

Using Twitter trust network for stock market analysis

Yefeng Ruan*, Arjan Durreesi, Lina Alfantoukh

Department of Computer and Information Science, Indiana University Purdue University Indianapolis, Indianapolis, IN 46202, USA



ARTICLE INFO

Article history:

Received 27 September 2017

Revised 15 January 2018

Accepted 19 January 2018

Available online 31 January 2018

Keywords:

Twitter trust network
Trust management
Stock abnormal return
Sentiment analysis
Linear regression

ABSTRACT

Online social networks are now attracting a lot of attention not only from their users but also from researchers in various fields. Many researchers believe that the public mood or sentiment expressed in social media is related to financial markets. We propose to use trust among users as a filtering and amplifying mechanism for the social media to increase its correlation with financial data in the stock market. Therefore, we used the real stock market data as ground truth for our trust management system. We collected stock-related data (tweets) from Twitter, which is a very popular Micro-blogging forum, to see the correlation between the Twitter sentiment valence and abnormal stock returns for eight firms in the S&P 500. We developed a trust management framework to build a user-to-user trust network for Twitter users. Compared with existing works, in addition to analyzing and accumulating tweets' sentiment, we take into account the source of tweets – their authors. Authors are differentiated by their power or reputation in the whole community, where power is determined by the user-to-user trust network. To validate our trust management system, we did the Pearson correlation test for an eight months period (the trading days from 01/01/2015 through 08/31/2015). Compared with treating all the authors equally important, or weighting them by their number of followers, our trust network based reputation mechanism can amplify the correlation between a specific firm's Twitter sentiment valence and the firm's stock abnormal returns. To further consider the possible auto-correlation property of abnormal stock returns, we constructed a linear regression model, which includes historical stock abnormal returns, to test the relation between the Twitter sentiment valence and abnormal stock returns. Again, our results showed that by using our trust network power based method to weight tweets, Twitter sentiment valence reflect abnormal stock returns better than treating all the authors equally important or weighting them by their number of followers.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Online social media (e.g. Twitter) is becoming more popular, as it is easier for users to post and spread information than with traditional media. With more users joining in online social networks, more data is available. Therefore, many data-driven applications, such as disaster detection [1], election predictions [2,3], information filtering [4], opinion mining [5–7] and so on benefit from this trend. Among them, financial market analysis is one of the most attractive fields and has attracted a lot of attention [8–12].

The stock market is a very hot topic in the field of finance and economics. Many researchers try to analyze and predict stock returns based on various types of theories [13,14]. For example, Chartist theory [15] assumes that the stock market's past behavior patterns will recur in the future. Thus we can predict future stock

returns by using historical data. In contrast to Chartist theory, Random Walk theory [16,17] considers stock returns as identical independent variables. Although these theories' assumptions are different, many existing works use historical stock market data, such as open price, close price, daily trade volume and so on, to predict future stock returns.

Besides historical stock market performance, investors' decisions can be affected by news [18] and media [8,10,19–21]. Also, public mood or sentiment which is reflected in media plays an important role in investors' decision making processes [22,23]. Investors' decisions in turn can affect stock market. Therefore, stock market is related with public mood in news or media.

With the popularity of Twitter and its easy-to-use open Application Programming Interfaces (APIs), there exist many works that use Twitter as a platform to analyze and predict stock market activities, including both indicator-level and firm-level analysis [9,11,12,24,25]. In addition to academic researchers, firms are also paying attention to Twitter for their commercial purposes. Many firms use Twitter to interact with their investors and customers

* Corresponding author.

E-mail addresses: yefruan@iupui.edu (Y. Ruan), adurreesi@iupui.edu (A. Durreesi), lalfanto@umail.iu.edu (L. Alfantoukh).

[26]. Compared with traditional media, Twitter is efficient. To use Twitter to analyze stock market, typically Twitter feeds (tweets) are first analyzed by sentiment analysis tools to extract their sentiment, then tweet sentiments are aggregated together. Aggregated Twitter sentiment valence is then used for financial market analysis. Most widely used sentiment analysis tools generate binary results (positive or bullish vs. negative or bearish), although some sentiment analysis can generate multi-level sentiment results.

The main hypothesis of this work is that the users reputation, built by the inter trust among them, using our trust management system, helps in making better predictions of the stock market. To verify this hypothesis and to validate our trust management system, we collected stock-related data from Twitter to see the correlation between Twitter sentiment valence and abnormal stock returns. Therefore, the correlation between Twitter sentiment valence, filtered by our trust management system, and abnormal stock returns served as ground truth for our trust management system. We selected eight firms which are the top eight mentioned firms (which have the largest number of tweets) in our data set. The reason we selected these eight firms is that, for other firms, the average number of daily tweets is low. Based on only a small number of tweets, we think that the analysis is not reliable. For the selected eight firms we collected their stock data correspondingly from Yahoo! Finance. As indicated in [27], the source (users) of tweets is also an important factor. Therefore, unlike many existing works which treat all the authors equally important or ignore authors' identities, in addition to analyzing tweets' sentiment, we also take into account tweets' authors. We adapted our trust management framework [28] and constructed a user-to-user trust network for Twitter users based on their tweeting behaviors. Then, users were differentiated by their reputation or power in the whole community, where reputation or power is determined by the user-to-user trust network. Furthermore, to aggregate tweets together for Twitter sentiment valence, each tweet is weighted by its author's power.

To compare our approach to other ones, we used the Pearson correlation tests among results for eight months time (the trading days from 01/01/2015 through 08/31/2015). Compared to treating all the authors equally important or weighting them by their number of followers, our trust network based reputation mechanism amplifies the correlation between a specific firm's Twitter sentiment valence and the firm's stock abnormal returns. To further consider the possible auto-correlation property of abnormal stock returns and to test the relation between Twitter sentiment valence and abnormal stock returns, we constructed a linear regression model, which includes historical stock abnormal returns. Again, our results show that by using our trust network power based method to weight tweets, Twitter sentiment valence reflect abnormal stock returns better than other two methods, that is treating all the authors equally important or weighting them by their number of followers.

The remaining portion of this paper is organized as follows: in Section 2, we introduce some background knowledge and literature works in this field. In Section 3, we introduce our trust management framework and adapt it to Twitter. Also, we propose a simple method to calculate for users' power or reputation. In Section 4, we illustrate how we aggregate Twitter sentiment valence for the firms. And we propose our trust network power based method as well as other two baseline methods. In Section 5, we give detailed information about the data sets we used in this work. Also, we compare our trust network power based method with other two baseline methods regarding Pearson correlation coefficients and a linear regression model. In Section 6, we conclude our work and list several limitations of applicability of this work as well as some potential future work.

2. Background and related works

Twitter, as one of the most popular online social media platforms, provides its users the ability to share and spread their opinions. It also enables users who have the same interests to form groups. The stock market is among one of the hottest topics among Twitter users. There are many stock market-related groups or gurus on Twitter, such as StockTwits, FinancialTimes, Market-Watch, and so on. Recent research has shown that investors are likely to post financial news or articles and share their opinions on Twitter [29]. Compared with traditional media, Twitter feeds can be incorporated instantly into stock prices. Therefore, Twitter has become a widely used platform for researchers to analyze and predict stock returns.

As Refs. [8,19,20] pointed out, investors' emotions or sentiments can be reflected by the stock market. Negative sentiment or pessimism on social media might induce a stock price to drop. Positive sentiment is more likely to induce stock prices to increase than neutral or pessimism sentiment. Therefore, given users' text (tweets), natural language processing methods are needed to analyze investors' emotions. There exist many sentiment analysis tools. Roughly, they can be divided into two categories: word count analysis strategy and machine learning strategy. Word count analysis strategy uses dictionaries to determine sentiment for each word and then aggregate words' sentiment together. Most commonly used dictionaries in this field include Harvard-IV dictionary [30] and Loughran and McDonald's financial dictionary [31]. Among machine learning methods, most of them are classifiers, such as Naive Bayes classifier, SVM classifier, and so on. One of the problems of the machine learning strategy is that it requires a set of labeled training data, which might need a huge load of manual work. In this work, we use an existing sentiment analysis tool – SentiStrength [32], which is designed for short informal text.

Twitter sentiment valence is then measured based on the detected positive and negative tweets. Various Twitter sentiment valence measurements are used in literature [8,11,29,33]. In principle, Twitter sentiment valence measures the ratio of positive tweets to negative tweets. To investigate the linear relation between Twitter sentiment valence and stock prices or stock returns, Pearson correlation coefficients [12,29,34] and beta coefficients of linear regression models [9,11,33] are widely used in literature.

Existing works in this field can be divided into two categories based on their focus. Indicator-level works mainly focus on indicators, such as Dow Jones Industrial Average Index, NASDAQ, S&P 500 index, and so on. This type of work focuses on the whole industry. Indicator-level works include [8,9,35,36]. More recently, researchers are also paying much attention to firm-level works; as the name itself indicates, instead of investigating the whole industry, this type of work focuses on specific firms. Refs. [11,29,33,34,37,38] belong to firm-level works. In this paper, we focus on specific firms.

Bollen et al., [9] used OpinionFinder and Google-Profile of Mood States (GPOMS) to measure sentiment for tweets. Rather than outputting binary sentiment results (OpinionFinder), GPOMS measures sentiment in six dimensions, which includes calm, alert, sure, vital, kind, and happy. And it showed that only calm is related to Dow Jones Industrial Average Index. Tetlock [8] did experiments with Wall Street Journal, and mainly focused on the pessimism score of the media. It showed that high media pessimism scores caused the drop in stock market prices. Zhang et al., [36] classified tweets into fear, worry, and hope based on the corresponding words. It showed that Twitter sentiment (fear, worry, and hope) is negatively correlated with Dow Jones Industrial Average Index, NASDAQ and S&P 500 index. Similarly, Gilbert and Karahalios [35] measured anxiety, worry, and fear in LiveJournal,

and it turned out that they were negatively related to the S&P 500 index.

Smailovic et al., [33] calculated positive sentiment probabilities by dividing the number of positive tweets by the total number of tweets. It then analyzed eight firms' stock returns and their positive sentiment probabilities by using the Granger causality test [39]. Instead of focusing on stock returns, Ranco et al., [34] measured the Pearson correlation coefficient between 30 firms' abnormal returns and their Twitter sentiment valence. Its focus was on event detection and events' relation with abnormal returns. Similarly, Sul et al., [11] measured the relation between Twitter sentiment valence and abnormal returns via analyzing beta coefficients of linear regression models. In [29], besides investigating the relation among Twitter sentiment valence, stock returns, message volume and trading volume, Sprenger et al., showed that tweets from users who always provide good advice are more likely to be retweeted more than other users.

However, none of those mentioned above works takes into account tweets' authors. Even in [29], it only investigated the relation between the advice quality and the number of retweets. It did not differentiate their authors. In the remainder of this work, given historical Twitter data, we adapted our trust management framework [28] to measure user-to-user trust relationships and then construct a trust network for the whole community. Based on the trust network within the community, we then derived users' reputation or power, which is used later as weights in the process of Twitter sentiment aggregation. Through weighting each tweet by its author's reputation or power, we can amplify the correlation between specific firms' Twitter sentiment valence and their abnormal stock returns. Also, we show that to get reliable analysis, a sufficient number of tweets must be available.

3. Trust network for Twitter

Trust, which is a subjective concept, plays an extremely important role in people's daily lives. Actually, we use our trust estimations or trust networks to make decisions in our lives [40–42]. For example, among several service providers, we might choose the one who has the highest rating. On Twitter, given a huge number of subscribers, trust is very important for users to differentiate among other users. A user might have thousands of followers or friends on Twitter; however, not all of them are acquaintances. Huberman et al., [43] differentiated “claimed friends” from “real friends” on Twitter by counting the number of interactive tweets that two users post towards each other. To handle trust on large online social media sites, such as Facebook and Twitter, we need support from computer systems. Therefore, we have to represent trust computationally, which we call trust modeling in this work. A lot of research has been done in this field [44–46].

Typically, only very few users are directly connected through trust network on Twitter [47]. Therefore, besides representing trust in a computational way, we also need a framework that can effectively infer Friend of a Friend (FOAF) relationships, as user-to-user direct trust relationships are not sufficient in sparsely connected online social networks. In this paper, we call this trust inference. In [28], we proposed a measurement theory based trust management framework which addresses both trust modeling and trust inference. We represent our trust management framework in Fig. 1.

Trust Modeling maps the available trust related raw data (interactive tweets) from Twitter into trust metrics. Trust Inference infers indirect trust relationships by propagating and aggregating direct trust relationships over the whole network or over the part of particular interest network. In Decision Making, which is stock market analysis in our case, we use the produced trust knowledge obtained by trust management framework to support decision-making processes. All three phases of trust processing are

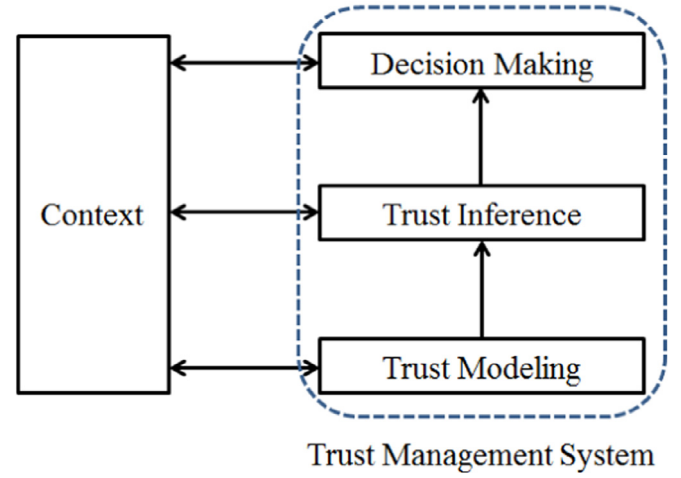


Fig. 1. Three phases of trust processing.

dependent on the context [48]. For example, depending on the type of decisions, or the risks involved, we would appropriately map the raw trust data into defined trust metrics.

3.1. Trust components

Our trust management framework is based on measurement theory. On Twitter, we treat each interactive tweet as a measurement that the truster has towards the trustee. Similar to physical measurement cases, after doing a set of measurements, we get a set of measured results, which are not necessarily the same due to several factors (i.e., measurement error, random noise, and so on). From a set of measured results, we derive a comprehensive summary of the trustee's impression or trustworthiness. Similar to the averaging of sample measurements in statistics, we use the mean of measurements to represent impression and denote it as m in this paper. Besides impression, similar to sampling in statistics, the truster would have a distribution of measured results in a range around the summarized impression. Such a distribution, which in fact shows to what extent the truster is confident about her/his estimation, is similar to “error” (standard error of the mean) in physical measurements [49]. We use confidence (denoted as c) to capture the measurement “error”. Therefore, our trust metric is composed of two components: impression (m) and confidence (c). Impression measures how good or trustworthy the trustee is in the truster's point of view. Confidence measures how confident the truster is about her/his impression estimation.

To consider impression and confidence together, we denote trust as $T(m, c)$, where m and c can be calculated in Eq. (1). Impression m is the mean of measurements. And confidence c is inversely related to an error. The higher the error, the lower is the confidence. For each measurement result m_i , we normalize it into the range of [0, 1]. Therefore, both m and c are in the range of [0, 1]. The higher the impression value, more trustworthy the trustee is. Similarly, the higher the confidence value, more confident is the truster is about the corresponding impression estimation.

$$m = \frac{\sum_{i=1}^N m_i}{N} \text{ and } c = 1 - 2 * r \text{ where } r = \sqrt{\frac{\sum_{i=1}^N (m_i - m)^2}{N * (N - 1)}} \quad (1)$$

3.2. Trust modeling for Twitter

Our trust management framework is based on measurements, so we need to first collect measurements on Twitter. Also, remember that trust relationship is user-to-user. We want to analyze the

truster's trust estimation towards the trustee. Therefore, we collect three special types of tweets: mentions, replies, and retweets. Honey et al., called them interactive tweets in [50]. They are used for communication among Twitter users. We consider them as conversations among users on Twitter and treat them as measurements in our framework. One common feature of these three types of tweets is that they all contain the symbol "@", followed by the trustees that the truster want to communicate with.

Tweets reflect the truster's opinion or sentiment towards the trustee. Similar to Kale [51], we build up trust based on sentiment analysis results. There are many existing sentiment analysis tools. As in [52], we select SentiStrength [32] as our sentiment analysis tool. To test the accuracy of SentiStrength, we used a Yelp dataset which includes both text reviews and ratings in [28]. We used the scalar outputs (from -4 to +4) of SentiStrength, and normalized sentiment results and Yelp ratings into a range of [0, 1]. Our test results showed that SentiStrength's sentiment results were very close to the actual Yelp ratings (the mean absolute error was equal to 0.224).

By treating interactive tweets as measurements, we can calculate impression m by following Eq. (1). Instead of just averaging all the tweets, we divide them into different time windows based on their posted date. We cluster tweets posted within the same month into the same window. As we will see later, we treat these windows differently. In each window, we group tweets based on their posted dates. On each day, we treat them equally and calculate the mean impression m_d for that specific day by following Eq. (1), as well as c_d .

After calculating m and c for each day, we use the weighted mean to combine the results for each month (time window) in Eq. (2). Here we use $w_i = \frac{1}{r_i^2}$, to assign higher weights to those who have higher confidence. Correspondingly, the error of the weighted mean is expressed in Eq. (3).

$$m_{\text{month}} = \frac{\sum_{i=1}^{i=31} w_i m_i}{\sum_{i=1}^{i=31} w_i} \quad (2)$$

$$e_{\text{month}}^2 = \frac{1}{\sum_{i=1}^{i=31} w_i} \quad (3)$$

Similar to Sun et al., [53], we assume that impression can fade with time. For example, if the truster just evaluated the trustee a few days ago, she/he may be quite confident about the estimation; however, if the truster "measured" the trustee several months ago, the impression has somehow faded. So we introduce a forgetting factor σ , where σ is less than 1, to capture this effect on the truster's confidence. As we are going to use tweets posted from 01/01/2015 through 08/31/2015 to analyze stock market, we use tweets which were posted in the year of 2014 (before the stock market analysis period) to construct the trust network. Therefore, we have 12 time windows in total. The confidence of December, which is the most recent month, is not discounted. The confidence of November is discounted by σ , and the confidence of October is discounted by σ^2 , and so on. We show this effect of forgetting factor in Eq. (4), where i is the number of the corresponding month (i.e. $i = 1$ for January).

$$c'_i = c_i * \sigma^{12-i} \quad (4)$$

Similar to Eqs. (2) and (3), we combine 12 time windows' results using the weighted mean where weights are confidence c' . In this work, we select one month as the length of the time window and the forgetting factor $\sigma = 0.95$. Further refinement of these parameters will be part of our future work.

3.3. Trust inference

As indicated by psychologists and sociologists, such as Wasserman and Faust [54] and Holland and Leinhardt [55], trust is prop-

agative. There are two types of inference operations used to capture trust propagation: trust transitivity and trust aggregation [28].

Based on the trust transitivity, if A trusts B, and B trusts Z, to some extent A will also trust Z. We denote this case as: $m_Z^{A:B} = m_B^A \otimes m_Z^B$. This is a concatenation of trust path (a chain of trust relationships) A-B and B-Z using node B as a connecting node for trust transitivity.

Given that there exist multiple paths between A and Z, we use trust aggregation to aggregate them. For example, if A can learn Z through B and C, then it can be denoted as $m_Z^{A:(B,C)} = m_Z^{A:B} \oplus m_Z^{A:C}$.

To calculate confidence for the propagated trust, we use the error propagation theory [56]. If we represent transitivity and aggregation operators in a general function as $f(m_1, m_2)$, then the error of function (r in Eq. (1)) can be computed as in Eq. (5) [57]:

$$r_f^2 = \left(\frac{\partial f}{\partial m_1} \right)^2 r_1^2 + \left(\frac{\partial f}{\partial m_2} \right)^2 r_2^2 + 2 \frac{\partial f}{\partial m_1} \frac{\partial f}{\partial m_2} \text{cov}(m_1, m_2) \quad (5)$$

Here $\text{cov}(m_1, m_2)$ is the covariance between m_1 and m_2 . In case that m_1 and m_2 are independent, the covariance becomes zero. We can see that r_f can be calculated for any format of arithmetic operators using the error propagation theory. Therefore, our framework is very flexible and can be adapted to various transitivity and aggregation operators.

In [28], we collected three trust transitivity operators and five trust aggregation operators. These are all commonly used operators in literature, and each operator has its meaning. We use the same operators in this work. In the following, we list transitivity operators as TP , and list aggregation operators as AP . Also, we list equations for calculating corresponding errors for the transitivity operators and aggregation operators. Note that, to calculate confidence, we assume that the covariances in Eq. (5) are equal to 0. For the meanings and sources of these operators, please refer to Ruan et al., [28].

$$m_B^A \otimes m_Z^B = m_B^A * m_Z^B \quad (TP1)$$

$$r_B^A \otimes r_Z^B = \sqrt{(m_Z^B)^2 * (r_B^A)^2 + (m_B^A)^2 * (r_Z^B)^2} \quad (TP1)$$

$$m_B^A \otimes m_Z^B = m_B^A * m_Z^B + (1 - m_B^A) * (1 - m_Z^B) \quad (TP2)$$

$$r_B^A \otimes r_Z^B = \sqrt{(2 * m_B^A - 1)^2 * (r_Z^B)^2 + (2 * m_Z^B - 1)^2 * (r_B^A)^2} \quad (TP2)$$

$$m_B^A \otimes m_Z^B = m_{\min} = \min(m_B^A, m_Z^B) \quad (TP3)$$

$$r_B^A \otimes r_Z^B = \min(r_i \text{ where } m_i = m_{\min}) \quad (TP3)$$

$$m_Z^{A:B} \oplus m_Z^{A:C} = \frac{m_Z^{A:B} + m_Z^{A:C}}{2} \quad (AP1)$$

$$r_Z^{A:B} \oplus r_Z^{A:C} = \sqrt{\frac{1}{2^2} ((r_Z^{A:B})^2 + (r_Z^{A:C})^2)} \quad (AP1)$$

$$m_Z^{A:B} \oplus m_Z^{A:C} = \frac{w_1 * m_Z^{A:B} + w_2 * m_Z^{A:C}}{\sum w_i} \quad (AP2)$$

$$r_Z^{A:B} \oplus r_Z^{A:C} = \sqrt{\frac{1}{(\sum w_i)^2} (w_1^2 * (r_Z^{A:B})^2 + w_2^2 * (r_Z^{A:C})^2)} \quad (AP2)$$

$$m_Z^{A:B} \oplus m_Z^{A:C} = m_Z^{A:B} + m_Z^{A:C} - m_Z^{A:B} * m_Z^{A:C} \quad (AP3)$$

$$r_Z^{A:B} \oplus r_Z^{A:C} = \sqrt{(1 - m_Z^{A:C})^2 * (r_Z^{A:B})^2 + (1 - m_Z^{A:B})^2 * (r_Z^{A:C})^2} \quad (AP3)$$

$$m_z^{A:B} \oplus m_z^{A:C} = m_{\max} = \max(m_z^{A:B}, m_z^{A:C}) \quad (\text{AP4})$$

$$r_B^{A:B} \oplus r_z^{A:C} = \max(r_i \text{ where } m_i = m_{\max}) \quad (\text{AP4})$$

$$r_B^{A:B} \oplus r_z^{A:C} = r_{\max}, \text{ where } c_{\max} = \max(c_z^{A:B}, c_z^{A:C}) \quad (\text{AP5})$$

$$m_z^{A:B} \oplus m_z^{A:C} = \max(m_i \text{ where } c_i = c_{\max}) \quad (\text{AP5})$$

As we indicated in [28], the user-to-user Twitter trust network built in Section 3.2 is sparsely connected. We use the above three transitivity operators and five aggregation operators, which results in 15 combinations of operators, to infer indirect trust relationships. We call the number of links from the truster to the trustee hops. For example, if A knows Z through B, in this case, we say that it has two hops. On the one hand, by increasing the number of hops, we can have more pairs of users being connected. On the other hand, the larger number of hops, the lower is the accuracy of the inferred indirect trust [58]. For the trade-off between them, in this work, we only infer indirect trust by two hops.

3.4. Users' power

Based on the built trust network on Twitter, we can calculate users' power or reputation. If a user is trusted by a large number of other users, she/he will have a high influence in the whole community. In other words, for a user to have a higher power or reputation, first of all, she/he needs to have a large number of friends or incoming trust links. Besides this, those incoming trust links have to be trustworthy. Remember that we represent impressions in the range of [0, 1], and 0.5 represents neutral sentiment. In other words, most of the positive incoming trust links' impressions should be larger than or equal to 0.5. Also, we want these incoming trust links to be confident as well as trustworthy. To consider the number of incoming trust links, their impressions and confidence together, in this work, we define a simple method to calculate power for users as in Eq. (6).

$$P_u = \sum_{ui \in IN_u \text{ \& } m_{ui} \geq 0.5} m_{ui} * c_{ui} \quad (6)$$

Here, IN_u is the set of users who have trust links toward user u . In other words, IN contains all the users that have trust estimations towards user u . Considering impression and confidence together, we use the product of them. In such a way, even given the same impression m , higher confident incoming trust links contribute more to users' power than lower confident trust links. By setting a threshold of 0.5 for impression, we only count trustworthy incoming trust links. For negative or neutral links, we do not want them to contribute to the power.

4. Twitter sentiment valence

4.1. Sentiment analysis for tweets

As many existing works [9,11,12,22,24,25,29,33,34] suggested, social media's emotional valence can be a helpful and important factor for stock market analysis. Given a tweet, its sentiment can be analyzed in two dimensions: valence or polarity and arousal [11,59]. As did many existing works, in this work we only use tweets' sentiment valence. In other words, for each tweet, we only analyze whether it is positive or negative. In some works in the field of stock market analysis, positive is also called bullish and negative is called bearish [20]. In this work, we use the terms of positive and negative.

In literature, there are two types of sentiment analysis tools: word count analysis strategy based tools and machine learning strategy based tools. For simplicity, many works use word count analysis strategy [9,11]. In this category, based on the given positive and negative dictionaries, each word is mapped into the positive, neutral, or negative tag. To aggregate sentiment valence, there exists two methods: document-level method or tweet-level method. In the document-level works, the numbers of positive, neutral and negative words are counted together for all the interesting tweets in the document. While in the tweet-level works, each tweet is first tagged as positive or negative, based on the number of positive and negative words that it contains. Then, sentiment tagged tweets are accumulated for document-level sentiment valence. In machine learning strategy based tools, words are first used to construct features, i.e., TF-IDF [60], which can be later used along with other features by classifiers [34,52].

In this work, we use an existing sentiment analysis tool – SentiStrength [32] which was also used by Zheludev et al., [52], to do tweet-level sentiment analysis for each tweet. SentiStrength has shown its good performance for informal short text [61]; however, it is not specially designed for financial text analysis. Therefore, in addition to its default lexicon, we also add Loughran and McDonald's financial dictionary [31], which is widely used for financial text sentiment analysis, into SentiStrength.

Note that SentiStrength has different types of output results. In the previous trust modeling stage, we used SentiStrength's multi-scales output results, which can provide more grained information about users' attitudes. While in the field of analyzing Twitter sentiment and stock market, only binary output results (positive or bullish vs. negative or bearish) are used [9,11,20,29,34,52]. Following this, we use SentiStrength's binary output results in the stage of stock market analysis.

4.2. Aggregation of Twitter sentiment valence

By using SentiStrength, we analyze sentiment for each tweet. To investigate the relation between the Twitter sentiment valence and stock returns, we need to accumulate tweets' sentiment valence on a daily basis. To aggregate daily sentiment valence, there are three widely used variables in literature [11,37]. Following Sprenger et al., [29], in this work, we select to use the log of the ratio of the number of positive tweets to the number of negative tweets, which is shown in Eq. (7). Here, P are the number of daily positive tweets, and N is the number of daily negative tweets.

$$TSV = \log \left(\frac{1 + P}{1 + N} \right) \quad (7)$$

Among existing works, all tweets are considered equally important regardless of their authors. Each tweet contributes to either the number of positive tweets or to the number of negative tweets in Eq. (7). However, in reality, the source of information is very important [27]. In the case of stock market analysis, we assume that users who have a higher power in the community should have more influence than users with lower power. Therefore, we adjust Eq. (7) by incorporating users' power, which is calculated in Section 3.4, into calculating Twitter sentiment valence. Instead of considering all tweets equally important, we weight tweets by their authors' power as in Eq. (8). Here, PS is the set of positive tweets, and NS is the set of negative tweets. up and un are the authors who post positive and negative tweets correspondingly.

$$TSV = \log \left(\frac{1 + \sum_{p \in PS} \text{Power}(up \mid up \text{ posts } p)}{1 + \sum_{n \in NS} \text{Power}(un \mid un \text{ posts } n)} \right) \quad (8)$$

Note that, it is possible that a single user appears two or more times in Eq. (8). For example, a user might post two positive tweets about a specific firm on the same day. Or she/he can even

post one positive tweet and one negative tweet about the same firm on the same day. In such cases, this user appears multiple times in Eq. (8).

To compare with existing works, we introduce two baseline methods to calculate Twitter sentiment valence. In the first one, authors' information is ignored such that all the tweets are considered equally important as in Eq. (7). We use TSV_{equal} to denote this method. Actually, TSV_{equal} is widely used by many existing works, including [11,29,33,34,37]. In the second method, instead of using users' power as weights in Eq. (8), we use the number of followers that the users have as weights. This is a straightforward way to differentiate users' influence, as the number of followers that a user has is directly available on Twitter. We denote this method as $TSV_{followers}$. Correspondingly, we denote our trust network power based method as TSV_{power} .

5. Results

5.1. Data

To investigate the relationship between Twitter sentiment valence and stock returns, we collected two sets of data: financial data set and Twitter dataset.

5.1.1. Financial data

To have sufficient information for firms that we were going to investigate, we selected eight firms that had the largest number of tweets in our Twitter dataset. On average, each firm had more than 40 daily tweets. Also, these eight firms were selected from the S&P 500 index. They were Apple Inc (AAPL), Amazon.com Inc (AMZN), Alphabet Inc Class C (GOOG), Facebook Inc (FB), Netflix Inc (NFLX), Gilead Sciences Inc (GILD), General Electric Corp (GE), and Microsoft Corp (MSFT). For the period that we were interested, we downloaded their daily stock market data from Yahoo!Finance, which included open price, highest price, lowest price, close price, adjusted the close price and trading volume. We analyzed Twitter and stock market information from 01/01/2015 through 08/31/2015, which included 167 trading days in total.

Like many other works [9,11,34], we focused on stock returns, which is defined in Eq. (9). Here, we used the adjusted close price in Eq. (9). Therefore, stock returns reflected stock price's change compared with the previous trading day.

$$R_d = \frac{Price_d - Price_{d-1}}{Price_{d-1}} \quad (9)$$

In the field of finance, people are more interested in abnormal returns than stock returns [34]. Abnormal returns are defined as the actual stock returns minus the expected stock returns (also called normal returns) [34,62], as shown in Eq. (10). Here, we use $E[R_d]$ to denote the expected returns or normal returns. From this definition, we can see that abnormal returns somehow reflect external events or news' influence on the stock portfolios. In other words, abnormal returns are more sensitive to external events and news than stock price itself.

$$AR_d = R_d - E[R_d] \quad (10)$$

In literature, there are many alternative methods and models used to calculate the expected stock returns, [63]. To evaluate the difference and performance of these models is beyond the scope of our work. As in [34], we use the market model to estimate the expected returns. It assumes that a firm's stock returns have a linear relation with the whole industry's stock returns. We use linear regression model to represent it in Eq. (11). In our case, we use the S&P 500 index as the independent variable RSP . α is the intercept, and β is the linear coefficient. As in [64] and [34], we use the previous 120 days as the training set to estimate α and

β . α and β are estimated following the ordinary least squares (OLS) procedure. Therefore, although we only investigated from 01/01/2015 through 08/31/2015, we also collected part of 2014s stock data to calculate the expected returns.

$$E[R_d] = \alpha + \beta * RSP_d \quad (11)$$

5.1.2. Twitter data

To collect stock market-related tweets, we found three official certificated accounts on Twitter. They are StockTwits, Financial-Times, and MarketWatch. All of them are stock market-related companies or organizations. We considered them as three groups, and their followers discuss stock market within the groups. We also collected all the followers of these three groups and combined them into a single group or community.

We developed an application using Twitter's open API as well as twitter4j library to collect data from Twitter. We first retrieved users' IDs and then used these user IDs to retrieve their tweets, which are written in English. Note that Twitter's open API limits data are collecting up to 3200 tweets from a single user's timeline. The dataset consists of users' screen names, locations, tweets, and the date and time when they posted the tweets. We took a snapshot of the group in September 2015. At that time, it has 2,898,756 users in total. And from users' timelines, we collected all the tweets posted before September 2015, for a total number of 775,928,121 tweets. In addition to their official accounts' followers, we also included users towards whom the followers had posted interactive tweets. To build the trust network among users, we used all the collected interactive tweets. Based on our definition of trust in Section 3.2, there are 20,916,112 pairs of users having trust relationships. And based on the definition of users' power, 3,929,933 users had their power calculated. So, we only considered tweets that were posted by these 3,929,933 users in the later stock market analysis stage.

After building the trust network, we filtered out tweets that were not related to the stock market in the stage of stock market analysis. Similar to many other works [9,11,24,34], we used the dollar sign (e.g. \$AAPL), to select stock market related tweets, since the dollar sign is commonly used on Twitter to tag stock market related tweets. For the eight firms that we had selected above, we collected their daily tweets from 01/01/2015 through 08/31/2015. All the tweets were grouped on a daily basis for each firm. We listed the number of tweets on trading days for each firm in Table 1.

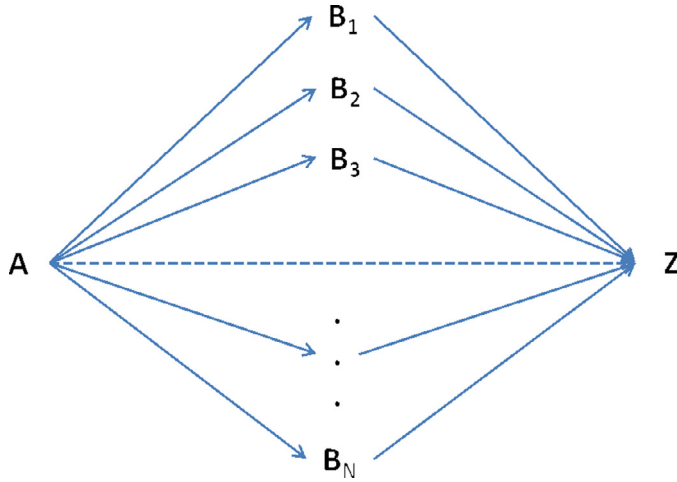
5.2. Trust inference validation experiment

To infer indirect trust relations among users on Twitter, we collected three transitive operators and five aggregation operators from literature in Section 3.3, which results in 15 possible combinations. Each of them might fit better specific applications. As [28,65] indicated, in different applications, users might exhibit different trust propagation behavior patterns. Therefore, among 15 combinations, we needed to select a combination that works for our Twitter application. To measure their trust inference accuracy, we used the leave-one-out cross-validation method [66]. Basically, in the leave-one-out cross-validation method, we compared the difference between the actual trust expressed by the truster and the inferred trust calculated by combinations of transitive operators and aggregation operators. For each leave-one-out case, we first hid the actual direct trust link (dash line in Fig. 2) from the truster towards the trustee and used all the remaining indirect trust links (solid line in Fig. 2) to infer indirect trust by trust transitivity and trust aggregation. An example of leave-one-out case is shown in Fig. 2, where there exists N indirect paths from truster A to trustee Z through B_1, B_2, \dots, B_N . In our dataset, we have 499,327 leave-one-out cases.

Table 1

Number of tweets on trading days from January 1st 2015 through August 31st 2015.

| Firm | Total number of tweets | Average number of daily tweets | Maximum number of daily tweets | Minimum number of daily tweets |
|------|------------------------|--------------------------------|--------------------------------|--------------------------------|
| AAPL | 61,807 | 370.1018 | 2653 | 101 |
| FB | 24,047 | 143.9940 | 1089 | 37 |
| GOOG | 19,461 | 116.5329 | 704 | 29 |
| NFLX | 15,964 | 95.5928 | 665 | 13 |
| AMZN | 13,943 | 83.4910 | 912 | 14 |
| GE | 9091 | 54.4371 | 491 | 10 |
| MSFT | 8087 | 48.4251 | 567 | 9 |
| GILD | 7329 | 43.8862 | 483 | 4 |

**Fig. 2.** An example of leave-one-out case.

For AP2, we used confidence as weights, $w = c$. Accuracy was measured by classical mean absolute error (MAE). We used *diffm* to represent the absolute difference between the inferred m and actual m , and used *diffc* to represent the absolute difference between the inferred c and actual c accordingly. Additionally, to consider *diffm* and *diffc* together, we also measured MAE for Manhattan distances, which is defined in Eq. (14). Note, the range of MAE of Manhattan distances is $[0, 2]$. As we have 499,327 leave-one-out cases, $n = 499,327$ in Eqs. (12)–(14).

$$MAE(diffm) = \frac{\sum_{i=1}^n |actual_m_i - inferred_m_i|}{n} \quad (12)$$

$$MAE(diffc) = \frac{\sum_{i=1}^n |actual_c_i - inferred_c_i|}{n} \quad (13)$$

$$MAE(Man) = \frac{\sum_{i=1}^n |diffm_i| + |diffc_i|}{n} \quad (14)$$

As Golbeck [58] pointed out, inferred indirect trust becomes unreliable when the length of the chains (the number of hops) increases. Therefore, we only took into account the chains containing two hops. We list the performance results of 15 combinations in Table 2.

From Table 2, we can see that TP1 and AP3's performance is significantly worse than other operators, which is consistent with our previous work [28]. Although many applications use multiplication (TP1) as the transitivity operator [58,65,67], in this application, it is not the best one. To consider impression m and confidence c together, we selected to use the combination of TP3AP2 in the remainder of this work, which has the smallest MAE(*man*) among 15 combinations. By selecting TP3, it means that we considered the minimum m in a trust path as the bottleneck. To aggregate trust paths, we used the weighted mean method AP2, where weights are trust paths' confidence. In other words,

Table 2

Comparison of operators' performances.

| Operators | MAE (diffm) | MAE (diffc) | MAE (Man) |
|----------------|---------------|---------------|---------------|
| TP1,AP1 | 0.2449 | 0.0793 | 0.3242 |
| TP1,AP2 | 0.2452 | 0.0795 | 0.3247 |
| TP1,AP3 | 0.2237 | 0.0786 | 0.3023 |
| TP1,AP4 | 0.2133 | 0.0816 | 0.2949 |
| TP1,AP5 | 0.2520 | 0.0811 | 0.3331 |
| TP2,AP1 | 0.0728 | 0.0946 | 0.1674 |
| TP2,AP2 | 0.0728 | 0.0947 | 0.1676 |
| TP2,AP3 | 0.2294 | 0.0961 | 0.3255 |
| TP2,AP4 | 0.0730 | 0.0907 | 0.1636 |
| TP2,AP5 | 0.0733 | 0.0977 | 0.1710 |
| TP3,AP1 | 0.0780 | 0.0851 | 0.1631 |
| TP3,AP2 | 0.0777 | 0.0850 | 0.1627 |
| TP3,AP3 | 0.2252 | 0.0893 | 0.3145 |
| TP3,AP4 | 0.0788 | 0.0919 | 0.1707 |
| TP3,AP5 | 0.0797 | 0.0916 | 0.1713 |

we assumed that higher confident trust paths are more important than lower confident trust paths in aggregating trust paths.

5.3. Users' power distribution

As stated above, in addition to the direct trust links constructed in Section 3.2, we also inferred indirect trust relations for users who originally were not directly connected by using TP3AP2 operators. Given the trust network which includes both direct and indirect inferred trust relations among users, we calculated users' power by following our definition presented in Section 3.4. In Fig. 3, we show the distribution of the number of users for 100 bins of power. We normalized users' power into the range of $[0, 1]$ using feature scaling. So each bin has a length of 0.01. Also, note that we use the log scale for the number of users in each bin. From Fig. 3, we can see that as in many online communities, this distribution follows the power law distribution [68]. Only a few users have high influence in the community. These powerful users can be professional investors or gurus in the field of stock market.

5.4. Pearson correlation

Pearson correlation [69] is widely used to measure the linear relationship between two variables, including time series variables. In this work, we use Pearson correlation coefficients (PCC) to measure the linear relation between the abnormal stock returns (AR) and Twitter sentiment valence (TSV). Remember that from 01/01/2015 through 08/31/2015, we have 167 trading days. Therefore, AR and TSV are two 167×1 vectors. Given these two vectors, Pearson correlation coefficients can be calculated as shown in Eq. (15), where E stands for the expectation value of the variable.

$$PCC = \frac{E[AR * TSV] - E[AR] * E[TSV]}{\sqrt{E[AR^2] - E[AR]^2} * \sqrt{E[TSV^2] - E[TSV]^2}} \quad (15)$$

In Table 3, we list the Pearson correlation coefficients between the selected eight firms' abnormal returns and their Twitter sen-

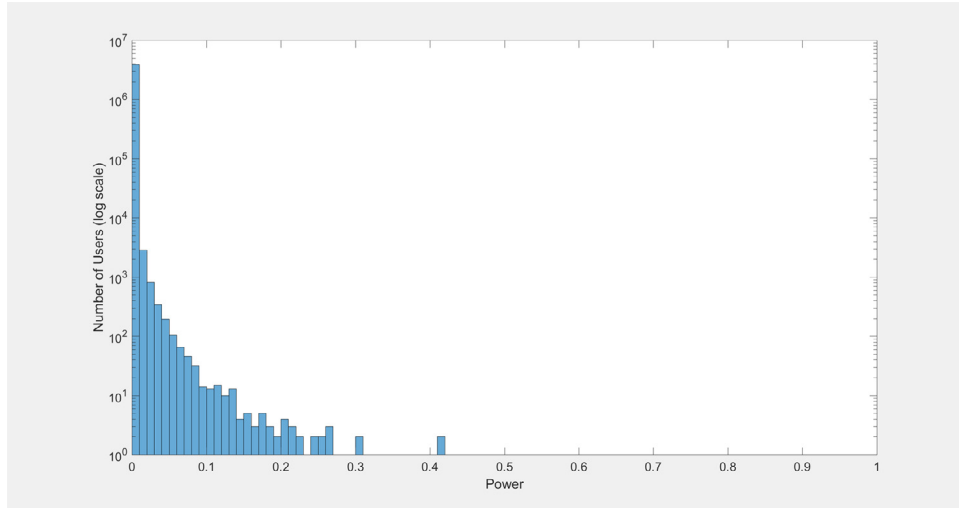


Fig. 3. Distribution of the number of users with regard to users' power.

Table 3
Comparison of Pearson correlation coefficients for eight firms.

| Firms | TSV_{equal} | | $TSV_{followers}$ | | TSV_{power} | |
|-------|---------------|----------------------|-------------------|----------------------|---------------|---|
| | PCC | <i>p</i> -value | PCC | <i>p</i> -value | PCC | <i>p</i> -value |
| AAPL | 0.3370 | 8.4×10^{-6} | 0.3969 | 1.1×10^{-7} | 0.4644 | 2.6×10^{-10} |
| FB | 0.0662 | 0.395 | 0.0544 | 0.485 | 0.0962 | 0.216 |
| GOOG | 0.1830 | 0.018 | 0.1295 | 0.095 | 0.2883 | 1.6×10^{-4} |
| NFLX | 0.1416 | 0.068 | 0.1758 | 0.023 | 0.4036 | 6.4×10^{-8} |
| AMZN | 0.1314 | 0.091 | 0.3949 | 1.3×10^{-7} | 0.5318 | 1.4×10^{-13} |
| GE | 0.0401 | 0.610 | 0.0043 | 0.956 | 0.1530 | 0.048 |
| MSFT | 0.0533 | 0.494 | 0.1035 | 0.183 | 0.3812 | 3.7×10^{-7} |
| GILD | 0.0969 | 0.213 | 0.0305 | 0.696 | 0.1702 | 0.028 |

timent valence. In addition to Pearson correlation coefficients, we also test the *p*-values for them. We compare our trust network power based method TSV_{power} with other two baseline methods TSV_{equal} (a widely used method by many existing works, such as [11,29,33,34,37].) and $TSV_{followers}$ that we mentioned in Section 4.

In Table 3, among TSV_{equal} , $TSV_{followers}$ and TSV_{power} we use bold font to represent the most linearly correlated method with the stock abnormal returns. We can see that our method TSV_{power} performs better than other two methods for all eight firms. By weighting tweets' sentiment by their authors' power, TSV_{power} has higher PCC (and correspondingly lower *p*-value) than other two methods. For many firms, such as AMZN, GE, MSFT and GILD, by using $TSV_{followers}$ or TSV_{equal} , the Pearson coefficient between their Twitter sentiment valence and abnormal returns is weak (*p*-values are greater than 0.05), which means that Twitter sentiment valence might not have linear relation with abnormal stock returns. However, by using our trust network power based method, Twitter sentiment valence is significantly linearly related to abnormal stock returns for all the firms except FB. This confirms that the source of information (tweets) is an important factor to consider in this field of study. Compared with $TSV_{followers}$, TSV_{power} not only takes the number of trust links into account, but it also considers the quality of trust links. Our trust network power based method highlights powerful users' tweets and opinions, such that the accumulated Twitter sentiment valence is more linearly related to the firms' abnormal returns.

To illustrate this, we compared three methods' of performance for NFLX's abnormal returns and its Twitter sentiment valence in Fig. 4. We can see that our trust network power based method reflects NFLX's abnormal returns better than other two methods. For example, for the abnormal returns' peak at day 13, our method

TSV_{power} follows the peak, while other two methods are not able. Note that, in Fig. 4, to compare abnormal returns and Twitter sentiment valence in the same scale, we converted both of them to Standard scores (also called *z*-scores) as shown in Eq. (16), whose means are 0 and standard deviations are 1. In Eq. (16), *z* is a Standard score, *x* is the original score, μ and σ are the mean and standard deviation of the population, respectively.

$$z = \frac{x - \mu}{\sigma} \quad (16)$$

To see how good Twitter sentiment valence is linearly related to the firms' abnormal returns, we selected AMZN as an example, which has the largest Pearson correlation coefficient among eight firms. We illustrate the relation between AMZN's abnormal returns and our trust network power based method in Fig. 5. We can see that our method captures abnormal returns' fluctuation very well, especially for three abnormal returns' big peaks in Fig. 5. Such kind of linear correlation can be used for other advanced analysis, for example, event study [34,70] and stock price prediction.

5.5. Linear regression correlation

In the above, Pearson correlation coefficient was used to measure the pairwise linear correlation between abnormal stock returns and Twitter sentiment valence. In addition to that, by taking into account that abnormal stock returns might exhibit auto-correlation property [27], we also constructed a linear regression model which includes both Twitter sentiment valence and historical abnormal returns, as in Eq. (17).

$$AR_d = \alpha + \beta * TSV_d + \gamma * CV + \varepsilon_d \quad (17)$$

Here, α is the intercept. β is the coefficient that we are going to investigate, and ε denotes a zero mean disturbance term. CV stands for control variables. Although there are many factors (i.e., trading volume, volatility) that can be considered as control variables [8,27], in this work, we considered the previous three days' abnormal returns as control variables. Thus, we can rewrite the regression Eq. (17) as in Eq. (18).

$$AR_d = \alpha + \beta * TSV_d + \sum_{i=1}^{i=3} \gamma_i * AR_{d-i} + \varepsilon_d \quad (18)$$

We tested Eq. (18) with TSV calculated by three methods we mentioned above. We list estimated coefficient β , standard error of the estimation SE, *t* statistic for a test that the coefficient is

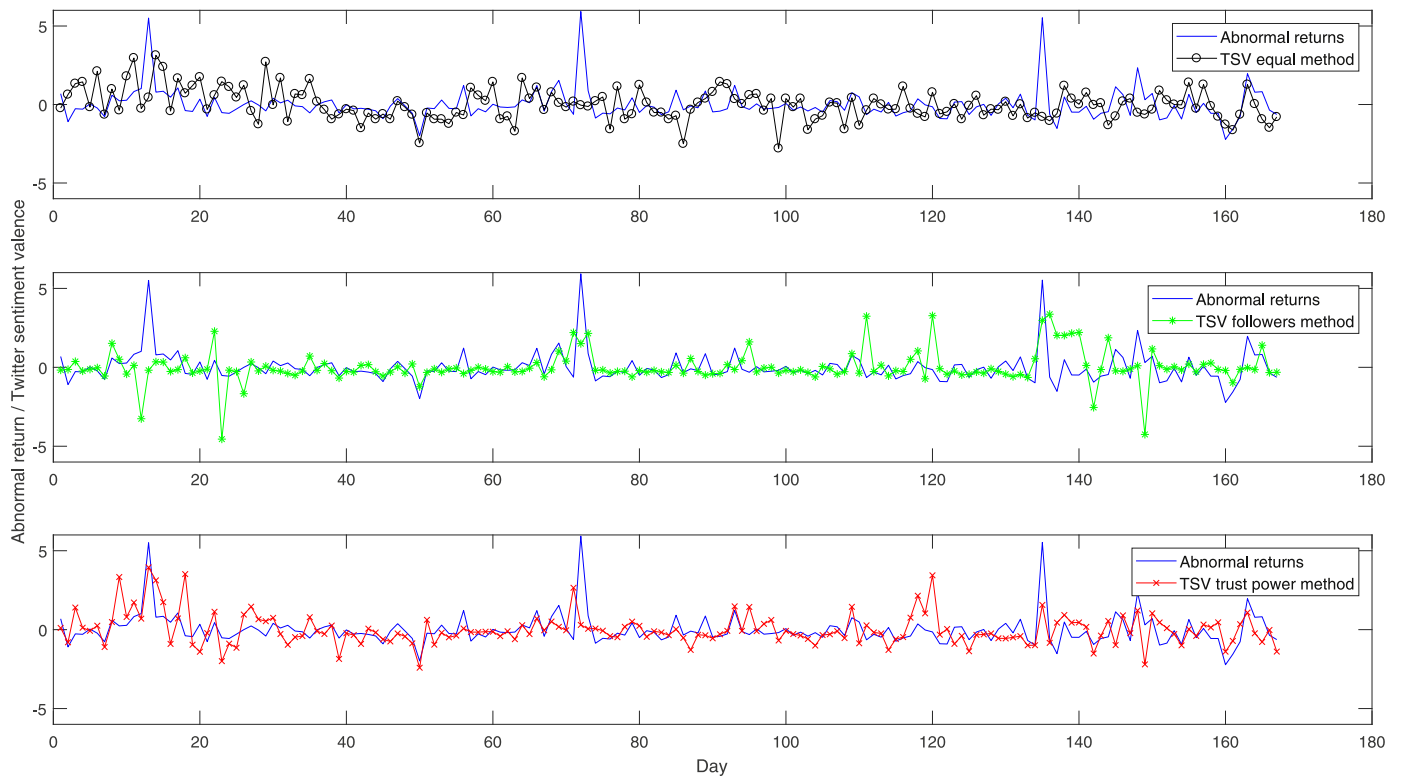


Fig. 4. Comparison of three methods for NFLX.

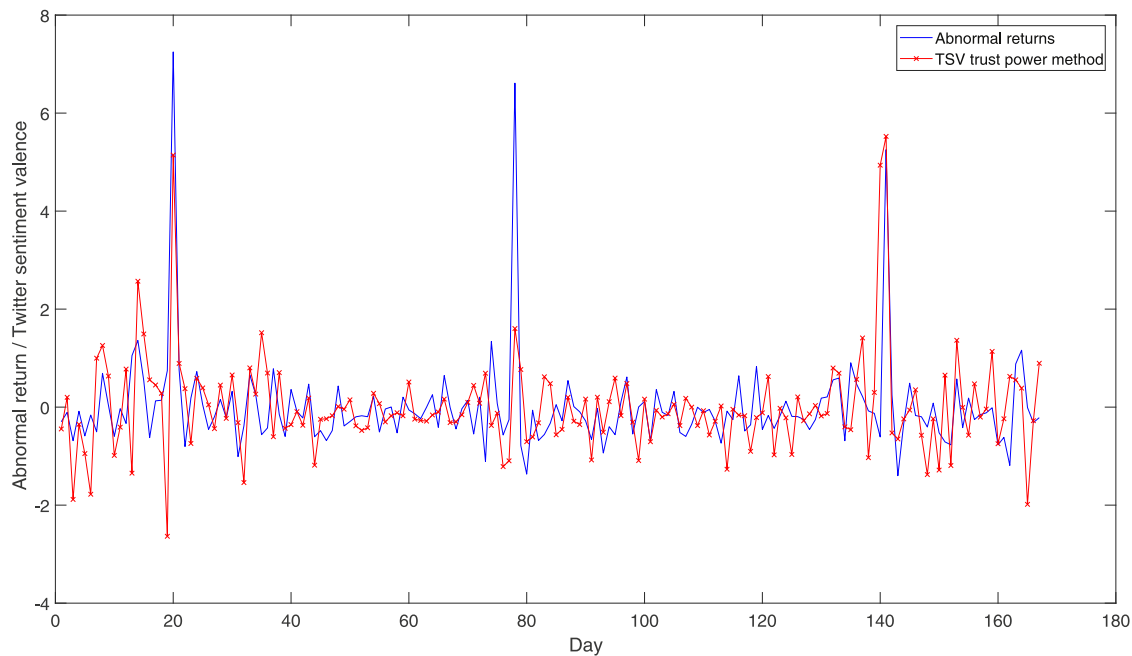


Fig. 5. Pearson correlation between AMZN's abnormal returns and trust network power based Twitter sentiment valence.

zero $tStat$, p -value for the t statistic $pValue$, and adjusted R-square $adjR^2$ in Table 4.

From Table 4, we observe the same performance pattern as in Pearson correlation coefficient test. In other words, after considering the stock abnormal returns' possible auto-correlation property, still, our trust network power based method outperforms other two methods. Similarly, in Table 4, we highlight the lowest p -value and the highest adjusted R-square among three methods with bold font. Therefore, the main hypothesis of this work that the users

reputation built by using our trust management system, helps in making better predictions of the stock market is confirmed.

5.6. A limitation – number of tweets

Although in the above experiments our trust network power based method outperformed other two baseline methods for all eight firms we selected, we found that to achieve this each firm must have enough number of daily tweets available. Remember

Table 4
Regression results of abnormal returns for eight firms.

| Firms | TSV methods | coefficient β | SE | tStat | p-Value | adjR ² |
|-------|--------------------------|---------------------|--------|---------------|--|-------------------|
| AAPL | TSV _{equal} | 0.0127 | 0.0026 | 4.9243 | 2.07×10^{-6} | 0.1241 |
| | TSV _{followers} | 0.1103 | 0.0187 | 5.9106 | 1.95×10^{-8} | 0.1717 |
| | TSV _{power} | 0.0365 | 0.0051 | 7.1752 | 2.45×10^{-11} | 0.2359 |
| FB | TSV _{equal} | 0.0016 | 0.0020 | 0.7919 | 0.4296 | 0.0372 |
| | TSV _{followers} | 0.0166 | 0.0246 | 0.6751 | 0.5006 | 0.0361 |
| | TSV _{power} | 0.0113 | 0.0086 | 1.3162 | 0.1900 | 0.0437 |
| GOOG | TSV _{equal} | 0.0066 | 0.0029 | 2.3166 | 0.0218 | 0.0115 |
| | TSV _{followers} | 0.0700 | 0.0407 | 1.7200 | 0.0873 | −0.0030 |
| | TSV _{power} | 0.0398 | 0.0105 | 3.8031 | 2.02×10^{-4} | 0.0624 |
| NFLX | TSV _{equal} | 0.0077 | 0.0045 | 1.6902 | 0.0929 | 0.0094 |
| | TSV _{followers} | 0.2252 | 0.0991 | 2.2725 | 0.0244 | 0.0231 |
| | TSV _{power} | 0.1388 | 0.0244 | 5.6836 | 5.99×10^{-8} | 0.1595 |
| AMZN | TSV _{equal} | 0.0048 | 0.0028 | 1.6936 | 0.0923 | 0.0204 |
| | TSV _{followers} | 0.3276 | 0.0565 | 5.7987 | 3.41×10^{-8} | 0.1744 |
| | TSV _{power} | 0.1228 | 0.0154 | 7.9768 | 2.57×10^{-13} | 0.2842 |
| GE | TSV _{equal} | 0.0007 | 0.0013 | 0.5824 | 0.5611 | −0.0133 |
| | TSV _{followers} | 0.0014 | 0.0347 | 0.0400 | 0.9681 | −0.0154 |
| | TSV _{power} | 0.0298 | 0.0150 | 1.9911 | 0.0481 | 0.0089 |
| MSFT | TSV _{equal} | 0.0010 | 0.0016 | 0.6095 | 0.5430 | −0.0184 |
| | TSV _{followers} | 0.0588 | 0.0423 | 1.3908 | 0.1662 | −0.0087 |
| | TSV _{power} | 0.0822 | 0.0155 | 5.2984 | 3.77×10^{-7} | 0.1300 |
| GILD | TSV _{equal} | 0.0023 | 0.0017 | 1.3519 | 0.1783 | −0.0050 |
| | TSV _{followers} | 0.0603 | 0.1161 | 0.5189 | 0.6045 | −0.0146 |
| | TSV _{power} | 0.0453 | 0.0189 | 2.3970 | 0.0177 | 0.0185 |

Table 5
Pearson correlation coefficients of BAC.

| Testing period | TSV _{equal} | | TSV _{followers} | | TSV _{power} | |
|----------------------|----------------------|--------------|--------------------------|---------|----------------------|--------------|
| | PCC | p-value | PCC | p-value | PCC | p-value |
| All 167 trading days | 0.1877 | 0.015 | 0.0244 | 0.755 | 0.1589 | 0.040 |
| Subset40 | 0.1471 | 0.464 | −0.1773 | 0.376 | 0.4378 | 0.022 |

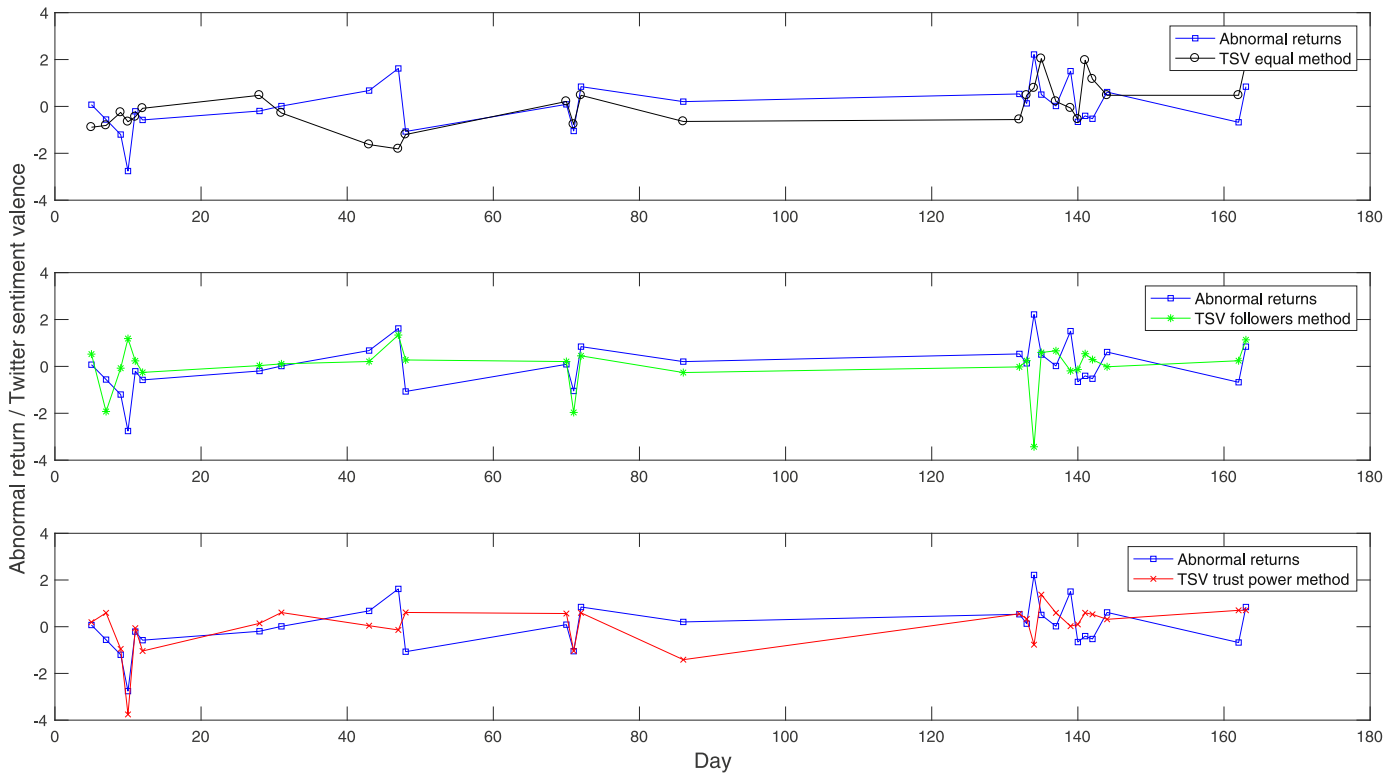


Fig. 6. Comparison of BAC's performance in Subset40.

that all eight firms we selected have more than 40 average daily tweets.

To see the influence of the number of daily tweets, we selected another firm – Bank of America Corp (BAC), as an example. BAC is the 9th most mentioned firm in our Twitter dataset. And it has an average of 31.4192 daily tweets during our testing period. As before, we did the Pearson correlation test for BAC with three methods for all the 167 trading days. We list the results in Table 5. From Table 5 we can see that, in this case, TSV_{equal} performs better than our method TSV_{power} .

Since BAC has only a few tweets on many trading days, instead of testing for all the 167 trading days, we selected a subset of trading days on which BAC has more than 40 tweets available (Subset40). By setting a threshold of 40 for the number of daily tweets, Subset40 has 27 trading days. Also, we tested the Pearson correlation for Subset40 and included its results in Table 5. We can see that if we have enough number of tweets (in this example more than or equal to 40 daily tweets) to infer Twitter sentiment valence for BAC, still our method can outperform other two methods. Besides performance, we think that to get reliable analysis results, it is necessary to have sufficient tweets. Note that, compared with the difference between TSV_{equal} and TSV_{power} for all 167 trading days, our method performs much better in Subset40. Also, we compared three methods' performance of BAC in Subset40 in Fig. 6. For example, our method TSV_{power} can predict for day10/s drop, while other two methods are not able to do such prediction.

6. Conclusions and future work

In this work, we used the abnormal stock returns as ground truth for our trust management system. For this reason we verified the hypothesis that the users reputation, built by the inter trust among them, using our trust management system, helps in making better predictions of abnormal stock returns. So, we collected a group of users who were interested in stock market activities from Twitter. Based on tweets posted by the users, we selected eight firms which were the top eight mentioned firms in the data set. Correspondingly, those eight firms' stock market data was collected from Yahoo! Finance. For the users on Twitter, we adapted our trust management framework [28] and constructed a user-to-user trust network. Based on this user-to-user trust network, we calculated for users' power or reputation in a simple way.

To see whether or not Twitter sentiment information could help to analyze stock market, for each firm, we analyzed Pearson correlation coefficients between Twitter sentiment valence and the firm's abnormal returns. Compared with existing works, when accumulating Twitter sentiments, we took into account tweets' authors. Authors were weighted and differentiated by their reputation or power in the whole community. Compared with treating all the authors equally or simply weighting authors by the number of their followers, we could see that our trust network based reputation mechanism could amplify the correlation between a specific firm's Twitter sentiment valence and the firm's stock abnormal returns.

To further consider the auto-correlation property of abnormal stock returns, we also constructed a linear regression model, in which the previous three days' abnormal returns were considered as control variables. Again, our results showed that by using our trust network power based method to weight tweets, we did linear regression better than other two methods.

However, our work also has some limitations. First of all, we did experiments only for a period (from 01/01/2015 through 08/31/2015). It is possible that the relation pattern we found here does not apply to other periods of time [27]. Therefore, testing our method on multiple data sets and periods of time is part of our future work. Furthermore, our study showed that when

the number of tweets about a firm was very small; the Twitter sentiment valence might not be able to reflect the stock market. So, in the future, we will consider collecting more data or think of how to use available data more effectively. Finally, we will further tune the used reputation algorithms.

7. Competing interests

The authors declare that they have no competing interests.

8. Author's contributions

YR initiated and proposed the framework. Also, YR drafted the paper. AD discussed about the framework, proofread the draft and provided comments and suggestions to improve the draft. LA collected the Twitter data set. YR was also in charge of submitting the paper and corresponding with the editors of the journal. All authors read and approved the final manuscript.

Acknowledgments

This work was partially supported by National Science Foundation under Grant No. 1547411 and National Institute of Food and Agriculture (NIFA) USDA Award 2017-67003-26057. Also, it was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute, and in part by the Indiana METACyt Initiative. The Indiana METACyt Initiative at IU is also supported in part by Lilly Endowment, Inc.

References

- [1] T. Sakaki, M. Okazaki, Y. Matsuo, Earthquake shakes Twitter users: real-time event detection by social sensors, in: Proceedings of the Nineteenth International Conference on World Wide Web, ACM, 2010, pp. 851–860.
- [2] A. Tumasjan, T.O. Sprenger, P.G. Sandner, I.M. Welpe, Predicting elections with Twitter: what 140 characters reveal about political sentiment 10 (1) (2010) 178–185.
- [3] H. Wang, D. Can, A. Kazemzadeh, F. Bar, S. Narayanan, A system for real-time Twitter sentiment analysis of 2012 us presidential election cycle, in: Proceedings of the ACL System Demonstrations, Association for Computational Linguistics, 2012, pp. 115–120.
- [4] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, M. Demirbas, Short text classification in Twitter to improve information filtering, in: Proceedings of the Thirty-Third International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2010, pp. 841–842.
- [5] E. Kouloumpis, T. Wilson, J.D. Moore, Twitter sentiment analysis: the good the bad and the omg! 11 (538–541) (2011) 164.
- [6] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, A. Perera, Opinion mining and sentiment analysis on a Twitter data stream, in: Proceedings of the International Conference on Advances in ICT for Emerging Regions (ICTer), IEEE, 2012, pp. 182–188.
- [7] F.H. Khan, S. Bashir, U. Qamar, TOM: Twitter opinion mining framework using hybrid classification scheme, Decis. Support Syst. 57 (2014) 245–257.
- [8] P.C. Tetlock, Giving content to investor sentiment: the role of media in the stock market, J. Finance 62 (3) (2007) 1139–1168.
- [9] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, J. Comput. Sci. 2 (1) (2011) 1–8.
- [10] L. Fang, J. Peress, Media coverage and the cross-section of stock returns, J. Finance 64 (5) (2009) 2023–2052.
- [11] H. Sul, A.R. Dennis, L.I. Yuan, Trading on Twitter: the financial information content of emotion in social media, in: Proceedings of the Forty-Seventh Hawaii International Conference on System Sciences, 2014, pp. 806–815, doi:10.1109/HICSS.2014.107.
- [12] Y. Ruan, L. Alfantoukh, A. Dursesi, Exploring stock market using Twitter trust network, in: Proceedings of the IEEE twenty-Ninth International Conference on Advanced Information Networking and Applications, 2015, pp. 428–433, doi:10.1109/AINA.2015.217.
- [13] C.-H. Chen, C.-H. Yu, A series-based group stock portfolio optimization approach using the grouping genetic algorithm with symbolic aggregate approximations, Knowl. Based Syst. 125 (2017) 146–163.
- [14] Y. Pan, Z. Xiao, X. Wang, D. Yang, A multiple support vector machine approach to stock index forecasting with mixed frequency sampling, Knowl. Based Syst. 122 (2017) 90–102.
- [15] A. Cowles 3rd, Can stock market forecasters forecast? Econom. J. Econom. Soc. 1 (1933) 309–324.
- [16] E.F. Fama, Random walks in stock market prices, Finan. Anal. J. 51 (1) (1995) 75–80.

- [17] E.F. Fama, The behavior of stock-market prices, *J. Bus.* 38 (1) (1965) 34–105.
- [18] P. Klibanoff, O. Lamont, T.A. Wizman, Investor reaction to salient news in closed-end country funds, *J. Finance* 53 (2) (1998) 673–699.
- [19] X. Luo, J. Zhang, W. Duan, Social media and firm equity value, *Inf. Syst. Res.* 24 (1) (2013) 146–163.
- [20] H. Chen, P. De, Y. Hu, B.-H. Hwang, Wisdom of crowds: the value of stock opinions transmitted through social media, *Rev. Financ. Stud.* 27 (5) (2014) 1367–1403.
- [21] Q. Li, Y. Chen, J. Wang, Y. Chen, H. Chen, Web media and stock markets : a survey and future directions from a big data perspective, *IEEE Trans. Knowl. Data Eng.* PP (99) (2017) 1, doi:10.1109/TKDE.2017.2763144.
- [22] J.R. Nofsinger, Social mood and financial economics, *J. Behav. Finance* 6 (3) (2005) 144–160.
- [23] C. Oh, O. Sheng, Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement, in: *Proceedings of the International Conference on Information Systems*, 2011.
- [24] S. Zhao, Y. Tong, X. Liu, S. Tan, Correlating Twitter with the stock market through non-Gaussian SVAR, in: *Proceedings of the Eighth International Conference on Advanced Computational Intelligence (ICACI)*, 2016, pp. 257–264, doi:10.1109/ICACI.2016.7449835.
- [25] N. Oliveira, P. Cortez, N. Areal, The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices, *Expert Syst. Appl.* 73 (2017) 125–144.
- [26] L.F. Lee, A.P. Hutton, S. Shu, The role of social media in the capital market: evidence from consumer product recalls, *J. Account. Res.* 53 (2) (2015) 367–404.
- [27] H.K. Sul, A.R. Dennis, L.I. Yuan, Trading on Twitter: using social media sentiment to predict stock returns, *Decis. Sci.* 48 (3) (2017) 454–488.
- [28] Y. Ruan, P. Zhang, L. Alfantoukh, A. Durresi, Measurement theory-based trust management framework for online social communities, *ACM Trans. Internet Technol.* 17 (2) (2017) 16:1–16:24, doi:10.1145/3015771.
- [29] T.O. Sprenger, A. Tumasjan, P.G. Sandner, I.M. Welp, Tweets and trades: the information content of stock microblogs, *Eur. Finan. Manag.* 20 (5) (2014) 926–957.
- [30] D.W. Jorgenson, K. Vu, Information technology and the world economy, *Scand. J. Econ.* 107 (4) (2005) 631–650.
- [31] T. Loughran, B. McDonald, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *J. Finance* 66 (1) (2011) 35–65.
- [32] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, Sentiment strength detection in short informal text, *J. Assoc. Inf. Sci. Technol.* 61 (12) (2010) 2544–2558.
- [33] J. Smilović, M. Grčar, N. Lavrač, M. Žnidaršič, Predictive sentiment analysis of Tweets: a stock market application, in: *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*, Springer, 2013, pp. 77–88.
- [34] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar, I. Mozetič, The effects of Twitter sentiment on stock price returns, *PLoS ONE* 10 (9) (2015) e0138441.
- [35] E. Gilbert, K. Karahalios, Widespread worry and the stock market, in: *Proceedings of the International Conference on Web and Social Media*, 2010, pp. 59–65.
- [36] X. Zhang, H. Fuehres, P.A. Gloor, Predicting stock market indicators through Twitter ãi hope it is not as bad as i fear, *Proc. Soc. Behav. Sci.* 26 (2011) 55–62.
- [37] P.C. Tetlock, M. Saar-Tsechansky, S. Macskassy, More than words: quantifying language to measure firms' fundamentals, *J. Finance* 63 (3) (2008) 1437–1467.
- [38] N. Evangelopoulos, M.J. Magro, A. Sidorova, The dual micro/macro informing role of social network sites: can Twitter macro messages help predict stock prices? *Inf. Sci.* 15 (2012) 247–268.
- [39] C.W. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econom. J. Econom. Soc.* 37 (3) (1969) 424–438.
- [40] A.G. Sutcliffe, D. Wang, R.I. Dunbar, Modelling the role of trust in social relationships, *ACM Trans. Internet Technol. (TOIT)* 15 (4) (2015) 16.
- [41] Q. Shambour, J. Lu, A hybrid trust-enhanced collaborative filtering recommendation approach for personalized government-to-business e-services, *Int. J. Intell. Syst.* 26 (9) (2011) 814–843.
- [42] Q. Shambour, J. Lu, A trust-semantic fusion-based recommendation approach for e-business applications, *Decis. Support Syst.* 54 (1) (2012) 768–780.
- [43] B.A. Huberman, D.M. Romero, F. Wu, Social networks that matter: Twitter under the microscope (2008), arXiv:0812.1045v1.
- [44] Y.A. Kim, H.S. Song, Strategies for predicting local trust based on trust propagation in social networks, *Knowl. Based Syst.* 24 (8) (2011) 1360–1371.
- [45] A. Jøsang, T. Azderska, S. Marsh, Trust transitivity and conditional belief reasoning 374 (2012) 68–83.
- [46] Y. Ruan, A. Durresi, A survey of trust management systems for online social communities—trust modeling, trust inference and attacks, *Knowl. Based Syst.* 106 (2016) 150–163.
- [47] G. Guo, J. Zhang, D. Thalmann, Merging trust in collaborative filtering to alleviate data sparsity and cold start, *Knowl. Based Syst.* 57 (2014) 57–68.
- [48] C. Castelfranchi, R. Falcone, *Trust Theory: A Socio-Cognitive and Computational Model*, 18, John Wiley & Sons, 2010.
- [49] M.G. Morgan, M. Henrion, M. Small, *Uncertainty: A Guide to Dealing With Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, 1992.
- [50] C. Honey, S.C. Herring, Beyond microblogging: conversation and collaboration via Twitter, in: *Proceedings of the Forty-Second Hawaii International Conference on System Sciences*, IEEE, 2009, pp. 1–10.
- [51] A. Kale, Modeling trust and influence on blogosphere using link polarity, University of Maryland, Baltimore, 2007 (Ph.D. thesis).
- [52] I. Zheludev, R. Smith, T. Aste, When can social media lead financial markets? *Sci. Rep.* 4 (2014) 4213.
- [53] Y. Sun, W. Yu, Z. Han, K.J.R. Liu, Trust modeling and evaluation in ad hoc networks, in: *Proceedings of the IEEE Global Telecommunications Conference*, 3, 2005, p. 6, doi:10.1109/GLOCOM.2005.1577971.
- [54] S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications*, 8, Cambridge University Press, 1994.
- [55] P. Holland, S. Leinhardt, Some evidence on the transitivity of positive interpersonal sentiment, *Am. J. Sociol.* 77 (1972) 1205–1209.
- [56] A.A. Clifford, *Multivariate Error Analysis: A Handbook of Error Propagation and Calculation in Many-Parameter Systems*, Wiley, 1973.
- [57] H.J. Berendsen, *A Student's Guide to Data and Error Analysis*, Cambridge University Press, 2011.
- [58] J. Golbeck, Computing and applying trust in web-based social networks, University of Maryland at College Park, College Park, MD, USA, 2005 (Ph.D. thesis).
- [59] M.M. Bradley, P.J. Lang, Measuring emotion: the self-assessment manikin and the semantic differential, *J. Behav. Ther. Exp. Psychiatry* 25 (1) (1994) 49–59.
- [60] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2016.
- [61] M. Thelwall, K. Buckley, G. Paltoglou, Sentiment strength detection for the social web, *J. Assoc. Inf. Sci. Technol.* 63 (1) (2012) 163–173.
- [62] S.J. Brown, J.B. Warner, Using daily stock returns: the case of event studies, *J. Financ. Econ.* 14 (1) (1985) 3–31.
- [63] A. Ilmanen, *Expected Returns: An Investor's Guide to Harvesting Market Rewards*, John Wiley & Sons, 2011.
- [64] J.Y. Campbell, A.W.-C. Lo, A.C. MacKinlay, *The Econometrics of Financial Markets*, Princeton University Press, 1997.
- [65] P. Victor, C. Cornelis, M. De Cock, *Trust Networks for Recommender Systems*, 4, Springer Science & Business Media, 2011.
- [66] P. Refaellizadeh, L. Tang, H. Liu, Cross-validation, in: *Encyclopedia of Database Systems*, Springer, 2009, pp. 532–538.
- [67] G. Theodorakopoulos, J.S. Baras, On trust models and trust evaluation metrics for ad hoc networks, *IEEE J. Sel. Areas Commun.* 24 (2) (2006) 318–328.
- [68] L.A. Adamic, B.A. Huberman, Power-law distribution of the world wide web, *Science* 287 (5461) (2000) 2115–2115.
- [69] F. Galton, Regression towards mediocrity in hereditary stature, *J. Anthropol. Inst. G. B. Irel.* 15 (1886) 246–263.
- [70] E. Boehmer, J. Masumeci, A.B. Poulsen, Event-study methodology under conditions of event-induced variance, *J. Financ. Econ.* 30 (2) (1991) 253–272.