

# Value and Misinformation in Collaborative Investing Platforms

TIANYI WANG, Tsinghua University and University of California, Santa Barbara

GANG WANG, University of California, Santa Barbara and Virginia Tech

BOLUN WANG, DIVYA SAMBASIVAN, and ZENGBIN ZHANG, University of California, Santa Barbara

XING LI, Tsinghua University

HAITAO ZHENG and BEN Y. ZHAO, University of California, Santa Barbara

It is often difficult to separate the highly capable “experts” from the average worker in crowdsourced systems. This is especially true for challenge application domains that require extensive domain knowledge. The problem of stock analysis is one such domain, where even the highly paid, well-educated domain experts are prone to make mistakes. As an extremely challenging problem space, the “wisdom of the crowds” property that many crowdsourced applications rely on may not hold.

In this article, we study the problem of evaluating and identifying experts in the context of SeekingAlpha and StockTwits, two crowdsourced investment services that have recently begun to encroach on a space dominated for decades by large investment banks. We seek to understand the quality and impact of content on collaborative investment platforms, by empirically analyzing complete datasets of SeekingAlpha articles (9 years) and StockTwits messages (4 years). We develop sentiment analysis tools and correlate contributed content to the historical performance of relevant stocks. While SeekingAlpha articles and StockTwits messages provide minimal correlation to stock performance in aggregate, a subset of experts contribute more valuable (predictive) content. We show that these authors can be easily identified by user interactions, and investments based on their analysis significantly outperform broader markets. This effectively shows that even in challenging application domains, there is a *secondary or indirect* wisdom of the crowds.

Finally, we conduct a user survey that sheds light on users’ views of SeekingAlpha content and stock manipulation. We also devote efforts to identify potential manipulation of stocks by detecting authors controlling multiple identities.

Categories and Subject Descriptors: H.3.5 [Information Storage and Retrieval]: Online Information Services; J.4 [Computer Applications]: Social and Behavioral Sciences

General Terms: Measurement, Management, Design

Additional Key Words and Phrases: Crowdsourcing, stock market, sentiment analysis

The first version of this article appeared in the Proceedings of the 18th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW 2015) [Wang et al. 2015]. This version extends the original by including Section 10 (1.5 pages), Section 11 (4.5 pages), and three more questions and discussions in Section 8.3 (1.5 pages). Besides, it includes a real-time demonstration to run the algorithm in this article. It also adds some more detailed explanations and discussions in the rest of the article.

This work is supported in part by the National Science Foundation under grants IIS-1321083 and CNS-1224100, by the DARPA GRAPHS program (BAA-12-01), and by the Department of State. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any funding agencies.

Authors’ addresses: T. Wang and X. Li, Electronic Engineering, Tsinghua University; emails: tsinghuawty@gmail.com, xing@cernet.edu.cn; T. Wang, G. Wang, B. Wang, D. Sambasivan, Z. Zhang, H. Zheng, and B. Y. Zhao, Computer Science Department, UC Santa Barbara, Santa Barbara, CA 93106; emails: {tianyi, gangw, bolunwang, divya\_sambasivan, zengbin, htzheng, ravenben}@cs.ucsb.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2017 ACM 1559-1131/2017/05-ART8 \$15.00

DOI: <http://dx.doi.org/10.1145/3027487>

**ACM Reference Format:**

Tianyi Wang, Gang Wang, Bolun Wang, Divya Sambasivan, Zengbin Zhang, Xing Li, Haitao Zheng, and Ben Y. Zhao. 2017. Value and misinformation in collaborative investing platforms. *ACM Trans. Web* 11, 2, Article 8 (May 2017), 32 pages.  
DOI: <http://dx.doi.org/10.1145/3027487>

**1. INTRODUCTION**

In the age of the Internet, the “wisdom of the crowd” phenomenon has revolutionized content generation, discovery, and curation. User-contributed content now dominates restaurant views (Yelp), travel and hospitality (TripAdvisor), encyclopedias (Wikipedia), general Q&A (Quora, Yahoo Answers), and even photography (Flickr). Even as content proliferates on these platforms, content curation (e.g., identifying high-quality content and the “experts” who generate such content) is a growing challenge [Adamic et al. 2008; Harper et al. 2008; Wang et al. 2013a].

This problem is especially critical in the rapidly growing area of collaborative investment analysis (i.e., user-contributed analysis on value and investment strategies of stocks, bonds, and commodities). For over 150 years, personal investment advice has been the exclusive domain of investment banks and wealth advisors such as Goldman Sachs and Lehman Brothers. Over the last decade, however, networks like CNBC and Bloomberg established roles as independent sources of financial news, and the financial crisis of 2008 led to the collapse of several of the oldest investment banks (Bear Stearns, Lehman Brothers). Filling in the void were rapidly growing services such as SeekingAlpha and StockTwits, where independent analysts and retail investors could contribute and share analysis for free. SeekingAlpha now reports more than 3 million users and 9 million unique monthly visits. This represents a significant portion of the US investment market, where more than 50 million estimated households own mutual funds or equities [Investment 2013].

In this article, we seek to understand the quality and impact of analysis shared on collaborative investment platforms, and how to distinguish high-quality analysis from biased or uninformed opinions. We target the two primary yet quite different social investment platforms, SeekingAlpha and StockTwits, and analyze the potential for investment returns following their recommendations versus the market baseline, the S&P 500 stock market index. We seek to understand how the expertise of contributors can affect the quality and utility of contributed content, using SeekingAlpha as an “expert” model (all content is contributed by fewer than 0.27% of users) and StockTwits as a “peer” model (any user can contribute).

Our work has three goals. First, we assess the quality of content in these platforms to determine how much value, if any, these collaborative platforms provide to individual investors. Second, we contrast two platforms using an “expert”-based model and a “peer”-based model, to explore the design principles and lessons for building collaborative social investing systems. Finally, we aim to evaluate techniques to identify the most effective “expert contributors” in the crowd while mitigating the impact from biased, manipulative contributors.

We summarize our contributions as follows.

*First*, we gather longitudinal datasets from both platforms since their inception (9 years for SeekingAlpha, 4 years for StockTwits). We develop sentiment analyzers on each dataset, using a mixture of keyword-processing and machine-learning classifiers. Validation shows our methods achieve high accuracy in extracting sentiments toward individual stocks (85.5% for SeekingAlpha, 76.2% for StockTwits).

*Second*, we analyze correlation between content sentiments from both services with stock returns at different time scales. We show that content from both SeekingAlpha and StockTwits provides minimal forward correlation with stock performance.

While the average article provides little value, we find that a subset of “top authors” in SeekingAlpha contribute content that shows significantly higher correlation with future stock performance.

*Third*, we evaluate the hypothetical performance of simple investment strategies following top authors from both platforms. We show that strategies based on sentiment from top SeekingAlpha authors perform exceptionally well, and significantly outperform broader markets. More importantly, we show that these experts can be easily identified without historical stock market data, using only user interactions with their articles as a guide. This suggests the design of a “meta-reputation” system for users that can serve to both curate content and simplify its discovery.

*Fourth*, we conduct a survey of SeekingAlpha users and contributors to understand their usage, reliance, and trust in the SeekingAlpha service. Results show that despite seeing potentially intentionally misleading or manipulative articles, most users still rely heavily on the site content for investment advice. Most consider SeekingAlpha unique and would not use a competing alternative in its absence.

*Finally*, to further sanitize the SeekingAlpha content, we explore a novel approach (i.e., stylometry analysis) to identify Sybil accounts that collude to publish manipulation articles. The key intuition is that multiple Sybil accounts controlled by the same attacker should exhibit unusually similar writing styles. Our approach produces promising results on ground-truth data and detected new suspicious authors.

A recent article in a financial journal also studied SeekingAlpha and showed statistical correlation between its content and earning surprises [Chen et al. 2014]. In contrast, our work contrasts the performance of expert- (SeekingAlpha) versus peer-based (StockTwits) systems, evaluates the performance of realistic and simple trading strategies, and reports user views of SeekingAlpha through detailed surveys.

We also implement a website<sup>1</sup> to run our algorithms in real time. On the website, we provide a tool to query the sentiment of a stock and present a live simulation of our trading strategy. Our strategy works well and outperforms the market when mining the data in real time.

In summary, the rise of collaborative investment analysis has significantly changed how retail investors manage their investments. Our analysis shows that even on curated sites such as SeekingAlpha, most articles are poor indicators of market performance. However, a subset of SeekingAlpha authors provide valuable content that can be leveraged to build trading strategies that significantly outperform the broader markets. More importantly, these authors can be identified not only by their statistical performance but also more easily by the feedback their articles generate from other users. This shows that even for a complex and domain-specific issue such as stock trading, broader input from the crowd can help identify high-quality content in a sea of data. Finally, results from our user survey confirm that most SeekingAlpha users have seen and have learned to distinguish biased or manipulative articles from useful articles. Also, automated tools like stylometry analysis can be useful to such platforms to identify colluded manipulation contributors.

## 2. BACKGROUND: SEEKINGALPHA AND STOCKTWITS

**Seeking Alpha.** Launched in 2004, SeekingAlpha (SA) is the most popular platform today for independent stock analysis. As of early 2014, SA has more than 8 million unique monthly viewers and 3 million registered users [Tweney 2013]. SA’s content is mainly contributed by roughly 8,000 registered *contributors* [SeekingAlpha 2014], and articles are vetted by an editorial board before publication on the site. Users can subscribe to stocks of interest to receive related articles and news summaries,

<sup>1</sup><http://hotcrp.cs.ucsb.edu:88/>.

*follow* contributors to receive their articles, and interact with contributors and other users through *comments* on articles. SA contributors include independent investors, portfolio managers, professional investors, and investment firms. Roughly 400 out of 8,000 contributors self-identify as investment firms. SA pays each contributor \$10 per 1,000 page views on articles.

**StockTwits.** StockTwits (ST) started in 2009 as a financial social network for sharing ideas among traders. Anyone on StockTwits can contribute content—short messages limited to 140 characters that cover ideas on specific investments—and post their messages to a public stream visible to all. There’s no editorial board or content curation, and users are not compensated for their messages. Like Twitter, ST users *follow* others to build directional social links, and also follow the stock symbols they are interested in. Unlike SeekingAlpha, StockTwits provides real-time streaming of investor sentiment toward individual stocks. As of the end of 2013, StockTwits has 300K registered users and its content reaches an audience of 40 million across the Internet [StockTwits 2014].

### 3. METHODOLOGY

Seeking Alpha and StockTwits represent the largest and most representative sites in expert- and peer-based investment analysis. In this study, we seek to quantify correlation between sentiment in user-contributed investment analysis to real movements in stock equities, how and if such correlations can be leveraged for investment gain, and how users view and utilize platforms such as SeekingAlpha in their investments. Our methodology is as follows:

- First, we gather complete datasets of contributed articles from SeekingAlpha and “twits” from StockTwits. We then develop sentiment analyzers for both datasets and evaluate their accuracy.
- Second, we compute statistical correlation between sentiment of contributed content and the performance of stocks they discuss. We do so for different time scales and both individual stocks and the aggregate market. We also sort authors by their performance to identify authors whose content consistently correlate with stock performance.
- Third, we propose strategies for identifying and trading stocks using sentiments from top authors in both platforms and evaluate them against baseline market indices. We explore the efficacy of strategies that identify top authors by historical performance and by interactions with other users.
- Finally, we use a large user survey of SeekingAlpha users and contributors to understand how they utilize social investing platforms and their views on stock manipulation. Stock “pump and dump” scams have been discovered on SA in the past [Feuerstein 2014; Kimelman 2014]. We then explore techniques (e.g., stylometry analysis) to identify Sybil authors that participate in stock manipulations.

### 4. DATA COLLECTION

**SeekingAlpha.** In April 2014, we crawled the complete list of historical articles published on SeekingAlpha since its launch in 2004. This produced 410,290 articles written by 8,783 authors and 2,237 news and conference transcripts (transcripts cover board meetings and conference calls). Our analyses focus on articles and do not include news in the form of SA “market currents” or transcripts. Our crawl also produced 4,115,719 total comments, of which 75% were written by 227,641 noncontributing users. The remaining 25% are from authors themselves. We crawl profiles of all authors and active users for their brief bio and number of followers and followees.

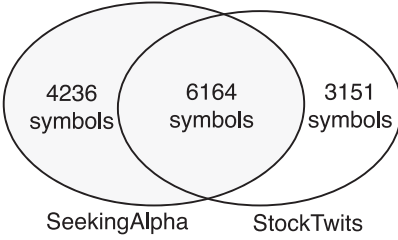


Fig. 1. Stock symbols extracted from SA and ST.

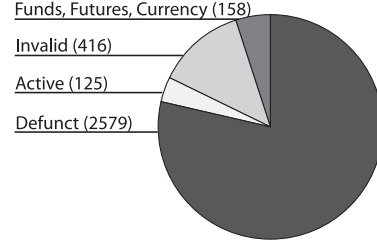


Fig. 2. Breakdown of symbols without history price.

Each SeekingAlpha article has an “about” field that lists what stock(s) the article discusses; 163,410 (about 40%) of our articles have at least one stock symbol in their about field. Articles without stock symbols usually discuss overall market trends or sectors of stocks. From our entire dataset of SeekingAlpha articles, we were able to extract 10,400 unique stock symbols.

**StockTwits.** We were fortunate to receive permission from StockTwits Inc. to access their historical message archive, including all messages posted from 2009 (initial launch) to February 2014. The dataset contains 12,740,423 messages posted by 86,497 users. Each message includes a messageID, author’s userID, author’s number of followers (followees), timestamp, and message text. Each message is limited to 140 characters, and stock symbols are preceded by a “CashTag” (\$). In our dataset, about 67% of StockTwits messages have at least one CashTag. From these messages, we extract 9,315 unique stock symbols.

StockTwits messages can also be labeled “bullish” or “bearish” by the author to label their sentiments toward the mentioned stocks—bullish (bearish) means investors believe the stock price will increase (decline). Ten percent of messages (1.3 million) have this label. We use these labeled messages later as ground truth to build and evaluate sentiment analysis tools. The sentiment analysis tools serve to extract author opinions toward the discussed stocks. More details can be found in Section 6.

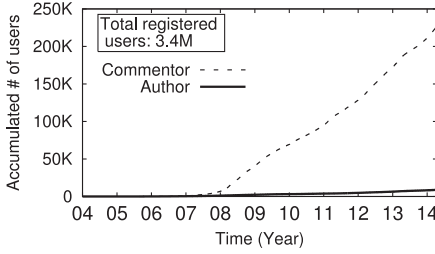
**Stock Historical Price Data.** Our two datasets include a total of 13,551 *unique* stock symbols. Symbols from two sites do not completely overlap (Figure 1): 6,164 symbols appear in both datasets, and most represent stocks on the NASDAQ and NYSE exchanges. SeekingAlpha-only symbols (4,236) are mostly small stocks sold on the Over-The-Counter Bulletin Board (OTCBB), while StockTwits-only symbols (3,151) are mostly from the Toronto Stock Exchange.

We use the Yahoo! Finance open API [Yahoo 2014] to crawl historical prices for all stock symbols. For each stock, we obtain its historical *daily* opening and closing prices, volume, and intraday price range. Of our 13,551 symbols, we found data for 10,273 symbols. We tracked down the 3,278 missing symbols using both Yahoo Finance and Bloomberg (see Figure 2). First, 2,579 missing symbols are defunct (i.e., symbols made invalid due to corporate breakups, merge/acquisitions, or bankruptcy). Second, 125 are active stocks, either on foreign exchanges or OTC stocks not covered by Yahoo. Third, 158 symbols are ETF or mutual funds, futures contracts, and currencies. Finally, we manually inspect the remaining 416 symbols and find they are often user-defined symbols such as \$CRASH or nonlisted companies such as \$QUORA. Missing symbols account for 7% of SeekingAlpha articles and 6% of StockTwits messages; thus, we believe it would not impact our overall conclusion. We summarize the final datasets used in our analysis in Table I.

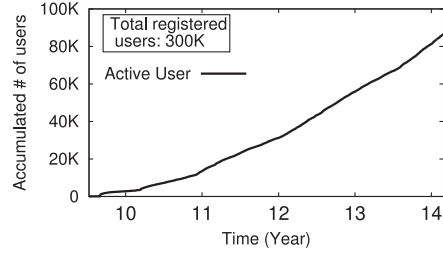


Table I. Basic Statistics of Collected Data

Site	Data Since	Total Posts	Posts w/ Stocks	Active Users (Authors)	Covered Stocks
SeekingAlpha	2004	410K	163K	228K (8783)	10.4K
StockTwits	2009	12.7M	8.5M	86K (86K)	9.3K

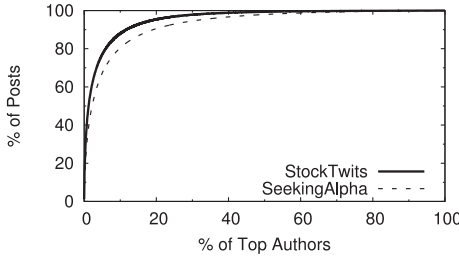


(a) SeekingAlpha

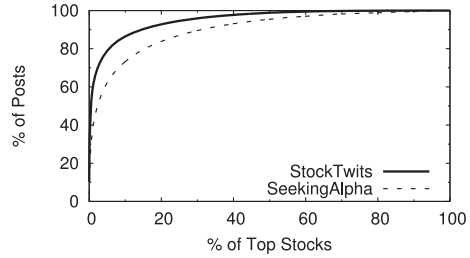


(b) StockTwits

Fig. 3. Total number of authors and active users over time. The number of active users is only a small portion of all registered users.



(a) Posts by authors.



(b) Posts per stock.

Fig. 4. Distribution of articles over authors and individual stocks.

## 5. PRELIMINARY DATA ANALYSIS

Here, we briefly analyze our datasets to understand the structure of user communities and content in these two systems. The two systems are quite different. SeekingAlpha focuses on detailed, blog-like contributions by a small group of “experts” further curated by editors, and the large majority of users consume and comment on articles. In contrast, StockTwits encourages short, terse contributions by all users. We compare and contrast the platforms on growth over time, skew of author contributions, and distribution of stocks covered.

**User Growth.** Figure 3 plots the growth of users over time for both systems. Recall that our SeekingAlpha data includes all users who have contributed or commented on at least one article. At both sites, active users are growing at a stable rate but only make up a small portion of all registered accounts (236,000 active vs. 3.4 million accounts in SeekingAlpha, and 86,000 active vs. 300K accounts for StockTwits).

**Distribution of Content over Authors and Stocks.** An author’s contribution to the platform is measured by the number of articles from the author. For both SeekingAlpha and StockTwits, we find the contribution per author is highly skewed (Figure 4(a)). On SeekingAlpha, 20% of the most active users contributed 80% of articles, while on StockTwits, 20% of active users contributed 90% of the messages. Even though StockTwits tries to leverage the power of the crowd, it has an even higher skewness in

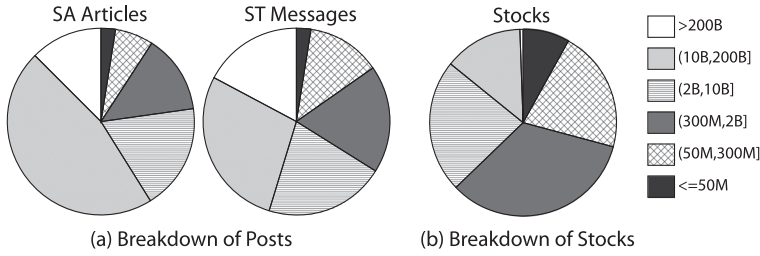


Fig. 5. Percentage of posts/stocks in different market cap categories.

content contribution than SeekingAlpha. This sheds concerns that the wisdom of the crowd is likely to be dominated by the most active authors.

Content posted on both sites is also highly skewed to a small portion of “popular” stocks (Figure 4(b)). More than 70% of SeekingAlpha articles cover the top 10% of the most popular stocks. The skew is even stronger in StockTwits, with 90% of messages focusing on the 10% most popular stocks.

Figure 5 shows the heavy emphasis of articles on large capitalization companies: 47% of SeekingAlpha articles and 28% of StockTwits messages cover stocks of companies between \$10 billion and \$200 billion in market cap, which account for only 14% of all stocks. The emphasis is stronger for the largest companies (market cap >\$200 billion). They account for only 0.3% of all stocks but are covered by 10% to 15% of the content on both platforms.

## 6. SENTIMENT EXTRACTION

Our analysis on the value of user-contributed investment analysis hinges on our interpretation of sentiment in SeekingAlpha articles and StockTwits messages. The first step in this process is developing reliable tools to extract sentiment (positive or negative opinion) on stocks from posted articles and messages. We discuss our sentiment analysis techniques here, and rely on them in later sections to compute stock performance correlation and to drive trading strategies.

Our approaches to extract sentiment from SeekingAlpha and StockTwits are quite different. More specifically, SeekingAlpha articles are sufficiently long to apply an approach using a keyword dictionary, while we applied a supervised machine-learning approach for the short messages in StockTwits, using messages with “bullish” or “bearish” labels as training data. Our validation results show we achieve an accuracy of 85.5% for SeekingAlpha and 76.2% for StockTwits. We note that these accuracy results are on par (Senti-Strength, 0.815) or significantly better than existing sentiment analysis techniques (e.g., Sentic-Net, 0.590, and Happiness Index, 0.639) [Chen et al. 2014; Gonçalves et al. 2013].

### 6.1. Sentiment Analysis: Seeking Alpha

We develop a dictionary-based method to extract sentiment from SeekingAlpha articles. At a high level, we measure author sentiment (toward a stock) based on the ratio of positive and negative keywords in the article. We rely on a widely used financial sentiment dictionary [Loughran and McDonald 2011] to identify positive and negative keywords in the article and calculate the sentiment score as  $S = \log \frac{1 + \sum P_i}{1 + \sum N_i}$ , where  $P_i$  ( $N_i$ ) is the number of positive (negative) words or phrases in sentence  $i$ . The sentiment score is a decimal value with high positive value indicating strong and positive sentiment, and vice versa. For example, an extremely positive article with 100 positive

words and 0 negative words gets a score of 4.6; an extremely negative article with 100 negative words gets a  $-4.6$ .

However, there are problems with applying this method naively to SeekingAlpha articles. First, many articles discuss multiple stocks, and sentiments may be quite different for each discussed stock. Generating one sentiment score for all stocks is clearly oversimplifying. Second, simple keyword counting can easily make mistakes. For instance, “low risk” contains negative keyword “risk” but the overall sentiment is positive. Also, negation usually changes the sentiment of words, such as “not good” or “no benefits.”

We make several refinements to our method to address these challenges. First, we partition multistock articles and assign individual sentences to each stock symbol (stock symbols are easily recognized as hyperlinks in each article). Our method is a simple distance-based slicing: we consider stock symbols (and company names) as landmarks, and we assign each sentence to the closest landmark in the article. Since cases where a single sentence contains multiple stocks are rare, we do not include them in our analysis to avoid error. Next, we make two adjustments to basic keyword counts. First, we identify the sentiment of noun phrases such as “higher revenue,” “low return,” and “low risk.” We extract frequent noun phrases that occur in more than 1% of articles and manually label their sentiment. Second, we reverse the sentiment of words or phrases affected by negation words [Pang et al. 2002].

To validate our method, we sample 300 articles from our article collection and manually label their sentiment as positive or negative. We have three graduate students read each article and vote for the final label. Then we run our sentiment extraction method on these articles to generate sentiment scores. The result shows our method achieves 85.5% accuracy. Note that this accuracy only considers the polarity of the scores, that is, whether an article is positive or negative. This accuracy is comparable with a prior paper on sentiment analysis [Gonçalves et al. 2013]. In our further analysis, we will aggregate the sentiment (opinion) of articles on the same stock. In this case, the integrated sentiment will achieve a higher accuracy [Sheng et al. 2008].

## 6.2. Sentiment Analysis: StockTwits

Roughly 10% of StockTwits messages already have sentiment labels, either “bullish” or “bearish.” Our goal is to extract sentiment for the remaining 90%. We choose to use supervised machine learning, since messages are too short for dictionary-based approaches [Gonçalves et al. 2013]. This is also confirmed by our experiments, and the accuracy is only 16.2% by dictionary-based methods for StockTwits.

To build a machine-learning classifier, we follow prior work [Pang and Lee 2008] to use the existence of unigrams (i.e., unique words) as features. To reduce noise in the machine-learning model, we exclude infrequent unigrams that occur fewer than 300 times over all messages, and remove stopwords, stock symbols, and company names from messages. We use the ground-truth messages as training data and empirically test multiple machine-learning models, including Naive Bayes, Supported Vector Machine (SVM), and Decision Trees. We randomly sample 50K messages labeled as “bearish” and 50K labeled as “bullish,” and run 10-fold cross-validation. We find the SVM model produces the highest accuracy (76.2%) and use SVM to build the final classifier.

The sentiment score of StockTwits messages is binary: 1 indicates positive sentiment and  $-1$  indicates negative sentiment. For rare messages ( $<5\%$ ) with multiple symbols, we attribute the same sentiment score to all symbols in the message (messages are too short to slice).

## 7. PREDICTING STOCK PRICE CHANGES

Using our sentiment analysis tools, we can now quantify the value of SeekingAlpha and StockTwits content, by measuring statistical correlation between their sentiment



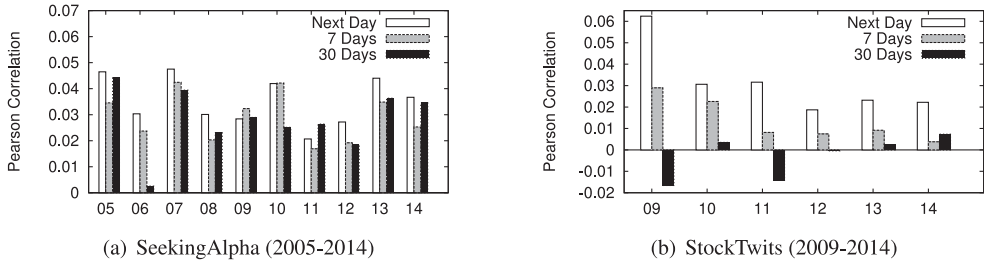


Fig. 6. Pearson correlation coefficient between article sentiment and near-term stock price movement (in 1, 7, 30 days).

toward individual stocks and each stock's near-term price performance. Our goal is to study such correlation for different time periods after the content was published, for both platforms and over different historical periods (to account for bull/bear cycles in the stock market).

### 7.1. Per-Article Sentiment and Stock Performance

We start by studying how well each article predicts the future price trends of the stocks it discusses. We compute the Pearson correlation coefficient [Pearson 1895] between an article's sentiment (positive or negative) and a stock's future price change. For simplicity, we ignore the magnitude of price movements and strength of sentiments, and reduce both metrics to binary (positive/negative) values. The Pearson correlation coefficient is widely used to measure the linear correlation between two variables.<sup>2</sup> Pearson correlation is suitable here because the variables are linear related. With binary values, Pearson correlation is equivalent to the phi correlation. Its value ranges from  $-1$  to  $1$ , where  $1$  means perfect positive correlation,  $0$  means no correlation, and  $-1$  means perfect negative correlation. In this context, the Pearson coefficient is  $1$  if a stock always increases in value after it is discussed positively by an article.

We compute the Pearson coefficient between two variables  $S$  and  $P$ .  $S$  is  $1$  ( $-1$ ) if the article's sentiment is positive (negative); and  $P$  is  $1$  or  $-1$  depending on whether the discussed stock goes higher or lower in price. We study stock price changes in different time windows after a relevant article is published, including the next day, the next week, and the next month. For articles with multiple stock symbols, we count each stock as one data point. We also group articles in each year and compute the per-year Pearson correlation coefficient to understand the consistency of correlation across different years. Results are shown in Figure 6.

First, we observe that correlation is extremely low across different time windows and different market years for both systems. To better understand the Pearson values, consider that a prediction history of 75% correlation would produce a Pearson coefficient of  $0.4$ . The most significant correlation in our results is  $0.05$ , which translates to a prediction accuracy of 53%, 3% better than a random guess. This means that taken as an aggregate, SeekingAlpha articles and StockTwits messages provide minimal value for investors.

Looking closer, SeekingAlpha generally does a bit better than random, while StockTwits has weaker, sometimes negative correlations. Clearly StockTwits is better as a gauge of instantaneous market sentiment and a poor predictor of even near-term performance. In contrast, SeekingAlpha is a bit more consistent over different time

<sup>2</sup>For two variables  $X$  and  $Y$ , the Pearson correlation coefficient  $\rho_{X,Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}$ , where  $\mu_X$  and  $\sigma_X$  is the mean and standard deviation of  $X$  and  $E$ , is the expectation.

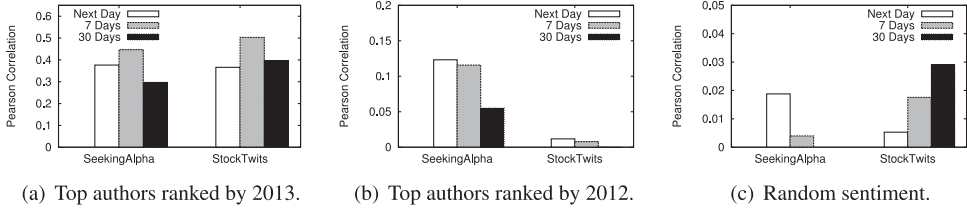


Fig. 7. Correlation analysis of stock performance in 2013 by top authors. We consider two sets of authors, top authors based on performance in 2013 (left) and top authors based on performance in 2012 (middle). As a reference, we also show correlation with random sentiment.

windows and still has some value for predicting price movements in the following month. We note that SeekingAlpha accuracy is a bit lower for market years with high volatility (2008–2009, 2011–2012).

Predicting price changes between two time points can be challenging given the high volatility of individual stock prices. Here, we simplify the metric by only predicting the up and down of *average* price within a certain time window before and after the article. For a simple experiment, we use 2013 data to compute Pearson correlation between article sentiment and the average price of 1 week before and after the article. We find the correlation scores are slightly improved (0.067 for SA and 0.035 for ST) but are still very weak correlations.

The natural follow-up question is: is the weak correlation consistent across all authors, or are there authors whose contributions provide consistently good predictors of stock performance but are buried in the noise? Here, we perform a simple test to determine if more accurate authors exist. We rank authors based on how well their articles predict stock returns in a single year (e.g., 2013). For each author, we measure the average hypothetical return per article for all his or her articles as a percentage after a time window  $W$ . If  $P(x)$  is the stock closing price of a given day  $x$  and the article is posted on day  $d$ , then return  $R$  from a positive article is  $R = \frac{P(d+W)-P(d)}{P(d)}$ , and the return on a negative article is  $R = -\frac{P(d+W)-P(d)}{P(d)}$ .

For our experiment, we set  $W$  to 1 week and compute the average return per article to rank authors in 2013. We then take a closer look at correlations of 500 stocks discussed by the top-ranking authors. Figure 7(a) clearly shows that correlation scores for top authors in SeekingAlpha and StockTwits are both very high (around 0.4); that is, top authors can predict stock movement within a week with  $\sim 75\%$  accuracy.

In practice, we cannot use data from the current year (e.g., 2013) to identify top authors for that year. We can only rely on past performance to guide us. Thus, we repeat the experiment using 2012 data to rank authors, and then we study the performance of their 2013 stock recommendations. As expected, correlation results for those authors' stocks in 2013 (see Figure 7(b)) are much lower. Top SeekingAlpha authors show a significant correlation score around 0.12 ( $p = 0.004$ ), which is still much better than the average. This confirms our intuition, that filtering out the “noise” does indeed reveal more accurate contributors in the crowd. Note that this does not hold for StockTwits; that is, no StockTwits authors can consistently predict stock performance over different time periods. StockTwits is peer based with a lower level of user expertise than Seeking Alpha on average, and no review or moderation on user posts.

To further validate our results, we also repeat our experiments with random sentiment. That is, we assign a random sentiment to each article, instead of using the sentiment score we calculated in Section 6. We plot the correlation using random sentiments in Figure 7(c). As shown in the figure, random sentiments give very weak correlations ( $< 0.03$ ). The result gives us two insights. First, the high correlations in

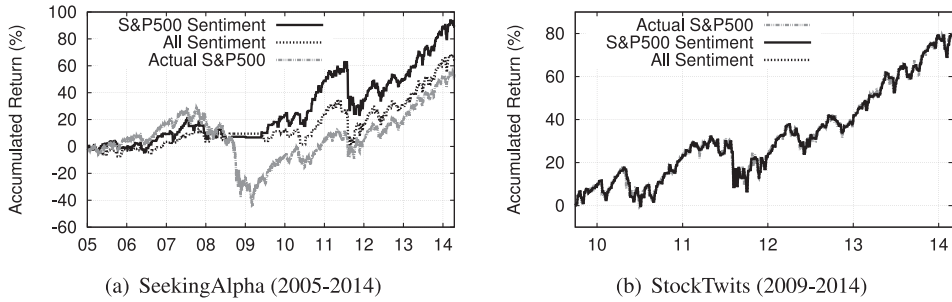


Fig. 8. Predicting S&P 500 index using aggregated sentiment of whole site versus only using sentiment related to the indexed 500 stocks.

Figure 7(b) come from the expertise of authors, instead of the results of good luck. Second, it confirms that our sentiment classification is accurate enough from the angle of real applications.

## 7.2. Aggregated Sentiment for Market Prediction

Since the correlation between articles and individual stocks is weak, we consider aggregated sentiment of all articles as a possible predictor for the market as a whole. Here we use the S&P 500 index as a metric for overall market performance, which is a widely accepted market index based on the market capitalizations of 500 large companies. We treat the S&P 500 index as a single stock and trade it based on the *aggregated* sentiment over time. In practice, this can be done using SPY, an Exchange Traded Fund (ETF) that tracks the S&P 500.

The process is intuitive: we start by holding a position in the S&P 500. After every  $K$  days, we check the aggregated sentiment of articles posted in the past  $K$  days and choose to either buy or sell the S&P 500. Sentiment over a window of  $K$  days is computed by first computing the average sentiment of all articles for each day, then setting the overall sentiment to positive if there are more net-positive days than net-negative days. If the sentiment for a window is negative, we sell our entire position (if any) in the S&P 500. If the sentiment is positive, we buy back our full position if we have none or hold it if we already have a position.

**Long-Term Performance.** We simulate this trading strategy using data from SeekingAlpha (January 2005 to March 2014) and StockTwits (September 2009 to February 2014).<sup>3</sup> We set time window  $K$  to 1 week.<sup>4</sup> On each dataset, we run two configurations, one using “all” sentiment from the entire network and the other only taking sentiment specifically about the 500 stocks listed in S&P 500 index.<sup>5</sup> As a baseline, we run a “buy-and-hold” strategy on the S&P 500 index, that is, holding the stock for the entire duration of the simulation.

Figure 8 plots the total return (normalized by initial investment) accumulated over time. For SeekingAlpha (Figure 8(a)), we find that both configurations outperform the actual S&P 500. Aggregated sentiment can generally predict market trends. Not surprisingly, sentiment specifically about the 500 stocks in the index produces more accurate results. A closer look shows that our strategy significantly outperforms the

<sup>3</sup>SeekingAlpha only had three total articles in 2004. Thus, we start our SeekingAlpha simulations from 2005.

<sup>4</sup>We tested with  $K$  as 1 day/week/month and find weekly aggregation produces a better return [Wang et al. 2015].

<sup>5</sup>The stocklist of the S&P 500 index changes periodically, and we adapt the list in our evaluation accordingly.

real market during 2008–2010, when the financial crisis caused the stock market (and the S&P 500) to lose more than half of its value. Given the overall negative sentiment in SeekingAlpha, our strategy held no positions and avoided much of the market losses. For StockTwits (Figure 8(b)), we find that all three lines completely overlap. In fact, after we aggregate the sentiment of the whole network, StockTwits' overall opinion toward the market is almost always positive. Our sentiment-driven trading is equivalent to buy-and-hold.

In summary, our analysis shows that sentiment to performance correlation is quite low for both SeekingAlpha and StockTwits. However, there are authors who consistently provide high-correlation analysis in their articles. The challenge is to identify them efficiently. Next, we address this challenge and develop practical sentiment-driven strategies for stock trading that can significantly outperform the market.

## 8. PRACTICAL SENTIMENT-BASED TRADING

Thus far, we have determined that while correlation to stock performance is low over all articles as a whole, certain subsets of the user population contribute content with much stronger correlation to stock performance. Two key questions remain. First, can these “valuable” authors be identified easily? Second, once identified, can their content be analyzed to form the basis of real stock trading strategies, and how well would such strategies perform?

We address these questions in three parts. First, we explore several possible ranking heuristics for identifying the valuable authors (and their analysis contributions) from both SeekingAlpha and StockTwits. Second, we consider possible stock trading strategies based on sentiment analysis of these contributions. Finally, we use historical stock data to drive simulations of these trading strategies and use empirical results to draw conclusions about the value of these top authors and the efficacy of our mechanisms to identify them.

### 8.1. Ranking Authors

To identify the (possibly small) subset of top authors in our systems, we explore two different sets of heuristics. First, we consider using empirical past performance (i.e., correlation between sentiment and stock performance) as a gauge to rank authors. While this is likely the most direct way to rank authors by performance, its computation requires access to significant resources, including past stock data and sentiment analysis tools. Second, we consider a simpler alternative based on user interactions (comments). The intuition is that user feedback and engagement with content provides a good indicator of valuable content.

**Ranking Authors by Prediction Accuracy.** Our first ranking heuristic is purely empirical: we rank authors based on how well their previous articles predict stock returns. For a given author and historical time period (e.g., a year), we compute the average hypothetical return of his or her articles posted during that given period. Recall that we used this in the previous section as a metric of an author's prediction ability. A variant of this ranking metric is an author's average hypothetical return *per stock*. Compared to the *per article* metric, this highlights authors who have consistently good performance over a range of stocks over those who write numerous articles on a small set of stocks. We consider both metrics in our experiments.

**Ranking Authors by Received Comments.** The challenge with empirical performance-based metrics is that they require significant resources in historical data and computation. Here, we also consider the value of a simpler approximation based on reader engagement. The intuition is that the audience in these systems is a valuable

asset, and we can observe reader responses to contributed content and indirectly infer the value of the content. More specifically, we use two heuristics that rank authors based on either total number of comments or comments per article. Without semantic analysis of comments, we use the number of comments over a prior period (e.g., a year) as an approximate indicator of user agreement. In our experiment, the top authors can receive as many as 10,000 comments each year, and a single popular article can even receive more than 1,000 comments.

## 8.2. Sentiment-Based Stock Trading Strategies

Given a ranking of top authors, the next step is to formulate a stock trading strategy that takes advantage of (hopefully) valuable and predictive sentiment on individual stocks. Our strategies select a group of stocks to trade each year from stocks mentioned by articles from top authors from the prior year (sorted by either correlation or comments). For simplicity, we build a portfolio for our simulations from the 500 stocks mentioned by the top-ranked authors. Experiments with smaller portfolios show highly consistent results and are omitted for brevity.

In terms of trading strategies, we implement two simple strategies: a basic “long” strategy (buy or sell based on sentiment) similar to the one used to trade the SPY in last section, and a more aggressive “long/short” strategy that allows investors to short stocks. For both strategies, we trade stocks on a weekly basis based on earlier results.

**Long Strategy.** Our long strategy builds a portfolio by initially spreading funds evenly to purchase  $N$  stocks, and  $N = 500$  in our examples. We make trading decisions on each stock independently on a weekly basis, using weekly sentiment from the same top authors identified from the prior year, that is, the same top authors used to select the 500 stocks. For each stock in the portfolio, we sell our entire position in the stock if the sentiment about this stock in the past week is negative. Otherwise, we hold the stock (or buy it back if we sold it earlier). If the stock was not covered by a “top author” in the past week, no action is taken. We use the same sentiment aggregation method as before but only consider the top author’s sentiment on each stock. The return of the portfolio is the sum of returns over all stocks. We use the adjusted stock price in case a stock was split, but we did not consider stock dividend in our experiments.

**Long/Short Strategy.** A more aggressive “long/short” strategy not only buys stocks with positive sentiment but also proactively “shorts” stocks with negative sentiments. Shorting is the practice of borrowing a stock to sell at the current price and buying it back at a lower price [Diamond and Verrecchia 1987]. Investors target stocks they believe will drop in value and profit from the drop between their sell price and repurchase price. Shorting is generally considered to be very risky, because the price of a stock can go up without limit; thus, there is no limit to the size of potential losses on a short position.

In our short strategy, if the aggregated sentiment on a stock is negative in the previous week, we not only sell any shares of the stock in our portfolio but also short the stock for a week (and buy shares back at the end of the week). We short a number of shares equal to a value of  $\frac{1}{N}$  of our total portfolio. If and when we have lost 100% of the value initially allocated to a stock, then we close our position on that stock permanently.

## 8.3. Empirical Evaluation

We evaluate trading strategies generated using a combination of author-ranking heuristics and long versus long/short trading strategies. For our author-ranking heuristics, we use average return per article (*PerA*), average return per stock (*PerS*), number of total comments (*AllCom*), and average comments per article (*AvgCom*). As described



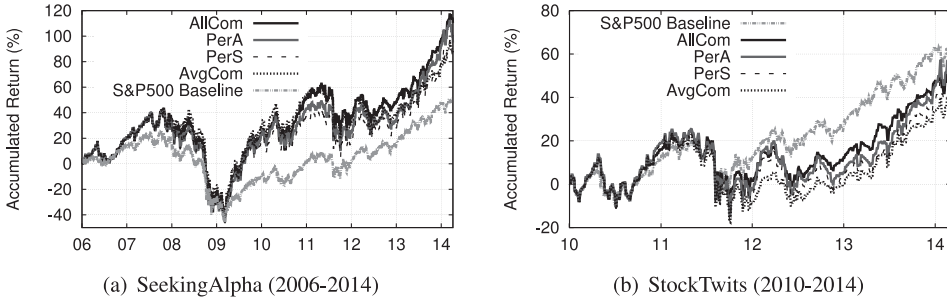


Fig. 9. Investing results on 500 stocks recommended by top authors over the years.

earlier, we choose 500 stocks mentioned by the top-ranked authors and split the funds of a hypothetical portfolio evenly among them. Each week, we trade them based on aggregated sentiment from our chosen top authors (the same authors used to select the 500 stocks) in the previous week. By default, we regard 1 year as a cycle, and rerank authors and reset the list of stocks for the portfolio at the beginning of each year (using the heuristics of previous year).

#### Q1: Do our strategies outperform the broader markets?

We simulate our strategies using historical stock price data and compare them to a baseline following a “buy-and-hold” strategy on the S&P 500 index. We ignore transaction fees in our simulations and initially focus on the long-only strategy. We plot results in Figure 9.

The first takeaway is that SeekingAlpha clearly outperforms the baseline market under all settings. For example, our “all comment” strategy produces a normalized total return of 108% at the end of the 8-year period, compared to 47.8% of the S&P 500. This represents a more than 10% annual compounded return, during a time period that includes two market crashes (2008, 2011), and *not including dividends*. Since StockTwits only started in 2009, its simulations ran on only 4 years of historical data. The same strategy on StockTwits produced a total return of around 54.5% from 2010 to 2014. This is a good return in absolute terms, but significantly below the baseline S&P 500 (64.1%) during the same timeframe. Results show that SeekingAlpha consistently generates higher annual returns than StockTwits.

**Implications.** This result is *significant*, because it means we can in fact use empirical methods to identify the articles of value from SeekingAlpha. More importantly, we *significantly* outperform the S&P 500 using a very simple trading strategy that ignores the semantic meaning of the articles and uses only binary sentiment values. As context, we consider hedge funds, which manage money for large investors, and charge annually 2% of assets managed and 20% commission on all gains. The \$2.5 *trillion* hedge fund industry has underperformed the S&P 500 for 5 years in a row, and has fallen behind the S&P by 97% since 2008 [Kishan and Bit 2013]. Significantly outperforming the S&P over a period of 10 years would be considered excellent performance for a managed hedge fund. Achieving this result by mining an open community like SeekingAlpha is ample evidence of the very high quality of analysis by the small set of expert authors.

#### Q2: What ranking method identifies experts most effectively?

The next big question is whether we can validate a simple methodology for identifying the top-performing authors. Among a number of author-ranking heuristics, we find the all-comment metric ( $R_{allC}$ ) to obtain the highest level of investment returns in our SeekingAlpha simulations (see Figure 9). Similarly, the same strategy also performs

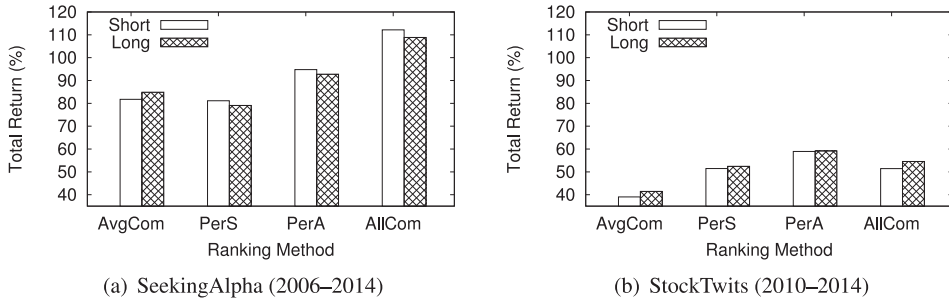


Fig. 10. The total return with long strategy versus short strategy.

the best for StockTwits, outperforming other metrics in most years. The number of top authors varies across different ranking methods. The typical number of top authors is between 200 and 300 (out of 8K authors). That is, the wisdom of around 2% of experts contributes to the high returns in our experiments.

**Implications.** This result implies that not only does something as simple as comment count do a great job of identifying top authors in SeekingAlpha but also it does even better than heuristics based on prior-year performance. Note that we are not leveraging sentiment analysis on comments, only the number of comments. This implies that the majority of comments are supportive notes that validate the article’s analysis and insights. More importantly, this highlights the value of the SeekingAlpha user population as a filter to identify the best analysts among the authors. Even for a subject as complex and domain specific as stock analysis, a reasonably knowledgeable crowd can serve to pinpoint valuable content in an otherwise mixed collection of content.

### Q3: Do more aggressive strategies improve performance?

We also study the impact of allowing shorts of stocks along with traditional long positions. Shorting stocks is a typical strategy used by hedge funds but rarely used by individual investors because of its potential for unbounded loss. We repeat the same experiment as earlier but add a long/short strategy in addition to the long strategy. The results are plotted in Figure 10 (note the different timeframes in the two subfigures). The high-level observation is that adding shorts improves performance for strategies based on SeekingAlpha, but not significantly enough to justify the added risk.

**Implications.** Shorting stocks is a highly valued tool for hedge funds, which are expected to use it to produce positive returns even in negative markets. Yet our results show that the majority of user-contributed short strategies do not produce significant gains over long-only strategies. Given the significant added risk, this suggests that sentiment-based trading strategies should focus on long-only strategies to minimize risk while achieving the gains of a long/short strategy.

### Q4: Does this approach support “real-time” trading?

We evaluate the impact of a smaller time window to aggregate sentiment and trade, especially given the real-time nature of StockTwits. We perform experiments using one-day time windows and show the results (long-only strategy) in Figure 11. For StockTwits, daily trading leads to incrementally higher returns compared to those using weekly windows. The opposite is true on SeekingAlpha, where the lower frequency of articles means that weekly trading outperforms daily trading strategies.

**Implications.** This again confirms the real-time nature of StockTwits messages. In practice, transaction fees will play a role in determining the trading frequency that

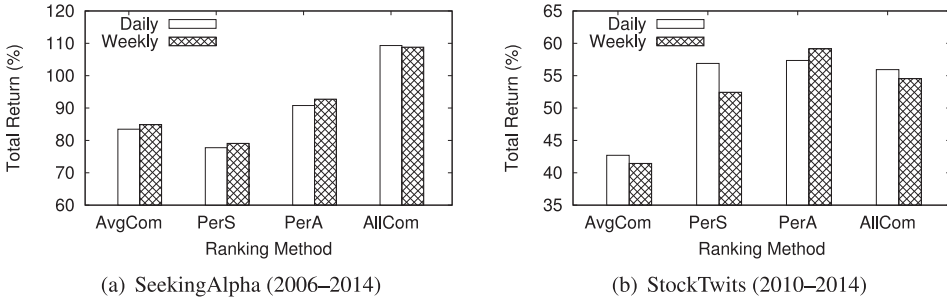


Fig. 11. The total return with daily trading and weekly trading.

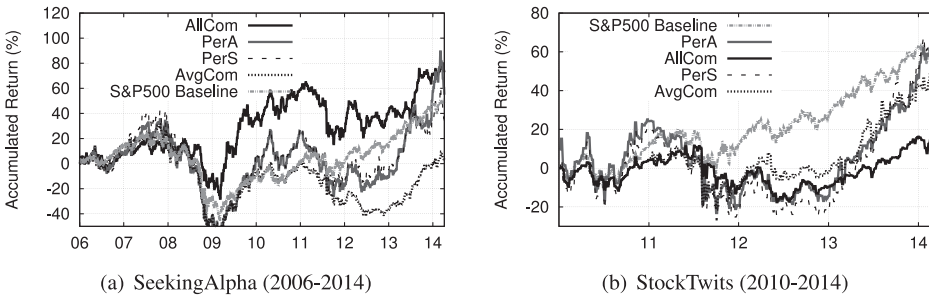


Fig. 12. Investing results on 30 stocks recommended by top authors.

maximizes gains. Optimizing that tradeoff would require more complex models beyond the scope of this article.

#### Q5: Can individual investors benefit from this trading strategy?

A key difference between individual investors and hedge funds is the number of stocks they operate on. Individual investors can only operate on tens of stocks. The above-500-stock experiments simulate the hedge funds scenario, and we now simulate the trading of individual investors with  $N = 30$  stocks. We choose the number 30 because highly followed stock indices such as the Dow Jones Industrial Average (DJIA) includes 30 stocks. As shown in Figure 12, SA still outperforms the market baseline and ST is slightly worse than the baseline. Compared with the hedge funds setting (Figure 9), the overall return for individual investors is a little lower. Intuitively, with only 30 stocks, there's less room to tolerate mistakes, and mispredictions can easily drag down the overall return of the portfolio.

**Implications.** Due to the small size of the portfolio, individual investors are more sensitive to the loss of each stock they have. As a result, they need to be more careful about their stock selection and use better strategies to avoid high-risk stocks (e.g., constrain the ratio of pink sheet stocks in the portfolio).

#### Q6: What if the investor chooses to update the author rankings and stock list more frequently?

We update the stock list in the portfolio every time we rerank authors. We keep the setting of  $N = 500$  stocks and run the strategy with the *AllCom* ranking metric and long trading strategy. The key variable is, instead of ranking authors every year, we update author ranking quarterly or monthly. Consistently, each time we rank authors, we still go back for a time window of 1 year to mine the author's history data. Essentially, we slide this 1-year time window to update author ranking every quarter (or month).

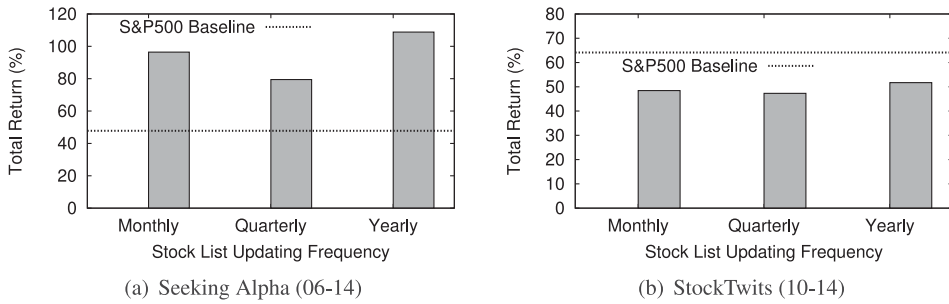


Fig. 13. Impact of the updating frequency of stock list toward the final return (S&P 500 as baseline).

We report the experiment result in Figure 13. We find the trend is consistent for SeekingAlpha and StockTwits: yearly and monthly updates achieve a higher return than the quarterly update. Regardless of the frequency of updating, SeekingAlpha still beats the performance of the S&P 500 baseline.

**Implications.** Frequent updating (monthly update) can help to catch up with the recent hot authors as well as stocks recommended by top authors. Holding stocks for a long time (yearly update) also allows enough time to accumulate the long-term gains. However, updating the portfolio in a compromised frequency (quarterly update) ends up with losing both short- and long-term gains.

#### Q7: How does the strategy work in practical scenarios with real-time data?

We've built a website<sup>6</sup> to present a live simulation of our strategy. The website has two major functions. First, it demonstrates our investment strategy in real time, with an update of strategy and returns every week. It presents detailed sell/hold/buy actions for each stock in real time, with the gains both in the week and accumulatively. Second, it allows users to query the sentiment of a stock and observe the dynamics of sentiment at the daily base. It also provides a sentiment score for each article on the queried stock. From the website, we see our strategy still outperforms the market in practice. For example, our strategy gains a total return of 10.72% on from January 1 to August 15, 2016, which is still higher than the S&P 500 baseline.

**Implications.** Our strategy works quite well in practice. Real investors can benefit from our strategy if they follow our suggestions to buy/sell stocks. They can also make more informative decisions if they know other investors' attitudes toward the stocks.

**Discussion and Limitations.** Given the complexities of financial investments, our models are clearly an oversimplification of the possible trading strategies. We ignore strength of sentiments and reduce strategies to a binary "buy all in" or "sell all shares." Our goal is to understand the lower-bound performance from extracting only the higher-order bits of analysis. Weighing sentiment from articles and more nuanced strategies like "dollar-cost averaging" or "cost-basis averaging" are likely to improve performance further, but we leave the confirmation of those strategies to future work.

One possible concern is whether our system will impact the overall health of the market—by allowing end-users to access the same "optimized" information for trading. We believe this is very unlikely. Intuitively, movements on the stock market are dominated by large hedge funds and pension funds. For example, assets under management of the entire hedge fund industry hit nearly \$3.2 trillion in 2015.<sup>7</sup> Even perfectly synchronized, retail investors following SeekingAlpha would not produce meaningful price

<sup>6</sup><http://hotcrp.cs.ucsb.edu:88/>.

<sup>7</sup><https://www.preqin.com/docs/samples/2016-Preqin-Global-Hedge-Fund-Report-Sample-Pages.pdf>.

movements for the vast majority of individual equities, because the amount of money they can invest into stock markets is too small.

Intuitively, our results suggest the wisdom that a “smaller, smarter crowd” can outperform a larger, more general crowd, especially for crowdsourcing tasks that require specific domain knowledge (financial knowledge in our case). Moreover, the larger crowd is still helpful as members collaboratively discover domain experts through their interactions. We hope future work will explore how to apply this insight to other domain-specific crowdsourcing applications. Most importantly, our results demonstrate that even when the overall crowd produces low-quality results, the crowd itself can help us identify the top experts among them. In other words, the crowd can *identify* good work even when it is incapable of producing it directly. This suggests the design of an interaction-based “meta-reputation” system that would greatly assist users in their discovery of high-quality content.

The implications on collaborative work systems are significant. It is often the case that the crowd as a whole produces suboptimal results, but subsets among them produce exceptionally good results. In some contexts, it is possible to identify these top experts using ground-truth “tests” [Wang et al. 2013c]. More often, however, there is no ground-truth data or there is no way of testing workers without inadvertently influencing them. Our work suggests that in certain contexts, workers’ meta-evaluation of each other’s results could provide a robust way to identify the most capable workers in a population.

## 9. SEEKINGALPHA USER SURVEY

After quantitative analysis, we next deploy a user survey on SeekingAlpha to better understand users’ levels of investment experience and how they feel about the utility and reliability of SeekingAlpha articles. These responses will help us understand the level of impact articles have on SA users and how they deal with any potentially manipulative articles.

### 9.1. Survey Setup

In May 2014, we sent out a user survey via private messages in SeekingAlpha to 500 authors and 500 noncontributing users. We set the number to 500 to reduce the survey bias without spamming over the website. We received 199 responses (95 from authors and 104 from normal users). We chose authors and users to ensure we captured a full range of activity levels. We sorted authors into buckets by the number of articles written and users into buckets by the number of comments, and randomly sampled from each bucket to choose our targets. Here the buckets are equally split in log scale. For example, if the minimal and maximal number of articles the authors write are  $x$  and  $y$ , respectively, we will create 10 equal buckets from  $[\log x, \log y]$ , and the bucket size is  $(\log y - \log x)/10$ . In this way, we can cover users/authors at different activity levels.

We asked normal users five questions and authors four questions (see the full questions in Table II). The questions are closed with a single choice, while for some questions we allow open responses by participants [Olson and Kellogg 2014]. First, to measure demographics, we ask both authors and users about their levels of investing experience (Q1). Second, to understand users’ perceived value of the platform, we ask whether normal users trade stocks based on SA articles (Q2) and whether they trust authors (Q3). For authors, we ask which platforms if any they would use to disseminate investing ideas if SeekingAlpha were no longer available (Q7). Third, in terms of risks in the platform, we ask both authors and users whether they have seen heavily biased (potentially manipulative) articles (Q4). Finally, we ask normal users for their reaction to stock manipulation (Q5), and authors whether they believe articles can actually impact



Table II. Survey Questions to SeekingAlpha Authors and Normal Users (We Limit the Number of Questions to Improve Our Chances of Getting Responses)

ID	Question	Tester	Answer Options
1	How many years of experience do you have in investing in the stock market?	Author & User	a. less than 1 year; b. 1–5 years; c. 5–10 years; d. more than 10 years.
2	How often do you trade stocks based on opinions and information gathered from reading SeekingAlpha articles?	User	a. Every trade; b. Very often; c. Only sometimes; d. Never.
3	Do you think you can trust the authors on SeekingAlpha?	User	a. Yes, all of them can be trusted; b. Most of them can be trusted; c. Only a few of them can be trusted; d. No, I don't trust them.
4	How often have you seen articles on SeekingAlpha that looked like they were heavily biased, written with the intent to manipulate a particular stock (to move its price up or down)?	Author & User	a. Often; b. Sometimes; c. Never; d. Not sure.
5	If and when you did see what looked like an article intent on stock manipulation, what was your general reaction?	User	a. Leave a comment to argue with the article author; b. Report the article to SeekingAlpha administrator; c. Do nothing; d. Not applicable; e. Other (Please specify)
6	Do you think SeekingAlpha articles can impact the future movement of stocks they focus on?	Author	a. Yes; b. No; c. Not sure.
7	If SeekingAlpha did not exist, where would you post your investment ideas?	Author	a. StockTwits; b. Twitter; c. Yahoo Finance Message Board; d. Motley Fool; e. Other (please list them here)

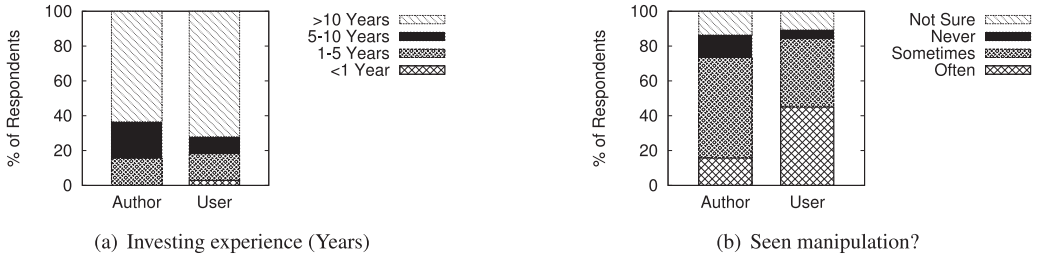


Fig. 14. Results of two questions for both authors and users.

stock price (Q6). In further analysis, we directly calculate the portion of each choice for close-ended questions. For open questions, we manually examined their responses and summarized the key points in the article.

## 9.2. User Responses

We first look at demographics of the survey respondents. We plot responses to Q1 in Figure 14(a). The large majority of respondents have over 10 years of investing experience. Surprisingly, we find that the portion of experienced investors (i.e., >10 years) is higher in normal users than in authors.

**Perceived Value.** First, in response to Q2, 70% of users stated that they “sometimes” or “very often” traded stocks following opinions from SeekingAlpha. Five percent of users stated they “always” relied on information from SA articles. The remaining 25% “never” followed views from articles. Two respondents added clarifying comments that they treated SeekingAlpha as an important information source for research but did

not follow opinions blindly. As a whole, it is clear that SeekingAlpha articles play an important role in most users' investment decisions.

In response to Q3, 83% of users stated they trusted "some" or "most" SA authors, and 1% stated they fully trusted authors. Sixteen percent did not trust authors at all. This suggests that while SeekingAlpha users generally acknowledge the value of the platform, hesitations and concerns remain.

Q7 asked authors where they would disseminate their ideas if SeekingAlpha were no longer available. Surprisingly, only 6% of authors chose StockTwits, and more chose Twitter (7%), Yahoo! Finance Message Board (9%), and Motley Fool (22%). Still, the majority of authors (56%) did not choose any existing platform but preferred a personal website or nothing at all. Clearly, SeekingAlpha provides a unique and valuable platform for its contributors.

**Biased Articles and Stock Manipulation.** SeekingAlpha's nature makes it possible for companies or individuals to profit by influencing users and manipulating stock prices. Some of these efforts were identified and publicly documented [Feuerstein 2014; Kimelman 2014]. Surprisingly, both authors and users (80%) stated that they have seen manipulative articles on SA (Figure 14(b)). More normal users reported frequently seeing manipulative articles. When asked about their response to manipulative articles (Q5), 42% (44 out of 104) stated that they would dispute the conclusions using comments on the article, 3% (three out of 104) would report to SA administrators, while 25% (26 out of 104) "do nothing." Ten percent (10 out of 104) did not see manipulation articles and thus the question is not applicable. Finally, 21 users chose other, and 20 users added further detailed comments to explain: 15 respondents stated biased articles are to be expected and ignored; two people said they would blacklist the author and never read his or her articles again; and two stated they could still extract value from dissenting comments on the article. Finally, one user stated that it was very difficult to distinguish between real stock manipulations from authors and strong personal preferences.

**Impact.** We ask if authors believe their articles have the power to impact stock prices (Q6). More than 62% of authors believed this was true, 14% said no, and the rest were unsure. One author commented that SeekingAlpha articles typically cannot impact large cap stocks like Apple (AAPL) but could affect small and micro cap stocks.

Finally, we also received unsolicited anecdotal feedback from past authors on additional types of author misbehavior. They identified some authors who generated large volumes of articles solely for the purpose of soliciting users to subscribe to their investment website memberships. They also identified others who produced detailed articles at extremely high volumes, too high to be produced even by a full-time professional. The assertion is that these accounts are managed by teams of writers who generate numerous articles to increase their reputation and ranking, only to then profit by manipulating one or two stocks they own in large volume. While we cannot confirm or disprove these assertions, they are consistent with other survey responses and could account for the apparent disconnect between poor average correlation of articles and high correlation of articles by top authors.

## 10. ANALYZING POTENTIAL STOCK MANIPULATION

Our survey indicates that stock manipulation articles have aroused wide concern among SeekingAlpha users. Today, SeekingAlpha is taking active steps to detect and remove manipulation articles from the site.<sup>8</sup> If detected, the content of such articles is replaced with text similar to "This author's articles have been removed from Seeking

<sup>8</sup><http://seekingalpha.com/article/2112703-what-seeking-alpha-is-doing-to-prevent-paid-stock-promotion>.

Table III. Dataset of Ground-Truth Manipulation Authors and Their Articles

Author	Total Articles	SA Deleted Articles (# Pro.)	404 Error	Still Available
john-mylant	881	880 (3)	1	0
stock-whisper	99	99 (1)	0	0
equity-options-guru	58	2 (0)	56	0
tech-guru	55	54 (9)	1	0
zorro-trades	47	47 (16)	0	0
ryan-s-view	15	15 (0)	0	0
kingmaker	8	1 (0)	7	0
john-eastman	45	8 (0)	1	36
andreas-spiro	27	7 (0)	2	18
alberto-savrieno	32	5 (0)	2	25
austrolib	15	4 (0)	0	11
wonderful-wizard	8	1 (0)	6	1

Alpha due to a Terms of Use violation.” We use this text as an indicator of a manipulative article that has been detected and removed by SeekingAlpha. To quantitatively understand manipulation behaviors, we collect a set of ground-truth articles that were detected by SeekingAlpha in the past. Then we analyze authors of these manipulation articles and explore potential features to distinguish them from benign authors.

**Ground-Truth Dataset.** To gather a set of ground-truth manipulation articles, in November 2014, we revisited the links of the 410K articles in our dataset (initially crawled in April 2014) and found that 1,654 articles were deleted. Among the deleted articles, 1,123 returned the message *“This author’s articles have been removed from Seeking Alpha due to a Terms of Use violation,”* and we treat them as ground-truth manipulation articles. The other 531 articles returned *“HTTP 404 error,”* which means the page does not exist any more. It is possible that these articles were deleted by the authors themselves. Since we cannot confirm whether they are truly manipulations, we exclude them from the ground-truth set. In total, there are 12 authors responsible for these 1,123 manipulation articles.<sup>9</sup>

Table III shows a detailed breakdown of the manipulation authors’ articles. First, we notice the top author john-mylant has 880 articles deleted, which accounts for more than 80% of the deleted articles by SeekingAlpha. Second, not all the articles of these authors got deleted; for example, the bottom five authors in Table III still have articles available on the site. Finally, we find some manipulation articles even got into the *Pro.* category, that is, high-quality articles hand-picked by SA editors. Particularly, zorro-trades has 16 *Pro.* articles and tech-guru has nine. This suggests manipulation articles can be highly deceptive, even to professional investors and editors.

**Characterizing Manipulation Authors.** We now analyze potential features to distinguish manipulation authors. These features are mostly derived from the insights gained in our user survey (Section 9), namely, the temporal burstiness of posting articles, the stocks they focused on, and comments.

First, we expect manipulation authors to post many articles in a short time to sway stock prices. However, only a few manipulation authors in our dataset do so. Figure 15(a) shows three authors got into the 90th percentile of the maximum number of articles in a day: john-mylant once posted seven articles a day (six for stock-whisper

<sup>9</sup>SA never reveals the technical details on how they detected manipulation articles. Anecdotes indicate that these manipulation articles were revealed by SA authors who infiltrated the companies that run stock promotion campaigns. <http://seekingalpha.com/article/2086173-behind-the-scenes-with-dream-team-cytrx-and-galena>.

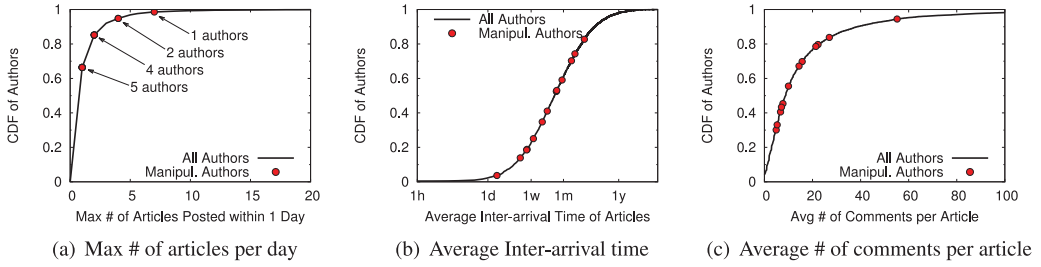


Fig. 15. Comparing manipulation authors (dots) with the overall distribution of all authors (line) in SeekingAlpha.

and equity-options-guru). All other manipulation authors post articles in a normal frequency (fewer than five articles per day). Also, their average interarrival time of articles follows a similar distribution as all authors (Figure 15(a)).

Second, most manipulation authors focus on stocks with low market cap. The exception is *john-mylant*, as 95% of his articles are about large companies with a more than \$2 billion market cap. For the other 11 authors, 69% of their articles focus on stocks less than \$2 billion (compared with 24% for all authors in Figure 5). Intuitively, the price of a small cap stock is easier to manipulate. However, this is not a deterministic feature to flag manipulation authors.

Finally, there’s no obvious difference in the number of received comments between manipulation authors and the rest (Figure 15(c)). The manipulation author *zorror-trades* received the most comments, who is also the one with most *Pro*. articles.

While our ground-truth data is likely incomplete, our results still suggest that manipulation authors vary significantly in behavior and often act similarly to normal authors. Using simple behavioral features alone, it is difficult to deterministically distinguish manipulation authors from others. In the next section, we explore using more sophisticated approaches (stylometric analysis) to detect Sybils and potential manipulations.

## 11. SYBIL IDENTITIES AND AUTHOR MISREPRESENTATION

Stock analysis is a challenging task where “success” is defined as any small relative improvement over random chance. Experts in the financial industry (sell-side analysts, hedge fund managers, media analysts) routinely make wrong predictions, often with severe consequences for their clients, followers, and themselves. All this makes manipulation of stock prices very difficult to determine on a forum like SeekingAlpha. Incorrect predictions can be the result of intentional misrepresentation by authors, or simply the result of an unexpected or random event. Proving ground truth would require definitive proof that the author’s actions produce incorrect results for self-gain (i.e., the author’s financial transactions that directly contradict his or her online statements). Even then, proving intent is difficult.

While proving direct manipulation is impractical, it is much easier to detect efforts inside SeekingAlpha to amplify or misrepresent authorship. First, recent manipulation “pump and dump” campaigns have been found on SeekingAlpha [Feuerstein 2014; Kimelman 2014]. Typically, these manipulation campaigns require multiple author accounts (*Sybils*) to write biased articles (or opinions) on certain stocks, seeking to cause an intended price change. Having multiple accounts working together can amplify the credibility of the “opinion” while reducing the suspicion cast on individual authors. This is similar to “astroturfing,” information dissemination campaigns that are sponsored by an organization but obfuscated so as to appear like spontaneous, decentralized

grass-root movements. Second, groups of manipulators can hide behind an online pseudonym, concentrating their efforts to boost followers and visibility while generating significant content under a single author name.

Evidence of either type of authorship misrepresentation can be detected by identifying fake accounts (i.e., Sybil accounts). In this section, we explore the use of *stylometry* (writing style) analysis to identify Sybil accounts controlled by the same attacker. The intuition is: when an attacker publishes multiple articles through different author accounts, these accounts would exhibit unusually high levels of stylometric similarity. Here, we first introduce background on stylometry analysis and how to quantify the similarity of author writing styles. Then we take a hybrid of supervised and unsupervised machine-learning approaches to capture author pairs (or groups) with highly similar writing styles and detect Sybil accounts among them.

### 11.1. Background: Stylometry

Linguistic stylometry (i.e., writing styles) is often considered unique for each individual person. As a result, stylometry is widely used to attribute authorship to anonymous or disputed content (e.g., a threatening letter), de-anonymize Internet users, or even detect frauds and deceptions [Narayanan et al. 2012; Afroz et al. 2012]. In the context of detecting stock manipulations, stylometry can be a promising tool to detect Sybil authors. Intuitively, when an attacker uses Sybils (multiple author accounts under his or her control) to publish manipulation articles, these Sybils will show unusually high similarity on their writing styles. Thus, by matching the writing style of different accounts, we can capture colluded Sybils among them. We seek to validate this intuition on SeekingAlpha data.

Next, we examine different stylometry techniques and select the suitable ones for our scenario. In general, stylometry techniques identify authorship of anonymous text based on the text of known authors. First, we don't consider techniques that are customized for short texts such as Tweets [Bhargava et al. 2013] and online messages [Zheng et al. 2006], since SA articles are sufficiently long. Second, we do not employ a "bag of words" feature or any other features (e.g., ngrams of characters) that are strongly associated to the specific context/topic under discussion [Afroz et al. 2012; George and Argiri 2007; Sohn et al. 2009]. The reason is that SA authors typically write about different stocks and companies, and we need to accurately attribute authorship regardless of which stocks (topics) these articles cover.

Our approach follows the method in Narayanan et al. [2012], and features are mostly linguistic and stylistic features that are independent from specific key words or topics. These features can capture more subtle writing differences among authors. For example, we have 769 features of syntactic pairs. These features model the sentence structures an author tends to use. We also count the frequency of words as features. However, we only consider the function words like "the," "of," and "then," which are common words that anyone might use. In total, we use 1,168 features, as listed in Table IV. With these features, we can build stylometry models for each SA author and detect Sybils with the same writing styles.

### 11.2. Sybil Detection Methodology

Now we apply stylometry analysis to detect Sybil identities in SeekingAlpha. Our methodology is intuitive: we assume articles written by the same author have a fairly unique writing style and can be differentiated from other authors. So with stylometry features (Table IV), we build a dedicated classifier for *each author account* and try to find other accounts with similar writing styles (i.e., Sybils). For account  $a$ , we train a classifier (e.g., SVM) using a training dataset of an author's own articles (label  $A$ ) and a random sample of articles from other accounts (label  $B$ ). Then we apply  $a$ 's classifier on



Table IV. Features to Characterize Author Stylometry

Category (# Features)	Description
Length (2)	Number of words/characters in post
Word shape (5)	Frequency of words with different combinations of upper- and lowercase letters
Digits (10)	Frequency of 0 to 9
Punctuation (11)	Frequency of . ? ! , ; : ( ) " ' -
Vocabulary richness (11)	Yule's $K$ , <sup>10</sup> frequency of <i>hapax legomena</i> , <i>dis legomena</i> , etc.
Word length (20)	Frequency of words that have 1–20 characters
Special characters (21)	Frequency of special characters # @ ` % ^ \$ & _ + = * \   / < > [ ] { }
Letters (26)	Frequency of a to z, ignoring case
Function words (293)	Frequency of words like “the,” “of,” and “then”
Syntactic pairs (769)	Frequency of every pair (A,B), where A is the parent of B in the parse tree <sup>11</sup>

all SA authors' articles to detect those with similar writing styles. If certain account  $x$  has sufficient articles classified as label  $A$  (e.g.,  $>50\%$ ), it indicates  $a$  and  $x$  have similar writing styles, and we mark this account pair  $(a, x)$  as a Sybil pair. To find all potential Sybil pairs, we iterate through every author to run this process.

We realize a problem in applying this method on real data: if we treat each article as a data point, not all the accounts have sufficient data points (e.g., 50% of authors have fewer than five articles) to train a classifier. So instead, we slice an article to several text blocks and treat each block as a data point, thus generating more data points per author. This is doable because SA articles are typically multiple pages long. We will examine the appropriate length for the text block shortly.

**Ground-Truth Experiments.** We first perform experiments to examine the effectiveness (accuracy) of our Sybil detection method and determine the right set of parameters. In order to compute detection accuracy, we need ground-truth data that contains labeled Sybil accounts. Note that directly labeling Sybil accounts (e.g., via manual examination) is extremely difficult. Instead, we construct a small ground-truth dataset as follows: First, we randomly select 50 accounts, split each account's articles into two even parts, and treat each part as a new account. These two newly generated accounts are essentially Sybil accounts, since their articles are written by the same author. In total, we generate 100 Sybil accounts (50 Sybil pairs). Then we select another random 100 accounts as the legitimate accounts (non-Sybils). This serves as our experiment dataset. We understand this ground-truth data is still not perfect: there could be a small chance that the 100 random legitimate accounts still contain Sybil pairs. But given the probability of such unlabeled Sybil pairs is low, this data should provide a good estimation on detection accuracy (lower bound).

There are three key parameters (or configurations) in our method. First is the choice of machine-learning algorithm to build the stylometry model. Second, when slicing an article into text blocks, we need to find the minimal length for the text block ( $L$ ) to compute stylometry features. Third, we need a threshold ( $P$ ) to determine whether two authors are a Sybil pair. Here  $P$  is the minimal percentage of one author's articles to be classified as the other author's.

First, we test several widely used classifiers to build stylometry models: Supported Vector Machine (SVM), C4.5 Decision Tree (Tree), and Bayesian Network (BN). In this experiment, we set  $L = 500$  and  $P = 0.5$  and compute two types of errors: false

<sup>10</sup>Yule's  $K$  measures the diversity of vocabulary. We compute Yule's  $K$  in the same way of Abbasi and Chen [2008].

<sup>11</sup>Parse tree [Klein and Manning 2003] is a tree representation of the syntactic structure of sentences, where leaf nodes are words or punctuation, and other nodes represent various syntactic categories (e.g., prepositional phrase, personal pronoun).

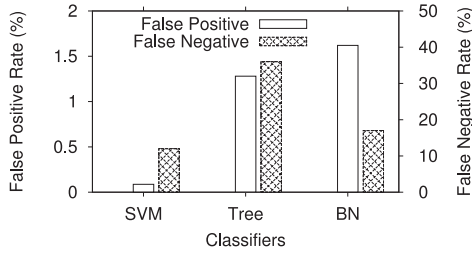


Fig. 16. Impact of classifiers on detection error rates ( $P = 0.5, L = 500$ ).

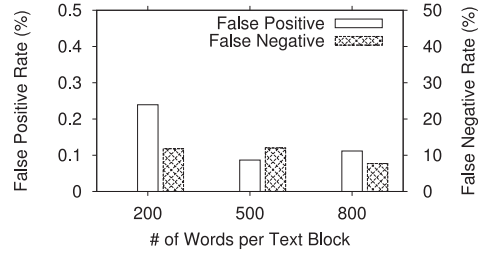


Fig. 17. Impact of the size of text block ( $L$ ) on detection error rates ( $P = 0.5$ ).

positives (classifying non-Sybil pairs as Sybil pairs) and false negatives (classifying Sybil pairs as non-Sybil pairs). As shown in Figure 16, SVM produces much smaller errors in detecting Sybil pairs compared with Decision Tree and Bayesian Network. For example, the false-positive rate is only 0.1% for SVM, while the corresponding values for Decision Tree and Bayesian Network are more than 1%. So we choose SVM as our classifier.

Then we examine how the text length  $L$  affects detection accuracy (with  $P$  set as 0.5). As shown in Figure 17, our approach can accurately detect Sybil pairs: with  $L = 500$ , the false-positive rate is 0.1% with a false-negative rate of 12%. Second, not too surprisingly, longer text (500 and 800 words) produces fewer errors than results using shorter text (200 words), since short text may not sufficiently capture an author's stylometry. Finally, we get comparable error rates with  $L = 500$  and  $L = 800$ . But considering that smaller  $L$  generates more data points for authors, we set  $L = 500$  (a moderate value) for the rest of the analysis.

Next, we tune threshold  $P$  to examine the tradeoffs between false-positive and false-negative rates. As shown in Figure 18, a more restrictive threshold (larger  $P$ ) reduces false positives but also increases the chance of missing true Sybil pairs. If we set  $P = 0.35$  (a relatively balanced value), we get a false-positive rate of 0.17% and a false-negative rate of 4.0%. In practice, how to configure  $P$  depends on the specific application scenario. For example, if SeekingAlpha uses this as a prefiltering system (before more in-depth investigations), it should flag as many Sybils as possible, while tolerating some false positives. But if SeekingAlpha relies on the detection result to issue direct penalties to Sybils, then it is crucial to keep the false-positive rate low.

Note that our ground-truth data is highly imbalanced, where the number of Sybil pairs (50) is much smaller than non-Sybil pairs (19,850). With such a dataset, even when the false-positive rate is low, the absolute number of false positives can still impact the efficiency of locating Sybils. For example, in our experiment, a 0.1% false-positive rate will introduce 20 benign pairs labeled as Sybil pairs. Thus, further investigation of detected pairs is usually needed. The key benefit of our method is that it dramatically reduces the scope of suspicious candidates for further investigation.<sup>12</sup>

### 11.3. Detecting Sybil Identities in the Wild

Given the good performance on ground-truth data, we now apply our method to the whole dataset of SA authors to detect Sybil accounts. We follow the parameter setting  $L = 500$  as discussed earlier. For  $P$ , we pick a conservative value as 0.5. In total, our method detected 67,236 suspicious Sybil pairs out of 31 million account pairs examined.

<sup>12</sup>SeekingAlpha can leverage the rich server-side data for in-depth investigation such as authors' IPs, phone number, and even credit card information. This information is acquired upon registering as an author.

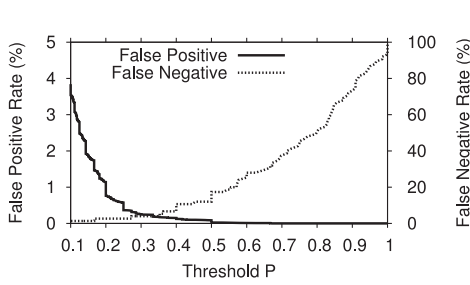


Fig. 18. Impact of the parameter  $P$  on detection error rates ( $L = 500$ ).

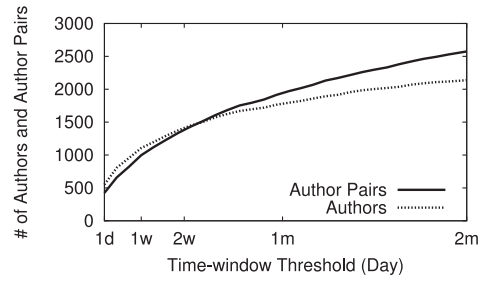


Fig. 19. Time-window threshold versus the number of author pairs left in the candidate set.

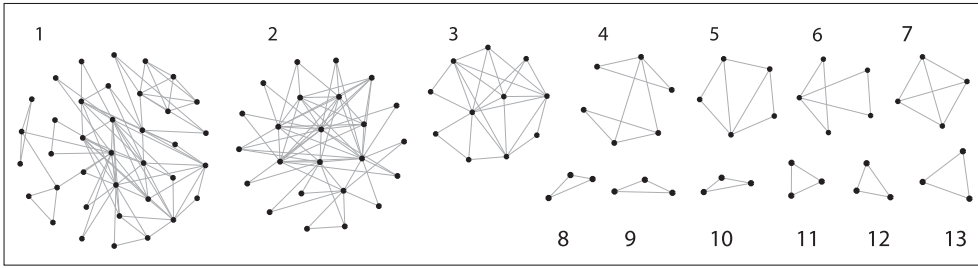


Fig. 20. Detected Sybil clusters.

Clearly, the result still contains false positives, and we need to further trim down the number of suspicious pairs. Here, we apply three filters based on key intuitions of Sybil behaviors. In practice, this step is very flexible and can be combined with many other features and criteria. For example, Seeking Alpha can also check with server-side information (e.g., if the accounts were operated with the same IP address or proxy).

*First*, since the purpose of Sybils is to collaboratively promote/demote stocks, Sybil accounts should have overlaps on the stocks they write about. So we exclude account pairs with zero overlap on stocks (11,883 pairs left). *Second*, to cause meaningful changes on stock prices (e.g., to perform “pump and dump”), Sybils need to post articles within a close time period. So we also exclude account pairs that have no articles (on the same stock) collocated within any time window  $T$ . With different  $T$ , Figure 19 shows the number of pairs left in our candidate set. We conservatively pick  $T$  as 2 weeks, which gives us 1,394 suspicious pairs. *Third*, we put all these pairs into a graph (authors are nodes, detected pairs are edges) and examine whether there are Sybil groups. We find many connected components in the graph are only loosely connected inside, which indicates that false-positive edges still exist. For a real Sybil group, the intraconnectivity should be high, since members in the group all share similar writing styles. To this end, we apply a simple filter by detecting triangles in the graph and removing those that don’t form triangles. Finally, we obtain 228 suspicious pairs forming 13 groups, as shown in Figure 20.

These filtering steps trim down the number of suspicious candidates so that we can perform further (manual) investigation. The drawback is we cannot estimate how many real Sybils are missed due to the filtering. In practice, it’s possible for SeekingAlpha to deploy more sophisticated filters with server-side data (e.g., author IP, user login logs).

**Examining Sybil Clusters.** Now that we have 13 groups of suspicious accounts, we examine their actions and explore how likely they are Sybil identities. We first

cross-compare our detected authors with the set of manipulation authors already flagged by SeekingAlpha (Table III). We find the four authors in cluster 7 (ryan-s-view, andreas-spiro, alberto-savrieno, austrolib) matched in the manipulation author set. These four authors have been suspected to participate in the promotion of Northwest Biotech (\$NWBO) together [Pearson 2014] and are very likely to be Sybils.

Then we examine cluster 1, which is the largest Sybil group with 37 authors. Interestingly, five authors in the group have the same last names (anup-singh, renu-singh, amal-singh, ayush-singh, vinay-kumar-singh), and all five authors started to post articles around September 2013. These five authors are highly likely Sybils. The rest of the authors in the group do not use the same last name, but 27 of them joined the site around the same time.

Clusters 2, 3, 6, 9, and 11 contain accounts that focus on the same set of stocks and post articles synchronously. These accounts are likely to be created to run manipulation campaigns (then never active again). For example, in cluster 6, three of the five authors only talk about a single stock, \$HLF, and the other two authors have more than half of their articles on \$HLF. It's worth noting that \$HLF (Herbalife) is an extremely controversial company often accused of being a sophisticated pyramid scheme. It is also public knowledge that very active hedge fund managers have been on both sides of the Herbalife battle, either owning (and praising) the stock or shorting (and publicly attacking) the stock. Thus, a manipulation scam targeting \$HLF would be unsurprising. Accounts in cluster 2 have a similar active period (first article in February 2012 and last article in December 2012).

We also find benign Sybil clusters: clusters 8 and 12 contain a group of SeekingAlpha editor accounts who host an article series called "wall-street-breakfast." This series of articles contain daily summaries on business and financial news. Since it is the same set of editors that are operating those accounts, it makes sense that they have similar writing styles. Even though the two clusters are benign Sybils, they confirmed the effectiveness of our methodology for Sybil detection.

For the remaining four clusters, we find certain suspicious signals in their actions but cannot determine whether they are Sybils. For example, authors in cluster 4 have similar interests in biotech and healthcare companies, but the specific stocks have few overlaps. All five authors in cluster 5 are financial writers, and one of them was once found writing stock promotion articles on Forbes.com.<sup>13</sup>

**Limitations and Discussion.** One limitation of our method is scalability, since we need to train a classifier per author and apply it to all other authors ( $O(n^2)$ ). But our method fits SeekingAlpha well—as an expert-based crowdsourcing platform, the contributor population is only a small portion of the total user base (0.2%), which is within our capacity. For example, to run a full test on 8K SA authors, it takes 2 days using seven Quad Core servers. Also, we can still potentially speed up this process (e.g., by only testing two accounts if they have overlaps on stocks).

Our result suggests stylometry is a promising tool to detect Sybil accounts. In a broader scope, Sybils pose a unique challenge to crowdsourcing systems like SeekingAlpha, where a small number of experts play a major role. In these systems, it is not a guarantee that honest users can always outnumber attackers (Sybils) on a specific domain (or topic). For example, in SeekingAlpha, many stocks have few articles on them and can be easily manipulated by Sybils. With several biased articles, attackers can make their opinions dominating. As a future work, we explore generalizing our Sybil detection method to other expert-based crowdsourcing platforms.

<sup>13</sup><http://www.goodetrades.com/2013/06/seekingalpha-and-other-websites-being-co-opted-by-paid-stock-promoters/>.

## 12. RELATED WORK

**Stock Market Prediction.** Stock market prediction via data mining has been explored in a variety of contexts, including Google trends [Preis et al. 2013], Wikipedia [Moat et al. 2013], online blogs [De Choudhury et al. 2008; Gilbert and Karahalios 2010], financial news [Fung et al. 2003; Schumaker and Chen 2009], and social content from Twitter [Bollen et al. 2011; Brown 2012; Rao and Srivastava 2012; Sprenger et al. 2013] and Facebook [Karabulut 2011]. Some have studied stock-centered social networks, that is, StockTwits [Oh and Sheng 2011; Oliveira et al. 2013] and SeekingAlpha [Chen et al. 2014]. Most of these draw conclusions based on short-term data of less than a year, despite the highly cyclical nature of bull and bear markets that lasts multiple years. In contrast, our data covers up to 9 years, long enough to cover both crashes (2008–2009) and strong bull markets (2013). In addition, our comparative analysis between StockTwits and SeekingAlpha helps us understand the impact of experts as contributors versus average users.

**Sentiment-Based Investing and Analysis.** There are a few works on the sentiment-based investment strategy. Most papers focus on the prediction of and investment in several selected stocks [Oliveira et al. 2013]. Other work invests in all possible stocks in the market [Makrehchi et al. 2013]. Prior works [Bar-Haim et al. 2011; Liao et al. 2014] use small datasets from Twitter/StockTwits to quantify the noisiness in overall sentiment and to motivate the need for contributions from experts. This is consistent with our results that show the best performance is achieved from identifying and relying on top experts.

Existing sentiment extraction methods vary widely, from dictionary-based methods [Esuli and Sebastiani 2007; Godbole et al. 2007; Loughran and McDonald 2011] to supervised machine-learning algorithms [Liu 2012; Pang et al. 2002; Barbosa and Feng 2010]. Dictionary-based methods require domain-specific dictionaries [Loughran and McDonald 2011] and do not work well with short texts like Tweets [Barbosa and Feng 2010] and reviews [Pang et al. 2002]. Researchers have applied supervised machine-learning algorithms to sentiment classification, with common features such as term frequency, parts of speech, and negations [Liu 2012].

**Crowd-Based Stock Prediction.** Estimote<sup>14</sup> is an open financial platform that aggregates estimates of company earnings reports from the opinions of independent, buy- and sell-side analysts and private investors. Estimote has contributions from 4,628 analysts covering over 900 stocks. Note that Estimote focuses on predictions of quarterly earnings results, not stock performance.

**Crowdsourcing and Collaborative Work.** Researchers have studied crowdsourcing and collaborative works under contexts ranging from translation and labeling [Sun and Dance 2012], to answering questions [Nichols et al. 2013; Wang et al. 2013a; Tausczik et al. 2014], to gathering opinions [Xu et al. 2014], and even to performing creative collaboration tasks [Kim et al. 2014]. One of the key challenges in collaborative work is the uneven expertise and low quality of contributions from large crowds [Kittur et al. 2013]. Recent works have explored different directions to design better crowdsourcing systems, including quality control on contributors and content [Sun and Dance 2012; Otterbacher 2009], identifying domain experts [Jacovi et al. 2014; Liao et al. 2012], and applying weighted methods to integrate crowd results [Forlines et al. 2014]. In the application context of collaborative investment analysis, our work demonstrates that more “skilled” experts can be identified in collaborative crowds through interactions

<sup>14</sup><http://www.estimize.com>.



with the broader crowd. Even crowds that cannot generate valuable content can identify those that can.

Finally, crowdsourcing platforms are making a significant impact on the security of online communities. While they can be used as effective tools for detecting fake accounts on online social networks [Wang et al. 2013c], they can also provide willing labor for performing tasks to bypass traditional security mechanisms [Wang et al. 2012]. While recent works are making progress against these malicious crowdsourcing systems [Wang et al. 2014], they remain one of the biggest challenges facing online systems today.

**Sybil Detection.** Researchers have devoted significant efforts to detect Sybils in web services like online social networks [Gao et al. 2010; Yang et al. 2011; Wang et al. 2013b], microblogs [Wang 2010; Benevenuto et al. 2010], and review sites [Mukherjee et al. 2013; Akoglu et al. 2013]. Early works [Yu et al. 2006; Viswanath et al. 2010] assume there are rare links between normal users and Sybils. So they use graph-based methods to identify clusters of Sybils. However, this assumption is not always true [Gao et al. 2010]. Some other works [Wang et al. 2013b; Benevenuto et al. 2010; Wang et al. 2014] make use of supervised machine-learning tools. The key idea is to identify abnormal behaviors/features of Sybil accounts from normal users. Our work also leverages the tools of supervised machine learning. Different from prior work, we focus on stylometry and aim to identify if articles are written by the same author.

### 13. CONCLUSIONS AND FUTURE WORK

In this article, we analyze the correlation between stock performance and user-contributed content sentiment over a period of 4 to 9 years. Our analysis shows that while expert-contributed stock analysis in SeekingAlpha provides more positive correlation than user-generated content from StockTwits, the correlation is very weak. We show that valuable content can be extracted using well-designed filters based on user comments and can lead to strategies that significantly outperform the broader stock market. We also take the initial step to the identification of manipulative behaviors by detecting Sybil accounts on SeekingAlpha. As future work, we plan to develop better algorithms to detect manipulation content. We also want to explore how to extend such data mining approaches to other platforms.

Our work suggests that even complex, domain-specific tasks such as stock analysis can benefit from collaboratively generated content. Even when only a small portion of users is capable of generating valuable content, others in the crowd help by indirectly identifying the valuable content through their interactions. This suggests a possible design for “meta-reputation” systems that highlight quality content by tracking user interactions in the crowd.

### ACKNOWLEDGMENTS

We would like to thank Chris Corriveau for providing access to the StockTwits dataset, and the editors and reviewers for their helpful comments.

### REFERENCES

- Ahmed Abbasi and Hsinchun Chen. 2008. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems (TOIS)* 26, 2 (2008), 7.
- Lada A. Adamic, Jun Zhang, Eytan Bakshy, and Mark S. Ackerman. 2008. Knowledge sharing and yahoo answers: Everyone knows something. In *Proc. of World Wide Web (WWW)*.
- Sadia Afroz, Michael Brennan, and Rachel Greenstadt. 2012. Detecting hoaxes, frauds, and deception in writing style online. In *Proc. of IEEE S&P*.

- Leman Akoglu, Rishi Chandy, and Christos Faloutsos. 2013. Opinion fraud detection in online reviews by network effects. In *Proc. of ICWSM*.
- Roy Bar-Haim, Elad Dinur, Ronen Feldman, Moshe Fresko, and Guy Goldstein. 2011. Identifying and following expert investors in stock microblogs. In *Proc. of EMNLP*.
- Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on Twitter from biased and noisy data. In *Proc. of COLING*.
- Fab rio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida. 2010. Detecting spammers on Twitter. In *Proc. of CEAS*.
- Mudit Bhargava, Pulkit Mehndiratta, and Krishna Asawa. 2013. Stylometric analysis for authorship attribution on Twitter. In *Proc. of International Conference on Big Data Analytics*. 37–47.
- Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2, 1 (2011), 1–8.
- Eric D. Brown. 2012. Will Twitter make you a better investor? A look at sentiment, user reputation and their effect on the stock market. In *Proc. of SAIS*.
- Hailiang Chen, Prabhuddha De, J. Hu, and Byoung-Hyoun Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies* 27, 5 (2014), 1367–1403.
- Munmun De Choudhury and others. 2008. Can blog communication dynamics be correlated with stock market activity? In *Proc. of HyperText*.
- Douglas W. Diamond and Robert E. Verrecchia. 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics* 18, 2 (1987), 277–311.
- Andrea Esuli and Fabrizio Sebastiani. 2007. Pageranking wordnet synsets: An application to opinion mining. In *Proc. of ACL*.
- Adam Feuerstein. 2014. Galena Biopharma Pays For Stock-Touting Campaign While Insiders Cash Out Millions. *TheStreet News*. (February 2014).
- Clifton Forlines, Sarah Miller, Leslie Guelcher, and Robert Bruzzi. 2014. Crowdsourcing the future: Predictions made with a social network. In *Proc. of CHI*.
- Gabriel Pui Cheong Fung, Jeffrey Xu Yu, and Wai Lam. 2003. Stock prediction: Integrating text mining approach using real-time news. In *Proc. of CIPHER*.
- Hongyu Gao and others. 2010. Detecting and characterizing social spam campaigns. In *Proc. of IMC*.
- Mikros K. George and Eleni K. Argiri. 2007. Investigating topic influence in authorship attribution. In *Proc. of PAN*.
- Eric Gilbert and Karrie Karahalios. 2010. Widespread worry and the stock market. In *Proc. of ICWSM*.
- Namrata Godbole, Manja Srinivasaiyah, and Steven Skiena. 2007. Large-scale sentiment analysis for news and blogs. In *Proc. of ICWSM*.
- Pollyanna Gonalves, Matheus Ara jo, Fab rio Benevenuto, and Meeyoung Cha. 2013. Comparing and combining sentiment analysis methods. In *Proc. of COSN*.
- F. Maxwell Harper, Daphne Raban, Sheizaf Rafaeli, and Joseph A. Konstan. 2008. Predictors of answer quality in online Q&A sites. In *Proc. of CHI*.
- Investment 2013. *2013 Investment Company Fact Book*. Technical Report. Investment Company Institute. Retrieved from [http://www.ici.org/pdf/2013\\_factbook.pdf](http://www.ici.org/pdf/2013_factbook.pdf).
- Michal Jacovi, Ido Guy, Shiri Kremer-Davidson, Sara Porat, and Netta Aizenbud-Reshef. 2014. The perception of others: Inferring reputation from social media in the enterprise. In *Proc. of CSCW*.
- Yigitcan Karabulut. 2011. Can Facebook predict stock market activity? *SSRN eLibrary* (2011).
- Joy Kim, Justin Cheng, and Michael S Bernstein. 2014. Ensemble: Exploring complementary strengths of leaders and crowds in creative collaboration. In *Proc. of CSCW*.
- John Kimelman. 2014. An insider’s tale of a stock promotion plan. *Barrons News*. (March 2014).
- Saijel Kishan and Kelly Bit. 2013. Hedge funds trail stocks by the widest margin since 2005. *Bloomberg News*. (December 2013).
- Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proc. of CSCW*.
- Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proc. of ACL*.
- Q. Vera Liao, Claudia Wagner, Peter Piroli, and Wai-Tat Fu. 2012. Understanding experts’ and novices’ expertise judgment of Twitter users. In *Proc. of CHI*.
- Wenhui Liao, Sameena Shah, and Masoud Makrehchi. 2014. Winning by following the winners: Mining the behaviour of stock market experts in social media. In *SBP*. 103–110.

- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* 5, 1 (2012), 1–167.
- Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66, 1 (2011), 35–65.
- Masoud Makrehchi, Sameena Shah, and Wenhui Liao. 2013. Stock prediction using event-based sentiment analysis. In *Proc. of WI-IAT*.
- Helen Susannah Moat, Chester Curme, Adam Avakian, Dror Y. Kenett, H. Eugene Stanley, and Tobias Preis. 2013. Quantifying wikipedia usage patterns before stock market moves. *Scientific Reports* 3 (2013).
- Arjun Mukherjee, Abhinav Kumar, Bing Liu, Junhui Wang, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013. Spotting opinion spammers using behavioral footprints. In *Proc. of SIGKDD*. ACM, 632–640.
- Arvind Narayanan, Hristo Paskov, Neil Zhenqiang Gong, John Bethencourt, Emil Stefanov, Eui Chul Richard Shin, and Dawn Song. 2012. On the feasibility of internet-scale author identification. In *Proc. of IEEE S&P*.
- Jeffrey Nichols, Michelle Zhou, Huahai Yang, Jeon-Hyung Kang, and Xiao Hua Sun. 2013. Analyzing the quality of information solicited from targeted strangers on social media. In *Proc. of CSCW*.
- Chong Oh and Olivia Sheng. 2011. Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. In *Proc. of ICIS*.
- Nuno Oliveira, Paulo Cortez, and Nelson Areal. 2013. On the predictability of stock market behavior using stockTwits sentiment and posting volume. In *Progress in AI*. 355–365.
- Judith S. Olson and Wendy A. Kellogg. 2014. *Ways of Knowing in HCI*. Springer.
- Jahna Otterbacher. 2009. “Helpfulness” in online communities: A measure of message quality. In *Proc. of CHI*.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval* 2, 1–2 (2008), 1–135.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: Sentiment classification using machine learning techniques. In *Proc. of ACL*.
- Karl Pearson. 1895. Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society A* 186 (1895), 343–414.
- Richard Pearson. 2014. Behind the promotion of Northwest Bio. *Seeking Alpha*. (July 2014).
- Tobias Preis, Helen Susannah Moat, and H. Eugene Stanley. 2013. Quantifying trading behavior in financial markets using google trends. *Scientific Reports* 3 (2013).
- Tushar Rao and Saket Srivastava. 2012. Analyzing stock market movements using Twitter sentiment analysis. In *Proc. of ASONAM*.
- Robert P. Schumaker and Hsinchun Chen. 2009. Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems (TOIS)* 27, 2 (2009), 12.
- SeekingAlpha. 2014. About SeekingAlpha. [http://seekingalpha.com/page/about\\_us](http://seekingalpha.com/page/about_us).
- Victor S. Sheng, Foster Provost, and Panagiotis G. Ipeirotis. 2008. Get another label? Improving data quality and data mining using multiple, noisy labelers. In *Proc. of KDD*. 614–622.
- Dae-Neung Sohn, Jung-Tae Lee, and Hae-Chang Rim. 2009. The contribution of stylistic information to content-based mobile spam filtering. In *Proc. of the ACL*. 321–324.
- Timm O. Sprenger, Andranik Tumasjan, Philipp G. Sandner, and Isabell M. Welpe. 2013. Tweets and trades: The information content of stock microblogs. *European Financial Management* (2013).
- StockTwits. 2014. About StockTwits. <http://stocktwits.com/about>.
- Yu-An Sun and Christopher R. Dance. 2012. When majority voting fails: Comparing quality assurance methods for noisy human computation environment. In *Proc. of Collective Intelligence*.
- Yla R. Tausczik, Aniket Kittur, and Robert E. Kraut. 2014. Collaborative problem solving: A study of math-overflow. In *Proc. of CSCW*.
- Dylan Tweney. 2013. Seeking Alpha: Who needs an acquisition when were doing so well? *VentureBeat News*. (October 2013).
- Bimal Viswanath and others. 2010. An analysis of social network-based sybil defenses. In *Proc. of SIGCOMM*.
- Alex Hai Wang. 2010. Don’t follow me: Spam detection in Twitter. In *Proc. of SECURITY*.
- Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng, and Ben Y. Zhao. 2013a. Wisdom in the social crowd: An analysis of quora. In *Proc. of WWW*.
- Gang Wang, Tristan Konolige, Christo Wilson, Xiao Wang, Haitao Zheng, and Ben Y. Zhao. 2013b. You are how you click: Clickstream analysis for sybil detection. In *Proc. of USENIX Security*.

- Gang Wang, Manish Mohanlal, Christo Wilson, Xiao Wang, Miriam Metzger, Haitao Zheng, and Ben Y. Zhao. 2013c. Social turing tests: Crowdsourcing sybil detection. In *Proc. of NDSS*.
- Gang Wang, Tianyi Wang, Bolun Wang, Divya Sambasivan, Zengbin Zhang, Haitao Zheng, and Ben Y. Zhao. 2015. Crowds on wall street: Extracting value from collaborative investing platforms. In *Proc. of CSCW*.
- Gang Wang, Tianyi Wang, Haitao Zheng, and Ben Y. Zhao. 2014. Man vs. machine: Practical adversarial detection of malicious crowdsourcing workers. In *Proc. of USENIX Security*.
- Gang Wang, Christo Wilson, Xiaohan Zhao, Yibo Zhu, Manish Mohanlal, Haitao Zheng, and Ben Y. Zhao. 2012. Serf and turf: Crowdturfing for fun and profit. In *Proc. of WWW*.
- Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: Generating structured feedback on visual designs using a crowd of non-experts. In *Proc. of CSCW*.
- Yahoo. 2014. Yahoo Finance API. <https://code.google.com/p/yahoo-finance-managed/>.
- Zhi Yang and others. 2011. Uncovering social network sybils in the wild. In *Proc. of IMC*.
- Haifeng Yu, Michael Kaminsky, Phillip B. Gibbons, and Abraham Flaxman. 2006. SybilGuard: Defending against sybil attacks via social networks. In *Proc. of SIGCOMM*.
- Rong Zheng, Jiexun Li, Hsinchun Chen, and Zan Huang. 2006. A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the American Society for Information Science and Technology* 57, 3 (2006), 378–393.

Received March 2016; revised September 2016; accepted December 2016