

A LONGITUDINAL APPROACH TO ONLINE “COLLECTIVE IDENTITY WORK”: THE CASE OF THE GILETS JAUNES IN THE VAR DEPARTMENT*

Davide Morselli, Maite Beramendi, Zakaria Bendali, and Olivier Fillieule†

Social movement theorists have highlighted the importance of accounting for the fluidity of collective identities and the ways in which they change over time. Capitalizing on the availability of social media data and the shift from collective to connective action, new methods can be used to model identity change over a medium-to-long time span. We analyze Facebook data to make the case that a complex and longitudinal approach to the study of collective identity is not only possible but also necessary. Our analyses explore the formation and transformation of identity in the context of a local branch of the Gilets Jaunes (Yellow Vests) movement in France. By employing a mixed-method design that combines automated topic modeling, content analysis, and dictionary-based linguistic inquiry, we show how collective identities are discussed through complex and conflictual processes that form perpetual identity work.

The role of identity is central in the field of social movement research (see Polletta and Jasper 2011; Reger, Myers, and Einwohner 2008; Snow 2001). Alberto Melucci (1989, 1996) was among the first in the social movement community to establish the construction of collective identity as a key feature of engagement in grassroots action, thus making possible an abundance of research pertaining to this “pivotal concept” (Snow and McAdam 2000). To date, the lion’s share of this research has been dedicated to examining the ways in which activists’ sense of identity shapes a variety of movement features, including emergence (Melucci 1989; Pfaff 1996), factors related to individual commitment (Calhoun 1994; Gamson 1992; Neuhauser 1998; Polletta 1998), conditions for sustained participation over time (Hirsch 1990; Nepstad 2004; Taylor 1989), strategic choices and framing strategies (Bernstein 1997; Hunt, Benford and Snow 1994), organizational forms (Clemens 1996) and outcomes (Einwohner 1999). Another fascinating stream of research has studied the factors contributing to the construction of activists’ identity, including organizational characteristics of activist groups and events (Gamson 1996; Reger 2002), the role of repression by the police (Drury and Reicher 2000; Drury, Reicher, and Stott 2003), labeling and stigmatization by opponents, the media and the public (Einwohner 2002; Haines 2006), and the changing contexts of action that directly impact the pace, scope, and nature of generational replacement within movements (Gusfield 1963; McAdam 1988; Roth 2008; Whittier 1995; Zwerman, Steinhoff, and della Porta 2000).

There is broad consensus in the literature that collective identity in social movements should not be regarded as a feature with clear-cut boundaries but rather as a process that unfolds through both incoming and outgoing communications and features a constant state of change. Identity processes involve activists’ development of a sense of themselves as a community that is united in their beliefs, claims, and tactical repertoires and stands in opposition to some other groups. Despite this general agreement concerning the fluid and evolving character of collective identities, the number of studies that directly address the transformation and evolution of

* This publication benefited from the support of the Swiss National Science Foundation (grant number: 10001A_192686) and the Swiss Centre of Expertise in Life Course Research LIVES (NCCR LIVES, grant number: 51NF40-185901). The authors are grateful to the Swiss National Science Foundation for its financial assistance.

† Davide Morselli is Assistant Professor in Life Course Social Psychology at the University of Lausanne, Maite Beramendi is Senior Researcher in Social and Political Psychology at the University of Lausanne and the University of Buenos Aires, Zakaria Bendali is Junior Researcher in Political Sociology at the University of Lausanne, Olivier Filleule is Professor of Political Sociology at the University of Lausanne. Contact: davide.morselli@unil.ch

collective identity in social movements remains relatively limited (see Whittier 1995; Reger 2002; Fillieule and Blanchard 2013 for examples).

In this article, we capitalize on the availability of social media data and the shift from collective to connective action (Bennett and Segerberg 2012) to have a new look at collective identity processes. As Paolo Gerbaudo and Emiliano Treré (2015) argued, most previous research has focused on the networked aspects of online activism instead of on the construction and unfolding of collective identities. In this article, we contribute to the task of addressing this shortcoming by proposing a mixed-method framework to extract insights from social media data concerning the evolution and transformation of collective identities. In particular, we use computational methods and grounded theory to investigate the changing elements of collective identities.

In the following sections, we first address the issue of collective identity and its embedding in social media. Then, we present our case study: a local branch of the *Gilets Jaunes* (Yellow Vests) movement in France. This relatively small-scale case was characterized by a connection between offline and online activism in which street protesters discussed and connected through Facebook groups, making it possible to observe the ways in which discussions unfolded.¹ Finally, we present our methodology and analysis of these Facebook group discussions.

COLLECTIVE IDENTITY AS A TRANSFORMATIVE PROCESS

Melucci defines collective identity as “an interactive and shared definition produced by a number of individuals (or groups at a more complex level) concerning the orientations of their actions” (1996, 70). According to this definition, collective identity is defined in terms of self-conceptions and is expressed through action and interaction (Flesher Fominaya 2019). Also, as Melucci (1995) notes, such a sense of unity is not a starting point for mobilization but rather a result thereof. As activists become embedded in networks and politicized, they renegotiate their notions of self, leading to changes in their collective identities. Finally, Melucci stresses the nonhomogeneous properties of collective identity: what appears at a given moment to be a collective identity is rather a product of the balance among competing and occasionally conflicting elements (see also Flesher Fominaya 2010; Kavada 2009).

For Melucci, collective identity is, therefore, not limited to a shared definition of the group. Collective identity refers to the process by which the collective distinguishes itself from its environment and assumes agency. Hence, this process, which he calls “identization,” is a cognitive framework related to the scope of the associated action, established through interactive exchanges among members, common practices, shared emotions, and a sense of solidarity. Beyond the superposition between being and acting, Taylor and Whittier (1992) stress that collective identities unfold through a process of definition of boundaries and the negotiation of meanings within and outside the movement. The task of defining the identities that are relevant to social movements is not linear. Such identities are instead fraught with contradictions and controversies (Reger, Myers, and Einwohner 2008).

Therefore, the construction and maintenance of identity require a great deal of “identity work,” a concept proposed by David Snow and Doug McAdam (2000) to describe more clearly the ways in which movement participants actively work to create their sense of who and what they are. Central to this concept are the notions of sameness and difference as well as an us-versus-them dialectic. While the us-versus-them separation has been shown to be a basic process of intergroup relations (Tajfel and Turner 2004) and plays a central role in mobilization (Klandermans 2014; Pozzi et al. 2022; Thomas, Mavor, and McGarty 2012; Van Zomeren, Postmes, and Spears 2008), the notion of identity work may extend beyond this mechanism. According to this conception, collective identities are emergent properties of collective action and are created by way of a process of “construction and maintenance through joint action, negotiation, and interpretive work” (Snow and McAdam 2000, 46). This interpretive aspect of identity construction cannot be encapsulated as simply defining allies and enemies. Rather, it

extends to how different aspects of movement life become concurrent in defining who and what the movement is.

Focusing on this constructivist approach (Snow and McAdam 2000), Anastasia Kavada (2014, 2015) argued that digital media not only represent new forms of political participation (Bennett and Segerberg 2012; George and Leidner 2019) or an easily accessible organizational tool (della Porta and Mosca 2005; Tüfekçi 2017) but are also a key factor in the construction of collective identities (Coretti and Pica 2018). While some research has provided empirical evidence that both offline activists and online activists communicate via internet platforms, thus activating symbolic resources to construct collective identities (Hara and Estrada 2005; Kavada, 2009, Nip 2004), other researchers have argued that online discussions have only had a marginal influence on the definition and maintenance of collective identities (Ayers 2003; Pickerill 2003; Wall 2007). It must be noted, however, that most of this research was conducted prior to the emergence of Web 2.0, that is, the widespread incorporation of social media into every aspect of life, including activism. Since its advent, social media has become part of the current practice of organizing protests, simplifying some of the tasks that made organizing traditional social movements too costly (Chen 2019; Tüfekçi 2017). However, the repercussions of social media with respect to other vital elements of movements, such as the construction of collective identities, remain an open question.

The spread of social media in the context of mobilization has given researchers access to a complex variety of textual and discursive data. Paradoxically, a major stream in social movement research has exhibited a decreasing emphasis on the collective identity framework in favor of a focus on the structural and organizational aspects of social media in the context of protests (Bennett and Segerberg 2012; Mattoni and Pavan 2018; Schumann and Klein 2015; Shirky 2011) or the links between offline and online protests (Freelon, McIlwain, and Clark 2018; Jost et al. 2018; Valenzuela 2013).

Gerbaudo and Treré (2015), introducing a special issue of *Information, Communication & Society* on social media and protest, argued that this paradox derives from the fact that big data are primarily analyzed via quantitative methods, while the processes operative in identification can only be understood via qualitative analysis. As a result, their special issue presented a collection of seminal works that applied various types of qualitative (as well as quantitative) analyses to investigate the formation and boundaries of collective identities in the context of various movements (Coretti and Pica 2015; 2018; Gerbaudo 2015; Kavada 2015; McDonald 2015; Milan 2015; Monterde et al. 2015).

Although this research provides a snapshot of the complexity involved in forming collective identities and their fluidity, it does not yet allow us to fully grasp their changes over time and, accordingly, to understand the idea of identity as a process. In other words, if the study of collective identity is to produce a sufficiently complex picture of the identity process (Treré 2015), it must also be analyzed via a longitudinal framework in order to investigate the changes that are encapsulated in the concept of process properly. In line with Melucci's (1995) idea that identity is rather a consequence than a starting point for mobilization, we can expect that different elements are prioritized at different moments. In particular, we can expect the identity work to become more central after repeated rounds of mobilization when some organizational routines have been established. Therefore, we can expect to observe a decrease in the centrality of organizational matters over time as the activists start building and negotiating common meanings. Thus, we expect collective identity features to become more relevant in discussions among the movement members, with activists defining allies and opponents and managing in-group differences. In particular, we theorize three different and complementary aspects of the identity work. First, we expect that a sense of "us" is initially prioritized. Second and complementary to "us" is the emergent definition of "them." We expect the opposition "us" to "them" to become increasingly more relevant across time, especially depending on the responses of the authorities to the mobilization (Combes and Fillieule 2012; Morselli, Passini, and McGarty 2021). Finally, and in line with the idea that collective identities are constituted by conflictual and contradictory positions (Flesher Fominaya 2010; Snow and McAdam 2000), we expect

“you” elements defining internal divisions and oppositions to also become increasingly relevant (Ghaziani 2008). We argue that these three aspects do not occur in a vacuum but are interrelated to contextual factors, such as political opportunities, repression, and/or lassitude of the authorities, that create a fertile ground for shaping the identity work. In line with Kavada (2009), we also argue that online activities are fundamental for these processes and that social media data offer great potential to facilitate their investigation, linking changes to other events, both external and internal to the movement.

THE VAR GILETS JAUNES

At the end of 2018, the French government, led by President Emmanuel Macron and his party (La République En Marche, LREM), presented a budget for 2019 that included an increase in fuel taxes. Widespread discontent began to emerge, to the point that a call to protest was posted on Facebook on October 10. On November 17, approximately 290,000 people responded to this call, thus giving birth to the largest and most enduring social movement in France in the past fifty years and widening the scope of the protestors’ grievances. Protests, including the occupation of roundabouts and motorway toll stations, were organized weekly across the country throughout 2019 (Collectif d’enquête sur les Gilets Jaunes 2019). By the end of the year, however, the movement had dwindled considerably, and dissipation was likely in the following months.

In this study, we focus on activity occurring in the Var department in southeastern France. Participation in Var was strong compared to other French departments (Boyer et al. 2020), with emblematic roundabout actions, such as that at Cannet-des-Maures, where an arch of triumph, an Eiffel Tower and a Louvre pyramid were erected, attracting media attention. In addition, the extreme right in the area had been achieving high electoral scores for a long time (several municipalities are held by National Rally (RN) elected officials). This department also has unique economic characteristics and social fabric. Var is an area featuring approximately 1 million inhabitants and a density of 180 people per square kilometer. The population is concentrated along the coastline and the main roads that cross the department from east to west. Occupations of roundabouts occurred along these routes, affecting people living in both urban and rural areas. The poverty rate in this department is higher than the national average (15.5% compared to 14%), particularly for individuals under thirty years of age (24%). According to the socioeconomic picture of Var collected by the administration in November 2018, the economy of the department is characterized by the importance of companies with fewer than ten employees (96% compared to 94% in metropolitan France) and by the dominance of the construction sector (14% of the total compared to 10% in metropolitan France). However, the economy of Var is primarily oriented toward producing goods and services for people residing in the area, whether as residents or tourists.

The sociology of the local Gilet Jaunes reflects this situation, including various predominant social positions: blue- and white-collar workers, workshop technicians, small artisans and traders, small farmers, unemployed persons, individuals in precarious situations, retired people with low incomes, and single-parent families. However, what is also striking is the number of people employed in services dedicated to persons, especially older individuals. In this sector, 87% of employees are women. Thus, whether in retirement homes or the home hospitalization sector, the aging population accelerates the formation of a proletariat employed in the so-called “silver economy,” which serves as the core of the Gilets Jaunes movement in Var. A factor that seems to have increased the solidarity among activists is a shared feeling that their living conditions have deteriorated, thus subjecting them to a situation of vulnerability that they consider unjust. Generally, the Var Gilets Jaunes were not previously involved in any political, union, or collective activity, including voting; therefore, their activism and political convictions were forged in the crucible of the movement.

In Var, as elsewhere in France, Facebook played a key role in organizing the Gilets Jaunes movement. By the end of 2018, more than 1,500 Facebook groups and pages were created

nationwide. Pierre Boyer and colleagues (2020) reported data drawn from Google Trends showing that Gilets Jaunes content was demanded more frequently on Facebook than other social media. These data also show that the number of unique subscribers to the Gilets Jaunes Facebook groups mirrored the official counts of street protesters reported by the French Ministry of the Interior and followed the same patterns of fluctuation.

The link between online and offline activism in this movement was particularly close (Boyer et al. 2020). Aside from organizing large-scale street protests in Paris, the Gilets Jaunes movement took the form of small units, which organized protests on a local scale and occupied roundabouts, where encampments were erected. These encampments made it possible to establish “protected” or “free” spaces that helped lift the barriers of fear, mobilize and maintain energy, and socialize political novices who had no experience of protest with political action (Fillieule and Tartakowsky 2013). The small scale of this movement might be of particular interest in the context of investigating the formation of collective identities. People know each other at the local level, and activism therefore contributes to a sense of community embedded in a larger context (the national movement). Unlike other studies concerning the Gilets Jaunes (see Bendali and Rubert 2020 for a review), we focus on local groups in Var and monitor their discussion on the relevant Facebook groups during a fourteen-month period covering the first wave of the movement.

DATA AND METHODS

We collected data from six active Facebook groups from October 2018 to December 2019 via the Facebook Group application programming interface (API) and CrowdTangle. The average number of active users per group was 607 (minimum = 140, maximum = 1495). We collected 56,954 posts, comments, and responses, of which 34,109 contained text and could be analyzed for the purpose of this study. Because we were interested in the discourse coconstructed by interactions among users, we grouped all comments and responses in accordance with the post to which they referred, thereby reconstructing discourse threads ($n = 6,339$). In this study, we adopted a mixed-method framework relying on computational social science techniques and in-depth qualitative analysis to investigate the content of these Facebook conversations. As Lewis, Zamith, and Hermida (2013) argue, a hybrid approach allows for an understanding of large-scale data while leveraging the strength of qualitative analysis. To this end, our analytical approach was divided into two steps, each mixing qualitative and quantitative methods. In the first step, we adopted an inductive approach to explore and understand the foci of these online conversations. Second, we used a theory-driven approach to investigate the prevalence and evolution of collective identity features in the data.

Mixed-Method Content Analysis

To investigate the text’s underlying dimensionality and highlight the content of the discussions among the targeted Facebook groups, we first adopt an inductive data-mining approach by applying topic modeling and content analysis together. In particular, we used the structural topic-modeling (stm) framework with spectral initialization as implemented in the stm package for the R language (Roberts, Stewart, and Tingley 2019). Topic modeling is a technique that models the co-occurrence of features in language and assumes that text can discuss several topics simultaneously. This approach estimates the probability that each topic is present in a document of the corpus. Therefore, the goal of topic modeling is to determine the number of topics that most suitably describe (fit) a corpus while describing the likelihood of these topics occurring in each document of the corpus. Similar to exploratory factor analysis, a series of indicators can be used to identify the optimal number of such topics and their interpretability. In this study, we relied on the procedure highlighted by Roberts, Stewart, and Tingley (2019), which compares the performance of several models based on the triangulation of semantic

coherence (Mimno et al. 2011), the held-out likelihood (Wallach et al. 2009) and residual analysis (Taddy 2012). To reduce linguistic complexity and improve model convergence, we performed lemmatization and part-of-speech tagging using the UDPipe framework as implemented in the R language (Wijffels 2020). The final corpus included only adjectives, adverbs, nouns, and verbs. Stop words, punctuation, numbers, and modal verbs were also removed. To overcome the known limitations of topic modeling in the context of short texts (Sbalchiero and Eder 2020), we group the threads according to activist group and week, following the weekly rhythm of the movement (each weekend was marked by an “act,” in the language of the movement, that complemented the permanent occupation of roundabouts with street demonstrations in urban centers) and maintaining the specificity of each group. The final model was then applied to the full ungrouped corpus to obtain the probabilities of each topic by post.

The results of the topic model are presented to the researcher as a series of groups of co-occurring words that she/he must interpret in a meaningful way. This task is not always straightforward, as words can have different meanings based on context. The usual practice is to name the topics based on the researcher’s interpretation of the semantic associations among words. Although some research has compared manual coding to automatic classification via topic modeling and shown an overall convergence between these two methods (Rodriguez and Storer 2020), the use of automated methods in isolation could cause important aspects of the interaction and co-construction of meaning that occurs through online conversations to be overlooked (Blok and Pedersen 2014; Lewis, Zamith, and Hermida 2013; Rodriguez and Storer 2020). To overcome this limitation, we integrated topic modeling with grounded-theory content analysis (Corbin 2017) to validate and understand the content of these topics in depth. Once the algorithm extracted the topics, we selected 90 of the longest discussion threads (those with the highest number of comments) and those with the highest topic probability. Before the analysis by category began, all threads were read twice. Then, we applied open coding, which involves decomposing the data into discrete parts to be examined in detail and compared in the search for similarities and differences. Concepts considered related or similar in nature were grouped into categories (Strauss and Corbin 2014). Finally, axial coding was carried out by relating categories to subcategories along the lines of their properties and dimensions and observing how they intersected and linked with each other (Strauss and Corbin 2014). Following Strauss and Corbin (2014), the coding process was not linear; rather, the different types of coding were superimposed and intertwined. The procedure was repeated until saturation, meaning changes were no longer being made to the definitions of the categories, and no additional text could be assigned to them.

Dictionary-Based Analysis of Collective Identity

We used a dictionary-based text inquiry method to complement this content analysis and investigate the prevalence of collective identity features in the discussions. While topic modeling is an inductive, data-driven method, dictionary-based approaches are theory-driven and allow for the investigation of preconstructed categories. These methods search the text for a list of words the researcher assigns to one or more categories (a dictionary) and assess their prevalence and relevance. While a number of precompiled dictionaries exist, it has been argued that such dictionaries should be context-specific and adapted to the corpus under investigation (Reveilhac and Morselli 2022). In this study, we constructed a dictionary using an iterative qualitative–quantitative approach.

First, two independent judges qualitatively analyzed a random selection of discussion threads to compile a list of words and expressions ($n = 39$) related to collective identity discourse. Terms about self-conceptions, common action, and sense of unity (Flesher Fominaya 2019; Melucci 1995, Taylor and Whittier 1992) in the context of the Var Gilets Jaunes were selected. The list included expressions of the definition of the self as a collective (e.g., “the Gilet Jaunes are,” “us together,” “family,” “to be a Gilet Jaune”) and words referring to ideological features that could be either part of or alien to the movement (e.g., “far left/right,”

“fascist,” “pacifist”) and were needed in the interaction with external actors, such as other movements, political counterparts, and media (Taylor and Whittier 1992). Hence, this list was used as a set of search keywords that were included alongside “gilets jaunes” to retrieve 265,688 tweets from the Twitter historical archive for 2018-2019 via the Academic Research API. We trained a word-embedding model using word2vec in Gensim for Python (Rehurek and Sojka 2011) and used this model to find words, bigrams, and trigrams with a similar vector representation to our hand-cured list and thus expand the dictionary.²

Finally, the resulting expanded dictionary was checked manually by reference to a further random selection of posts from the corpus. This iterative procedure allowed us to expand the initial dictionary to include features not considered by the researchers and correct it by removing ambiguous terms. The final dictionary included fifty-one words and expressions (see appendix A available at <https://osf.io/k4dmz>). Therefore, this dictionary was applied to the text using the Linguistic Inquiry and Word Count method (Tausczik and Pennebaker 2010), as implemented in the *quanteda* package for the R language (Benoit et al. 2018). Examples of utterances highlighted by this procedure were “[Joanne,] do not play their game of dividing and conquering. We have to stick to our claims and be united in solidarity and freedom”; “A militant of the right-wing party is not necessarily a fascist”; and “It’s inappropriate to confound the Black Blocks and the Gilets Jaunes.”³

To further investigate the construction of identity in terms of “us” and “them” from a linguistic point of view, we applied the LIWC procedure to score references to plural first person (“us”) and singular and plural third person (“him/her/them”) pronouns. Plural second-person pronouns were also used to investigate eventual group division within the discussions. Examples of utterances with the highest “us” scores were “France, wake up! We are the country of human rights, let’s rediscover our values, our strength, our honor”; “It’s our fault as consumers of these services if we don’t intervene. We can’t wait anymore for a change”; and “They are cheating us: They want to make our history, our roots, and us disappear at the same time!” Examples of “them” were “He [Macron] doesn’t care about the French people. Shame on him. France doesn’t need guys like him” and “They [the police] are blocking the way to the harbor. Let them sink with their president, poor cowards”. Finally, examples of top “you” utterances were “You are only here to criticize the Gilets Jaunes! It’s easy for you when you get up in the morning and don’t need to earn your day” and “Madame, keep your opinion to yourself. You only listen to the media, it is better if you properly inform yourself first”.

Longitudinal Analysis

To investigate the development of themes over time, the evolution of collective identity features, and their relationships to contextual events, we transformed the content and collective identity analysis into a time series that indicated the distribution of each shared topic across time as well as the collective identity reference within each topic. All these analyses were conducted using French text, with all authors being either native speakers or fluent in French. To make this text more accessible to the international research community, it has been translated into English throughout this article.

Results Extracting Topics from the Discussion Threads

Following the procedure developed by Roberts, Stewart, and Tingley (2019), we tested models that included two to thirty topics. The best solution in terms of triangulation of statistical fit indexes and interpretability of the result was provided by a model including eight topics. Figure 1 shows the semantic coherence (the probability that words are used close to each other in the text) and exclusivity (the probability that words are not shared with other topics) indexes for these eight topics, confirming that the topics have either internal consistency or are different from other themes: four topics exhibit high coherence and below-average exclusivity, three feature high exclusivity but average coherence, and one demonstrates high exclusivity but low coherence. The most relevant words for each topic are displayed in figure 2.

Figure 1. Semantic Coherence and Exclusivity of the Eight-Topics Model

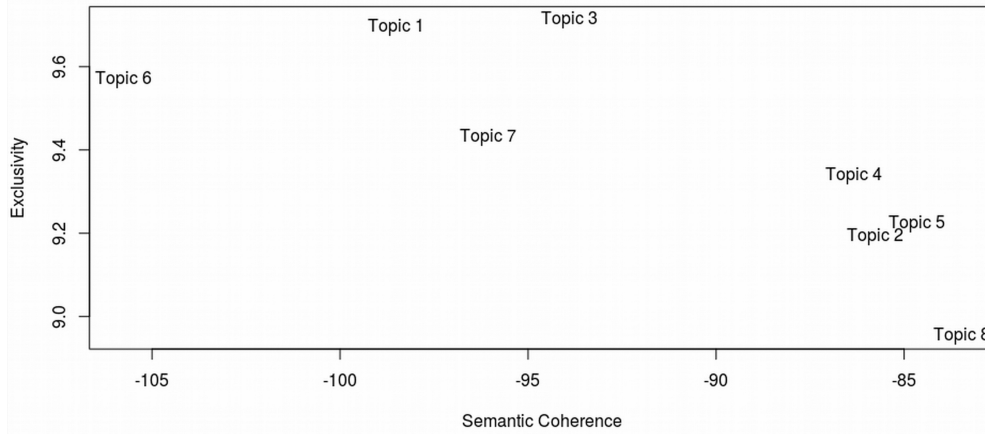
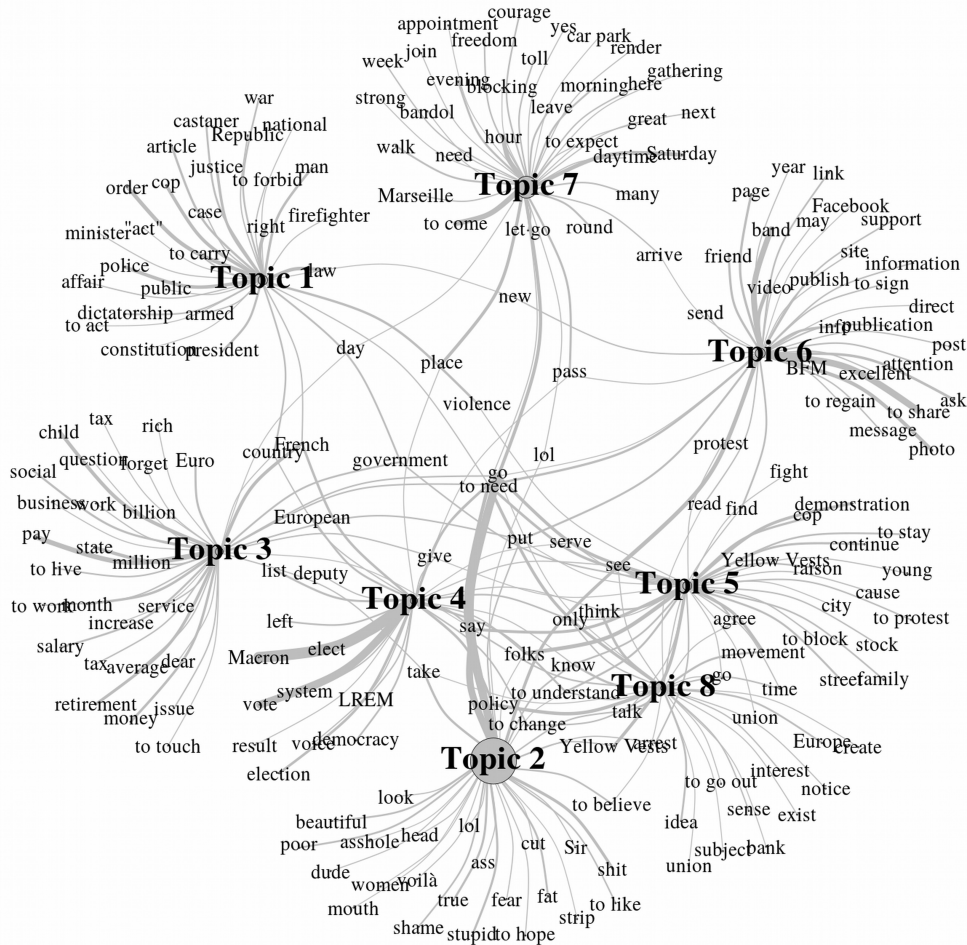


Figure 2. Graphical Representation of the Topics and Their Most Relevant Words



Note: The wider the edge, the higher the frequency in the topic; the larger the vertex, the higher the topic distribution in the corpus.

Based on these results, we conducted content analysis concerning selected threads for each topic. A summary of the results of this analysis is displayed in table 1.⁴ Consistent with what could be inferred from the topic model results (figure 2), the discussions extracted from topic 1 concerned political abuse from the government, politicians, and the police. In relation to the abuse by the government, the Gilets Jaunes commented on situations featuring an abuse of rights, such as the prohibition of assembly and protest. A more general discussion of the legitimacy of the government's actions, such as unfair international commercial trade agreements, was also conducted. Abuse of power by politicians also emerged as a theme in the analysis, in which context cases of politicians spending public money on expensive and unjustified dinners were discussed. Finally, this topic included legal abuses by law enforcement, in which context Gilets Jaunes experienced direct violence. Violent acts of repression committed by the police during the movement's demonstrations were noted and discussed, as was the government's com-

Table 1. Content of the Discussion Threads

<i>Topic Label</i>	<i>Content of the Threads</i>
1 Political and Police abuses	Police discrediting the Gilets Jaunes demonstrations; Discussions about police violence in demonstrations; Government covering up police violence; Disadvantageous international trade agreements for France; Government forbidding demonstrations; Abuse of representatives' spending.
2 Anger and outrage	Anger against police; Anger over the bad organization of the Grand Débat; Anger over the Macron's comments on the Gilets Jaunes; Anger over the lack small shop support; Mocking politicians; Discussion about the people's choices in the elections; Anger over a fine given to a Gilet Jaune; Anger over domestic violence and the laws around it; Negative expressions toward Macron and Castaner; Negative expressions towards BFMTv; Anger for the presidential spending; Anger over tax evasion by politicians.
3 Social claims	Against the minimum wage increase; Discussion about oil taxes; Reform of pensions and retirement; Discussion about respecting the public space; Tax evasion by a politician; Macron renouncing part of his retirement; Protest actions for the increments in oil taxes.
4 Elections	Invitation to a voting counting; Discussion about election candidates; Elections; European elections.
5 Identity	Backing/lack of support from the French public personalities; Support to a Gilets Jaunes member for being attacked for not wearing her vest in a meeting; Invitations to mobilize and share ideas; Discussion about Gilets Jaunes mobilizations; Discussion about backing and opposition towards the movement; Police as an enemy; Critics and proposals for mobilizations; Petitions to support street doctors; Discussion about police demonstrations; Discussion on the lack of commitment by some Gilets Jaunes; A homeless Gilets Jaunes sharing his FB homeless-group page design.
6 Online participation	Comments about the Eurovision; Collecting money for a Gilet Jaune to pay a fine; Adoption of a dog; Annual homeless death rate.
7 Protest organization and coordination	Organizational discussion about gathering signatures and a meeting; boycotting the financial sector; Grand Débat; demonstrations; and a train strike.
8 Political discussions	Discussion of alliances with ecology movement; Debt of France; Banks; Leaving EU; Red scarf movement (a short-lived movement aimed at supporting Macron's policy against the Gilets Jaunes); Leaving EU: Free trade and other aspects; The LREM alliances in the parliamentary elections: Money support; Representation of the Gilets Jaunes; Ecology and Greta Thunberg; Discussion of alliances with the movement; Conversations about trades and banks; Notre-Dame fire.

plicity in covering up investigations that might have resulted in the condemnation of illegal procedures and in allowing the use of heavy weapons by police during the protests. Hence, we named topic 1 *political and police abuse*.

Conversations extracted from topic 2 were also characterized by violence. In these discussions, participants primarily expressed emotions of anger and outrage. The qualitative analysis showed that such negative emotions were mostly triggered by police abuse and opinions regarding politicians. A lack of support for the movement from different sectors of society also generated discomfort. Insults and expressions of anger were directed toward out-group actors. However, some of these comments triggered fights among participants and created internal conflicts. In general, the themes that inspired the most aggressive and violent emotions were elections, discussions of political parties, the ideologies of *Gilets Jaunes* members, and the modalities of protest. Some of these discussions were quite harsh, but they always included people trying to moderate, calm and redirect the discussion to a productive conversation. These results converged with those from the topic modeling, which included angry words in this topic. The topic model also showed that this topic was the largest in the corpus. Topic 2 was therefore named *anger and outrage*.

The threads in topic 3 focused on the movement's demands to reduce income inequality and social injustice. As the movement stemmed from protests against increased taxes on diesel and petrol, the injustice of the taxation system was extensively discussed from the start. Users wanted to roll back Macron's tax cuts for the wealthy and his pro-business economic agenda. Participants argued that a fair tax reform would allow for a redistribution of the national budget to benefit the majority of French people, not merely the elite. A number of concrete measures were advanced, notably including an increase in the minimum wage and a reform of the pension and retirement system alongside the improvement of salaries and the allocation of better funding for education, the healthcare system, and social security. Words connected to problematic situations concerning salaries, retirement, and taxes also emerged in topic 3, which shared some terms referencing country-level themes with other topics (1 and 4). Topic 3 was therefore labeled *social claims*.

Topic 4 was labeled *elections*, as it focused on party politics and elections at various levels ranging from the European Parliament to municipal elections. Within this category, users discussed various aspects of democracy: the organization and characteristics of European Parliament elections, the candidates involved in various elections, and the people who were asked to count the votes. Words such as "elections," "European," "voice," and "democracy" were associated with this topic in the automatic analysis. However, the qualitative content analysis also allowed us to identify two positions: (a) individuals who believed that the *Gilets Jaunes* should continue to be an apolitical and independent movement and (b) those who believed that the *Gilets Jaunes* should become a political party. In line with this latter position, debates arose concerning who could be a legitimate candidate for the movement and who should receive votes in the absence of a clear *Gilets Jaunes* candidate. Different ideological and party positions emerged. Some participants declared their intention to vote for the party that ran candidates whose proposals were least opposed to the principles of the *Gilets Jaunes*. Other participants expressed the intention to vote for ideological reasons, especially for parties on the political left. Finally, this topic also included discussions concerning whether the movement should join other movements to gain more strength and power.

Topic 5 included words such as "*Gilets Jaunes*," "demonstration," and "cause," as well as terms such as "family," "to stay," and "to believe." The content analysis confirmed that this category pertained to the *identity* of the movement. The relevant discussion threads debated topics such as who supported or opposed the movement and what its strength and critical points were. In these debates, people attempted to characterize the social actors who supported the movement and those who opposed it. While some actors were considered to be enemies of the movement (e.g., Macron, LREM, and the police), the status of others, such as supporting citizens or the international community, changed over time, as their categorization depended more on the media portrayal of the demonstrations and the modes of those demonstrations at

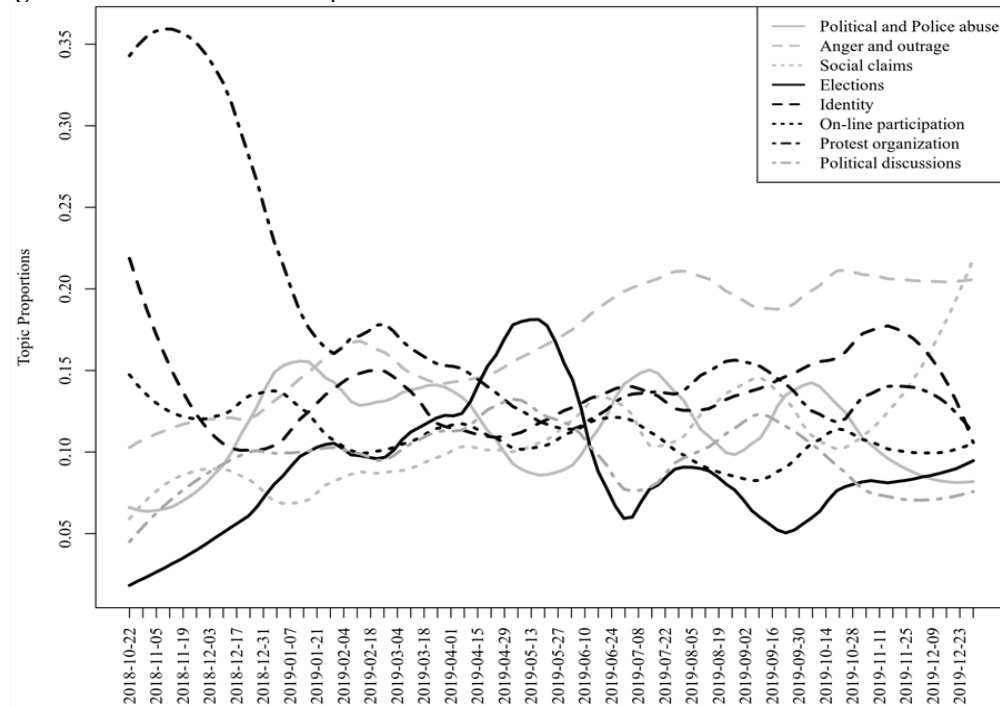
different times (e.g., occupation of roundabouts, blockades, and manifestations). In these threads, the characteristics of the movement, the modalities of protest to be employed, and the political and ideological positions of the Gilets Jaunes were also discussed.

Explicit reference to aspects of the internet (e.g., “Facebook,” “site,” “to share,” and “to sign”) defined topic 6 in the automatic analysis. The content of these conversations was very diverse, although they all included the common goal of sharing information, photos, and videos. Sharing information had three different functions in conversations: (a) denouncing people or situations, (b) providing data and statistics to expand knowledge concerning a political or social event, and (c) requests for collaboration. The last category included various actions, such as signing petitions, invitations to participate in events, and collecting money for a cause. In this topic, people discussed the role and functions of Facebook as a platform that facilitates online political participation and social actions for the movement. We found threads focused on inspiring concrete political actions to promote political and social change: people reported information related to social and political issues, promoted online petitions, and asked others to collaborate for a cause. In this specific topic of *online participation*, conversations did not become debates.

In topic 7, discussions focused mainly on organizing protest activities, meetings, and debates. Words associated with this topic included “gathering,” “blocking,” “to come,” “toll” (station), “roundabout,” and “Saturday,” all of which refer to specific modalities of the Gilets Jaunes protest: occupying roundabouts and toll stations every Saturday. The content analysis confirmed this group and showed that the analyzed threads focused on three types of actions: street demonstrations, boycotts, and gathering signatures for petitions. Topic 7 was the second largest in the corpus and was labeled *protest organization*. In discussions concerning these activities, people expressed opinions about the best protest modality, the best city to organize a demonstration, and the allies of the demonstrators (e.g., discussions concerning whether to protest alongside the General Confederation of Labor (CGT) trade union). People also expressed their concerns regarding possible infiltration, which could leak information to the *blacpolice* and allow the authorities to suppress participants’ activity.

Finally, topic 8 concerned *political discussions*. In this topic, we found themes related to national and international politics, such as the following: (a) Gilets Jaunes’ political partners, (b) national debt, (c) banks, (d) the possibility of breaking with the EU, (e) international trade, and (f) national political actors. Depending on the subject, various degrees of agreement were attained. However, despite differences in the movement’s constitution and national party policy, there was broad consensus regarding international policy. France’s membership in the European Union was identified as the ultimate cause of the country’s problems, as such membership prevented sovereignty over the national budget, international trade policy, and international politics in general.

After these topics had been interpreted through the convergence of the automatic and qualitative analysis results, we calculated the prevalence per week of each topic to explore changes in these discussions on Facebook over time and to relate such changes to contextual events. In line with our expectations, figure 3 on the next page shows that *protest organization* was predominant, especially during the first period of the movement. Some topics, such as *online participation* and *political discussion*, remained relatively stable throughout the whole period, with a probability of 5% to 10% that a randomly extracted post would address these themes. Events ignited other discussions: the *election* topic dominated the discussion during the period surrounding the European Parliament elections on May 26; *political and police violence* was boosted by the “Affaire Andrieux” and remained quite central overall. In that case, Didier Andrieux, a chief police officer in Toulon and the commander of the local riot control units, had been filmed assaulting apparently peaceful protesters on multiple occasions. Linear growth trends were observable for *social claims* and *anger and outrage*. These two topics, especially the latter, became progressively more central to the discussion. Finally, after the first period, the *identity* topic, which had been the second most likely subject of conversation, decreased during the first half of 2019. In line with our expectations, this topic regained centrality in the latter period of the movement.

Figure 3. Distribution of the Topics across Time

Note: Time is expressed on the x-axis in weeks starting on Monday (the displayed date).

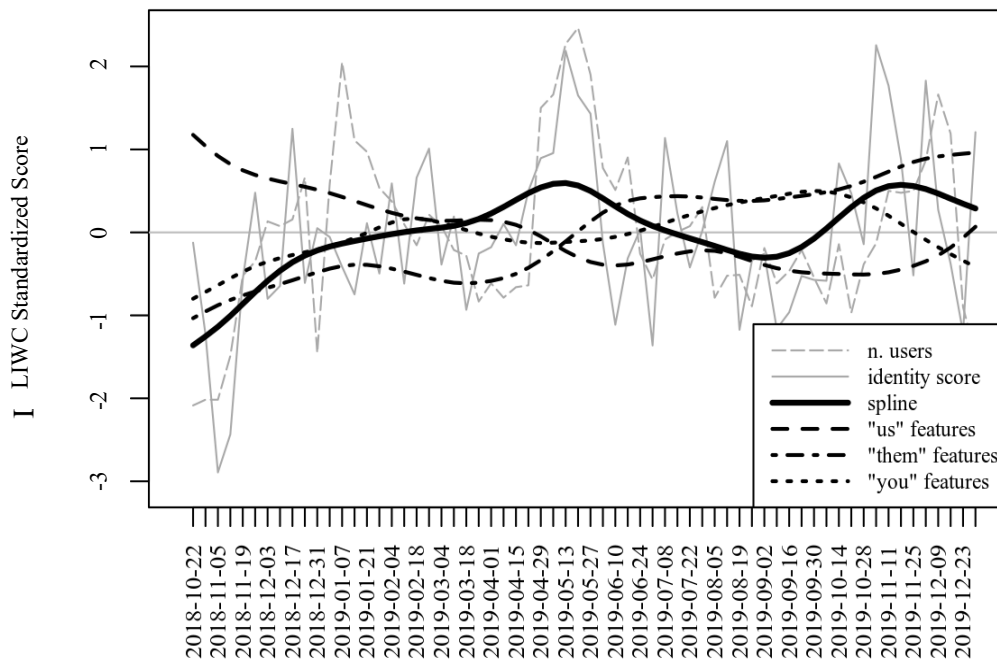
Collective Identity across Time and Topics

To investigate the progress of identity discussions in further detail, we applied the LIWC algorithm to the corpus by way of our dictionary of collective identity. Compared to the topic analysis, which focused on the themes of the discussion (i.e., the subjects that users discussed), this subsequent analysis investigated whether certain identity-related features were referenced within and across topics. In other words, this analysis examined whether the language used was identity related, even when discussing subjects that were not directly related to collective identity. Figure 4 reports the results of this analysis. The solid grey line represents the standardized score, and the bold black line indicates the spline, which is easier to interpret. Both lines show an increasing trend, with identity features becoming more central as the movement evolves.

Because participation in the movement had increased during the second half of 2019 (including a specific surge in participation following the anniversary of the movement toward the end of the year), it could be that only core members of the groups remained active, as the topic of identity may have been more central to these members than to other members who had ceased to participate. To control this selection bias, we calculated the number of active users (i.e., the number of users posting or interacting with other posts) per week. If selection bias were a relevant factor, we should have observed a correlation between an increase in collective identity scores (higher presence of the topic in the discussions) and a decrease in the number of users (only core members remain). Figure 4 shows that no such correlation was found. A Pearson partial correlation trend test ($r = .26$, $p = 0.034$) controlling for the number of users confirmed the growth trend of the identity score.

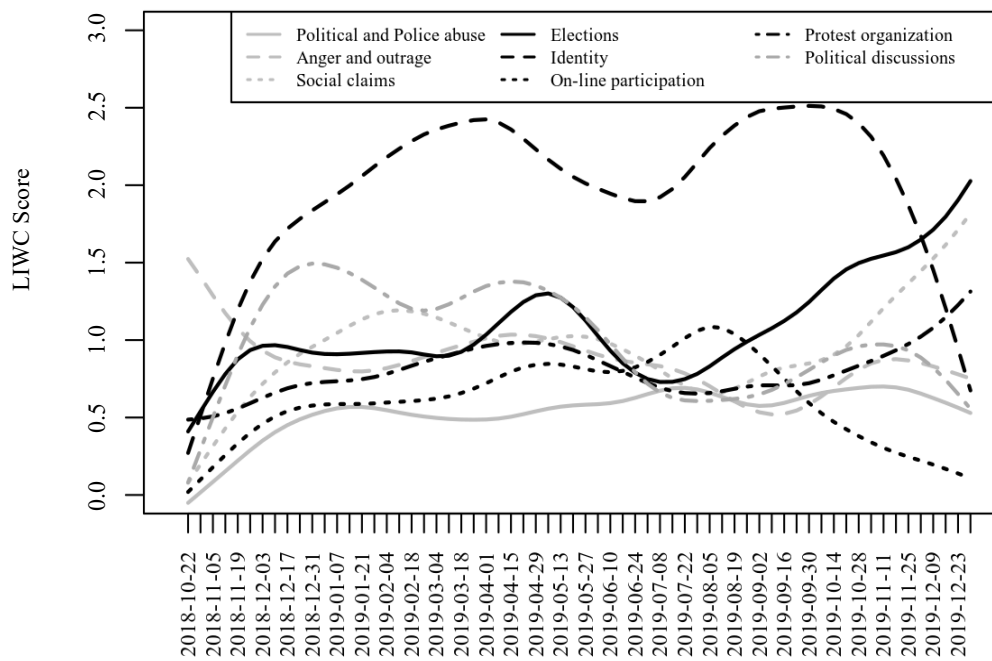
To further investigate the dynamic of collective identity, figure 4 also reports the variation over time of “us,” “them,” and “you” features. The use of “us” was less relevant after the first

Figure 4. Distribution of the Collective Identity Scores Overall



Note: Time is expressed on the x-axis in weeks starting on Monday (the displayed date)

Figure 5. Distribution of the Collective Identity Scores by Topic



Note: Time is expressed on the x-axis in weeks starting on Monday (the displayed date)

period of the movement. In line with our expectations, both “them” (denoting outgroup opposition) and “you” (denoting within-group opposition) features became more frequent, following the collective identity trend. This result shows an increase in language referring to outgroup actors and confirms the qualitative analysis of the collective identity topic by showing that references to within-group division became more frequent during the definition of allies and foes.

Because we were interested in assessing whether identity features increased as the movement developed and in understanding which discussion themes became more identity-centered over time, we performed the LIWC analysis for each topic. The results are reported in figure 5. Unsurprisingly, confirming the previous analyses, the *identity* topic had the most identity features. In line with our content analysis, the figures also show that collective identity was marginal during the first period, with the exception of the *anger and outrage* topic. Identity features were then found for the *political discussions*, *social claims*, and *election* topics. The centrality of identity to these topics increased during the latter part of the year. Similarly, we observed an increase in the centrality of identity in discussions related to *protest organization*.

DISCUSSION

In this study, we used a mixed-method design to investigate changes in collective identity in the context of online discussions of the Gilets Jaunes movement in the Var department. Although our analysis does not allow us to identify all the facets of the Gilets Jaunes identity, it frames the question of identity as part of a specific timeframe, which analyses of this topic often fail to do; in addition, our analysis highlights a series of key features that can allow us to understand the Gilets Jaunes movement at the local level.

Our topic and content analysis showed that Facebook should be regarded as an essential instrument for the constitution of the Gilets Jaunes in Var. This platform was fundamental not only to the spontaneous creation of this horizontal political movement but also to the movement’s process of collective identity construction. In line with Melucci’s (1996) definition of collective identity as a dynamic process, Facebook had different functions depending on the orientation of the movement’s actions. In the first period, the Var Gilets Jaunes movement progressed through the first two phases of the collective identity construction process, which Polletta and Jasper (2001) define as emphasizing collective claims and focusing on member recruitment. Facebook was predominantly used to organize protests, discuss group identity, promote online participation, and express and share anger and outrage with respect to the government.

After this first period, as in any process of identity construction concerning a social movement (Polletta and Jasper 2001), the Var Gilets Jaunes needed both to plan strategies and tactics to represent the movement in the political sphere and to discuss the results of their actions. As part of this process, Facebook was used to a greater extent to talk about national or European elections, social claims, participants’ experiences with mobilizations, their feelings of injustice in the face of government and police abuses, and their identity as well as to facilitate participation in online political actions. The elections topic peaked immediately before the May 26, 2019, European elections. These elections are usually shunned by less politicized citizens in France, who view the European Parliament as distant, technocratic, and unimpactful (Muxel 2005). The 2019 elections featured an increase in participation, likely due partly to the Gilets Jaunes movement. The question of whether to vote in the elections was frequently discussed in the virtual spaces we investigated. Qualitative research (Ravelli 2020; Rubert 2021) has identified the three main positions on this topic. Some activists wanted to vote for Marine Le Pen to deny a clear victory to Macron’s party, arguing that she was the only real challenger. Other activists wanted to boycott the election entirely to express their discontent with the representative system. Supporters of the third position ultimately considered voting for Gilets Jaunes candidates, which sparked a great deal of controversy due to the movement’s ambitions

toward horizontality. Some participants believed that running for an election contradicted the core tenets of the *Gilets Jaunes*.

Although the social network was also used for political debate, as shown in the longitudinal topic analysis, this function was not the most central use of Facebook discussions. It would seem that this mode of conventional political participation (i.e., participating in debates and meetings) could not be recorded by or accomplished through the Facebook platform. A variety of reasons could explain the relative lack of political debates in the online discussions. According to Coretti and Pica (2018), people tend to communicate on Facebook via effortless information exchanges because the main goal of communication is to establish a social presence rather than to engage in informed discussion. If this argument applies to political debates, it must be noted that our analysis showed that the *Var Gilets Jaunes* shared information from the web, including videos, photos, and forwarded messages from other Facebook pages, thus contributing to the production of common knowledge. The shared information focused on defining social claims and other topics rather than lengthy discussions of political ideas.

According to Humprecht, Hellmueller, and Lischka (2020), deliberate conversations are lacking in social media because hostility undermines the preconditions for productive exchange. Discussions of political topics on social media seem to inspire hostile emotions such as anger, contempt, disgust, frustration, and hate. Social media conversations are characterized by few political debates and many hostile situations where users express polarized political positions. In contrast, Miller and Conover (2015) suggested that, although hostility undermines debate, this emotion is an indicator of people's commitment to and identification with a political party. In these conversations, support for the in-group and attacks on the out-group are more important than participation in thoughtful political debate. In line with the position of Miller and Conover (2015), our results show that, on the one hand, hostile expressions directed toward consensus enemies (e.g., the government and the police) reinforced the bonds among members. On the other hand, political discussions pertaining to elections triggered violent discussion and created internal conflicts. According to Bendali, Challier, Della Sudda, and Fillieule (2019), the *Gilets Jaunes* movement, being ideologically diverse, needed to defuse conflicts to maintain unity, which might have created a strong incentive to avoid political discussions, thus providing a possible explanation for the relatively low amount of political discussion that we observed. In line with this possibility, our results showed that the *Var Gilets Jaunes* discussed their grievances more extensively than political matters. In other words, the activists found it easier to produce consensus regarding their demands than with respect to the best way of achieving those demands politically. Qualitative research concerning the use of Facebook by the French popular classes has indeed emphasized that people tend to share consensus opinions to reinforce a "shared moral common ground" (Pasquier 2018) and strengthen the in-group by expressing commonly agreed-upon values.

In line with Humprecht, Hellmueller, and Lischka (2020), our longitudinal findings showed an opposite trend between the anger and outrage and the political debate topics, with the latter decreasing as the former increased. This result shows that hostile exchanges decreased the possibility of debating antagonistic political or ideological positions. The *Var Gilets Jaunes* used Facebook to express emotions and interactive frames, including attempts to inform and share information to promote strategies to engage in collective actions to advance their demands. An increase in the frequency of the topic *anger and outrage* also corresponded to increases in the *political and police abuse* and *identity* topics. Ire and indignation were common features in discussions regarding violence and abuses of power. According to the *Gilets Jaunes* users in our analyses, democracy had transformed into a dictatorship, and the police became aggressors instead of figures representing order and justice. These users felt betrayed by the government and the police, who turned protesters into criminals by taking away their right to protest peacefully with their families.

In contrast, the identity topic sparked within-group confrontation and even aggressive conflict. Our content analysis showed that group identity characteristics were quite general, allowing people with very dissimilar ideological positions to meet in the same space. As shown

by our data, this laxity with respect to identity resulted in constant discussions concerning the course of the movement and the identity of a prototypical Gilet Jaune. The within-group distinctions, captured by the text's increase in "you" features, highlight this dynamic, showing how it gained more centrality over time. A Gilets Jaunes identity was constructed first through practices of direct action, but this process then shifted to focus on interactions that constantly sought to provide a definition of an "us," although such a definition was not necessarily achieved.

This constant identity work was highlighted in both of our analyses. First, the triangulation between the automatic and qualitative topic analysis showed that the Gilets Jaunes constantly and spontaneously discussed their identity online. Even if, on the one hand, they agreed that the movement was to be considered a large and inclusive family with well-defined enemies (e.g., the government, Macron, and the police), less consensus was found with respect to defining the movement's allies, such as the media, bystanders, trade unions and political parties. Our analysis showed that participants were mostly able to define an identity based on what they were not instead of what they were. The increase in references to external actors in the longitudinal analysis supports this finding, in line with previous studies on the "us" versus "them" dynamic in social movements (e.g., Klandermans 2014). However, our analysis also expands this literature by providing empirical support for the hypothesis that the definition of "us" is related to external actors and internal group divisions. Conflictual positions emerged after the first half-year. Though it could be expected that internal division would lead to a selection of movement members, with some dropping out, our data showed this was not the case in the time window we analyzed. On the contrary, despite these within-group divisions, participation rose again, though this was likely due to external factors (e.g., the anniversary of the movement). Nevertheless, the Var Gilets Jaunes remained a heterogeneous, contradictory, and internally divided group.

In addition to these findings, the dictionary-based analysis showed that identity features were present in a variety of topics. Not only was identity discussed as a topic per se, but several other themes became identified, to use Melucci's term. Our results show that identity work was not confined to the way the movement defined itself but interfaced with the context through which the Gilets Jaunes moved. In other words, identity features constituted the frame used to interact and discuss key aspects of the movement, such as claims and elections. The ingredients of identity were far from fixed and constant; instead, they shifted over time. In the first phase, they were more related to the definition of a sense of collective, identified by the wide use of the "us" feature in the discussions and identity features related to anger and outrage. However, with the aging of the movement, other aspects became infused with identity salience, notably linked to the core claims of the movement, opposition to external actors, and internal contestation and negotiation.

Collective identity is thus a complex and nonlinear phenomenon in which several factors are concurrent in defining the movement identity. To paraphrase complexity theory (Taylor 2001), if we consider identity to be the outcome that the movement "needs" to achieve, several resources compete and overtake each other to achieve this task. Our results show that the identity of the Var Gilets Jaunes was multifaceted, included contradictory positions, and was channeled through different aspects of movement life that changed in importance and priority over time. If some of the increase in identity-related features in our data can be identified as a response to external events (e.g., the increase in police violence or the disappointing results of the European elections), this finding provides empirical support for the claim that the construction of a collective identity is essential to the movement and becomes more central as the movement matures (Melucci 1996; Polletta and Jasper 2001). Indeed, it could be argued that this rise in identity discourse and controversy is related to the waning of the national movement from the spring of 2019 onward. The growing difficulty of staging large-scale protest actions and the dwindling numbers of supporters and participants alike could have shifted the focus of the local Gilets Jaunes from an outward perspective to an inward perspective. As noted by Freeman (1972: 159), for activist groups with no clear operational aim, "the mere act of staying together becomes the reason for their staying together," and the need to justify their togetherness leads to in-group criticism and attempts at "personal control," according to which activists increasingly clash with respect to the task of defining the real identity of the group. When people are focused in unison on a common

goal, as was the case at the beginning of the *Gilets Jaunes* movement, these topics seem far less important in the face of the challenges at hand.

To conclude, Facebook was not merely a platform that allowed activists to coordinate protest actions, it was also a place to communicate and exchange information, opinions, and experiences, thus facilitating political participation. These activities shaped the movement's identity and were accompanied by offline protest actions (Ramaciotti Morales, Cointet, and Froio 2022). Although it has been argued that spontaneous social movements created through social networks are less costly and complex to organize than political parties or traditional social movements (Chen 2019), we showed that the *Gilets Jaunes* in the Var had to face constant identity challenges to become institutionalized as a social movement (Coretti and Pica 2018). Their horizontal structure and overt lack of ideological and political positions forced constant and increasingly frequent discussions concerning the definition of the movement's boundaries and claims. The identity work was constantly in play during different phases of the movement, though the focus of the work and its intensity changed with time.

These processes could be seen in the context of the Var *Gilets Jaunes* only when the large time window we analyzed is considered, which was possible by way of the triangulations between in-depth qualitative analysis and quantitative methods that are facilitated by big data. Far from being the only possible method for studying collective identity, the one we propose has advantages and limitations. Concerning the latter, automated analysis has the disadvantage of being somehow noisy. For instance, in analyzing the presence of "us" and "them" features, it is difficult to properly assess whether they are used consistently in speech to refer to intergroup dynamics. To minimize the risks of measurement error, we adopted a mixed-method approach by manually checking the results of the automated analysis. Although we believe that this is a good practice, it cannot be performed on the entire corpus; thus, measurement error cannot be completely excluded. Another limitation of using social media analysis is that we can observe and analyze only active users, that is, the internet population who not only use the platform passively but actively engage in interactions. In our case study, this was probably less of a problem given the relatively small size of the groups and the online/offline connection that characterized this movement (Ramaciotti Morales, Cointet, and Froio 2022); however, silent participants disappeared from our analyses. On the other hand, the analysis of large corpora has the advantage of providing an overview of the movement over an extended period. This method has several advantages compared to interviews or survey data. First, it allows us to adopt a continuous definition of time. While surveys and interviews can be performed only at specific intervals, social media allow events and changes to be monitored in real-time. Similarly, social media data are spontaneously generated in the sense that a researcher does not solicit them. Hence, the analysis of social media is intended as an ethnographic process in which the researcher interrogates the data and not the respondent. For these reasons, social media and more classic data are not concurrent but complement each other (Reveillac, Steinmetz, and Morselli 2022). In this study, we have shown that a mixed-method approach can validate old hypotheses and produce new insight into identity processes. Further applications of this methodology to other contexts and movements will allow clarification of whether our observations are specific to the *Gilets Jaunes* of the Var district or are generalizable dynamics of identity construction.

NOTES

¹ This article is part of an ongoing research project funded by the Swiss National Research Fund (MYF project—The Mystery of the Yellow France) on the biographical consequences of participation in the Yellow Vests movement in the department of Var in southeastern France.

² The use of Twitter data in the word2vec model was imposed by the need for a large corpus to train the model. Facebook and Twitter posts can have very different lengths and have different linguistic contexts for the target words. However, in this research, we have examined comments and replies to posts and found that the average length was 20.8 words, which is similar to a tweet length. In addition, the word2vec model was trained with a fifteen-tokens window, meaning that only the fifteen words before and after the target word were considered. Concerning differences in the language used on Facebook and Twitter, some studies (Herdağdelen and Marelli 2017) have shown that the two platforms do not differ

drastically in terms of linguistic properties. However, it still cannot be completely ruled out that word2vec could pick up synonyms on Twitter that are absent on Facebook and ignore Facebook-specific words, thus underestimating the identity features in the text. To overcome this shortcoming, every step was supervised by the researchers; the initial list was drawn from an in-depth reading of the discussion threads and was reviewed and revised after the automatic procedure.

³ To protect users' safety and anonymity and in agreement with Facebook/Meta's terms of service, all reported examples are adaptations and translations of text that are consistent with the original posts but do not actually exist in the corpus.

⁴ A more detailed version of this table reporting dates and thread examples is available at <https://osf.io/k4dmz>.

REFERENCES

- Ayers, Michael D. 2003. "Comparing Collective Identity in Online and Offline Feminist Activists." Pp. 145-164 in Martha McCaughey and Michael D. Ayers, eds. *Cyberactivism: Online Activism in Theory and Practice*. New York: Taylor and Francis.
- Bendali, Zakaria, and Aldo Rubert. 2020. "Les Sciences Sociales en Gilet Jaune". *Politix* 32(4): 177-215.
- Bendali, Zakaria, Raphaël Challier, Magali Della Sudda, and Olivier Fillieule. 2019. "Le Mouvement des Gilets Jaunes : Un Apprentissage en Pratique(s) de la Politique ?" *Politix* 31(4): 143-177.
- Bennett, W. Lance, and Alexandra Segerberg. 2012. "The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics." *Information, Communication & Society* 15(5): 739-768.
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "Quanteda: An R Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3(30): 774. <https://joss.theoj.org/papers/10.21105/joss.00774>
- Bernstein, Mary. 1997. "Celebration and Suppression: The Strategic Uses of Identity by the Lesbian and Gay Movement." *American Journal of Sociology* 103(3): 531-565.
- Blok, Anders, and Morten Axel Pedersen. 2014. "Complementary Social Science? Quali-Quantitative Experiments in a Big Data World." *Big Data & Society* 1(2) <https://doi.org/10.1177/2053951714543908>
- Boyer, Pierre C., Thomas Delemotte, Germain Gauthier, Vincent Rollet, and Benoît Schmutz. 2020. "Les Déterminants de la Mobilisation des Gilets Jaunes." *Revue Économique* 71(1): 109-138.
- Chen, Simin Michelle. 2019. "Women's March Minnesota on Facebook: Effects of Social Connection on Different Types of Collective Action." *New Media & Society* 22(10): 1785-1807.
- Calhoun, Craig. 1994. "Social Theory and the Politics of Identity." Pp. 9-36 in Craig Calhoun, ed. *Social Theory and the Politics of Identity*. Oxford: Blackwell.
- Clemens, Elisabeth S. 1996. "Organizational Form as Frame: Collective Identity and Political Strategy in the American Labor Movement, 1880-1920." Pp 205-226 in Doug McAdam, John D. McCarthy and Mayer N. Zald, eds. *Comparative Perspectives on Social Movements. Political Opportunities, Mobilizing Structures, and Cultural Framings*. New York: Cambridge University Press.
- Collectif d'enquête sur les Gilets jaunes. 2019. "Enquêter in Situ par Questionnaire sur une Mobilisation en Cours : une Étude sur les Gilets Jaunes." *Revue Française de Science Politique* 5(69): 869-892.
- Combes, Hélène, and Olivier Fillieule. 2012. "Repression and Protest. Structural Models and Strategic Interactions." *Revue Française de Science Politique* 61(6): 1047-1072.
- Corbin, Juliet. 2017. "Grounded Theory." *Journal of Positive Psychology* 12(3): 301-302.
- Coretti, Lorenzo, and Daniele Pica. 2018. "Facebook's Communication Protocols, Algorithmic filters, and Protest: A Critical Socio-Technical Perspective." Pp. 84-97 in Mette Mortensen, Christina Neumayer, and Thomas Poell, eds. *Social Media Materialities and Protest. Critical Reflection*. London: Routledge.
- . 2015. "The Rise and Fall of Collective Identity in Networked Movements: Communication Protocols, Facebook, and the Anti-Berlusconi Protest." *Information, Communication & Society* 18(8): 951-967.
- della Porta, Donatella, and Lorenzo Mosca. 2005. "Global-Net for Global Movements? A Network of Networks for a Movement of Movements." *Journal of Public Policy* 25(1): 165-190.
- Drury, John, and Steve Reicher. 2000. "Collective Action and Psychological Change: The Emergence of New Social Identities." *British Journal of Social Psychology* 39(4): 579-604.
- Drury, John, Steve Reicher, and Clifford Stott. 2003. "Transforming the Boundaries of Collective Identity: From the 'Local' Anti-Road Campaign to 'Global' Resistance?" *Social Movement Studies* 2(2): 191-212.
- Einwohner, Rachel L. 1999. "Gender, Class and Social Movement Outcomes: Identity and Effectiveness in Two Animal Rights Campaigns." *Gender & Society*, 13(1): 56-76.

- . 2002. "Bringing The Outsiders In: Opponents' Claims and The Construction of Animal Rights Activists' Identity." *Mobilization: An International Quarterly*, 7(3): 253-268
- Fillieule, Olivier, and Philippe Blanchard. 2013. "Fighting Together. Assessing Continuity and Change in Social Movement Organizations Through the Study of Constituencies' Heterogeneity." Pp, 79-110 in Niilo Kauppi, ed. *The New European Political Sociology*. London: Routledge.
- Fillieule, Olivier, and Danielle Tartakowsky. 2013. *Demonstrations*. Winnipeg: Fernwood Publishing.
- Flesher Fominaya, Cristina. 2010. "Collective Identity in Social Movements: Central Concepts and Debates." *Sociology Compass* 4(6): 393-404.
- . 2019. "Collective Identity in Social Movements: Assessing the Limits of a Theoretical Framework." Pp. 429-445 in David A. Snow, Sarah A. Soule, Hanspeter Kriesi, and Holly J. McCammon, eds. *The Wiley Blackwell Companion to Social Movements*, 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc.
- Freelon, Deen, Charlton McIlwain, and Meredith Clark. 2018. "Quantifying the Power and Consequences of Social Media Protest." *New Media & Society*, 20(3): 990-1011.
- Freeman, Jo. 1972. "The Tyranny of Structurelessness." *Berkeley Journal of Sociology* 17: 151-164.
- Gamson, Joshua. 1996. "The Organizational Shaping of Collective Identity: The Case of Lesbian and Gay Film Festivals in New York." *Sociological Forum* 11(2): 231-61.
- Gamson, William A. 1992. *Talking Politics*. New York: Cambridge University Press.
- George, Jordana J., and Dorothy E. Leidner. 2019. "From Clicktivism to Hacktivism: Understanding Digital Activism." *Information and Organization* 29(3): 1-45.
- Gerbaudo, Paolo. 2015. "Protest Avatars as Memetic Signifiers: Political Profile Pictures and the Construction of Collective Identity on Social Media in the 2011 Protest Wave." *Information, Communication & Society*, 18(8): 916-929.
- Gerbaudo, Paolo, and Emiliano Treré. 2015. "In Search of the 'We' of Social Media Activism: Introduction to the Special Issue on Social Media and Protest Identities." *Information, Communication & Society* 18(8): 865-871.
- Ghaziani, Amin. 2008. *The Dividends of Dissent, How Conflict and Culture Works in Lesbian and Gay Marches on Washington*. Chicago and London: Chicago University Press.
- Gusfield, Joseph R. 1963. *Symbolic Crusade: Status Politics and the American Temperance Movement*. Urbana: University of Illinois Press.
- Hara, Noriko, and Zilia Estrada. 2005. "Analyzing the Mobilization of Grassroots Activities via the Internet: A Case Study." *Journal of Information Science* 31(6): 503-514.
- Haines, Herbert H. 2006. "Dangerous Issues and Public Identities: The Negotiation of Controversy in Two Movement Organizations." *Sociological Inquiry* 76(2): 231-263.
- Herdağdelen, Amaç, and Marco Marelli. 2017. "Social Media and Language Processing: How Facebook and Twitter Provide the Best Frequency Estimates for Studying Word Recognition." *Cognitive Science* 41(4): 976-995.
- Hirsch, Eric L. 1990. "Sacrifice for the Cause: Group Processes, Recruitment, and Commitment in a Student Social Movement." *American Sociological Review* 55(2): 243-254.
- Hunt, Scott A., Robert D. Benford, and David A. Snow. 1994. "Identity Fields: Framing Processes and The Social Construction Movement Identities." Pp 185-208 in Enrique Laraña, Hank Johnston, and Joseph R. Gusfield, eds. *New Social Movements: From Ideology to Identity*. Philadelphia: Temple University Press.
- Humphrecht, Edda, Lea Hellmueller, and Juliane A. Lischka. 2020. "Hostile Emotions in News Comments: A Cross-National Analysis of Facebook Discussions." *Social Media + Society* 6(1) <https://doi.org/10.1177/2056305120912481>
- Jost, John T., Pablo Barberá, Richard Bonneau, Melanie Langer, Megan Metzger, Jonathan Nagler, Joanna Sterling, and Joshua A. Tucker. 2018. "How Social Media Facilitates Political Protest: Information, Motivation, and Social Networks." *Political Psychology* 39(1): 85-118.
- Kavada, Anastasia. 2009. "Email Lists and the Construction of an Open and Multifaceted Identity: The Case of the London 2004 European Social Forum." *Information, Communication & Society* 12(6): 817-839.
- . 2014. "Transnational Civil Society and Social Movements." Pp. 351-369 in Karin Gwinn Wilkins, Thomas Tufté, and Rafael Obregon, eds. *Handbook of Development Communication & Social Change*. Hoboken, NJ: John Wiley & Sons, Inc.
- . 2015. "Creating the Collective: Social Media, the Occupy Movement and Its Constitution as a Collective Actor." *Information, Communication & Society* 18(8): 872-886.

- Klandermans, Bert. 2014. "Identity Politics and Politicized Identities: Identity Processes and the Dynamics of Protest." *Political Psychology* 35(1): 1-22.
- Lewis, Seth C., Rodrigo Zamith, and Alfred Hermida. 2013. "Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods." *Journal of Broadcasting & Electronic Media* 57(1): 34-52.
- Mattoni, Alice, and Elena Pavan. 2018. "Politics, Participation and Big Data. Introductory Reflections on the Ontological, Epistemological, and Methodological Aspects of a Complex Relationship." *Partecipazione e Conflitto* 11(2): 313-331.
- McAdam, Doug. 1988. *Freedom Summer*. New York: Oxford University Press.
- McDonald, Kevin. 2015. "From Indymedia to Anonymous: Rethinking Action and Identity in Digital Cultures." *Information, Communication & Society* 18(8): 968-982.
- Melucci, Alberto. 1989. *Nomads of the Present*. London: Hutchinson Radius.
- . 1995. "The Process of Collective Identity." Pp. 41-63 in Hank Johnston, and Bert Klandermans, eds. *Social Movements and Culture*. Minneapolis: University of Minnesota Press.
- . 1996. *Challenging Codes. Collective Action in the Information Age*. Cambridge: Cambridge University Press.
- Milan, Stefania. 2015. "From Social Movements to Cloud Protesting: The Evolution of Collective Identity." *Information, Communication & Society* 18(8): 887-900.
- Miller, Patrick R., and Pamela Johnston Conover. 2015. "Red and Blue States of Mind: Partisan Hostility and Voting in the United States." *Political Research Quarterly* 68(2): 225-239.
- Mimno, David, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. "Optimizing Semantic Coherence in Topic Models." In *EMNLP '11: Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 262-272. Edinburgh, Scotland, UK: Association for Computational Linguistics.
- Monterde, Arnau, Antonio Calleja-López, Miguel Aguilera, Xabier E. Barandiaran, and John Postill. 2015. "Multitudinous Identities: A Qualitative and Network Analysis of the 15M Collective Identity." *Information, Communication & Society* 18(8): 930-950.
- Morselli, Davide, Stefano Passini, and Craig McGarty. 2021. "SOS Venezuela: An Analysis of the Anti-Maduro Protest Movements Using Twitter." *Social Movement Studies* 20(5): 509-530.
- Muxel, Anne. 2005. "Les Abstentionnistes le Premier Parti Européen." Pp. 45-76 in Pascal Perrineau, ed. *Le Vote Européen 2004-2005: De L'élargissement Au Référendum Français*. Paris: Presses de Sciences Po.
- Nepstad, Sharon Erickson. 2004. "Persistent Resistance: Commitment and Community in the Plowshares Movement." *Social Problems* 51(1): 43-60.
- Neuhouser, Kevin. 1998. "'If I Had Abandoned My Children': Community Mobilization and Commitment to the Identity of Mother in Northeast Brazil." *Social Forces* 77(1): 331-358.
- Nip, Joyce Y. M. 2004. "The Queer Sisters and Its Electronic Bulletin Board." *Information, Communication & Society* 7(1): 23-49.
- Pasquier, Dominique. 2018. *L'internet des Familles Modestes. Enquête dans la France Rurale*. Paris: Presses des Mines.
- Pfaff, Steven. 1996. "Collective Identity and Informal Groups in Revolutionary Mobilization: East Germany in 1989." *Social Forces* 75(1): 91-117.
- Pickerill, Jenny. 2003. *Cyberprotest: Environmental Activism Online*. Manchester, UK: Manchester University Press.
- Polletta, Francesca. 1998. "'It Was Like a Fever...': Narrative and Identity in Social Protest." *Social Problems* 45(2): 137-159.
- Polletta, Francesca, and James M. Jasper. 2001. "Collective Identity and Social Movements." *Annual Review of Sociology* 27(1): 283-305.
- Pozzi, Maura, Stefano Passini, Maria Chayinska, Davide Morselli, Adriano Mauro Ellena, Anna Włodarczyk, and Carlo Pistoni. 2022. "'Coming Together to Awaken Our Democracy': Examining Precursors of Emergent Social Identity and Collective Action Among Activists and Non-Activists in the 2019-2020 'Chile Despertó' Protests." *Journal of Community & Applied Social Psychology* 32(5): 830-845.
- Ramaciotti Morales, Pedro, Jean-Philippe Cointet, and Caterina Froio. 2022. "Posters and Protesters. The Networked Interplay Between Onsite Participation and Facebook Activity in the Yellow Vests Movement in France." *Journal of Computational Social Science*, 5(1): 1129-1157.
- Ravelli, Quentin. 2020. "Un Chaos Organisé. La Grappe de Ronds-Points Comme Structure Politique Souple." *Condition Humaine/Conditions* <https://revues.mshparisnord.fr/chcp/index.php?id=141&lang=es>

- Reger, Jo. 2002. "Organizational Dynamics and the Construction of Multiple Feminist Identities in the National Organization for Women." *Gender and Society* 16(5): 710-727.
- Reger, Jo, Daniel J. Myers, and Rachel L. Einwohner, eds. 2008. *Identity Work in Social Movements*. Minneapolis: University of Minnesota Press.
- Rehurek, Radim, and Petr Sojka. 2011. "Gensim–Python Framework for Vector Space Modelling." *NLP Centre, Faculty of Informatics, Brno: Masaryk University, Brno, Czech Republic* 3(2). <https://citeserx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b55fe23d7290f59d14e51e7813f5950f5ff08b2b>
- Reveillac, Maud, and Davide Morselli. 2022. "Dictionary-Based and Machine Learning Classification Approaches: A Comparison for Tonality and Frame Detection on Twitter Data." *Political Research Exchange* 4(1): 1-24.
- Reveillac, Maud, Stephanie Steinmetz, and Davide 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(7): 10107-10142.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. "Stm: An R Package for Structural Topic Models." *Journal of Statistical Software* 91(2): 1-40.
- Rodriguez, Maria Y., and Heather Storer. 2020. "A Computational Social Science Perspective on Qualitative Data Exploration: Using Topic Models for the Descriptive Analysis of Social Media Data." *Journal of Technology in Human Services* 38(1): 54-86.
- Roth, Silke. 2008. "Dealing with Diversity: The Coalition of Labor Union Women." Pp 213-231 in Jo Reger, Daniel J. Myers, and Rachel L. Einwohner, eds. 2008. *Identity Work in Social Movements*. Minneapolis: University of Minnesota Press.
- Rubert, Aldo. 2021. "Retrouver sa Propre Radicalité : Démobilisation, Remobilisation et Cohabitation Jaunes." *Condition Humaine / Conditions Politiques*, 2. <https://doi.org/10.56698/chcp.510>
- Sbalchiero, Stefano, and Maciej Eder. 2020. "Topic Modeling, Long Texts and the Best Number of Topics. Some Problems and Solutions." *Quality & Quantity* 54(4): 1095-1108.
- Schumann, Sandy, and Olivier Klein. 2015. "Substitute or Stepping Stone? Assessing the Impact of Low-Threshold Online Collective Actions on Offline Participation." *European Journal of Social Psychology* 45(3): 30-322.
- Shirky, Clay. 2011. "The Political Power of Social Media: Technology, the Public Sphere, and Political Change." *Foreign Affairs* 90(1): 28-41.
- Snow, David A., and Doug McAdam. 2000. "Identity Work Processes in the Context of Social Movements: Clarifying the Identity/Movement Nexus." Pp 41-67 in Sheldon Stryker, Timothy J. Owens, and Robert W. White, eds. *Self, Identity, and Social Movements*. Minneapolis: University of Minnesota Press.
- Snow, David A. 2001. "Collective Identity and Expressive Forms." *University of California, Irvine: Center for the Study of Democracy, Working Papers*. Last accessed on March 8, 2023 from <https://escholarship.org/uc/item/2zn1t7bj>
- Strauss, Anselm, and Juliet Corbin. 2014. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, Fourth edition*. Newbury Park: SAGE.
- Taddy, Matt. 2012. "On Estimation and Selection for Topic Models." *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research* 22: 1184-1193.
- Tausczik, Yla R., and James W. Pennebaker. 2010. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." *Journal of Language and Social Psychology* 29(1): 24-54.
- Tajfel, Henri, and John C. Turner. 2004. "The Social Identity Theory of Intergroup Behavior." P, 7-24 in John T. Jost, and Jim Sidanius, eds. *Political Psychology*. New York: Psychology Press.
- Taylor, Verta. 1989. "Social Movement Continuity: The Women's Movement in Abeyance." *American Sociological Review* 54(5): 761-775.
- Taylor, Mark C. 2001. *The Moment of Complexity: Emerging Network Culture*. Chicago: University of Chicago Press.
- Thomas, Emma F., Kenneth I. Mavor, and Craig McGarty. 2012. "Social Identities Facilitate and Encapsulate Action-Relevant Constructs: A Test of the Social Identity Model of Collective Action." *Group Processes & Intergroup Relations* 15(1): 75-88.
- Treré, Emiliano. 2015. "Reclaiming, Proclaiming, and Maintaining Collective Identity in the #Yosoy132 Movement in Mexico: an Examination of Digital Frontstage and Backstage Activism Through Social Media and Instant Messaging Platforms." *Information, Communication & Society* 18(8): 901-915.
- Tüfekçi, Zeynep. 2017. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. New Haven: Yale University Press.

- Valenzuela, Sebastián. 2013. "Unpacking the Use of Social Media for Protest Behavior: The Roles of Information, Opinion Expression, and Activism." *American Behavioral Scientist* 57(7): 920-942.
- Van Zomeren, Martijn, Tom Postmes, and Russell Spears. 2008. "Toward an Integrative Social Identity Model of Collective Action: A Quantitative Research Synthesis of Three Socio-Psychological Perspectives." *Psychological Bulletin* 134(4): 504-535.
- Wall, Melissa A. 2007. "Social Movements and Email: Expressions of Online Identity in the Globalization Protests." *New Media & Society* 9(2): 258-277.
- Wallach, Hanna M., Iain Murray, Ruslan Salakhutdinov, and David Mimno. 2009. "Evaluation Methods for Topic Models." Pp. 1105-1112 in *Proceedings of the 26th Annual International Conference on Machine Learning*. New York: Association for Computing Machinery.
- Whittier, Nancy. 1995. *Feminist Generations: The Persistence of the Radical Women's Movement*. Philadelphia: Temple University Press.
- Wijffels, Jan. 2020. "udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit. R package version 0.8.5." Last accessed March 8, 2023 at <https://cran.r-project.org/web/packages/udpipe/index.html>
- Zwerman, Gilda, Patricia G. Steinhoff, and Donatella della Porta. 2000. "Disappearing Social Movements: Clandestinity in the Cycle of New Left Protest in the U.S., Japan, Germany, and Italy." *Mobilization: An International Quarterly* 5(1): 85-104.