derived from the data are, and how the results that are obtained can be generalised to other contexts (e.g., another group of users or another platform).

6.6. Concluding remarks

To date, there is evidence that the content of social media has become pervasive in traditional news outlets (Dubois et al., 2020). It is also heavily relied on to appeal to party supporters and frame the political agenda (Kreiss, 2014). In addition, it has become an important means of acquiring, sharing, and discussing political information for numerous (non)political organizations, social movements and citizens (Kwak et al., 2010; Neuman et al., 2014; Tucker et al., 2016). Against this background, social scientists can no longer ignore social media data as influencing or mirroring public opinion.

This thesis makes theoretical and methodological contributions to the study of public opinion by highlighting best practices in complementing social media data and survey data. The ease of access to social media data allied with increasing computing capabilities has created a paradigm shift in the way opinions and behaviours are studied. Early studies relying on social media for answering social and political questions were prompt to claim the potential of these data to replace survey data. Although there have been successful attempts to replace surveys to predict electoral outcomes, this enthusiasm was quickly balanced by replicability issues and a lack of methodological guidelines. At the same time, surveys have been suffering from declining response rates while social media have become widely used as a medium for sharing opinions and as a source of information.

Although there seems to be a consensus that social media will not replace survey research, but rather serve as a complement, research designs studying social phenomena based on social media data are not fully developed yet and the difficulty in complementing both data sources remains an important challenge. The different chapters provide practical guidelines how the best complement these data sources and proposed a critical assessment of existing text-as-data approaches from a social science perspective.

This thesis highlights the need to adopt a methodology that iterates between human work (e.g., interpretation, validation) and computation work (e.g., classification, aggregation) to combine the interpretative skills, contextual knowledge and computational power. Doing so enables us to better justify the use of text analysis methods in our empirical research (i.e., What constructs are relevant? Can textual data be a relevant source to measure them? What are important features to be extracted from the text?). It also help to assess the validity of the findings (i.e., Is the measured quantity/aspect measuring what it is supposed to measure? Is the measured quantity/aspect adhering to a set of excepted patterns?).

To date, significant challenges remain for researchers interested in complementing social media and survey data for studying public opinion. In this thesis, special emphasis is placed on how to design data complementarity (e.g., through data comparison and integration), thus providing useful suggestions on enhancing public opinion measurement with new data sources. However, in addition to social media data, future research should try to include other data sources. This thesis provides solid guidelines for future research aiming to assess the how much alternative data sources can contribute to the study of public opinion.

Annexes

Appendix to section 2.1

Table 2.1.2: List of theoretical papers focusing on the combination of survey and social media data (continues next pages)

Author(s)	Date	Title	Focus
Blumenthal	2005	Toward an open-source methodology: What we can learn from the blogosphere	General
Murphy et al.	2011	Social media, new technologies, and the future of health survey research	Ethics
Birks et al.	2011	Shifting the Boundaries of Research	Ontology
Gayo-Avello	2011	Don't turn social media into another Literary Digest poll	Prediction
Sobkowicz et al.	2012	Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web	Prediction
Boyd & Crawford	2012	Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon	General
Metaxas & Mustafaraj	2012	social media and the elections	Prediction
Gayo-Avello	2012	I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper	Prediction
Gayo-Avello	2012	No, You Cannot Predict Elections with Twitter	Prediction
Sobkowicz et al.	2012	Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web	Prediction
Smith	2013	Survey-research paradigms old and new	Ontology
Couper	2013	Is the sky falling? New technology, changing media, and the future of surveys	General
Couper et al.	2013	Report of the aapor task force on non-probability sampling	General
Schoen et al.	2013	The Power of Prediction with Social Media	Prediction
Gayo-Avello	2013	A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data	Prediction
Stieglitz & Dang-Xuan	2013	Social media and political communication: a social media analytics framework	General
Gayo-Avello et al.	2013	Understanding the predictive power of social media	Prediction
Scime & Murray	2013	Social science data analysis: The ethical imperative	Ethics
Ruths & Pfeffer	2014	Social Media for Large Studies of Behavior	Behaviour

Tufekci	2014	Big Questions for social media Big Data: Representativeness, Validity, and Other Methodological Pitfalls	Error sources
Murphy et al.	2014	Social media, sociality, and survey research	Ontology
Murphy et al.	2014	Social media in public opinion research: Executive summary of the aapor task force on emerging technologies in public opinion research	General
Hill & Dever	2014	The Future of Social Media, Sociality, and Survey Research	General
Tang et al.	2014	Mining social media with social theories: a survey	Ontology
Zagheni & Weber	2015	Demographic Research with Non-representative Internet Data	Demograp hic
Resnick et al.	2015	What social media data we are missing and how to get it	General
Ampofo et ak.	2015	Text Mining and Social Media: When Quantitative Meets Qualitative, and Software Meets Humans	Ontology
Hargittai	2015	Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites	General
Schober et al.	2016	Social media analyses for social measurement	General
Olteanu et al.	2016	Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries	Error sources
Junngherr	2016	Twitter use in election campaigns: A systematic literature review	Prediction
Spiro	2016	Research opportunities at the intersection of social media and survey data	Linking
Conway & O'Connor	2016	Social media, big data, and mental health: current advances and ethical implications	Ethics
RJ Dalton	2016	The potential of big data for the cross-national study of political behavior	Behaviour
Johnson & Smith	2017	Big Data and Survey Research: Supplement or Substitute?	Linking
Hsieh & Murphy	2017	Total Twitter error	Error sources
Salleh	2017	From survey to social media: Public opinion and politics in the age of big data	General

Pal	2017	Studying political communication on Twitter: the case for small data	Small data
Salunkhe et al.	2017	A review: Prediction of election using twitter sentiment analysis	Prediction
Klašnja et al.	2018	Measuring Public Opinion with social media Data	General
Jungherr	2018	Normalizing digital trace data	General
Kwak & Cho	2018	Analyzing public opinion with social media data during election periods: A selective literature review	Prediction
Szreder	2018	Will big data affect opinion polls?	General
Freelon	2019	Inferring individual-level characteristics from digital trace data: Issues and recommendations	Demographic
Trottier	2019	A research agenda for social media surveillance	General
Sen et al.	2019	A Total Error Framework for Digital Traces of Humans	Error sources
Salvatore et al.	2020	Social Media and Twitter Data Quality for New Social Indicators	Error sources
Stier et al.	2020	Integrating survey data and digital trace data: key issues in developing an emerging field	Linking
Romele et al.	2020	Digital hermeneutics: from interpreting with machines to interpretational machines	Ontology
Skoric et al.	2020	Electoral and Public Opinion Forecasts with social media Data: A Meta-Analysis	Prediction
Rousidis et al.	2020	Social media prediction: a literature review	Prediction
Chauhan et al.	2020	The emergence of social media data and sentiment analysis in election prediction	Prediction

Table 2.1.3: List of publications combining social media and survey data for prediction purposes (continues next pages)

Author(s)	Date	Title	Source	SM	Topic	Level of analysis
Tumasjan et al.	2011	*Election forecasts with Twitter: How 140 characters reflect the political landscape*	Social Science Computer Review	Twitter	election	national
Aparaschivei	2011	*The use of new media in electoral campaigns: Analysis on the use of blogs, Facebook, Twitter and YouTube in the 2009 Romanian presidential campaign*	Journal of Media Research-Revista de Studii Media	Twitter & Facebook & Youtube	election	national
Jungherr et al.	2012	*Why the pirate party won the German election of 2009 or the trouble with predictions: A response to Tumasjan*	Social science computer review	Twitter	election	national
González-Bailón et al.	2012	*Emotions, public opinion, and US presidential approval rates: A 5-year analysis of online political discussions*	Human Communication Research	Usenet	election	national
Borondo et al.	2012	*Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish Presidential election as a case study*	Chaos An Interdisciplinary Journal of Nonlinear Science	Twitter	election	national
Jungherr et al.	2012	*Why the pirate party won the German election of 2009 or the trouble with predictions*	Social science computer review	Twitter	election	national
Borondo et al.	2012	*Characterizing and modeling an electoral campaign in the context of twitter: 2011 Spanish presidential election as a case study*	Chaos: An Interdisciplinary Journal of Nonlinear Science	Twitter	election	national
Choy et al.	2012	*US Presidential Election 2012 Prediction using Census Corrected Twitter Model*	arXiv Preprint	Twitter	election	national
González-Bailón et al.	2012	*Emotions, Public Opinion and U.S. Presidential Approval Rates: A 5 year Analysis of Online Political Discussions*	Human Communication Research	other social media	presidential approval	national

Franch	2013	*(Wisdom of the Crowds)2: 2010 UK Election Prediction with* Social Media	Journal of Information Technology & Politics	Facebook, Twitter, Google, & YouTube	election	national
Kermanidis & Maragoudakis	2013	*Political sentiment analysis of tweets before and after the Greek elections of May 2012*	Int. J. Social Network Mining	Twitter	election	national
Fu & Chan	2013	*Analyzing online sentiment to predict telephone poll results*	Cyberpsychology, Behavior, and Social Networking	online discussion forums, personal blogs, and microblogs	election	local
DiGrazia et al.	2013	*More Tweets, More Votes: social media as a Quantitative Indicator of Political Behavior*	PloS one	Twitter	politics	district
Paul & Dredze	2014	*Discovering health topics in social media using topic models*	PloS one	Twitter	health	national
Cheng & Chen	2014	*Global* social media*, local context: A case study of Chinese-language tweets about the 2012 presidential election in Taiwan*	Aslib Journal of Information Management	Twitter	politics	regional
Ceron et al.	2014	*Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France*	New media & society	Twitter	politics	national
Ceron et al.	2015	*Using sentiment analysis to monitor electoral campaigns: Method matters—evidence from the United States and Italy*	Social Science Computer …	Twitter	election	national
Murthy	2015	*Twitter and elections: are tweets, predictive, reactive, or a form of buzz?*	Information, Communication & Society	Twitter	election	State
Eom et al.	2015	*Twitter-based analysis of the dynamics of collective attention to political parties*	PloS one	Twitter	election	national
Wong et al.	2015	*Twitter sentiment predicts Affordable Care Act marketplace enrollment*	Journal of medical …	Twitter	health	national

Durahim & Coşkun	2015	*# iamhappybecause: Gross National Happiness through Twitter analysis and big data*	Technological Forecasting and Social Change	Twitter	well-being	national & regional
Huberty	2015	*Can We Vote with Our Tweet? On the Perennial Difficulty of Election Forecasting with S*ocial Media	International Journal of Forecasting	Twitter	election	national
Jungherr et al.	2015	*Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support*	Social Science Computer Review	Twitter	election	national
Burnap et al.	2016	*140 characters to victory?: Using Twitter to predict the UK 2015 General Election*	Electoral Studies	Twitter	election	national
Cody et al.	2016	*Public opinion polling with Twitter*	arXiv Preprint	Twitter	presidential approval	national
Beauchamp	2017	*Predicting and interpolating state-level polls using Twitter textual data*	American Journal of Political Science	Twitter	election	State
Yaqub et al.	2017	*Analysis of political discourse on twitter in the context of the 2016 US presidential elections*	Government Information Quarterly	Twitter	election	national
Lopez et al.	2017	*Predicting the Brexit vote by tracking and classifying public opinion using twitter data*	Statistics, Politics and Policy	Twitter	Brexit	national
Vepsäläinen et al.	2017	*Facebook likes and public opinion: Predicting the 2015 Finnish parliamentary elections*	Government Information Quarterly	Facebook	election	national
Feng et al.	2017	*Twitter analysis of California's failed campaign to raise the state's tobacco tax by popular vote in 2012*	Tobacco Control	Twitter	health	national
Kristensen et al.	2017	*Parsimonious data: How a single Facebook like predicts voting behavior in multiparty systems*	PloS one	Facebook	politics	national
Beauchamp	2017	*Predicting and interpolating state-level polls using Twitter textual data*	American Journal of Political Science	Twitter	election	national

Oliveira et al.	2017	*Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls*	Journal of Information Technology & Politics	Twitter	election	national
Masoomali et al.	2018	*Using Facebook Ad Data to Track the Global Digital Gender Gap*	World Development	Facebook	equality (gender, immigration, LGB, etc.)	international
Bastos & Mercea	2018	*Parametrizing Brexit: mapping Twitter political space to parliamentary constituencies*	Information, Communication & Society	Twitter	Brexit	constituencies
Chmielewska-Szlajfer	2018	*Opinion dailies versus Facebook fan pages: the case of Poland's surprising 2015 presidential elections*	Media, Culture & Society	Facebook	election	national
Pasek et al.	2018	*The stability of economic correlations over time: identifying conditions under which survey tracking polls and Twitter sentiment yield similar conclusions*	Public Opinion Quarterly	Twitter	economic satisfaction	national
Heredia et al.	2018	Social media *for polling and predicting United States election outcome*	Social Network Analysis and Mining	Twitter	election	national
Zhang	2018	Social media *popularity and election results: A study of the 2016 Taiwanese general election*	PloS one	Facebook	election	national
Bansal & Srivastava	2018	*On predicting elections with hybrid topic based sentiment analysis of tweets*	Procedia Computer Science	Twitter	election	State
Grimaldi	2019	*Can we analyse political discourse using Twitter? Evidence from Spanish 2019 presidential election*	Social Network Analysis and Mining	Twitter	election	national
Awais et al.	2019	*Leveraging big data for politics: predicting general election of Pakistan using a novel rigged model*	Journal of Ambient Intelligence and Humanized Computing	Twitter	election	national
Pasek et al.	2019	*Who's Tweeting About the President? What Big Survey Data Can Tell Us About Digital Traces?*	Social Science Computer Review	Twitter	presidential approval	national

Jaidka et al.	2019	*Predicting elections from social media: a three-country, three-method comparative study*	Asian Journal of Communication	Twitter	election	international
Pasek et al.	2020	*Who's tweeting about the president? What big survey data can tell us about digital traces?*	Social Science Computer Review	Twitter	presidential approval	national
Chin & Wang	2020	*A New Insight into Combining Forecasts for Elections: The Role of Social Media*	Journal of Forecasting	Facebook	election	County & city
Stieglitz et al.	2020	*Going back in time to predict the future- the complex role of the data collection period in social media analytics*	Information Systems Frontiers	Twitter	election	international
Gong et al.	2020	*Measuring relative opinion from location-based* social media*: A case study of the 2016 US presidential election*	Plos one	Twitter	election	State
Sepúlveda & Norambuena	2020	*Twitter sentiment analysis for the estimation of voting intention in the 2017 Chilean elections*	Intelligent Data Analysis	Twitter	election	national

Table 2.1.4: List of publications combining social media and survey data for enrichment purposes

Author(s)	**Date**	**Title**	**Source**	**SM**	**Topic**	**Level of analysis**
Vaccari et al.	2013	Social media *and political communication: A survey of Twitter users during the 2013 Italian general election*	Rivista italiana di scienza politica	Twitter	election	national
Vaccari et al.	2015	*Political expression and action on* social media*: Exploring the relationship between lower- and higher-threshold political activities among Twitter users in Italy*	Journal of Computer-Mediated Communication	Twitter	politics	national
Karlsen & Enjolras	2016	*Styles of social media campaigning and influence in a hybrid political communication system: Linking candidate survey data with Twitter data*	The International Journal of Press/Politics	Twitter	election	national
Hofstra et al.	2017	*Sources of segregation in social networks: A novel approach using Facebook*	American Sociological Review	Facebook	equality (gender, immigration, LGB, etc.)	national
Quinlan et al.	2018	*'Show me the money and the party!' – variation in Facebook and Twitter adoption by politicians*	Information, Communication & Society	Twitter & Facebook	political communication	national
Stier et al.	2018	*Election campaigning on* social media*: Politicians,audiences, and the mediation of political communication on Facebook and Twitter*	Political communication	Twitter & Facebook	politics	national
Cardenal et al.	2019	*Is Facebook eroding the public agenda? Evidence from survey and web-tracking data*	International Journal of Public Opinion Research	Facebook	politics	national
Jacbs & Spierings	2019	*A populist paradise? Examining populists' Twitter adoption and use*	Information, Communication & Society	Twitter	politics	national
De Sio & Weber	2020	*Issue yield, campaign communication, and electoral performance: a six-country comparative analysis*	West European Politics	Twitter	politics	international
Shin	2020	*How Do Partisans Consume News on* Social Media*? A Comparison of Self-Reports With Digital Trace Measures Among Twitter Users*	Social Media + Society	Twitter	politics	national

Table 2.1.5: List of publications using survey as proxy with social media data (continues next page)

Author(s)	Date	Title	Source	SM	Topic	Level of analysis
Vaccari & Nielsen	2013	*What drives politicians' online popularity? An analysis of the 2010 US midterm elections*	Journal of Information Technology & Politics	Facebook, Twitter, and YouTube	election	national
LaMarre & Suzuki-Lambrecht	2013	*Tweeting democracy? Examining Twitter as an online public relations strategy for congressional campaigns'*	Public relations review	Twitter	election	users
Jensen & Anstead	2013	*Psephological investigations: Tweets, votes, and unknown unknowns in the Republican nomination process*	Policy & Internet	Twitter	election	State
Larsson	2015	*The EU Parliament on Twitter—Assessing the permanent online practices of parliamentarians*	Journal of Information Technology & Politics	Twitter	political communication	international
Theocharis et al.	2016	*A bad workman blames his tweets: the consequences of citizens' uncivil Twitter use when interacting with party candidates*	Journal of Communication	Twitter	politics	national
Ceron & d'Adda	2016	*E-campaigning on Twitter: The effectiveness of distributive promises and negative campaign in the 2013 Italian election*	New media & society	Twitter	politics	national
Park et al.	2017	*Cultural values and cross-cultural video consumption on YouTube*	PLoS one	Youtube	other	international
Ernst et al.	2017	*Extreme parties and populism: an analysis of Facebook and Twitter across six countries*	Information, Communication & Society	Twitter & Facebook	politics	national
Stier et al.	2018	*Election campaigning on* social media*: Politicians, audiences, and the mediation of political communication on Facebook and Twitter*	Political communication	Twitter & Facebook	political communication	national
Barberá & Zeitzoff	2018	*The new public address system: why do world leaders adopt* social media*?*	International Studies Quarterly	Twitter & Facebook	political communication	international

Rossini et al.	2018	*The relationship between race competitiveness, standing in the polls, and social media communication strategies during the 2014 U.S. gubernatorial campaigns*	Journal of Information Technology & Politics	Twitter & Facebook	political communication	national
Rossini et al.	2018	Social Media*, Opinion Polls, and the Use of Persuasive Messages During the 2016 US Election Primaries*	Social Media + Society	Twitter & Facebook	political communication	national
Rossini et al.	2018	*Social media, opinion polls, and the use of persuasive messages during the 2016 US election primaries*	Social Media + Society	Twitter & Facebook	political communication	national
Plescia et al.	2019	*Filling the Void? Political Responsiveness of Populist Parties*	Representation	Twitter	politics	international
Wells et al.	2020	*Trump, Twitter, and news media responsiveness: A media systems approach*	New Media & …	Twitter	politics	national
Lazarus & Thornton	2020	*Bully Pulpit? Twitter Users' Engagement With President Trump's Tweets*	Social Science Computer Review	Twitter	politics	national
Daniel & Obholzer	2020	*Reaching out to the voter? Campaigning on Twitter during the 2019 European elections*	Research & Politics	Twitter	political communication	international
Eberl et al.	2020	*What's in a post? How sentiment and issue salience affect users' emotional reactions on Facebook*	Journal of Information Technology & Politics	Facebook	politics	national

Table 2.1.6: List of publications combining social media and survey data for comparison purposes (continues next pages)

Author(s)	**Date**	**Title**	**Source**	**SM**	**Topic**	**Level of analysis**
King et al.	2013	*Twitter and the health reforms in the English National Health Service*	Health policy	Twitter	health	national
Tsou et al.	2013	*Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): a case study in 2012 US Presidential Election*	Cartography and Geographic Information Science	Twitter (+ web pages)	election	national
Kim et al.	2013	*Can tweets replace polls? A US health-care reform case study*	Book chapter (3): Social Media, Sociality, and Survey Research	Twitter	politics	national
Ceron et al.	2014	*Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France*	New media & society	Twitter	election	national
Barry	2014	*Using social media to discover public values, interests, and perceptions about cattle grazing on park lands*	Environmental management	FlickrTM (pictures+comments)	climate & environment & energy	national
Van Dalen et al.	2015	*Policy considerations on Facebook: Agendas, coherence, and communication patterns in the 2011 Danish parliamentary elections*	Journal of Information Technology & Politics	Facebook	politics	national
Jungherr et al.	2016	*The mediation of politics through Twitter: An analysis of messages posted during the campaign for the German federal election 2013*	Journal of Computer-Mediated Communication	Twitter	election	national
Bhattacharya et al.	2016	*Perceptions of presidential candidates' personalities in twitter*	Journal of the Association for Information Science and Technology	Twitter	politics	national

Diaz et al.	2017	*Online and social media Data As an Imperfect Continuous Panel Survey*	PloS one	Twitter	politics	international
Grčar et al.	2017	*Stance and influence of Twitter users regarding the Brexit referendum*	Computational social networks	Twitter	Brexit	national
Davis et al.	2017	*Public response to Obamacare on Twitter*	Journal of medical Internet research	Twitter	health	national
Bajaj	2017	*The use of Twitter during the 2014 Indian general elections: Framing, agenda-setting, and the personalization of politics*	Asian Survey	Twitter	political communication	national
Wang et al.	2018	*Comparing social media Data and Survey Data in Assessing the Attractiveness of Beijing Olympic Forest Park*	Sustainability	other social media	other	local
Farhadloo et al.	2018	*Associations of topics of discussion on Twitter with survey measures of attitudes, knowledge, and behaviors related to Zika: probabilistic study in the United States*	JMIR Public Health Surveillance	Twitter	health	national
Howell et al.	2018	*National Academies of Sciences, Engineering, and Medicine report on genetically engineered crops influences public discourse*	Politics and the Life Sciences	Twitter	health	national
Wainger et al.	2018	*Evidence of a shared value for nature*	Ecological Economics	Twitter	climate & environment & energy	national
Nawa et al.	2018	*Analysis of public discourse on heart transplantation in Japan using social network service data*	American Journal of Transplantation	Twitter	health	national
Scarborough	2018	*Feminist Twitter and Gender Attitudes: Opportunities and Limitations to Using Twitter in the Study of Public Opinion*	Socius	Twitter	equality (gender, immigration, LGB, etc.)	region & State & national
Mancosu & Bobba	2019	*Using deep-learning algorithms to derive basic characteristics of social*	PloS one	Facebbok	politics	national

		media users: The Brexit campaign as a case study				
Merkley et al.	2020	*A Rare Moment of Cross-Partisan Consensus: Elite and Public Response to the COVID-19 Pandemic in Canada*	Canadian Journal of Political Science	Twitter (+GoogleTrend)	health	national
Loureiro & Alló	2020	*Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the U.K. and Spain*	Energy Policy	Twitter	climate & environment & energy	international
Amaya et al.	2020	*Measuring the Strength of Attitudes in Social Media Data*	book chapter (5): Big Data Meets Survey Science: A Collection of Innovative Methods	Reddit	health	national

Table 2.1.7: List of publications combining social media and survey data for generating new insights (continues next pages)

Author(s)	**Date**	**Title**	**Source**	**SM**	**Topic**	**Reason to complement**	**Level of analysis**
Ampofo et al.	2011	*Trust, confidence, and credibility: Citizen responses on twitter to opinion polls during the 2010 UK general election*	Information, Communication & Society	Twitter	election	what citizens think about surveys	national
Robillard et al.	2013	*Utilizing social media to study information-seeking and ethical issues in gene therapy*	Journal of Medical Internet Research	Yahoo! Answers	health	capture emergent opinions	international
Cavazos-Rehg et al.	2014	*Characterizing the followers and tweets of a marijuana-focused Twitter handle*	Journal of Medical Internet Research	Twitter	health	capture emergent opinions	international
Russell Neuman et al.	2014	*The dynamics of public attention: Agenda-setting theory meets big data*	Journal of Communication	Twitter, blogs, forum commentaries, and traditional media news stories	politics	alternative to self-reported measures	national
Kim & Kim	2014	*Public Opinion Sensing and Trend Analysis on Social Media: A Study on Nuclear Power on Twitter*	International Journal of Multimedia and Ubiquitous Engineering	Twitter	climate & environment & energy	capture emergent opinions	national
Trilling	2015	*Two different debates? Investigating the relationship between a political debate on TV and simultaneous comments on Twitter*	Social science computer review	Twitter (+ transcript of TV debate)	election	more nuanced approach of PO	national
Williams et al.	2015	*Network analysis reveals open forums and echo chambers in social media discussions of climate change*	Global Environmental Change	Twitter	climate & environment & energy	more dynamic perspective of PO	international
Kirilenko et al.	2015	*People as sensors: Mass media and local temperature influence*	Global Environmental Change	Twitter	climate & environment & energy	"passive survey" of PO	national & regional

		climate change discussion on Twitter					
Thompson et al.	2015	*Prevalence of marijuana-related traffic on Twitter, 2012–2013: a content analysis*	Cyberpsychology, Behavior, and Social Networking	Twitter	health	capture emergent opinions	national
Sajuria & Fábrega	2016	*Do we need polls? why Twitter will not replace opinion surveys, but can complement them*	Digital Methods for Social Science	Twitter	politics	alternative to self-reported measures	national
Settle et al.	2016	*From posting to voting: The effects of political competition on online political engagement*	Political Science Research and Methods	Facebook	politics	alternative to self-reported measures	State
Marchetti & Ceccobelli	2016	*Twitter and television in a hybrid media system: the 2013 Italian election campaign*	Journalism Practice	Twitter	election	more nuanced approach of PO	national
Krauss et al.	2017	*"Get drunk. Smoke weed. Have fun.": a content analysis of tweets about marijuana and alcohol*	American Journal of Health Promotion	Twitter	health	capture emergent opinions	national
Barisione & Ceron	2017	*A Digital Movement of Opinion? Contesting Austerity Through Social Media*	Social Media and European Politics	Twitter	economic satisfaction		national
Chan & Fu	2017	*The relationship between cyberbalkanization and opinion polarization: Time-series analysis on Facebook pages and opinion polls during the Hong Kong Occupy*	Journal of Computer-Mediated Communication	Facebook	politics	more dynamic perspective of PO	city (Hong Kong)
Flores	2017	*Do anti-immigrant laws shape public sentiment? A study of Arizona's SB 1070 using Twitter data*	American Journal of Sociology	Twitter	equality (gender, immigration, LGB, etc.)	causal inference	State
Stautz et al.	2017	*Reactions on Twitter to updated alcohol guidelines in the UK: a content analysis*	BMJ open	Twitter	health	more nuanced approach of PO	national

Chadwick & Dennis	2017	Social media*, professional media and mobilisation in contemporary Britain: Explaining the strengths and weaknesses of the Citizens' Movement 38 Degrees*	Political Studies	Twitter (+ campaign emails & online news articles)	politics	more nuanced approach of PO	national
Etter at al.	2018	*Measuring organizational legitimacy in* social media*: Assessing citizens' judgments with sentiment analysis*	Business & Society	Twitter	economic satisfaction	expand the scope of survey focus	national
Karami et al.	2018	*Mining public opinion about economic issues: Twitter and the us presidential election*	International Journal of Strategic Decision Sciences	Twitter	election	more nuanced approach of PO	national
Clark et al.	2018	*Using Twitter to study public discourse in the wake of judicial decisions: Public reactions to the Supreme Court's same-sex-marriage cases*	Journal of Law and Courts	Twitter	equality (gender, immigration, LGB, etc.)	expand the scope of survey focus	national
Couper et al.	2019	*Developing a global indicator for Aichi Target 1 by merging online data sources to measure biodiversity awareness and engagement*	Biological Conservation	Twitter (+ GoogleSearches & media)	climate & environment & energy	expand the scope of survey focus	International
Aydogan et al.	2019	*Ideological congruence and social media text as data*	Journal of Representative Democracy	Twitter	politics	novel approach	national
Hatipoğlu et al.	2019	*Automated text analysis and international relations: The introduction and application of a novel technique for Twitter*	All Azimuth: A Journal of Foreign Policy and Peace	Twitter	politics	expand the scope of survey focus	national
Vidal-Alaball et al.	2019	*A New Tool for Public Health Opinion: Using Twitter Polls for Insight into Telemedicine*	JMIR Formative Research	Twitter	health	validate survey measurements	international

Barberá et al.	2019	*Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data*	American Political Science Review	Twitter	politics	expand the scope of survey focus	national
Dahlberg et al.	2020	*Democracy in context: using a distributional semantic model to study differences in the usage of democracy across languages and countries*	Zeitschrift für Vergleichende Politikwissenschaft	Different social media (+ online news)	politics	validate survey measurements	international
Lovari et al.	2020	*Blurred Shots: Investigating the Information Crisis Around Vaccination in Italy*	American Behavioral Scientist	Facebook	health		national
Adams-Cohen	2020	*Policy Change and Public Opinion: Measuring Shifting Political Sentiment with social media Data*	American Politics Research	Twitter	equality (gender, immigration, LGB, etc.)	causal inference	national & State
Tavoschi et al.	2020	*Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy*	Human Vaccines & Immunotherapeutics	Twitter	health	capture emergent opinions	national
Guan et al.	2020	*Chinese views of the United States: evidence from Weibo*	International Relations of the Asia-Pacific	Weibo	politics	capture emergent opinions	national
Kinra et al.	2020	*Examining the potential of textual big data for public policy decision-making on driverless cars: A case study from Denmark*	Transport Policy	Twitter	other	expand the scope of survey focus	national

Table 2.1.8: List of publications using social media as a recruitment tool.

Author(s)	**Date**	**Title**	**Source**	**SM**	**Topic**	**Level of analysis**
Bekafigo & McBride	2013	*Who Tweets About Politics?: Political Participation of Twitter Users During the 2011 Gubernatorial Elections*	Social Science Computer Review	Twitter	election	State
Bode & Dalrymple	2014	*Politics in 140 characters or less: Campaign communication, network interaction, and political participation on Twitter*	Journal of Political Marketing	Twitter	politics	national
Vaccari et al.	2014	*Social media and political communication: A survey of Twitter users during the 2013 Italian general election*	Rivista Italiana di Scienza Politica	Twitter	politics	National
Vaccari et al.	2015	*Dual screening the political: Media events,* social media*, and citizen engagement*	Journal of Communication	Twitter	politics	national
Vaccari et al.	2015	*Political expression and action on* social media*: Exploring the relationship between lower-and higher-threshold political activities among Twitter users in Italy*	Journal of Computer-Mediated Communication	Twitter	politics	national
Vaccari et al.	2016	*Of echo chambers and contrarian clubs: Exposure to political disagreement among German and Italian users of Twitter*	Social Media + Society	Twitter	politics	national

Appendix to section 3.1

Annex 3.1.1: List of search-queries to collect the tweets

SEARCH-QUERIES	FRENCH	GERMAN	ITALIAN	HASHTAGS
FROM PRESS REVIEW AND THEORETICAL INPUTS				
1	democratie	demokratie	democrazia	
2	democratique OR democratiques	demokratisch OR demokratisch OR demokratischen	democratico OR democratica	
3	anti-democratique OR antidemocratique	undemokratisch OR undemokratische OR undemokratischen	anti-democratico OR antidemocratico OR anti-democratica OR antidemocratica	
4			antidemocratica	
5	democratie directe	direkte demokratie	democracia diretta	
6				
7	politique	politik	politico OR politica	
8	politicien OR politiciens	politiker OR politikern	politici OR politiche	
9	politicienne OR politiciennes	politikerin OR politikerinnen	deputato OR deputata OR deputati OR deputate	
10			parlamentari	
11	votants	wähler	elettori	
12		wahler		
13	citoyennete	staatsbürgerschaft	cittadinanza	
14		staatsburgerschaft		
15	citoyens	staatsbürgern	cittadini	
16		staatsburgern		
17	citoyen	stimmbürger	cittadino	
18		stimmburger		
19	peuple	volk	popolo	
20			gente	
21	elite OR elite	eliten	élite	
22			elite	
23	elu OR elus OR elue OR elues	volksvertreter OR volksvertretern	rappresentanti	
24	conseiller OR conseillers OR conseillère OR conseillères	abgeordnete OR abgeordneten	rappresentante	
25	parlementaire OR parlementaires			
26	gouvernement	regierung	governo	
27	conseil federal	bundesrat	consiglio federale	

28	parlement	parlament	parlamento
29	assemblee federale	bundesversammlung	assemblea federale
30	representation AND politique	vertretung	
31	constitutionnel OR constitutionnelle	verfassung	costituzionale
32	federal OR federale	federal	federale
33		bundes	federali
34	federales OR federaux	staat OR staatlich	
35		staats	
36	droits populaires	volksrecht	diritti popolari
37	initiative	initiativ	iniziativa
38			iniziative
39	votation	abstimmung	voto
40	votations	abstimmungen	voti
41	votation populaire	volksabstimmung	voto populare
42	consultation populaire	volksentscheidung	
43	scrutin	volksbefragung	
44	votations populaires	volksabstimmungen	voti populari
45		abstimmungskampf	
46	elections	wahlen	elezioni
47	campagne AND electorale	wahlkampagne	campagna AND elettorale
48			
49	bataille AND electorale	wahlkampf	concorrenza AND elettorale
50			
51	referendum	referendum	referendum
52			
53	referendaires		
54			
55	décision populaire	stimmvolk	decisione popolare
56		volksentscheid	
57	institutionnel OR institutionnelle	institutionel OR institutionnelle	istituzionale OR istituzionali
58	vote OR voter	stimmen	votare
59	contre-projet	gegenvorschlag	controprogetto
60	parti politique OR partis politiques	partei OR parteien	partito OR partiti
61	le parti OR les partis		

62	droite AND parti OR partis OR politique	rechte AND partei OR parteien OR politik	destra AND partito OR partiti OR politica	
63	gauche AND parti OR partis OR politique	linke AND partei OR parteien OR politik	sinistra AND partito OR partiti OR politica	
64	centre AND parti OR partis OR politique	mitte AND partei OR parteien OR politik	centro AND partito OR partiti OR politica	
65	extrême OR extreme AND droite OR gauche	rechtsextrem OR linksextrem	estrema AND destra OR sinistra	
66	droits politiques	politische rechte OR politischen rechten	diritti politici	
67	coalition	koalition OR coalition	coalizione	
68	souverainete	soveränität	sovranità OR sovranita	
69	souverain	souverän OR souveränen	sovrano	
FROM HASHTAGS REVIEW				
1				#abst18
2				#abst19
3				#Abstimmungssonntag
4				#AccordCadre
5				#Bundeshaus
6				#Bundesrat
7				#Cantonali19
8				#cantonali2019
9				#chvote
10				#Conseilfédéral
11				#ConsigliodiStato
12				#ConsiglioFederale
13				#contreprojet
14				#controprogetto
15				#deardemocracy
16				#democratie
17				#démocratieparticipative
18				#democrazia
19				#democraziadiretta
20				#Demokratie
21				#DirekteDemokratie
22				#eDemocracy
23				#ef19
24				#ef2019
25				#eGov
26				#eGovernment
27				#elezioni2019
28				#evoting

29		#eVotingMoratorium
30		#facciamolo
31		#gegenvorschlag
32		#granconsiglio
33		#iniziativa
34		#Lobby
35		#Lobbying
36		#nationalratswahlen
37		#NeinzumEURahmenabkommen
38		#NeinzurEU
39		#NoEvoting
40		#nrw19
41		#parlCH
42		#parldigi
43		#participation
44		#politica
45		#politik
46		#politique
47		#Rahmenabkommen
48		#Rahmenvertrag
49		#Rechtsstaat
50		#referendum
51		#sociétécivile
52		#souveraineté
53		#Souveränität
54		#Sovranità
55		#SwissEUrelations
56		#swissregulations
57		#transparence
58		#transparenz
59		#Versicherungslobby
60		#verwaltung
61		#Volksabstimmungen
62		#Volksinitiativen
63		#vorwärts
64		#vot18
65		#vot19
66		#votations
67		#votazioni
68		#wahlCh19
69		#wahlen19
70		#wahlen2019
71		#wahlenCH19

Appendix to section 3.2

Annex 3.2.1: Glossary of features

categories	definitions	labels	examples	sources
Stance, tonality, and entity				
	stance words	stance		custom dictionary
	tonality words	tonality		custom + 'off-the-shelf' dictionaries
	entity words	entity		custom dictionary
	target entity words	main_entity		custom dictionary
	unambiguous stance words for each target	no_doubt		custom dictionary
Linguistic features				
	unspecific words	non_specific	*like, some, somebody, someone*	external lists
	expression of personal view	personal_view	*opinion, view, thinking, personally*	external lists
	vague words	vagueness	*overall, quite, rather, just*	external lists
	frequency adverbs	frequency_adv	*always, frequently, generally, in general*	external lists
	quantifiers	quantifiers	*most, many, much, rare*	external lists
	degree words	degree	*about, almost, apparently, approximately*	external lists
	introductory verbs	introductory_verb	*allege, alleged, appear, appear to be*	external lists
	modal adverbs	modal_adv	*certainly, conceivably, fairly, hopefully*	external lists
	modal nouns	modal_nouns	*assumption, diagnostic, possibility, probability*	external lists
	evidence words	evidences	*actually, assuredly, avowedly, clearly*	external lists
	modal adjective conveying certainty	modal_adj_certain	*certain, clear, definite, suggestive*	external lists
	modal adjective conveying uncertainty	modal_adj_uncertain	*apparent, doubtful, improbable, inconclusive*	external lists
	positive modals	modals_pos	*ca, can, could, may*	external lists
	negative modals	modals_neg	*cannot, cant, can't, couldn't*	external lists
	contrast words	contrast	*after, although, because, before*	external lists
	negators	negators	*aint, ain't, didn't, doesn't*	external lists

	words expressing eventualities	eventuality	*if, imagining, supposing, unless*	external lists
	neutral stance adverbs	stance_adverbs_ntr	*briefly, broadly, confidentially, curiously*	external lists
	negative stance adverbs	stance_adverbs_neg	*candidly, foolishly, incorrectly, oddly*	external lists
	positive stance adverbs	stance_adverbs_pos	*amazingly, appropriately, artfully, cleverly*	external lists
Additional features				
	interrogation mark	interrogation	*'?'*	manual
	supportive words	support	*favour, support, accept, admit*	manual
	rejection words	reject	*against, reject, discard, refuse*	manual
	normative words	normative	*urge, require, lack, need*	manual
	factive indication given by the verb	factive	VBP or VBZ or VVZ	parser
	passive indication given by the verb	passive	VBN	parser
	explicit words (direct person)	explicit	*I or we*	manual
	explicit words (indirect person)	explicit_ext	*you, she, he, they*	manual

Annex 3.2.2: List of the features and their associated weights (ordered by weight and by label)

features	weight	features	weight	features	weight
main_entity_con_stance_neg	3	quantifiers_pos	1	evidences_neg	-1
main_entity_stance_pos	3	stance_adverbs_pos	1	explicit_ext_neg	-1
no_doubt_pos	3	stance_pos	1	explicit_normative_neg	-1
entity_stance_pos	2	support	1	factive_neg	-1
main_entity_con_neg	2	contrast	0	frequency_adv_neg	-1
main_entity_con_no_HoC_neg	2	degree	0	hedges_neg	-1
main_entity_no_HoC_pos	2	entity	0	main_entity_con	-1
main_entity_pos	2	eventual	0	modal_adj_uncertain	-1
reject_neg	2	evidences	0	modals_neg	-1
stance_pos_neg	2	explicit	0	modals_neg_pos	-1
stance_pos_pos	2	explicit	0	modals_pos_neg	-1
support_pos	2	explicit_ext	0	neg_pos_gd	-1
contrast_neg_pos	1	explicit_normative	0	nega_pos	-1
contrast_pos	1	factive	0	negators	-1
degree_pos	1	frequency_adv	0	normative_neg	-1
entity_pos	1	hedges	0	passive_neg	-1
evidences_pos	1	interrog	0	pos_neg_gd	-1
explicit_ext_pos	1	introductory_verb	0	quantifiers_neg	-1
explicit_neg	1	main_entity	0	reject	-1
explicit_normative_pos	1	modal_adv	0	stance_adverbs_neg	-1
explicit_pos	1	modal_nouns	0	stance_neg	-1
factive_pos	1	modals	0	entity_stance_neg	-2
frequency_adv_pos	1	non_specific	0	main_entity_con_no_HoC_pos	-2
hedges_pos	1	normative	0	main_entity_con_pos	-2
modal_adj_certain	1	passive	0	main_entity_neg	-2
modals_neg_neg	1	quantifiers	0	main_entity_no_HoC_neg	-2
modals_pos	1	stance_adv	0	reject_pos	-2
modals_pos_pos	1	stance_adverbs	0	stance_neg_neg	-2
neg_neg	1	stance_adverbs_ntr	0	stance_neg_pos	-2
nega_neg	1	vagueness	0	support_neg	-2
normative_pos	1	contrast_neg	-1	main_entity_con_stance_pos	-3
passive_pos	1	contrast_pos_neg	-1	main_entity_stance_neg	-3
personal_view	1	degree_neg	-1	no_doubt_neg	-3
pos_pos	1	entity_neg	-1		

Annex 3.2.3: Detailed correction rules

Features	Rules
Sum of main features	- If the sum of stance_pos, stance_neg, main_entity, no_doubt_pos, and no_doubt_neg equal 0, then the stance of the tweet is "NONE"
Main entity	- compute main_entity_f_pos = sum(main_entity_pos, main_entity_stance_pos, main_entity_con_neg, main_entity_con_stance_neg, main_entity_no_HoC_pos, main_entity_con_no_HoC_neg) - compute main_entity_f_neg = sum(main_entity_neg, main_entity_stance_neg, main_entity_con_pos, main_entity_con_stance_pos, main_entity_no_HoC_neg, main_entity_con_no_HoC_pos) - if main_entity_f_pos > main_entity_f_neg, then the stance of the tweet is "FAVOR" - if main_entity_f_pos < main_entity_f_neg, then the stance of the tweet is "AGAINST"
Stance	- compute stance_f_pos = sum(stance_pos, stance_pos_pos, stance_neg_neg, entity_stance_pos, main_entity_stance_pos, main_entity_con_stance_neg) - compute stance_f_neg = sum(stance_neg, stance_pos_neg, stance_neg_pos, entity_stance_neg, main_entity_stance_neg, main_entity_con_stance_pos) - if stance_f_pos > stance_f_neg, then the stance of the tweet is "FAVOR" - if stance_f_pos <stance_f_neg, then the stance of the tweet is "AGAINST"
No doubt	- If no_doubt_pos > no_doubt_neg, then the stance of the tweet is "FAVOR" - If no_doubt_pos < no_doubt_neg, then the stance of the tweet is "AGAINST"

Annex 3.2.4: Previous studies using ML on the SemEval dataset

year	authors	title	features	method	average F1 score
2016	P. Sobhani, S. M. Mohammad, S. Kiritchenko	Detecting stance in tweets and analyzing its interaction with sentiment	n-grams, word embeddings, sentiment features	SVM	59.21
2016	J. Ebrahimi, D. Dou, D. Lowd	A joint sentiment-target-stance model for stance classification in tweets	n-grams for sentiment and target variables (in multi-way interactions)	Maximum entropy	61.04
2016	W. Wei, X. Zhang, X. Liu, W. Chen, T. Wang	A specific convolutional neural network system for effective stance detection	Word embeddings and hashtags (use of vote scheme for prediction)	CNN	67.33
2017	Y. Hacohen-Kerner, Z. Ido, R. Ya'akobov	Stance classification of tweets using skip char ngrams	Skip char n-grams	SVM	77.11
2017	S. M. Mohammad, P. Sobhani, S. Kiritchenko	Stance and sentiment in tweets	Sentiment features and additional unlabelled data through distant supervision and word embeddings	SVM	70.30
2017	K. Dey, R. Shrivastava, S. Kaushik	Twitter stance detection-a subjectivity and sentiment polarity inspired two-phase approach	Subjectivity and sentiment features (two-phase feature-driven model)	SVM	74.44
2017	Y. Zhou, A. Cristea, L. Shi	Connecting targets to tweets: Semantic attention-based model for target-specific stance detection	Word embeddings for assessing the implicitness of target in text (attention mechanism)	BiGRU, CNN	67.40
2017	J. Du, R. Xu, Y. He, L. Gui	Stance classification with target-specific neural attention networks	Target-specific information (attention mechanism)	RNN, LSTM	68.79
2018	U. A. Siddiqua, A. N. Chy, M. Aono	Stance detection on microblog focusing on syntactic tree representation	Hahstag segmentation	SVM tree kernel, majority voting	70.03
2018	Q. Sun, Z. Wang, Q. Zhu, G. Zhou	Stance detection with hierarchical attention network	Sentiment, dependency and argument sequences	LSTM	61.00
2018	P. Wei, W. Mao, D. Zeng	A target-guided neural memory model for stance detection in twitter	Tweet-target pars (attention mechanism)	BiGRU	71.04
2019	A. AlDayel, W. Magdy	Your stance is exposed! analysing possible factors for stance detection on social media	Network features	SVM	72.49
2019	S. Zhou, J. Lin, L. Tan, X. Liu	Condensed convolution neural network by attention over self-attention for stance detection in twitter	Word embeddings (stance-indicative words)	CNN	62.45
2019	L. Sun, X. Li, B. Zhang, Y. Ye, B. Xu	Learning stance classification with recurrent neural capsule network	Word embeddings (three layers: embedding, encoding and capsule)	RNN	69.44
2020	A. I. Al-Ghadir, A. M. Azmi, A. Hussain	A novel approach to stance detection in social media tweets by fusing ranked lists and sentiments	Features selected with ranked lists of tf-idf scores and the sentiment information	KNN	76.45
2020	Y. Yang, B. Wu, K. Zhao, W. Guo	Tweet stance detection: A two-stage dc-bilstm model based on semantic attention	Word embeddings and attention layer based on text similarity (two-stage model)	BiLSTM	69.21

2020	M. Lai, A. T. Cignarella, D. I. H. Farías, C. Bosco, V. Patti, P. Rosso	Multilingual stance detection in social media political debates	Stylistic, structural, affective and contextual features	SVM, Logistic regression, CNN, LSTM, biLSTM	64.51 (for Hillary Clinton only)
2020	M. Ahmed, A. N. Chy, N. K. Chowdhury	Incorporating hand-crafted features in a neural network model for stance detection on microblog	Semantic and hand-crafted features	Random forest, MLP, CNN, biLSTM	70.46
2021	S. Vychegzhanin, E. Kotelnikov	A new method for stance detection based on feature selection techniques and ensembles of classifiers	Word and character n-grams, dependency features, target features, stance-indicative features, linguistic features, stylistic features, sentiment features, and word embeddings	Ensemble model	71.24
2021	P. Chen, K. Ye, X. Cui	Integrating n-gram features into pre-trained model: A novel ensemble model for multi-target stance detection	n-grams	Ensemble model	73.77
2021	B. Schiller, J. Daxenberger, I. Gurevych	Stance Detection Benchmark: How Robust is Your Stance Detection?	Word embeddings (single- and multi-dataset models)	BERT versus MT-DNN	0.67
2022	L. H. X. Ng, K. M. Carley	Is my stance the same as your stance? A cross validation study of stance detection datasets	Tweet body (single- and multi-dataset, and leave-one-out models)	BERT models	0.71

Appendix to section 4.1

Annex 4.1.1: contribution of variables and modalities to the formation of the MCA axes.

	modalities	dim.1	dim.2	freq
Chating 2013	*no*	1	0	1722
	yes	6.4	0.1	283
	Total	7.5	0.1	2005
Chating 2016	*no*	1.1	0	1659
	yes	5.5	0	346
	Total	6.6	0	2005
Freenews 2013	*no*	2.2	1.4	918
	yes	1.8	1.2	1087
	Total	4	2.6	2005
Freenews 2016	*no*	1.3	1.9	923
	yes	1.1	1.6	1082
	Total	2.4	3.5	2005
Magazines 2013	*no*	0.3	7.8	924
	yes	0.2	6.6	1081
	Total	0.5	14.4	2005
Magazines 2016	*no*	0.4	8.7	980
	yes	0.4	8.3	1025
	Total	0.8	17	2005
Newspaper 2013	*no*	2.8	14	332
	yes	0.5	2.8	1673
	Total	3.3	16.8	2005
Newspaper 2016	*no*	3.1	11.6	410
	yes	0.8	3	1595
	Total	3.9	14.6	2005
Online news 2013	*no*	6.6	5	690
	yes	3.5	2.6	1315
	Total	10.1	7.6	2005
Online news 2016	*no*	6.6	5.6	558
	yes	2.6	2.1	1447
	Total	9.2	7.7	2005
	no	1.3	0.4	1577

Radio and TV	*yes*	4.9	1.6	428
2013	*Total*	6.2	2	2005
Radio and TV	*no*	1.7	0.6	1498
2016	*yes*	5	1.7	507
	Total	6.7	2.3	2005
Social media 2013	*Facebook/Twitter*	9.9	2.1	775
	Other	1.4	3.5	97
	None	8.7	0.4	1133
	Total	20	6	2005
Social media 2016	*Facebook/Twitter*	8.9	2.1	801
	Other	1.1	2.9	146
	None	8.9	0.4	1058
	Total	18.9	5.4	2005

Note: Significance levels defined as $^{**}p < 0.01$, $^{*} < 0.05$, ^{a} $p < 0.08$; N = 1416.

Appendix to section 4.2

Annex 4.2.1: text classification of policy issues

To identify the policy issues discussed in the tweets included in our corpus, we set up a classification system to assign tweets to policy areas. The policy areas are similar to those used in the Swiss election surveys *Selects* (*Swiss Electoral Studies*) conducted in 2015 and 2019 (see: https://forsbase.unil.ch/).

We rely on two supervised ensemble machine learning algorithms to classify the main policy issue content for the tweets. To train the ensemble model, we selected a sample of tweets based on a dictionary approach. The dictionary consists of manually curated terms from two main sources. The first source relies on open-ended survey answers to the item "*What is the most important political problem affecting Switzerland nowadays?*" The answers are manually assigned by the survey team into 17 policy issue categories (see first column of the Table below). For each category, we extract the most frequent and the most discriminative words. We then review the entire list of words and keep only the unambiguous ones. The second source considers the 5,000 most frequent German and French hashtags in the corpus of tweets. Again, we review the entire list of hashtags and keep only those that can be unambiguously assigned to a policy issue category. The final dictionary is composed of 1,987 German and French words.

The training dataset thus involves tweets that are labelled by the dictionary using the *liwcalike()* function from the *quanteda* package of the *R* programming language (Benoit et al., 2018). The training dataset is composed of 131,525 German and 27,819 French tweets. To decrease the computational time, we further randomly sample 20,000 tweets from each language to train the ensemble models separately for both languages. The algorithms used for the ensemble model are *gradient boosting machine* and *random forest* using the *R* programming language *h2o* package (LeDell, 2020).

Before we can use the training data for classification, we transform the texts into a numerical format. We use the approach *Word2Vec* that takes the large text corpus (training data). It produces a vector space of multiple dimensions (600 in our case) where each word is assigned to a corresponding vector space. Thereby, words with similar contexts have a close proximity to one another in this vector space. We use the implementation provided in *H2O* (see: https://www.h2o.ai) that maximizes the classification of a word based on another word in the same context using the *Skip-Gram* model. The ensemble model uses the sum of all probabilities of the individual policy issues classification and selects the class with the highest probability. The results of the classification models are displayed in the Table below by policy issue categories and by the original tweet language.

Policy issues	German precision	German recall	German F1	German accuracy	French precision	French recall	French F1	French accuracy
Class: agriculture	0.54	0.42	0.47	0.71	0.80	0.66	0.72	0.83
Class: economy	0.46	0.38	0.41	0.67	0.63	0.72	0.67	0.84
Class: education & culture	0.45	0.24	0.31	0.61	0.76	0.66	0.70	0.82
Class: environment & energy	0.60	0.75	0.66	0.84	0.78	0.78	0.78	0.88
Class: eu, europe	0.82	0.81	0.81	0.90	0.78	0.75	0.77	0.87
Class: finances & taxes	0.43	0.05	0.08	0.52	0.80	0.66	0.72	0.83
Class: gender issues & discrimination	0.63	0.61	0.62	0.79	0.71	0.77	0.74	0.87
Class: immigration & asylum	0.47	0.44	0.45	0.71	0.89	0.55	0.68	0.78
Class: international relations & conflicts, foreign policy & army	0.61	0.63	0.62	0.79	0.69	0.76	0.72	0.87
Class: labour market	0.67	0.18	0.29	0.59	0.86	0.69	0.76	0.84
Class: law & order	0.44	0.56	0.49	0.74	0.64	0.66	0.65	0.82
Class: other	0.50	0.45	0.47	0.71	0.70	0.75	0.72	0.86
Class: political system, parties & politicians	0.48	0.67	0.56	0.79	0.70	0.79	0.74	0.87
Class: public health	0.56	0.45	0.50	0.72	0.74	0.79	0.76	0.89
Class: public services & infrastructure	0.56	0.50	0.52	0.74	0.81	0.81	0.81	0.90
Class: regions & national cohesion	1.00	0.14	0.25	0.57	0.55	0.23	0.32	0.61
Class: social security/welfare state	0.70	0.54	0.61	0.76	0.79	0.70	0.74	0.85

Annex 4.2.2: Descriptive statistics of our corpus of tweets

	2011-15			2015-19		
	Election campaign	**Whole period**	**Citizens (Selects W3)**	**Election campaign**	**Whole period**	**Citizens (Selects W3)**
Topics						
Economy	24%	25%	6%	19%	21%	3%
Environment & energy	45%	40%	4%	47%	43%	31%
EU, Europe	10%	12%	11%	15%	13%	19%
Immigration & asylum	12%	14%	53%	8%	11%	10%
Public health	5%	5%	1%	6%	5%	5%
Social security & welfare State	4%	4%	6%	5%	7%	21%
Number of tweets	16405	42632		22897	90858	
Tweeting frequency	56.7	29.3		78.4	62.1	
Number of retweets	67851	124215		392054	1271588	
Number of likes	16138	31294		194736	394876	

Annex 4.2.3: Description of the categories for manual coding

Category	Examples	Description
Politics		
committee/campaign page	konzern_vi, Atomausstieg_JA, greeneconomy_ch	page dedicated to a voting object (e.g. referendum or initiative)
government	alain_berset, Violapamher, ignaziocassis	government accounts including accounts of ministers and government organizations
foreign party/movement	Die_Gruenen, CDU, APFfrancophonie	page from foreign parties
foreign politician	sven_giegold, JunckerEU cem_oezdemir	foreign politicians
lobby	cloudista, GSoASchweiz, Kinderlobbyist	lobbyists and other accounts clearly focused on public interests,
national politician	bglaettli, zac1967, RegulaRytz	Swiss politician with a federal mandate
officials	KasparSchuler, BR_Sprecher, fljan	political officials including ambassadors, general secretary of organizations, or directors of public affairs
party/movement	GrueneCH, spschweiz, FDP_Liberalen	page from Swiss parties at the federal, cantonal, or local level
political institutions/ambassy	ParlCH, BAG_OFSP_UFSP, vbs_swiss_un	page from political institutions (e.g. courts, ambassies, etc) in Switzerland or abroad
politician	StrebelLuca, Michael_Koepfli, benoitgaillard	Swiss politician who are not in the list of politicians with a federal mandate
Public organisations		
Association/organisation/ONG/unions/platform	alliance_F, swisscleantechD, Amnesty_Schweiz	group representation in associations, as well as collective organisations, nongovernmental organizations, or trade unions
canton/region	kanton_bern, Kanton_BL, Graubunden	cantonal or regional institutions (e.g. cultural office, local parliament, etc)
Expertise		
consultant/communication manager	Mark_Balsiger, dani_graf, redder66	independent advisory service providers (especially in the communication domain)
experts/scientists	Claudelongchamp, vecirex, schlegel_stefan	science and research experts, as well as professors and other academic employees, but also non-university private institutes or centres
lawyer	martinsteiger, MarcSchinzel, sebastienfanti	lawyers or jurists
university	ETH, EPFL, EUErasmusPlus	university faculty, institutes, labs, or offices

Media		
foreign journalist	StephanIsrael, christof_moser, SebRamspeck	foreign journalists (including redactors, moderators, editorials, etc)
foreign media	spiegelonline, welt, lemondefr	foreign official media accounts (including television, newspapers, radio, etc)
journalist	thomas_ley, alex_baur, SandroBrotz	journalists (including redactors, moderators, editorials, etc)
media	NZZ, Tagesanzeiger, srfnews	official media accounts (including television, newspapers, radio, etc)
Business		
enterprise	migros, RailService, verkehrsclub	private business or corporations
head of business/entrepreneur	andre_schaad, Philipp_Straehl, ahugi	leaders of private business or corporations
think-tank	Avenir_Suisse, foraus, sanudurabilitas	group of experts within a private law structure
Civil society		
activist	klimastreik, postcovid_CH, tiretpointtiret	proclaimed activists (clear involvement in collective action)
artist/exhibitions/museums/etc	PattiBasler, michaelelsener, kindlimann	artistic accounts (painters, sculptors, photographers, etc.), museums, galleries, or exhibitions
citizen		lay citizens
foreign citizen		lay foreign citizens
sport	bwertli, BSC_YB, VolleyNaefels	sport organisations or athletes
Not Applicable		
suspended or not found		profile could not be retrieved (removal of the account)
no description		no profile description nor personal url is provided

Annex 4.2.4: Poisson regression models without the responsiveness to public concerns.

	Politicians' ranking (whole)			Politicians' media coverage (election)			Politicians' ranking (election)		
	Std. Coef. (Std. Error)	P-value		Std. Coef. (Std. Error)	P-value		Std. Coef. (Std. Error)	P-value	
Constant	0.879 (0.226)	0.000	***	-1.618 (0.542)	0.003	**	0.656 (0.221)	0.003	**
Communication style:									
Proportion of replies	-0.003 (0.003)	0.171		-0.001 (0.006)	0.854		**0.006 (0.003)**	**0.017**	*
Proportion of replies to journalists	-0.002 (0.003)	0.454		-0.005 (0.005)	0.354		-0.003 (0.003)	0.255	
Proportion of replies to media	-0.012 (0.004)	0.005	**	-0.014 (0.01)	0.184		-0.003 (0.003)	0.364	
Proportion of replies to national politicians	-0.004 (0.002)	0.114		0.006 (0.003)	0.033	*	-0.003 (0.002)	0.115	
Proportion of replies to local politicians	0.003 (0.003)	0.304		-0.003 (0.004)	0.374		-0.003 (0.002)	0.150	
Proportion of replies to parties	**-0.01 (0.005)**	**0.033**	*	-0.012 (0.008)	0.150		-0.003 (0.003)	0.312	
Proportion of replies to citizens	-0.003 (0.002)	0.258		-0.008 (0.006)	0.139		-0.007 (0.003)	0.009	**
Responsiveness to public concerns	not included			not included			not included		
Proportion of links	-0.001 (0.002)	0.724		-0.003 (0.004)	0.464		**0.000 (0.002)**	**0.968**	
Reactions to politicians' tweets:									
Proportion of retweeted politicians' messages	0.003 (0.001)	0.038	*	0.009 (0.004)	0.011	*	**0.006 (0.001)**	**0.000**	***
Proportion of favourited politicians' messages	-0.002 (0.001)	0.185		0.007 (0.003)	0.029	*	**-0.004 (0.001)**	**0.005**	**
Legislature dummy:									
Tweeting frequency	-0.006 (0.023)	0.784		-0.024 (0.09)	0.791		-0.035 (0.024)	0.149	
Legislature dummy: 2019-22 (ref. 2015-19)	-0.298 (0.087)	0.001	***	0.075 (0.43)	0.861		-1.339 (0.18)	0.000	***
Tweeting frequency x legislature dummy	0.099 (0.031)	0.001	**	0.316 (0.093)	0.001	***	0.193 (0.035)	0.000	***
Control variables:									
Gender: woman (ref. man)	0.088 (0.061)	0.147		0.139 (0.122)	0.253		0.072 (0.062)	0.245	
Regions: Latin (ref. German-speaking)	-0.063 (0.08)	0.432		-0.437 (0.18)	0.015	*	**-0.015 (0.08)**	**0.855**	
Left-right position	0.081 (0.012)	0.000	***	0.013 (0.024)	0.591		0.097 (0.013)	0.000	***
Incumbent: yes (ref. no)	0.975 (0.068)	0.000	***	-0.554 (0.16)	0.001	***	1.037 (0.067)	0.000	***
National Council (ref. Council of States)	-0.123 (0.135)	0.360		0.441 (0.17)	0.009	**	-0.124 (0.139)	0.375	
Adjusted R2:	0.25 (25%)			0.54 (54%)			0.28 (28%)		
Number of observations:	339 observations			321 observations			321 observations		

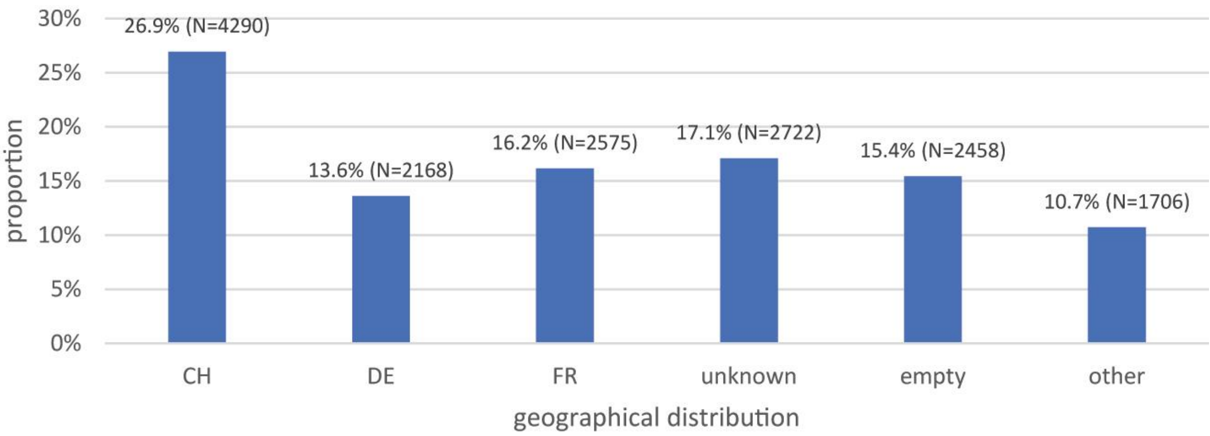
Note: significativity levels read as '***' for p<0.001, '**' for p<0.01, and '*' for p<0.05; changes compared to the model including the variable "responsiveness to public concerns" are marked in bold.

Appendix to section 4.3

Annex 4.3.1: Description of the Political Accounts by Left–Right Position

Party Abbreviation	Left-right score	Number of accounts	Number of tweets	Tweeting frequency
POP (called 'extreme left')	1	20	188	9.0
PdA	2	/	/	/
CSP	3	3	14	4.7
SP	4	199	2176	10.9
Grüne	5	107	927	8.4
Sub-total left		**329**	**3305**	
GLP	6	65	242	3.7
EVP	7	16	57	3.6
CVP	8	62	283	4.6
BDP	9	12	53	4.4
Sub-total centre		**15**	**635**	
FDP	10	67	207	3.1
SVP	11	41	187	4.6
Lega	12	/	/	/
MCG	13	/	/	/
EDU	14	2	12	6.0
Sub-total right		**110**	**406**	

Annex 4.3.2: Distribution of Accounts According to Geolocation



Appendix to section 5.1

Annex 5.1.1: Topic modelling results (manual labelling, topic weight, and top terms)

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
1	patient needs	patients	patients, can, need, know, don, like, think, just, people, healthcare	0,04533	0,02712	0,02367	0,02962	0,01770	0,03451
2	/	/	health, forward, looking, great, digital, amp, today, event, day, innovation	0,03731	0,01924	0,02988	0,02225	0,02373	0,02409
3	/	/	time, patients, get, will, can, now, long, need, just, right	0,03511	0,01946	0,01824	0,02241	0,01382	0,02491
4	digital transformation	innovations	digital, healthcare, health, new, technology, will, transformation, future, care, innovation	0,03449	0,02283	0,02094	0,02454	0,02099	0,02080
5	public system	health system	health, public, amp, people, medicine, science, need, will, just, don	0,03322	0,01903	0,01892	0,02383	0,01199	0,02978
6	patient technology/support	patients	care, patients, health, can, help, home, providers, technology, learn, patient	0,03059	0,02080	0,01402	0,01531	0,02069	0,01455
7	specialists (CEO/professor/etc)	actors	health, amp, ceo, director, prof, president, healthcare, professor, john, founder	0,02784	0,01502	0,01352	0,01365	0,01640	0,01168
8	crisis response	security	health, covid, pandemic, crisis, response, coronavirus, public, healthcare, care, can	0,02745	0,01729	0,01387	0,01614	0,01587	0,01518
9	services/programs	innovations	health, care, nhs, digital, amp, across, innovation, support, new, social	0,0274	0,00660	0,03375	0,01683	0,01926	0,01425

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
10	patient teams	patients	great, work, patient, amp, team, patients, safety, thanks, thank, see	0,02728	0,01248	0,02057	0,01362	0,01437	0,01846
11	care system	health system	care, health, healthcare, value, based, patient, system, amp, approach, systems	0,02711	0,01772	0,01326	0,01424	0,01576	0,01562
12	report/litterature	information	health, new, read, report, research, article, amp, paper, published, medicine	0,02711	0,01390	0,01688	0,01438	0,01368	0,01804
13	webinars	information	join, webinar, register, amp, health, free, will, learn, now, next	0,02657	0,01453	0,01933	0,01248	0,02085	0,01241
14	patient safety/experience	patients	patient, care, improve, can, healthcare, outcomes, safety, amp, experience, quality	0,02643	0,01989	0,01279	0,01437	0,01871	0,01316
15	partient access to records	patients	patients, can, app, health, patient, access, online, video, nhs, help	0,0262	0,01354	0,01888	0,01666	0,01720	0,01534
16	patient care	patients	patients, can, data, make, amp, making, patient, need, care, decision	0,02617	0,01544	0,01450	0,01462	0,01459	0,01509
17	quality systems (social needs)	health system	health, care, amp, access, social, need, services, quality, systems, communities	0,02596	0,01563	0,01268	0,01314	0,01556	0,01672
18	impact of technology	innovations	healthcare, technology, make, will, can, future, look, change, like, impact	0,02554	0,01447	0,01434	0,01575	0,01317	0,01487
19	telemedicine	innovations	telehealth, telemedicine, care, virtual, patients, remote, visits, covid, pandemic, patient	0,0249	0,02012	0,01027	0,01788	0,01459	0,01553
20	health solutions	innovations	healthcare, amp, challenges, together, solutions, innovation, health, technology, industry, can	0,02394	0,01265	0,01518	0,01232	0,01515	0,01132

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
21	mental health	health domains	health, mental, can, help, day, awareness, mentalhealth, week, world, support	0,02246	0,01308	0,01549	0,01222	0,01664	0,01344
22	health panels/discussions	information	join, health, will, amp, register, healthcare, live, panel, today, don	0,02229	0,01704	0,01301	0,01635	0,01675	0,01192
23	european public system	health system	health, digital, global, amp, european, public, systems, europe, national, policy	0,02198	0,00887	0,01793	0,01323	0,01408	0,01328
24	research papers	information	health, digital, new, evidence, study, review, interventions, based, use, research	0,02177	0,01058	0,01556	0,00995	0,01158	0,01708
25	artificial intelligence	innovations	tech, health, intelligence, artificial, healthcare, learning, digital, technology, machine, via	0,02039	0,01281	0,01786	0,02307	0,01202	0,01561
26	big data	innovations	data, health, analytics, use, research, patient, can, big, real, using	0,01953	0,01197	0,01139	0,01099	0,01158	0,01107
27	partnerships	industry	health, excited, proud, team, announce, new, healthcare, work, see, part	0,0195	0,01160	0,01056	0,01118	0,01157	0,01054
28	blockchain industry	industry	healthcare, technology, technologies, blockchain, industry, market, via, trends, will, tech	0,01896	0,01362	0,01181	0,01342	0,01297	0,01108
29	hiring opportunities	industry	health, team, apply, looking, join, amp, research, opportunity, interested, work	0,01863	0,00921	0,01749	0,00992	0,01342	0,01290
30	/	/	health, get, day, one, time, just, week, can, amp, today	0,01848	0,01039	0,00981	0,01159	0,00918	0,01090
31	insurance	health system	health, healthcare, care, insurance, survey, patients, costs, cost, new, study	0,01839	0,01636	0,00645	0,01440	0,01033	0,01176

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
32	patient health record	patients	health, data, patient, records, ehr, electronic, record, systems, information, platform	0,01822	0,01264	0,01016	0,01211	0,01132	0,01126
33	world health (future of health)	innovations	health, role, amp, play, people, healthy, can, future, work, world	0,01818	0,00906	0,01155	0,00878	0,01014	0,01049
34	health workers (e.g., nurses)	actors	health, day, thank, nurses, care, patients, healthcare, workers, amp, world	0,01813	0,01355	0,00922	0,00960	0,01230	0,01176
35	family doctor (physician)	actors	patient, patients, experience, family, care, doctor, voice, physician, engagement, can	0,01811	0,01279	0,00827	0,01203	0,00918	0,01182
36	virtual events	information	health, conference, will, event, join, healthcare, register, annual, visit, week	0,01805	0,01036	0,01358	0,01008	0,01547	0,00801
37	risks (pandemic, mental, etc)	security	health, mental, social, people, amp, issues, risk, impact, can, covid	0,01798	0,01138	0,01005	0,01022	0,00975	0,01387
38	tracing (for covid)	covid	health, public, covid, coronavirus, cases, testing, contact, new, will, tracing	0,01784	0,01111	0,00860	0,01256	0,00898	0,01362
39	learning ressources	innovations	health, care, learning, resources, amp, new, free, available, healthcare, professionals	0,01775	0,00715	0,01598	0,00856	0,01257	0,00886
40	project development	industry	research, health, new, amp, funding, innovation, will, support, projects, project	0,01771	0,00775	0,01430	0,00798	0,01295	0,00854
41	school/university	education	medicine, health, students, medical, school, amp, program, university, faculty, research	0,01771	0,01441	0,00568	0,00616	0,01132	0,01008
42	equity (gender/race/etc)	education	health, women, amp, black, gender, racism, equity, disparities, sexual, diversity	0,01767	0,01223	0,00833	0,00886	0,00978	0,01203

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
43	hospital/emergency	actors	patients, hospital, care, covid, hospitals, patient, home, emergency, icu, new	0,01729	0,01052	0,00996	0,00991	0,00929	0,01170
44	surveys	information	please, survey, health, help, share, want, know, can, take, get	0,01728	0,00762	0,01275	0,00900	0,00957	0,01146
45	patient stories	patients	patient, patients, one, just, life, people, like, story, medicine, amp	0,01717	0,01089	0,00830	0,01233	0,00664	0,01388
46	problem solving	innovations	healthcare, system, health, technology, problem, need, one, can, care, change	0,01646	0,01004	0,00860	0,01130	0,00734	0,01161
47	precision medicine (cell/genomics/etc)	health domains	medicine, precision, new, technology, research, amp, technologies, cell, disease, cancer	0,01618	0,00973	0,01036	0,01008	0,00957	0,00999
48	challenge	innovations	health, apply, challenge, now, digital, tech, innovation, deadline, open, healthcare	0,01599	0,00972	0,01667	0,01056	0,01481	0,00846
49	cardiovascular diseases	health domains	patients, heart, disease, risk, study, can, stroke, failure, chronic, diseases	0,01584	0,01093	0,00919	0,00928	0,00961	0,01101
50	medicare/medicaid	health system	health, telehealth, medicare, state, care, new, healthcare, medicaid, services, will	0,01526	0,01606	0,00433	0,01286	0,00973	0,00918
51	gratulations/awards	information	health, award, awards, congratulations, year, innovation, best, tech, healthcare, winners	0,01432	0,00964	0,01101	0,00888	0,01182	0,00787
52	health costs/fundings	health system	health, year, healthcare, million, billion, report, per, funding, growth, digital	0,01431	0,00966	0,00746	0,00994	0,00770	0,00958

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
53	women in tech	education	health, tech, healthcare, companies, digital, big, via, digitalhealth, women, startups	0,01431	0,01057	0,00824	0,01164	0,00743	0,01033
54	/	/	years, last, year, week, one, months, ago, two, past, next	0,01424	0,00772	0,00666	0,00744	0,00695	0,00707
55	startups	industry	health, tech, amp, startups, innovation, companies, medtech, digitalhealth, healthcare, healthtech	0,01421	0,00684	0,01189	0,01031	0,01077	0,00680
56	medical market (devices/regulations/etc)	industry	medtech, medical, amp, tech, device, med, industry, devices, companies, innovation	0,01403	0,00577	0,01107	0,01151	0,00915	0,00630
57	funding platforms	education	health, digital, tech, startup, via, million, raises, funding, healthcare, platform	0,01398	0,01080	0,00957	0,01212	0,00862	0,00999
58	donations/conferences	information	health, conference, now, register, don, event, miss, join, get, digital	0,01384	0,00781	0,01216	0,00918	0,01240	0,00714
59	(cyber)security	security	data, privacy, security, health, healthcare, patient, cybersecurity, cyber, information, amp	0,01376	0,01032	0,00784	0,01112	0,00853	0,00879
60	aging	health domains	healthy, hearing, health, aging, can, ageing, older, sleep, amp, exercise	0,01335	0,00891	0,00754	0,00615	0,00907	0,00779
61	/	/	podcast, health, listen, episode, healthcare, digital, ceo, new, amp, latest	0,0132	0,00975	0,00713	0,01352	0,00859	0,00670
62	covid testing	covid	patients, covid, test, study, testing, positive, symptoms, coronavirus, tests, new	0,01316	0,00998	0,00815	0,00918	0,00718	0,01246

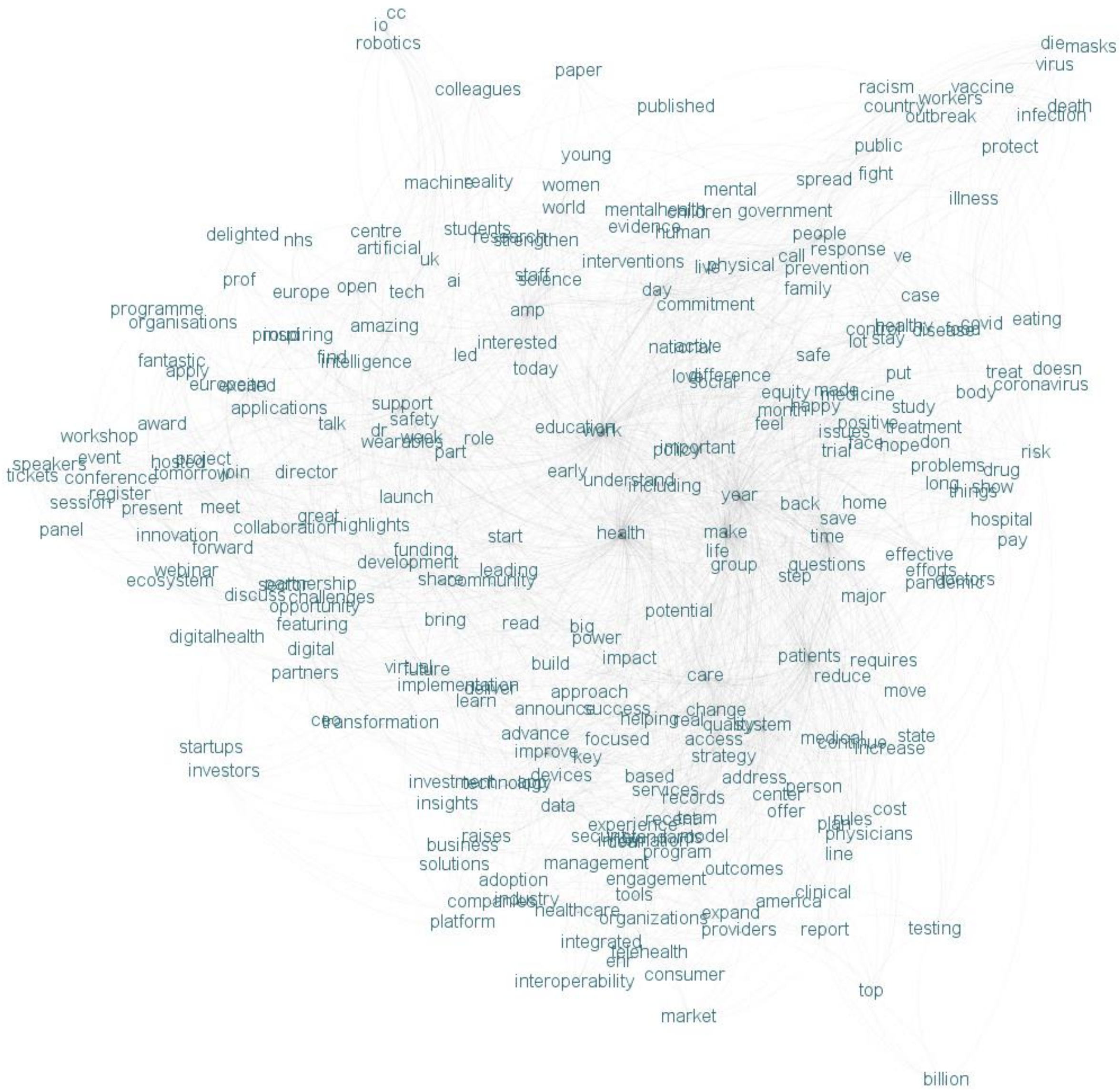
n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
63	healthcare company (e.g., Amazon cited)	industry	health, healthcare, company, amazon, care, digital, new, telehealth, via, united	0,0131	0,01404	0,00554	0,01918	0,00659	0,01023
64	head of medicine (chief/officer/director/etc)	actors	chief, officer, health, healthcare, director, medical, ceo, amp, technology, digital	0,01231	0,00819	0,00609	0,00739	0,00817	0,00540
65	youth wellbeing	health domains	health, people, mental, young, support, amp, help, can, social, services	0,01195	0,00390	0,01238	0,00651	0,00833	0,00700
66	safety (covid distancing)	covid	stay, healthy, home, keep, safe, can, health, amp, help, people	0,01186	0,00759	0,00633	0,00697	0,00701	0,00796
67	children health	health domains	health, children, mental, amp, child, school, kids, youth, schools, young	0,01186	0,00735	0,00611	0,00455	0,00786	0,00634
68	depression/anxiety	health domains	mental, long, health, term, patients, depression, study, anxiety, therapy, can	0,0116	0,00724	0,00707	0,00633	0,00666	0,00797
69	latest news	information	latest, newsletter, healthcare, blog, health, read, technology, news, post, featuring	0,01122	0,00791	0,00659	0,00611	0,00786	0,00679
70	food/diet/nutrition	health domains	healthy, food, health, diet, can, nutrition, eating, amp, eat, foods	0,01115	0,00879	0,00580	0,00634	0,00816	0,00826
71	surgery	health domains	patients, patient, cancer, surgery, treatment, new, first, therapy, brain, technology	0,01112	0,00836	0,00784	0,00719	0,00780	0,00801
72	clinical trial	education	patient, clinical, research, trials, patients, trial, amp, engagement, involvement, public	0,01069	0,00575	0,00704	0,00668	0,00576	0,00648
73	disrupting issues/innovations	innovations	latest, thanks, health, daily, healthcare, technology, innovation, news, disrupting, global	0,01068	0,01236	0,00530	0,00657	0,01144	0,01554

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
74	health and environment (climate/air/etc)	health domains	health, amp, global, climate, change, public, diseases, world, disease, air	0,01023	0,00592	0,00621	0,00530	0,00717	0,00608
75	covid staff	covid	workers, healthcare, care, ppe, covid, health, amp, help, patients, support	0,00964	0,00604	0,00505	0,00568	0,00569	0,00653
76	wearables	innovations	health, monitoring, wearable, wearables, patient, new, data, test, apple, devices	0,00949	0,00769	0,00578	0,00734	0,00649	0,00593
77	headlines	information	health, today, thanks, hit, watch, connect, headlines, edition, stay, amp	0,00946	0,00750	0,00491	0,01012	0,00484	0,00674
78	health technology	innovations	technology, health, design, digital, following, thanks, user, technologies, solutions, game	0,00943	0,00473	0,00788	0,00546	0,00670	0,00460
79	NHS	health system	nhs, health, read, digital, trust, digitalhealth, news, full, story, healthtech	0,00899	0,00328	0,01024	0,01196	0,00619	0,00365
80	mobile apps	innovations	health, mental, apps, app, help, digital, support, can, mobile, tools	0,0089	0,00553	0,00590	0,00479	0,00603	0,00501
81	chronic conditions/diseases	health domains	diabetes, patients, chronic, management, health, disease, conditions, type, people, care	0,00854	0,00491	0,00508	0,00422	0,00530	0,00458
82	international relations (Sinai/Australia/Germany/Canada/etc)	industry	health, digital, minister, sinai, national, new, australia, system, today, first	0,00844	0,00294	0,00467	0,00629	0,00438	0,00504
83	saving lifes (with technology)	innovations	lives, life, help, save, people, technology, healthier, can, saving, live	0,00797	0,00365	0,00538	0,00417	0,00480	0,00366
84	cancer	health domains	cancer, patients, treatment, breast, screening, amp,	0,00773	0,00562	0,00346	0,00318	0,00518	0,00425

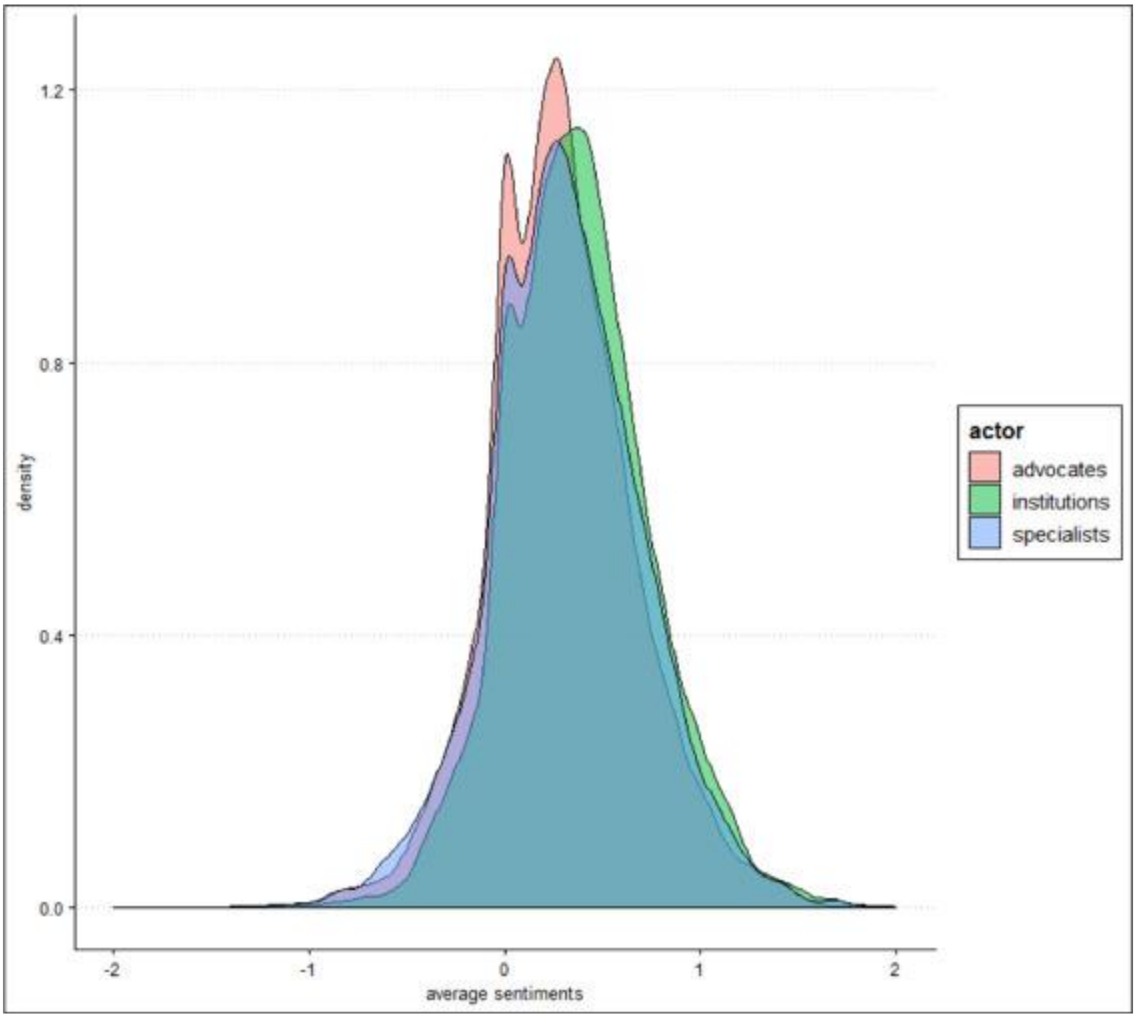
n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
			diagnosis, oncology, skin, awareness						
85	interoperability	innovations	health, interoperability, data, patient, information, access, healthcare, rule, onc, amp	0,00758	0,00803	0,00230	0,00462	0,00586	0,00382
86	medications/substances (for pain and addictions)	health domains	patients, use, medication, opioid, treatment, adherence, drug, addiction, pain, substance	0,00755	0,00664	0,00288	0,00349	0,00533	0,00422
87	engineering	innovations	health, research, science, amp, technology, university, students, engineering, based, course	0,00752	0,00293	0,00588	0,00295	0,00474	0,00462
88	developing countries (india/africa/etc)	health system	health, india, amp, africa, healthcare, bharat, ayushman, pmjay, corona, south	0,00723	0,00196	0,00220	0,00252	0,00487	0,00500
89	/	/	happy, year, health, healthy, new, family, day, holiday, christmas, good	0,00694	0,00493	0,00347	0,00389	0,00484	0,00431
90	vaccines (covid/flu/etc)	covid	vaccine, vaccines, health, covid, flu, vaccination, get, amp, public, workers	0,00692	0,00472	0,00335	0,00459	0,00386	0,00477
91	generic drugs	health domains	medicines, good, latest, generic, drug, use, fda, drugs, safety, thanks	0,00653	0,00272	0,00470	0,00412	0,00499	0,00337
92	smoking	health domains	health, media, social, digital, publications, smoking, today, quit, web, twitter	0,00594	0,00545	0,00234	0,00410	0,00540	0,00251
93	training and education (simulation/imaging/radiology/etc)	education	technology, simulation, healthcare, training, radiology, imaging, aid, medical, education, amp	0,00576	0,00435	0,00278	0,00294	0,00436	0,00222
94	maternity/motherhood	health domains	healthcare, icymi, health, harlow, women, maternal,	0,00519	0,00534	0,00179	0,00165	0,00525	0,00193

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
			pregnant, pregnancy, mothers, babies						
95	covid in UK (stocks management)	covid	via, coronavirus, healthinnovations, health, pharma, stocks, healthcare, brexit, stories, news	0,0051	0,00246	0,00756	0,00302	0,00296	0,00678
96	eyes (ophthalmology/optometry/etc)	health domains	eye, vision, video, patient, watch, peek, amp, ophthalmology, patients, videos	0,0046	0,00270	0,00281	0,00202	0,00289	0,00270
97	(international) congress	information	digital, now, world, book, health, congress, online, london, healthcare, conference	0,00317	0,00075	0,00654	0,00117	0,00499	0,00100
98	private clinics (arrayit)	health system	healthcare, arrayit, san, sales, team, life, usa, sciences, top, markets	0,00289	0,00392	0,00086	0,00139	0,00341	0,00090
99	chronic pain (hand/back/shoulders/etc)	health domains	pain, health, free, call, randolph, screening, hand, foot, back, register	0,00266	0,00288	0,00120	0,00099	0,00283	0,00110
100	services (premium)	patients	health, clinical, healthcare, services, testing, sales, team, providing, premium, reports	0,00148	0,00193	0,00123	0,00059	0,00233	0,00038

Annex 5.1.2: Semantic relations based on the top terms in our corpus of tweets.



Annex 5.1.3: Density plot of sentiment score by actor type.



Appendix to section 5.2

Annex 5.2.1: Phases of the Covid-19 pandemic in Switzerland

Phase	MOSAïCH waves	Label	Covid-19 wave	Explanation
January to March 2020		W0		Detection of the first cases and a rapid increase in the number of cases.
March to April 2020		W1	1st wave	Peak in the epidemic and the establishment of a state of emergency. At the end of March 2020, the Confederation appoints a scientific advisory board called the *Taskforce* to find the best approach to overcoming the pandemic. A first measure establishes a policy on testing, isolation, and quarantine.
end of April to mid-June 2020	1st wave (start)	N1		Decrease in the number of cases and a relaxation of measures. The Confederation demonstrates its interest in an official mobile contact tracking application *SwissCovid (*launch of the testing phase on May 25, and public availability of the application on June 25). The application had been downloaded by 2.1 million inhabitants, or 25% of the population, 6 weeks after its release.
mid-June to September 2020	1st wave (end)	N2		End of the state of emergency and a further increase in cases. A nationwide face-mask rules for public transportation is introduced on July 6th.
October to December 2020	2nd wave (start & end)	W2	2nd wave	Tracing is ensured in the face of the very big increase in cases during the second wave of infections in the fall of 2020. At the end of December 2020, Switzerland begins a national vaccination campaign against Covid-19.
January to May 2021	3rd wave (start & end)	W3	3rd wave	Teleworking becomes compulsory, and shops not selling everyday consumer goods are closed. The third wave takes place in April 2021. The relaxations will be phased in until the end of May 2021.
mid-May to August 2021		N4		Public spaces (e.g. restaurant's terraces, cinemas, theatres or sport stadiums) partially reopen. Wearing masks and collecting contact details remains compulsory.
September to mid-November 2021		W4	4th wave	The fourth wave takes place in September 2021. The Federal Council announces that the covid certificate will be mandatory in indoor restaurants and other public spaces.
mid-November to mid-December 2021		W5	5th wave	Switzerland observes the first case of the *Omicron* variant which also coincide with the beginning of the fifth pandemic wave.
mid-December 2021		N5		The "*2G*" rule (for *geimpft* and *genesen*, i.e. vaccinated or cured) becomes the norm on December 20 for visits to public establishments and the so-called "2G+" rule comes into force implying that the last dose of vaccine must have been given within the last four months.

Annex 5.2.2: List of Twitter accounts included as seed actors

Groups of seed actors	List of accounts
University research centres	ETH_Rat, Conseil_EPF, ETH, EPFL, snf_ch, fns_ch, CH_universities, academies_ch, SAGW_CH, UniBasel, unibern, UniFreiburg, UNIGEnews, USI_university, unil, UniLuzern, UniNeuchatel, HSGStGallen, UZH_ch
News media	19h30RTS, 20min, 20minutesOnline, 24heuresch, 52minutesRTS, AargauerZeitung, Ageficom, AppenzellerZeit, arcinfo, BauernZeitung1, bazonline, BernerZeitung, bielertagblatt, bilanmagazine, blickamabend, Blickch, bodenseewoche, bote_online, CdT_Online, chmediaag, Der_Landbote, derbund, die_weltwoche, energy_ch, Forum_RTS, Friburgera, Gauchebdo, giorndelpopolo, gruyere_journal, GTGrenchen, gvaobserver, heidi_news, info_sept, JournalduJura, LaCoteJournal, laliberte, LaNotizia, laregione, LaRegionNV, LausanneCites, lecourrier, lemanbleutv, Lematinch, lematindimanche, lenouvelliste, letemps, Lillustre, LuzernerZeitung, LuzeRund, mag_bonasavoir, MigrosMagazin, migrosmagazine, miseaupoint, Mittellaendisch, News_Luzern, Nouvo, NZZ, NZZaS, OstschweizamSon, radio24, radio3i, radiortn, RadioTeleSuisse, RepublikMagazin, RSInews, RSIonline, RTSinfrarouge, RTSpresse, RTSredaction, RTSUnDeux, SchweizerBauer, schweizerillu, SHN_News, SoBlick, sonntagsblatt, sonntagszeitung, SRF, srf_ostschweiz, srfaarau, srfbasel, srfbern, srfdata, srfkultur, srfluzern, srfnews, srfzuerich, suedostschweiz, swissinfo, swissinfo_de, swissinfo_en, swissinfo_fr, swissinfo_it, SZSolothurn, tagesanzeiger, tdgch, TeleBaernTV, Teleticino, TeleZueri, tempsprésent, Ticino7_CH, ticinonews, Ticinonline, watson_news, Weltwoche, weltwocheonline, WillisauerBote, Wochenzeitung, ZSZonline, zt_info, zuerisee
Business and industry	foraus, Avenir_Suisse, Avenir_Suisse_f, Avenir_Suisse_i, SGE, gewerbeverband, BaumeisterCH, arbeitgeber_ch, economie_suisse, economiesuisse, GewerkschaftSGB, SyndicatUSS, TravailsuisseCH, usp, sbv, santesuisse, doctorfmh, spitexch, publichealth_ch, SwissBankingSBA, GastroSuisseCH, hs_politik
Cantonal authorities	CantonduJura, Etat_Neuchatel, EtatdeVaud, GE_chancellerie, Etat_Fribourg, cantondeberne, kanton_bern, KantonSolothurn, Kanton_BL, BaselStadt, kantonaargau, KantonLuzern, CantonduValais, Kanton_Obwalden, KantonNW, infokantonuri, KantonZug, KantonZuerich, Kanton_Thurgau, kantonsg, AppAusserrhoden, Kanton_GR, ti_SIC
Swiss Parliament	ParlCH
Political parties	BDPSchweiz, Mitte_Centre, evppev, FDP_Liberalen, GrueneCH, grunliberale, LEGAdeiTicinesi, LesVertsSuisses, GrueneCH, PBDSuisse, PLR_Suisse, PSSuisse, pst_pop, solidariteS_CH, spschweiz, SVPch, UDCch, vertliberaux
Federal Council	BR_Sprecher, alain_berset, Violapamherd, ignaziocassis, s_sommaruga, ParmelinG, EDA_DFAE, EDI_DFI, EJPD_DFJP_DFGP, vbs_ddps, efd_dff, DefrWbf, UVEK_DETEC
Federal Office of Public Health	BAG_OFSP_UFSP
Task force	SwissScience_TF, TanjaStadler_CH

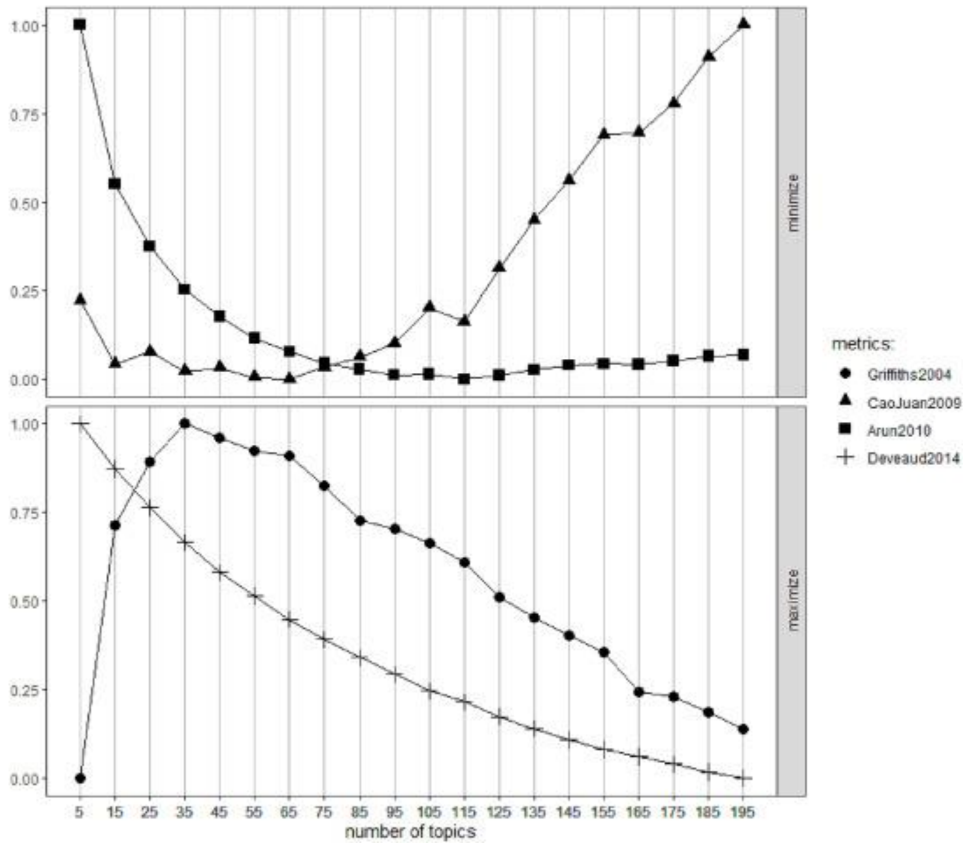
Annex 5.2.3: Keywords for tweets selection

generic	Distancing	face mask	contact tracing	quarantine	vaccine	pass
.*covid.*	social distancing.*	atemschutz.*	covid-app.*	lockdown.*	geimpf.*	hygieneausweis.*
.*corona.*	social-distancing.*	ffp2	dp-3t	lock-down.*	impfstoff.*	pass sanitaire
.*cov19.*	distanciation sociale	hygienemaske	dp 3t	quarantaine.*	impfen	passanitaire
.*cov2019.*	distance sociale	maske.*	dp3t	quarantän.*	impffrei.*	passe sanitaire
.*sars-cov.*	soziale distanz.*	maskenpflicht.*	kontakt.*rückverfolg.*	stayhome.*	impf-frei.*	passsanitaire
.*sarscov.*	distanzregel.*	maskenwahn.*	kontakt.*verfolg.*	confinement.*	impfnebenwirkung.*	passe-sanitaire
.*ncov.*	Distanzierung	maskenzwang.*	swisscovid.*	confiner	impf-nebenwirkung.*	pass-sanitaire
.*n-cov.*	Distanciation	masque.*	contact.*tracing.*	confiné.*	impfpässe	sanitary pass
anticovid.*		masquer	tracing	lockerung.*	impfplicht.*	zertifikatspflicht
anti-covid.*		schutzmaske.*	traçage	isolement.*	impf-plicht.*	.*zertifikat.*
.*taskforce.*		masks	tracer	quarantine	impfquote.*	.*certificat.*
.*task-force.*			contact-tracing.*	isoliering	impfrückstand.*	.*passcovid.*
ausserordendliche lage			corona app		impf-rückstand.*	.*covidpass.*
besondere lage			coronaapp.*		impfung.*	.*covid-pass.*
situation extraordinaire			corona-app.*		impfzwang.*	.*pass-covid.*
crise sanitaire			coronawarn.*		impf-zwang.*	passe-covid.*
crisesanitaire			corona warn.*		moderna	passe covid.*
gesundheitskrise.*			corona-warn.*		pfizer	
.*cv19.*			coviddapp.*		.*vaccin.*	
.*infektion.*			covid-codes		verimpf.*	
infection.*			covid app.*		anticorps	
pandemie			kontakt-rückverfolg.*		anti-coprs	
pandémie			kontakt-verfolg.*		antikörper.*	
beatmungsgerät.*					.*imunität	
respirateur.*					immunité	
respirator.*						
epidemie						
epidémie						
hospitalis.*						

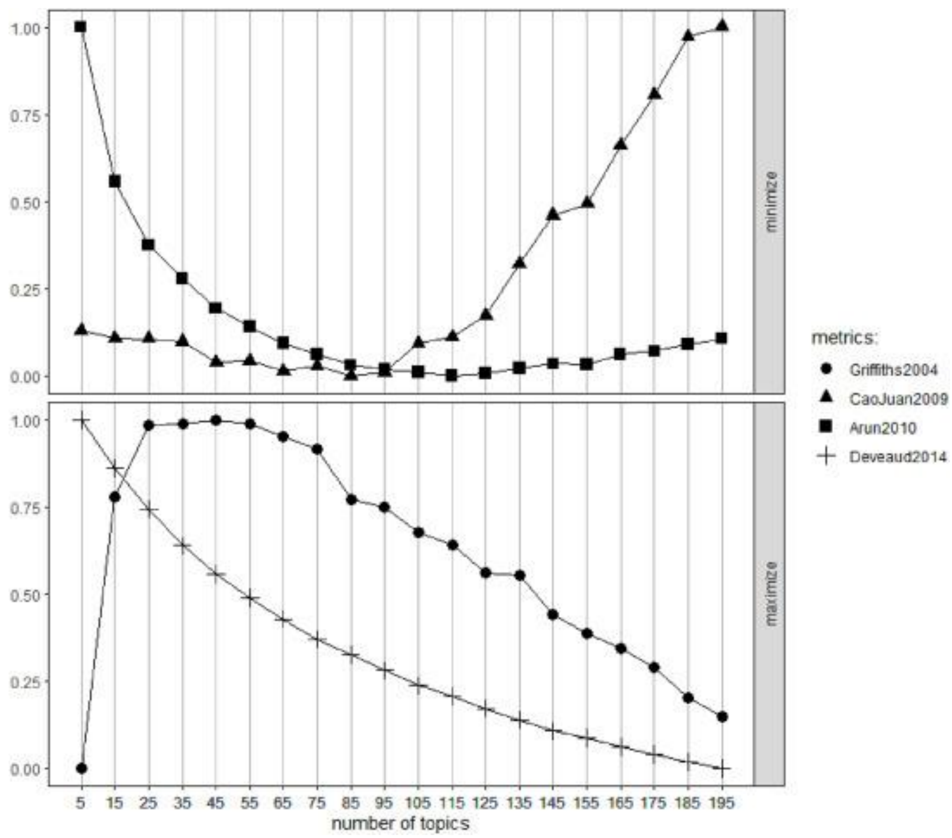
Annex 5.2.4: Description of our corpus of tweets and of the triggered reactions by actor group

Actor information			Covid-19 tweets		Popularity measures		
Actor groups	**#accounts**	**#tweets**	**#Covid-19 tweets**	**#cleaned tweets in German or French**	**Mean reply**	**Mean retweet**	**Mean like**
Business & industry	22	13472	2385 (18%)	2036	0,90	1,26	3,41
Cantonal authorities	23	28624	7647 (27%)	6405	1,46	2,61	6,64
Federal Council	13	15894	2735 (17%)	1946	13,08	13,57	43,99
FOPH	1	5115	3359 (66%)	2157	14,85	15,63	36,35
News media	91	596820	87746 (15%)	59126	1,49	1,48	3,82
Political parties	17	17005	1783 (10%)	1733	6,83	7,09	28,46
Politicians	152	57563	7208 (13%)	6195	6,55	7,28	44,46
Swiss Parliament	1	2102	338 (16%)	213	0,74	1,99	5,57
Taskforce	1	283	176 (62%)	80	6,56	10,52	25,93
Taskforce board	1	253	97 (38%)	30	3,33	20,30	58,28
University research centres	19	23160	2126 (9%)	1745	0,35	3,49	7,94
Total	189	760291	115600		5,10	7,75	24,08

Annex 5.2.5: Best topic number for French tweets (k = 60)



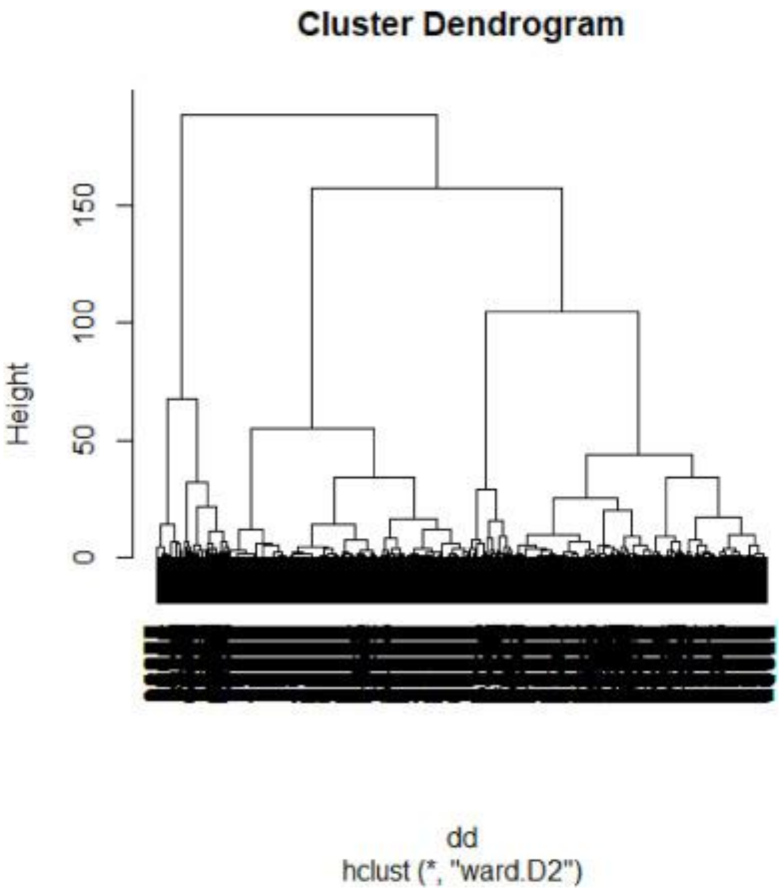
Annex 5.2.6: Best topic number for German tweets (k = 70)



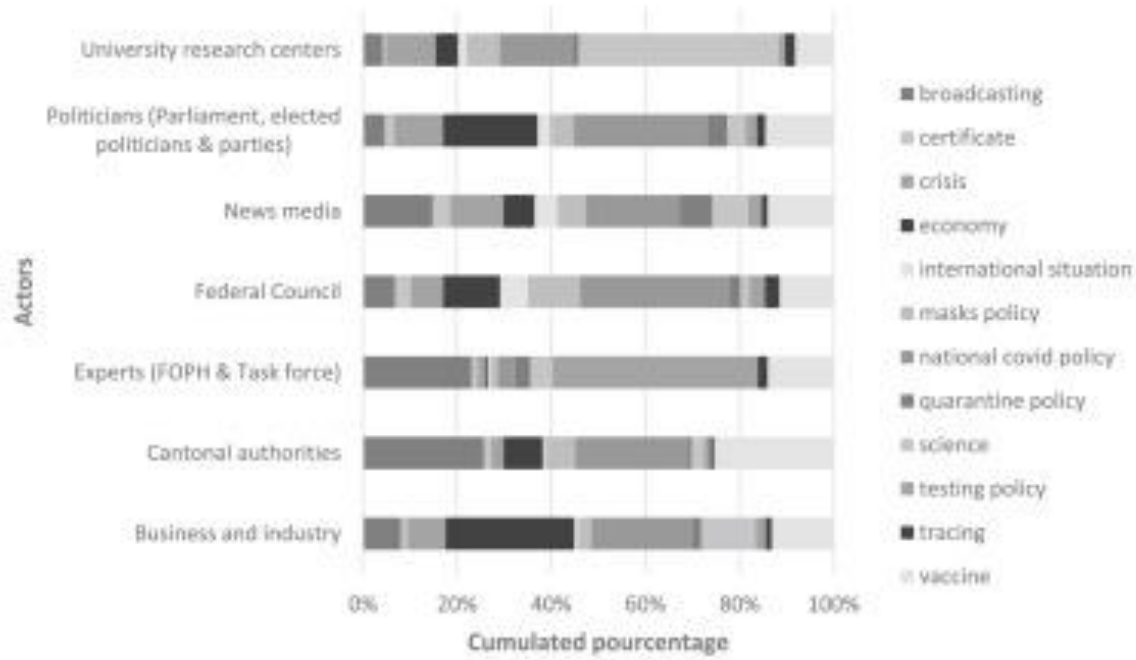
Annex 5.2.7: List of search queries for the target populations

language	children	Elderly	women	patients	adultes
French	enfant.*	personne.*âgée.*	femme.*	patients	adultes
	étudiant.*	personne.*agée.*	enceinte.*	hospitalisé.*	population
	bébé.*	retraité.*	mère.*		peuple
	école.*	maison.* de retraite	maman.*		
	écolier.*	EMS			
	écolière.*				
	gymnase.*				
	gymnasien.*				
German	kind	alte person.*	frau.*	patient.*	erwachsene.*
	kinder.*	ruhestand.*	schwanger.*	hospitalisierte.*	population.*
	baby	altersheim.*	mutter.*		Volk
	student.*	pflegeheim.*	mama.*		
	schüler.*	APH			
	schule.*				
	gymnasium				

Annex 5.2.8: Results of the Hierarchical Clustering with the Ward algorithm



Annex 5.2.9: Topic distribution by actor



References

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. Journal of medical Internet research, 22(4), e19016. https://doi.org/10.2196/19016

Adams-Cohen, N. J. (2020). Policy Change and Public Opinion: Measuring Shifting Political Sentiment with Social Media Data. American Politics Research, 48 (5), 612-21. https://doi.org/10.1177/1532673X20920263

Albaugh, Q., Sevenans, J., Soroka, S., & Loewen, P. J. (2013). The automated coding of policy agendas: A dictionary-based approach. In 6th Annual Comparative Agendas Conference, Atnwerp, Beligum.

AlDayel, A., & Magdy, W. (2021). Stance detection on social media: State of the art and trends. Information Processing & Management, 58(4), 102597. https://doi.org/10.1016/j.ipm.2021.102597

Alejandro, J. (2010). Journalism in the age of social media. Reuters Institute Fellowship Paper, 5.

Al-Ghadir, A. I., Azmi, A. M., & Hussain, A. (2021). A novel approach to stance detection in social media tweets by fusing ranked lists and sentiments. Information Fusion, 67, 29-40. https://doi.org/10.1016/j.inffus.2020.10.003

Amaya, A., Bach, R., Kreuter, F., & Keusch, F. (2020). Measuring the strength of attitudes in social media data. In Hill, C., Biemer, P., Buskirk, T., Japec, L., Kirchner, A., Kolenikov, S., & Lyberg, L. (eds), Big data meets survey science, pp. 163-192. Wiley, Hoboken

Amsler, M. (2020). Using Lexical-Semantic Concepts for Fine-Grained Classification in the Embedding Space. Doctoral dissertation. University of Zurich, Faculty of Arts.

Amsler, M., & Wüest, B. (2016). Legitimacy of New Forms of Governance in Public Discourse - An Automated Media Content Analysis Approach Driven By Techniques of Computational Linguistics. In Proceedings of the International Conference on the Advances in Computational Analysis of Political Text (PolText), 1-7.

Annales. (2017). Qualité Des Médias. Zürich: Forschungsinstitut Öffentlichkeit und Gesellschaft. https://www.foeg.uzh.ch/dam/jcr:15172213-4895-42c5-aaf9-850d9a412d9c/Broschur_Jahrbuch_foeg_franz_2017_ohne_Sperrvermerk.pdf

Anstead, N., & O'Loughlin, B. (2015). Social media analysis and public opinion: the 2010 UK general election. J Comput-Mediat Commun, 20(2): 204-220. https://doi.org/10.1111/jcc4.12102

Archer, A. M. N., & Kam, C. D. (2020). Modern Sexism in Modern Times Public Opinion in the #Metoo Era. Public Opinion Quarterly, 84 (4), 813-37.

Arnold, T., & Tilton, L. (2016). Wrappers Around Stanford CoreNLP Tools. R package, version 0.4-2. https://cran.r-project.org/web/packages/coreNLP/coreNLP.pdf

Baden, C., Kligler-Vilenchik, N., & Yarchi, M. (2020). Hybrid content analysis: Toward a strategy for the theory-driven, computer-assisted classification of large text corpora. Communication Methods and Measures, 14(3), 165-183. https://doi.org/10.1080/19312458.2020.1803247

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. G. (2021). Three Gaps in Computational Text Analysis Methods for Social Sciences: A Research Agenda. Communication Methods and Measures, 16(1), 1-18. https://doi.org/10.1080/19312458.2021.2015574

Baik, J. M., T. H. Nyein, & Modrek, S. (2021). Social Media Activism and Convergence in Tweet Topics After the Initial #MeToo Movement for Two Distinct Groups of Twitter Users. Journal of Interpersonal Violence, 37 (15-16), NP13603-22.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature, 533, 452-454. https://doi.org/10.1038/533452a

Bakker, R., Hooghe, L., Jolly, S., Marks, G., Polk, J., Rovny, J., Steenbergen, M., & Vachudova, M. (2021). Chapel Hill expert survey (CHES) Europe 1999-2019 trend file. EUI research data, Robert Schuman Centre for Advanced Studies, European University Institute. https://cadmus.eui.eu//handle/1814/69975

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130-1132. https://doi.org/10.1126/science.aaa1160

Baldassarri, D., & Park, B. (2020). Was there a Culture War? Partisan Polarisation and Secular Trends in US Public Opinion. The Journal of Politics, 82 (3), 809-27.

Ball-Rokeach, S. J., & DeFleur, M. L. (1976). A dependency model of mass-media effects. Communication research, 3(1): 3-21.

Barberá, P., & Rivero, G. (2015). Understanding the political representativeness of Twitter users. Social Science Computer Review, 33(6), 712-729.

Barberá, P., Boydstun, A. E., Linn, S., McMahon, R., & Nagler, J. (2021). Automated Text Classification of News Articles: A Practical Guide. Political Analysis, 29 (1), 19-42. https://doi.org/10.1017/pan.2020.8

Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., & Tucker, J. A. (2019). Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data. American Political Science Review, 113(4), 883-901. https://doi.org/10.1017/S0003055419000352

Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., & Tucker, J. A. (2019). Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data. American Political Science Review, 113(4), 883-901. https://doi.org/10.1017/S0003055419000352

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? Psychological Science, 26(10), 1531-1542. https://doi.org/10.1177/0956797615594620

Barberá, P., & Steinert-Threlkeld, Z. (2020). How to Use Social Media Data for Political Science Research. In Curini, L., & Franzese, R (eds.), The SAGE Handbook of Research Methods in Political Science and International Relations, pp. 404–21. Thousand Oaks, CA: SAGE Publications.

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: A survey of techniques, tools and platforms. Ai & Society, 30(1), 89-116. https://doi.org/10.1007/s00146-014-0549-4

Beauchamp, N. (2016). Predicting and interpolating state-level polls using Twitter textual data. American Journal of Political Science, 61(2), 490-503. https://doi.org/10.1111/ajps.12274

Bekafigo, M. A., & McBride, A. (2013). Who Tweets About Politics?: Political Participation of Twitter Users During the 2011Gubernatorial Elections. Social Science Computer Review, 31(5), 625-643. https://doi.org/10.1177/0894439313490405

Beninger, K., Fry, A., Jago, N., Lepps, H., Nass, L., & Silvester, H. (2014). Research using social media; users' views. NatCen Social Research, 20.

Bennett, J., Rachunok, B., Flage, R., & Nateghi, R. (2021). Mapping climate discourse to climate opinion: An approach for augmenting surveys with social media to enhance understandings of climate opinion in the United States. PLOS ONE, 16(1), e0245319. https://doi.org/10.1371/journal.pone.0245319

Bennett, W. L. (1993). Constructing publics and their opinions. Taylor & Francis.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R Package for the Quantitative Analysis of Textual Data. Journal of Open Source Software, 3 (30), 1-4. https://doi.org/10.21105/joss.00774

Berinsky, A. J. (2017). Measuring public opinion with surveys. Annual review of political science, 20, 309-329. https://doi.org/10.1146/annurev-polisci-101513-113724

Bessi, A., Petroni, F., Del Vicario, M., Zollo, F., Anagnostopoulos, A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Homophily and polarization in the age of misinformation. The European Physical Journal Special Topics, 225(10), 2047-2059. https://doi.org/10.1140/epjst/e2015-50319-0

Bethard, S., Carpuat, M., Cer, D., Jurgens, D., Nakov, P., & Zesch, T. (2016). Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). San Diego: ACL. https://aclanthology.org/S16-1000

Biber, D., & Finegan, E. (1988). Adverbial stance types in English. Discourse processes, 11(1), 1-34. https://doi.org/10.1080/01638538809544689

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). Longman grammar of spoken and written English, (Vol. 2). London: Longman.

Biemer, P. P. 2010. Total survey error: Design, implementation, and evaluation. Public Opinion Quarterly,74(5), 817–848. https://doi.org/10.1093/poq/nfq058

Biemer, P. P., & Christ, S. L. (2008). Weighting survey data. International Handbook of Survey Methodology, 2008, 317-341.

Bivens, R. K. (2008). The Internet, mobile phones and blogging: How new media are transforming traditional journalism. Journalism practice, 2(1), 113-129. https://doi.org/10.1080/17512780701768568

Blei, D. M. (2012). Probabilistic Topic Models. Communications of the ACM, 55(4), 77-84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993-1022.

Blinder, S. and Richards, L. (2020). UK public opinion toward immigration: overall attitudes and level of concern. Oxford: Migration Observatory. https://migrationobservatory.ox.ac.uk/wp-content/uploads/2020/01/Briefing-

UK-Public-Opinion-toward-Immigration-Overall-Attitudes-and-Level-of-Concern.pdf

Blondiaux L. (1998). La fabrique de l'opinion. Une histoire sociale des sondages. Paris: Seuil.

Bode, L., & Dalrymple, K. E. (2016). Politics in 140 Characters or Less: Campaign Communication, Network Interaction, and Political Participation on Twitter. Journal of Political Marketing, 15(4), 311-332. https://doi.org/10.1080/15377857.2014.959686

Boumans, J. W., & Trilling, D. (2016). Taking stock of the toolkit. Digital Journalism, 4(1), 8-23. https://doi.org/10.1080/21670811.2015.1096598

Bourdieu, P. (1973). L'opinion publique n'existe pas. Les Temps Modernes, 318.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878

Brandt, C. J., Søgaard, G. I., Clemensen, J., Sndergaard, J., & Nielsen, J. B. (2018). General Practitioners' Perspective on eHealth and Lifestyle Change: Qualitative Interview Study. JMIR mHealth and uHealth, 6(4), e88. https://doi.org/10.2196/mhealth.8988

Brick, J.M., & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. Ann Am Acad Political Soc Sci, 645(1):36-59

Brickman, D., & Peterson, D. A. M. (2006). Public Opinion Reaction to Events: Citizen Response to Multiple Supreme Court Abortion Decisions. Political Behavior, 28 (1), 87-112.

Brüggemann, M., & D'Angelo, P. (2018). Defragmenting news framing research: Reconciling generic and issue-specific frames. In D'Angelo, P. (Ed.), Doing news framing analysis, pp. 90-111. New York: Routledge.

Bundschus, M., Tresp, V., & Kriegel, H. P. (2009). Topic models for semantically annotated document collections. NIPS workshop: Applications for Topic Models: Text and Beyond, pp. 1-4.

Burgess, J., & Bruns, A. (2013). Cumulative and comparative social media analytics for crisis communication. Selected Papers of Internet Research 14.0, 1-5. https://eprints.qut.edu.au/67133/

Camerer, C.F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B.A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E.-J., & Wu, H.. (2018). Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015. Nature Human Behaviour, 2(9), 637-644. https://doi.org/10.1038/541562-0180399-2

Card, D., Gross, J. H., Boydstun, A., & Smith, N. A. (2016). Analyzing Framing Through the Casts of Characters in the News. In Proceedings of the conference on empirical methods in natural language processing, 1410-1420.

Cardenal, A. S., Galais, C., & Majó-Vázquez, S. (2019). Is Facebook Eroding the Public Agenda? Evidence From Survey and Web-Tracking Data. International Journal of Public Opinion Research, 31(4), 589-608. https://doi.org/10.1093/ijpor/edy025

Castanho Silva, B., & Proksch, S. (2021). Politicians unleashed? Political communication on Twitter and in parliament in Western Europe. Political Science Research and Methods, 1-17. https://doi.org/10.1017/psrm.2021.36

Ceron, A. (2017). Intra-party politics in 140 characters. Party Politics, 23(1): 7-17.

Chan, C. H., Bajjalieh, J., Auvil, L., Wessler, H., Althaus, S., Welbers, K., van Atteveldt, W., & Jungblut, M. (2021). Four best practices for measuring news sentiment using 'off-the-shelf'dictionaries: A large-scale p-hacking experiment. Computational Communication Research, 3(1), 1-27. https://doi.org/10.5117/CCR2021.1.001.CHAN

Chan, C.-H., Zeng, J., Wessler, H., Jungblut, M., Welbers, K., Bajjalieh, J. W., van Atteveldt, W., & Althaus, S. L. (2020). Reproducible extraction of cross-lingual topics. Communication Methods and Measures, 14(4), 285-305. https://doi.org/10.1080/19312458.2020.1812555

Charteris-Black, J. (2006). Britain as a Container: Immigration Metaphors in the 2005 Election Campaign. Discourse & Society, 17(5): 563-581. https://doi.org/10.1177/0957926506066345

Chauhan, P., Sharma, N., & Sikka, G. (2020). The emergence of social media data and sentiment analysis in election prediction. Journal of Ambient Intelligence and Humanized Computing, 12(2), 2601-2627. https://doi.org/10.1007/s12652-020-02423-y

Chen, K., & Tomblin, D. (2021). Using Data from Reddit, Public Deliberation, and Surveys to Measure Public Opinion about Autonomous Vehicles. Public Opinion Quarterly, 85(S1), 289-322. https://doi.org/10.1093/poq/nfab021

Chen, K., Duan, Z., & Yang, S. (2022). Twitter as research data: Tools, costs, skill sets, and lessons learned. Politics and the Life Sciences, 41(1), 114-130. https://doi.org/10.1017/pls.2021.19

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PloS One, 5(11), e14118.

Cinelli, M., G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, & Starnini, M. (2021). The Echo Chamber Effect on Social Media. Proceedings of the National Academy of Sciences, 118 (9), e2023301118.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. Scientific reports, 10(1), 1-10. https://doi.org/10.1038/s41598-020-73510-5

Clark, T. S., Staton, J. K., Wang, Y., & Agichtein, E. (2018). Using Twitter to Study Public Discourse in the Wake of Judicial Decisions: Public Reactions to the Supreme Court's Same-Sex-Marriage Cases. Journal of Law and Courts, 6(1), 93-126. https://doi.org/10.1086/695423

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. Journal of Communication, 64 (2), 317-32. https/doi.org/10.1111/jcom.12084.

Combei, C. R., Farnè, M., Pinto, L., & Giannetti, D. (2020). Populism and Policy Issues: Examining Political Communication on Twitter in Italy 2018-2019. Italian Political Science, 15(2): 223-241.

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. Fifth International AAAI Conference on Weblogs and Social Media.

Converse, P.E. (1970). Attitudes and non-attitudes: continuation of a dialogue. In E.R. Tufte (Ed.), The Quantitative Analysis of Social Problems, pp. 168–89. Reading, MA: Addison-Wesley

Cortes, C., Jackel, L. D., Solla, S., Vapnik, V., & Denker, J. (1994). Learning Curves: Asymptotic Values and Rate of Convergence. In Proceedings of the sixth International conference on neural information processing system, 327-334.

Couper, M. P. (2013). Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys. Survey Research Methods, 7(3), 145-156. https://doi.org/10.18148/srm/2013.v7i3.5751

Crawley, H. (2005). Evidence on attitudes to asylum and immigration: what we know, don't know and need to know. COMPAS Working Paper 23. Oxford: Centre on Migration, Policy and Society. www.compas.ox.ac.uk/2005/wp-2005-023-crawley_attitudes_asylum_immigration

Crawley, H., & McMahon, S. (2016). Beyond fear and hate: mobilising people power to create a new narrative on migration and diversity. Coventry: Centre for Trust, Peace and Social Relations. https://pureportal.coventry.ac.uk/en/publications/beyond-fear-and-hate-mobilising-people-powerto-create-a-new-nar

Czajka, J. L., & Beyler, A. (2016). Declining response rates in federal surveys: trends and implications (background paper). Mathematica Policy Research. https://aspe.hhs.gov/sites/default/files/private/pdf/255531/Decliningresponserates.pdf

D'Enza, A. I., & Greenacre, M. (2012). Multiple Correspondence Analysis for the Quantification and Visualization of Large Categorical Data Sets. In Advanced Statistical Methods for the Analysis of Large Data-Sets. Studies in Theoretical and Applied Statistics, edited by A. Di Ciaccio, M. Coli, and J. Angulo Ibanez. Berlin: Springer.

Daas, P., Roos, M., van de Ven, M., & Neroni, J. (2012). Twitter as a potential source for official statistics in the Netherlands. Paper for the 67th AAPOR 2012 Conference Vom, 17-20.

Dahlberg, S., Axelsson, S., & Holmberg, S. (2020). Democracy in context: Using a distributional semantic model to study differences in the usage of democracy across languages and countries. Zeitschrift Für Vergleichende Politikwissenschaft, 14(4), 425-459. https://doi.org/10.1007/s12286-020-00472-3

Davis, M. A., Zheng, K., Liu, Y., & Levy, H. (2017). Public Response to Obamacare on Twitter. Journal of Medical Internet Research, 19(5), e6946. https://doi.org/10.2196/jmir.6946

De Heer, W., & De Leeuw, E. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A., Survey nonresponse (Eds.), vol. 41, pp. 41-54. New York, NY: Wiley.

de Rosa, A. S., Bocci, E., Bonito, M., & Salvati, M. (2021). Twitter as social media arena for polarised social representations about the (im)migration: The controversial discourse in the Italian and international political frame. Migration Studies, mnab001. https://doi.org/10.1093/migration/mnab001

De Sio, L. D., & Weber, T. (2020). Issue yield, campaign communication, and electoral performance: A six-country comparative analysis. West European Politics, 43(3), 720-745. https://doi.org/10.1080/01402382.2019.1655968

de Vreese, C.H., Peter, J. & Semetko, H. A. (2001). Framing Politics at the Launch of the Euro: A Cross-National Comparative Study of Frames in the News. Political Communication, 18: 107-122.

De Wilde, P., & Zürn, M. (2012). Can the Politicization of European Integration be Reversed? JCMS: Journal of Common Market Studies 50: 137-53. van Dijck, J., and T. Poell. 2013. Understanding Social Media Logic. Media and Communication, 1 (1), 2-14.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., … Quattrociocchi, W. (2016). The spreading of misinformation online. Proceedings of the National Academy of Sciences, 113, 554-559.

Dempster, H., Leach, A., & Hargrave, K. (2020). Public attitudes towards immigration and immigrants. Working paper 588. London: ODI. https://cisp.cachefly.net/assets/articles/attachments/83372_202009_odi_public_attitudes_towards_immigration_wp_final_0.pdf

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

Diaz, F., Gamonm, M., Hofman, J.M., Kıcıman, E., & Rothschild, D. (2016). Online and social media data as an imperfect continuous panel survey. PLoS One, 11(1):e0145406. https://doi.org/10.1371/journal.pone.0145406

Digital Global Report. (2020). Digital use around the world in July 2020. https://wearesocial.com/blog/2020/07/digital-use-around-the-world-in-july-2020

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting Affinities Between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of US Government Arts Funding. Poetics 41 (6): 570-606. https://doi.org/10.1016/j.poetic.2013.08.004

Dixon, K. (2014). Feminist Online Identity: Analyzing the Presence of hashtag Feminism. Journal of Arts & Humanities, 3 (7), 34-40.

Dobbrick, T., Jakob, J., Chan, C. H., & Wessler, H. (2021). Enhancing Theory-Informed Dictionary Approaches with Glass-box Machine Learning: The Case of Integrative Complexity in Social Media Comments. Communication Methods and Measures. https://doi.org/10.1080/19312458.2021.1999913

Doherty, B. (2015). Call me illegal: the semantic struggle over seeking asylum in Australia. Oxford: Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/our-research/call-me-illegal-semantic-struggle-over-seeking-asylum-australia

Domahidi, E., Yang, J., Niemann-Lenz, J., & Reinecke, L. (2019). Outlining the way ahead in computational communication science: An introduction to the IJOC special section on Computational methods for communication science: Toward a strategic roadmap. International Journal of Communication, 13, 1-9. https://ijoc.org/index.php/ijoc/article/view/10533

Duan, Z., Li, J., Lukito, J., Yang, K. C., Chen, F., Shah, D. V., & Yang, S. (2022). Algorithmic Agents in the Hybrid Media System: Social Bots, Selective Amplification, and Partisan News about COVID-19. Human Communication Research. https://doi.org/10.1093/hcr/hqac012

Dubois, E., & Szwarc, J. (2018). Self-Censorship, Polarization, and the—Spiral of Silence on Social Media. In Policy & Politics Conference. https://blogs.oii.ox.ac.uk/policy/wp-content/uploads/sites/77/2018/08/IPP2018-Dubois.pdf

Dubois, E., Gruzd, A., & Jacobson, J. (2020). Journalists' Use of Social Media to Infer Public Opinion: The Citizens' Perspective. Social Science Computer Review, 38(1), 57-74. https://doi.org/10.1177/0894439318791527

Duffy, B. (2019). The perils of perception: why we're wrong about nearly everything. London: Atlantic Books.

Eberl, J.-M., C. E. Meltzer, T. Heidenreich, B. Herrero, N. Theorin, F. Lind, R. Berganza, H. G. Boomgaarden, C. Schemer, and S. Strömbäck. (2018). The European Media Discourse on Immigration and its Effects: A Literature Review. Annals of the International Communication Association, 42(3): 207-223. https://doi.org/10.1080/23808985.2018.1497452

Eberl, J.-M., Tolochko, P., Jost, P., Heidenreich, T., & Boomgaarden, H. G. (2020). What's in a post? How sentiment and issue salience affect users' emotional reactions on Facebook. Journal of Information Technology & Politics, 17(1), 48-65. https://doi.org/10.1080/19331681.2019.1710318

Ebrahimi, J., Dou, D., & Lowd, D. (2016). A joint sentiment-target-stance model for stance classification in tweets. In Proceedings of COLING 2016, the 26th international conference on computational linguistics, Technical papers, pp. 2656-2665.

Ehret, K., & Taboada, M. (2021). The interplay of complexity and subjectivity in opinionated discourse. Discourse Studies, 23(2), 141-165. https://doi.org/10.1177/1461445620966923

Eisenegger, M. (2020). Yearbook 2020: The Quality of the Media. University of Zurich: Research Center for the Public Sphere and Society (fög). https://www.foeg.uzh.ch/dam/jcr:efe9e096-c9a1-4c33-b255-876200f5e40c/Main_Findings_2020.pdf

Eisner, L., D. Spini, & Sommet, N. (2020). A Contingent Perspective on Pluralistic Ignorance: When the Attitudinal Object Matters. International Journal of Public Opinion Research, 32 (1), 25-45.

Eisner, L., Settersten, R., Turner-Zwinkels, F., & Hässler, T. (2022). Perceptions of intolerant norms both facilitate and inhibit collective action among sexual minorities. Group Processes & Intergroup Relations, 25(7), 1797-1818. https://doi.org/10.1177/13684302211024335

Ekman, M. (2019). Anti-immigration and racist discourse in social media. European Journal of Communication, 34(6): 606-618. https://doi.org/10.1177/0267323119886151

Ellison, N. B., Vitak, J., Steinfield, C., Gray, R., & Lampe, C. (2011). Negotiating privacy concerns and social capital needs in a social media environment. In S. Trepte & L. Reinecke (Eds.), Privacy onlin, pp. 19-32. London: Springer.

Eltantawy, N., & Wiest, J. B. (2007). The Arab Spring | Social Media in the Egyptian Revolution: Reconsidering Resource Mobilization Theory. Journal of Communication, 5, 1207-24.

Engel, U. (2021). Causal and predictive modeling in computational social science. In Engel, U. (ed.), Handbook of Computational Social Science, pp. 131-149, Vol 1. Taylor & Francis.

Ennser-Jedenastik, L., Haselmayer, M., Huber, L. M., & Scharrer, M. E. (2022). Issue substitution or volume expansion? How parties accommodate agenda change. Electoral Studies, 76, 102437.

Entman, R. M. (2007). Framing bias: Media in the distribution of power. Journal of communication, 57(1), 163-173. https://doi.org/10.1111/j.1460-2466.2006.00336.x

Ernst, N., Engesser, S., Büchel, F., Blassnig, S., & Esser, F. (2017). Extreme parties and populism: An analysis of Facebook and Twitter across six countries. Information, Communication & Society, 20(9), 1347-1364. https://doi.org/10.1080/1369118X.2017.1329333

Ernst, N., Esser, F., Blassnig, S., & Engesser, S. (2019). Favorable opportunity structures for populist communication: Comparing different types of politicians and issues in social media, television and the press. The International Journal of Press/Politics, 24(2): 165-188.

European Commission, Brussels (2015). Flash Eurobarometer 404 (European Citizens´ Digital Health Literacy). GESIS Data Archive, Cologne. ZA5948 Data file Version 1.0.0. https://doi.org/10.4232/1.12194

Evans, J. A., & Aceves, P. (2016). Machine translation: mining text for social theory. Annual Review of Sociology, 42, 21-50. https://doi.org/10.1146/annurev-soc-081715-074206

Eveland, W. P., Marton, K., & Seo, M. (2004). Moving Beyond 'Just the Facts': The Influence of Online News on the Content and Structure of Public Affairs Knowledge. Communication Research, 31 (1), 82-108. https/doi.org/10.1177/0093650203260203.

Fagerlund, AJ, Holm, IM, & Zanaboni, P. (2019). General practitioners' perceptions towards the use of digital health services for citizens in primary care: a qualitative interview study. BMJ Open, 9(e028251), 1-17. https://doi.org/10.1136/bmjopen-2018-028251

Fellbaum, C. (2005). WordNet and wordnets. In K. Brown (Ed.), Encyclopedia of Language and Linguistics, (2nd edition), pp. 665-670. Oxford: Elsevier.

Ferrín, M. (2018). European Social Survey Round 10 Module Design Teams (QDT) Stage 2 Application. London: Centre for Comparative Social Surveys, City University London. http://www.europeansocialsurvey.org/docs/round10/questionnaire/ESS10_ferrin_proposal.pdf

Firth, J. R. (1957). A synopsis of linguistic theory 1930-1955. Studies in Linguistic Analysis, 1-32.

Fishman, R. M. (2016). Rethinking Dimensions of Democracy for Empirical Analysis: Authenticity, Quality, Depth, and Consolidation. Annual Review of Political Science ,19, 289-309. https://doi.org/10.1146/annurev-polisci-042114-015910

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. Public Opinion Quarterly, 80 (S1), 298-320. https/doi.org/10.1093/poq/nfw006.

Fög. (2016). Qualität der Medien: Hauptbefunde Jahrbuch 2016. https://www.qualitaet-der-medien.ch/hauptbefunde

Fög. (2017). Qualität der Medien: Hauptbefunde Jahrbuch 2017. https://www.qualitaet-der-medien.ch/hauptbefunde

Fög. (2018). Qualität der Medien: Hauptbefunde Jahrbuch 2018. https://www.qualitaet-der-medien.ch/hauptbefunde

Freelon, D. & Karpf, D. (2015). Of big birds and bayonets: Hybrid Twitter interactivity in the 2012 Presidential debates. Information, Communication & Society, 18(4), 390–406. https://doi.org/10.1080/1369118X.2014.952659

Gamson, W. A., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. American journal of sociology, 95(1), 1-37.

García-Díaz, J. A., Apolinario-Arzube, Ó., Medina-Moreira, J., Luna-Aveiga, H., Lagos-Ortiz, K., & Valencia-García, R. (2018). Sentiment Analysis on Tweets related to infectious diseases in South America. In Proceedings of the Euro American Conference on Telematics and Information Systems, Brazil, pp. 1-5. https://doi.org/10.1145/3293614.3293647

García-Marín, J., & Calatrava, A. (2018). The use of Supervised Learning Algorithms in Political Communication and Media Studies: Locating Frames in the Press. Comunicación y Sociedad, 31(3), 175-188.

García-Orosa, B. (2021). Disinformation, social media, bots, and astroturfing: the fourth wave of digital democracy. Profesional De La información, 30(6). https://doi.org/10.3145/epi.2021.nov.03

Garg, R., Kiwelekar, A. W., Netak, L. D., & Bhate, S. S. (2021a). Personalization of news for a logistics organisation by finding relevancy using NLP. In Gunjan, V. K., & Zurada, J. M. (Eds.), Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough: Latest Trends in AI, vol. 2, pp. 215-226. Cham: Springer. https://doi.org/10.1007/978-3-030-68291-0_16

Garg, R., Kiwelekar, A. W., Netak, L. D., & Ghodake, A. (2021b). i-Pulse: A NLP based novel approach for employee engagement in logistics organization. International Journal of Information Management Data Insights, 1(1), Article 100011. https://doi.org/10.1016/j.jjimei.2021.100011

Gayo-Avello, D. (2011). Don't turn social media into another 'Literary Digest' poll. Communications of the ACM, 54(10), 121-128. https://doi.org/10.1145/2001269.2001297

Gayo-Avello, D. (2013). A Meta-analysis of state-of-the-art electoral prediction from twitter data. Soc Sci Comput Rev, 31(6):649-679. https://doi.org/10.1177/0894439313493979

Gayo-Avello, D., Metaxas, P., & Mustafaraj, E. (2011). Limits of Electoral Predictions Using Twitter. Proceedings of the International AAAI Conference on Web and Social Media, 5(1). https://ojs.aaai.org/index.php/ICWSM/article/view/14189

Genner, S. (2017). Does age determine media habits in Switzerland? https://www.swissinfo.ch/eng/youth-and-media_does-age-determine-media-habits-in-switzerland--/43551692

Gerring, J. (2012). Mere Description. British Journal of Political Science, 42(4), 721-746. https://doi.org/10.1017/S0007123412000130

Ghanem, S. (1997). "Filling the Tapestry: The Second Level of Agenda Setting." In Communication and Democracy: Exploring the Intellectual Frontiers in Agenda-Setting Theory, ed. by Maxwell McCombs, Donald L. Shaw and David Weaver. 3-14. Mahwah, NJ: Lawrence Erlbaum.

Ghosh, S., Singhania, P., Singh, S., Rudra, K., & Ghosh, S. (2019). Stance Detection in Web and Social Media: A Comparative Study. In Fabio Crestani et al. (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Lecture Notes in Computer Science, 11696. Springer, Cham. https://doi.org/10.1007/978-3-030-28577-7_4

Giger, N., D. Traber, F. Gilardi, & Bütikofer, S. (2021). The Surge in Women's Representation in the 2019 Swiss Federal Elections. Swiss Political Science Review, 28, 361-76.

Gilardi, F., & Wüest, B. (2018). Text-as-data methods for comparative policy analysis. Zurich: University of Zurich. Department of Political Science. https://doi.org/10.5167/uzh-207867

Gilardi, F., Gessler, T., Kubli, M., & Müller, S. (2021). Social media and policy responses to the COVID-19 pandemic in Switzerland. Swiss Political Science Review, 27(2), 243-256. https://doi.org/10.1111/spsr.12458

Gilardi, F., Gessler, T., Kubli, M., & Müller, S. (2022). Social media and political agenda setting. Political Communication, 39(1), 39-60. https://doi.org/10.1080/10584609.2021.1910390

Gilardi, F., Shipan, C. R., & Wüest, B. (2021). Policy Diffusion: The Issue-definition Stage. American Journal of Political Science, 65 (1), 21-35. https://doi.org/10.1111/ajps.12521

Goffman, E. (1974). Frame Analysis: An Essay on the Organization of Experience. New York, NY: Harper & Row.

Gold, N. (2020). “Using Twitter data in research: Guidance for researchers and ethics reviewers”. Available at: https://www.ucl.ac.uk/data-protection/sites/data-protection/files/using-twitter-research-v1.0.pdf

Golder, S., Ahmed, S., Norman, G., & Booth, A. (2017). Attitudes toward the ethics of research using social media: a systematic review. Journal of medical internet research, 19(6), e7082. https://doi.org/10.2196/jmir.7082

González-Bailón, S., & Paltoglou, G. (2015). Signals of public opinion in online communication. The ANNALS of the American Academy of Political and Social Science, 659(1), 95-107. https://doi.org/10.1177/0002716215569192

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. Social Networks, 38, 16-27. https://doi.org/10.1016/j.socnet.2014.01.004

Gonzalez-Barrera, A., & Connor, P. (2019). Around the World, More Say Immigrants Are a Strength Than a Burden. Report from Pew Research Center. https://www.pewresearch.org/global/2019/03/14/around-the-world-more-say-immigrants-are-a-strength-than-a-burden/

González-Ibánez, R., Muresan, S., & Wacholder, N. (2011). Identifying sarcasm in Twitter: A closer look. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2, 581-586. Association for Computational Linguistics.

Gordon, S. R. (2015). Fostering Social Movements with Social Media. SSRN Electronic Journal, 1-30. https://doi.org/10.2139/ssrn.2708079

Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from# Brexit and# USElection. European Economic Review, 136, 103772. https://doi.org/10.1016/j.euroecorev.2021.103772

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. Educational Evaluation and Policy Analysis, 11(3), 255-274. JSTOR. https://doi.org/10.2307/1163620

Grimmer, J., & B. M. Stewart. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, 21(3), 267-97. https://doi.org/10.1093/pan/mps028

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton, New Jersey Oxford: Princeton University Press.

Grover, P., Kar, A. K., & Davies, G. (2018). Technology enabled Health - Insights from twitter analytics with a socio-technical perspective. International Journal of Information Management, 43, 85-97. 10.1016/j.ijinfomgt.2018.07.003.

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. Public Opinion Quarterly, 70(5), 646-675. https://doi.org/10.1093/poq/nfl033

Groves, R. M. (2011). Three eras of survey research. Public opinion quarterly, 75(5), 861-871. https://doi.org/10.1093/poq/nfr057

Groves, R. M., & Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. Public Opinion Quarterly, 74(5), 849-879. https://doi.org/10.1093/poq/nfq065

Gründl, J. (2022). Populist ideas on social media: A dictionary-based measurement of populist communication. New Media & Society, 24(6), 1481-1499. https://doi.org/10.1177/1461444820976970

Guess, A. M. (2021). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. American Journal of Political Science, 65(4), 1007-1022. https://doi.org/10.1111/ajps.12589

Guido, Z., & Marsh, A. (2016). MITRE at SemEval-2016 task 6: Transfer learning for stance detection. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). 458-463.

Habermas, J. (1991). The structural transformation of the public sphere: An inquiry into a category of bourgeois society. Cambridge, UK: MIT press.

Hameleers, M. (2019). The populism of online communities: Constructing the boundary between “blameless” people and “culpable” others. Communication Culture & Critique, 12(1), 147-165. https://doi.org/10.1093/ccc/tcz009

Hamilton, W. L., Clark, K., Leskovec, J., & Jurafsky, D. (2016). Inducing Domain Specific Sentiment Lexicons from Unlabeled Corpora. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, 595-605.

Hanitzsch, T., Hanusch, F., Ramaprasad, J., & de Beer, A. S. (2019). Worlds of journalism: Journalistic cultures around the globe. Columbia University Press.

Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. The ANNALS of the American Academy of Political and Social Science, 659(1), 63-76. https://doi.org/10.1177/000271621557086

Hart, R. P. (1999). Seducing America: How Television Charms the Modern Voter. Sage.

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing Automated Text Classification Methods. International Journal of Research in Marketing, 36 (1), 20-38. https://doi.org/10. 1016/j.ijresmar.2018.09.009

Harzing, A.-W. (2007). Publish or Perish. https://harzing.com/resources/publish-or-perish

Hatipoğlu, E., Gökçe, O. Z., Arin, İ., & Saygin, Y. (2019). Automated Text Analysis and International Relations: The Introduction and Application of a Novel Technique for Twitter. All Azimuth: A Journal of Foreign Policy and Peace, 8(2), 183-204. https://doi.org/10.20991/allazimuth.476852

Hatton, T. (2017). Public opinion on immigration in Europe: preference versus salience. IZA Discussion Paper N° 10838. Bonn: IZA. http://ftp.iza.org/dp10838.pdf

Häuptli, A., Schranz, M., & Schneider, J. (2017). Situation actuelle et perspectives pour le paysage médiatique dans les régions de montagne. Fribourg: Journée d'étude SAB. http://www.sab.ch/fileadmin/user_upload/customers/sab/Tagungen/2017/Fribourg/01_Andrea_General-Praesentation_BergregionenTagung_10-11-2017fr.pdf

Hedman, U., & Djerf-Pierre, M. (2013). The Social Journalist: Embracing the Social Media Life or Creating a New Digital Divide? Digital Journalism, 1(3), 368-385.

Heidenreich, T., Eberl, J. M., Lind, F., & Boomgaarden, H. (2020). Political migration discourses on social media: a comparative perspective on visibility and sentiment across political Facebook accounts in Europe. Journal of Ethnic and Migration Studies, 46(7), 1261-1280. https://doi.org/10.1080/1369183X.2019.1665990

Hemphill, L., Schöpke-Gonzalez, A. & Panda, A. (2022). Comparative sensitivity of social media data and their acceptable use in research. Sci Data, 9, 643. https://doi.org/10.1038/s41597-022-01773-w

Herbst, S. (1998). Reading public opinion: How political actors view the democratic process. Chicago: University of Chicago Press.

Hill, C. A., Dean, E., & Murphy, J. (2013). Social media, sociality, and survey research. John Wiley & Sons.

HIMSS Analytics. (2019). Annual European eHealth Survey 2019. eHealth Trendbarometer. https://europe.himssanalytics.org/sites/himssanalytics_europe/files/eHealth%20TRENDBAROMETER%20-%20HIMSS%20Analytics%20Annual%20European%20eHealth%20Survey%202019.pdf

Hogenraad, R., McKenzie, D. P., & Péladeau, N. (2003). Force and influence in content analysis: the production of new social knowledge. Quality and Quantity, 37(3), 221-238. https://doi.org/10.1023/A:1024401325472

Holzinger, A. (2014). Social Media Mining and Social Network Analysis: Emerging Research. Online Information Review, 38(1), 157-158. https://doi.org/10.1108/OIR-08-2013-0200

Hooks, B. (2000). Feminist Theory: From Margin to Centre. London: Pluto Press.

Hothorn, T., & Zeileis, A. (2015). partykit: A modular toolkit for recursive partytioning in R. The Journal of Machine Learning Research, 16(1), 3905-3909. https://www.jmlr.org/papers/volume16/hothorn15a/hothorn15a.pdf

Hsieh, Y. P., & Murphy, J. (2017). Total Twitter Error. In Biemer, P. P., de Leeuw, E. D., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., ... & West, B. T. (Eds.), Total Survey Error in Practice, pp. 23-46. Hoboken, NJ: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119041702.ch2

Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 168-177.

Huges, A., & Wojcik, S. (2019). 10 Facts About Americans and Twitter. Pew Research Center. https://www.pewresearch.org/fact-tank/2019/08/02/10-facts-about-americans-and-twitter/

Husson, F., J. Josse, S. Le, & Mazet, J. (2013). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R. R package, version 1(1.29). https://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. Applied Linguistics, 17, 433-454.

Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated Text Analysis in Psychology: Methods, Applications, and Future Developments. Language and Cognition, 7 (2), 265-290. https://doi.org/10.1017/langcog.2014.30

Jacobs, K., & Spierings, N. (2016). Saturation or maturation? The diffusion of Twitter and its impact on preference voting in the Dutch general elections of 2010 and 2012. Journal of Information Technology & Politics, 13(1), 1-21. https://doi.org/10.1080/19331681.2015.1131652

Jacobs, K., & Spierings, N. (2019). A populist paradise? Examining populists' Twitter adoption and use. Information, Communication & Society, 22(12), 1681-1696. https://doi.org/10.1080/1369118X.2018.1449883

Jaidka, K. (2022). Talking politics: Building and validating data-driven lexica to measure political discussion quality. Computational Communication Research, 4(2), 486-527. https://doi.org/10.5117/CCR2022.2.005.JAID

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O'Neill, C. & Usher, A. (2015). Big data in survey research: AAPOR task force report. Public Opinion Quarterly, 79(4): 839-880.

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C.,Lane, J., O'Neil, C., & Usher, A. (2015). Big data in survey research: AAPOR task force report. Public Opinion Quarterly, 79(4), 839-880. https://doi.org/10.1093/poq/nfv039

Jiang, S. (2019). Functional interactivity in social media: an examination of Chinese health care organizations' microblog profiles. Health Promotion International, 34(1), 38-46. https://doi.org/10.1093/heapro/dax056

Jiang, T., Hou, Y., & Wang, Q. (2016). Does Micro-blogging Make Us shallow? Sharing Information Online Interferes with Information Comprehension. Computers in Human Behavior, 59: 210-214.

Johnson, T. P., & Smith, T. W. (2017). Big data and survey research: Supplement or substitute? In Thakuriah, P. Tilahun, N., & Zellner, M. (Eds.), Seeing cities through big data: Research, methods and applications in urban informatics, pp. 113-125. New York: Springer. https://www.springer.com/gp/book/9783319409009

Joseph, K., Friedland, L., Hobbs, W., Lazer, D., & Tsur, O. (2017), ConStance: Modeling Annotation Contexts to Improve Stance Classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark: Association for Computational Linguistics, pp. 1115-1124. http://aclweb.org/anthology/D17-1116

Joseph, K., Shugars, S., Gallagher, R., Green, J., Mathé, A. Q., An, Z., & Lazer, D. (2021). (Mis) alignment Between Stance Expressed in Social Media Data and Public Opinion Surveys. arXiv preprint. https://doi.org/2109.01762

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). Bag of Tricks for Efficient Text Classification. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2, 427-431.

Jungherr, A. (2016). Twitter use in election campaigns: a systematic literature review. J Inform Technol Polit, 13(1):72-91. https://doi.org/10.1080/19331681.2015.1132401

Jungherr, A., & Theocharis, Y. (2017). The empiricist's challenge: Asking meaningful questions in political science in the age of big data. Journal of Information Technology & Politics, 14(2), 97-109. https://doi.org/10.1080/19331681.2017.1312187

Kamocki, P., Hannesschläger, V., Hoorn, E., Kelli, A., Kupietz, M., Lindén, K., & Puksas, A. (2022). Legal issues related to the use of Twitter data in language research. In Selected papers from the CLARIN Annual Conference 2021. Linköping University Electronic Press.

Kannangara, S., & Wobcke, W. (2021). Determining political interests of issue-motivated groups on social media: joint topic models for issues, sentiment and stance. J Comput Soc Sc. https://doi.org/10.1007/s42001-021-00146-4

Kantola, J., & Lombardo, E. (2020). Strategies of Right Populists in Opposing Gender Equality in a Polarized European Parliament. International Political Science Review, 42 (5), 565-79.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business horizons, 53(1), 59-68. https://doi.org/10.1016/j.bushor.2009.09.003

Kar, A. K., & Dwivedi, Y. K. (2020). Theory building with big data-driven research-Moving away from the What towards the Why. International Journal of Information Management, 54, 102205.

Karlsen, R., & Enjolras, B. (2016). Styles of social media campaigning and influence in a hybrid political communication system: Linking candidate survey data with Twitter data. The International Journal of Press/Politics, 21(3), 338-357.

Katz & Lazarsfeld (1955). "Personal Influence". New York: Free Press.

Keane, J. (1991). Democracy and the Media. International Social Science Journal, 129 (s1), 523-540.

Keeter, S., Kennedy, C., Clark, A., Tompson, T., & Mokrzycki, M. (2007). What's Missing from National Landline RDD Surveys?: The Impact of the Growing Cell-Only Population. Public Opinion Quarterly, 71(5), 772-792. https://doi.org/10.1093/poq/nfm053

Keller, T. R. (2020). To Whom Do Politicians Talk and Listen?: Mapping Swiss Politicians' Public Sphere on Twitter. Computational Communication Research, 2(2), 175-202.

Kertzer, J. D., & Zeitzoff, T. (2017). A bottom-up theory of public opinion about foreign policy. American Journal of Political Science, 61(3), 543-558.

Key, V. O. (1961). Public opinion and the decay of democracy. The Virginia Quarterly Review, 37(4), 481-494.

Kim, D., & Kim, J. (2014). Public Opinion Sensing and Trend Analysis on Social Media: A Study on Nuclear Power on Twitter, International Journal of Multimedia and Ubiquitous Engineering, 9(11): 373-384. http://dx.doi.org/10.14257/ijmue.2014.9.11.36

King, G. (2011). Ensuring the data-rich future of the social sciences. Science, 331(6018), 719-721. https://doi.org/10.1126/science.1197872

Kirkwood, S., V. Cree, D. Winterstein, A. Nuttgens, & Sneddon, J. (2018). Encountering #feminism on Twitter: Reflections on a Research Collaboration Between Social Scientists and Computer Scientists. Sociological Research Online, 23 (4), 763-79.

Klašnja, M., Barberá, P., Beauchamp, N., Nagler, J., & Tucker, J.A. (2018). Measuring public opinion with social media data. In Atkeson, L.R., & Alvarez, R.M. (eds.), The Oxford handbook of polling and survey methods, pp. 555-582. Oxford: Oxford University Press.

Kletecka-Pulker, M., Mondal, H., Wang, D., Parra, R. G., Maigoro, A. Y., Lee, S., . . . & Atanasov, A. G. (2021). Impacts of biomedical hashtag-based Twitter campaign:# DHPSP utilization for promotion of open innovation in digital health, patient safety, and personalized medicine. Current Research in Biotechnology, 3, 146-153. https://doi.org/10.1016/j.crbiot.2021.04.004

Koteyko, N, Hunt, D, & Gunter, B. (2015). Expectations in the field of the internet and health: an analysis of claims about social networking sites in clinical literature. Sociol Health Illn. Mar, 37(3), 468-484. https://doi.org/10.1111/1467-9566.12203

Kraft, P. W. (2018). Measuring Morality in Political Attitude Expression. Journal of Politics, 80 (3), 1028-1033. https://doi.org/10.1086/696862

Krasodomski-Jones, A. (2016). Talking to ourselves? Political debate online and the echo chamber effect. London: Demos.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. Public Opinion Quarterly, 72(5), 847-865. https://doi.org/10.1093/poq/nfn063

Kriesi, H., Morlino, L., Magalhães, P., Alonso, S., & Ferrín, M. (2013). European Social Survey Round 6 Module on Europeans' Understandings and Evaluations of Democracy - Final Module in Template. London: Centre for Comparative Social Surveys, City University London. https://www.europeansocialsurvey.org/docs/round6/questionnaire/ESS6_final_understandings_and_evaluation_of_democracy_module_template.pdf

Küçük, D., & Can, F. (2020). Stance detection: A survey. ACM Computing Surveys, 53(1), 1-37. https://doi.org/10.1145/3369026

Kushwaha, A. K., Kar, A. K., & Dwivedi, Y. K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. International Journal of Information Management Data Insights, 1(2), Article 100017. https://doi.org/10.1016/j.jjimei.2021.100017

Kyawa, T., & Aungb, S. S. (2020). Classification of the Stance in Online Debates Using the Dependency Relations Feature. International Journal of Computer, 38(1), 153-163. https://core.ac.uk/download/pdf/327266594.pdf

Ladd, J., Ryan, R., Singh, L., Bode, L., Budak, C., Conrad, F., ... & Traugott, M. (2020). Measurement considerations for quantitative social science research using social media data. White Paper. https://doi.org/10.31234/osf.io/ga6nc

Lasorsa, D. L. (1991). Political Outspokenness: Factors Working Against the Spiral of Silence. Journalism Quarterly, 68 (1), 131-40.

Latzer, M. (2017). Strong Digital Well-Being in Switzerland. University of Zürich. https://www.media.uzh.ch/en/Press-Releases/2017/Digital-Well-Being-.html

Laver, M., Benoit, K., & Garry, J. (2003). Extracting Policy Positions from Political Texts Using Words as Data. American Political Science Review, 97 (2), 311-331.

Lawrence, E., & Ringrose, J. (2018). @NoToFeminism, #FeministsAreUgly, and Misandry Memes: How Social Media Feminist Humor is Calling Out Antifeminism. In Keller, J., & Ryan, M. (Eds.), Emergent Feminisms: Complicating a Postfeminist Media Culture, pp. 211-32. New York, NY: Routledge.

Lazer, D. M., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., ... & Wagner, C. (2020). Computational social science: Obstacles and opportunities. Science, 369(6507), 1060-1062. https://doi.org/10.1126/science.aaz8170

Lazer, D., & Radford, J. (2017). Data ex machina: Introduction to big data. Annual Review of Sociology, 43, 19-39.

LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., ... & H2O. ai. (2018). Package 'h2o'. https://cran.r-project.org/web/packages/h2o/h2o.pdf

Ledford, H. (2020). How Facebook, Twitter and other data troves are revolutionizing social science. Nature, 582(7812), 328-331. https://www.nature.com/articles/d41586-020-01747-1

Lee, J., Kim, J., Hong, Y. J., Piao, M., Byun, A., Song, H., & Lee, H. S. (2019). Health Information Technology Trends in Social Media: Using Twitter Data. Healthcare informatics research, 25(2), 99-105. https://doi.org/10.4258/hir.2019.25.2.99

Lee, J.-S., & Nerghes, A. (2018). Refugee or Migrant Crisis? Labels, Perceived Agency, and Sentiment Polarity in Online Discussions. Social Media + Society. https://doi.org/10.1177/2056305118785638

Lee, T. (2002). Mobilizing Public Opinion: Black Insurgency and Racial Attitudes in the Civil Rights Era (Studies in Communication, Media, and Public Opinion): Taeku Lee: 9780226470252: Amazon.com: Books (1st edition). Chicago: University of Chicago Press. https://www.amazon.com/Mobilizing-Public-Opinion-Insurgency-Communication/dp/0226470253

Li, Y., & Caragea, C. (2019). Multi-task stance detection with sentiment and stance lexicons. In Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), pp. 6299-6305. https://aclanthology.org/D19-1657

Lind, F., Eberl, J.-M., Heidenreich, T., & Boomgaarden, H. G. (2019). When the journey is as important as the goal: A roadmap to multilingual dictionary construction. International Journal of Communication, 13. https://ijoc.org/index.php/ijoc/article/view/10578

Lorenzetti, M. I. (2020). Right-Wing Populism and the Representation of Immigrants on Social Media. A Critical Multimodal Analysis. Iperstoria, 15: 59-95.

Loughran, T., & McDonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance, 66 (1), 35-65. https://doi.org/10.1111/j.1540-6261.2010. 01625.x

Loureiro, M. L., & Alló, M. (2020). Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the U.K. and Spain. Energy Policy, 143, 111490. https://doi.org/10.1016/j.enpol.2020.111490

Loyola-Gonzalez, O. (2019). Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. IEEE Access, 7, 154096-154113. https://doi.org/10.1109/ACCESS.2019.2949286

Lukoianova, T., & Rubin, V. L. (2013). Veracity Roadmap: Is Big Data Objective, Truthful and Credible? Advances in Classification Research Online, 24(1), 4-15. https://doi.org/10.7152/acro.v24i1.14671

Lupton, D. (2012). M-health and health promotion: the digital cyborg and surveillance society. Social Theory and Health, 10(3), 229-244.

Lyon, D. (2010). Surveillance, power and everyday life. In Kalantzis-Cope, P., & Gherab-Martin, K. (Eds.), Emerging Digital Spaces in Contemporary Society, pp. 107-120. Houndsmills, UK: Palgrave Macmillan.

Maerz, S. F. (2019). Simulating Pluralism: The Language of Democracy in Hegemonic Authoritarianism. Political Research Exchange, 1 (1), 1-23. https://doi.org/10.1080/2474736X.2019.1605834

Mahdikhani, M. (2021). Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of COVID-19 pandemic. International Journal of Information Management Data Insights, 100053. https://doi.org/10.1016/j.jjimei.2021.100053

Mahoney, J., Le Louvier, K., Lawson, S., Bertel, D., & Ambrosetti, E. (2022). Ethical considerations in social media analytics in the context of migration: lessons learned from a Horizon 2020 project. Research Ethics, 18(3), 226–240. https://doi.org/10.1177/17470161221087542

Maia, R. C. M. (2017). Politicization, new media, and everyday deliberation. In Fawcett, P. Flinders, M., Hay, C., & Wood, M. (Eds.), Anti-politics, depoliticization, and governance, pp. 68-90. Oxford: Oxford University Press. http://doi.org/10.1093/oso/9780198748977.003.0004

Mancosu, M., & Bobba, G. (2019). Using deep-learning algorithms to derive basic characteristics of social media users: The Brexit campaign as a case study. Plos one, 14(1), e0211013. https://doi.org/10.1371/journal.pone.0211013

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60. https://stanfordnlp.github.io/CoreNLP

Matthes, J. & Kohring, M. (2008). The Content Analysis of Media Frames: Toward Improving Reliability and Validity. Journal of Communication, 58, 258-279.

McCallum, A. K. (2002). MALLET: A Machine Learning for Language Toolkit. http://mallet.cs.umass.edu

McClure, D. (2015). Textplot refresh - Python 3, PyPI, CLI app. http://dclure.org/tutorials/textplot-refresh/

McGregor, S.C. (2019). Social media as public opinion: how journalists use social media to represent public opinion. Journalism, 20(8):1070-1086. https://doi.org/10.1177/1464884919845458

Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. Research & Politics, 4(3), 2053168017720008.

Merklé, P. (2017.) Une traversé de l'adolescence. Culture, classes, réseaux. Thesis, Institut d'études politiques, Université de Paris, FR.

Metzler, K., Kim, D. A., Allum, N., & Denman, A. (2016). Who is doing computational social science? Trends in Big Data research.

Michalke, M., Brown, E., Mirisola, A., Brulet, A., Hauser, L., & Michalke, M. M. (2021). Package 'koRpus'. https://www.reaktanz.de/R/pckg/koRpus/koRpus.pdf

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Google Books Team, ... & Aiden, E. L. (2010). Quantitative Analysis of Culture Using Millions of Digitized Books. Science, 331 (6014), 176-182.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and Their Compositionality. arXiv preprint arXiv:1310.4546.

Milioni, D. L., Spyridou, L. P., & Vadratsikas, K. (2015). Framing immigration in online media and television news in crisis-stricken Cyprus. The Cyprus Review, 27(1): 155-185.

Miller, G. A. (1995). WordNet: A Lexical Database for English. Communications of the ACM, 38 (11), 39-41. https://doi.org/10.1145/219717.219748

Mirbabaie, M., F. Brünker, M. Wischnewski, & Meinert, J. (2021). The Development of Connective Action During Social Movements on Social Media. ACM Transactions on Social Computing, 4 (1), 1-21.

Mislove, A., Lehmann, S., Ahn, Y. Y., Onnela, J. P., & Rosenquist, J. (2011). Understanding the demographics of Twitter users. In Proceedings of the International AAAI Conference on Web and Social Media, 5(1), 554-557. https://ojs.aaai.org/index.php/ICWSM/article/view/14168

Modrek, S., & Chakalov, B. (2019). The #MeToo Movement in the United States: Text Analysis of Early Twitter Conversations. Journal of Medical Internet Research, 21 (9), e13837.

Mohammad, S. M., & Turney, P. D. (2013). Nrc emotion lexicon. National Research Council, Canada, 2. http://www.saifmohammad.com/WebDocs/NRCemotionlexicon.pdf.

Mohammad, S. M., Sobhani, P., & Kiritchenko, S. (2017). Stance and sentiment in tweets. ACM Transactions on Internet Technology, 17(3), 1-23.

Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., & Cherry, C. (2016). Semeval-2016 task 6: Detecting stance in tweets. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), pp. 31-41. https://aclanthology.org/S16-1003.pdf

Monroe, B. L., & Schrodt, P. A. (2008). Introduction to the special issue: The statistical analysis of political text. Political Analysis, 16(4), 351-355. https://doi.org/10.1093/pan/mpn017

Morselli, D., S. Passini, & McGarty, C. (2021). Sos Venezuela: An Analysis of the Anti-Maduro Protest Movements Using Twitter. Social Movement Studies, 20 (5), 509-30.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. (2013). Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. In Proceedings of the international AAAI conference on web and social media, 7(1), 400-408. https://ojs.aaai.org/index.php/ICWSM/article/view/14401

Moy, P., & Murphy, J. (2016). Problems and Prospects in Survey Research. Journalism & Mass Communication Quarterly. (Sage CA: Los Angeles, CA). https://doi.org/10.1177/1077699016631108

Munro, G. D., & Ditto, P. H. (1997). Biased Assimilation, Attitude Polarization, and Affect in Reactions to Stereotype-Relevant Scientific Information. Personality and Social Psychology Bulletin, 23 (6), 636-53.

Murphy, J., Dean, E., Hill, C. A., & Richards, A. (2011). Social media, new technologies, and the future of health survey research. National Center for Health Statistics (Ed.), Proceedings of the 10th Conference on Health Survey Research Methods, pp. 231-241.

National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Committee on Science, Engineering, Medicine, and Public Policy; Board on Research Data and Information; Division on Engineering and Physical Sciences; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Analytics; Division on Earth and Life Studies; Nuclear and Radiation Studies Board; Division of Behavioral and Social Sciences and Education; Committee on National Statistics; Board on Behavioral, Cognitive, and Sensory Sciences; Committee on Reproducibility and Replicability in Science. Reproducibility and Replicability in Science. Washington (DC): National Academies Press (US). (2019). Understanding Reproducibility and Replicability. https://www.ncbi.nlm.nih.gov/books/NBK547546

Neisser, U. (1976). Cognition and reality: Principles and Implications of Cognitive Psychology. New York: WH Freeman/Times Books/Henry Holt and Co.

Neuendorf, K. A. (2016). The Content Analysis Guidebook. Thousand Oaks: Sage.

Ng, A. (2019). Machine learning yearning: Technical strategy for ai engineers in the era of deep learning. https://github.com/ajaymache/machine-learning-yearning

Ng, L. H. X., & Carley, K. M. (2022). Is my stance the same as your stance? A cross validation study of stance detection datasets. Information Processing & Management, 59(6), 103070. https://doi.org/10.1016/j.ipm.2022.103070

Nicholls, T., & Culpepper, P. D. (2020). Computational identification of media frames: Strengths, weaknesses, and opportunities. Political Communication, 38(1), 159-181. https://doi.org/10.1080/10584609.2020.1812777

Nielsen, F. A. (2011). AFINN. Denmark: Informatics; Mathematical Modelling. Technical University of Denmark. https://bit.ly/2pZzWL4

Noelle-Neumann, E. (1984). The Spiral of Silence: Public Opinion-Our Social Skin. Chicago: University of Chicago Press.

O'Connor, B., Balasubramanyan, R., Routledge, B.R., & Smith, N.A. (2010). From tweets to polls: linking text sentiment to public opinion time series. Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, 4(1), 122-129. https://ojs.aaai.org/index.php/ICWSM/article/view/14031

O'Regan, V., & Riordan, E. (2018). Comparing the representation of refugees, asylum seekers and migrants in the Irish and UK press: A corpus-based critical discourse analysis. Journal of Language and Politics, 17(6): 744-768.

Obembe, D., Kolade, O., Obembe, F., Owoseni, A., & Mafimisebi, O. (2021). COVID-19 and the tourism industry: an early stage sentiment analysis of the impact of social media and stakeholder communication. International Journal of Information Management Data Insights, 1(2). https://doi.org/10.1016/j.jjimei.2021.100040

ODI, & Chatham House. (2017) 'Changing public perceptions of refugees and migrants: the role of politicians, the media and civil society'. Event summary. London: ODI and Chatham House. www.chathamhouse.org/event/changing-public-perceptions-refugees-and-migrants-role-politicians-media-and-civil-society

OECD (Organisation for Economic Co-operation and Development). (2019). The Heavy Burden of Obesity: The Economics of Prevention. Paris: OECD Publishing. https://doi.org/10.1787/67450d67-en

OECD (Organisation for Economic Co-operation and Development). (2020). ICT Access and Usage by Households and Individuals. OECD Telecommunications and

Internet Statistics. Database accessed in January 2021. https://stats.oecd.org/Index.aspx?DataSetCode=ICT_HH2

Olteanu, A., Castillo, C., Diaz, F., & Kiciman, E. (2016). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries (SSRN Scholarly Paper No. ID 2886526). Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2886526

Opgenhaffen, M., & d'Haenens, L. (2011). The Impact of Online News Features on Learning From News. A Knowledge Experiment. International Journal of Internet Science, 6(1), 8-28.

Pandey, R., Purohit, H., Castillo, C., & Shalin, V. L. (2022). Modeling and mitigating human annotation errors to design efficient stream processing systems with human-in-the-loop machine learning. International Journal of Human-Computer Studies, 160, 102772. https://doi.org/10.1016/j.ijhcs.2022.102772

Pariser, E. (2011). The Filter Bubble: What the Internet Is Hiding from You. London: Penguin UK. Price, Vincent. 1988. On the Public Aspects Of Opinion: Linking Levels of Analysis in Public Opinion Research. Communication Research, 15(6), 659-679.

Park, M., Park, J., Baek, Y. M., & Macy, M. (2017). Cultural values and cross-cultural video consumption on YouTube. PLOS ONE, 12(5), e0177865. https://doi.org/10.1371/journal.pone.0177865

Pasek, J., McClain, C. A., Newport, F., & Marken, S. (2020). Who's tweeting about the President? What big survey data can tell us about digital traces?. Social Science Computer Review, 38(5), 633-650. https://doi.org/10.1177/0894439318822007

Patra, B. G., Das, D., & Bandyopadhyay, S. (2016). JU_NLP at semeval-2016 task 6: Detecting stance in Tweets using support vector machines. In Proceedings of the 10th International Workshop on Semantic Evaluation, pp. 440-444.

Pavalanathan, U., & De Choudhury, M. (2015). Identity management and mental health discourse in social media. Proceedings of the 24th International Conference on World Wide Web, 315-321. ACM.

Peluchette, J. V., Karl, K. A., & Coustasse, A. (2016). Physicians, patients, and Facebook: Could you? Would you? Should you? Health marketing quarterly, 33(2), 112-126. https://doi.org/10.1080/07359683.2016.1166811

Pennebaker, J. W., Francis, M., & Booth, R. (2001). Linguistic Inquiry and Word Count: LIWC 2001. Mahwah, NJ: Erlbaum.

Pennebaker, J. W.,Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The Development and Psychometric Properties of LIWC2015. https://repositories.lib.utexas.edu/handle/2152/31333.

Pentland, S. J., Fuller, C. M., Spitzley, L. A., & Twitchell, D. P. (2022). Does accuracy matter? Methodological considerations when using automated speech-to-text for social science research. International Journal of Social Research Methodology, 1-17. https://doi.org/10.1080/13645579.2022.2087849

Persily, N., & Tucker, J. A. (2020). Social media and democracy: The state of the field, prospects for reform. Cambridge: Cambridge University Press.

Pew Research Center. (2016). Social media causes some users to rethink their views on an issue. Consulted March 2020: https://www.pewresearch.org/fact-tank/2016/11/07/social-media-causes-some-users-to-rethink-their-views-on-an-issue/

Pew Research Center. (2020). News Use Across Social Media Platforms in 2020. Consulted on February 2022: https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/

Plescia, C., Kritzinger, S., & Sio, L. D. (2019). Filling the Void? Political Responsiveness of Populist Parties. Representation, 55(4), 513-533. https://doi.org/10.1080/00344893.2019.1635197

Poell, T., R. Abdulla, B. Rieder, R. Woltering, & L. Zack. (2015). Protest Leadership in the Age of Social Media. Information, Communication & Society, 19 (7), 994-1014.

Posegga, O., & Jungherr, A. (2019). Characterizing Political Talk on Twitter: A Comparison Between Public Agenda, Media Agendas, and the Twitter Agenda with Regard to Topics and Dynamics. In Proceedings of the 52nd Hawaii International Conference on System Sciences.

Potamias, R. A., Siolas, G., & Stafylopatis, A. G. (2020). A transformer-based approach to irony and sarcasm detection. Neural Computing and Applications, 32(23), 17309-17320. https://doi.org/10.1007/s00521-020-05102-3

Quattrociocchi, W., A. Scala, & Sunstein, C. R. (2016). Echo Chambers on Facebook. SSRN Electronic Journal, 1-15. https://doi.org/10.2139/ssrn.2795110

Radford, J., & Joseph, K. (2020). Theory in, theory out: the uses of social theory in machine learning for social science. Frontiers in big Data, 3, 18. https://doi.org/10.3389/fdata.2020.00018

Rajendran, P., Bollegala, D., & Parsons, S. (2016). Contextual stance classification of opinions: A step towards enthymeme reconstruction in online reviews. In Proceedings of the Third Workshop on Argument Mining. Berlin, Germany: Association for Computational Linguistics, pp. 31-39. http://aclweb.org/anthology/W16-2804

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Truthy: Mapping the Spread of Astroturf in Microblog Streams. Proceedings of the 20th international conference companion on World wide web, 249-252. ACM.

Rauchfleisch, A., & Metag, J. (2016). The special case of Switzerland: Swiss politicians on Twitter. New Media & Society, 18(10), 2413-2431. https://doi.org/10.1177/1461444815586982

Rauchfleisch, A., Vogler, D., & Eisenegger, M. (2021). Public Sphere in crisis mode: How the COVID-19 pandemic influenced public discourse and user behaviour in the Swiss twitter-sphere. Javnost-The Public, 28(2), 129-148. https://doi.org/10.1080/13183222.2021.1923622

Reason, J. (2000). Human error: models and management. Bmj, 320(7237), 768-770. https://doi.org/10.1136/bmj.320.7237.768

Reuters. (2016). Digital News Report. Switzerland. http://www.digitalnewsreport.org/survey/2016/switzerland-2016

Reuters. (2018). Digital News Report. Switzerland. http://www.digitalnewsreport.org/survey/2018/switzerland-2018

Reveilhac, M., & Morselli, D. (2022). Dictionary-based and machine learning classification approaches: A comparison for tonality and frame detection on Twitter data. Political Research Exchange. https://doi.org/10.1080/2474736X.2022.2029217

Reveilhac, M., & Schneider, G. (2022). Complementing Kernel Density Estimation and Topic Modelling to Visualise Political Discourse. In Proceedings of the Digital Research Data and Human Sciences (DRDHum) - Diversity of Methods and Materials, Jyväskylä, Finland, 1-3 December 2022.

Reveilhac, M., Steinmetz, S., & Morselli, D. (2022). A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion. Multimedia Tools and Applications, 1-36.

Riekert, M., Riekert, M., & Klein, A. (2021). Simple Baseline Machine Learning Text Lassifiers for Small Datasets. SN Computer Science, 2 (178), 1-16.

Rinker, T. (2021). Package 'sentimentr': Calculate Text Polarity Sentiment. https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf

Rinker, T. W. (2018). Textstem: Tools for Stemming and Lemmatizing Text. http://github.com/trinker/textstem

Rinker, T. W. (2019). sentimentr: Calculate Text Polarity Sentiment. http://github.com/trinker/sentimentr

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... & Rand, D. G. (2014). Structural topic models for open-ended survey responses. American journal of political science, 58(4), 1064-1082. https://doi.org/10.1111/ajps.12103

Robillard, J. M., Whiteley, L., Johnson, T. W., Lim, J., Wasserman, W. W., & Illes, J. (2013). Utilizing Social Media to Study Information-Seeking and Ethical Issues in Gene Therapy. Journal of Medical Internet Research, 15(3), e2313. https://doi.org/10.2196/jmir.2313

Rocchio, J. (1971). Relevance Feedback in Information Retrieval. In Dierk, S. F. (Ed.), The SMART Retrieval System: Experiments in Automatic Document Processing, pp. 313-323. Englewood Cliffs, NJ: Prentice-Hall Inc.

Roggeband, C. (2018). The Good, The Bad, and the Ugly: Making Sense of Opposition to Feminisms from a Social-Movement Perspective. In Varieties of Opposition to Gender Equality in Europe, edited by M. Verloo, 19-37. London: Routledge.

Rooduijn, M. (2020). Immigration attitudes have barely changed - so why is far right on rise?. The Guardian. www.theguardian.com/world/2020/mar/02/immigration-attitudes-havebarely-changed-why-far-right-on-rise

Rooduijn, M., & Pauwels, T. (2011). Measuring Populism: Comparing two Methods of Content Analysis. West European Politics, 34 (6), 1272-1283. https://doi.org/10.1080/01402382.2011.616665

Ross, G. M. (2021). I use a COVID-19 contact-tracing app. Do you? Regulatory focus and the intention to engage with contact-tracing technology. International Journal of

Information Management Data Insights, 1(2), Article 100045. https://doi.org/10.1016/j.jjimei.2021.100045
Rossier, T. (2018). Construire au moyen d'une analyse des correspondances multiples dynamique l'espace des devenirs soixante-huitards. In Fillieule, O., Sommier, I., Masclet, C., & Hirsch, T. (Eds.), Changer le monde, changer sa vie, pp. 1087-1106. Paris: Acte sud.
Rossier, T., & Fillieule, O. (2019). Devenir(s) militants: Proposition de méthode pour une exploration des conséquences biographiques de l'engagement des soixante-huitards. Revue Française de Science Politique, 69(4), 631-683.
Rothschild, D., & Wolfers, J. (2011). Forecasting elections: Voter intentions versus expectations. Available at SSRN, 1884644.
Rousidis, D., Koukaras, P., & Tjortjis, C. (2020). Social media prediction: A literature review. Multimedia Tools and Applications, 79, 1-33. https://doi.org/10.1007/s11042-019-08291-9
Russell, H. M., W. Tervo, D. L. Ariail, & Smith, L. M. (2020). Relationship of Economic Freedom to Economic Performance, Gender Equality, and Social Progress. The World Economic, 21, 171-90.
Rutter, J. and Carter, R. (2018). The national conversation on immigration: final report. London: British Future. www.britishfuture.org/publication/national-conversation-immigration-final-report
Sahlgren, M. (2006). The Word-Space Model: Using distributional Analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces. Ph.D. thesis, Stockholm University.
Sahlgren, M., & Lenci, A. (2016). The Effects of Data Size and Frequency Range on Distributional Semantic Models. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 975-980. Austin/Texas: Association for Computational Linguistics. https://aclanthology.org/D16-1099/
Sajuria, J., Fábrega, J. (2016). Do We Need Polls? Why Twitter Will Not Replace Opinion Surveys, but Can Complement Them. In Snee, H., Hine, C., Morey, Y., Roberts, S., & Watson, H. (Eds.), Digital Methods for Social Science, pp. 87-104. Palgrave Macmillan, London. https://doi.org/10.1057/9781137453662_6
Salganik, M. J. (2019). Bit by bit: Social research in the digital age. Princeton University Press.

Scarborough, W. J. (2018). Feminist Twitter and Gender Attitudes: Opportunities and Limitations to Using Twitter in the Study of Public Opinion. Socius, 4, 1-16.

Scarborough, W. J. (2018). Feminist Twitter and Gender Attitudes: Opportunities and Limitations to Using Twitter in the Study of Public Opinion. Socius, 4, 2378023118780760. https://doi.org/10.1177/2378023118780760

Schiller, B., Daxenberger, J., & Gurevych, I. (2021). Stance detection benchmark: How robust is your stance detection?. KI-Künstliche Intelligenz, 35(3), 329-341. https://doi.org/10.1007/s13218-021-00714-w

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing, Manchester, UK.

Schmidt, B. (2017). Word Vectors. An R package for building and exploring word embedding models. https://github.com/bmschmidt/wordVectors

Schneider, G. (2022). Medical topics and style from 1500 to 2018. In Turo Hiltunen and Irma Taavitsainen (eds.) Corpus pragmatic studies on the history of medical discourse. Amsterdam: Benjamins.

Schneider, G., El-Assady, M., & Lehmann, H. M. (2017). Tools and methods for processing and visualizing large corpora. Studies in Variation, Contacts and Change in English, 19.

Schober MF, Conrad FG, Holzberg J, Ferg RA, Katz J, Childs JH, Beatty PC, Gagnon-Bartsch JA (2020) Better understanding when and how social media posts can augment public opinion surveys. Paper presented at the BigSurv20 big data meets survey science. http

Schober, M. F., Pasek, J., Guggenheim, L., Lampe, C., & Conrad, F. G. (2016). Social Media Analyses for Social Measurement. Public Opinion Quarterly, 80 (1), 180-211. https://doi.org/10.1093/poq/nfv048

Schober, M. F., Pasek, J., Guggenheim, L., Lampe, C., & Conrad, F. G. (2016). Social Media Analyses for Social Measurement. Public Opinion Quarterly, 80(1), 180-211. https://doi.org/10.1093/poq/nfv048

Schober, M.F., Conrad, F.G., Holzberg, J., Ferg, R.A., Katz, J., Childs, J.H., Beatty, P.C., & Gagnon-Bartsch, J.A. (2020). Better understanding when and how social media posts can augment public opinion surveys. Paper presented at the BigSurv20 big data meets survey science. Accessed 13 November 2020: https://www.bigsurv20.org/conf20/program/?sess=11#100

Schoonvelde, M., Schumacher, G., & Bakker, B. N. (2019). Friends with text as data benefits: Assessing and extending the use of automated text analysis in political science and political psychology. Journal of Social & Political Psychology, 7(1), 124-143. https://doi.org/10.5964/jspp.v7i1.964

Schumacher, S., & Kent, N. (2020). 8 charts on internet use around the world as countries grapple with COVID-19. Pew Research Center https://www.pewresearch.org/fact-tank/2020/04/02/8-charts-on-internet-use-around-the-world-as-countries-grapple-with-COVID-19/

Schwartz, C., Simon, M., Hudson, D. and van-Heerde-Hudson, J. (2020). A populist paradox? How Brexit softened anti-immigrant attitudes. British Journal of Political Science, 51(3): 1160-1180. https://doi.org/10.1017/S0007123419000656

Schwartz, H. A., & Ungar, L. H. (2015). Data-Driven Content Analysis of Social Media: A Systematic Overview of Automated Methods. The ANNALS of the American Academy of Political and Social Science, 659 (1), 78-94. https://doi.org/10.1177/0002716215569197

Scott, B. K., Miller, G. T, Fonda, S. T., Yeaw, R. E., Gaudaen, J. C., Pavliscsak, H. H., Quinn, M. T., & Pamplin, J. C. (2020). Advanced Digital Health Technologies for COVID-19 and Future Emergencies. Telemedicine and e-Health, 26(10), 1226-1233. http://doi.org/10.1089/tmj.2020.0140

Segerberg, A. (2017). Online and social media campaigns for climate change engagement. In Oxford Research Encyclopedia of Climate Science. https://doi.org/10.1093/acrefore/9780190228620.013.398

Segerberg, A., & Bennett, W. L. (2011). Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests. The Communication Review, 14(3), 197-215. https://doi.org/10.1080/10714421.2011.597250

Sen, I., Flöck, F., & Wagner, C. (2020). On the reliability and validity of detecting approval of political actors in tweets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 1413-1426.

Sen, I., Floeck, F., Weller, K., Weiss, B., & Wagner, C. (2019). A Total Error Framework for Digital Traces of Humans. ArXiv:1907.08228 [Cs]. http://arxiv.org/abs/1907.08228

Shen, D., Wang, G., Wang, W., Min, M. R., Su, Q., Zhang, Y., Li, C., Henao, R., & Carin, L. (2018). Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 1, 440-450.

Shin, J. (2020). How Do Partisans Consume News on Social Media? A Comparison of Self-Reports With Digital Trace Measures Among Twitter Users. Social Media + Society, 6(4), 2056305120981039. https://doi.org/10.1177/2056305120981039

Shmueli, G. (2010). To explain or to predict?. Statistical science, 25(3), 289-310. https://doi.org/10.1214/10-STS330

Singh, K. N., Devi, S. D., Devi, H. M., & Mahanta, A. K. (2022). A novel approach for dimension reduction using word embedding: An enhanced text classification approach. International Journal of Information Management Data Insights, 2(1), Article 100061. https://doi.org/10.1016/j.jjimei.2022.100061

Sini, R. (2017). #MeToo, #TakeAKnee and #Covfefe: Hashtags that Dominated in 2017. BBC News. http://www.webcitation.org/76UCvB69Z

Skoric, M. M., Liu, J., & Jaidka, K. (2020). Electoral and public opinion forecasts with social media data: A meta-analysis. Information, 11(4), 187. https://doi.org/10.3390/info11040187

Sloan, L., Morgan, J., Burnap, P., & Williams, M. (2015). Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. PloS one, 10(3), e0115545. https://doi.org/10.1371/journal.pone.0115545

Soares, C. D. M., & Joia, L. A. (2015). The Influence of Social Media on Social Movements: An Exploratory Conceptual Model. Lecture Notes in Computer Science, 9249. Cham: Springer.

Sobhani, P., Mohammad, S., & Kiritchenko, S. (2016). Detecting stance in tweets and analyzing its interaction with sentiment. In Proceedings of the fifth joint conference on lexical and computational semantics, pp. 159-169. https://aclanthology.org/S16-2021.pdf

Sönning, L., & Werner, V. (2021). The replication crisis, scientific revolutions, and linguistics. Linguistics, 59(5), 1179-1206. https://doi.org/10.1515/ling-2019-0045

Sterling, R., & LeRouge, C. (2019). On-demand telemedicine as a disruptive health technology: qualitative study exploring emerging business models and strategies among early adopter organizations in the United States. Journal of medical Internet research, 21(11), e14304. https://doi.org/10.2196/14304

Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018). Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter. Political communication, 35(1), 50-74. https://doi.org/10.1080/10584609.2017.1334728

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2019). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. Social Science Computer Review, 0894439319843669. https://doi.org/10.1177/0894439319843669

Stone, P. J., Dunphy, D. C., & Smith, M.S. (1966). The General Inquirer: A Computer Approach to Content Analysis. Cambridge: MIT Press.

Street, J. (2005). Politics Lost, Politics Transformed, Politics Colonised? Theories of the Impact of Mass Media. Political Studies Review, 3(1), 17-33.

Sunstein, C. (2017). Republic.com 2.0. Princeton, NJ: Princeton University Press.

Swiss Federal Council. (2017). Un cadre juridique pour les médias sociaux: nouvel état des lieux. Bern. https://www.bakom.admin.ch/dam/bakom/fr/dokumente/informationsgesellschaft/social_media/social%20media%20bericht.pdf.download.pdf/rapport-media-sociaux-2017-FR.pdf

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267-307. https://doi.org/10.1162/COLI_a_00049

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology, 29(1), 24-54.

Tavoschi, L., Quattrone, F., D'Andrea, E., Ducange, P., Vabanesi, M., Marcelloni, F., & Lopalco, P. L. (2020). Twitter as a sentinel tool to monitor public opinion on vaccination: An opinion mining analysis from September 2016 to August 2017 in

Italy. Human Vaccines & Immunotherapeutics, 16(5), 1062-1069. https://doi.org/10.1080/21645515.2020.1714311

Thompson, L., Rivara, F. P., & Whitehill, J. M. (2015). Prevalence of Marijuana-Related Traffic on Twitter, 2012-2013: A Content Analysis. Cyberpsychology, Behavior, and Social Networking, 18(6), 311-319. https://doi.org/10.1089/cyber.2014.0620

Törnberg A and Wahlström M (2018) Unveiling the radical right online: Exploring framing and identity in an online anti-immigrant discussion group. Sociologisk Forskning, 55(2-3): 267-292. https://sociologiskforskning.se/sf/article/view/18193

Tornes, A., & Trujillo, L. (2021). "Enabling the future of academic research with the Twitter API". Available at: https://blog.twitter.com/developer/en_us/topics/tools/2021/enabling-the-future-of-academicresearch-with-the-twitter-api.html

Tourangeau, R., & Galešić, M. (2008). Conceptions of attitudes and opinions. The SAGE Handbook of Public Opinion Research, 141-154.

Tran, H. L. (2015). More or less? Multimedia Effects on Perceptions of News Websites. Electronic News, 9(1), 51-67.

Trepte, S., & Reinecke, L. (Eds.). (2011). Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web. Berlin Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-21521-6

Tsai, J. M., Cheng, M. J., Tsai, H. H., Hung, S. W., & Chen, Y. L. (2019). Acceptance and resistance of telehealth: The perspective of dual-factor concepts in technology adoption. International Journal of Information Management, 49, 34-44. 10.1016/j.ijinfomgt.2019.03.003.

Tucker, J. A., A. Guess, P. Barberá, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, & Nyhan, B. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. SSRN Electronic Journal, 1-95. https://doi.org/10.2139/ssrn.3144139

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., ... & Nyhan, B. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. Technical report, The Hewlett Foundation.

https://www.hewlett.org/wp-content/uploads/2018/03/ Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf

Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. ArXiv:1403.7400 [Physics]. http://arxiv.org/abs/1403.7400

Turow, J., Hennessy, M., & Draper, N. (2015). The Tradeoff Fallacy: How Marketers are Misrepresenting American Consumers and Opening Them Up to Exploitation (SSRN Scholarly Paper No. ID 2820060). Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2820060

Vaccari, C., & Augusto V. (2021). Outside the Bubble: Social Media and Political Participation in Western Democracies. New York: Oxford Academic.

Vaccari, C., & Valeriani, A. (2015). Follow the leader! Direct and indirect flows of political communication during the 2013 Italian general election campaign. New media & society, 17(7), 1025-1042. https://doi.org/10.1177/146144481351103

Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., & Tucker, J. (2013). Social media and political communication. A survey of Twitter users during the 2013 Italian general election. Rivista italiana di scienza politica, 43(3), 381-410.

Vaccari, C., Valeriani, A., Barberá, P., Jost, J. T., Nagler, J., & Tucker, J. A. (2016). Of Echo Chambers and Contrarian Clubs: Exposure to Political Disagreement Among German and Italian Users of Twitter. Social Media + Society, 2(3), 2056305116664221. https://doi.org/10.1177/2056305116664221

van Aelst, P., J. Strömbäck, T. Aalberg, F. Esser, C. De Vreese, J. Matthes, D. Hopmann, S. Salgado, N. Hubé, A. Stępińska, S. Papathanassopoulos, R. Berganza, G. Legnante, C. Reinemann, T. Sheafer, & Stanyer, J. (2017). Political Communication in a High-Choice Media Environment: A Challenge for Democracy? Annals of the International Communication Association, 41 (1), 3-27.

van Atteveldt, W., van der Velden, M. A. C. G., & Boukes, M. (2021). The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. Communication Methods and Measures, 15(2), 121-140. https://doi.org/10.1080/19312458.2020.1869198

van Atteveldt, W., Welbers, K., & van der Velden, M. (2019). Studying political decision making with automatic text analysis. Oxford Research Encyclopedia of Politics, 1-11. https://doi.org/10.1093/acrefore/9780190228637.013.957

Van Dijck, J., & Poell, T. (2013). Understanding social media logic. Media and communication, 1(1), 2-14. https://doi.org/10.17645/mac.v1i1.70

Van Gorp, B. (2005). Where is the frame? Victims and intruders in the Belgian press coverage of the asylum issue. European journal of communication, 20(4): 484-507.

van Vliet, Livia. (2020). The Twitter Parliamentarian Database. figshare. Dataset. https://doi.org/10.6084/m9.figshare.10120685.v2

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. ArXiv:1703.03107 [Cs]. http://arxiv.org/abs/1703.03107

Venturini, T., & Rogers, R. (2019). API-based research or how can digital sociology and journalism studies learn from the Facebook and Cambridge Analytica data breach. Digital Journalism, 7(4), 532-540. https://doi.org/10.1080/21670811.2019.1591927

Verma, B., & Thakur, R. S. (2018). Sentiment analysis using lexicon and machine learning-based approaches: A survey. Proceedings of international conference on recent advancement on computer and communication, pp. 441-447. Springer, Singapore.

Vicente, I. S., & Saralegi, X. (2016). Polarity Lexicon Building: To What Extent Is the Manual Effort Worth?. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 938-942.

Vliegenthart, R., and H. G. Boomgaarden. (2007). Real-world Indicators and the Coverage of Immigration and the Integration of Minorities in Dutch Newspapers. European Journal of Communication, 22(3): 293-314. https://doi.org/10.1177/0267323107079676

Wallaschek, S., K. Kaushik, M. Verbalyte, A. Sojka, G. Sorci, H. J. Trenz, & Eigmüller, M. (2022). Same Same But Different? Gender Politics and (Trans-)National Value Contestation in Europe on Twitter. Politics and Governance, 10 (1), 146-60.

Weber, S. (2018). The Great American Search for Healthcare

Weeks, B. E., A. Ardèvol-Abreu, & Gil de Zúñiga, H. (2017). Online Influence? Social Media Use, Opinion Leadership, and Political Persuasion. International Journal of Public Opinion Research 29 (2): 214-39.

WHO (World Health Organization). (2015). Global survey on health technology assessment by national authorities. Main findings. Geneva: WHO. https://www.who.int/publications/i/item/9789241509749

Wijffels, J., M. Straka, & Straková, J. (2018). Package 'udpipe'. https://cran.microsoft.com/snapshot/2018-06-12/web/packages/udpipe/udpipe.pdf

Willemin, N. (2018). Médias Suisses, Le Virage Numérique. Lausanne: Livreo-Alphil.

Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation. Sociology, 51(6), 1149-1168.

Williams, M. L., Burnap, P., Sloan, L., Jessop, C., & Lepps, H. (2017). Users' views of ethics in social media research: Informed consent, anonymity, and harm. In Woodfield, K. (ed.), The ethics of online research, pp.27-52. Emerald Publishing Limited.

Wlezien, C. (1995). The public as thermostat: Dynamics of preferences for spending. American Journal of Political Science, 39, 981-1000. https://doi.org/10.3886/icpsr01104

Wojcieszak, M. (2010). 'Don't Talk to Me': Effects of Ideologically Homogeneous Online Groups and Politically Dissimilar Offline Ties on Extremism. New Media & Society, 12 (4), 637-55.

Woolley, S. C. (2016). Automating Power: Social Bot Interference in Global Politics. First Monday, 21(4).

Wüest, B., Amsler, M., & Schneider, G. (2017). SIFT-A Language Technology Toolkit to Assess the Print Media Coverage of New Forms of Governance. Working paper series/NCCR-Democracy, (95).

Xiang, C., Zhang, J., Li, F., Fei, H., & Ji, D. (2022). A semantic and syntactic enhanced neural model for financial sentiment analysis. Information Processing & Management, 59(4), 102943. https://doi.org/10.1016/j.ipm.2022.102943

Xiong, Y., M. Cho, & Boatwright, B. (2019). Hashtag Activism and Message Frames Among Social Movement Organizations: Semantic Network Analysis and Thematic Analysis of Twitter During the #metoo Movement. Public Relations Review, 45 (1), 10-23.

Xu, R., Zhou, Y., Wu, D., Gui, L., Du, J., & Xue, Y. (2016). Overview of NLPCC shared task 4: Stance detection in Chinese microblogs. In Proceedings of International

Conference on Computer Processing of Oriental Languages and National CCF Conference on Natural Language Processing and Chinese Computing, pp. 907-916.

Yantseva, V. (2020). Migration Discourse in Sweden: Frames and Sentiments in Mainstream and Social Media. Social Media + Society. https://doi.org/10.1177/2056305120981059

Young, L., & Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. Political Communication, 29 (2), 205-231. https://doi.org/10.1080/10584609.2012.671234

Young, L., & Soroka, S. (2012). Lexicoder sentiment dictionary. McGill University, Montreal, Canada.

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Algorithmic decision-making and the control problem. Minds and Machines, 29(4), 555-578. https://doi.org/10.1007/s11023-019-09513-7

Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-Level Convolutional Networks for Text Classification. In Advances in Neural Information Processing Systems, 649-657.

Zhang, Y., Cao, B., Wang, Y., Peng, T. Q., & Wang, X. (2020). When Public Health Research Meets Social Media: Knowledge Mapping From 2000 to 2018. Journal of medical Internet research, 22(8), e17582. https://doi.org/10.2196/17582

Zhao, J., Liu, K., & Xu, L. (2016). Sentiment analysis: mining opinions, sentiments, and emotions. Computational Linguistics, 42(3), 595–598. https://doi.org/10.1162/COLI_r_00259

Zhu, L., He, Y., & Zhou, D. (2020). Neural opinion dynamics model for the prediction of user-level stance dynamics. Information Processing & Management, 57(2), 102031. https://doi.org/10.1016/j.ipm.2019.03.010

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political effects of the internet and social media. Annual Review of Economics, 12: 415-438.

Zubiaga, A., Kochkina, E., Liakata, M., Procter, R., Lukasik, M., Bontcheva, K., Cohn, T., & Augenstein, I. (2018). Discourse-aware rumour stance classification in social media using sequential classifiers. Information Processing & Management, 54(2), 273-290. https://doi.org/10.1016/j.ipm.2017.11.009

Zuiderveen Borgesius, F. J., D. Trilling, J. Möller, B. Bodó, C. H. De Vreese, & Helberger, N. (2016). Should We Worry About Filter Bubbles? Internet Policy Review, 5 (1), 1-16. https://doi.org/10.14763/2016.1.401