Year : 2023

Three Essays on Financial Social Media

Goutte Maud Rose

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE FINANCE

Three Essays on Financial Social Media

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès Sciences Économiques, mention « Finance »

par

Maud Rose GOUTTE

Directeur de thèse
Prof. Michael Rockinger

Jury

Prof. Rafael Lalive, président
Prof. Norman Schürhoff, expert interne
Prof. Daron Acemoglu, expert externe
Prof. Rüdiger Fahlenbrach, expert externe

LAUSANNE
2023

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE FINANCE

Three Essays on Financial Social Media

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès Sciences Économiques, mention « Finance »

par

Maud Rose GOUTTE

Directeur de thèse
Prof. Michael Rockinger

Jury

Prof. Rafael Lalive, président
Prof. Norman Schürhoff, expert interne
Prof. Daron Acemoglu, expert externe
Prof. Rüdiger Fahlenbrach, expert externe

LAUSANNE
2023

UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

IMPRIMATUR

Sans se prononcer sur les opinions de l'autrice, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Madame Maud Rose GOUTTE, titulaire d'un bachelor en Economie de l'Université de Lausanne, et d'un master en Finance de l'Université de Lausanne, en vue de l'obtention du grade de docteure ès Sciences économiques, mention « finance ».

La thèse est intitulée :

THREE ESSAYS ON FINANCIAL SOCIAL MEDIA

Lausanne, le 9 juin 2023

La Doyenne

Marianne SCHMID MAST

Members of the Thesis Committee

Prof. Rafael LALIVE
Vice-Dean of the Haute Etudes Commerciales of the University of Lausanne and Professor of Economics at the University of Lausanne
President of the thesis committee

Prof. Michael ROCKINGER
Professor of Finance at the University of Lausanne
Thesis supervisor

Prof. Norman SCHUERHOFF
Professor of Finance at the University of Lausanne
Internal member of the doctoral committee

Prof. Daron ACEMOGLU
Professor of Economics at the Massachusetts Institute of Technology
External member of the doctoral committee

Prof. Rüdiger FAHLENBRACH
Professor of Finance at Ecole Polytechnique Fédérale de Lausanne
External member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Maud Rose GOUTTE

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________ Date: 11. 05. 2023

Prof. Michael ROCKINGER
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Maud Rose GOUTTE

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________ Date: 12.05.2023

Prof. Norman SCHUERHOFF
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Maud Rose GOUTTE

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Date: May 11, 2023

Prof. Daron ACEMOGLU
External member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Maud Rose GOUTTE

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Rüdiger Fahlenbrach Date: 05/15/2023

Prof. Rüdiger FAHLENBRACH
External member of the doctoral committee

Contents

Introduction

This thesis focuses on examining the influence of social media on financial markets, particularly the predictive power of financial network users, and delves into the skills of social media investors on two widely used platforms. The study also explores the presence of echo chambers in financial social media and investigates the determinants of investors' beliefs concerning individual stocks. Echo chambers occur when individuals are surrounded by like-minded peers and consume information that reinforces their beliefs, strengthening biases and potentially contributing to herd behavior. The research findings emphasize social media's crucial role in shaping trading patterns and investment decisions, underscoring the need to evaluate its impact on the finance industry. Each of the three chapters in this thesis is the result of single-authored work.

The first chapter aims to investigate whether the investment recommendations made by social media users have predictive power for future returns in financial markets. Specifically, I compare the skills of users on two popular social media platforms, namely StockTwits (ST) and Seeking Alpha (SA). ST is a free platform that leverages the collective intelligence of the crowd, while SA is a subscription-based platform that draws upon the expertise of financial professionals. To measure the skills of social media investors, I use the average abnormal returns associated with a change in recommendation as a proxy. The results indicate that the skills of ST users surpass those of SA authors, even when I consider only those recommendations that are not made simultaneously with company news. I further explore potential sources of users' skills, including their experience on the platform, the period of time, their audience, and the proportion of recommendations made following news announcements. The regression analyses reveal that users with a higher proportion of recommendations made after analyzing news, more time spent on the platform, and a large community of followers tend to have higher skills.

The second chapter takes a network perspective to investigate the impact of social media on financial markets and the predictive power of financial network users. By analyzing the relationship between users' skills, centrality, and network characteristics, the study aims to identify the key drivers of their predictive power. Furthermore, the paper seeks to distinguish between competing hypotheses: the wisdom of crowds and echo chambers. Echo chambers can have negative social consequences and pose a risk to financial markets, making it crucial to understand their presence in financial social media. To address these issues, the paper uses data from the

StockTwits platform to construct two finance investors' networks, identifying influential users and ranking them based on their prediction power and demonstrated skill. The analysis reveals that a user's centrality in the network significantly influences their performance in predicting financial markets, with the most central users reducing the diversity within the network and creating echo chambers. These echo chambers pose a risk to investors and prevent them from making well-informed decisions. Overall, this study emphasizes the need to understand better how social media platforms can influence trading patterns and investment decisions. By highlighting the risks associated with the popularization of financial social networks, this paper underscores the importance of evaluating the role of social media in finance.

The third chapter of this thesis, published in the Journal of Behavioral and Experimental Finance in 2022, focuses on the determinants of investors' beliefs regarding individual stocks at a daily frequency using extensive data available on the StockTwits platform. Specifically, millions of beliefs on different stocks are analyzed to determine the sources of changes in belief and examine how individual sentiment affects future sentiment for a particular stock. The study tests the hypothesis of whether investors' future beliefs are affected by current sentiment, trades in the previous period, the amount of communication, and the size of the message's audience. I find that higher average sentiment in tweets leads to convergence in beliefs, while an increase in communication leads to more diversity of opinion. Additionally, it is discovered that an increase in general recommendation quality does not help reduce uncertainty. Furthermore, the analysis suggests that social media sentiment and uncertainty are not always aligned with newspaper articles, highlighting the importance of incorporating social media data in financial analysis. Overall, the findings provide insights into the role of social media in shaping investor beliefs and have significant implications for market efficiency and financial decision-making.

Chapter 1

When the ‘Dumb’ Crowd Beats the ‘Smart’ Crowd

This paper focuses on the skills of social media users in issuing investment recommendations. On average, when users make recommendations, the resulting abnormal returns tend to be positive. The cross-section of skills across stocks outperforms, on average, the cross-section of skills across users. Modeling users’ performance with Gaussian mixture distributions for several skill groups shows that the large majority of users exhibit positive skills, outperforming the skills of qualified financial website authors. Social media user communities demonstrate heterogeneous skills due to several determinants, including analyzing news events, experience, the number of recommendations issued, and the number of users’ followers.

1.1 Introduction

In finance, is it worth spending money to earn money? Historically, professionals have paid analysts, financial specialists, and consultants to guide their investment strategies. At present, despite the broader access to information in general (new websites, real-time notifications, social media forums), analysts are still paid to make recommendations, and several websites offer reliable investment advice at a given cost.[1] In contrast to these fee-based sources of information, the information presented on social media is free. However, information on these platforms is considerably noisier, and the large amount of data available requires time to aggregate and process. Furthermore, the informal language found on these platforms raises concerns about the credibility of their financial advice. Undoubtedly, anyone who consults investment recommendations in the form of tweets will think that the information is substantially less trustworthy than the well-written recommendations found on specialized websites. Conversely, in line with the idea of the wisdom of the crowd, one can argue that the aggregated recommendations of a large group might be more valuable than the recommendation of an individual finance professional. Recent real-life examples supporting the merits of the quality of information available through social networks are innumerable: investors have obtained information on Twitter before financial news providers, and tweets from financial gurus can move markets by billions (Elon Musk, Carl Icahn),[2] and Reddit traders recently caused billions of dollars in losses on Wall Street.[3] According to the literature, several financial and social media platforms have predictive power for financial markets (Antweiler and Frank (2004), Bollen et al. (2011), Das and Chen (2007)). Therefore, conducting a more thorough analysis of the users' behavior on these platforms is worth considering.
Several questions arise from these considerations: Is it still worthwhile to pay subscription fees for financial analyses given by sites that are seemingly more credible? Alternatively, is the information that the 'crowd' provides through social media of the same economic value? What are the skills of these social media investors, and what are their determinants?

According to the assumption that investors' recommendations on social media platforms (free of charge or for a fee) are informative about future returns, this paper builds on the literature on analyst skills (Crane and Crotty (2020)) and hedge fund skills (Chen et al. (2017)) to estimate the skills of social media investors. I compare two social media platforms: one free site powered by the crowd and another accessible via subscription powered by paid authors. Given that the paid

[1] As an example, the Seeking Alpha (SA) platform provides, apart from its free news services, professional investment advice for an approximately 499$ subscription fee per year.

[2] https://www.cnbc.com/2013/04/25/Twitter-Trading:-8-Tweets-That-Moved-Markets.html

[3] https://www.ft.com/content/56e8b33a-d9b6-4f74-998b-327ef54c4d5a

authors are mostly financial professionals, they contrast with the crowd investors and represent in this study a group of "qualified investors." [4]

To represent the "crowd investors," I use the users' recommendations on the platform StockTwits (ST hereafter). The "smart crowd investors" are represented by the authors of the site Seeking Alpha (SA hereafter). Both platforms are widely used, with millions of visitors per month. Given the frequency of the recommendations issued, SA data represent adequate professional recommendation data to be compared to ST users. In Section 1.4.4, I further motivate the choice of SA authors to represent a group of qualified investors. In the context of this study, I focus on S&P500 stocks.

ST recommendations are labeled through machine learning classifiers, while the SA website directly offers investment recommendations labeled as 'Long ideas' and 'Short ideas.' After preprocessing the data, there are 2,389,745 and 34,946 recommendation events stemming from 2014 to 2020 for ST and SA, respectively. The number of users from ST is 62,346, and there are 2,099 authors for SA.

Following Crane and Crotty (2020), I measure skill using average abnormal returns associated with a change in recommendation. A positive (negative) change in a recommendation event is defined by a user issuing a negative (positive) recommendation and, later a positive (negative) recommendation. Buy and hold abnormal returns are estimated using the market model with a three-day event window (other potential event windows are included in Section 1.4.4). When capturing the price movement following a recommendation revision, I exclude the day of the recommendation.

To convert a change in the recommendation measure to a skill measure, I use the buy and hold abnormal returns associated with buying stocks following an upgrade and selling stock following a downgrade. Individual skill refers to the average abnormal performance following the user/author recommendations (user-level aggregation). Crowd skill corresponds to the average abnormal performance following the recommendations of all the users on a stock (stock-level aggregation).

Descriptive statistics indicate that, on average, individual skills are higher on the ST platform, although the average skills are positive for SA authors and ST users. Notably, the differences in skills between the two platforms persist when we remove from the sample the recommendations made simultaneously with company news. For groups of more active users on the platform, average individual performance increases while the variability of the users' skills decreases. At the stock level, 92% of the abnormal performances are positive for ST, indicating that aggregated recommendations from many users result in positive abnormal performance in nearly all cases. For both platforms, the average crowd skill exceeds the average individual user skill. The differences in the contributors' skills on the two platforms remain similar when running the same analysis

[4] https://seekingalpha.com/who-to-follow

with stocks of the same type (same size, sector, and B/M ratio).

Given the size of the groups of contributors and the considerable diversity between the characteristics of users, there is heterogeneity in users' skills. This heterogeneity, together with the strong rejection of the normality hypothesis for the distribution of skills on the two platforms, motivates using a mixture distribution to model the skills of users/authors. Mixture distributions are appropriate to estimate the proportion of skilled versus unskilled investors because of the group structure in the data (McLachlan et al. (2019)).

Following Crane and Crotty (2020) and Chen et al. (2017), I model the performance of ST users and SA authors as a mixture distribution of multiple skill groups. The Bayesian information criterion (BIC) suggests a mixture distribution of three normal distributions for ST and two normal distributions for SA. To simultaneously estimate the fraction of users belonging to each skill group, the average skill of each group, and the dispersion in skills within all groups, I use the expectation-maximization (EM) algorithm. The results show that the large majority of ST users are drawn from a distribution centered around a positive abnormal return, with 54% of users reporting an abnormal performance of more than 0.14% (over a three-day horizon). The majority of SA authors are drawn from a distribution centered at zero abnormal returns, and 22% of the authors come from a distribution centered on 0.08%.

All estimates in this paper suggest that the skills of ST users, although representing a more heterogeneous group of users, exceed the skills of SA authors. These results have important implications, illustrating that although the recommendations on ST may seem unreliable, they provide information with substantial economic value. This information is, on average, even more trustworthy than that of sites considered legitimate sources of financial advice. For this purpose, the last analysis in this paper focuses on the sources of the skills of ST users.

Potential sources of ST users' skills include experience (i.e., the amount of time a user spends on the platform), the period of time, the users' audience, or from users issuing recommendations both from following news (information processing) and without any prior news on a particular stock (information production). I use the SA 'Top news' and 'Latest news' headlines as news data. The SA news data are published in real-time, with, on average, 200 news items posted per day. Descriptive statistics show that although most recommendations occur independently of a news event, the highest average abnormal performances are achieved due to a reaction to news and not from users' own analysis. The regression results demonstrate that higher skills are attainable for users with a higher proportion of recommendations made after analyzing news, more time spent on the platform, and a large community of followers, regardless of the number of recommendations made. The abnormal performance of all ST users aggregated at the stock level also increases with the proportion of recommendations made following news announcements. I show in Section 1.4.4 that my results are robust to different choices of investment horizons. In the last part of the paper, by comparing the recommendations of SA authors to the later recommendations of financial analysts, I argue that SA authors can be considered qualified investors while having the most similar publication frequency to that of social networks.

Although numerous papers examine the informational value of recommendations shared in investment-related online communities (Das and Chen (2007), Bollen et al. (2011), Skuza and Romanowski (2015) and Mittal and Goel (2012)), this paper contributes to the literature by showing how skills vary among users of financial social media and by focusing on the determinants of their skills. To the best of my knowledge, this study is the first to model the skills of financial social network users using mixture distributions, which are appropriate in their case precisely because of the heterogeneity in user groups (Crane and Crotty (2020), Chen et al. (2017)). In addition, this study compares the value of free investment advice with fee-based advice, which have a similar frequency of issues. The unexpectedness of the results, although not contradicting past research finding predictive power in SA data (Chen et al. (2014)), is a compelling incentive to investigate financial social media users in greater depth. Another important contribution of this study is beyond the estimation of social media users' skills by studying the determinants of the users' skills.

The remainder of the paper is structured as follows. First, Section 2 discusses the corresponding literature review. Section 3 describes the data. Section 4 presents the analysis. Finally, I conclude in Section 5.

1.2 Literature Review

With the growth of financial social networks, the value of the information available on these platforms remains unclear. While some studies document the predictive power of social media users, others find no informational value.

Consistent with this paper's findings, Chen et al. (2014) use SA data to investigate the extent to which investors' opinions disclosed on social media predict future stock returns. They find that opinions transmitted by investors through this platform predict future stock returns and earnings surprises. SA authors are paid to make recommendations and share valuable investment ideas. According to Crawford et al. (2017), skilled investors have incentives to share profitable ideas to correct mispricing by additional investors.

Using free recommendation websites, Bollen et al. (2011), Skuza and Romanowski (2015) and Mittal and Goel (2012) find that public mood on Twitter can predict next-day up-and-down changes in the Dow Jones Industrial Average Index. These results support the use of short time horizon analysis to assess the performance of social media users. Avery et al. (2016) study positive and negative stock picks submitted by users to the "CAPS" website. They find that negative picks strongly predict future stock price declines. By focusing on the predictive power of users at the aggregate level (i.e., the crowd skill in this paper), the findings of this paper are concordant with Das and Chen (2007) who find that time-series and cross-sectional aggregation of sentiment are better at predicting stock returns. With ST data, Oh and Sheng (2011) and Rao et al. (2012) show that stock sentiment expressed on the platform has predictive power for future market direction. Several research papers have reached opposite conclusions. Dewally (2003) study stock recommendations distributed through major newsgroups and find that the recommendations have no informational value. Moreover, they document that investors on the Internet mostly follow a momentum strategy by recommending stocks based on past performance. Using a larger dataset, Kim and Kim (2014) and Oliveira et al. (2013) find no evidence that investor sentiment can predict future stock returns, volatility, or trading volume. Although these results are not aligned with those of this paper, social networks were not equally popular, and the number of recommendations before 2014 is insufficient to estimate the predictive power of social media platforms Goutte (2022).

The difference between the above studies and this paper is the estimation of the individual performance of social media users in terms of skills and the investigation of the sources of users' skills. The methodology of this study relies on Crane and Crotty (2020) and Chen et al. (2017). Crane and Crotty (2020) model analysts' skills using a mixture of multiple skill distributions. I follow their approach to estimate the users' skills and the proportion of skilled versus unskilled users. The large number of users on financial networks naturally results in a wide dispersion of their skills, as these platforms cater to a diverse audience, ranging from finance professionals seeking advanced market insights to casual investors seeking news and updates on the markets. This group structure in the data motivates the use of a mixture distribution to model users' skills. Furthermore, this paper relates the source of analyst performance by differentiating between performance following news events and performance due to the production of additional analysis on a stock. The authors find that almost all analysts are skilled and that those skills come from both news analysis and the production of new information. Performance resulting from news coverage benefits my study because financial social media users are very likely to follow the news on the stocks they recommend.
Following the same methodology, Chen et al. (2017) use the mixture distribution of multiple skill groups and the EM algorithm to evaluate the performance of hedge funds. The skills measure is based on hedge funds' estimated alphas, and they use the BIC to find the number of skill groups that best fit the data.

One possible explanation for the predictive power of the social media users documented in this paper is that they are herding. According to Ammann and Schaub (2021), who study a social trading platform, unsophisticated investors are likely to trade after posting a comment about a shared portfolio. These findings, coupled with herding behavior, could potentially create spikes in volume, explaining the positive abnormal performance of the crowd after a recommendation. Additionally, Sprenger et al. (2014) find a relationship between message volume and trade volume, and according to Bandara (2016), an increase in the number of messages can significantly strengthen the relationship between sentiment and return, leading to a doubling of the explanatory power of social media. In line with these results, Section 1.4.3 provides evidence of high abnormal trading activity around users' recommendations. However, after removing the effect of news, the abnormal trading volume is significantly reduced. Section 1.4.3 delves into alternative sources of user skills.

1.3 Data

1.3.1 Crowd Data: StockTwits

Founded in 2008, StockTwits is the most extensive social network for investors and traders, with over one million registered community members and approximately three million monthly visitors. The platform enables users to express recommendations in real-time. In the dataset, the following variables are associated with each tweet: the content of the message, the ticker of the stock related to the message, the user's name, the time of the post (with split-second accuracy), the number of users liking or sharing the tweet, the number of people following the user, the number of tweets already posted by the user, the date from which the user has been active on the platform, a logical variable called 'official account' (if the account is related to an official media figure or a guru, i.e., an expert or authority figure in the field who is widely recognized and respected for their knowledge and advice) and a logical variable called 'market leader.'
All users can post a message of a maximum of 140 characters about a particular stock. Since 2012, users have also been able to classify their messages as bearish or bullish. Table 1.1 displays examples of self-classified tweets for both categories.
Following Renault (2020), I use these bullish/bearish labels to train the machine learning classifier in Section 1.3.3, which splits the tweets into positive and negative recommendations.

Table 1.1: Example of Tweets for each category
This table displays examples of self-classified tweets for both the bullish and bearish categories. Each tweet is associated with the stock's ticker discussed in the message. For example, for the stock of Apple Inc., each tweet is referenced with the ticker "$AAPL."

Category	**Tweets**
Bullish	$AAPL highly attractive $AAPL Fly me to The Moon......YESSSSSS $AAPL Deutsche Bank sees stocks up just 3% in 2014 http://stks.co/ayfQ $AAPL AH up $2.08 do we see $570.00 tomorrow. Very possible
Bearish	$AAPL See today's downgrade coming. What a drama. $AAPL Looks like another day in the red, more waiting $AAPL bad EOD for AAPL bad all day for AAPL bad bad bad bad bad bad bad $AAPL probably should sell

I first consider all the messages posted on StockTwits linked to S&P500 tickers (18,280,088 tweets on 504 stocks). To obtain a sufficient number of messages per day for each stock, I removed the stocks associated with fewer than 400 messages from the sample. Additionally, I analyze only messages posted after 2013 due to the low posting level at the platform's inception. The final dataset consists of 16,883,881 tweets from 501 stocks.

1.3.2 Qualified Investor Data and News Data: Seeking Alpha

The Seeking Alpha platform is a crowdsourced content service for financial markets founded in 2004. Investors and industry experts provide the content of the website. Authors are paid for exclusive articles.[5] The website attracts more than 8 million unique visitors a month. I download article headlines from four categories present on the website: 'Long ideas,' 'Short ideas,' 'Latest news,' and 'Top news.' Headlines from an article contain all the essential information, with an average length of eight words. Table 1.2 presents examples of headlines from the four categories. The description of the headline categories is the following:

- 'Long ideas': Displays articles related to stocks associated with a potential return upside. The 'Long ideas' dataset contains 86,639 headlines from January 2014 to April 2020.
- 'Short ideas': Displays articles related to stocks associated with a potential downside in returns. The 'Short ideas' dataset contains 20,729 headlines from April 2008 to November 2020. All the headlines are related to one or several tickers (approximately 6,000 unique tickers).
- 'Latest news': This category groups all recent news on companies. It contains 120,200 headlines discussing companies' or markets' daily events from November 2018 to April 2020. The headlines from this category are not associated with a ticker by the website.
- 'Top news': This category represents the articles that have received the most significant attention from Seeking Alpha subscribers. The 'Top news' dataset contains 22,509 headlines, of which 55% relate to the stock market in general and 45% relate to a particular ticker. I collect 'Top news' headlines from April 2011 to April 2020.

For each headline, I download the article posting date, the author's name, the Ticker if the article refers to a particular stock rather than the overall market, and the number of comments related to the article. The Seeking Alpha database allows for evaluating qualified investors' skills and features a database of financial news.
I use both the 'Long ideas' and 'Short ideas' headlines categories to evaluate the skills of qualified investors. This classification on the site allows us to directly extract positive or negative recommendations made by investors without analyzing the article's content. Figure 1.1 shows an example of the content of a Seeking Alpha article that belongs to the 'Long Ideas' category.

The 'Top news' and 'Latest news' headlines categories represent the financial news data. News data can be used to differentiate the sources of users' skills: Are users' skills in predicting the

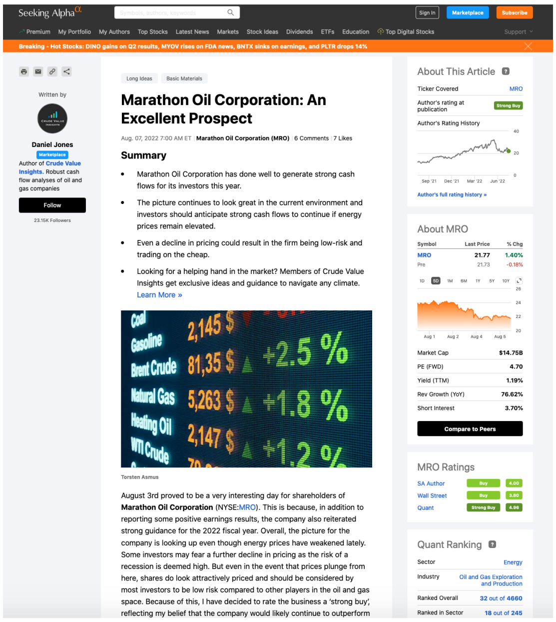
[5] Authors can be paid up to $1000 for an exclusive article. Further details about the article's submission guidelines can be found on the Seeking Alpha website: https://seekingalpha.com/page/article-submission-guidelines.

performance of actions due to their own analysis or to the interpretation of news (information processing versus information creation)?

Table 1.2: Example of Headlines for each category

This table presents examples of headlines from four categories present on the Seeking Alpha website: 'Long ideas,' 'Short ideas,' 'Latest news,' and 'Top news.' Headlines are, on average, eight words summary of the article content.

Category	**Headlines**
Long ideas	Nvidia: Time To Raise Estimates Alibaba Should Be As Big As Amazon Exxon Mobil: Prudent 2020 Management Decisions Should Help Performance Continental Resources: Still Going Strong
Short ideas	There's A Dark Cloud On Dropbox's Horizon Tesla: Where Battery Day Went Wrong Apple: Most Expensive Valuation In A Decade Netflix - Debt Increasing, Subscriber Growth Decelerating, And Competition Intensifying
Latest news	Amazon: Cyber Monday breaks records Google employees continue China protest WesBanco declares $0.29 dividend Hilton Grand Vacations adds $200M in buyback firepower
Top news	JPMorgan gives up on this major planemaker Nike +7% after earnings topper Fed may fire another bullet tonight - BofA Eurozone bond yields dive after ECB move



(a) Article Content Summary

(b) Part of the Article Content

Figure 1.1: Screenshot of SA 'Long Ideas' Article Content

This figure shows a screenshot of the content of one of the SA 'Long Ideas' articles. This article concerns the ticker MRO. The first subfigure shows the summary of the article, and the second subfigure shows part of the article's content that indicates a portion of the author's analysis.

1.3.3 Text Preprocessing and Text Classification

A message on StockTwits can easily be considered a recommendation from a platform user to the community.

Before the classification of tweets, I perform different text preprocessing tasks. First, I put the text in lower cases, removed HTML codes, and removed some punctuations and stopwords. Following Renault (2020), I keep several punctuation marks ! ? % + ' = : ;) (. and several key stopwords used in finance to increase the accuracy of the classifier.[6] According to Renault (2020), POS tagging and Stemming reduce the accuracy of a sentiment classifier. Tickers ($AAPL) are replaced by the word cashtag, users mentioned in tweets (@user) by the word usertag, and links by the word linktag. Words replace smileys and emoticons using the emoji package in python.[7]

One of the key functionalities of StockTwits is that users can classify their message as bearish or bullish when posting their tweets. The 'self-classified' tweets represent 30% of the number of messages posted.

To train the machine learning classifier, I use an unbalanced dataset of 5,680,911 tweets with 77% positive tweets.[8] The unequal proportion of long and short headlines is consistent with the past decade's market movement. Moreover, Antweiler and Frank (2004) and Dewally (2003) report that the majority of investors on social media have a bullish view of the market. After removing tweets with fewer than three words from the dataset, I obtained 4,948,721 tweets. I divide the dataset into an 80% training set and a 20% test set.

I use various algorithms for text classification: naive Bayes (NB), support vector machines (SVMs), and the maximum entropy classifier (MaxEnt). To obtain optimal performance, I optimize parameters with 'GridSearchCV'.[9] Table 1.3 displays the performance of the different classifiers on the test set. The highest accuracy (86.31%) is obtained using the maximum entropy algorithm with a bigram coupled with TF-IDF feature values.

After attributing a recommendation score to each message with the selected classifier (+1 for a positive and -1 for a negative recommendation), I obtain an intermediate dataset of 240,258 users.

As the overall analysis of the paper is about individuals' skills, it is necessary to ensure that only the recommendations of the users who have posted enough messages are considered. On average, users post 108 recommendations (approximately 10 per stock). At the stock level, the average number of messages posted is 32,234. Deleting messages from users that post fewer than

[6]Indeed, NLTK Stopwords corpus includes words like 'up,' 'down,' 'below,' and 'above,' which give essential meaning to a sentence in finance.

[7]For example, the smiley with the face with tears of joy is replaced by the word face_with_tears_of_joy

[8]According to Renault (2020), an unbalanced dataset provides greater prediction accuracy than a balanced dataset with the same number of tweets.

[9]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

Table 1.3: Performance of classification models

This table presents precision, recall, and accuracy measures for various machine learning methods. The dataset contains 4,948,271 messages (77% positive and 33% negative) with a minimum of three words per message. The dataset is divided into 80% training set and 20% test set.

Model	Precision	Recall	Accuracy
NB	0.82	0.57	0.79
MaxEnt	0.83	0.76	0.86
SVM	0.81	0.51	0.77

two messages on each stock narrows the number of users to 148,606 for a total of 16,149,692 recommendations.

1.4 Analysis

1.4.1 Estimation of Skills

Estimation of Skills across users: StockTwits

In this section, I present the estimation of average skill measures at the individual level.

In line with research on analysts' recommendations (Crane and Crotty (2020), Jegadeesh and Kim (2010), Boni and Womack (2006)), I use the change in recommendation rather than the recommendation level as it is more informative. Positive recommendations (+1) correspond to positive tweets, and negative recommendations (-1) correspond to negative tweets. For user i, recommendation event j and stock s, the change variable $\Delta^s_{i,j}$ equals +1 for a positive change (upgrade event) and -1 for a negative change (downgrade event). For the rest of the paper, I only consider users that issue at least two recommendations and one change in recommendation for a given stock, leading to 2,389,745 recommendation events. In total, there are 1,190,006 upgrade events and 1,199,738 downgrade events.

To convert a change in recommendation measure to a skill measure, I use the buy and hold abnormal returns associated with buying stocks following an upgrade and selling stock following a downgrade over a three-day horizon:

$$BHAR^s_{i,j} = \Delta^s_{i,j}(\prod_{t=1}^{4}(1 + r_{s,t}) - \prod_{t=1}^{4}(1 + R_{s,t})). \tag{1.1}$$

With $R_{s,t}$ the expected return on stock s estimated with the market model:
$r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event. $r_{s,t}$ is the return of stock s on day t, i the user and, j the change in recommendation event. The market factor is the daily value-weighted return of all NYSE, AMEX, and NASDAQ firms from the Fama/French website.

In Section 1.4.4, I provide justification for using a three-day event window and also examine alternative investment horizons ranging from one to five days. Certain contributors may 'piggy-back' on market performance (i.e., they issue positive recommendations when the market goes up and vice versa). To mitigate this concern, I exclude the date on which the recommendation was issued (i.e., day 0). Some recommendations take place before and after trading hours. For a recommendation that occurs after trading hours, the day 0 return is the return on the following day. For a recommendation prior to trading hours, the day 0 return is the return on the day of the recommendation.

To estimate users' skills denoted by $\hat{\alpha_i}$, I aggregate the buy and hold abnormal returns to the user level using the average buy and hold abnormal return:

$$\hat{\alpha_i} = ABHAR_i = \frac{1}{n_i} \sum_{j=1}^{n_i} BHAR^s_{i,j}, \tag{1.2}$$

with n_i being the number of changes in recommendation j made by user i.

Estimation of Skills across Qualified Authors: Seeking Alpha

To analyze the skills of social media users compared to finance professionals, I use Seeking Alpha (SA) long and short recommendation data. The size of the headlines is close to the size of a tweet, and the posting frequency of the platform is the closest to that of a social media platform. Therefore, the recommendations on this website are the most appropriate to compare to social media data, the main difference being the characteristics of the recommendation issuers. All types of investors can post tweets, whereas SA recommendations are issued only by authors providing professional financial analysis. The definition of a change in recommendation events is different for SA authors. The platform directly gives categories to the articles posted by the authors. Thus, a positive change in a recommendation event corresponds to a 'Long Idea' posting article, whereas a negative recommendation event corresponds to a 'Short Idea' posting article. For SA authors, $\Delta^s_{i,j}$ equals +1 and -1 if the analysis is published in the Long Ideas and Short Ideas categories, respectively. Removing the Long and Short Ideas unrelated to S&P500 tickers leads to 2,809 short recommendations on 315 tickers from 828 authors and 32,126 long recommendations on 496 tickers from 3,421 authors. The overall SA recommendation data contain 34,946 recommendations from 498 tickers written by 3,659 authors. On average, an SA author issues ten recommendations (with a maximum of 564 recommendations issued for the same author). To obtain an adequate number of recommendation changes, I consider 493 stocks for ST and 481 stocks for SA.

I employ two approaches to identify positive and negative recommendations on the two platforms. To identify recommendations on ST, I use user classifications and machine learning. I rely on the authors' classifications to identify recommendations on Seeking Alpha. To ensure that differences in the approaches used to identify recommendations are not explaining differences in the results across the two platforms, I use the same machine learning algorithm to classify Seeking Alpha authors' recommendations. The machine learning and author classification align in 92% of cases, mitigating the above concern. Furthermore, the difference in approaches does not affect the differences in the results across platforms.

Estimation of Skills across Stocks: Skill of the Crowd

It is reasonable to assume that the aggregate opinion of the population on social networks or of all SA authors is a better predictor of market movements than the opinion of a single individual. One may wonder if the skills of individual users on the platform are not exceeded by the skills of all users combined.

To evaluate the skill of the crowd for each stock, the abnormal returns related to each recommendation are aggregated at the stock level.

To estimate stock skills denoted $\hat{\alpha_s}$, I calculate average abnormal returns at the stock level:

$$\hat{\alpha_s} = ABHAR_s = \frac{1}{n_s} \sum_{j=1}^{n_s} BHAR_{i,j}^{s}, \tag{1.3}$$

with n_s being the number of changes in recommendation j made for stock s.

1.4.2 Descriptive Statistics

Table 1.4 presents descriptive statistics of average abnormal returns across ST users and SA authors and descriptive statistics of average abnormal returns across S&P500 stocks. ST accounts for 62,346 users and 493 stocks. SA accounts for 2,099 authors and 481 stocks. For ST, at the user level, 55.94% of the $ABHAR_i$ are positive. At the stock level, 92.09% of the $ABHAR_s$ are positive. These values indicate that aggregated recommendations from a vast number of users at the stock level on platforms with many users result in positive abnormal performance in nearly all cases. In line with Das and Chen (2007), cross-sectional aggregation of recommendations has greater predictive power. For SA, values at the author level are similar to those of the ST users, with 49.50% of $ABHAR_i$ being positive. Nevertheless, the results at the stock level are lower than those of ST, with 54.98% of $ABHAR_s$ being positive. As expected, the effect of the crowd's wisdom is less pronounced for SA because of the lower number of authors and recommendations for each stock.

The average $ABHAR_i$ for both platforms is positive. Positive average users' skills are consistent with the findings in Chen et al. (2014), that show that recommendations on Seeking Alpha predict future stock returns as well as Oh and Sheng (2011) and Rao et al. (2012) documenting the performance of StockTwits users to predict future market direction. The average, $ABHAR_i$, is more than ten times greater for ST than for SA (0.12% and 0.01%, respectively). Those different average skills favor the wisdom of the crowd's hypothesis. Alternatively, in line with Ammann and Schaub (2021), a large number of unsophisticated investors on the ST platform coupled with a large-group effect may react simultaneously and trade. The skills of SA authors are less scattered than those of ST users: the standard deviations are 0.014 and 0.015, respectively.

The aggregated results at the stock level are again higher for ST, and the cross-sectional variation in $ABHAR_s$ is lower (0.002 and 0.010, respectively). These values are again consistent with the wisdom of the crowd hypothesis.

The cross-sectional 90th percentile $ABHAR_i$ for ST is ten times higher than the cross-sectional average $ABHAR_i$ by 1.50%. It is worth noting that this performance still involves over 6,000 users of the platform.

Cross-sectional statistical distributions of ST users' average abnormal returns are positively skewed. The p values resulting from the Kolmogorov Smirnov test against the normal distribution of average abnormal returns are in line with the use of a mixture distribution to model the skills of ST users and SA authors in Section 1.4.3.

Table 1.4: Descriptive Statistics: Users/Authors Skills ($ABHAR_i$) versus Crowd Skill ($ABHAR_s$)

This table reports descriptive statistics of average buy and hold abnormal return $ABHAR_i$ and $ABHAR_s$ following ST users' and SA authors' recommendations. There are 2,389,745 and 34,946 recommendation events from 2014 to 2020 for ST and SA, respectively. Buy and hold abnormal returns are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The expected return for each event is estimated with the market mode with a three-day event window: $r_{f,t}+\hat{\beta}_s(r_{M,t}-r_{f,t})$, using an estimation window of [-250, -10] days before the event. $ABHAR_s$ represents average buy and hold abnormal returns of stock s over n_s changes in recommendation. $ABHAR_i$ represents average buy and hold abnormal returns of user/author i over n_i changes in recommendation. n represents the number of users/authors and stocks, respectively. $\bar{n}_i$ accounts for the average number of recommendation events made by all users/authors. Fraction positive represents the fraction of positive $ABHAR_i$ and $ABHAR_s$. The K-S test $p-values$ are the p values resulting from a Kolmogorov-Smirnov test against normal distribution.

	StockTwits		**Seeking Alpha**	
	User Level	**Stock Level**	**Author Level**	**Stock Level**
n	62,346	493	2,099	481
$\bar{n}_i/\bar{n}_s$	38.06	4,850.79	15.59	71.03
Fraction positive	55.94%	92.09%	49.50%	54.98%
Mean	0.12%	0.17%	0.01%	0.09%
Std	0.015	0.002	0.014	0.008
Median	0.05%	0.15%	-0.01%	0.07%
Min	-36.28%	-2.33%	-10.83%	-2.68%
Max	22.33%	0.77%	10.77%	6.60%
Skewness	0.23	0.59	-0.28	1.69
q10%	-1.23%	0.00%	-1.34%	-0.77%
q20%	-0.57%	0.05%	-0.78%	-0.48%
q80%	0.78%	0.27%	0.82%	0.53%
q90%	1.50%	0.36%	1.47%	0.92%
K-S test p-value	0.00		0.00	

Table 1.5: Descriptive Statistics: Users/Authors Skills ($ABHAR_i$) versus Crowd Skill ($ABHAR_s$) Without news

This table reports descriptive statistics of average buy and hold abnormal return $ABHAR_i$ and $ABHAR_s$ following ST users' and SA authors' recommendations made without any news on the recommendation day. There are 523,060 and 31,004 recommendation events without concurrent news from 2014 to 2020 for ST and SA, respectively. Buy and hold abnormal returns are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The expected return for each event is estimated with the market model with a three-day event window: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event. $ABHAR_s$ represents average buy and hold abnormal returns of stock s over n_s changes in recommendation. $ABHAR_i$ represents average buy and hold abnormal returns of user/author i over n_i changes in recommendation. n represents the number of users/authors and stocks, respectively. $\bar{n}_i$ accounts for the average number of recommendation events made by all users/authors. Fraction positive represents the fraction of positive $ABHAR_i$ and $ABHAR_s$. The K-S test $p-values$ are the p values resulting from a Kolmogorov-Smirnov test against normal distribution.

	StockTwits		**Seeking Alpha**	
	User Level	**Stock Level**	**Author Level**	**Stock Level**
n	21,597	364	1967	362
$\bar{n}_i/\bar{n}_s$	23.84	1436.98	15.01	85.40
Fraction positive	55.88%	75.00%	49.77%	54.82%
Mean	0.07%	0.14%	0.01%	0.04%
Std	0.028	0.003	0.014	0.007
Median	0.00%	0.12%	0.00%	0.04%
Min	-57.74%	-0.59%	-10.83%	-2.61%
Max	35.51%	1.09%	13.91%	3.77%
Skewness	-0.07	0.46	-0.13	0.47
q10%	-1.98%	-0.14%	-1.33%	-0.72%
q20%	-0.08%	-0.04%	-0.81%	-0.44%
q80%	0.99%	0.33%	0.81%	0.48%
q90%	2.07%	0.44%	1.46%	0.81%

The statistics from Table 1.4 represent the performance of all SA authors and ST users in the dataset. Given the heterogeneity in users' performance, it is interesting to study the evolution of these statistics through the number of recommendations issued. Figure 1.2 plots the evolution of cross-sectional statistics of average buy and hold abnormal return $ABHAR_i$ following ST users' and SA authors' recommendations as a function of the minimum number of recommendations issued. Subfigures 1.2a and 1.2b focus on the cross-sectional average and the fraction of positive $ABHAR_i$ for ST and SA, respectively. Two observations can be made. For both platforms, the performance increases if we consider the groups with a higher number of recommendations. These results support Wang et al. (2015) who find that StockTwits and Seeking Alpha users with high levels of interaction on those platforms have greater predictive power. Additionally, Bandara (2016) highlights that a greater number of messages doubles the explanatory power of social media.

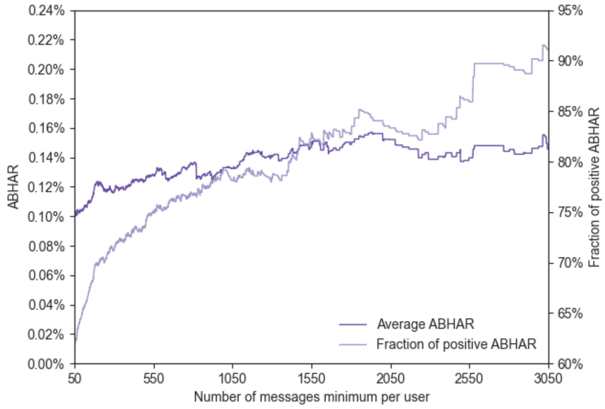
Furthermore, it has been observed that once a certain number of recommendations have been posted on ST, the users' performance converges to an average of approximately 0.14% and 92% positive $ABHAR_i$. For the SA authors, although the performance improves, the small number of authors remaining after 200 recommendations issued does not allow us to identify a clear trend. The evolution of the cross-sectional quantiles is shown by the subgraphs 1.2c and 1.2d. ST users in the 5th and 10th percentiles of performance tend to see improvement in their performance as they issue more recommendations. On the contrary, the best performing groups (90% and 95% percentiles) remain well above average, regardless of the number of recommendations issued. The cross-sectional average is superior to the cross-sectional median at all times, indicating that above-median skilled users drive average performance. For SA authors, the performances of the worst and best performer groups converge to the average performance as the minimum number of recommendations increases.

Subgraphs 1.2e and 1.2f show the decrease of the cross-sectional standard deviation as a function of the minimum number of recommendations issued for both platforms. The heterogeneity of users'/authors' performances is reduced with the number of recommendations posted. From subgraph 1.2g, one can observe that the increase in the number of recommendations issued translates into an increase in the skewness and the number of followers. When considering user groups with more than 1000 recommendations, skewness is significantly reduced. This reduction is due to the decrease in the size of the user groups.

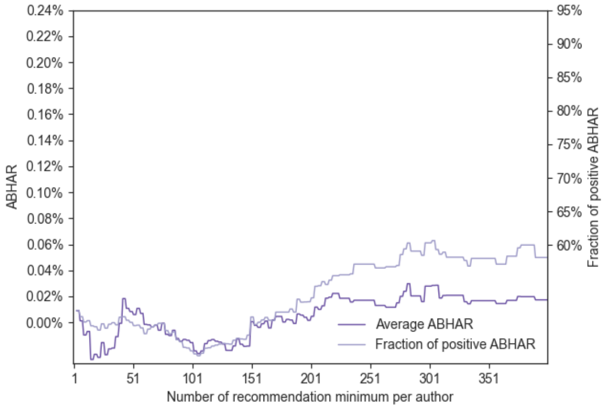
All the figures and cross-sectional statistics in this section indicate that, on average, individual skills are higher on the ST platform. These statistics increase when considering groups of users who are more active on the platform. At the same time, the variability of the users' skills decreases.

Section 1.4.3 presents the estimation of the fraction of users/authors stemming from different skill groups using mixture distributions.

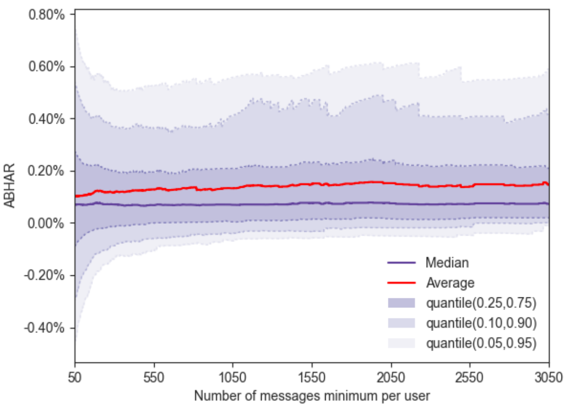
The performance documented above may be attributed to the market reacting to news events. While this effect is mitigated by removing the day of the recommendation from the event window, the impact of company news on returns can persist and explain the positive abnormal performance of social media contributors. The statistics from Table 1.5 represent the performance of all SA authors and ST users when making recommendations without any news on the same day. I use SA's 'Top news' and 'Latest news' headlines as news data, described in Section 1.4.3. There are 523,060 and 31,004 recommendation events without concurrent news from 2014 to 2020 for ST and SA, respectively. The proportion of recommendations made the same day as news is higher for ST users (78%) than for SA authors (11%). The average performance of ST users decreases once the news effect is taken into account, while the performance of SA authors remains steady. Nevertheless, the skill differences between the two platforms do not change, and ST users still outperform SA authors at the user and stock levels.



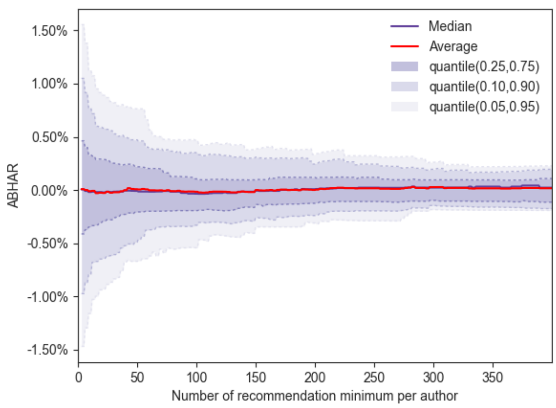
(a) ST Users Skills



(b) SA Authors Skills



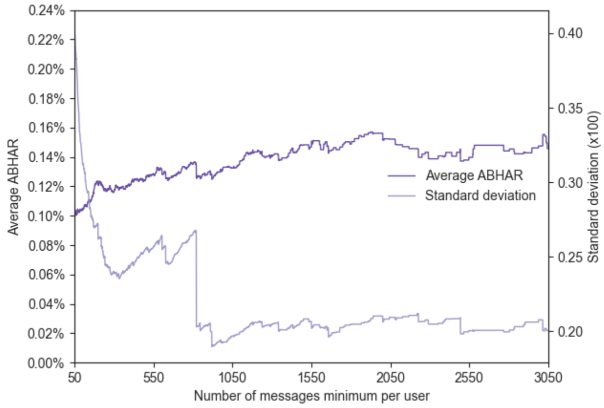
(c) ST Users Skills: Quantiles



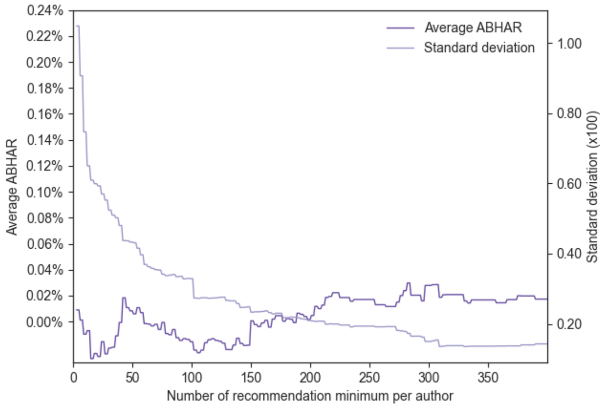
(d) SA Authors Skills: Quantiles

Figure 1.2: Evolution of ST Users and SA Authors Skills With Minimum Number of Recommendations Issued

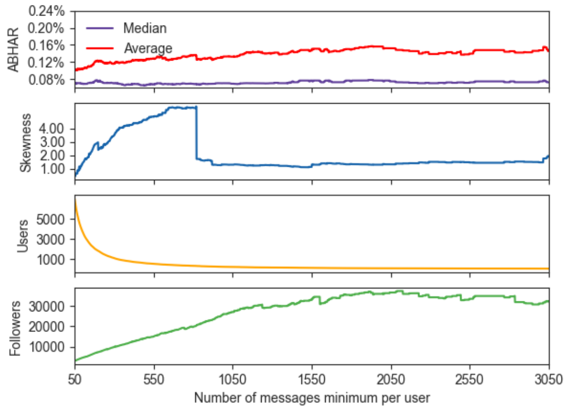
These figures plot the evolution of cross-sectional statistics of average buy and hold abnormal return $ABHAR_i$ following ST users' and SA authors' recommendations as a function of the minimum number of recommendations issued. There were 2,389,745 and 34,946 recommendation events from 2014 to 2020 for ST and SA, respectively. Buy and hold abnormalities are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The expected return for each event is estimated with the market model: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event. Followers represent the average number of followers of the ST user group.



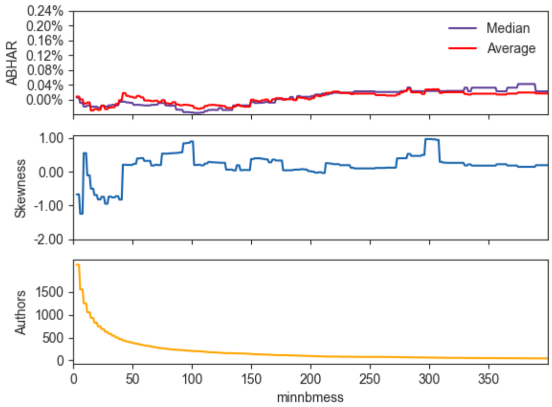
(e) ST Users Skills: Standard Deviation



(f) SA Authors Skills: Standard Deviation



(g) ST Users Skills: Skewness, Number of Users and Followers



(h) SA Authors Skills: Skewness and Number of Authors

Figure 1.2: Evolution of ST Users and SA Authors Skills With Minimum Number of Recommendations Issued

These figures plot the evolution of cross-sectional statistics of average buy and hold abnormal return $ABHAR_i$ following ST users' and SA authors' recommendations as a function of the minimum number of recommendations issued. There were 2,389,745 and 34,946 recommendation events from 2014 to 2020 for ST and SA, respectively. Buy and hold abnormalities are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The expected return for each event is estimated with the market model: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event. Followers represent the average number of followers of the ST user group.

1.4.3 Modelling Skills

Distribution of Users Skills

In this subsection, I follow the Crane and Crotty (2020) and Chen et al. (2017) mixture modeling approaches to estimate the fraction of ST users and SA authors belonging to a skill group with a specific skill level.

As described in the previous section, each user or author is assigned a skill measure α_i based on their average buy-and-hold abnormal return following a change in recommendation. These sections focus on the proportion of skilled vs. unskilled users/authors. Following Crane and Crotty (2020) and Chen et al. (2017), I model the performance of StockTwits users as a mixture distribution of J skill groups. Mixture distributions are appropriate when there is a group structure in the data (McLachlan et al. (2019)), which is the case for social media users, as seen with the descriptive statistics. This method implies that each user's skill α_i is part of a skill group $j \in \{1, J\}$, with J being the number of groups. The skills of each group follow a normal distribution, with mean skill μ_j and standard deviation of skills σ_j representing the dispersion in skills within the group. The true skill of user/author i that belongs to group j is defined as $\alpha_i = \mu_j + \omega_i$, with ω_i driving dispersion in user skills, a normally distributed random variable with mean 0 and standard deviation σ_j. The abnormal buy and hold returns are estimated with error $\epsilon_i \sim N(0, s_i)$. The estimated skill of user/author i is defined as $\hat{\alpha_i} = \mu_j + \omega_i + \epsilon_i$.

The proportion of users associated with skill group j is defined as λ_j, with $\sum_{j=1}^{J} \lambda_j = 1$. The cross-sectional distribution of user/author skills is a mixture distribution of J distributions, and its density function is the following:

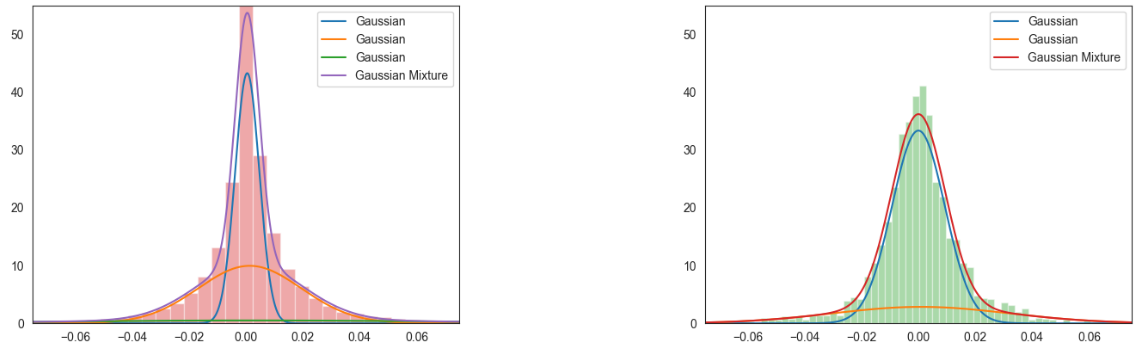
$$f(\alpha_i) = \sum_{j=1}^{J} \lambda_j \phi(\alpha_i; \mu_j, \sigma_j), \tag{1.4}$$

where α_i is the skill of user i and $\phi(\alpha_i; \mu_j, \sigma_j)$ is the normal probability density function for group j evaluated at α_i.

To simultaneously estimate the set of parameters $\{\lambda_j, \mu_j, \sigma_j\}$, I use expectation-maximization (EM) algorithm. EM is appropriate to estimate parameters from multiple skill groups simultaneously.

To find the number of skill groups in both ST and SA data, I estimate the model for $J = 2, ..., 5$ and select the model with the lowest Bayesian information criterion (BIC). For ST data, the BIC values of groups 2, 3, and 4 are -339042, -346139, and -345182, respectively. For SA data, the BIC of 2, 3, and 4 groups are -12471, -12449, and -12450, respectively, and continue to increase thereafter.

The BIC suggests a mixture of three skill groups for StockTwits users: Very High, High, and Neutral. For Seeking Alpha authors, BIC suggests a mixture of two skill groups: High and Low. Figures 1.3a and 1.3b plot the estimated mixture distribution for ST user skills and SA author skills, respectively.



(a) Estimated Mixture Distribution of ST User Skills

(b) Estimated Mixture Distributions of SA Author Skills

Figure 1.3: Estimated Mixture Distributions of ST User Skills and SA Author Skills
These figures plot the estimated mixture distribution for ST user skills and SA author skills with the EM algorithm. Skills correspond to the average buy and hold return $ABHAR_i$ following multiple recommendations made by user/author i. Buy-and-hold abnormal returns are estimated using the market model with a three-day event window. The sample period is from January 2014 to December 2020. The SA database includes 2,099 users, and the ST database contains 62,346 users. The BIC criterion suggests a mixture distribution of four normal distributions for ST and two normal distributions for SA.

Table 1.6 presents the parameter estimates for ST users' skill distribution and SA authors' skill distribution. For ST, the three distribution locations lie in the positive area. Taking all users into account, 46% of users belong to the neutral-type distribution, centered at 0.047%. The remaining 54% of users belong to either the high or very high distribution group, centered at 0.14% and 0.30% abnormal returns, respectively. The dispersion of the neutral-type distribution is 43 bps, reflecting that some users may have true negative skills. Keeping in mind that the average buy and hold abnormal returns result from a calculation over a three-day horizon, these results are economically significant. Although the standard deviations range from 180 bps to 680 bps, all skill groups have positive average skills. This result indicates that on the ST platform, most users have positive skills.

For SA authors, the results are less favorable. The great majority of authors (81%) come from a neutral-type distribution centered at 0.02% abnormal returns. The standard deviation of this distribution is 84 bps, indicating that some of the authors that belong to the low-type distribution still exhibit true positive ability. The proportion of SA authors coming from the low-type distribution is 19% ($\hat{\mu} = 0.001\%$).

In summary, the fraction of SA authors with positive skills from the estimation is below that of ST users with positive skills. These results align with the cross-sectional statistics from Section 1.4.2. Therefore, the remaining part of the paper will focus on the sources of the skills and the performance of ST users.

Table 1.7 presents the parameter estimates for ST users' skill distribution and SA authors' skill distribution without news. While the average of the different skill groups decreases for both platforms, the overall interpretation of the results remains the same for ST. The great majority of SA authors (79%) come from a low-type distribution centered at 0.00% abnormal returns. For ST, removing the effect of news events harms the social media users' performance while not preventing the majority of the users from having positive skills.

Table 1.6: Parameter Estimates for Different Skill Group Distributions

This table presents the parameter estimates for the different skill groups in the ST user sample and the SA author sample. The BIC criterion suggests a mixture of three skill groups for ST (Very High, High, and Neutral) and two skill groups for SA (Neutral and Low). Skills correspond to the average buy and hold returns $ABHAR_i$ following multiple recommendations made by user/author i. Buy-and-hold abnormal returns are estimated using the market model with a three-day event window. The sample period is from January 2014 to December 2020. The SA database includes 2,099 users, and the ST database contains 62,346 users. $\hat{\lambda}_j$ denotes the fraction of users belonging to skill group j, $\hat{\mu}_j$ accounts for the average skill of group j and, $\hat{\sigma}_j$is the dispersion in skills within group j.

StockTwits Users	$\hat{\lambda}_j$	$\hat{\mu}_j$	$\hat{\sigma}_j$
Very High	9%	0.300%	0.068
High	45%	0.140%	0.018
Neutral	46%	0.047%	0.004
Seeking Alpha Authors	$\hat{\lambda}_j$	$\hat{\mu}_j$	$\hat{\sigma}_j$
Neutral	81%	0.020%	0.008
Low	19%	0.001%	0.028

Sources of Skills: User Skills

The preceding analysis focuses on whether ST users show skills in predicting stock returns and whether these skills surpass those of qualified authors who issue fee-based recommendations. I found that from the perspective of cross-sectional statistics as well as from the perspective of mixture distribution estimation results, the ST users outperform the SA authors. In addition, when we aggregate the buy and hold returns following a change in recommendation of all ST users at the stock level (Crowd Skill), the economic value of the average abnormal returns increases significantly. As a result, the following section analyses the sources of skills for individual ST

Table 1.7: Parameter Estimates for Different Skill Group Distributions: Without News

This table presents the parameter estimates for the different skill groups in the ST user sample and the SA author sample. Skills are derived from users' recommendations on days without any company news. The BIC criterion suggests a mixture of three skill groups for ST (Very High, High, and Neutral) and two skill groups for SA (Neutral and Low). Skills correspond to the average buy and hold returns $ABHAR_i$ following multiple recommendations made by user/author i. Buy-and-hold abnormal returns are estimated using the market model with a three-day event window. The sample period is from January 2014 to December 2020. The SA database includes 1,967 users, and the ST database contains 21,597 users. $\hat{\lambda}_j$ denotes the fraction of users belonging to skill group j, $\hat{\mu}_j$ accounts for the average skill of group j and, $\hat{\sigma}_j$is the dispersion in skills within group j.

StockTwits Users	$\hat{\lambda}_j$	$\hat{\mu}_j$	$\hat{\sigma}_j$
Very High	8%	0.340%	0.086
High	49%	0.045%	0.005
Neutral	43%	0.041%	0.021
Seeking Alpha Authors	$\hat{\lambda}_j$	$\hat{\mu}_j$	$\hat{\sigma}_j$
Neutral	21%	0.007%	0.027
Low	79%	0.000%	0.008

users and for the overall platform.

Several questions naturally arise in assessing the sources of user skills: Do influential recommendations come from the users' reaction to news about the stock or from users' own analyses? Following Crane and Crotty (2020), the two different sources of skills stem from differentiating recommendations following news events from recommendations that do not follow news events. In other words, if the user revises his recommendation after a news release, the recommendation quality is likely due to information-processing skills. If the recommendation change occurs without any concurring news about the company, the recommendation likely comes from the users' information-providing skills. Wysocki (1998) argues that news induces users in investment-related online communities to post comments. If this is the case, the positive abnormal returns around changes in recommendations are possibly driven by users responding to the news. By distinguishing between recommendations that follow news and those that do not, it becomes possible to isolate users' ability to interpret news.

Since posting tweets occurs at a higher frequency than traditional financial advice providers, I use Seeking Alpha's 'Top news' and 'Latest news' headlines as news data. The real-time nature of Seeking Alpha news makes it a good source of news data to analyze alongside StockTwits data, as Seeking Alpha platform post over 200 news items per day.

'Top news' directly links headlines to the corresponding ticker. For the 'latest news' category, headlines are not directly linked to tickers on the website. I label the headlines to the corresponding tickers by looking for the ticker or the company name in the text. In total, the news data on S&P500 tickers contain 10,855 news items from January 2014 to May 2020 (approximately four news items a day). For news to correspond to a recommendation, the recommendation on ST must concern the same ticker later on the same day. I found that 126,411 recommendation events in the sample are made after a news is issued on the same ticker on the same day.

Table 1.8 presents cross-sectional statistics of ST users' average abnormal returns for the two types of users: users issuing recommendations following news (information processing) and users issuing recommendations without any prior news about that stock (information production). Due to the high frequency with which users post on the ST platform, it is not surprising that most recommendations do not stem from news (522,654 recommendations versus 126,411 recommendations, respectively). The results indicate that the cross-sectional average $ABHAR_i$ is four times greater for recommendations following news (0.17%) than for recommendations not following news (0.07%). The same conclusion applies to the fraction of positive $ABHAR_i$, with 59.05% of users showing positive average performance reacting to news versus 55.88% for the group of users issuing recommendations without related news. These results support the hypothesis that the skills of ST users are a result of the reaction to and processing of the news rather than from their own analysis. Subsequently, I perform regression analysis on the different sources of skills in Table 1.10.

Another question that naturally arises is whether the skills of ST users can be explained by the amount of time spent on the platform or the period in which the recommendation is issued. The popularity of the StockTwits platform increases over time, as does the user base. From the perspective of the wisdom of the crowd, one can expect the predictive power of the users on the platform to grow as the number of users grows, even if the users do not individually present better skills. Table 1.9 reports the evolution of the cross-sectional statistics of average buy and hold abnormal return $ABHAR_i$ between the period before 2018 and the period after 2018. The pool of users after 2018 is twice the number before 2018 (24,346 vs. 45,761). The number of recommendations issued per user has decreased over time, with an average of 38.35 recommendations per user before 2018 and an average of 31.26 after January 2018. In parallel, the average performance of users and the fraction of positive average buy and hold returns are lower after 2018 than before 2018: the cross-sectional average $ABHAR_i$ is 0.16% and 0.09%, respectively. These results suggest that new users of the platform may have inferior skills compared to older users with more experience.

Table 1.8: Descriptive Statistics: News vs. No News Following Events

This table reports descriptive statistics of average buy and hold abnormal return $ABHAR_i$ for the two types of users: users issuing recommendations following news (Information processing) and users issuing recommendations without any prior news about that stock (Information production). Buy and hold abnormal returns are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. Expected returns for each event are estimated with the market model with a three-day event window: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days relative to the event. $ABHAR_i$ represents average buy and hold abnormal returns of user/author i over n_i changes in recommendations. $\bar{n}_i$ accounts for the average number of recommendation events made by all users. The fraction positive represents the fraction of positive $ABHAR_i$. In total, the news data on S&P500 tickers contains 10,855 news items from January 2014 to May 2020. For this table of descriptive statistics, the recommendation events belong to two categories: "News following events" groups the recommendation events which take place the same day following a news announcement (126,411 recommendations events), and "No News following events" gather the recommendations events that take place without any recent news on this stock (522,654 recommendations events). Only the recommendations events mentioning a ticker present in the news database are accounted for in this table (364 tickers).

	News following events	**No News following events**
Number Of Users	10,649	21,564
$\bar{n}_i$	11.30	23.84
Number of Recommendations	126,411	522,654
Fraction positive	59.05%	55.88%
Mean	0.17%	0.07%
Std	0.032	0.028
Median	0.00	0.00
Min	-32.17%	-57.73%
Max	33.87%	35.51%

Table 1.9: Descriptive Statistics: StockTwits Users' Skills Over Time

This table reports descriptive statistics of average buy and hold abnormal return $ABHAR_i$ for two periods: Pre-2018 and Post-2018. Pre-2018 data is from January 2014 to December 2017, with 946,332 recommendations from 24,492 users. Post-2018 data goes from January 2018 to May 2020, with 1,445,111 recommendations from 45,761 users. Buy and hold abnormal returns are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The expected return for each event is estimated with the market model with a three-day event window: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event. $ABHAR_i$ represents the average buy and hold abnormal returns of user/author i over n_i change in recommendation. $\bar{n}_i$ accounts for the average number of recommendation events made by all users. The fraction positive represents the fraction of positive $ABHAR_i$ in the sample.

	Pre 2018	**Post 2018**
Number Of Users	24,492	45,761
$\bar{n}_i$	38.35	31.26
Number of Recommendations	946,332	1,445,111
Fraction positive	59.25%	56.70%
Mean	0.16%	0.09%
Std	0.019	0.025
Median	0.01%	0.00%
Min	-23.16%	-57.74%
Max	27.13%	39.72%

After establishing leads on the sources of skills, including news events and experience, one can infer causal links with a regression.
To identify sources of ST user skills, I use the following specification:

$$ABHAR_i = \alpha + \gamma X_i + \epsilon_i, \tag{1.5}$$

$$\text{with } X_i = (News_i, History_i, GainFollowers_i, Followers_i, Recommendation_i).$$

The variables are standardized to facilitate comparison. The estimates measure the change in the average user's skill corresponding to a one-standard-deviation change in the independent variable. The variable $News_i = \frac{1}{n_i}\sum_{i=1}^{n_i} HasNews_i$ represents the proportion of recommendations made following a news announcement by user i, with $HasNews_i$ being a binary variable equal to one if the recommendation made on the stock follows news about the same stock earlier the same day and n_i being the number of recommendation changes made by user i. $History_i$ represents the time difference between the first and last recommendation of user i, $GainFollowers_i$ reflects the number of new followers gained by user i by posting his recommendations, and $Followers_i$ is the absolute number of followers of user i. $Recommendations_i$ is the total number of recommendations issued by user i.

Table 1.10 presents the results for this specification. The first column shows a significant positive relationship between the proportion of recommendations made following news by user i and the average user skill i. The news coming before the recommendation was made rules out a possible endogeneity bias. A one-standard-deviation change in the proportion of recommendations made following news corresponds to a 2% change in user i's average buy and hold abnormal returns. By showing that a higher proportion of recommendations following news leads to greater skill, this result indicates that ST users, on average, demonstrate skills in predicting the stock price direction following news about that stock. Column two shows that an increase in the user's time on the platform corresponds to a rise in the user's skill. In line with Oh and Sheng (2011) and Rao et al. (2012), the number of new followers gained by user i by posting his recommendations has a positive and significant relationship with the user's i skill. It seems that the community on ST can identify the users to follow, i.e., those who exhibit higher skills. This result is consistent with Sprenger et al. (2014), who show that users who provide higher-quality investment advice have more influence on the social media platform. Column 5 presents the results for the specification, including all variables. All variables discussed above, except the News variable, still have a positive and significant relationship with the average skill of a user. To detect multicollinearity in the regression analysis, I calculate the variance inflation factors for all predictors. The results indicate little or no correlation between the variables, with Variance Inflation Factor (VIF) estimates, i.e., a measure of the amount of multicollinearity in regression analysis, ranging between

1 and 1.14.

Table 1.10: Sources of Skills: All Users

The table presents parameter estimates from OLS regression on $ABHAR_i$, the average skill of StockTwits user i with several different specifications. Skills correspond to the average buy and hold return $ABHAR_i$ following multiple recommendations made by user i. Buy-and-hold abnormal returns are estimated using the market model with a three-day event window.

Each variable is standardized; reported estimates measure the change in $ABHAR_i$ corresponding to a one-standard-deviation change in X. The sample consists of all 10,556 users posting on the platform from 02/01/2014 to 13/01/2021. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. The variable $News_i = \frac{1}{n_i}\sum_{j=1}^{n_i} HasNews_{i,j}$, represents the proportion of recommendations j made following news by the user i, with $HasNews_{i,j}$ a binary variable equal to one if recommendation j made on the stock follows a news item about the same stock earlier the same day and, n_i the number of recommendation changes made by user i. $History_i$ represents the time difference between user i's first and last recommendation, $GainFollowers_i$ reflects the number of new followers gained by user i by posting his recommendations and, $Followers_i$ is the absolute number of followers of user i. $Recommendations_i$ is the total number of recommendations issued by user i. The last column represents the variance inflation factors related to each explanatory variable from the last specification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) $ABHAR_i$	(2) $ABHAR_i$	(3) $ABHAR_i$	(4) $ABHAR_i$	VIF
$News_i$	0.019*			0.018	1.00
	(1.754)			(1.634)	
$History_i$		0.051***		0.047***	1.03
		(4.267)		(4.271)	
$GainFollowers_i$			0.043**	0.032***	1.09
			(2.458)	(3.237)	
$Followers_i$				0.026**	1.09
				(2.334)	
$Recommendations_i$				-0.019*	1.100
				(-1.651)	
N	10,556	10,556	10,556	10,556	
$F-statistic$	7.328	15.120	7.847	9.318	
R^2	0.001	0.002	0.002	0.005	

In contrast, the total number of recommendations issued has a negative and significant effect on a user's skill. This result is not surprising because the large pool of users shows a wide range of recommendation frequencies. A user making a few recommendations may benefit more from increasing his recommendation frequency than a user who already makes numerous recommendations. To compare the different sources of skills according to the frequency of recommendation emission, I estimate the same specification on two groups of users in Table 1.11. Users are separated into groups according to their posting level: the 10% quantile group corresponds to 1,542 users with less than nine recommendations in total. The 90% quantile group accounts for 1,042 users with more than 103 recommendations. The first column presents the results for the group of users with the fewest recommendations issued. The only variable that has a significant positive effect on the skills of these users is the amount of time spent on the platform. Accordingly, even a user who does not have a large posting level would benefit from spending time on the platform. Although the number of recommendations issued has no significant effect on $ABHAR_i$, the estimated coefficient is negative, which suggests that a user who issues very few recommendations does not benefit from issuing more.

The second column focuses on the skill determinants of the users who make the most recommendations. The results indicate that among these users, their platform activity - measured by time spent on the platform, number of followers, and followers gained - significantly impacts their skill levels. On the other hand, an increase in the number of recommendations made has a negative effect on user skills. Therefore, users who already make many recommendations do not benefit from additional recommendations, and this effect is larger than for those who post at a low frequency.

In summary, the primary sources of their skills for all users are the number of recommendations following a news analysis, the time spent on the platform, and their follower base. Considering that not all users are equal, the skills of low-frequency users improve more with experience, while the skills of high-frequency users increase with the time spent on the platform and the number of followers.

Table 1.11: Sources of Skills: 10% Quantile and 90% Quantile

The table presents parameter estimates from OLS regression on $ABHAR_i$ for the 10% quantile (users with less than nine recommendations) and 90% (users with more than 103 recommendations) quantile based on posting level and the average skill of StockTwits user i with several different specifications. Skills correspond to the average buy and hold return $ABHAR_i$ following multiple recommendations made by user i. Buy-and-hold abnormal returns are estimated using the market model with a three-day event window.
Each variable is standardized; reported estimates measure the change in $ABHAR_i$ corresponding to a one-standard-deviation change in X. The sample consists of all 1,542 (q10%) and 1,042 (q90%) users posting on the platform from 02/01/2014 to 13/01/2021. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. The variable $News_i = \frac{1}{n_i}\sum_{j=1}^{n_i} HasNews_{i,j}$, represents the proportion of recommendations j made following news by the user i, with $HasNews_{i,j}$ a binary variable equal to one if the recommendation j made on the stock follows a news item about the same stock earlier the same day and, n_i the number of recommendation changes made by user i. $History_i$ represents the time difference between the first and last recommendation of user i, $GainFollowers_i$ reflects the number of new followers gained by user i by issuing his recommendations, and, $Followers_i$ is the absolute number of follower of user i. $Recommendations_i$ is the total number of recommendations issued by user i. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) $ABHAR_i$: $q_{10\%}$	(2) $ABHAR_i$: $q_{90\%}$
$News_i$	0.028	0.014
	(0.871)	(0.462)
$History_i$	0.077**	0.092***
	(2.359)	(3.054)
$GainFollowers_i$	0.042	0.086 ***
	(1.204)	(2.681)
$Followers_i$	-0.038	0.208***
	(-1.109)	(6.711)
$Recommendations_i$	-0.023	-0.057*
	(-0.698)	(-1.836)
N	1,542	1,042
$F-statistic$	2.912	16.800
R^2	0.010	0.073

Sources of Skills: Crowd Skills

The ST community's skill corresponds to the abnormal returns related to each recommendation aggregated at the stock level. In other words, this measure refers to the abnormal performance of all users combined for one stock.

The previous analysis in section 1.4.2 implies that the average skills of the ST community $ABHAR_s$ exceed the individual users' skills $ABHAR_i$. In terms of trading strategy, it is, therefore, more profitable to follow the recommendations of the entire community, aggregated at the

stock level. As a result, this research also needs to focus on the sources of the abnormal performance of the crowd: Does the performance relate to the news coverage of this stock, the number of recommendations issued on this stock, or the audience reached by these recommendations?
To identify sources of abnormal ST community performance, I use the following specification:

$$ABHAR_s = \alpha + \gamma X_s + \epsilon_s, \tag{1.6}$$

$$\text{with } X_s = (News_s, Followers_s, Recommendation_s).$$

The variable $News_s = \frac{1}{n_s}\sum_{j=1}^{n_s} HasNews_{s,j}$ represents the proportion of recommendations j made following news for stock s, with $HasNews_{s,j}$ being a binary variable equal to one if the recommendation j made on stock s follows news on the same stock earlier the same day and n_s being the number of recommendation changes made by all users for stock s. $Followers_s$ account for the total number of followers of all users making recommendations on stock s. $Recommendations_s$ is the total number of recommendation events related to stock s from all ST users.
Table 1.12 reports the results of the average abnormal performance sources of all users with respect to a particular stock. The sample consists of 364 stocks. The sample is reduced to account for only the tickers included in the news database on Seeking Alpha. Columns one to three show the results for each variable individually, and column four presents the specifications' results. The proportion of recommendations made following news has a significant relationship with the performance of the ST community. A one-standard-deviation increase in the proportion of recommendations following news corresponds to an average increase of 11% in the abnormal performance of the ST community. This effect is economically significant, showing that after news about a stock has been released, it is worthwhile to follow the aggregated recommendations of users on ST. The total number of followers of all users making recommendations on stock s has a positive and significant relationship with the abnormal performance of ST users making valuable recommendations about stock s. However, one cannot establish a causal relationship because of possible endogeneity bias.
Although the aggregate performance of ST users at the stock level is higher than their individual-level performance, this is not due to a simple increase in the volume of recommendations per stock. In fact, there is a positive relationship between the number of recommendations issued on a stock and user engagement on ST, but this does not always translate into better user performance, as shown by a significant and negative estimated coefficient. Therefore, while the overall superior performance is consistent with the wisdom of the crowd hypothesis, other factors are likely contributing to this outcome.
Finally, users show positive skills mainly due to news analysis and the time spent on the platform. Individual performance is surpassed by the performance of the community in response to news,

regardless of the number of recommendations made by the crowd.

Table 1.12: Sources of Skills: Skill of the Crowd

The table presents parameter estimates from OLS regression on $ABHAR_s$ and the average skill of StockTwits symbol s with several different specifications. $ABHAR_s$ represents average buy and hold abnormal returns of stock s over n_s changes in recommendations. Buy-and-hold abnormal returns are estimated using the market model with a three-day event window.

Each variable is standardized; reported estimates measure the change in $ABHAR_s$ corresponding to a one-standard-deviation change in X. The sample consists of all 364 stocks. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. The variable $News_s = \frac{1}{n_s}\sum_{j=1}^{n_s} HasNews_{s,j}$, represents the proportion of recommendations j made following news for stock s, with $HasNews_{s,j}$ a binary variable equal to one if the recommendation j made on stock s follows a news announcement about the same stock made earlier the same day and, n_s the number of recommendation changes made by all users for stock s. $Followers_s$ account for the total number of followers of all users making recommendations on the stock s. $Recommendations_s$ is the total number of recommendation events related to stock s. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) $ABHAR_s$	(2) $ABHAR_s$	(3) $ABHAR_s$	(4) $ABHAR_s$
$News_s$	0.112**			0.170***
	(2.355)			(3.100)
$Followers_s$		0.206***		0.201***
		(3.746)		(3.601)
$Recommendations_s$			-0.079***	-0.148**
			(-3.399)	(-2.560)
N	364	364	364	364
$F-statistic$	5.546	14.030	15.820	13.330
R^2	0.010	0.039	0.001	0.057

Trading Volume around Recommendation Events

Another possible explanation for ST users' superior performance is the audience size. Since unsophisticated investors are likely to trade after engaging in social media communication (Ammann and Schaub (2021)), the large user base using ST can induce the stock market movements observed around recommendation events. This is in line with previous research on social media recommendations and trading volume (Antweiler and Frank (2004), Sprenger et al. (2014), Kim and Kim (2014)).
To investigate trading volume around changes in recommendation events, I use the Devos et al. (2015) definition of market-model abnormal volume:

$$AV_{s,t} = V_{s,t} - (\alpha_s - \beta_s V_{m,t}). \tag{1.7}$$

Where $V_{s,t}$ is the number of shares of firm s traded on day t divided by the number of shares outstanding and $V_{m,t}$ is the aggregate number of shares traded on day t divided by the total shares outstanding on day t.[10] The market model parameters are estimated over the (-150, -30) period relative to the recommendation event date. Table 1.13 reports the average and median abnormal trading volume around the recommendation event dates. The first column refers to all the recommendation event dates, and the second column refers to the recommendation event dates that are not accompanied by any news about the stock. $\bar{V}_{s,t}$ are the average abnormal trading volume around recommendation revisions and $V_{s,t}^{med}$ are the median abnormal trading volume around recommendation revisions.

The first column demonstrates that, on average, abnormal trading activity starts increasing one day prior to the recommendation and peaks on day one. Median abnormal trading activity is lower but also starts rising from the day previous to the recommendation. To mitigate the effect of trading activity due to news posting, the second column only considers the recommendations event taking place without any concurrent news. The magnitude of the abnormal trading activity is lower, but there is still a peak on day one, with the effect dissipating thereafter. The findings are consistent with Giannini et al. (2019), documenting that Twitter sentiment is associated with higher trading volume around earnings announcements. Collectively, these statistics confirm that one possible explanation for the superior skill of ST users is the correlated trading behavior of a large group of users.

[10]The trading volume data corresponds to the stock daily trading volume divided by shares outstanding (VOL/SHROUT) from the CRSP database.

Table 1.13: Average Abnormal Volume around Recommendations

This table reports average and median abnormal trading volume around the recommendation event dates. The first column refers to all the recommendation event dates, and the second column refers to the recommendation event dates that are not accompanied by any news about the stock. $\bar{V}_{s,t}$ are the average abnormal trading volume around recommendation revisions and $V_{s,t}^{med}$ are the median abnormal trading volume around recommendation revisions. The parameters of the market model are estimated over the (-150, -30) period relative to the recommendation event date.

	(1) All Rec		**(2) No News**	
	$\bar{V}_{s,t}$	$V_{s,t}^{med}$	$\bar{V}_{s,t}$	$V_{s,t}^{med}$
Days				
-2	3.47%	-0.09%	3.61%	-0.08%
-1	4.08%	0.05%	3.99%	0.03%
0	6.63%	0.57%	5.91%	0.40%
1	7.88%	0.64%	7.40%	0.47%
21	4.24%	0.18%	4.18%	0.13%

1.4.4 Robustness

Predicting Ability within Similar Type of Stocks

In line with the results from Table 1.4 and the literature, the predictive power of social network users differs across stocks (Sprenger et al. (2014), Bandara (2016)). To enable a meaningful comparison between SA and ST, it is important to consider that some stocks may be more popular and receive greater attention and analysis on one platform versus the other. Therefore, a more accurate comparison can be made by evaluating the skills of users across similar stocks. Table 1.14 presents descriptive statistics of average abnormal returns across ST users and SA authors for different sectors (Technology, Healthcare, Consumer Defensive, Utilities, Financial Services, Real Estate, Basic Materials, Energy, and Communication Services). Consistent with Tumarkin and Whitelaw (2001), ST users exhibit above-average predictive skills for the stock returns of technology companies (0.13% versus -0.02% average $ABHAR_i$). In addition, the ST users show above-average predictive power for the Utilities, Real Estate, and Basic Materials sectors stocks. Except for the stocks of the Consumer Cyclical and Real Estate sectors, the average predictive power of Seeking Alpha authors is always lower than that of StockTwits users. Seeking Alpha authors have negative average skills in Technology, Consumer Defensive, Utilities, and Financial Services sector stocks. These negative skills statistics contribute to the different predictive power between the two platforms.

Table 1.15 presents descriptive statistics of average abnormal returns for small- and large-capitalization stocks. Small-capitalization and large-capitalization stocks account for the 10% size quantile and 90% size quantile of S&P500 stocks. The average skills of contributors are lower for

Table 1.14: Descriptive Statistics: Users/Authors Skills ($ABHAR_i$) versus Crowd Skills ($ABHAR_s$) for Different Sectors

This table reports descriptive statistics of average buy and hold abnormal return $ABHAR_i$ and $ABHAR_s$ following ST users and SA authors' recommendations for different sectors (I: Industrials, CC: Consumer Cyclical, T: Technology, H: Healthcare, CD: Consumer Defensive, U: Utilities, FS: Financial Services, RE: Real Estate, BM: Basic Materials, E: Energy, CS: Communication Services). There are 2,389,745 and 34,946 recommendation events from 2014 to 2020 for ST and SA, respectively. Buy and hold abnormal returns are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The expected return for each event is estimated with the market model: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event. n represents the number of users/authors and stocks, respectively. FP accounts for the fraction of positive $ABHAR_i$.

					StockTwits						
Sector	**I**	**CC**	**T**	**H**	**CD**	**U**	**FS**	**RE**	**BM**	**E**	**CS**
n	19372	24780	39131	9753	3352	918	8887	791	1969	3991	22042
FP	55.44%	56.19%	56.34%	57.76%	56.86%	57.73%	55.64%	56.89%	56.02%	54.87%	55.87%
Mean	0.09%	0.18%	0.13%	0.19%	0.10%	0.23%	0.13%	0.28%	0.22%	0.04%	0.08%
Std	0.056	0.040	0.032	0.034	0.040	0.027	0.023	0.039	0.042	0.045	0.030
Median	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%
Min	-67.60%	-40.29%	-38.56%	-30.11%	-53.52%	-13.58%	-16.00%	-18.70%	-34.82%	-40.54%	-20.27%
Max	67.60%	40.29%	38.56%	36.25%	32.56%	32.57%	25.32%	17.81%	26.22%	40.54%	20.27%
					Seeking Alpha						
Sector	**I**	**CC**	**T**	**H**	**CD**	**U**	**FS**	**RE**	**BM**	**E**	**CS**
n	705	1184	993	590	326	190	536	72	218	294	741
FP	49.50%	52.87%	50.35%	52.37%	50.31%	48.42%	51.68%	50.00%	51.38%	53.06%	45.34%
Mean	0.05%	0.21%	-0.02%	0.16%	-0.09%	-0.28%	-0.07%	0.29%	-0.04%	0.16%	-0.02%
Std	0.031	0.027	0.025	0.024	0.022	0.023	0.020	0.029	0.030	0.027	0.020
Median	-0.01%	0.08%	0.01%	0.09%	0.07%	-0.09%	0.05%	0.01%	0.03%	0.17%	-0.12%
Min	-34.72%	-32.26%	-24.53%	-20.17%	-17.01%	-11.97%	-14.86%	-4.88%	-11.01%	-14.83%	-11.10%
Max	15.10%	18.25%	25.54%	15.17%	7.98%	10.14%	9.43%	12.92%	19.41%	13.87%	13.28%

large-capitalization stocks than for small-capitalization stocks for both platforms. Unexpectedly, SA authors' average skills $ABHAR_i$ are close to 0.

Table 1.16 presents descriptive statistics of average abnormal returns for low and high book-to-market ratio (B/M) stocks. Low B/M stocks and high B/M stocks account for the 10% B/M quantile and 90% B/M quantile of S&P500 stocks. While the predictive performance is similar for both SA and ST contributors for the low B/M stocks, ST users' average skills are significantly higher for high B/M stocks.

Collectively, the predictive power of social media platforms varies across different types of stocks, with higher performance for low B/M and small-capitalization stocks.

Evaluating Real Performance

Throughout the paper, individual skills are estimated by calculating average buy and hold abnormal returns following a change in recommendation. One way to ensure this measure of skill is appropriate is to evaluate the performance of a trading strategy that relies on these estimates. The performance of a trading strategy driven by changes in user recommendations should increase with the skill level of the users. To test this claim, I divide the users into three groups

Table 1.15: Descriptive Statistics: Users/Authors Skills ($ABHAR_i$) versus Crowd Skill ($ABHAR_s$) for Small and Large Capitalization Stocks

This table reports descriptive statistics of average buy and hold abnormal return $ABHAR_i$ and $ABHAR_s$ following ST users' and SA authors' recommendations. There are 2,389,745 and 34,946 recommendation events from 2014 to 2020 for ST and SA, respectively. Buy and hold abnormal returns are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The expected return for each event is estimated with the market model: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event. n represents the number of users/authors and stocks, respectively. FP accounts for the fraction of positive $ABHAR_i$.

	Small Cap		**Large Cap**	
	StockTwits	**Seeking Alpha**	**StockTwits**	**Seeking Alpha**
n	17809	761	50279	1807
FP	56.66%	48.88%	56.05%	48.70%
Mean	0.16%	0.09%	0.12%	0.00%
Std	0.040	0.030	0.025	0.018
Median	0.00%	-0.06%	0.00%	-0.02%
Min	-40.05%	-14.06%	-30.71%	-19.39%
Max	34.96%	25.59%	30.71%	12.64%

Table 1.16: Descriptive Statistics: Users/Authors Skills ($ABHAR_i$) versus Crowd Skill ($ABHAR_s$) Low and High B/M Stocks

This table reports descriptive statistics of average buy and hold abnormal return $ABHAR_i$ and $ABHAR_s$ following ST users' and SA authors' recommendations for low and high B/M stocks. There are 2,389,745 and 34,946 recommendation events from 2014 to 2020 for ST and SA, respectively. Buy and hold abnormal returns are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The expected return for each event is estimated with the market model: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event. n represents the number of users/authors and stocks, respectively. FP accounts for the fraction of positive $ABHAR_i$.

	Low B/M		**High B/M**	
	StockTwits	**Seeking Alpha**	**StockTwits**	**Seeking Alpha**
n	33086	972	18000	1194
FP	55.56%	50.10%	57.12%	50.59%
Mean	0.09%	0.09%	0.18%	-0.03%
Std	0.026	0.024	0.034	0.025
Median	0.00%	0.00%	0.00%	0.02%
Min	-32.29%	-16.28%	-40.05%	-24.53%
Max	34.21%	20.11%	34.96%	14.85%

according to their skill level: The top 75% user group represents users with $ABHAR_i$ higher than 25% quantile, the above-median user group represents users with $ABHAR_i$ higher than the 50% quantile and, the top 25% user group represents users with $ABHAR_i$ higher than the 75% quantile. All $ABHAR_i$ estimates in this section are from January 2014 to December 2018. For each user of the different groups, I calculate the cumulative returns after December 2018 of a trading strategy based on buying stocks following an upgrade and selling stocks following a downgrade. Similar to the skill measure, the positions are liquidated after three days. For user i and change in recommendation event j, $r^3_{i,j}$ denotes the realized return after three days (excluding day 0). To estimate the trading performance following all recommendations for each user, I calculate the cumulative return at the user level:

$$R_i^{cum} = (\prod_{j=1}^{n_i}(1 + r^3_{i,j})) - 1, \tag{1.8}$$

with n_i the number of changes in recommendation made by user i.

Table 1.17 presents the average cumulative returns $\bar{R}^{cum}$ and standard deviations of cumulative returns $s_{R^{cum}}$ for the different user groups over a one-year horizon. The average cumulative returns for all users is 1.3% and increases as we consider groups of users with higher estimated skills. Regarding the Top 25% users, the annual cumulative performance is, on average, 9.3%. The above results confirm that social media recommendations generate positive return strategies (Sul et al. (2014), Zhang and Skiena (2010)), and users with higher skills on those platforms have greater predictive power (Wang et al. (2015)). In parallel, the performance dispersion decreases as we consider superior-skill users, with the standard deviation of cumulative returns going from 0.914 for the Top 75% users to 0.393 for the Top 25% users. These results motivate the choice of the paper skill measure to evaluate the skills of ST users and emphasize the economic value of the information available on the ST platform.

Table 1.17: Users Real Performance

The table presents the average cumulative returns $\bar{R}^{cum}$ and standard deviations of cumulative returns $s_{R^{cum}}$ for the different user groups over a one-year horizon. n represents the number of users within each group. The user groups are defined according to their estimated skills ($ABHAR_i$). The top 75% user group represents users with $ABHAR_i$ higher than the 25% quantile, and the above median user group represents users with $ABHAR_i$ higher than the 50% quantile. The top 25% user group represents users with $ABHAR_i$ higher than the 75% quantile. All $ABHAR_i$ estimates are from January 2014 to December 2018.

	All Users	**Top 75% Users**	**Top 50% Users**	**Top 25% Users**
$\bar{R}^{cum}$	1.30%	3.50%	6.70%	9.30%
$s_{R^{cum}}$	0.770	0.914	0.414	0.393

Event Window Choice

This study uses three days event windows for ST users and SA authors. Alternatively, I consider various investment horizons from one-day to five-days. Crane and Crotty (2020) calculate average abnormal returns using a five-day event window; therefore, we can expect that the choice of the appropriate event window for financial websites with real-time recommendations is less than five days. Table 1.18 reports descriptive statistics of average buy and hold abnormal returns $ABHAR_i$ and $ABHAR_s$ following ST users' and SA authors' recommendations for different choices of event windows. Average ST users' skills are maximized with a three-day investment horizon. This result is consistent with Renault (2017), documenting a positive abnormal return following a spike in message activity on social media platforms, followed by a significant price reversal during the next five trading days. The short-term reversal in the stock performance can be explained by the size of the ST crowds or the large proportion of news events simultaneous to ST users' recommendations. For SA authors, average skills are maximized with a four-day investment horizon and decrease with a longer investment horizon. Most notably, regardless of the choice of event window in this study, the statement that, on average, the skills of ST users exceed those of SA authors remains valid.

Table 1.18: Descriptive Statistics: Users/Authors Skills ($ABHAR_i$) versus Crowd Skills ($ABHAR_s$) with Different Event Windows

This table reports descriptive statistics of average buy and hold abnormal return $ABHAR_i$ and $ABHAR_s$ following ST users' and SA authors' recommendations. There are 2,389,745 and 34,946 recommendation events from 2014 to 2020 for ST and SA, respectively. Buy and hold abnormal returns are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The expected return for each event is estimated with the market model: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event.

	StockTwits					Seeking Alpha				
Event window	**1-day**	**2-day**	**3-day**	**4-day**	**5-day**	**1-day**	**2-day**	**3-day**	**4-day**	**5-day**
User Level										
Mean	0.12%	0.12%	0.12%	0.11%	0.10%	0.01%	0.01%	0.01%	0.03%	0.03%
Std	0.016	0.020	0.015	0.025	0.026	0.008	0.012	0.014	0.017	0.018
Median	0.02%	0.01%	0.05%	0.00%	0.00%	0.00%	0.02%	-0.01%	0.00%	0.01%
Min	-35.58%	-36.01%	-36.28%	-53.27%	-65.39%	-7.89%	-11.72%	-10.83%	-14.21%	-15.99%
Max	25.89%	35.03%	22.33%	48.34%	39.93%	7.56%	8.46%	10.77%	9.90%	12.27%
Fraction Positive	58.77%	58.16%	55.94%	57.25%	57.26%	50.17%	50.98%	49.50%	50.05%	50.74%
Stock Level										
Mean	0.16%	0.16%	0.17%	0.16%	0.16%	0.01%	0.07%	0.09%	0.10%	0.11%
Std	0.001	0.002	0.002	0.002	0.002	0.004	0.007	0.008	0.011	0.011
Median	0.15%	0.15%	0.15%	0.15%	0.14%	0.01%	0.02%	0.07%	0.06%	0.05%
Min	-1.14%	-1.89%	-2.33%	-2.79%	-3.66%	-2.38%	-2.44%	-2.68%	-2.87%	-4.01%
Max	0.62%	0.68%	0.77%	0.69%	0.70%	2.36%	5.93%	6.60%	5.75%	7.84%
Fraction Positive	94.73%	91.28%	92.09%	89.45%	86.82%	51.24%	52.28%	54.98%	55.81%	53.53%

Following Crane and Crotty (2020), a five-day event window is appropriate to evaluate the skill of financial analysts. However, one may wonder if the type of recommendations on SA is not intended for longer-term investments. Figure 1.4 depicts the evolution of the average abnormal performance of the SA authors as a function of the duration of the investment. After 30 days, the abnormal performance reaches its peak with average $ABHAR_i$ reaching 0.12%. These results support that after sufficient time the performance of the SA authors converges to that of the ST users. However, the standard deviation of the authors' skills doubles after 30 days. From an opportunity cost perspective, these results are insufficient to change this study's overall conclusions.

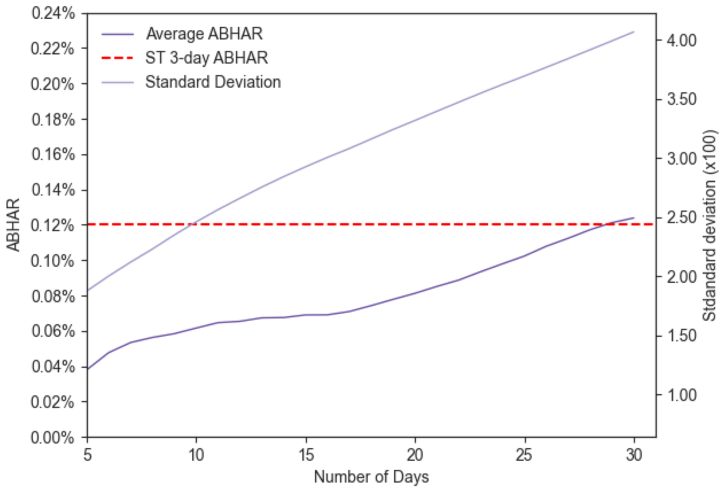


Figure 1.4: Evolution of SA Authors Skills with Event Window Size

These figures plot the evolution of cross-sectional statistics of average buy and hold abnormal return $ABHAR_i$ following SA authors' recommendations as a function of the number of days of the event window. There are 34,946 recommendation events from 2014 to 2020. Buy and hold abnormalities are derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The expected return for each event is estimated with the market model: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event. The red line corresponds to the average buy and hold abnormal return $ABHAR_i$ for ST users over a three-day horizon.

Seeking Alpha Data Choice

Although Seeking Alpha (SA) charges users for access to certain information and compensates authors for their articles, it is not the sole source of fee-based stock recommendations. Among those many outlets, SA is one of the largest investment-related social media websites in the U.S. The SA website was awarded a Forbes' Best of Web designation and was ranked number one in Constantine von Hoffman's list of Essential Economic blogs. Chen et al. (2014) argue that incentives to share valuable information are particularly strong on SA and document that the prediction power of the investment advice on SA outperforms Wall Street analysts' reports and financial news articles.

In contrast with SA, the ST platform allows for low-cost communication, reducing barriers to entry. Moreover, articles published on SA are reviewed by a panel and subject to editorial changes (Chen et al. (2014)). This review process and the content of SA articles are closer to the work of a financial analyst (see Figure 1.1) while having a frequency of publication closer to that of a social network.

To highlight the qualifications of these authors, I compare their predictions to the future recommendations of financial analysts. For obvious reasons related to differences in publication frequency (SA authors are faster than financial analysts at adjusting their recommendations), financial analysts may not adapt their recommendations as quickly as social network investors do when a financial report or news about a company is released.

Therefore, I suggest that if the recommendations of SA authors have a positive and significant relationship with the future recommendations of financial analysts, it is because SA authors can be considered a representative group of investment professionals. Notably, the same analysis with the ST users does not make sense because, in most cases, these users do not express recommendations following in-depth analyses, as do SA authors.

To obtain data on consensus analyst stock recommendations, I used the Institutional Brokers Estimates (I/B/E/S) database from Thomson Reuters. This database includes data on S&P500 stocks and covers the period from 2014 to 2020. The data frequency is monthly, and I use the mean recommendation score $MeanRec_{s,t}$, the percentage of buy recommendations made for stock s in month t, $BuyPercentage_{s,t}$, and the percentage of sell recommendations made for stock s in month t, $SellPercentage_{s,t}$. The summary recommendation mean corresponds to the average of each analyst recommendation score from 1 to 5: 1. Strong buy, 2. Buy, 3. Hold, 4. Underperform, and 5. Sell. For SA data, I assign a score of 1 for a long recommendation and -1 for a short recommendation. Monthly SA recommendations for the month t on stock s, $RecSA_{t,s}$, correspond to the average of all recommendation scores of SA authors in this month for stock s. I merge monthly analysts' stock recommendation variables with the past month's average

SA author recommendations for each S&P500 stock, s. The merged dataset comprises 11,945 summary analyst stock recommendations matching the previous month's SA recommendations. I assume that if the SA authors are a representative group of financial professionals, then they should, on average express recommendations in the same direction as financial analysts (with a maximum lag of 3 months for the financial analysts).

To study the relationship between summary analysts' recommendations and SA recommendations, I use the following specification:

$$Y_{s,t} = \alpha + \gamma X_{s,t} + \epsilon_{s,t}, \tag{1.9}$$

$$\text{with } X_{s,t} = (RecSA_{t-1,s}, RecSA_{t-2,s}, RecSA_{t-3,s}) \text{ and,}$$
$$Y_{s,t} = (MeanRec_{s,t}, BuyPercentage_{s,t}, SellPercentage_{s,t}).$$

The variables are standardized to facilitate comparison.

Table 1.19: Consensus Analyst Recommendations & Seeking Alpha Recommendations

The table presents parameter estimates from an OLS regression on $MeanRec_{s,t}$, $BuyPercentage_{s,t}$, $SellPercentage_{s,t}$ with different specifications. Each variable is standardized; reported estimates measure the change in $Y_{s,t}$ corresponding to a one-standard-deviation change in $X_{s,t}$. The sample consists of 11,944 summary analyst stock recommendations matching the previous month's SA recommendations. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. $MeanRec_{s,t}$ is the mean recommendation score (from 1 for a Strong Buy recommendation to 5 for a Sell recommendation). $BuyPercentage_{s,t}$ is the percentage of buy recommendations made for stock s in month t, and $SellPercentage_{s,t}$ is the percentage of sell recommendations made for stock s in month t. $RecSA_{t,s}$ corresponds to the average of all recommendation scores of SA authors made in month t for stock s.

	(1) $MeanRec_{s,t}$	(2) $BuyPercentage_{s,t}$	(3) $SellPercentage_{s,t}$
$RecSA_{t-1,s}$	-0.037***	0.048***	-0.059***
	(-3.765)	(4.892)	(-6.357)
$RecSA_{t-2,s}$	-0.039***	0.050***	-0.016
	(-3.813)	(4.921)	(-1.588)
$RecSA_{t-3,s}$	-0.039***	0.044***	-0.022**
	(-3.976)	(4.472)	(-2.236)
N	11'944	11'944	11'944
$F-statistic$	25.980	36.430	14.450
R^2	0.008	0.012	0.006

Table 1.19 presents the results for these specifications. Columns one to three show the results for the three different dependent variables. The average recommendations of the SA authors over the last three months have a significant relationship with the current financial analysts' mean recommendation scores. In addition, the signs of the coefficients point in the right direction. Note that the lower the mean recommendation score, the better the average recommendation of financial analysts (going from 1 for a Strong Buy recommendation to 5 for a Sell recommendation). An increase in the bullishness of the recommendations on SA correlates with a better future stock ranking by financial analysts. These observations hold over a 3-month horizon. The results in columns two and three show that a larger average SA author recommendation score translates to an increase in the percentage of buy recommendations and a decrease in the percentage of sell recommendations in the next three months.
While these results are not intended to claim that the SA authors can predict financial analysts' future recommendations, they acknowledge the choice of using the SA data as a representative group of financial professionals.

1.5 Conclusion

In this paper, I evaluate the skills of social media investors at making valuable recommendations. Social media investors can be either qualified investors (Seeking Alpha authors) or crowd investors with different levels of financial knowledge (StockTwits users). The primary purpose of this paper is to study the skills of crowd investors at the individual and aggregate levels and compare these skills to those of professionals issuing recommendations at a similar frequency. Second, I identify the sources of the skills of these users. The user's skill is defined as the average buy and hold returns following all recommendations issued by this user.
First, the cross-sectional statistics indicate that, on average, individual skills are greater for ST users than for SA authors. These statistics increase when considering groups of more active users on the platform while the users' skill heterogeneity decreases. The aggregated abnormal performance at the stock level of all users/authors outperforms their individual abnormal performance on both platforms.
Second, I use a mixture modeling approach to estimate the fraction of ST users and SA authors of different skill groups with specific skill levels. The estimation indicates that the large majority of ST users are drawn from a distribution centered at a positive abnormal return. In contrast, the large majority of SA authors are drawn from a distribution centered around zero abnormal return. All estimates in this paper suggest that the skills of ST users, although representing a more heterogeneous group of users, exceed the skills of SA authors. The importance of the results is emphasized by the illustration that the SA authors form a representative group of qualified investors.

Last, I analyze the sources of skills for individual ST users and for the overall ST platform. Descriptive statistics show that although most recommendations occur independently of a news event, the highest average abnormal performances are achieved as a result of a reaction to news and not as a result of their own analysis. Regression results show that higher skills are attainable for users with a higher proportion of recommendations made after analyzing news, more time spent on the platform, and a large community of followers, regardless of the number of recommendations made. The abnormal performance of all ST users aggregated to the stock level also increases with the proportion of recommendations following news. From an economic standpoint, it is still valuable to analyze StockTwits data even if the community is not considered as trustworthy as the curated authors of finance sites. From an investment perspective the recommendations of all users aggregated at the stock level have the best performance on average, enhanced by reactions to news, users with a long history, and a large follower base. Finally, if we consider a trading strategy based on the recommendations of ST users, the performance improves by taking into account the most skilled users' recommendations.

Bibliography

Ammann, M. and Schaub, N. (2021). Do individual investors trade on investment-related internet postings?, *Management Science* **67**(9): 5679–5702.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards, *The Journal of Finance* **59**(3): 1259–1294.

Avery, C. N., Chevalier, J. A. and Zeckhauser, R. J. (2016). The "caps" prediction system and stock market returns, *Review of Finance* **20**(4): 1363–1381.

Bandara, W. (2016). The information content of pre-open social media, *Available at SSRN 2758915* .

Bollen, J., Mao, H. and Zeng, X. (2011). Twitter mood predicts the stock market, *Journal of Computational Science* **2**(1): 1–8.

Boni, L. and Womack, K. L. (2006). Analysts, industries, and price momentum, *Journal of Financial and Quantitative Analysis* **41**(1): 85–109.

Chen, H., De, P., Hu, Y. J. and Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media, *The Review of Financial Studies* **27**(5): 1367–1403.

Chen, Y., Cliff, M. T. and Zhao, H. (2017). Hedge funds: The good, the bad, and the lucky, *Journal of Financial and Quantitative Analysis (JFQA)* **52**(3): 1081–1109.

Crane, A. and Crotty, K. (2020). How skilled are security analysts?, *The Journal of Finance* **75**(3): 1629–1675.

Crawford, S. S., Gray, W. R. and Kern, A. E. (2017). Why do fund managers identify and share profitable ideas?, *Journal of Financial and Quantitative Analysis* **52**(5): 1903–1926.

Das, S. R. and Chen, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web, *Management Science* **53**(9): 1375–1388.

Devos, E., Hao, W., Prevost, A. K. and Wongchoti, U. (2015). Stock return synchronicity and the market response to analyst recommendation revisions, *Journal of Banking & Finance* **58**: 376–389.

Dewally, M. (2003). Internet investment advice: Investing with a rock of salt, *Financial Analysts Journal* **59**(4): 65–77.

Giannini, R., Irvine, P. and Shu, T. (2019). The convergence and divergence of investors' opinions around earnings news: Evidence from a social network, *Journal of Financial Markets* **42**: 94–120.

Goutte, M.-R. (2022). Do actions speak louder than words? Evidence from microblogs, *Journal of Behavioral and Experimental Finance* **33**: 100619.

Jegadeesh, N. and Kim, W. (2010). Do analysts herd? An analysis of recommendations and market reactions, *The Review of Financial Studies* **23**(2): 901–937.

Kim, S.-H. and Kim, D. (2014). Investor sentiment from internet message postings and the predictability of stock returns, *Journal of Economic Behavior & Organization* **107**: 708–729.

McLachlan, G. J., Lee, S. X. and Rathnayake, S. I. (2019). Finite mixture models, *Annual Review of Statistics and its Application* **6**: 355–378.

Mittal, A. and Goel, A. (2012). Stock prediction using twitter sentiment analysis, *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)* **15**.

Oh, C. and Sheng, O. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement.

Oliveira, N., Cortez, P. and Areal, N. (2013). On the predictability of stock market behavior using stocktwits sentiment and posting volume, *Portuguese Conference on Artificial Intelligence*, Springer, pp. 355–365.

Rao, T., Srivastava, S. et al. (2012). Analyzing stock market movements using twitter sentiment analysis.

Renault, T. (2017). Market manipulation and suspicious stock recommendations on social media, *Available at SSRN 3010850* .

Renault, T. (2020). Sentiment analysis and machine learning in finance: A comparison of methods and models on one million messages, *Digital Finance* **2**(1): 1–13.

Skuza, M. and Romanowski, A. (2015). Sentiment analysis of twitter data within big data distributed environment for stock prediction, *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, IEEE, pp. 1349–1354.

Sprenger, T. O., Tumasjan, A., Sandner, P. G. and Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs, *European Financial Management* **20**(5): 926–957.

Sul, H., Dennis, A. R. and Yuan, L. I. (2014). Trading on twitter: The financial information content of emotion in social media, *2014 47th Hawaii International Conference on System Sciences*, IEEE, pp. 806–815.

Tumarkin, R. and Whitelaw, R. F. (2001). News or noise? internet postings and stock prices, *Financial Analysts Journal* **57**(3): 41–51.

Wang, G., Wang, T., Wang, B., Sambasivan, D., Zhang, Z., Zheng, H. and Zhao, B. Y. (2015). Crowds on wall street: Extracting value from collaborative investing platforms, *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 17–30.

Wysocki, P. D. (1998). Cheap talk on the web: The determinants of postings on stock message boards, *University of Michigan Business School Working Paper* (98025).

Zhang, W. and Skiena, S. (2010). Trading strategies to exploit blog and news sentiment, *Fourth International aAAI Conference on Weblogs and Social Media*.

Chapter 2

Your Network Is Your Tweet Worth

This paper examines the relationship between financial network users' prediction power and their position in the network, using communication and like networks. Results indicate that network centrality measures are positively associated with users' future performance and effectively identify high-quality users. The study investigates the grounds of financial social media prediction power and finds that the most influential investors' performance is not driven by the wisdom of crowds but rather by the emergence of echo chambers. Echo chambers arise because central users reduce diversity within the network, with a stronger impact on users with extreme positions on stocks.

2.1 Introduction

Social media platforms have revolutionized the financial landscape, providing a faster and more accessible means of disseminating information, leveling the playing field for retail investors, and challenging the dominance of Wall Street titans. Academic research has confirmed that financial social media platforms offer valuable insights into predicting stock returns and other financial variables (Chen et al. (2014); Broadstock and Zhang (2019)). The exponential growth of social media use among individual investors has given rise to a new breed of influential market participants: the financial "influencers" or "web gurus," whose impact on the market has been widely documented.

However, empirical research on social networks has been hindered by the lack of data on connections and social communication. This issue extends to empirical papers in Network Theory since network nodes and degrees are not fully observable. Financial social media data can help address these issues since it allows the construction of a network of investors by capturing their communication and linking activity. In this sense, a financial social media platform can be viewed as a network where users are the nodes, and the communication between them forms the links or degrees.

This paper examines the relationship between financial network users' skills, centrality, and network characteristics to determine the most likely drivers of their predictive power. I also investigate the impact and risk of financial social networks, highlighting the implications for financial markets. The first objective is to understand the influence of network centrality on an investor's ability to predict future stock returns. I focus on identifying influential users by ranking investors based on their centrality in the network.

This study contributes to the growing body of literature on financial social media's impact on market prediction by exploring the factors underlying online network users' predictive power. Two concurrent hypotheses underlie opinion formation in social networks. On the one hand, there is the wisdom of crowds, which arises from the divergence of opinions that add up when several users communicate or from the aggregation of individual opinions. On the other hand, social media platforms can foster homophily, or the tendency to cluster around similar opinions, leading to echo chambers. Such chambers have negative social consequences and pose a particular risk in financial social networks since they can lead to poor investment decisions and inefficient markets. The second objective of this paper is to distinguish between the two competing hypotheses. Understanding the origin of social media users' opinions is crucial because opinion transmission through social media affects trading patterns, particularly among unsophisticated investors engaged in attention-induced trading (Barber et al. (2021)). Network graphs and topology measures can help differentiate between the two hypotheses. For instance, a central user on a financial social network who typically agrees with the rest of the network and communicates with users who share similar opinions would support the echo chambers hypothesis. Conversely, a central user who tends to have a different opinion than the rest of the network and communicates with a diverse group of users would support the wisdom of the crowd hypothesis.

I use data from the StockTwits platform, including 20.9 million tweets from 326,686 users on S&P500 stocks, to construct two finance investors' networks. To build the network connections, I use the likes between users and their communication (retweet, reply, or mention) in their tweets. The StockTwits platform allows users to tag the stock discussed in their tweets and the date of the post, allowing to construct ticker and monthly sub-networks. I use various network centrality measures to identify influential users, including degree, closeness, and betweenness.

I rank users based on two measures: their prediction power and demonstrated skill. The analysis indicates that social media users' predictive performance improves with centrality in the network, regardless of whether users communicate or like each other. Moreover, advanced centrality measures can better identify high-quality network users beyond simple scaling effects. Users who have the potential to impact the entire network most significantly exhibit an increase in their average skill.

To assess the presence of homophily and echo chambers in financial social media, I use network graphs to construct two diversity variables: the diversity of a user with their neighbors and the diversity among a user's neighbors. Users in a ticker sub-network are likely to be part of an echo chamber if their average opinion on a stock is very similar to their nearest neighbors. Assessing the heterogeneity of the population a user reaches while making a recommendation is a crucial aspect of financial social media dynamics. From an individual perspective, user diversity should be high to support the wisdom of the crowd hypothesis, but diversity should also be high among the user's neighbors.

The findings on the diversity of all users' opinions in the network are mixed. Increasing a user's centrality leads to a decrease in the diversity among their neighbors and an increase in their diversity relative to their neighbors. When focusing only on the most central users, evidence supporting the wisdom of the crowd hypothesis fades. The most central users are associated with a reduction in diversity, up to -18% for the StockTwits influencers, resulting in thousands of users expressing similar opinions. The quality of like-minded opinions has significant consequences on financial markets. It is crucial to understand what types of recommendations are more widely transmitted to the network, i.e., moderate or extreme opinions. To detect the presence of echo chambers in financial social media, I study the evolution of diversity for the most central users being extremely bullish and bearish. I find that the negative effect of central users on diversity is most pronounced for those with extreme opinions. Furthermore, the negative effect is the largest, reducing diversity by up to one-third for the extremely bullish users and up to 40% for the extremely bearish users. This finding provides evidence of the worst types of echo chambers in financial social media, which pose significant risks for investors as they are characterized by an absence of exposure to opposing views. These echo chambers prevent investors from making well-informed decisions, and given the high stakes involved in financial decision-making, they can be particularly detrimental. In light of these findings, it is crucial to evaluate the role of social media in finance and its impact on market prediction. This study contributes to the growing body of literature exploring the effect of financial social media on market prediction by investigating the factors that underpin the predictive power of online network users. Additionally, this research highlights the potential risks associated with the popularization of financial social networks and the need to understand better how these platforms can influence trading patterns and investment decisions.

In summary, this paper demonstrates that an investor's position within an online network significantly influences their performance in predicting financial markets. Additionally, I study how financial social media may impact how information spreads and how echo chambers emerge. The results suggest that the performance of the most influential investors in the network is not driven by the wisdom of crowds but by the formation of echo chambers. These echo chambers emerge as the most central users reduce the diversity within the network, reinforced by users with extreme positions on stocks.

The remainder of the paper is structured as follows. First, section 2 presents the corresponding literature review. Section 3 describes the data and the methodology. Section 4 discusses the results. Finally, I conclude in Section 5.

2.2 Literature Review

The existing body of literature on social networks has established that individuals with more extensive and central network connections have access to superior information. Jackson (2007) summarizes research that links social networks to economics. Specifically, numerous studies focusing on social networks in finance have consistently found that individuals with more extensive and central network connections outperform their peers in various domains such as M&A deals (Cai and Sevilir (2012)), VC companies (Hochberg et al. (2007)), compensations (Engelberg et al. (2013)), and loan characteristics (Engelberg et al. (2012)). Network connections also influence the holdings of mutual funds (Pool et al. (2015)) and enhance their performance (Cohen et al. (2008), Coval and Moskowitz (2001)). At the individual level, connections also positively impact analysts' performance (Cohen et al. (2010)).

Given the advantages of social networks, many studies use network topology measures to uncover the sources of better performance.[1] For example, Li and Schürhoff (2019) use networks of dealers

[1]See Kenett and Havlin (2015) for an overview of applications of network science in economics.

in OTC municipal bond markets to show that network centrality increases dealer execution costs (larger markups). Babus and Kondor (2018) developed a model of trade in OTC markets using traders' networks and found that central dealers tend to learn more, trade at lower costs, and earn higher expected profits. Similarly, Chen et al. (2021) build a user co-attention network based on investors' searches of stocks on the web and show that an increase in network centrality is associated with higher abnormal returns. Cortez et al. (2016) study user influence using StockTwits data and network graph structure. Direct interactions in StockTwits are used to build network centrality measures that allow ranking the quality and influence of the user.

Moreover, network variables on social media data can identify quality recommendations and influential users, as well as differentiate the wisdom of the crowd hypothesis from the echo chamber hypothesis. In recent years, a large body of finance literature has studied the link between social media content and stock market variables. For instance, Chen et al. (2014) demonstrates that users' recommendations on StockTwits predict future stock returns and earnings surprises. Broadstock and Zhang (2019) show that sentiment from social media carries' pricing power' against the stock market. Tu et al. (2018) study how to estimate qualities of investment recommendation on StockTwits, which helps improve the prediction of stock recommendations. Cookson and Niessner (2020) rely on StockTwits data to study sources of investor disagreement and find that disagreement leads to higher trading volume, driven more by different information sets than by differences across investment philosophies. In recent research,Cookson et al. (2021) further show that disagreement leads to higher informed trading by activists, insiders, and short sellers. Bartov et al. (2018) found that aggregated user sentiment from Twitter successfully predicts firms' earnings and announcement returns. The prediction power remains after controlling concurrent news from traditional media sources. Additionally, Hossain et al. (2022) use StockTwits data to identify informed tweets, i.e., tweets with hyperlinks to original sources of information, which leads to stock price crash risk deduction and a reduction in managers' ability to hoard bad news.

The predictive power of financial social media can be attributed to the wisdom of crowds, as suggested by Hong and Page (2004), who showed that a diverse group of intelligent problem solvers could outperform a group of the best problem solvers. Frey and Van de Rijt (2021) introduced the "social dilemma" and demonstrated that decision quality increases when individuals are influenced by a majority that thinks differently and when individuals have independent voices. In line with this, Du et al. (2017) found that the aggregation of opinions weighted to account for individual dependence outperforms other aggregation methods for prediction. Empirically, Chen et al. (2014) documented the wisdom of crowds and found that opinions transmitted on the SeekingAlpha website strongly predict future stock returns and earnings surprises. Similarly, Breitmayer et al. (2019) reported the wisdom of crowds using "Shares" stock recommendations, providing explanatory power for future and abnormal stock returns. Hong et al. (2020) investigated the impact of crowd characteristics on future crowd prediction performance using StockTwits data and found that experience diversity, participant independence, and network decentralization are all positively associated with future crowd performance. Although the wisdom of the crowd hypothesis is plausible, the predictive value of social media data may come from correlated actions resulting from behavioral biases, such as copying influential users, rather than the aggregation of diverse opinions. Hirshleifer (2020) discussed social transmission bias in finance and how social interaction can shape economic thinking and behaviors. Molavi et al. (2018) modeled investors learning over social networks and found that unsophisticated investors are likely to herd. Pedersen (2022) presented a model on investment ideas propagation through social networks and demonstrated that fanatics, i.e., naive investors learning from social networks, and rational investors dominate over time, with their dominance emphasized depending on financial influencers' blessing.

Empirical evidence suggests that social interactions contribute to the disposition effect, as demonstrated by Heimer (2016) using financial social network data. Hong et al. (2005) found that investors spread information on stocks through word of mouth, and mutual fund managers from similar cities tend to have similar holdings. Cookson et al. (2022) found evidence of selective exposure to confirmatory information, i.e., echo chambers, across users of StockTwits. Specifically, they found that users are more likely to follow a user with similar views, which holds for less active and professional users. Bakshy et al. (2011) studied cascades on Twitter and argued that influential users are more likely to create significant message cascades in the future. Della Rossa et al. (2020) argued that when users tend to conform their beliefs to those of few leaders, herding can harm market efficiency. On the other hand, when users consider the plurality of opinions and act in line with the wisdom of the crowds, it enhances market efficiency.

Behavioral biases can generate correlations in beliefs and actions that lead to the formation of echo chambers when taken to the extreme. In the social science field, Iyengar and Hahn (2009) have demonstrated that the proliferation of new media and multiple media options contributes to polarization in the news audience through selective exposure. In their research on Reddit communities, Kumar et al. (2018) found that negative mobilizations have long-term detrimental effects on inter-community interactions and conflict. Bail et al. (2018) studied echo chambers on Twitter by exposing a large group of Democrats and Republicans to bots that retweeted messages from opposing political views. They found that opposing views on social media can further exacerbate the user's pre-existing beliefs and increase political polarization. Du and Gregory (2016) studied echo chambers and polarization on Twitter and found that Twitter communities are becoming more polarized with time. Their findings relied on the Twitter follows network at two different times, with the more recent edges in the network coming from inside existing network communities rather than between network communities. In their study on rumors spreading on online social media, Choi et al. (2020) found that rumors spread by echo chamber members tend to spread faster than those not distributed by echo chamber members, highlighting the potential adverse effect of social media echo chambers through amplifying rumor propagation. Barberá et al. (2015) used Twitter data to estimate users' ideological preferences and evaluate whether communication on the platform referred to an echo chamber or a national conversation. They found that the aggregate level of political polarization depends on the issue being discussed. Previous research has shown that social media platforms encourage the formation of echo chambers by enabling users to create communities.

2.3 Data and Methodology

2.3.1 StockTwits Data

Participants in the capital markets use social media to discuss stock ideas, communicate about investments, gain information, and connect with others. I use StockTwits data for this study. StockTwits is the most extensive social network for investors and traders, with over one million registered members and approximately three million monthly visitors. The platform enables users to express recommendations in real-time. In the dataset, the following variables are associated with each tweet: the content of the message, the ticker of the stock related to the message, the user's name, the time of the post (with split-second accuracy), the number of users liking or sharing the tweet, the number of people following the user, the number of tweets already posted by the user, the date from which the user has been active on the platform and, a logical variable called 'official account.' Since 2012, users can classify their messages as bearish or bullish. Those self-classified tweets represent approximately 20% of the sample. I use this bullish/bearish label to train the machine learning classifier, which splits the remaining tweets into positive and negative recommendations.

All the tweets of S&P500 stocks from December 2016 to January 2022 comprise 20'983'763 tweets from 326'686 users and 506 tickers. Once I only consider the users communicating with each other, the dataset represents 1'033'113 tweets from 88'615 users (around 5% of the tweets are communication tweets).

A message on StockTwits can easily be considered a recommendation from a platform user to the community. Before the classification of tweets, I perform different text preprocessing tasks.[2] First, I put the text in lower cases, removing HTML codes, some punctuation, and stopwords. Following Renault (2020), I keep several punctuation marks ! ? % + '= : ;) (. and several key stopwords used in finance to increase the accuracy of the classifier.[3] According to Renault (2020), POS tagging and Stemming reduce the accuracy of a sentiment classifier. Tickers ($AAPL) are replaced by the word cashtag, users mentioned in tweets (@user) by the word usertag, and links by the word link tag. Words replace smileys and emoticons using the emoji package in python.[4]

One of the key functionalities of StockTwits is that users can classify their message as bearish or bullish when posting their tweets. The 'self-classified' tweets account for 30% of the messages posted. Table 3.6 displays examples of self-classified tweets for both categories.

To train the machine learning classifier, I use an unbalanced dataset of 13'585'586 tweets with

[2]More details on the text-preprocessing tasks are provided in Goutte (2022).

[3]Indeed, NLTK Stopwords corpus includes words like 'up,' 'down,' 'below,' and 'above,' which give important meaning to a sentence in finance.

[4]For example, the smiley with the face with tears of joy is replaced by the word face_with_tears_of_joy.

77.34% positive tweets.[5] The unequal proportion of long and short headlines is consistent with the past decade's market movements. Moreover, Antweiler and Frank (2004) report that the majority of investors on social media have a bullish view of the market. I divide the dataset into 80% training and 20% test sets.

I use various algorithms for text classification: naive Bayes (NB), support vector machines (SVMs), and the maximum entropy classifier (MaxEnt). To obtain optimal performance, I optimize parameters with 'GridSearchCV'.[6] After attributing a recommendation score to each message with the selected classifier (+1 for a positive recommendation and -1 for a negative recommendation), the final StockTwits data comprises 20'983'763 tweets from 326'686 users.

Table 2.1: Example of Tweets for each category
This table displays examples of self-classified tweets for both bullish category and bearish category. Each tweet is associated with the ticker of the stock discussed in the message. For example, for the stock of Apple Inc., each tweet is referenced with the ticker "$AAPL."

Category	Tweets
Bullish	$AAPL highly attractive $AAPL Fly me to The Moon......YESSSSSS $AAPL Deutsche Bank sees stocks up just 3% in 2014 http://stks.co/ayfQ $AAPL AH up $2.08 do we see $570.00 tomorrow. Very possible
Bearish	$AAPL See today's downgrade coming. What a drama. $AAPL Looks like another day in the red, more waiting $AAPL bad EOD for AAPL bad all day for AAPL bad bad bad bad bad bad bad $AAPL probably should sell

2.3.2 StockTwits Variables

Skill Measure and Prediction Power Measure

To rank users according to the goodness of their recommendations, I choose to use those two measures: prediction power and skill. The skill measure is based on the paper by Crane and Crotty (2020), which uses this measure to evaluate analysts' skills when making recommendations. It corresponds to the average buy-and-hold abnormal return following each user's recommendation (bullish or bearish). The second variable, Prediction Power, is based on the measure of the quality of recommendations used in Zhang (2009) and Sprenger et al. (2014). This measure is then aggregated at the user level to compute the prediction power variable.

[5] According to Renault (2020), an unbalanced dataset provides greater prediction accuracy than a balanced dataset with the same number of tweets.

[6] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

- **Skill:**

$$BHAR_{i,j} = \Delta^{s}_{i,j}(\prod_{t=1}^{4}(1 + r_{s,t}) - \prod_{t=1}^{4}(1 + R_{s,t})),$$

With $R_{s,t}$ the expected return on stock s estimated with the market model: $r_{f,t} + \hat{\beta}_s(r_{M,t} - r_{f,t})$, using an estimation window of [-250, -10] days before the event. $r_{s,t}$ is the return of stock s on day t, i the user and, j the change in recommendation event. The market factor is the daily value-weighted return of all NYSE, AMEX, and NASDAQ firms from the Fama/French website.

Certain contributors may "piggyback" on market performance (i.e., they issue positive recommendations when the market goes up and vice versa). To mitigate this concern, I exclude the date on which the recommendation was issued (i.e., day 0). Some recommendations take place before and after trading hours. For a recommendation after trading hours, the day 0 return is the return the following day. For a recommendation before trading hours, the day 0 return is the return on the day of the recommendation.
To estimate users' skills denoted by $\hat{\alpha}_i$, I aggregate the buy and hold abnormal returns to the user level using the average buy and hold abnormal return:

$$\hat{\alpha}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} BHAR_{i,j},$$

with n_i being the number of changes in recommendation j made by user i.

The event window is set to three days. This choice is consistent with Crane and Crotty (2020) using a five-day event window for analysts' recommendations, and Renault (2017), documenting a positive abnormal return the days following a spike in message activity on social media platforms, followed by a significant price reversal during the next five trading days. I provide further justification for the choice of time horizon in a previous study.

- **Prediction Power:**

$$Quality_{i,j} = \begin{cases} 1 & \text{if } \frac{\Delta^{i}_{s,t,j}}{R_{s,t+1}} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

with $\Delta^{i}_{s,t}$ equals +1 for bullish and -1 for bearish recommendation on stock s at time t by user i for recommendation j.

$$PP_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Quality_{i,j}$$

Disagreement Measure

Following the measure in Cookson and Niessner (2020) of cross-group disagreement, i.e., the standard deviation of average sentiment across different investment approaches, I focus on the standard deviation of average sentiment across all the users posting messages on a particular stock on the same day. The disagreement measure for user i's recommendation made at time t with respect to the rest of the users in the network is calculated as:

$$Disagreement_{i,t,s} = \frac{\sqrt{(\Delta^i_{s,t} - AverageSentiment_{s,t})^2}}{2},$$

with $\Delta^i_{s,t}$ equals $+1$ for bullish and -1 for bearish recommendation on stock s at time t by user i, $AverageSentiment_{s,t}$ is the average sentiment of all user for stock s on day t. $Disagreement_{i,t,s}$ equals 0 when the user agrees with the crowd ($\Delta^i_{s,t}$ equals -1 or 1, and the $AverageSentiment_{s,t}$ equals to -1 or 1 respectively) and equals 1 in case of maximum disagreement ($\Delta^i_{s,t}$ equals -1 or 1, and the $AverageSentiment_{s,t}$ equals to 1 or -1 respectively).

Figure 2.1 reflects how the disagreement measure varies for a user i across ten recommendations issued on firm s on the same day. If the user's recommendation is bullish (i.e., $\Delta^i_{s,t}$ =1), the disagreement measure decreases as the number of bullish recommendations posted increases. For a user issuing a bearish recommendation, the disagreement measure increases as the number of bullish messages and recommendations posted increases. The disagreement measure differs from the diversity measure as it does not imply that the users discussing the same stock are connected in the network.

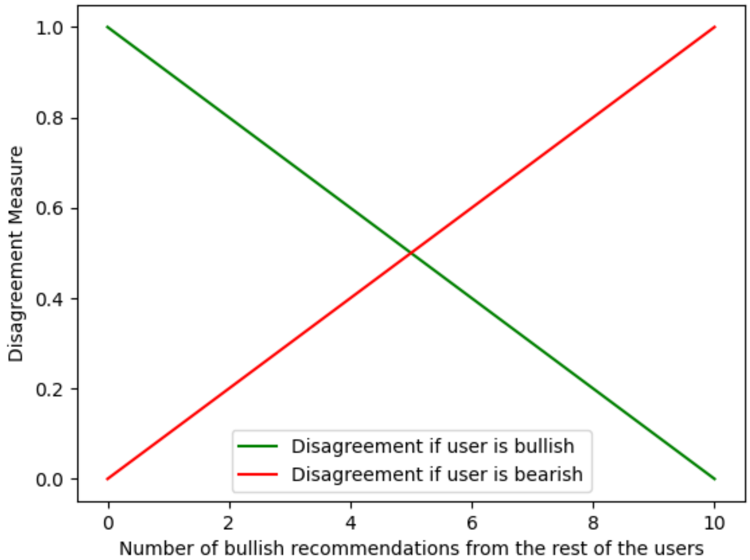


Figure 2.1: Example of Disagreement Measure with 10 Users
This figure illustrates how the disagreement measure of user i depends on the average sentiment of the rest of the users in the network. This example presents the disagreement measure of user i depending on the number of bullish recommendations issued by the rest of the crowd for the case of 10 total messages.

The disagreement measure is aggregated at the user level, i.e., the extent to which a user disagrees on average with the rest of the StockTwits community on all stocks the user i discusses in his tweets.

$$Disagreement_i = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{T}\sum_{t=1}^{T} Disagreement_{i,t,s},$$

where T accounts for the number of recommendations from user i on stock s and S accounts for the number of stocks discussed by user i. Users with higher degrees of disagreement do not align with the crowd. These characteristics are relevant for users with high centrality, given that they are repeatedly confronted with opposing views but decide to make contrasting recommendations. The combination of high centrality and a strong level of disagreement for a user is closer to the wisdom of the crowd hypothesis. In contrast, the combination of high centrality and low levels of disagreement favors the echo chambers hypothesis.

I compute all the variables at the user-month level and the user-stock level. For example, $PP_{i,s}$ is the average prediction power of user i on stock s and $Disagreement_{i,tm}$ is the average disagreement of user i with the rest of the StockTwits user making a recommendation on month tm on all stocks.

Summary Statistics

Table 2.2 presents summary statistics of StockTwits users' related variables at the user-level, user-stock level, and user-month level. The variables are calculated using the recommendations from

326,696 users on S&P500 stocks. Users' average skill $\hat{\alpha_i}$, i.e., the average buy and hold abnormal return following user i recommendations, is positive for all aggregation levels. The highest average user skill amount to 0.186%, being the average skill of users who are issuing recommendations for a particular stock. Month-level values are in the same range as user-level and user-stock level values, reflecting that users' skills and prediction power are consistent across stocks and time.
The users' average prediction power is higher than 0.5 for all three levels (going from 0.514 to 0.519), indicating that, on average, users are more often correct than not at predicting returns over a one-year horizon. These results are economically significant because users make, on average more than 60 recommendations over the length of the database. The average values are consistent with Chen et al. (2014), highlighting prediction power in the aggregated recommendation of social media users. User skills and prediction power exhibit high heterogeneity, with the standard deviation of users' skills ranging from 0.034 to 0.038 and the standard deviation of users' predicting power varying from 0.298 to 0.358.
The cross-sectional 90th percentile user prediction power is 0.9 for all three levels, meaning that 10% of the users (more than 30,000 users) recommendations predict next-day return in all cases. Average user disagreement with the rest of the StockTwits community is slightly lower than 0.5 for all three levels, reflecting that users are, on average more in agreement about their opinion on a stock than in disagreement.

2.3.3 StockTwits Network

A graph $G = (V, E)$ is a set of vertices V and edges E where each edge (i, j) is a connection between vertices where $(i, j) \in V$. In Python, graphs are depicted using the nodes and edges. Nodes in a graph can represent any feature, whereas the edges represent how features interact. To populate the network graphs in this paper, I use NetworkX, a python package, to create and analyze the structure of complex networks.

Financial participants' decisions are significantly influenced by factors like investors' structural relationships with other investors and whether those additional investors have connected. StockTwits data is appropriate for building a finance investors network because it gives access to nodes (the users) and edges (the connection between those users). As edges, I use likes between users and communication (retweet, reply, or mention) between users in their tweets.

The name data of users involved in liking activities is unavailable directly through the StockTwits API. As a result, collecting the data by separately web scrapping each user's page is highly time-consuming (months of data collection). Using like data first ensures that the tweet has been read and each like is associated with a date. Alternatively, since users can tag other users in their tweets, a degree connection can be defined as the communication between users.

Table 2.2: Summary Statistics

This table reports summary statistics of StockTwits users' related variables at the user-level, user-stock level, and user-month level. Variables are calculated using 20'983'763 recommendations by 326'686 users from 2014 to 2022. $\hat{\alpha_i}$ is the average buy and hold abnormal return following user i recommendation, i.e. the average skill of user i at making valuable recommendations. $PP_i \in [0,1]$ is the prediction power of user i. $Disagreement_i \in [0,1]$ is the average disagreement of user i, i.e., the extent to which a user disagrees on average with the rest of the StockTwits community for all stocks user i discusses in his tweets. Variables are aggregated at the user-stock level and user-month level. For example, $PP_{i,s}$ is the average prediction power of user i on stock s and $Disagreement_{i,tm}$ is the average disagreement of user i with the rest of the StockTwits user making a recommendation on month tm on all stocks.

	User Level					
	Mean	**SD**	**q50**	**Min**	**Max**	**q90**
$\hat{\alpha_i}$	0.142%	0.034	0.001%	-59.652%	81.255%	2.762%
PP_i	0.514	0.298	0.500	0.000	1.000	1.000
$Disagreement_i$	0.491	0.138	0.497	0.000	0.947	0.727
	User-Stock Level					
	Mean	**SD**	**q50**	**Min**	**Max**	**q90**
$\hat{\alpha_{i,s}}$	0.186%	0.037	0.000	-60.24%	81.255%	3.130%
$PP_{i,s}$	0.519	0.347	0.500	0.000	1.000	1.000
$Disagreement_{i,s}$	0.473	0.162	0.494	0.000	0.954	0.732
	User-Month Level					
	Mean	**SD**	**q50**	**Min**	**Max**	**q90**
$\hat{\alpha_{i,tm}}$	0.160%	0.038	0.000	-81.125%	81.126%	3.268%
$PP_{i,tm}$	0.517	0.358	0.500	0.000	1.000	1.000
$Disagreement_{i,tm}$	0.484	0.175	0.499	0.000	0.947	0.744

Communication Network

From the 20'983'764 tweets from 326'686 users on S&P500 tickers from January 2014 to May 2022, 1'008'997 tweets from 85'737 users include tagging or responding to other users (1'308'904 communication pairs because a user can communicate with more than one other user in his tweet). Even though the communication data is smaller than the like data, it represents more assertive communication (interpersonal communication) and allows me to compute the sentiment of the communication. On average, a user communicates with 15 other users. Table 2.3 displays examples of tweets with communication.

Table 2.3: Example of Communication Tweets

This table displays examples of tweets with communication. Each tweet is associated with the user's username in the message. For example, for mentioning the username "maximum profits," the tweet is referenced with the usertag "@maximumprofits."

Tagged User	Tweets
@maximumprofits	@maximumprofits you were wrong about $TSLA, you're prolly wrong about this too
@Lowko08	@Lowko08 Bad analogy friend. The truth is $AAPL has yearly revenues in the billions
@SuperBigBoss	@SuperBigBoss $AAPL will see higher highs soon. Earnings will be great.
@NVXI	@NVXI what do you think about $TSLA
@DiamondHandsMatt	@DiamondHandsMatt yeah, all is waste except $TSLA
@ClaranceSale	@ClaranceSale I sold $TSLA at $600 to buy this SPAC. Hopefully that makes you feel better!
@minalex	@minalex it's too late for $NVDA and for me
@StonkRip	@StonkRip You're too late! $TSLA and $NVDA is overstretched, I exited.
@Vol888	$TSLA rsi on the daily has never been this high in the history of the stock @Vol888
@MeetJoeBlackBeard	@MeetJoeBlackBeard Stupid! $TSLA recent spike is not caused by Elon

The network graph is built as follows: I establish an edge to connect two nodes if there is any interaction (retweet, reply, or mention). I use an undirected graph to focus on the existing interaction between two users rather than its orientation. There are 212'184 nodes and 605'840 edges in the graph. The degree of a node is the number of edges that occur on that node. The average degree of the nodes in the graph is 5.7, and the most frequent degree found is 1. The communication network density, i.e., the ratio of actual edges in the network to all possible edges in the network, is 0.002. This number indicates that the network is not dense, meaning there are few connections. From the 212'184 nodes, there are 5'512 connected components. A connected component is a subset of connected nodes, i.e., when a path exists between each subset component.

Figure 2.2 presents a graphical representation of the resulting graph using Gephi software. I use the OpenOrd layout algorithm to place the nodes into the graphic space. This algorithm draws large-scale undirected graphs and better distinguishes clusters (Martin et al. (2011)). Since close nodes are grouped, and remote nodes are spread apart, this layout helps to understand the network's topology. The sub-graphs relates to the graph for the overall network, the sub-network for the nodes with degree above the 50th percentile (more than three connections), the sub-network for the nodes with degree above the 75th percentile (more than seven connections), and the sub-network for the nodes with degree above the 90th percentile (more than 17 connections) respectively. Several nodes lie on the periphery of the networks and show no communication with the rest of the network. Different colors represent different clusters, i.e., constellations of nodes that are more densely connected than the rest of the nodes in the network. Although there are other methods for determining network clusters, modularity is the most popular. A modularity class has a high density with other nodes within its modularity class but a low density with those outsides. Modularity is a measure of relative density in the network, and the modularity score

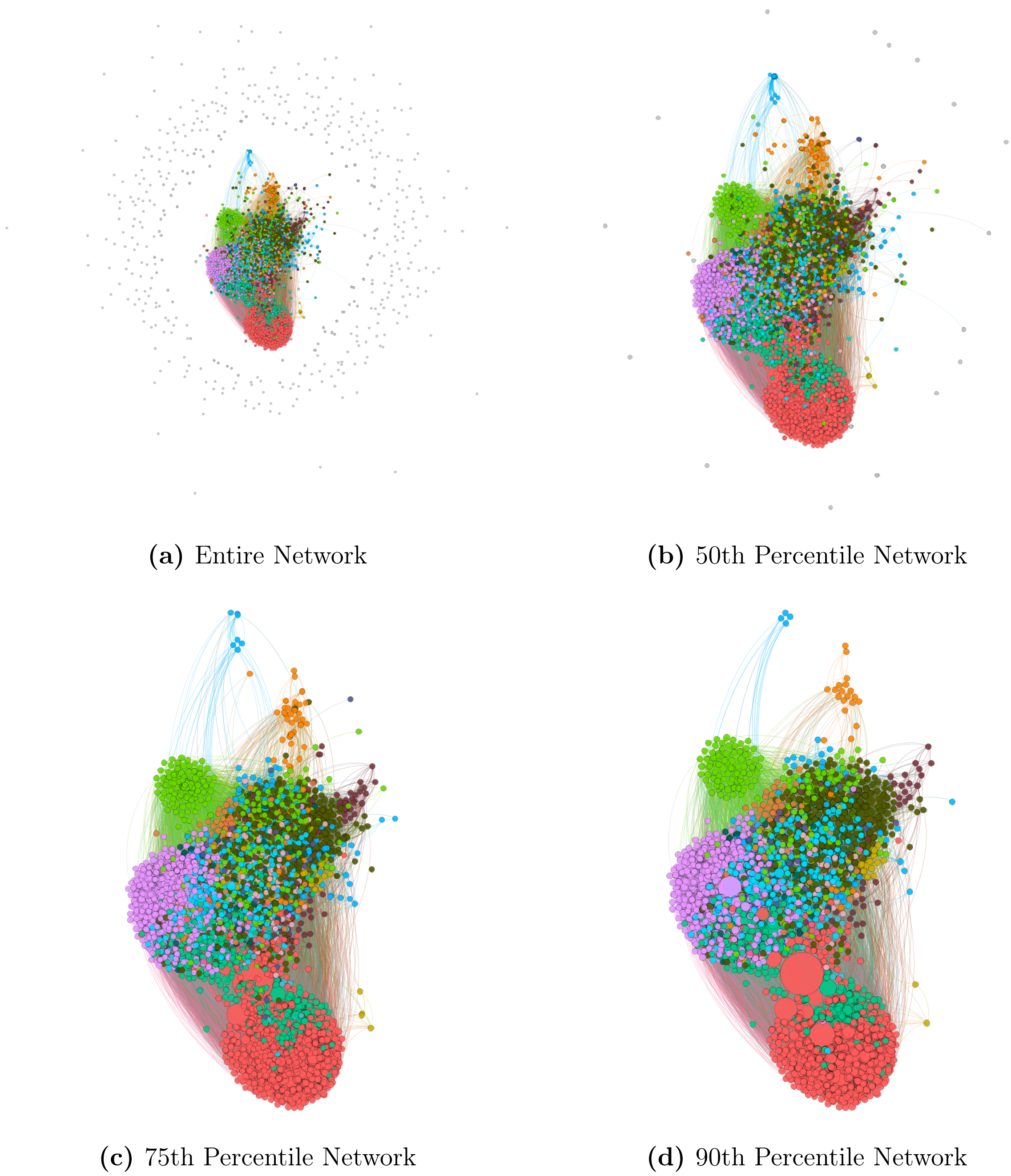
allows the division of the network into its constituent clusters.
The size of a node is proportional to its eigenvector centrality. The StockTwits communication network highlights six large clusters of nodes and a few central users. A small number of users, around 30, communicate with more than 1,000 users, while most of the network communicates with about 3.

Like Network

By downloading all tweets on S&P500 tickers from January 2014 to May 2022, I obtain a list of 326'686 users. From this sample of users, 216'068 engage in liking activities. For each user, I collected their liked list (i.e., the list of other users they liked and the date of the like). A like pair is the username of the user who likes the post, the username of the user who is liked, and the date of the like (with minute accuracy). There are 29'353'406 like pairs in total to build the network of users. On average, a user likes 136 other users' recommendations. The maximum number of likes for one user is 12'757. The median number of likes per user is 20.

The like network graph is built as follows: I establish an edge to connect to nodes if there is a like between them. There are 606'017 nodes and 17'455'534 edges present in the graph. The degree of a node is the number of edges that occur on that node. The average degree of the nodes in the graph is 57.6, and the most frequent degree found is 1. The communication network density, i.e., the ratio of actual edges in the network to all possible edges in the network, is 0.00009, i.e., the network is not dense. From the 606'017 nodes, there are only 199 connected components, consistent with the near zero density.
Figure 2.3 displays a graphical representation of the resulting graph. Compared to the communication network graph, the nodes are drawn far from each other on the network, illustrating the low number of connected components. The sub-graphs relates to the graph for the overall network, the sub-network for the nodes with degree above the 50th percentile (more than eight connections), the sub-network for the nodes with degree above the 75th percentile (more than 24 connections), and the sub-network for the nodes with degree above the 90th percentile (more than 71 connections) respectively. Many nodes cluster lie on the periphery of the networks and show no communication with the central clusters. Different colors represent different clusters, and the size of a node is proportional to its eigenvector centrality. The StockTwits like network overall highlights around eight large clusters of nodes and a few central users for each cluster. Around 15 users like more than 5,000 users, approximately 2,000 users like more than 1,000 users, and most of the network communicates with about eight users.

Overall both communication network and like network show different topological characteristics highlighting the difference in the networks' densities. These differences motivate the interest to study the interactions in the two networks separately to isolate social network features and



(a) Entire Network

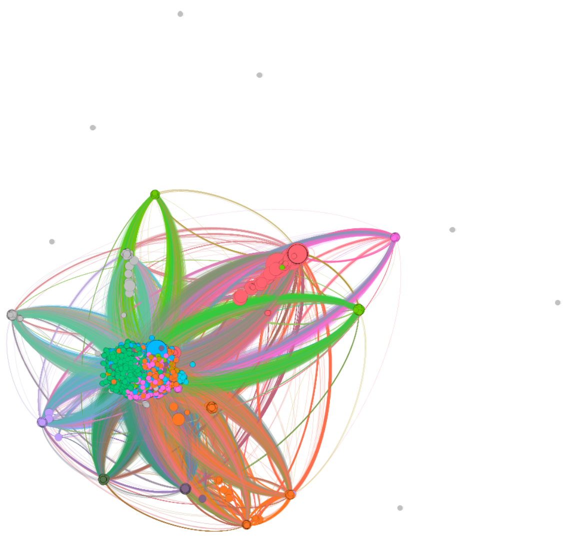
(b) 50th Percentile Network

(c) 75th Percentile Network

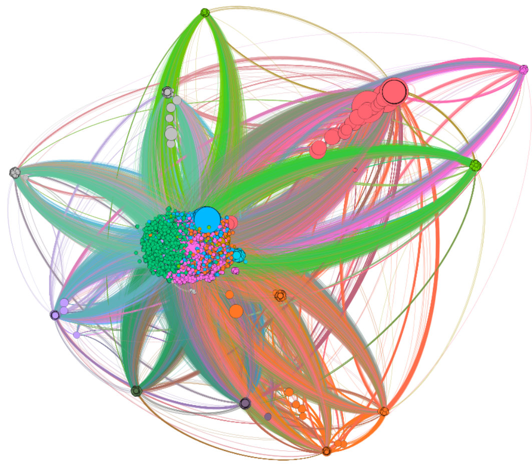
(d) 90th Percentile Network

Figure 2.2: Communication Network Graph

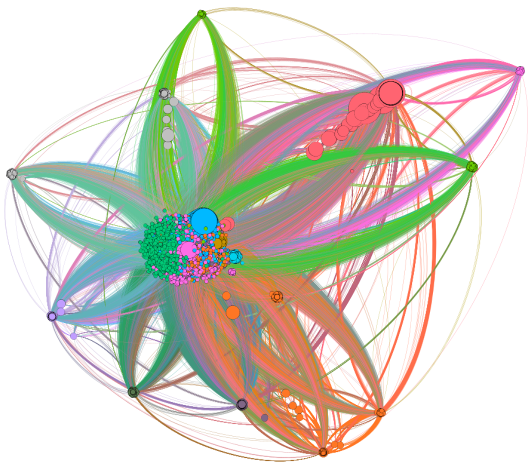
These figures illustrate the network structure of the StockTwits communication network. There are 212'184 nodes and 605'840 edges in the graph. The sub-graphs relate to the graph for the overall network, the sub-network for the nodes with degrees above the 50th percentile (more than three connections), the sub-network for the nodes with degrees above the 75th percentile (more than seven connections), and the sub-network for the nodes with degree above the 90th percentile (more than 17 connections) respectively. Different colors represent different clusters, i.e., constellations of nodes that are more densely connected than the rest of the nodes in the network. The size of a node is proportional to its eigenvector centrality.



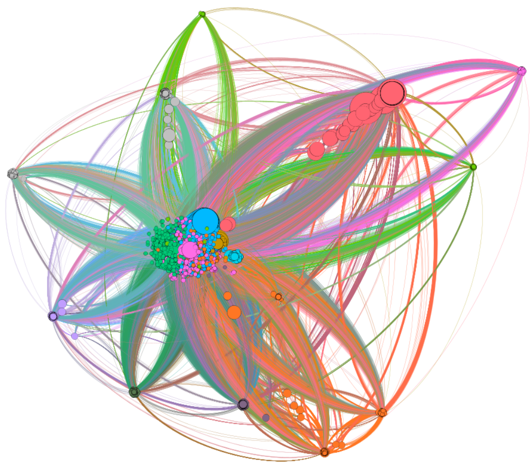
(a) Entire Network



(b) 50th Percentile Network



(c) 75th Percentile Network



(d) 90th Percentile Network

Figure 2.3: Like Network Graph
These figures illustrate the network structure of the StockTwits communication network. There are 606'017 nodes and 17'455'534 edges in the graph. The sub-graphs relate to the graph for the overall network, the sub-network for the nodes with degrees above the 50th percentile (more than eight connections), the sub-network for the nodes with degrees above the 75th percentile (more than 24 connections), and the sub-network for the nodes with degree above the 90th percentile (more than 71 connections) respectively. Different colors represent different clusters, i.e., constellations of nodes that are more densely connected than the rest of the nodes in the network. The size of a node is proportional to its eigenvector centrality.

evaluate how interaction via communication versus like changes how information is spread in the community.

Ticker Sub-network and Month Sub-network

Users tend to focus on particular stocks and issue recommendations on these tickers more than others. To strengthen the analysis, I build ticker sub-networks for communication and like networks, i.e., the sub-network of nodes communicating or liking other nodes per ticker. There are 506 ticker sub-networks for both like and communication networks for all stocks included in the S&P500 index from 2014 to 2022. The StockTwits user base changes over time so does the StockTwits users' network. I build monthly sub-networks, pairing likes and communication between nodes each month. There are 98 monthly sub-networks for both like and communication networks from January 2014 to January 2022.

2.3.4 StockTwits Network Variables

Centrality

Several measures of centrality are widely used in network analysis. They describe either the local connectivity of a node (users) or its global importance. In graph analytics, the centrality concept refers to the significance of a node's position in the network by taking into account: the degree on the premise that an important node will have a larger number of connections, closeness on the premise that important nodes are close to other nodes, and betweenness on the premise that important nodes are strategically placed and connect other nodes. Previous studies show that more central nodes in a network of financial agents have advantages (Hochberg et al. (2007), Cohen et al. (2008), Cohen et al. (2010)).

In the case of StockTwits, ranking users according to their centrality is not just about the size of the audience of users. For example, TESLABULL93 is a StockTwits user; he issues a bullish recommendation on Tesla Stock. The impact of his recommendation is different depending on the criteria used. Let's consider the case where this user's recommendation receives 1000 likes from other random users (evaluation based on the size of the user's network). In the second step, let's consider the case where this user's recommendation receives only 100 likes from other users. Still, one of these users is, in fact, Elon Musk (evaluation based on the user's network characteristics). It is reasonable to think that his recommendation will have more impact in the second case, although the size of his network is smaller than in the first case. This example illustrates that the network measures in this paper, i.e., centrality variables, are not a proxy for size but take into account the relative importance of each user in the network.

It is worth noting that the calculation of the network variables introduced below is extremely time and computationally intensive. For example, for a network of about 100,000 nodes, the computation of the closeness centrality variable takes more than a week on a regular computer. The time complexity of the network centrality algorithms increases with the number of nodes and edges in the network.[7]

To overcome this computational hurdle and compute the network variables for the two leading networks and all sub-networks, I use Amazon EC2 instances from Amazon Web Services (AWS).

- **First-degree Centrality:** First-degree centrality assigns an importance score based on the number of links each node holds, identifying very connected users. This measure is similar to like count. Degree centrality for node i, C_i^D is defined as:

  $$C_i^D = \sum_j m(i,j),$$

  where $m(i,j) = 1$ if a link exists from node i to node j.

- **Betweenness Centrality:** Betweenness centrality measures the number of times a node lies on the shortest path between other nodes or 'bridge nodes' (how many times each node is identified on the shortest paths). Betweenness centrality varies from 0 to 1: the closer to one, the greater the centrality. Nodes with high betweenness centrality are essential for communication and cohesion in a graph because they link many nodes. The network's connectivity or communication would be severely disrupted if these nodes were removed.

  Betweeness centrality for node i, C_i^B is defined as:

  $$C_i^B = \sum_{a \neq b \neq i} \frac{g_{a,b}(i)}{g_{a,b}},$$

  where (a,b) is any pair of nodes in the graph, $g_{a,b}(i)$ is the number of shortest paths from node a to node b passing through i and, $g_{a,b}$ is the number of shortest paths from node a to node b. Betweenness centrality takes longer to calculate than other measures because it must first determine every possible shortest path in the network.

- **Closeness Centrality:** Closeness centrality scores each node based on their 'closeness' to all other nodes in the network (assigns each node a score based on its sum of shortest paths) equals one over the sum of the distance to each other nodes. This measure helps find the individuals best placed to influence the entire network most quickly, i.e., estimates the speed at which the flow of information would be through a given node to other nodes. This measure is therefore interesting to identify the influential users in this paper. Influential users'

[7] For betweenness centrality calculation with the NetworX python package, the time complexity is $O(VE)$ with V the number of nodes and E the number of edges in the network.

recommendations could spill over into decisions made by other market participants, leading to conformity in the network. From the point of view of the network of users communicating with each other, the most significant effect on user performance is due to the ability of users to share their recommendations with the rest of the network quickly. For communication to affect the short-term performance of users, the speed of information delivery is crucial.

To conclude, an increase in user centrality, regardless of the type of centrality, has a positive relationship with current and future user performance. The centrality measures capture significantly more than a side effect of the user's network, which, therefore, better explains the predictive powers of the users.

Closeness centrality for node i, C_i^C is defined as :

$$C_i^C = \sum_j d(i,j),$$

with $d(i,j)$ the number of edges between i and j on the shortest path from i to j if a path exists. $d(i,j)$ corresponds to the geodesic distance between i and j.

- **Eigenvector Centrality:** Eigenvector Centrality evaluates the transitive effect of nodes, a node's score is increased more by connections coming from high-scoring nodes than by connections coming from nodes with low scores. A node with a high eigenvector score is linked to numerous others with high scores. Eigenvector centrality for node i, C_i^{EC} is defined as:

$$C_i^{EC} = \frac{1}{\lambda} \sum_{j \in M(i)} C_j^{EC} = \frac{1}{\lambda} \sum_j m(i,j) C_j^{EC},$$

  where $m(i,j) = 1$ if a link exists from node i to node j, $M(i)$ is the set of neighbors or node i and, λ is a constant.

I perform the same calculations for the ticker sub-networks and the monthly sub-networks. The network variables for the ticker sub-networks and the monthly sub-networks are denoted by $C_{i,s}^D$, $C_{i,s}^B$, $C_{i,s}^C$, $C_{i,s}^{ECC}$ and $C_{i,t}^D$, $C_{i,t}^B$, $C_{i,t}^C$, $C_{i,t}^{ECC}$, respectively.

Spreading capacity

One of the main aspects of social networks is that they allow for spreading information quickly and widely. Potential echo chambers may impact how information transmits through the users' network. I build on the methodology in Cota et al. (2019) to measure a user's spreading capacity. This study uses classical epidemic processes to better understand echo chambers' role in information propagation. The use of their approach can reveal the existence of echo chambers.

The general intuition is that the spreading capacity of a user i is a function of the "set of influence" of node i, F_i. The "set of influence F_i is the set of users that can be reached by a recommendation issued by user i. I use direct connections so that the set of influence of node i corresponds to the degree centrality of node i.
The spreading capacity $Spread_{i,s}$ of user i for stock s sub-network is defined as:

$$Spread_{i,s} = \frac{F_i}{N^s} = \frac{\sum_j m^s(i,j)}{N^s}, \tag{2.1}$$

with N^s the number of nodes in the stock s sub-network and $m^s(i,j) = 1$ if a link exists from node i to node j in the stock s sub-network. To better understand the dynamics of information spreading in financial social networks, I evaluate what type of users have the largest spreading capacity, i.e., skilled versus unskilled users or extreme opinion versus moderate opinion. $Spread_i = \frac{1}{S}\sum_s Spread_{i,s}$ refers to the average spreading capacity across all stocks S for user i.

Homophily

Homophily, or the propensity to associate with people who share our opinions, is a characteristic of social network platforms (Du and Gregory (2016), Cookson et al. (2022)). Since users tend to engage with others who hold the same beliefs as they do, smaller diversity of opinions in the financial social network indicates the presence of homophily. A node i in stock s sub-network is likely to be part of an echo chamber if its average opinion on stock s, Δ_s^i, is very close to the average opinions on stock s of its nearest neighbors, $\mu_{i,s}$. Thus, the diversity of user i for stock s sub-network with respect to its nearest neighbors is computed as follows:

$$Diversity_{i,s}^{user} = |\bar{\Delta}_s^i - \mu_{i,s}|, \tag{2.2}$$

with $\mu_{i,s} = \sum_j m^s(i,j)\frac{\bar{\Delta}_s^j}{F_i}$ the average opinions on stock s of node i's nearest neighbors. and $\bar{\Delta}_s^j \in (-1,1)$ the average opinion of user i on stock s. $Diversity_i^{user} = \frac{1}{S}\sum_s Diversity_{i,s}^{user}$ refers to the average diversity for user i across all stock sub-networks S.

Diversity of the crowd reached by the user

Another critical factor in the dynamics of financial social media is evaluating the diversity of the crowd the user reaches when issuing a recommendation. To support the hypothesis of the wisdom of the crowd from an individual point of view, it is not sufficient for the diversity among one's neighbors to be large but for the diversity among the neighbors themselves to be large.

The diversity of the neighbors reached by node i for stock s sub-network is computed as follows:

$$Diversity_{i,s}^{neighbors} = \sum_j m^s(i,j) \frac{(\bar{\Delta}_s^j - \mu_{i,s})^2}{F_i}, \tag{2.3}$$

with $\mu_{i,s} = \sum_j m^s(i,j)\frac{\bar{\Delta}_s^j}{F_i}$, the average learning of the users reached by node i, $m^s(i,j) = 1$ if a link exists from node i to node j in the stock s sub-network, F_i is the number of users that can be directly reached by a recommendation issued by user i and $\bar{\Delta}_s^j \in (-1, 1)$ the average opinion of user j on stock s. If $\bar{\Delta}_s^j$ equals -1 or 1, the user is extremely bearish or bullish on stock s, respectively. If $Diversity_{i,s}^{neighbors}$ is small, the crowd reached by node i is slightly diverse and a larger $Diversity_{i,s}^{neighbors}$ indicates higher diversity. In other words, the smaller $Diversity_{i,s}^{neighbors}$ relates to stronger echo chambers (Cota et al. (2019)) if the recommendations on stock s are either extremely bullish or bearish. $Diversity_i^{neighbors} = \frac{1}{S}\sum_s Diversity_{i,s}^{neighbors}$ refers to the average diversity of the neighbors reached by node i across all stock sub-networks S.
The diversity of neighbors has key implications for this study because, according to the social dilemma introduced by Frey and Van de Rijt (2021), the quality of users' decisions increases when they are around a group that thinks differently.

Summary Statistics

Table 2.4 presents summary statistics of the StockTwits Network Variables for both like and communication networks. Unsurprisingly, the average values of centrality are low, in line with the two networks' subsequent size and low density. The density and the number of connected components are low for both networks and even more so for the like network as observable on the graphical representations of the two networks in Figures 2.3 and 2.2. Summary statistics on users' degree centrality C_i^D are higher for the like network. This result largely results from the network size, as the degree centrality is simply a sum of the directly connected nodes. Concerning the other centrality measures, i.e., C_i^B and C_i^{EC}, users communicating with each other are more central on average. The communication network has a larger fraction of bridge nodes and influential nodes. For the subsequent three variables, i.e. $Spread_{i,s}$, $Diversity_{i.s}^{user}$ and, $Diversity_{i.s}^{neighbors}$, summary statistics are calculated at the sub-network level s. This choice matches the idea that the diversity of a user with respect to his neighbors can only be determined if the topic of conversation belongs to the same ticker. The average spreading capacity for user i in sub-network s is 0.037 for the communication network and 0.029 for the like network. Unsurprisingly, expressing ideas by communicating with other users has, on average, a higher spreading capacity on the network than simply liking other users' ideas.
Building network ties through communication is associated with lower diversity than building network connections through like. The maximum achievable diversity of a user compared to its neighbors is 2.00, i.e., when the user is extremely bullish or bearish about stock, s ($\Delta_s^i = 1$ or

$\Delta_s^i = -1$) and the user's neighbors are all extremely bearish or bullish about stock s respectively. The average diversity of a user with respect to its neighbors on stock s is 0.666 for the communication network and 0.749 for the like network. The results point in the same direction, the average diversity of the neighbors reach by node i for stock s sub-network is 0.114 for the communication network and 0.136 for the like network. 90% of the diversity of the neighbors reached by nodes i is lower than 0.235 for the communication network and 0.254 for the like network. Considering that the maximum diversity attainable is 1.00, these values indicate that the diversity of the neighbors reached by nodes i is particularly low in financial social networks.

Overall, building network connections by communicating seems stronger than liking other users due to a more significant number of influential nodes, increasing the ability to spread investment ideas and lowering diversity in investors' opinions about a particular stock.

Table 2.4: Summary Statistics: StockTwits Network Variables

This table reports summary statistics of StockTwits users' network variables for the communication network and the like network. Variables are calculated from 606'017 nodes for the like network and 212'184 for the communication network. C_i^D, C_i^B, C_i^C and C_i^{EC} represents degree, betweenness, closeness and eigenvector centrality of user i respectively. $Spread_{i,s}$ is the spreading capacity of user i for stock s sub-network, $Diversity_{i,s}^{user}$ is the diversity of user i for stock s sub-network with respect to its nearest neighbors and, $Diversity_{i,s}^{neighbors}$ is the diversity of the neighbors reach by node i for stock s sub-network.

	Communication Network						Like Network					
$Nodes$	212'184						606'017					
$Edges$	605'840						17'455'534					
$Density$	0.002						0.00009					
$ConnComp$	5'512						199					
	Mean	**SD**	**q50**	**Min**	**Max**	**q90**	**Mean**	**SD**	**q50**	**Min**	**Max**	**q90**
$C_i^D \times 100$	0.003	0.014	0.0004	0.0004	2.989	0.005	0.009	0.033	0.002	0.0001	2.150	0.019
$C_i^B \times 100$	0.001	0.029	0.000	0.000	9.199	0.001	0.0004	0.005	0.000	0.000	1.544	0.0002
$C_i^C \times 100$	19.462	5.375	20.505	0.000	33.972	23.842	30.584	2.618	30.375	0.000	47.186	34.064
$C_i^{EC} \times 100$	0.0004	0.002	0.0002	0.0001	0.404	0.001	0.0001	0.0004	0.000	0.000	0.038	0.0003
$Spread_{i,s}$	0.037	0.110	0.003	0.000	1.000	0.087	0.029	0.097	0.002	0.000	1.000	0.056
$Diversity_{i,s}^{user}$	0.666	0.348	0.723	0.000	2.000	1.000	0.749	0.305	0.827	0.000	2.000	1.000
$Diversity_{i,s}^{neighbors}$	0.114	0.106	0.112	0.000	1.000	0.235	0.136	0.099	0.134	0.000	1.000	0.254

2.4 Results

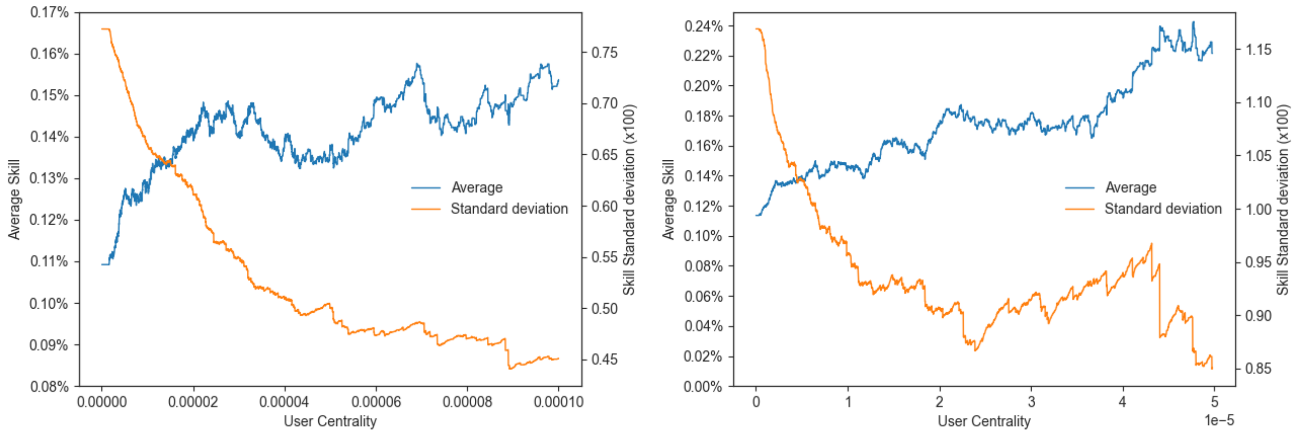
2.4.1 Relationship Between User Performance and Network Variables

In the finance literature, studies have shown that a higher centrality is positively associated with better performance (Cohen et al. (2010), Cortez et al. (2016)). This section of the study specifically examines the impact of a user's position within the social network on their ability to gather superior information about future stock returns. Given the potential spillover effects of influential user recommendations on the decisions of other market participants, a primary objective of this research is to investigate how an investor network's structure influences its participants' predictive power. The predictive ability of a StockTwits user is assessed using two variables: User skill, defined as the three-day average buy and hold return following a user's recommendation, and user average prediction power for the next-day stock return.

Skill

This subsection investigates the association between network centrality and investors' ability to predict future stock returns over three days. Goutte (2020) find that user skill, measured by the three-day average buy-and-hold return following a recommendation, is positively related to the number of followers. However, the number of followers only measures the size, whereas network variables can reveal a user's position in the network and their influence on others. By analyzing the predictive performance of the most influential users or those who spread information widely, I can evaluate the quality of the information in large financial networks.

To further explore the relationship between user centrality and skill, I examine the evolution of cross-sectional statistics of average user skill $\hat{\alpha}_i$ following user recommendations in relation to minimum user centrality. Figures 2.4a and 2.4b illustrate that an increase in user centrality corresponds with an increase in average user skill by up to 0.22% for the communication network and 0.16% for the like network. Moreover, the cross-sectional standard deviation of user skill decreases with minimum user centrality for both communication and like networks, suggesting a reduction in the heterogeneity of users' performances with centrality.



(a) Like Network **(b)** Communication Network

Figure 2.4: Evolution of Users Skills With Minimum User Centrality
These figures plot the evolution of cross-sectional statistics of average user skill $\hat{\alpha}_i$ following users' recommendations as a function of minimum user centrality. Skills are average buy-and-hold abnormal returns derived from buying stocks following a recommendation upgrade and selling stocks following a recommendation downgrade. The chosen measure is eigenvector centrality, C_i^{EC}. A high eigenvector score refers to a node connected to many high-score nodes. Variables are calculated from 606'017 nodes for the like network and 212'184 for the communication network.

To evaluate the relationship between investors' network centrality and investors' skills, I use the following specification:

$$\hat{\alpha}_i = \alpha + \gamma X_i + \epsilon_i, \tag{2.4}$$

$$\text{with } X_i = (C_i^D, C_i^B, C_i^C, C_i^{EC}).$$

C_i^D, C_i^B, C_i^C and C_i^{EC} represent degree, betweenness, closeness and eigenvector centrality of user i respectively. The variables are standardized to facilitate comparison. The estimates measure the average user's skill change corresponding to a one-standard-deviation change in the independent variable.

Table 2.5 reports the regression results of user centrality measures on average user skill for the like and communication networks. The first column of the like network shows a significant positive relationship between user degree centrality and the average user skill. Column two depicts that an increase in user betweenness centrality corresponds to an increase in user skill. Column four indicates that user eigenvector centrality has the most substantial significant impact on the average user's skill. A one-standard-deviation change in the eigenvector centrality of user i results in a 1.6% change in user i's three-day average buy and hold abnormal returns. In the communication network, although all user centrality measures positively impact the average user's skill, the coefficients are lower and less significant. The strongest relationship with user skill in the communication network is observed with the closeness centrality. A one-standard-deviation change in the closeness centrality of user i leads to a 1.3% change in user i's skill. The results demonstrate that more sophisticated centrality measures identify the most skilled users of a network rather than being just a scaling effect. Users with the potential to influence the entire network most quickly demonstrate more significant increases in their average skill than users with more network connections.

Table 2.5: User Skill and Centrality: Like Network and Communication Network

The table presents parameter estimates from OLS regression on $\hat{\alpha}_i$, the average skill of StockTwits user i with several different specifications.C_i^D, C_i^B, C_i^C and C_i^{EC} represent degree, betweenness, closeness and eigenvector centrality of user i respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Like Network	(1)	(2)	(3)	(4)
	$\hat{\alpha}_i$	$\hat{\alpha}_i$	$\hat{\alpha}_i$	$\hat{\alpha}_i$
C_i^D	0.009**			
	(2.417)			
C_i^B		0.009***		
		(2.778)		
C_i^C			0.019***	
			(3.604)	
C_i^{EC}				0.016***
				(3.111)
N	93,357			
$F - stat$	5.842	7.716	12.990	9.675

Communication Network	(1)	(2)	(3)	(4)
	$\hat{\alpha}_i$	$\hat{\alpha}_i$	$\hat{\alpha}_i$	$\hat{\alpha}_i$
C_i^D	0.008**			
	(2.220)			
C_i^B		0.003		
		(1.212)		
C_i^C			0.013*	
			(1.940)	
C_i^{EC}				0.0078**
				(2.223)
N	65,135			
$F - stat$	4.927	1.470	3.765	4.944

Prediction Power

This subsection examines the association between the network centrality of investors and their ability to forecast future stock returns within a one-day horizon.

To evaluate the relationship between investors' network centrality and prediction power, I use the following specification:

$$PP_i = \alpha + \gamma X_i + \epsilon_i, \tag{2.5}$$

$$\text{with } X_i = (C_i^D, C_i^B, C_i^C, C_i^{EC}).$$

C_i^D, C_i^B, C_i^C and C_i^{EC} represent degree, betweenness, closeness and eigenvector centrality of user i respectively. The variables are standardized to facilitate comparison. The estimates measure the average user's skill change corresponding to a one-standard-deviation change in the independent variable.

Table 2.6 reports the regression results for the like and communication networks. For the like network, all measures of user centrality are positively and significantly related to the user's predictive power. As with the user skills analysis, eigenvector centrality has the greatest positive relationship with user predictive power, with a one-standard-deviation change in user eigenvector centrality corresponding to a 1.2% change in user i's prediction power. The results point in the same direction for the communication network but are less significant. Column one shows a significant positive relationship between the degree centrality of user i and the user's prediction power. In contrast, column three shows a positive and significant relationship between the eigenvector centrality of user i and the user's prediction power. A one-standard-deviation change in the degree and eigenvector centrality of user i corresponds to a 0.5% change in user i's prediction power.

Overall, although the summary statistics of the network variables were higher for the communication network, the analysis suggests that an increase in centrality in the like network has a greater effect on the predictive abilities of users. These findings align with previous research on mutual fund performance (Cohen et al. (2008), Coval and Moskowitz (2001)) and analysts' performance (Cohen et al. (2010)), which suggest that social media users with higher centrality tend to have better performance.

Table 2.6: User Prediction Power and Centrality: Like Network and Communication Network

The table presents parameter estimates from OLS regression on PP_i, the average prediction power of StockTwits user i with several different specifications. C_i^D, C_i^B, C_i^C and C_i^{EC} represents degree, betweenness, closeness and eigenvector centrality of user i respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Like Network	(1) PP_i	(2) PP_i	(3) PP_i	(4) PP_i
C_i^D	0.010**			
	(2.525)			
C_i^B		0.006**		
		(2.188)		
C_i^C			0.022***	
			(4.194)	
C_i^{EC}				0.012***
				(3.368)
N	93,357			
$F - stat$	6.374	4.788	17.590	11.350

Communication Network	(1) PP_i	(2) PP_i	(3) PP_i	(4) PP_i
C_i^D	0.005*			
	(1.909)			
C_i^B		0.002		
		(1.007)		
C_i^C			0.010	
			(1.528)	
C_i^{EC}				0.005*
				(1.835)
N	65,135			
$F - stat$	3.643	1.161	2.334	3.367

Analysis of monthly sub-network

To address potential endogeneity bias, the present study regresses past month centrality on the user performance measures for month t. Furthermore, this section investigates whether higher

initial performance leads users to become more central in the network or if greater centrality improves user performance. I use a one-month horizon to have sufficient user recommendations and enough nodes in the monthly sub-networks. In addition, using a monthly horizon limits the persistence in the time-series data.
To evaluate the relationship between investors' past month network centrality and current month user performance, I use the following specifications:

$$\begin{aligned} \hat{\alpha_{i,t}} &= \alpha + \gamma X_{i,t-1} + \epsilon_{i,t}, \\ PP_{i,t} &= \alpha + \gamma X_{i,t-1} + \epsilon_{i,t}, \end{aligned} \tag{2.6}$$

$$\text{with } X_{i,t-1} = (C^D_{i,t-1}, C^B_{i,t-1}, C^C_{i,t-1}, C^{EC}_{i,t-1}, Like_{i,t-1}).$$

$Like_{i,t-1}$ represents the number of like collected from user i in month $t-1$. All the variables in the regression are standardized.

Table 2.7 and Table 2.8 display the month t user skill and prediction power results, respectively. The past centrality measures prove to be more effective in explaining user skills and prediction power for both networks and user performance variables than like counts. The past eigenvector centrality has the strongest positive and significant relationship with the current user's skill, as in column 4 of Table 2.7, for both the like and communication networks. A one-standard-deviation change in the eigenvector centrality of user i corresponds to a 1.00% change in user i's average skill within the like network. For the communication network, the past closeness centrality measure has the strongest relationship with users' prediction power, resulting in a 1.37% increase per standard deviation unit of centrality.

The main results remain consistent for the networks in period t and the monthly sub-networks in time $t-1$. The like network shows the most significant improvement in user performance for an increase in eigenvector centrality. Users with a higher eigenvector centrality can exert greater influence over the network and are more likely to receive recommendations from highly connected users. The Durbin-Watson test statistics indicate the absence of autocorrelation.

Table 2.7: User Skill and Past Centrality: Like Network and Communication Network

The table presents parameter estimates from OLS regression on $\hat{\alpha_{i,t}}$, the average skill of StockTwits user i at month t with several different specifications.$C^D_{i,t-1}$, $C^B_{i,t-1}$, $C^C_{i,t-1}$ and $C^{EC}_{i,t-1}$ represent previous month degree, betweenness, closeness and eigenvector centrality of user i respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Like Network	(1)	(2)	(3)	(4)	(5)
	$\hat{\alpha_{i,t}}$	$\hat{\alpha_{i,t}}$	$\hat{\alpha_{i,t}}$	$\hat{\alpha_{i,t}}$	$\hat{\alpha_{i,t}}$
$C^D_{i,t-1}$	0.006*				
	(1.797)				
$C^B_{i,t-1}$		0.009***			
		(4.012)			
$C^C_{i,t-1}$			0.009*		
			(1.677)		
$C^{EC}_{i,t-1}$				0.010***	
				(3.113)	
$Like_{i,t-1}$					0.006
					(1.486)
N	295,858				
$F-stat$	3.228	16.100	2.813	9.691	2.207
$DW-stat$	1.965	1.965	1.956	1.965	1.965

Communication Network	(1)	(2)	(3)	(4)	(5)
	$\hat{\alpha_{i,t}}$	$\hat{\alpha_{i,t}}$	$\hat{\alpha_{i,t}}$	$\hat{\alpha_{i,t}}$	$\hat{\alpha_{i,t}}$
$C^D_{i,t-1}$	0.005				
	(1.299)				
$C^B_{i,t-1}$		-0.001			
		(-0.190)			
$C^C_{i,t-1}$			0.011		
			(1.561)		
$C^{EC}_{i,t-1}$				0.006 *	
				(1.651)	
$Like_{i,t-1}$					-0.0006
					(-0.107)
N	155,488				
$F-stat$	0.600	0.036	2.438	2.726	0.011
$DW-stat$	1.935	1.935	1.935	1.935	1.935

Table 2.8: User Prediction Power and Past Centrality: Like Network and Communication Network

The table presents parameter estimates from OLS regression on $PP_{i,t}$, the average prediction power of StockTwits user i on month t with several different specifications. $C^{D}_{i,t-1}$, $C^{B}_{i,t-1}$, $C^{C}_{i,t-1}$ and $C^{EC}_{i,t-1}$ represent previous month degree, betweenness, closeness and eigenvector centrality of user i respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

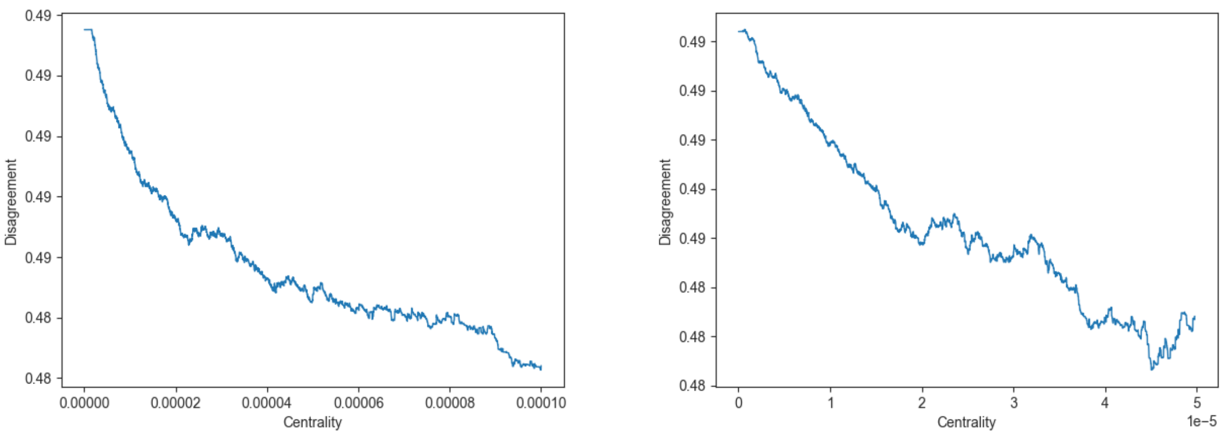
Like Network	(1)	(2)	(3)	(4)	(5)
	$PP_{i,t}$	$PP_{i,t}$	$PP_{i,t}$	$PP_{i,t}$	$PP_{i,t}$
$C^{D}_{i,t-1}$	0.005				
	(1.371)				
$C^{B}_{i,t-1}$		0.009***			
		(3.515)			
$C^{C}_{i,t-1}$			0.006		
			(1.134)		
$C^{EC}_{i,t-1}$				0.008**	
				(2.066)	
$Like_{i,t-1}$					0.005
					(1.192)
N	295,858				
$F-stat$	1.878	12.360	1.286	4.269	1.420
$DW-stat$	1.966	1.966	1.921	1.966	1.966

Communication Network	(1)	(2)	(3)	(4)	(5)
	$PP_{i,t}$	$PP_{i,t}$	$PP_{i,t}$	$PP_{i,t}$	$PP_{i,t}$
$C^{D}_{i,t-1}$	0.0002				
	(0.037)				
$C^{B}_{i,t-1}$		-0.004			
		(-0.789)			
$C^{C}_{i,t-1}$			0.013 **		
			(1.880)		
$C^{EC}_{i,t-1}$				0.0002	
				(0.035)	
$Like_{i,t-1}$					0.003
					(0.441)
N	155,488				
$F-stat$	0.001	0.621	2.438	0.001	0.194
$DW-stat$	1.934	1.934	1.934	1.934	1.934

2.4.2 Relationship between Disagreement and Network Variables

The previous section provides evidence that more influential users in social networks perform better in making valuable recommendations about future movements in financial markets. However, prior research suggests that this influence comes at a cost to the diversity of opinions on online platforms. Influential users can create cascades and herding behaviors among the rest of the community. Section 2.3.2, presents a measure of user disagreement, which captures the extent to which a user disagrees on average with the rest of the StockTwits community on all stocks the user discusses in their tweets.

Figures 2.5a and 2.5b show how average user disagreement $Disagreement_i$ evolves as a function of minimum user centrality for the communication network and the like network, respectively. Both figures reveal a negative relationship between average disagreement and higher centrality user groups, supporting the hypothesis that the plurality of opinions decreases as one considers the most central groups of users.



(a) Like Network **(b)** Communication Network

Figure 2.5: Evolution of Users Disagreement With Minimum User Centrality

These figures plot the evolution of average user disagreement $Disagreement_i$ as a function of minimum user centrality. $Disagreement_i$ is the extent to which a user disagrees on average with the rest of the StockTwits community on all stocks user i discusses in his tweets. The chosen measure is eigenvector centrality, C_i^{EC}. A high eigenvector score refers to a node connected to many high-score nodes. Variables are calculated from 606'017 nodes for the like network and 212'184 for the communication network.

To further analyze the relationships between the different measures of user centrality with the extent to which a user disagrees on average with the rest of the StockTwits community on all stocks user i, $Disagreement_i$, I use the following specification:

$$Disagreement_i = \alpha + \gamma X_i + \epsilon_i, \tag{2.7}$$

$$\text{with } X_i = (C_i^D, C_i^B, C_i^C, C_i^{EC}).$$

C_i^D, C_i^B, C_i^C and C_i^{EC} represent degree, betweenness, closeness and eigenvector centrality of user i respectively. The variables are standardized.

Table 2.9 presents the regression results for the like and communication networks, respectively, which reinforce the interpretations of the figures. In both networks, there is a strong and significant negative relationship between user centrality and the average user disagreement, with the eigenvector centrality showing the strongest association. Precisely, a one-standard-deviation change in the eigenvector centrality of a user corresponds to a 3.30% and 3.00% decrease in the user's average disagreement for the like and communication networks, respectively.

In light of these results, I aim to differentiate whether the findings of this section reflect behavioral biases or a global shared trend in social media users' recommendations. It is worth acknowledging that although network connections allow for differentiation between these possibilities, the rest of the network may not have visibility into the recommendations of central users. Therefore, the subsequent section focuses on the diversity of opinions among users linked directly to the more central users.

Table 2.9: User Disagreement and Centrality: Like Network and Communication Network

The table presents parameter estimates from OLS regression on $Disagreement_i$ with several different specifications. $Disagreement_i$ is the extent to which a user disagrees on average with the rest of the StockTwits community on all stocks user i discusses in his tweets. C_i^D, C_i^B, C_i^C and C_i^{EC} represents degree, betweenness, closeness and eigenvector centrality of user i respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Like Network	(1)	(2)	(3)	(4)
	Dis_i	Dis_i	Dis_i	Dis_i
C_i^D	-0.020 ***			
	(-3.884)			
C_i^B		-0.029***		
		(-5.723)		
C_i^C			-0.006	
			(-1.082)	
C_i^{EC}				-0.033 ***
				(-7.322)
N	93,357			
$F-stat$	15.090	32.760	1.172	53.610

Communication Network	(1)	(2)	(3)	(4)
	Dis_i	Dis_i	Dis_i	Dis_i
C_i^D	-0.027***			
	(-3.902)			
C_i^B		-0.014*		
		(-1.674)		
C_i^C			-0.032 ***	
			(-4.347)	
C_i^{EC}				-0.0301***
				(-4.353)
N	65,135			
$F-stat$	15.220	2.802	18.900	18.950

2.4.3 Wisdom of the Crowd versus Homophily

Investors may gain valuable knowledge and insights by sharing information through social interactions. However, influential users who propagate information through word-of-mouth can distort information and cause it to be misleading, thus harming market efficiency (Della Rossa et al. (2020)).

For the wisdom of crowds to prevail, it is crucial to have diversity in opinions and aggregate opinions appropriately, avoiding the most extreme opinions. Under such circumstances, each user's knowledge contributes, and individual errors cancel out over time, leading to the majority opinion being more accurate than individual users' opinions. Furthermore, diversity within a network enhances future crowd performance (Hong et al. (2020)).

On the other hand, homophily, which refers to a lower diversity of opinions, leads to biased aggregation of opinions as users tend to overweight opinions of users closer within the network. This behavioral bias is more pronounced for unsophisticated investors within social media platforms (Eaton et al. (2021)). Additionally, users tend to consume information from a similar view about the stock market in the presence of homophily (Cookson et al. (2022)). Homophily could result in a central user introducing bias in the beliefs of the majority of users in the network.

The focus of this section is to distinguish between the two competing hypotheses. The diversity of the most central users with respect to their neighbors and the diversity of the users surrounding them provide valuable insights into the sources of the central users' performance results presented above. This section uses variables at the user network and the ticker sub-network levels.

User Network

The following specification investigates the relationship between the diversity of StockTwits users' average opinions and the centrality of users within the entire network:

$$Y_i = \alpha + \gamma X_i + \epsilon_i, \tag{2.8}$$

$$\text{with } Y_i = (Div_i^{neighbor}, Div_i^{user}) \text{ and, } X_i = (C_i^D, C_i^B, C_i^C, C_i^{EC}).$$

Div_i^{user} is the average diversity of user i with respect to its nearest neighbors. $Div_i^{neighbor}$ is the average diversity of the neighbors reached by node i. C_i^D, C_i^B, C_i^C and C_i^{EC} represent degree, betweenness, closeness and eigenvector centrality of user i respectively. The variables are standardized.

Table 2.10 reports the results for the like and communication networks. For the like network, columns 1 to 4 show a strongly significant negative relationship between user i centrality and the average diversity of the nodes reached by user i. The most substantial negative relationship is observed for user eigenvector centrality, where a one-standard-deviation increase corresponds to a 1.7% decrease in the average diversity of the users liked by user i. In contrast, columns 5 to 8 reveal a significant positive relationship between user centrality and the average diversity of user i with respect to its nearest neighbors. For the communication network, the results are the opposite. Columns 1 to 4 exhibit a strongly significant positive relationship between user centrality and the average diversity of the nodes reached by user i. The strongest relationship pertains to user closeness centrality, where a one-standard-deviation increase corresponds to a 15.2% increase in the average diversity of the nodes reached by user i. In contrast, columns 5 to 7 indicate a significant negative relationship between user centrality and the average diversity of user i with respect to its nearest neighbors. The strongest negative relationship pertains to the closeness centrality of users, where a one-standard-deviation increase corresponds to a 15.3% decrease in the average diversity of user i with respect to its nearest neighbors.

The results differ across networks. Specifically, for the like network, being more central implies a greater diversity of the user with respect to their neighbors and a smaller diversity of the neighbors among themselves. On the other hand, for the communication network, an increase in the closeness centrality of users leads to a significant decrease in the diversity of users with their neighbors and a large increase in the diversity of users among themselves. When considering all the network users, verifying or disproving the wisdom of the crowd's hypothesis is impossible. However, given that not all users are equally important, performing the same analysis on the most central users is more informative as they have the most weight in the network.

This analysis now turns to focus on the 10% most central users in both the like and communication networks. The results from OLS regression on Div_i^{user} and $Div_i^{neighbor}$ for these users are presented in Table 2.11, and they are similar for both networks. Specifically, for the like network, the results indicate a negative and significant relationship between user centrality and the average diversity of the nodes reached by the user as well as between the user's centrality and the average diversity of the user with respect to its nearest neighbors. On the other hand, for the communication network, columns 1 to 4 display a positive and strongly significant relationship between user i and the average diversity of the nodes reached by user i, with the strongest relationship corresponding to the user eigenvector centrality, where a one-standard-deviation change in the eigenvector centrality of user i corresponds to a 10.5% increase in the average diversity of the nodes reached by user i. Columns 5 to 8 illustrate a significant negative relationship between the user centrality and the average diversity of user i with respect to its nearest neighbors.

When considering the diversity of all users, the results are mixed depending on the network type. The diversity of opinions is negatively impacted in both networks, whether users communicate or create network ties by liking each other. However, when only the most central users are considered, there is no indication of the wisdom of the crowd hypothesis.

Possible explanations for the reduction in diversity include the belief among users that influential users have superior information. With eigenvector centrality, a high score refers to central users themselves connected with central users. Consequently, the reduction in diversity may create an informational cascade where all the connected users express recommendations in tandem. Another possible explanation is that users fear underperforming their peers and therefore tend to copy their recommendations.

Table 2.10: Diversity and Centrality: Like Network and Communication Network

The table presents parameter estimates from OLS regression on Div_i^{user} and $Div_i^{neighbor}$ with several different specifications. Div_i^{user} is the average diversity of user i with respect to its nearest neighbors. $Div_i^{neighbor}$ is the average diversity of the neighbors reached by node i. C_i^D, C_i^B, C_i^C and C_i^{EC} represents degree, betweenness, closeness and eigenvector centrality of user i respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Like Network

	(1) $Div_i^{neighbor}$	(2) $Div_i^{neighbor}$	(3) $Div_i^{neighbor}$	(4) $Div_i^{neighbor}$	(5) Div_i^{user}	(6) Div_i^{user}	(7) Div_i^{user}	(8) Div_i^{user}
C_i^D	-0.009***				0.026***			
	(-6.040)				(15.816)			
C_i^B		-0.013***				0.005***		
		(-8.772)				(3.338)		
C_i^C			-0.0036				-0.1050***	
			(-1.368)				(-40.779)	
C_i^{EC}				-0.017***				0.016***
				(-10.951)				(9.767)
N	215,836							
$F-stat$	36.480	76.950	1.872	119.900	250.200	11.140	1663.000	95.390

Communication Network

	(1) $Div_i^{neighbor}$	(2) $Div_i^{neighbor}$	(3) $Div_i^{neighbor}$	(4) $Div_i^{neighbor}$	(5) Div_i^{user}	(6) Div_i^{user}	(7) Div_i^{user}	(8) Div_i^{user}
C_i^D	0.033***				-0.024***			
	(4.314)				(-4.647)			
C_i^B		0.007***				-0.003		
		(2.562)				(-1.528)		
C_i^C			0.152***				-0.153***	
			(32.633)				(-31.821)	
C_i^{EC}				0.016***				0.003
				(3.811)				(0.969)
N	85,735							
$F-stat$	18.610	6.562	1065	14.530	21.600	2.336	1013	0.938

Table 2.11: Diversity and Centrality Most Central Users (Top 10%): Like Network and Communication Network

The table presents parameter estimates from OLS regression on Div_i^{user} and $Div_i^{neighbor}$ for the 10% most central users with several different specifications. Div_i^{user} is the average diversity of user i with respect to its nearest neighbors. $Div_i^{neighbor}$ is the average diversity of the neighbors reached by node i. C_i^D, C_i^B, C_i^C and C_i^{EC} represents degree, betweenness, closeness and eigenvector centrality of user i respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Like Network

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Div_i^{neighbor}$	$Div_i^{neighbor}$	$Div_i^{neighbor}$	$Div_i^{neighbor}$	Div_i^{user}	Div_i^{user}	Div_i^{user}	Div_i^{user}
C_i^D	-0.128***				-0.082***			
	(-6.533)				(-3.786)			
C_i^B		-0.144***				-0.039		
		(-6.013)				(-1.396)		
C_i^C			-0.190***				-0.146***	
			(-27.398)				(-19.018)	
C_i^{EC}				-0.043***				-0.012*
				(-7.027)				(-1.893)
N	20,570							
$F-stat$	42.680	36.160	750.600	49.380	14.330	1.949	361.700	3.582

Communication Network

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Div_i^{neighbor}$	$Div_i^{neighbor}$	$Div_i^{neighbor}$	$Div_i^{neighbor}$	Div_i^{user}	Div_i^{user}	Div_i^{user}	Div_i^{user}
C_i^D	-0.073***				-0.057***			
	(-2.442)				(-4.355)			
C_i^B-		-0.018				-0.011**		
		(-1.613)				(-2.213)		
C_i^C			-0.094***				-0.108***	
			(-5.600)				(-23.352)	
C_i^{EC}				-0.105***				-0.066***
				(-2.866)				(-3.696)
N	7,947							
$F-stat$	5.963	2.601	31.360	8.217	18.960	4.897	545.300	13.660

Ticker Network

While the previous section mainly focused on the average diversity of users, it is essential to examine how users diverge in their recommendations and opinions on a per-discussion topic basis, which varies for each ticker sub-network. The following specification investigates the relationship between the diversity of StockTwits users' opinions about a particular stock and the centrality of users within the ticker sub-networks:

$$Y_{i,s} = \alpha + \gamma X_{i,s} + \epsilon_{i,s}, \tag{2.9}$$

$$\text{with } Y_{i,s} = (Div_{i,s}^{neighbor}, Div_{i,s}^{user}) \text{ and, } X_{,si} = (C_{i,s}^{D}, C_{i,s}^{B}, C_{i,s}^{C}, C_{i,s}^{EC}).$$

$Div_{i,s}^{user}$ is the diversity of user i for stock s sub-network with respect to its nearest neighbors. $Div_{i,s}^{neighbor}$ is the diversity of the neighbors reach by node i for stock s sub-network. $C_{i,s}^{D}$, $C_{i,s}^{B}$, $C_{i,s}^{C}$ and $C_{i,s}^{EC}$ represents degree, betweenness, closeness and eigenvector centrality of user i on sub-network s respectively. The variables are standardized.

Table 2.12 presents the results of the ticker sub-network analysis for both the like and communication networks. The direction of the results is similar to the user-level analysis, but the findings are much more significant. In the like network, columns 1 to 4 exhibit a strong and negative relationship between user i centrality and the diversity of the nodes reached by user i in sub-network s, except for the user closeness centrality. The most substantial negative relationship pertains to the user eigenvector centrality, where a one-standard-deviation increase in user i's eigenvector centrality corresponds to a 23.1% reduction in the diversity of users liked with user i in sub-network s. Columns 5 to 8 display a significant positive relationship between user centrality and the diversity of user i with respect to its nearest neighbors in sub-network s, except for user i closeness centrality.

Similarly, for the communication network, the results are the same. The strongest relationship pertains to the user eigenvector centrality, where a one-standard-deviation change in user i's eigenvector centrality corresponds to a 17.3% decrease in the diversity of the nodes reached by user i in sub-network s and an 11.3% increase in the diversity of user i with respect to its nearest neighbors in sub-network s.

Table 2.11 presents parameter estimates from OLS regression on $Div_{i,s}^{user}$ and $Div_{i,s}^{neighbor}$ for the 10% most central users. Although the results do not differ from the user-level analysis, the diversity reduction is more significant for these users. In the like network, a one-standard-deviation increase in the eigenvector centrality of user i results in a 17.24% reduction in the diversity of the nodes reached by user i in sub-network s. In the communication network, a one-standard-deviation increase in the eigenvector centrality of user i corresponds to an 18.76% decrease in the diversity of the nodes reached by user i in sub-network s. For both networks, there is a strong and significant decrease in the diversity of user i with respect to its nearest neighbors in sub-network s as the user's centrality increases.

When considering the difference in the tickers discussed, the reduction in user diversity is even more pronounced for the most central users in the ticker sub-network analysis. The results no longer differentiate between the like and communication networks for all users. An increase in user centrality corresponds to an increase in the diversity of user i with respect to its nearest neighbors in sub-network s and a decrease in the diversity of the nodes reached by user i in sub-network s. From the user's perspective, gaining popularity on the platform improves the diversity compared to their neighbors, but it also leads to being linked to less diverse surroundings. However, whether at the user level or considering the conversation topic, prevalent users negatively affect the network's diversity by creating new connections.

In summary, the findings in section 2.4.3 suggest that financial social networks exhibit homophily, particularly for users who significantly influence the community.

Table 2.12: Diversity and Centrality in Ticker Sub-Networks: Like Network and Communication Network

The table presents parameter estimates from OLS regression on $Div_{i,s}^{user}$ and $Div_{i,s}^{neighbor}$ with several different specifications. $Div_{i,s}^{user}$ is the diversity of user i for stock s sub-network with respect to its nearest neighbors. $Div_{i,s}^{neighbor}$ is the diversity of the neighbors reach by node i for stock s sub-network. $C_{i,s}^{D}$, $C_{i,s}^{B}$, $C_{i,s}^{C}$ and $C_{i,s}^{EC}$ represents degree, betweenness, closeness and eigenvector centrality of user i on sub-network s respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all tweets and network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Like Network							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$
$C_{i,s}^{D}$	-0.148***				0.029***			
	(-107.170)				(16.483)			
$C_{i,s}^{B}$		-0.197***				0.043 ***		
		(-154.518)				(25.083)		
$C_{i,s}^{C}$			0.182***				-0.124***	
			(104.503)				(-70.940)	
$C_{i,s}^{EC}$				-0.231***				0.064***
				(199.280)				(38.904)
N	335,622							
$F-stat$	1.150×10^{4}	2.390×10^{4}	1.090×10^{4}	3.970×10^{4}	271.700	629.200	5032	1513

	Communication Network							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$
$C_{i,s}^{D}$	-0.136***				0.081***			
	(-64.357)				(37.035)			
$C_{i,s}^{B}$		-0.127***				0.077***		
		(-32.660)				(26.076)		
$C_{i,s}^{C}$			0.164***				-0.154***	
			(77.065)				(-78.068)	
$C_{i,s}^{EC}$				-0.173***				0.113***
				(-38.486)				(34.165)
N	257,452							
$F-stat$	4142	1067	5939	1481	1372	680	6095	1167

Table 2.13: Diversity and Centrality in Ticker Sub-Networks Most Central Users (Top 10%): Like Network and Communication Network

The table presents parameter estimates from OLS regression on $Div_{i,s}^{user}$ and $Div_{i,s}^{neighbor}$ for the 10% most central users with several different specifications. $Div_{i,s}^{user}$ is the diversity of user i for stock s sub-network with respect to its nearest neighbors. $Div_{i,s}^{neighbor}$ is the diversity of the neighbors reach by node i for stock s sub-network. $C_{i,s}^{D}$, $C_{i,s}^{B}$, $C_{i,s}^{C}$ and $C_{i,s}^{EC}$ represents degree, betweenness, closeness and eigenvector centrality of user i on sub-network s respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Like Network

	(1) $Div_{i,s}^{neighbor}$	(2) $Div_{i,s}^{neighbor}$	(3) $Div_{i,s}^{neighbor}$	(4) $Div_{i,s}^{neighbor}$	(5) $Div_{i,s}^{user}$	(6) $Div_{i,s}^{user}$	(7) $Div_{i,s}^{user}$	(8) $Div_{i,s}^{user}$
$C_{i,s}^{D}$	-0.0263***				-0.0142***			
	(-5.374)				(-2.515)			
$C_{i,s}^{B}$		-0.1529***				-0.0794***		
		(-30.894)				(-4.016)		
$C_{i,s}^{C}$			-0.0509***				-0.0501***	
			(-9.475)				(-9.204)	
$C_{i,s}^{EC}$				-0.1724***				-0.0541
				(-5.432)				(-1.229))
N	32,242							
$F-stat$	28.87	954.4	89.77	29.51	6.327	16.13	84.72	1.509

Communication Network

	(1) $Div_{i,s}^{neighbor}$	(2) $Div_{i,s}^{neighbor}$	(3) $Div_{i,s}^{neighbor}$	(4) $Div_{i,s}^{neighbor}$	(5) $Div_{i,s}^{user}$	(6) $Div_{i,s}^{user}$	(7) $Div_{i,s}^{user}$	(8) $Div_{i,s}^{user}$
$C_{i,s}^{D}$	-0.1315***				-0.0456**			
	(-17.232)				(-2.066)			
$C_{i,s}^{B}$		-0.1227***				-0.0414		
		(-13.516)				(-1.562)		
$C_{i,s}^{C}$			-0.1394***				-0.0867***	
			(-22.461)				(-39.123)	
$C_{i,s}^{EC}$				-0.1876***				0.0598
				(-19.134)				(1.615)
N	21,321							
$F-stat$	296.9	182.7	504.5	366.1	4.268	2.439	1531	2.608

2.4.4 Extreme opinions

In financial social media, users' performance is influenced by their position within the network. Specifically, the most central users reduce the diversity of opinions among their neighbors, which can lead to the creation of echo chambers, where opposing views are suppressed. As shown by Cookson et al. (2020) in the case of StockTwits, partisan identity can also affect how investors update their market beliefs. Additionally, the spreading capacity of a user, as determined by the sentiment of their recommendation, affects their power to their recommendations. Figure 2.6 shows that bullish recommendations can reach a larger audience than pessimistic recommendations, which can lead to negative consequences as the vast majority of recommendations on financial social networks are bullish. Introducing malicious agents into the network can also enhance the spreading of extreme opinions, posing potential risks on social media platforms.

To further explore the relationship between spreading capacity, diversity, sentiment, and activity, Figures 2.7a and 2.7b display heat maps of the average spreading capacity and diversity as functions of sentiment and activity. The spreading capacity is higher for more active users on the platform. At the same time, the diversity of opinions is lower for users with the highest spreading capacity, illustrating that the users able to influence the network the most lower the diversity within the platform. This finding suggests that the most influential users are associated with a reduction in diversity, which challenges the conditions for verifying the wisdom of crowds hypothesis.

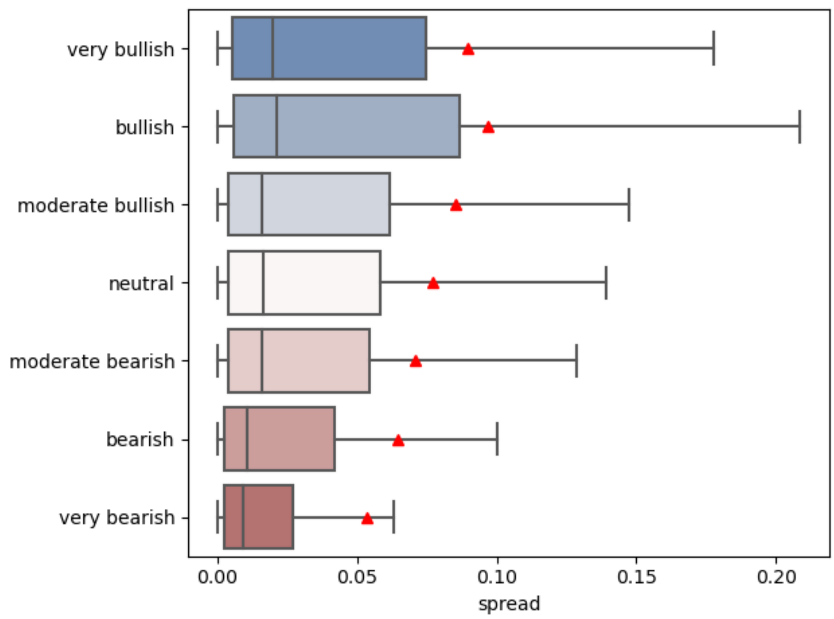
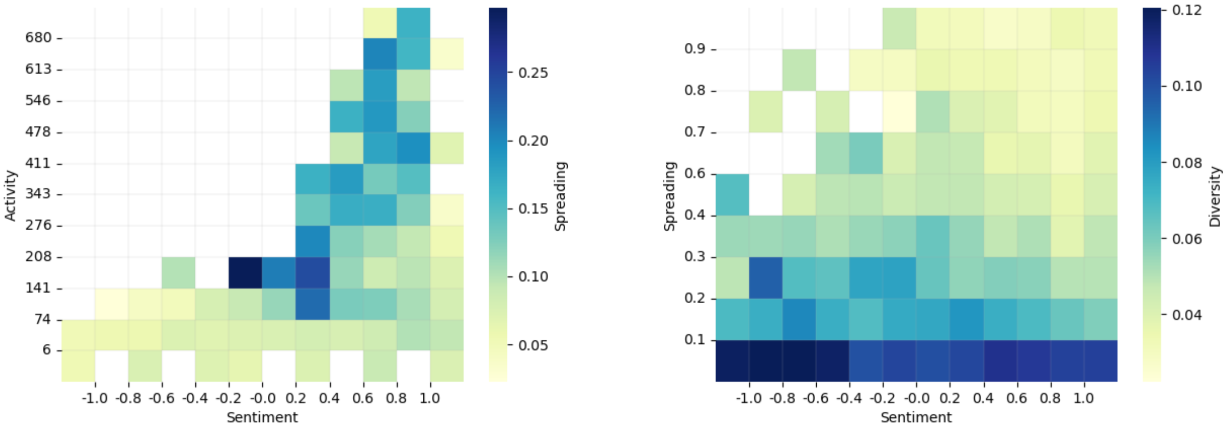


Figure 2.6: Box Plots of Spreading Capacity
This figure displays box plots of the spreading capacity as a function of sentiment. Sentiment is divided into five categories: very bullish (+1), bullish (0.5 to 1), moderate bullish (0.1 to 0.5), neutral (-0.1 to 0.1), moderate bearish (-0.5 to -0.1), bearish (-1 to - 0.5) and very bearish (-1). Boxes show the lower quartile, the median, and the upper quartile. The red triangles indicate the mean.



(a) Average Spreading Capacity **(b)** Average Diversity

Figure 2.7: Heat Maps of Spreading Capacity and Diversity
This figure displays heat maps of the average spreading capacity (panel A) and the average diversity (panel B) as functions of sentiment and activity. On the x-axis, the variable sentiment is divided into ranges of 0.2, except for the ends representing -1 and 1. Activity is divided into 12 equally-spaced groups.

Previous results suggest that financial social networks exhibit homophily, particularly for users who have a rapid and significant impact on the rest of the community. To detect the presence of echo chambers in financial social media, I use the exact specification as in Section 2.4.3 with the ticker sub-networks for the most central users being extremely bullish and extremely bearish users. This approach is motivated by the findings of Barberá et al. (2015), which suggest that echo chamber formation depends on the nature of the issue being discussed, i.e., the ticker in the case of this study.

Table 2.14 reports the results of OLS regression estimates for $Div_{i,s}^{user}$ and $Div_{i,s}^{neighbor}$ using data for the 10% most central and extremely bullish users. This study defines extremely bullish users as those with an average sentiment on stock s greater than 0.7. The findings indicate that the centrality of the most influential users is negatively associated with overall diversity, with the effect being the most significant in this study so far. Specifically, the results show a strong negative relationship between user centrality and the diversity of the nodes reached by user i in the sub-network s for the like network. For instance, a one-standard-deviation increase in the closeness centrality of user i corresponds to a 14.7% decrease in the diversity of user i with respect to its nearest neighbors in sub-network s. A similar relationship is observed for the communication network. However, the relationship between user centrality and the diversity of user i in sub-network s with respect to its nearest neighbors is not statistically significant for the eigenvector centrality.

Table 2.15 presents parameter estimates from OLS regression on $Div_{i,s}^{user}$ and $Div_{i,s}^{neighbor}$ for the 10% most central and extremely bearish users. Extremely bearish users on stock s have an average sentiment on stock s lower than -0.7. The results suggest that the negative effect of centrality on user diversity is just as strong for extremely bearish users as it is for extremely bullish users. For the like network, a one-standard-deviation increase in the eigenvector, betweenness, and eigenvector centrality of user i corresponds to a 28.88%, 39.90%, and 19.90% decrease in the diversity of the users liked from user i in sub-network s, respectively. Furthermore, the findings indicate a negative relationship between user centrality and the diversity of user i in sub-network s with respect to its nearest neighbors. A one-standard-deviation increase in the eigenvector, betweenness, closeness and eigenvector centrality of user i corresponds to an 8.00%, 15.30%, 11.60%, and 11.50% reduction in the diversity of the users liked from user i in sub-network s, respectively. Similar results are obtained for the communication network, where a one-standard-deviation increase in the eigenvector, betweenness, closeness and eigenvector centrality of user i corresponds to a 14.30%, 13.60%, 12.10%, and 17.40% reduction in the diversity of the users liked from user i in sub-network s, respectively. Moreover, the results show that a one-standard-deviation increase in the eigenvector, betweenness, closeness and eigenvector centrality of user i corresponds to a 12.10%, 10.20%, 12.90%, and 10.00% reduction in the diversity of user i in sub-network s with respect to its nearest neighbors.

The results are consistent with Acemoglu et al. (2010), who demonstrated that the presence of "stubborn" agents can lead to persistent disagreements due to different individuals being within the "sphere of influence" of distinct stubborn agents and being influenced to varying degrees. Similarly, this study shows that the negative effect of central users is most pronounced for those with extreme opinions, providing evidence of the worst types of echo chambers in financial social media, characterized by an absence of exposure to opposing views and posing significant risks for investors.

The findings of this study reveal the existence of echo chambers among financial social networks and indicate that the low diversity of opinion on these platforms may help explain why echo chambers are becoming more prevalent on social media. Low diversity, illustrated by selective exposure and confirmation bias, impacts the most central users in the network, who exhibit the best predictive abilities on future stock returns. This finding implies that the predictive value on social networks is not driven by the wisdom of crowds but rather by the influence of the most central users in the network.

Table 2.14: Extremely Bullish Users: Diversity and Centrality in Ticker Sub-Networks Most Central Users (Top 10%): Like Network and Communication Network

The table presents parameter estimates from OLS regression on $Div_{i,s}^{user}$ and $Div_{i,s}^{neighbor}$ for the 10% most central and extremely bullish users with several different specifications. $Div_{i,s}^{user}$ is the diversity of user i for stock s sub-network with respect to its nearest neighbors. $Div_{i,s}^{neighbor}$ is the diversity of the neighbors reach by node i for stock s sub-network. $C_{i,s}^{D}$, $C_{i,s}^{B}$, $C_{i,s}^{C}$ and $C_{i,s}^{EC}$ represents degree, betweenness, closeness and eigenvector centrality of user i on sub-network s respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Like Network

	(1) $Div_{i,s}^{neighbor}$	(2) $Div_{i,s}^{neighbor}$	(3) $Div_{i,s}^{neighbor}$	(4) $Div_{i,s}^{neighbor}$	(5) $Div_{i,s}^{user}$	(6) $Div_{i,s}^{user}$	(7) $Div_{i,s}^{user}$	(8) $Div_{i,s}^{user}$
$C_{i,s}^{D}$	-0.154***				-0.079***			
	(99.554)				(-2.691)			
$C_{i,s}^{B}$		-0.166***				-0.079		
		(-46.731)				(-1.125)		
$C_{i,s}^{C}$			-0.040***				-0.147***	
			(-6.593)				(-76.299)	
$C_{i,s}^{EC}$				-0.157***				-0.027
				(-4.633)				(-0.528)
N	27,782							
$F-stat$	9911	2184	43.470	21.460	7.244	1.265	5822	0.278

Communication Network

	(1) $Div_{i,s}^{neighbor}$	(2) $Div_{i,s}^{neighbor}$	(3) $Div_{i,s}^{neighbor}$	(4) $Div_{i,s}^{neighbor}$	(5) $Div_{i,s}^{user}$	(6) $Div_{i,s}^{user}$	(7) $Div_{i,s}^{user}$	(8) $Div_{i,s}^{user}$
$C_{i,s}^{D}$	-0.143***				-0.132***			
	(-16.201)				(-4.936)			
$C_{i,s}^{B}$		-0.149***				-0.031***		
		(-13.013)				(-3.781)		
$C_{i,s}^{C}$			-0.142***				-0.039***	
			(-19.491)				(-15.726)	
$C_{i,s}^{EC}$				-0.336***				0.111***
				(-16.620)				(14.048)
N	20,359							
$F-stat$	262.500	169.300	379.900	276.200	16.290	14.300	247.300	197.300

Table 2.15: Extremely Bearish Users: Diversity and Centrality in Ticker Sub-Networks Most Central Users (Top 10%): Like Network and Communication Network

The table presents parameter estimates from OLS regression on $Div_{i,s}^{user}$ and $Div_{i,s}^{neighbor}$ for the 10% most central and extremely bearish users with several different specifications. $Div_{i,s}^{user}$ is the diversity of user i for stock s sub-network with respect to its nearest neighbors. $Div_{i,s}^{neighbor}$ is the diversity of the neighbors reach by node i for stock s sub-network. $C_{i,s}^{D}$, $C_{i,s}^{B}$, $C_{i,s}^{C}$ and $C_{i,s}^{EC}$ represents degree, betweenness, closeness and eigenvector centrality of user i on sub-network s respectively. Each variable is standardized; reported estimates measure the change in the dependent variable corresponding to a one-standard-deviation change in X. The sample consists of all network connections from January 2014 to January 2022. The t-statistics (in parentheses) are computed using standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Like Network							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$
$C_{i,s}^{D}$	-0.288***				-0.080***			
	(-15.630)				(-4.057)			
$C_{i,s}^{B}$		-0.399***				-0.153***		
		(-22.752)				(-7.166)		
$C_{i,s}^{C}$			-0.029				-0.116***	
			(-1.447)				(-5.181)	
$C_{i,s}^{EC}$				-0.199***				-0.150***
				(-54.015)				(-7.475)
N	3,306							
$F-stat$	244.300	517.700	2.093	2918	16.460	51.350	26.850	55.870

	Communication Network							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{neighbor}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$	$Div_{i,s}^{user}$
$C_{i,s}^{D}$	-0.143***				-0.121***			
	(-16.557)				(-16.133)			
$C_{i,s}^{B}$		-0.136***				-0.102***		
		(-13.895)				(-12.360)		
$C_{i,s}^{C}$			-0.121***				-0.129***	
			(-4.031)				(-3.892)	
$C_{i,s}^{EC}$				-0.174***				-0.100***
				(-22.446)				(-9.859)
N	1,423							
$F-stat$	274.100	193.100	16.250	503.800	260.300	152.800	15.140	97.200

2.5 Conclusion

In this study, I explore the predictive power of financial social networks and investigate their origins by analyzing the network characteristics of their users. Specifically, I investigate the relationship between investors' network centrality and their ability to predict future stock returns, and I identify the most influential users of the platform. By examining the viewpoints of these users, I distinguish between the wisdom of the crowd hypothesis and the behavioral bias hypothesis, which leads to the formation of echo chambers.

First, I construct communication and like networks between StockTwits users and find that social media users' performance is higher for more central users. I observe a positive association between a user's centrality and their future performance. Moreover, more complex centrality measures, such as eigenvector and closeness centrality, provide useful information in identifying high-quality users.

Next, I utilize network graphs to develop two diversity variables: the diversity of a user with their neighbors and the diversity among a user's neighbors. I find that for the most influential users, the support for the wisdom of crowds hypothesis is diminishing. These users are associated with a significant reduction in diversity, leading to a large number of users expressing opinions in agreement about a particular stock.

Lastly, I analyze the types of recommendations that are more widely transmitted across the network, particularly moderate or extreme opinions. I find that for extremely bullish and bearish influential users, an increase in centrality is associated with the most substantial decrease in overall diversity.

In conclusion, users' performance within financial social networks is heavily influenced by their position within the network. Rather than being the result of the aggregation of financial market information from a large number of users, superior performance is a reflection of similar recommendations expressed in tandem. The emergence of echo chambers within social media platforms is driven by the growth of user influence, particularly for users with extreme opinions on a specific stock. These findings are crucial in finance because the formation of investors' opinions directly affects their investments. As the size and reach of financial social networks continue to expand, the value of information on these platforms will increase, but so will the challenges associated with the emergence of these financial influencers.

Bibliography

Acemoglu, D., Como, G., Fagnani, F. and Ozdaglar, A. (2010). Persistence of disagreement in social networks.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards, *The Journal of Finance* **59**(3): 1259–1294.

Babus, A. and Kondor, P. (2018). Trading and information diffusion in over-the-counter markets, *Econometrica* **86**(5): 1727–1769.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F. and Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization, *Proceedings of the National Academy of Sciences* **115**(37): 9216–9221.

Bakshy, E., Hofman, J. M., Mason, W. A. and Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on twitter, *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 65–74.

Barber, B. M., Huang, X., Odean, T. and Schwarz, C. (2021). Attention induced trading and returns: Evidence from robinhood users, *Journal of Finance, Forthcoming* .

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A. and Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber?, *Psychological science* **26**(10): 1531–1542.

Bartov, E., Faurel, L. and Mohanram, P. S. (2018). Can twitter help predict firm-level earnings and stock returns?, *The Accounting Review* **93**(3): 25–57.

Breitmayer, B., Massari, F. and Pelster, M. (2019). Swarm intelligence? Stock opinions of the crowd and stock returns, *International Review of Economics & Finance* **64**: 443–464.

Broadstock, D. C. and Zhang, D. (2019). Social-media and intraday stock returns: The pricing power of sentiment, *Finance Research Letters* **30**: 116–123.

Cai, Y. and Sevilir, M. (2012). Board connections and M&A transactions, *Journal of Financial Economics* **103**(2): 327–349.

Chen, H., De, P., Hu, Y. J. and Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media, *The Review of Financial Studies* **27**(5): 1367–1403.

Chen, X., Shangguan, W., Liu, Y. and Wang, S. (2021). Can network structure predict cross-sectional stock returns? Evidence from co-attention networks in china, *Finance Research Letters* **38**: 101422.

Choi, D., Chun, S., Oh, H., Han, J., Kwon, T. et al. (2020). Rumor propagation is amplified by echo chambers in social media, *Scientific reports* **10**(1): 1–10.

Cohen, L., Frazzini, A. and Malloy, C. (2008). The small world of investing: Board connections and mutual fund returns, *Journal of Political Economy* **116**(5): 951–979.

Cohen, L., Frazzini, A. and Malloy, C. (2010). Sell-side school ties, *The Journal of Finance* **65**(4): 1409–1437.

Cookson, J. A., Engelberg, J. E. and Mullins, W. (2020). Does partisanship shape investor beliefs? Evidence from the covid-19 pandemic, *The Review of Asset Pricing Studies* **10**(4): 863–893.

Cookson, J. A., Engelberg, J. and Mullins, W. (2022). Echo chambers, *Available at SSRN 3603107* .

Cookson, J. A., Fos, V. and Niessner, M. (2021). Does disagreement facilitate informed trading? Evidence from activist investors, *Evidence from Activist Investors (April 12, 2021)* .

Cookson, J. A. and Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors, *The Journal of Finance* **75**(1): 173–228.

Cortez, P., Oliveira, N. and Ferreira, J. P. (2016). Measuring user influence in financial microblogs: Experiments using stocktwits data, *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, pp. 1–10.

Cota, W., Ferreira, S. C., Pastor-Satorras, R. and Starnini, M. (2019). Quantifying echo chamber effects in information spreading over political communication networks, *EPJ Data Science* **8**(1): 1–13.

Coval, J. D. and Moskowitz, T. J. (2001). The geography of investment: Informed trading and asset prices, *Journal of political Economy* **109**(4): 811–841.

Crane, A. and Crotty, K. (2020). How skilled are security analysts?, *The Journal of Finance* **75**(3): 1629–1675.

Della Rossa, F., Giannini, L. and DeLellis, P. (2020). Herding or wisdom of the crowd? Controlling efficiency in a partially rational financial market, *Plos one* **15**(9): e0239132.

Du, Q., Hong, H., Wang, G. A., Wang, P. and Fan, W. (2017). Crowdiq: A new opinion aggregation model.

Du, S. and Gregory, S. (2016). The echo chamber effect in twitter: Does community polarization increase?, *International workshop on complex networks and their applications*, Springer, pp. 373–378.

Eaton, G. W., Green, T. C., Roseman, B. and Wu, Y. (2021). Zero-commission individual investors, high frequency traders, and stock market quality, *High Frequency Traders, and Stock Market Quality (January 2021)* .

Engelberg, J., Gao, P. and Parsons, C. A. (2012). Friends with money, *Journal of Financial Economics* **103**(1): 169–188.

Engelberg, J., Gao, P. and Parsons, C. A. (2013). The price of a CEO's Rolodex, *The Review of Financial Studies* **26**(1): 79–114.

Frey, V. and Van de Rijt, A. (2021). Social influence undermines the wisdom of the crowd in sequential decision making, *Management science* **67**(7): 4273–4286.

Goutte, M.-R. (2020). When the 'dumb' crowd beats the 'smart' crowd, *Available at SSRN 3918336* .

Goutte, M.-R. (2022). Do actions speak louder than words? Evidence from microblogs, *Journal of Behavioral and Experimental Finance* **33**: 100619.

Heimer, R. Z. (2016). Peer pressure: Social interaction and the disposition effect, *The Review of Financial Studies* **29**(11): 3177–3209.

Hirshleifer, D. (2020). Presidential address: Social transmission bias in economics and finance, *The Journal of Finance* **75**(4): 1779–1831.

Hochberg, Y. V., Ljungqvist, A. and Lu, Y. (2007). Whom you know matters: Venture capital networks and investment performance, *The Journal of Finance* **62**(1): 251–301.

Hong, H., Kubik, J. D. and Stein, J. C. (2005). Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, *The Journal of Finance* **60**(6): 2801–2824.

Hong, H., Ye, Q., Du, Q., Wang, G. A. and Fan, W. (2020). Crowd characteristics and crowd wisdom: Evidence from an online investment community, *Journal of the Association for Information Science and Technology* **71**(4): 423–435.

Hong, L. and Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers, *Proceedings of the National Academy of Sciences* **101**(46): 16385–16389.

Hossain, M. M., Mammadov, B. and Vakilzadeh, H. (2022). Wisdom of the crowd and stock price crash risk: Evidence from social media, *Review of Quantitative Finance and Accounting* **58**(2): 709–742.

Iyengar, S. and Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use, *Journal of communication* **59**(1): 19–39.

Jackson, M. O. (2007). The study of social networks in economics, *The missing links: Formation and decay of economic networks* **76**: 210–225.

Kenett, D. Y. and Havlin, S. (2015). Network science: A useful tool in economics and finance, *Mind & Society* **14**(2): 155–167.

Kumar, S., Hamilton, W. L., Leskovec, J. and Jurafsky, D. (2018). Community interaction and conflict on the web, *Proceedings of the 2018 world wide web conference*, pp. 933–943.

Li, D. and Schürhoff, N. (2019). Dealer networks, *The Journal of Finance* **74**(1): 91–144.

Martin, S., Brown, W., Klavans, R. and Boyack, K. (2011). Openord: An open-source toolbox for large graph layout, *Proc SPIE* **7868**: 786806.

Molavi, P., Tahbaz-Salehi, A. and Jadbabaie, A. (2018). A theory of non-bayesian social learning, *Econometrica* **86**(2): 445–490.

Pedersen, L. H. (2022). Game on: Social networks and markets, *Journal of Financial Economics* .

Pool, V. K., Stoffman, N. and Yonker, S. E. (2015). The people in your neighborhood: Social interactions and mutual fund portfolios, *The Journal of Finance* **70**(6): 2679–2732.

Renault, T. (2017). Market manipulation and suspicious stock recommendations on social media, *Available at SSRN 3010850* .

Renault, T. (2020). Sentiment analysis and machine learning in finance: A comparison of methods and models on one million messages, *Digital Finance* **2**(1): 1–13.

Sprenger, T. O., Tumasjan, A., Sandner, P. G. and Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs, *European Financial Management* **20**(5): 926–957.

Tu, W., Yang, M., Cheung, D. W. and Mamoulis, N. (2018). Investment recommendation by discovering high-quality opinions in investor based social networks, *Information Systems* **78**: 189–198.

Zhang, Y. (2009). Determinants of poster reputation on internet stock message boards, *American Journal of Economics and Business Administration* **1**(2): 114.

Chapter 3

Do Actions Speak Louder than Words? Evidence from Microblogs

This research identifies the determinants of investors' future beliefs by analyzing more than 50 million tweets on thousands of stocks from the microblogging platform StockTwits. To do so, I divide tweets into two categories: beliefs, representing the average sentiment of all investors regarding a particular stock, and actions, representing the actual transactions disclosed by StockTwits' users in their tweets. The results show that investors' next-period beliefs are positively impacted both by beliefs and actions of the current day. The quality of the investment advice strengthens this effect. More communication between investors leads to greater diversity in beliefs and more uncertainty.

3.1 Introduction

"*We currently have a large position in APPLE. We believe the company to be extremely undervalued. Spoke to Tim Cook today. More to come.*"

Carl Icahn, 13th August 2013

This tweet from Carl Icahn caused the market capitalization of Apple to increase by $12 billion and is just another example of how the information content of tweets can have a great influence on the valuation of a firm. Since 2013, information on social media has become more valuable for finance. Indeed, the Security and Exchange Commission (SEC) announced that year that it had "issued a report that makes clear that companies can use social media outlets like Facebook and Twitter to announce key information in compliance with Regulation Fair Disclosure (Regulation FD) as long as investors have been alerted about which social media will be used to disseminate such information."[1] In the past year, the popularity of microblogging platforms has increased drastically. Financial investors are increasingly interested in social media publications and several companies are selling social media sentiment data.[2] Consequently, the prediction of social media sentiment can have substantial economic value.

Historically, the lack of appropriate data renders the empirical identification of sources of changes in beliefs difficult.[3] However, modern social media data seem to be an adequate tool to identify the determinants of future beliefs. Since individuals remain important players in financial markets, one key ingredient to understand price formation is to learn how investors form their opinions.

This study focuses on the determinants of investors' beliefs concerning one individual stock at a daily frequency. The extensive data available on the StockTwits platform, the most extensive social network for investors and traders, allows me to analyze millions of beliefs on different stocks and determine the sources of changes in belief.

Unlike newspapers and traditional media, social media platforms allow the transmission of information coupled with interpersonal communication. According to Shiller (2015), opinions are more affected through interpersonal communication, i.e., when users can reply and communicate with each other, in contrast to reading newspapers, which represent unilateral communication. Indeed, the StockTwits platform allows users to post messages and the microblogging community to react to these messages by retweeting, liking, or responding to the messages.

One concern is that users may be dishonest in their tweets (particularly in regards to their actions) when posting messages for other reasons. However, microbloggers have incentives to behave honestly to gain popularity on the platform (measured by retweets, followers, and likes) and

[1] https://www.sec.gov/news/press-release/2013-2013-51htm.

[2] For example tweetfeel (http://www.tweetfeel.com/) and Twitratr (https://twitrratr.com/).

[3] For example analysts recommendations frequency does not allow for daily analysis.

to preserve their reputation. Sprenger et al. (2014) show that users who provide higher-quality investment advice have more influence on the social media platform. This finding favors the incentives of users to give honest opinions in their messages. To mitigate this concern, I weight each message sentiment score by the popularity of the user on the platform. Thus, I put more weight on messages posted by users who are more trustworthy according to the whole community.

This research focuses on the content of 50 million tweets regarding 2,154 stocks from January 2013 to August 2018. To the extent of my knowledge, previous studies on social-media financial content analyze only up to 2 years of data or a substantially lower number of tweets. I classify messages as beliefs (messages expressing opinions) and actions (messages disclosing actual transactions). Using different natural language processing techniques, I give each of those messages a sentiment score (bullish/bearish for opinions and buy/sell for trades).
Rather than studying the traditional "sentiment-predicts-returns" relationship, like previous studies on the relationships between tweet sentiment and aggregated performance (Batra and Daudpota (2018), Antweiler and Frank (2004)) as well as between tweet sentiment and individual stock market indicators (Sprenger et al. (2014), Agrawal et al. (2018)), I focus on how individual current sentiment affects future sentiment for a particular stock.
The objective of this research is to answer the following questions: Are investors' future beliefs affected by current sentiment? Are the beliefs affected by trades in the previous period? Are those effects influenced by the amount of communication and the size of the message's audience? Does the quality of communication and of the trades affect investors' beliefs and uncertainty?

To determine the nature of the effect observed on next-period beliefs, namely, whether the effect reflects rational updating or behavioral biases, I first perform a rationality test. I find that members of the StockTwits community are not fully rational Bayesian updaters because they are likely to underweight their prior beliefs. After rejecting the rational updating hypothesis, I answer the abovementioned questions by testing the corresponding hypothesis described in Section 3.3.

I find that average beliefs, as well as the average trade direction, can predict next-day average beliefs. Higher average sentiment in the tweets leads to convergence in beliefs, and the effect of the average trade direction is similar but smaller. Furthermore, an increasing amount of communication between the users on the platform leads to more diversity of opinion. Additionally, the prediction power is higher after taking into account the popularity of the message sender, confirming that convergence of beliefs is higher once the credibility of the source is taken into consideration. Finally, the quality of advice regarding the trades disclosed in the messages helps predict next-period investor sentiment, although an increase in general recommendation quality does not help reduce uncertainty.

Moreover, the market-related uncertainty measure computed from newspaper articles has no significant relationship with the uncertainty variable calculated from social media data. This result highlights the fact that social media data and sentiment are not always in line with newspaper articles.

The remainder of the paper is structured as follows. First, Section 2 presents the corresponding literature review. Section 3 discusses the research question and hypotheses. Section 4 describes the message classification and some basic features of the data. Section 5 presents the results of the regressions. Finally, I conclude in Section 6.

3.2 Literature Review

There is a growing body of literature on social media and their relationship with stock market data. Antweiler and Frank (2004) use messages from Yahoo! Finance and Raging Bull to show that messages predict volatility and that disagreement between messages induces trading volume. In line with Antweiler and Frank (2004), Sprenger et al. (2014) find some relationship between stock returns and sentiment as well as between disagreement in messages and volatility and between message volume and trade volume. They also demonstrate that higher-quality users (who make higher-quality predictions) have a larger audience and more popularity on the platform.

More recently, Berry et al. (2019) compare the performance of the GARCH model with stock returns versus the performance of the NAGARCH model with stock returns and StockTwits' sentiment. The authors show that in most cases, the NAGARCH model with sentiment improves the prediction of volatility. Batra and Daudpota (2018) use sentiment analysis of StockTwits messages to predict next day's stock movement. They identify a positive relationship between users' sentiment and market data. On the same topic, Broadstock and Zhang (2019) use messages on Twitter to construct sentiment measures. Their results show that sentiment from Twitter carries 'pricing power' against the stock market. Cookson and Niessner (2018) use StockTwits data to understand sources of disagreement between investors. They find that disagreement induced by different information sets is more likely to induce trading than disagreement induced by differences in investor characteristics.
Agrawal et al. (2018) and Hill-Kleespie (2018) use StockTwits data to test theories on momentum. Additionally, Agrawal et al. (2018) show that liquidity demand and supply are affected by investors' sentiment. Chen et al. (2019) analyze the impact of investor sentiment on returns in the cryptocurrency market during the 2017 bubble using StockTwits data, and found that the relationship between sentiment and returns depended on whether or not a bubble was present. Cortez et al. (2016) study user influence measures from StockTwits social network. They find

that direct interaction from financial microblogs, such as the number of messages posted, can be used to measure the quality and influence of users.
Using different sentiment data, Karampatsas et al. (2017) study the relationship between PsychSignal (a company that analyzes sentiment data from social media and other sources to generate insights and signals for investment decision-making) firm-specific measures of sentiment and market reaction to earnings announcement. The results illustrate that firm-specific sentiment affects abnormal stock returns on the announcement of earnings surprises. On the same topic, Bartov et al. (2018) evaluate whether users' opinions on Twitter preceding a firm's earnings announcement can predict its earnings and announcement returns.

In his book, Shiller (2015) states that people who interact with each other regularly tend to think similarly, either because they are reacting to the same information or out of non-fully independent judgment. It is undeniable that others influence us in our actions and opinions; the finance industry is no exception. More recently, Da and Huang (2020) study the effect of herding on the accuracy of earnings forecast consensuses from Estimize.com, a crowdsourcing corporate earning forecast platform. They find that users underestimate their private information as they see more public information on the platform. Moreover, herding increases with the proportion of influential users' estimates in the public information set.

Before assuming that investors change their beliefs through communication, persuasion, or herding behavior, one has to make sure investors are not fully rational Bayesian updaters. According to the rationality test proposed by Augenblick and Rabin (2018), it is possible to distinguish between the two competing theories regarding investor rationality and irrationality. The test results, which were applied to daily average social media beliefs on thousands of stocks, suggest that the StockTwits community members may not always exhibit fully rational Bayesian updating behavior.
The convergence of beliefs through communication occurs when investors review their opinions after communicating with their peers. According to DeMarzo et al. (2003), agents are subject to persuasion bias. These authors model opinion formation where agents repeatedly communicate within a social network and fail to account for the repetition of information. In their model, beliefs tend to converge to a weighted average of common beliefs, with the weights depending on the position in the network. In other words, persuasion occurs through communication, and persuasion power depends on the position of the persuader in the network. In terms of the relevance of the previous statement for the research question in the current study, the convergence of beliefs should depend on the popularity of the user expressing his opinion, as more network connections represent a larger audience. Foerster (2018) build a boundedly rational model of opinion formation in which agents are subject to persuasion bias and repeatedly communicate with their social network neighbors. They find that convergence of beliefs occurs when agents communicate using

the same language.

Theoretically, Hong et al. (2005) and DeMarzo et al. (2003) show that investors are often influenced by word of mouth when forming their beliefs or when making decisions. Ellison and Fudenberg (1995) build a model to study word-of-mouth communication in a nonstrategic environment. Their model predicts that more word-of-mouth communication leads to diversity, whereas less word-of-mouth communication leads to conformity, with conformity defined as the act of matching actions and attitudes. In Andrei and Cujean (2017), the authors show that word-of-mouth communication induces short-term momentum and long-term reversal. When this finding is extrapolated to microblogging platforms, it implies that in the short term, past beliefs should have a positive effect on next-period beliefs, whereas, in the long term, this effect should be negative. However, few papers empirically relate the word-of-mouth hypothesis to the stock market. In Hong et al. (2005), investors spread information on a particular stock through word-of-mouth communication. By looking at mutual fund managers' holdings and trades, the authors find that trades are more correlated among investors that are located in the same city. These findings imply that communication induces correlations in behaviors. Banerjee (1992) provides a sequential decision model where agents take into account previous agents' actions when making decisions. The paper illustrates herding behavior by taking as an example the decision on which restaurant to choose when looking at other individuals' reviews. Herding is broadly defined as correlated behavior among investors. Herding can take place through observation of a signal (public announcements, price movements) or the behavior of other investors. According to Devenow and Welch (1996), the most general explanation for herding may be information cascades. When investors value the actions of others to the extent that they will disregard their own previous information or belief if it contradicts those of others, an information cascade occurs. Thus, for an information cascade to emerge, investors' actions need to be observable. Molavi et al. (2018) study the behavioral foundations of non-Bayesian models of learning over social networks. They assume that agents treat the beliefs of their neighbors as sufficient statistics for the entire history of their observation. Bordalo et al. (2019) use a model of belief formation based on the representativeness heuristic to explain the overreaction of analysts to public information when they forecast earnings.

Graham and Harvey (1997) show that investors can earn superior returns by following the advice of newsletters that show a substantial ability to time the market. It is then straightforward to assume, based on this finding, that social media users who disclose successful trades influence future average beliefs. By differentiating between the three sources of changes in opinions (communication, actions, and consequences of actions), one should be able to determine which source is the strongest signal that sparks changes investor beliefs and thereby better understand opinion formation.

To uncover patterns of correlated trading behaviors among investors, Colla and Mele (2009) model information distance between individuals by looking at the amount of common information they share. These local connections are defined as information linkages. Close informational neighbors share similar information about long-term asset value. In line with DeMarzo et al. (2003), these authors find that for close informational neighbors, the correlation among beliefs and actions is high but decreases over time. Moreover, investors' beliefs and opinions diverge when information distance increases.

Further, Ozsoylev et al. (2013) show by evaluating the trading behavior of investors in the Istanbul Stock Exchange that trading behavior depends on an investor's position in the network. Consequently, in the variable construction process of the current study, beliefs and actions are weighted depending on how many of an investor's close neighbors have seen the information (corresponding to the number of likes and followers in the StockTwits platform). If users read a message on a microblog platform, they share the same information and become close informational neighbors. The finding that the convergence of beliefs decreases with uncertainty implies, in the context of microblogging, that if there are fewer tweets posted, prior beliefs and actions should have a more substantial effect on next-period average beliefs. This statement is consistent with the results of Ellison and Fudenberg (1995), who find that less communication leads to more conformity.
Siganos et al. (2017) introduce the concept of divergence of sentiment to the behavioral finance literature. The results show that diverging views are positively related to trading volume and stock price volatility. Those relationships are stronger when investors are likely to trade. Pouget et al. (2017) propose a model where traders are subject to confirmatory bias. Empirical findings using data on analysts' earnings forecasts show that traders update beliefs depending on the sign of past signals and previous beliefs.

3.3 Research question

The theoretical foundations of the following set of hypotheses assume that investors are not fully rational. To validate the setting of the study, I perform a rationality test in Section 3.5.1.

- **Hypothesis 1:** Agents' future beliefs are affected by previous **beliefs** and by previous **actions**.

  Consistent with Shiller (2015) and DeMarzo et al. (2003), interaction through communication leads to convergence in beliefs. Alternatively, following Barberis et al. (2015), many investors hold extrapolative expectations. They believe that stock prices will continue rising/falling once they have already risen/fallen. Thus, we can expect bullishness/bearishness to increase after previous bullish/bearish periods. Using analyst's earnings forecasts, Pouget et al. (2017) show that traders' current beliefs depend on previous beliefs.

According to Enke and Zimmermann (2019), beliefs are too sensitive to the repetition of stories and exhibit excessive swings. The more we are exposed to an opinion, the more our own opinion tends to converge to this opinion, neglecting the correlation of beliefs. I thus incorporate the number of messages posted into the specification as a control as a proxy for repetition and communication.

Actions are differentiated from beliefs in the model by representing the former through trades. Trades are actual transactions disclosed by StockTwits users in their message posting. This differentiation is important when evaluating the determinants of next-period beliefs, as trades involve spending money for the users and thus enhance credibility. Accordingly, I assume that actual transactions give a stronger signal of bullishness/bearishness than do opinions to the rest of the community.

Theoretically, Banerjee (1992) show that observing actions can affect investors' beliefs. Bursztyn et al. (2014) identify two separate channels of social influence: investors can learn from other investors' choices, defined as "social learning" or investors' utility can increase through possessing the same asset as their peers, defined as "social utility." Both channels predict a positive relationship between average trade sentiment and next-period average sentiment.

- **Hypothesis 2:** Agents' future beliefs are affected by the **quality of advice** related to stock actions on the previous day.

In line with Sprenger et al. (2014), users providing better-quality advice on the microblog receive more likes and follower subscriptions. When a user discloses a successful trading strategy, it should enhance his or her credibility and give a stronger signal to his or her network.

- **Hypothesis 3:** More communication between agents leads to more diversity in beliefs.

Using a conformity measure similar to the "agreement" measure in Antweiler and Frank (2004), the Stocktwits data allow for direct testing of the theory elaborated in Ellison and Fudenberg (1995). The conformity measure is differentiated from the agreement measure in that it refers to actions rather than beliefs.

- **Hypothesis 4:** An increase in the advice quality of messages is associated with lower uncertainty.

According to Shiller (2015), "Limitation in the effectiveness of conversation in conveying information may be a source of market volatility." Herding behavior is mitigated by users' effectiveness in generating a message of high quality. Quality is defined as the ability of the message sentiment to predict stock returns. Therefore, higher-quality investment advice should be associated with lower uncertainty. Furthermore, as mentioned in Sprenger et al. (2014) and Antweiler and Frank (2004), message volumes can predict volatility. It follows that a proxy of message volume should be incorporated as a control variable when testing hypothesis 4.

3.4 Data and methodology

3.4.1 StockTwits Data

Founded in 2008, StockTwits is the most extensive social network for investors and traders, with over one million registered community members and approximately three million monthly visitors. Because it enables users to express opinions and sentiments in real time, the Stocktwits dataset is suitable to test behavioral finance theories. In the dataset, the following variables are associated with each tweet: the content of the message, the ticker of the stock related to the message, the user's name, the time of the post (with split-second accuracy), the number of users liking or sharing the tweet, the number of people following the user, the number of tweets already posted by the user, the date from which the user has been active on the platform, a logical variable called "official account" (if the account is related to an official media figure or a guru) and a logical variable called "market leader."[4]

All users can post a message of a maximum of 140 characters about a particular stock. Since 2012, users have also been able to classify their message as bearish or bullish. However, the majority of the tweets do not have this classification. Approximately 20% of the tweets in the data for the Apple stock include the bearish/bullish tag. To extract bullish/bearish sentiment from all tweets, I use the natural language processing (NLP) techniques presented in Section 3.4.2.

I first consider all the messages posted on StockTwits since the creation of the platform (57,225,057 tweets on 8599 stocks). The stocks that are the subject of the greatest number of posts are presented in Table 3.1.

To obtain a sufficient number of messages per day for each stock, I removed the stocks associated with fewer than 3,000 messages from the sample. Additionally, I analyze only messages posted after 2013 due to the low posting level at the platform's inception (see Table 3.2).

Table 3.3 shows the average number of tweets posted each year on the platform. Consequently, I use messages from January 2013 to August 2018. The final dataset consists of 50,159,831 messages

[4] For example for the stock of Apple Inc. each tweet is referenced with the ticker "$AAPL."

Table 3.1: Stocks with the greatest posting level

Ticker	Company, Index	Message volume
BB	BlackBerry Ltd	655’324
AMD	Advanced Micro Devices, Inc.	655’206
TSLA	Tesla Inc.	655’133
FB	Facebook, Inc.	655’033
AAPL	Apple Inc.	655’018
SPY	SPDR S&P 500 ETF Trust	654’737
DRYS	DryShips Inc.	609’757
TWTR	Twitter Inc	562’034
JNUG	Daily JR Gold Miner	556’791
BHC	Bausch Health Companies Inc	545’687
NFLX	Netflix, Inc.	535’541
AMZN	Amazon.com, Inc.	513’828

from 2,154 stocks. Table 3.4 presents the summary statistics for the number of tweets posted for each stock.

Table 3.2: Descriptive statistics: Number of tweets per stock: Initial Sample

Mean	Median	Std	Min	Max	p25	p75
6’654	872	32’714	1	655’324	156	3’006

Table 3.3: Descriptive statistics: Number of tweets each year

	Mean	Median	Std	Min	Max	N
2009	14	0	62	0	1,511	29,705
2010	105	14	427	0	9,420	227,133
2011	291	52	1,084	0	18,385	627,755
2012	670	102	3,727	0	106,822	1,443,269
2013	1,645	335	1,096	0	359,949	3,544,213
2014	2,356	554	9,394	0	221,487	5,075,826
2015	2,831	751	8,536	0	130,558	6,098,646
2016	4,349	1,013	14,425	0	254,400	9,368,173
2017	7,063	1,800	23,248	0	500,617	15,214,374
2018	5,041	1,321	20,153	0	508,016	10,858,599

Table 3.4: Descriptive statistics: Number of tweets per stocks for final dataset

Mean	Median	Std	Min	Max	p25	p75
24,368	7,695	62,073	3,001	655,324	4,539	17,603

3.4.2 Variables construction

Before classifying tweets into sentiments (beliefs/actions), I perform the following steps:

- Put the text in lower-cases and remove dots.
- Remove all HTML tags from the body of the tweet and replace them with the ASCII decoded version of the given HTML tag.[5]
- Remove all non-ASCII characters.
- Use Porter Stemmer: remove morphological affixes from words, leaving only the word stem. For example, "generously" becomes "generous." This step helps matching a maximum number of possible words in the tweets with the lexicon.
- Remove Stopwords (words that are not expressing emotions). This step is not adding accuracy to the classification.
- Add Bigram.
- Identify negated words to match with the lexicon. Negated words like "not good" are replaced with "good_NEGFIRST." The NRC Emoticon Lexicon is composed of two lexicons, a negative context lexicon, and a positive context lexicon. Instead of reversing the polarity of the score of the word, the same word has two different scores depending on the context.
- Manually add some famous StockTwits abbreviations and expressions ("hod", "lod", "btfd", "dip", "glta", etc).

All tweets are then classified into two categories: beliefs and actions. Tweets in the action category represent actual transactions, whereas tweets in the beliefs category represent opinions. To use all the available information in the StockTwits data, belief tweets and action tweets are weighted by the number of likes and followers to test the difference of visibility of information.

[5] For example: """, "&", "'" are HTML tags.

Based on previous papers, I assume that more visible messages should have more credibility. After computing the beliefs and actions score for all tweets and weighting them with respect to user importance, I aggregate both measures at a daily frequency.

With the help of Quandl's Python API, I download the financial data series of 2,000 stocks.[6] To the extent of my knowledge, past studies building on the entire StockTwits dataset focus on a maximum of one year of data. An alternative is to consider a smaller universe of stocks.

Beliefs: Sentiment Analysis

Sentiment analysis is the process of identifying and categorizing opinions expressed in a sequence of words using NLP techniques. There have been different methods used to extract sentiment from tweets (see Man et al. (2019) and Mishev et al. (2020) for an overview of the different techniques used in the literature as well as performance of various sentiment analysis approaches).
This research makes use of a tokenized lexicon to predict sentiment, similar to the one developed by Loughran and McDonald (2011). Each word in the lexicon is given a score (positive or negative), and the overall sentiment of the tweets is the sum of all lexicon word's score matching the words in the tweet.
As shown in Table 3.5, the language used on StockTwits is highly informal, with sentiment expressed with common language or, alternatively, with hashtags and emojis. Thus, using the LM lexicon or another standard sentiment lexicon to classify sentiment is not appropriate here.
Following Hill-Kleespie (2018), I decide to use a sentiment and emoticon lexicon from the National Research Center of Canada developed by Mohammad and Turney (see Mohammad and Turney (2013)). The NRC Emoticon Lexicons was automatically generated from 1.6 million tweets with emoticons. I use both bigram and unigram lexicons to classify tweet sentiment. Each word of the lexicon is given a score. Indeed, it is fair to assume that some positive or negative words are stronger than others. Therefore, a score that differs from the usual rating (+1, -1) gives greater precision to the overall sentiment score of the tweet. Mohammad and Turney (2013) compute the score for each word as the difference between the positive pointwise mutual information (PMI) and the negative PMI.
The NRC Emoticon Lexicon allows for the identification of terms in negated contexts, with the prefix "_NEGFIRST" attached to terms that directly follow a negator. Identifying terms in negated contexts allows for better classification. I thus transform all the words following a negator to "word_NEGFIRST" to match the lexicon.[7] Given the uniqueness of the language on this platform, I manually add some words that are often used in StockTwits to be able to obtain a very specific lexicon for my study (e.g., "cheers", "hod" meaning high of day, "lod" meaning low of day, "glta" meaning good luck to all, "lmao", etc.).

[6] https://www.quandl.com/.
[7] For example the word "good" is positive but the bigram "not good" should indicate negative sentiment.

Messages are classified into 3 types: "bullish" representing positive sentiment, "bearish" representing negative sentiment, and "neutral" representing either noninformative messages or messages that are neither bullish nor bearish. The final sample size of beliefs-related tweets is 20 million (64% bullish, 32% bearish, and 4% neutral).

Table 3.5 presents examples of tweets that are first classified as beliefs and then as bullish, bearish, or neutral.

Table 3.5: Example of tweets that are classified as "Beliefs"

Belief	Tweet
bullish	$AAPL highly attractive
bullish	$AAPL Fly me to The Moon......YESSSSSS
bullish	$AAPL be patient here.. hold the line.. 651 far_NEGFIRST away, can take your profits there
bullish	$AAPL from USA Today 10 sold out in top 20 cities. WOW
bullish	$AAPL Extremely Bullish adoption rates for iPhone 5s and iPhone 5c
bullish	$AAPL Among other reasons for bullishness: There are few signs equities are extended ...
bullish	$AAPL Deutsche Bank sees stocks up just 3% in 2014 http://stks.co/ayfQ
bullish	$AAPL lets bet on tomorrow , green or red by end of the day?, I'm green
bullish	$AAPL AH up $2.08 do we see $570.00 tomorrow. Very possible
bullish	$AAPL boom goes the dynamite! Calling 575 manana!
bearish	$AAPL probably should sell
bearish	$AAPL really great_NEGFIRST setup
bearish	$AAPL my new x sucks! It sucks!!!
bearish	$AAPL ok just when i asked why it does_NEGFIRST go down, it goes down. LOL
bearish	$AAPL This is one weird day so far ...
bearish	$AAPL See today's downgrade coming. What a drama.
bearish	$AAPL Apple had a huge gap down on latest iPhone announcement. I am sure everyone remembers
bearish	$AAPL Looks like another day in the red, more waiting
bearish	$AAPL Release a time machine and disappoint analyst because it only works with iPhone.
bearish	$AAPL bad EOD for AAPL bad all day for AAPL bad bad bad bad bad bad bad
neutral	$AAPL aapl
neutral	$AAPL pivot http://stks.co/tR2K
neutral	$AAPL link?
neutral	@mryvoncedrac: $AAPL

Actions

To classify tweets as action-related, I follow the methodology in Bar-Haim et al. (2011), where messages are classified into facts and opinions. These authors were able to identify "trades" as an actual purchase or sale of the stock. Tweets reporting actual transactions can be identified based on recurrent language. For example, expressions such as "bot ... @" or "back in" or "sold 2day at" represent recurrent trade disclosures on StockTwits. I look at thousands of messages that represent actual actions and find some similar patterns and modes of speech that are used on StockTwits to represent trades.[8]

[8]Examples of rules are: "bot" or "bought" or "bght" in tweet AND ("yesterday" , " ago ") not in tweet.

Table 3.6 represents tweets that are first classified as actions and then classified as either "buy" or "sell" trades.

Table 3.6: Example of tweets that are classified as "Actions"

Action	Tweet
buy	$AMZN bought some 245 weekly calls
buy	$AMZN loaded up today at -$23 I love this damn stock. In since $585
buy	$AAPL just bought 615 shares
buy	$AAPL bought 500 pre mkt @ $169.05.
buy	$AAPL I bought 1000 share @ 169.42 this morning.
buy	$AAPL bought more here. Holding long
buy	$AAPL bought loads more here
sell	$AAPL sold at $4.20
sell	$AAPL sold at 1.95 for $1,200 profit
sell	$AMZN sold it all . Goodbye
sell	$AAPL Thanks. Just sold 100k
sell	$AAPL sold at $177...yest time to rebalance.
sell	$AAPL took the cash and ran!!! sold boys
sell	$AAPL out 90% sold for profit.

I classify self-disclosed hold sentiment as optimistic rather than neutral. Thus, all the trades that report "holding" or "still holding" on a stock are classified as "buy" actions.

I additionally hand-classified 1200 actions for future use with machine learning classification techniques as a training set and a test set. Nevertheless, the algorithm matching the essential vocabulary of action-related tweets seems appropriate and gives satisfactory results. The performance of classifiers is presented in the Appendix (see Table 3.14 and 3.15).

In total, I identify 440,105 actions (34% sales and 76% buys) for all the stocks. When actions are classified, there is a trade-off between the number of tweets classed as actions, increasing the robustness of the results, and the strictness of the classification algorithm. As the inclusion of more than 10,000 tweets is sufficient to obtain an average of 13 trades per day (see Table 3.7), I choose to focus on the precision of the algorithm.

User importance

It is fair to say that not all users have the same importance or the same audience. Thus, the beliefs or actions of users with a broader audience should give a stronger signal, translating into a larger influence power over the StockTwits investor community. In the same way, putting less weight on the sentiment/actions of users who are not popular on the platform is expected to reduce noise in the bullishness/bearishness of the signal. In addition, Bar-Haim et al. (2011) show that the correlation between tweet content and stock returns depends on the user's expertise. Based on Sprenger et al. (2014) findings, StockTwits users who provide recommendations of higher quality receive more likes and more followers. Thus, weighting each tweet with respect to user importance

increases the quality of the signal extracted from the message. Followers represent the regular audience of the users, whereas likes represent the current audience of a particular tweet. Those two measures should be highly correlated; however, they do not capture the same information. According to DeMarzo et al. (2003), agents' ability to persuade others with their opinions/actions depends on their position in the network. In addition to that statement, Da and Huang (2020) document that herding increases with the proportion of influential users' estimates in the public information set. I assume that users with a larger followership and like counts have a more central position in the network and a greater ability to persuade others.
The variables are weighted the following way for each tweet:

$$WeightedVariable_{tweet} = Variable_{tweet}(1 + \frac{\#like}{max(like)})(1 + \frac{\#followers}{max(followers)}), \quad (3.1)$$

where $max(like)$ and $max(followers)$, represent the maximum number of likes and followers, respectively, on all tweets discussing a given stock.

3.4.3 Descriptive Statistics

Table 3.7 presents summary statistics for the variables constructed with the messages on all stocks, aggregated at a daily frequency. Table 3.8 presents the correlation table. Several features of the statistics are worth noting. On average, there are more than 13 messages posted each day on each company, with 50% of the companies associated with more than two tweets per day. The average sentiment ($Sentiment_{i,t}$) and trade direction ($Action_{i,t}$) are higher than 0, indicating that the overall sentiment is bullish. Once weighted by user importance, the average daily sentiment and trade direction are higher still. In other words, popular users tend to be more bullish. Those averages are consistent with the US market trend in recent years, which has tended to be more bullish.
Conformity is on average equal to 0.78, meaning that the actions disclosed each day on the platform are often in the same direction. This result is consistent with the idea that members of the same community tend to act similarly. On average, 40% of the trades disclosed were successful, and 65% of the beliefs predicted next-day returns (the construction of the $SuccessfulTrading$ and $Quality$ variables is shown in Section A).

Table 3.7: Summary Statistics

The sample consists of all 2,154 stocks from 01/01/2013 to 07/08/2018. Variable definitions are provided in the Appendix.

	N	Mean	StdDev	Min	Median	Max
$MarketUncertainty_t$	1,376,152	32.65	52.92	4.80	15.65	716.48
$EcoUncertainty_t$	1,376,152	81.59	46.65	3.32	71.03	366.82
$\#Tweet_{i,t}$	1,376,152	13.64	79.77	0	2	10,530
$\#Like_{i,t}$	1,376,152	1,297.99	3,282.35	0	271.50	480,024
$\#Follower_{i,t}$	1,376,152	13,664.23	25,377.99	0	2,994	489,763
$Sentiment_{i,t}$	1,376,152	0.29	0.53	-1	0.25	1
$SentimentWeighted_{i,t}$	1,376,152	0.30	0.54	-2	0.25	2
$Action_{i,t}$	1,376,152	0.10	0.89	-32	0	158
$\#Bullish_{i,t}$	1,056,991	11.26	55.40	0	3	6,477
$\#Sell_{i,t}$	1,056,991	0.14	0.98	0	0	141
$\#Bearish_{i,t}$	1,056,991	5.68	31.22	0	1	3,862
$ActionWeighted_{i,t}$	1,056,991	0.13	1.05	-34.50	0	159.03
$\#Action_{i,t}$	1,056,991	0.41	2.48	0	0	304
$SuccessfulTrading_{i,t}$	1,376,152	0.40	0.49	0	0	1
$Quality_{i,t}$	1,376,152	0.65	0.47	0	1	1
$Conformity_{i,t}$	165,760	0.78	0.39	0	1	1
$ProbaBelief_{i,t}$	1,291,300	0.43	0.28	0	0.50	1
$Uncertainty_{i,t}$	1,291,300	0.16	0.10	0	0.23	0.25

Table 3.8: Correlation table

The sample consists of all 2,154 stocks from 01/01/2013 to 07/08/2018. Variable definitions are provided in the Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) $SentimentWeighted_{i,t}$	1										
(2) $MarketUncertainty_t$	-0.011	1									
(3) $\#Tweet_{i,t}$	-0.082	-0.006	1								
(4) $Sentiment_{i,t}$	0.997	-0.011	-0.082	1							
(5) $Action_{i,t}$	0.013	-0.010	0.564	0.013	1						
(6) $ActionWeighted_{i,t}$	0.015	-0.010	0.558	0.015	0.997	1					
(7) $\#Action_{i,t}$	-0.064	-0.005	0.931	-0.064	0.614	0.609	1				
(8) $SuccessfulTrading_{i,t}$	0.060	0.007	-0.030	0.061	0.014	0.014	-0.023	1			
(9) $Quality_{i,t}$	0.102	0.022	-0.016	0.103	0.001	0.001	-0.012	0.398	1		
(10) $Conformity_{i,t}$	0.116	0.001	-0.393	0.118	-0.114	-0.113	-0.405	0.181	0.019	1	
(11) $Uncertainty_{i,t}$	-0.836	0.008	0.156	-0.839	0.032	0.031	0.121	-0.091	-0.101	-0.215	1

3.5 Results

3.5.1 Rationality test

To determine the nature of the effect observed on next-period beliefs, namely, rational updating or behavioral biases, I perform a rationality test following Augenblick and Rabin (2018). Bayesian inference relies on the fact that changes in beliefs should, on average, be equal to a reduction in uncertainty (following the martingale hypothesis).

Given an agent observes the signals realized in each period $t > 0$ regarding two states $x \in$ ["*buy*","*sell*"], the uncertainty reduction and belief movement between t and $t + 1$ are given by:

$$UR_{t,t+1} = \pi_t(1 - \pi_t) - \pi_{t+1}(1 - \pi_{t+1}), \tag{3.2}$$

$$BM_{t,t+1} = (\pi_{t+1} - \pi_t)^2. \tag{3.3}$$

Where π_t is the probabilistic belief about a particular stock price going up (being in the "*buy*" state) at time t. Uncertainty at time t is $\pi_t(1 - \pi_t)$. The "*buy*" state represents the event of a rise in the stock price, whereas the "*sell*" state represents the event of a drop in the stock price.

If, on average, belief movement is larger than uncertainty reduction, the results are in favor of the base-rate neglect hypothesis (underweighting of priors). Underweighting of prior beliefs is in line with the persuasion and word-of-mouth theories.

To convert the beliefs calculated by sentiment analysis $B_t \in [-1, 1]$ into probabilistic beliefs $\pi_t \in [0, 1]$, I use the following transformation for $B_t \in [-1, 1]$:[9]

$$\pi_t = \frac{e^{tan(\frac{\pi}{2}B_t)}}{1 + e^{tan(\frac{\pi}{2}B_t)}}. \tag{3.4}$$

Under the null hypothesis of Bayesian updating, excess movement (i.e., the average difference between movement and uncertainty reduction for each belief change) is equal to zero, and the ratio of average movement divided by average uncertainty reduction is equal to one. I test this hypothesis for all the stocks in my sample. Two facts emerge from this exercise. First, members of the StockTwits community may not be fully rational Bayesian updaters. Second, on average, the excess movement is significantly higher than 0, indicating the nature of the bias: underweighting of prior beliefs. Thus, if, on average, beliefs tend to converge in the same direction, this is unlikely to be the consequence of investors overweighting prior beliefs. As a numerical example, for Apple Stock, the mean of average excess movement is equal to 0.0098, and the null hypothesis

[9]The reason for using an exponential transformation is that it can effectively magnify smaller differences between values at the lower end of the range, leading to a more uniform distribution of transformed values across the entire range.

is rejected with $Z = 5.46$ and $p < 0.001$. For the whole sample, I reject the null hypothesis of Bayesian updating for all stocks in the final sample.

3.5.2 Hypothesis

- **Hypothesis 1:** Agents' future beliefs are affected by previous **beliefs** and by previous **actions**.

  First, I estimate the impact of past beliefs on beliefs in the current day with the following regression:

$$Sentiment_{i,t} = \alpha + \beta X_{i,t-1} + \gamma SentimentWeighted_{i,t-1} + \nu_i + \epsilon_{i,t}, \tag{3.5}$$

  where t indexes day; i indexes the firm; X is a set of 1-day lagged control variables; ν_i is a firm fixed effect and $\epsilon_{i,t}$ is the random error term. $\epsilon_{i,t}$ is assumed to be heteroskedastic. The variables are standardized to facilitate comparison. The estimates measure the change in next-day average sentiment corresponding to a one-standard-deviation change in the independent variables. The variable $SentimentWeighted_{i,t}$ is defined as follows for each tweet:

$$SentimentWeighted_{i,t} = Sentiment_{i,t}(1 + \frac{\#like_{i,t}}{max(like_i)})(1 + \frac{\#followers_{i,t}}{max(followers_i)}). \tag{3.6}$$

  The coefficient of interest is γ, which measures the importance of average sentiment in the previous day in determining sentiment in the current day. $max(like_i)$ and $max(followers_i)$ are the maximum number of likes and followers, respectively, on all tweets discussing a given stock.

  Table 3.9 presents the results for this specification. The first column presents the results consisting of average belief in the previous period only. The results show a significant positive relationship between average sentiment in $t-1$ weighted by user importance and average sentiment in t. A one-standard-deviation change in average sentiment in t corresponds to an average change of 7% in future average sentiment. These findings are consistent with the notion of beliefs converging within the social media community.

Table 3.9: The effect of sentiment on next day sentiment

The table presents parameter estimates from panel regressions on Average $Sentiment_t$ with several different specifications. Each variable is standardized; reported estimates measure the change in $Sentiment_t$ corresponding to a one-standard deviation change in X. Firm fixed effects are included in the specification. The sample consists of all 2,154 stocks from 01/01/2013 to 07/08/2018. The t-statistics (in parentheses) are computed using standard errors robust to both clustering at the firm level and heteroskedasticity. Variables definitions are provided in the Appendix. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	(1) $Sentiment_t$	(2) $Sentiment_t$	(3) $Sentiment_t$	(4) $Sentiment_t$
$SentimentWeighted_{t-1}$	0.070***	0.069***	0.069***	0.070***
	(35.65)	(35.34)	(35.32)	(35.93)
$\#Tweet_{t-1}$		0.026***	0.026***	0.026***
		(18.54)	(18.49)	(18.14)
$MarketUncertainty_{t-1}$			-0.003***	-0.004***
			(-3.60)	(-4.16)
$SentimentWeighted_{t-1} \times Quality_{t-1}$				0.020***
				(13.35)
$Intercept_t$	-0.001***	-0.001***	-0.001***	-0.001***
	(-314.12)	(-71.47)	(-49.69)	(-12.33)
N	1,077,167	1,077,167	1,077,167	1,077,167
Firm fixed effects	Yes	Yes	Yes	Yes
R^2	0.004	0.005	0.005	0.006

Next, I incorporate multiple determinants in the specification. To measure communication, I use the number of tweets posted each day on a stock. The amount of communication each day on a particular stock presents a significant positive relationship with the next-day average sentiment. Thus, communication as a consequence of stock popularity is associated with an increase in bullishness. Antweiler and Frank (2004) show that bullish (buy) signals are stronger than bearish (sell) signals for investors. Thus, the more significant proportion of bullish messages in the data can explain the positive impact on next-day sentiment. To ensure that the effect of past beliefs is not just a reaction to public news, I use a daily market-related economic uncertainty measure developed by the Federal Reserve of St Louis based on newspapers in the United States. An increase in market uncertainty corresponds to a decrease in next-day average sentiment.

Second, I estimate the impact of past actions on current beliefs with the following regression:

$$Sentiment_{i,t} = \alpha + \beta X_{i,t-1} + \gamma_1 SentimentWeighted_{i,t-1} + \gamma_2 ActionWeighted_{i,t-1} + \nu_i + \epsilon_{i,t}. \tag{3.7}$$

The coefficients of interest are γ_1 and γ_2, measuring the impact of past beliefs and past actions on current period beliefs. Both parameters are positive and significant.

Table 3.10 allows a comparison of the absolute effect of average sentiment and average direction in trades in time $t-1$ on average sentiment in time t: The absolute effect of average sentiment is higher than the average sentiment of actions. One possible explanation is that only a low proportion of tweets are classified as actions compared to the number of tweets classified as beliefs. Another explanation is that the proportion of disclosed actions that predict market returns correctly is, on average, 40%, whereas the proportion of sentiment that predicts market returns correctly is on average 65%. In terms of market returns prediction, beliefs provide, on average, the most trustworthy signal. The daily numbers of trades and the daily number of buy trades are used as control variables. More communication through more daily trading disclosure leads to lower average sentiment. The previous effect is more substantial for the daily number of buy trades. An increase in the number of buy trades disclosed on StockTwits is associated with an increase in bearishness.

Table 3.10: The effect of trades on next day sentiment

The table presents parameter estimates from panel regressions on Average $Sentiment_t$ with several different specifications. Each variable is standardized; reported estimates measure the change in $Sentiment_t$ corresponding to a one-standard deviation change in X. Firm fixed effects are included in the specification. The sample consists of all 2,154 stocks from 01/01/2013 to 07/08/2018. The t-statistics (in parentheses) are computed using standard errors robust to both clustering at the firm level and heteroskedasticity. Variables definitions are provided in the Appendix. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	(1) $Sentiment_t$	(2) $Sentiment_t$	(3) $Sentiment_t$	(4) $Sentiment_t$
$SentimentWeighted_{t-1}$	0.0519***	0.052***	0.053***	0.052***
	(32.24)	(32.21)	(32.73)	(32.83)
$ActionWeighted_{t-1}$	0.001	0.003*	0.004**	0.009***
	(0.40)	(2.27)	(2.95)	(6.18)
$\#Actions_{t-1} \times ActionWeighted_{t-1}$		-0.001***	-0.001***	-0.001***
		(-3.59)	(-4.49)	(-5.63)
$\#Action_{t-1}$			-0.009***	-0.004**
			(-6.50)	(-2.95)
$\#Tweet_{t-1}$			0.018***	0.019***
			(14.36)	(16.39)
$MarketUncertainty_{t-1}$				-0.003*
				(-2.43)
$\#Buy_{t-1}$				-0.046***
				(-8.09)
$Intercept$	0.028***	0.029***	0.026***	0.005*
	(121.76)	(114.99)	(90.97)	(2.06)
N	828,838	828,838	828,838	828,838
Firm fixed effects	Yes	Yes	Yes	Yes
R^2	0.003	0.004	0.005	0.005

- **Hypothesis 2:** Agents' future beliefs are affected by the **quality of advice** related to stock actions on the previous day.

  Following Sprenger et al. (2014) measure of quality, the consequences of actions are defined as follows:

$$SuccessfulTrading_{i,t} = \begin{cases} 1 & \text{if } \frac{A_{i,t}}{R_{i,t+1}} > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{3.8}$$

  with $A_{i,t} = \{1, 0, -1\}$ being the sentiment of the average StockTwits action at time t (buy, no actions or sell) for each stock at time t and R_t the stock return at time t.

  To test the second hypothesis, I use the following specification:

$$
\begin{aligned}
Sentiment_{i,t} = \alpha + \gamma_1 SentimentWeighted_{i,t-1} + \gamma_2 ActionWeighted_{i,t-1} \\
+ \gamma_3 SuccessfulTrading_{i,t-1} + \beta X_{i,t-1} + \nu_i + \epsilon_{i,t}.
\end{aligned}
\tag{3.9}
$$

Table 3.11 presents the results for hypotheses 1 and 2 with the whole set of one-day lagged control variables. The quality of the average trading behavior at time $t-1$ negatively impacts the average investor sentiment at time t. One possible explanation for this result is the more significant proportion of buy trades in the action sample. After observing successful buy trades, day traders become more bearish as long as they expect a reversal in stock returns. The largest impact on next-day average sentiment comes from the success of the trades disclosed on the platform. This result is consistent with the credibility hypothesis. In contrast to the results for past beliefs, the average sentiment of the actions at time $t-1$ negatively predicts average sentiment in time t in this specification. Once the quality of the trades is taken into account by considering only the direction of past successful actions, average beliefs converge in the same direction, and the effect is ten times larger. This result aligns with the findings of Bursztyn et al. (2014), where choices made by peers affect investor decisions either through social learning or through social utility (when investors want to possess the same asset as their peers). To incorporate the effect of a different size audience, average numbers of likes and followers are added into the specification as control variables. The average number of likes per day serves as a proxy for the daily audience (i.e., the users who like have certainly seen the message), whereas the average number of daily followers represents the general audience. It turns out that an increase in the size of the current audience (likers) on the platform leads to an increase in bullishness. In concert with the relationship between the number of messages posted and next-day average sentiment, an increase in investor attention to a particular stock leads to an increase in bullishness. The second measure is difficult to interpret because the size of the followership of an account that posts a given message does not mean that the followers have seen the message. In this specification, the number of trades disclosed in the platform is associated with an increase in bullishness.

Table 3.11: The effect of consequences of trades on next day sentiment

The table presents parameter estimates from panel regressions on Average $Sentiment_t$ with several different specifications. Each variable is standardized; reported estimates measure the change in $Sentiment_t$ corresponding to a one-standard deviation change in X. Firm fixed effects are included in the specification. The sample consists of all 2,154 stocks from 01/01/2013 to 07/08/2018. The t-statistics (in parentheses) are computed using standard errors robust to both clustering at the firm level and heteroskedasticity. Variables definitions are provided in the Appendix. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	(1) $Sentiment_t$	(2) $Sentiment_t$	(3) $Sentiment_t$	(4) $Sentiment_t$
$SentimentWeighted_{t-1}$	0.053***	0.053***	0.052***	0.053***
	(33.64)	(33.60)	(33.49)	(34.22)
$ActionWeighted_{t-1}$	0.020***	0.005***	0.005***	-0.009***
	(17.36)	(4.23)	(4.19)	(-6.75)
$SuccessfulTrading_{t-1}$	-0.197***	-0.201***	-0.201***	-0.207***
	(-45.52)	(-47.31)	(-47.28)	(-47.70)
$ActionWeighted_{t-1} \times SuccessfulTrading_{t-1}$		0.037***	0.037***	0.038***
		(36.48)	(36.54)	(37.65)
$\#Like_{t-1}$			0.008***	0.008***
			(7.26)	(6.51)
$\#Follower_{t-1}$			0.006***	0.007***
			(4.90)	(5.75)
$MarketUncertainty_{t-1}$				-0.001
				(-1.35)
$\#Buy_{t-1}$				-0.024***
				(-3.94)
$\#Action_{t-1}$				0.042***
				(31.10)
$Intercept$	-0.034***	-0.039***	-0.040***	-0.053***
	(-26.16)	(-30.58)	(-31.98)	(-18.75)
N	828,838	828,838	828,838	828,838
Firm fixed effects	Yes	Yes	Yes	Yes
R^2	0.036	0.038	0.038	0.039

- **Hypothesis 3:** More communication between agents leads to more diversity.

  Following Antweiler and Frank (2004), conformity is defined as:[10]

$$Conformity_{i,t} = 1 - \sqrt{1 - (\frac{\#buy_{i,t} - \#sell_{i,t}}{\#buy_{i,t} + \#sell_{i,t}})^2}, \tag{3.10}$$

[10] Antweiler and Frank (2004) initially define a similar measure of "agreement" by taking sentiment instead of actions.

with $Conformity_{i,t} \in [0,1]$, and $\#buy_{i,t}$ and $\#sell_{i,t}$ are the number of "buy" trades and "sell" trades, respectively. If there are only "buy" trades, or "sell" trades respectively, reported on day t, the conformity measure is equal to 1. On days where the diversity of communication is highest (for example one "buy" trade and one "sell" trade are disclosed) the conformity measure is equal to 0.

In order to test Hypothesis 3, I use the following specification:

$$Conformity_{i,t} = \alpha + \beta X_{i,t-1} + \gamma_1 \#Tweets_{i,t-1} + \gamma_2 \#Actions_{i,t-1} + \nu_i + \epsilon_{i,t}. \tag{3.11}$$

The conformity measure helps to test hypothesis 3 directly in Table 3.12. A higher number of messages posted on the platform predicts lower conformity, in line with Ellison and Fudenberg (1995). Additionally, a higher number of trades predicts higher diversity. This result is consistent with the social learning channel of Bursztyn et al. (2014), where investors are influenced by the choices of others when making decisions. Social media users learn from other investors' choices when performing trading actions. The measure of the current audience on StockTwits also has a significant negative impact on conformity in behavior. Concerning conformity, the effect of an increasing number of buy trades is negative and significant.

Table 3.12: The effect of communication on conformity

The table presents parameter estimates from panel regressions on Average $Confomity_t$ with several different specifications. Each variable is standardized; reported estimates measure the change in $Conformity_t$ corresponding to a one-standard deviation change in X. Firm fixed effects are included in the specification. The sample consists of all 2,154 stocks from 01/01/2013 to 07/08/2018. The t-statistics (in parentheses) are computed using standard errors robust to both clustering at the firm level and heteroskedasticity. Variables definitions are provided in the Appendix. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	(1)	(2)	(3)
	$Conformity_t$	$Conformity_t$	$Conformity_t$
$\#Tweet_{t-1}$	-0.131***	-0.103***	-0.102***
	(-41.75)	(-26.90)	(-27.09)
$\#Action_{t-1}$		-0.041***	-0.013**
		(-10.71)	(-2.97)
$\#Like_{t-1}$			-0.021***
			(-5.02)
$\#Follower_{t-1}$			0.031***
			(6.16)
$\#Buy_{t-1}$			-0.101***
			(-14.93)
$Intercept$	0.056***	0.053***	0.028***
	(33.69)	(30.11)	(11.37)
N	130,163	124,731	124,731
Firm fixed effects	Yes	Yes	Yes
R^2	0.040	0.054	0.066

- **Hypothesis 4:** An increase in the advice quality of messages is associated with lower uncertainty.

  To test this hypothesis, I use the uncertainty measure computed for the rationality test. The quality variable is computed as follows, following Sprenger et al. (2014):

$$Quality_{i,t} = \begin{cases} 1 & \text{if } \frac{S_{i,t}}{R_{i,t+1}} > 0, \\ 0 & \text{otherwise}, \end{cases} \tag{3.12}$$

with $S_{i,t} = \{1, 0, -1\}$ being the average StockTwits sentiment at time t (bullish, neutral or bearish) and $R_{i,t}$ the stock return at time t. For the measure of quality and successful trading, I use the next-day return because, according to Antweiler and Frank (2004), the majority of users on StockTwits are "day traders" (with a trading horizon of one day).

In order to test Hypothesis 4, I estimate the following equation:

$$Uncertainty_{i,t} = \alpha + \beta X_{i,t-1} + \gamma Quality_{i,t-1} + \nu_i + \epsilon_{i,t}. \tag{3.13}$$

As controls for uncertainty, I use:

- the Economic Policy Uncertainty Index for the United States: from the Federal Reserve Bank of St Louis. This daily news-based index is based on newspapers in the United States.
- the Equity Market-related Economic Uncertainty Index, constructed in Baker et al. (2016).[11] This measure is similar to the first but without the policy-related term in newspapers. This measure is highly correlated with the VIX.

Additionally, I control for message volume (i.e., the number of tweets per day) following Sprenger et al. (2014), who find a significant positive relationship between volatility and message volume (i.e., the number of tweets per day), and Antweiler and Frank (2004), who show that message volume can predict volatility.

Table 3.13 shows that the average daily quality of the messages posted on StockTwits is associated with higher uncertainty. In other words, an increase in the quality of investment advice on the platform does not help reduce uncertainty. In line with Sprenger et al. (2014) and Antweiler and Frank (2004), message volume positively predicts next-period average uncertainty. An increase in bullishness/bearishness predicts a decrease/increase in next-day uncertainty, and this effect is significant at the 99% confidence level. Economic policy uncertainty, as well as market-related uncertainty, has no significant relationship with the uncertainty variable calculated from social media data. This result highlights the fact that social media data and sentiment are not always in line with newspaper articles.

[11] https://www.policyuncertainty.com/equity_uncert.html

Table 3.13: The effect of quality of communication on uncertainty

The table presents parameter estimates from panel regressions on Average $Uncertainty_t$ with several different specifications. Each variable is standardized; reported estimates measure the change in $Uncertainty_t$ corresponding to a one-standard deviation change in X. Firm fixed effects are included in the specification. The sample consists of all 2,154 stocks from 01/01/2013 to 07/08/2018. The t-statistics (in parentheses) are computed using standard errors robust to both clustering at the firm level and heteroskedasticity. Variables definitions are provided in the Appendix. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	(1) $Uncertainty_t$	(2) $Uncertainty_t$	(3) $Uncertainty_t$	(4) $Uncertainty_t$
$Quality_{t-1}$	0.126***	0.122***	0.132***	0.131***
	(36.41)	(38.35)	(40.76)	(41.34)
$Quality_t$		0.023***	0.030***	0.029***
		(11.19)	(14.42)	(13.82)
$\#Tweet_{t-1}$			0.080***	0.082***
			(53.06)	(57.39)
$MarketUncertainty_{t-1}$				-0.001
				(-0.46)
$SentimentWeighted_{t-1}$				-0.062***
				(-36.20)
$EcoUncertainty_{t-1}$				-0.001
				(-0.54)
$Intercept$	0.003***	0.003***	0.001***	0.001 ***
	(142.09)	(76.04)	(26.28)	(26.17)
N	1,011,163	1,011,163	1,011,163	1,011,163
Firm fixed effects	Yes	Yes	Yes	Yes
R^2	0.015	0.016	0.023	0.026

3.6 Conclusion

This paper studies how messages posted on the StockTwits platform affect investors' future beliefs. More than 50 million tweets on 2,154 stocks are classified as either beliefs (messages expressing opinions) or actions (messages disclosing actual transactions). The two message categories are assigned a sentiment score of -1 and 1, representing negative and positive views, respectively. The main questions are as follows: Are investors' future beliefs affected by their beliefs in the previous period? Are the beliefs affected by trades in the previous period? Are these effects influenced by the amount of communication and the size of the message's audience? Does the quality of communication and trades affect beliefs and uncertainty?

First, average sentiment on the microblog platform positively predicts next-day average sentiment, and this relationship is stronger once I account for the popularity of the message. Second, a change in average trade sentiment leads to a convergence of beliefs. Nevertheless, as the effect of actions is smaller than the effect of beliefs, it cannot be shown that actions speak louder than words. Third, the average daily quality of trades disclosed in messages helps predict investor sentiment in the next period but does not help reduce uncertainty. Fourth, an increasing amount of communication between users in the platform leads to diversity of opinion. Finally, I show that members of the StockTwits community are not entirely rational when changing their beliefs. This result provides evidence in favor of the theory of behavioral bias in opinion formation. To the extent of my knowledge, previous studies on StockTwits used a maximum data period of two years for all stocks or the entire data history for a smaller selection of stocks. This research is thus the first to analyze all the tweets available from StockTwits data.

3.7 Apppendix

A Variables definitions

- $Sentiment_{i,t}$: Average sentiment from StockTwits users $\in [-1, 1]$ (from the most bearish to the most bullish sentiment) for stock i at day t.
- $SentimentWeighted_{i,t}$: Average sentiment from StockTwits users $\in [-1, 1]$ for stock i at day t, weighted by user importance (corresponding to the number of likes associated to the tweet and the number of followers of the user).
- $Action_{i,t}$: Sum of trades directions from StockTwits users for stock i at day t.
- $ActionWeighted_{i,t}$: Sum of trades directions from StockTwits users for stock i at day t, weighted by user importance.
- $\#Tweet_{i,t}$: Number of messages posted for stock i at day t.
- $\#Action_{i,t}$: Number of trades disclosed on StockTwit for stock i at day t.
- $\#Buy_{i,t}$: Number of buy trades disclosed on StockTwit for stock i at day t.
- $\#Sell_{i,t}$: Number of sell trades disclosed on StockTwit for stock i at day t.
- $\#Like_{i,t}$: Sum of the likes associated with messages posted for stock i at day t.
- $\#Follower_{i,t}$: Sum of the followers associated with messages posted by Stocktwits users for stock i at day t.
- $Quality_{i,t}$: Binary variable equal to one if the StockTwits average sentiment for stock i at day t can predict the direction of the next day return for stock i, and 0 otherwise.
- $SuccessfulTrading_{i,t}$: Binary variable equal to one if the StockTwits average trades directions for stock i at day t can predict the direction of the next day return for stock i, and 0 otherwise.
- $Conformity_{i,t}$: Conformity measure on all trades disclosed on the platform for stock i at day t.
- $MarketUncertainty_t$: Economic Policy Uncertainty Index for the United States, from the Federal Reserve Bank of St Louis.
- $EcoUncertainty_t$: Equity Market-related Economic Uncertainty Index, from the Federal Reserve Bank of St Louis.

B Action classification algorithm

To evaluate the performance of the action classifier, I present Accuracy measure in Table 3.14. Accuracy is defined as the number of correctly classified tweets.

Table 3.14: Performance evaluation for "Action" classifier

	% Correct Classification
Apple	89.3%

I use different measures to evaluate the performance of the "buy" and "sell" classifier. This step of the classification takes place once messages are already classified as actions to identify the direction of the trade.

Precision measure is defined as the probability of a message disclosing a "buy" given that the classifier assigns the action message to a "buy" category:

$$Precision = \frac{TP}{(TP + FP)}, \tag{3.14}$$

with TP the true positives (messages that are correctly classified as "buy") and FP the false positives (messages that the classifier has incorrectly classified as "buy").

The Recall measure is defined as the probability that the classifier assigns the action message to a "buy" category given that the message is disclosing a "buy":

$$Recall = \frac{TP}{(TP + FN)}, \tag{3.15}$$

with FN the false negatives (messages that the classifier has incorrectly classified as "sell").

Finally, F1-score combine both Precision and Recall:

$$F1score = \frac{2}{(\frac{1}{Precision} + \frac{1}{Recall})}. \tag{3.16}$$

Table 3.15: Performance evaluation for "Buy" and "Sell" classifier

	% Correct Classification	Precision	Recall	F1-measure
Apple	82.5%	90.6%	96.9%	93.2%

Bibliography

Agrawal, S., Azar, P. D., Lo, A. W. and Singh, T. (2018). Momentum, mean-reversion, and social media: Evidence from stocktwits and twitter, *The Journal of Portfolio Management* **44**(7): 85–95.

Andrei, D. and Cujean, J. (2017). Information percolation, momentum and reversal, *Journal of Financial Economics* **123**(3): 617–645.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards, *The Journal of Finance* **59**(3): 1259–1294.

Augenblick, N. and Rabin, M. (2018). Belief movement, uncertainty reduction, and rational updating, *UC Berkeley-Haas and Harvard University Mimeo* . Working Paper.

Baker, S. R., Bloom, N. and Davis, S. J. (2016). Measuring economic policy uncertainty, *The Quarterly Journal of Economics* **131**(4): 1593–1636.

Banerjee, A. V. (1992). A simple model of herd behavior, *The Quarterly Journal of Economics* **107**(3): 797–817.

Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M. and Goldstein, G. (2011). Identifying and following expert investors in stock microblogs, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 1310–1319.

Barberis, N., Greenwood, R., Jin, L. and Shleifer, A. (2015). X-CAPM: An extrapolative capital asset pricing model, *Journal of Financial Economics* **115**(1): 1–24.

Bartov, E., Faurel, L. and Mohanram, P. S. (2018). Can twitter help predict firm-level earnings and stock returns?, *The Accounting Review* **93**(3): 25–57.

Batra, R. and Daudpota, S. M. (2018). Integrating stocktwits with sentiment analysis for better prediction of stock price movement, *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, IEEE, pp. 1–5.

Berry, S., Mitra, G. and Sadik, Z. (2019). Improved volatility prediction and trading using stocktwits sentiment data, *Available at SSRN 3527557* .

Bordalo, P., Gennaioli, N., Porta, R. L. and Shleifer, A. (2019). Diagnostic expectations and stock returns, *The Journal of Finance* **74**(6): 2839–2874.

Broadstock, D. C. and Zhang, D. (2019). Social-media and intraday stock returns: The pricing power of sentiment, *Finance Research Letters* **30**: 116–123.

Bursztyn, L., Ederer, F., Ferman, B. and Yuchtman, N. (2014). Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions, *Econometrica* **82**(4): 1273–1301.

Chen, C. Y.-H., Despres, R., Guo, L. and Renault, T. (2019). What makes cryptocurrencies special? Investor sentiment and return predictability during the bubble, *Investor Sentiment and Return Predictability During the Bubble (June 3, 2019)* .

Colla, P. and Mele, A. (2009). Information linkages and correlated trading, *The Review of Financial Studies* **23**(1): 203–246.

Cookson, J. A. and Niessner, M. (2018). Why don't we agree? Evidence from a social network of investors, *Evidence from a Social Network of Investors (August 6, 2018)* .

Cortez, P., Oliveira, N. and Ferreira, J. P. (2016). Measuring user influence in financial microblogs: Experiments using stocktwits data, *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, pp. 1–10.

Da, Z. and Huang, X. (2020). Harnessing the wisdom of crowds, *Management Science* **66**(5): 1847–1867.

DeMarzo, P. M., Vayanos, D. and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions, *The Quarterly Journal of Economics* **118**(3): 909–968.

Devenow, A. and Welch, I. (1996). Rational herding in financial economics, *European Economic Review* **40**(3-5): 603–615.

Ellison, G. and Fudenberg, D. (1995). Word-of-mouth communication and social learning, *The Quarterly Journal of Economics* **110**(1): 93–125.

Enke, B. and Zimmermann, F. (2019). Correlation neglect in belief formation, *The Review of Economic Studies* **86**(1): 313–332.

Foerster, M. (2018). Finite languages, persuasion bias, and opinion fluctuations, *Journal of Economic Behavior & Organization* **149**: 46–57.

Graham, J. R. and Harvey, C. R. (1997). Grading the performance of market-timing newsletters, *Financial Analysts Journal* **53**(6): 54–66.

Hill-Kleespie, A. (2018). The good the bad and the trending: Microblogging sentiment and short term momentum, *9th Miami Behavioral Finance Conference*.

Hong, H., Kubik, J. D. and Stein, J. C. (2005). Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, *The Journal of Finance* **60**(6): 2801–2824.

Karampatsas, N., Malekpour, S. and Mason, A. (2017). Beyond market mood: Stock sentiment and the response to corporate earnings announcements, *Technical report*, University of Surrey.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-ks, *The Journal of Finance* **66**(1): 35–65.

Man, X., Luo, T. and Lin, J. (2019). Financial sentiment analysis (fsa): A survey, *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, IEEE, pp. 617–622.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T. and Trajanov, D. (2020). Evaluation of sentiment analysis in finance: From lexicons to transformers, *IEEE Access* **8**: 131662–131682.

Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon, *Computational intelligence* **29**(3): 436–465.

Molavi, P., Tahbaz-Salehi, A. and Jadbabaie, A. (2018). A theory of non-bayesian social learning, *Econometrica* **86**(2): 445–490.

Ozsoylev, H. N., Walden, J., Yavuz, M. D. and Bildik, R. (2013). Investor networks in the stock market, *The Review of Financial Studies* **27**(5): 1323–1366.

Pouget, S., Sauvagnat, J. and Villeneuve, S. (2017). A mind is a terrible thing to change: Confirmatory bias in financial markets, *The Review of Financial Studies* **30**(6): 2066–2109.

Shiller, R. J. (2015). *Irrational exuberance: Revised and expanded third edition*, Princeton University Press.

Siganos, A., Vagenas-Nanos, E. and Verwijmeren, P. (2017). Divergence of sentiment and stock market trading, *Journal of Banking & Finance* **78**: 130–141.

Sprenger, T. O., Tumasjan, A., Sandner, P. G. and Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs, *European Financial Management* **20**(5): 926–957.