

## Echo Chambers in Investment Discussion Boards

Shiliang Tang,<sup>†</sup> Qingyun Liu,<sup>†</sup> Megan McQueen,<sup>‡</sup>  
 Scott Counts,<sup>\*</sup> Apurv Jain,<sup>\*</sup> Haitao Zheng,<sup>†</sup> Ben Y. Zhao<sup>†</sup>

<sup>†</sup>UC Santa Barbara, <sup>‡</sup>Yale University, <sup>\*</sup>Microsoft Research  
 {shiliang\_tang, qingyun\_liu, htzheng, ravenben}@cs.ucsb.edu,  
 megan.mcqueen@yale.edu, {counts, apurvj}@microsoft.com

### Abstract

We examine the quality of information and communication in online investment discussion boards. We show that positivity bias and skewed risk/reward assessments, exacerbated by the insular nature of the community and its social structure, contribute to underperforming investment advice and unnecessary trading. Discussion post sentiment has negligible correlation with future stock market returns, but does have a positive correlation with trading volumes and volatility. Our trading simulations show that across different timeframes, this misinformation leads 50-70% of users to underperform the market average. We then examine social structure in communities, and show that the majority of market sentiment is produced by a small number of community leaders, and that many members actively resist negative sentiment, thus minimizing viewpoint diversity. To improve generated information content in online investment communities, we suggest designing to increase diversity of opinion, minimize friction around incorporating new information, and provide performance feedback for self-correction.

### Introduction

Investment discussion boards are popular online forums where communities of individual investors share and discuss their opinions on investments (commonly stocks) of mutual interest. These forums support millions of investors, and are often amongst the first results returned by search engines in response to queries for stock symbols. As such, they can be very influential for investors looking for financial advice.

But are these forums providing a useful service to their millions of users? What is the quality of discourse, and how effective are they as mechanisms of disseminating useful financial information? As platforms for social discourse where participation is self-selected, there is a distinct possibility of naturally forming *echo chambers* that can dramatically affect the quality of information exchanged on these forums. Echo chambers are discussion environments where ideas tend to be mutually reinforced by like-minded people, often contrary to external inputs (Sustein 2001), and have been found to evolve organically in a variety of contexts, including political forums, blogs, and social networks

such as Twitter (Gilbert, Bergstrom, and Karahalios 2009; Yardi and Boyd 2010; Bakshy, Messing, and Adamic 2015).

While there could be rational reasons to “herd” when the community knows something the individual does not, our results indicate this is not the case. Instead, by reinforcing existing biases and insulating users from objective facts, echo chambers are likely to induce poor investment decisions with potentially severe consequences. In the context of finance theory, echo chambers could play a role in producing *noise traders*, investors who make decisions driven by biases and emotions that push prices of investments away from their true value. In aggregate, noise traders can negatively impact markets by increasing price volatility and contributing to price bubbles (Bloomfield, O’hara, and Saar 2009; Kogan et al. 2006).

In this paper, we try to determine the presence of echo chambers in investment discussion boards using an empirical approach driven by data from two of the largest stock discussion boards, Yahoo Stock Message Boards and InvestorHub (iHub), that together claim hundreds of millions of registered users and tens of millions of visitors per month. We download message logs for discussion boards on both sites covering publicly listed stocks of US-based companies, totaling 34 million message posts across nearly 13,000 boards. Our work focuses on three questions:

- Does sentiment in stock discussion boards reflect an “echo chamber” effect, where sentiment decouples from external information, *e.g.* stock price fluctuations? To what extent, if any, are these effects correlated with level of user attention, participation, and bias in the forum?
- What is the internal structure of communication inside these boards, and what roles do individuals play in the depth and length of specific discussions? Is there any relationship between communication across specific groups and insulation of sentiment against external factors?
- What empirical impact do discussions on investment boards have on trading performance? Can investors rely on discussion sentiment to outperform stock market indices? Does investment board activity correlate with other properties of stock trading such as volume and volatility?

We summarize our contributions below.

*First*, we prune our data to identify message boards that focus on publicly traded companies listed on US exchanges.

We analyze gross message volume for both Yahoo and iHub, and show that the highest levels of user activity often reside on small market cap stocks with higher risk and volatility.

*Second*, using customized sentiment analysis tools, we extract aggregate sentiment out of daily discussions from each of the message boards in our analysis. We demonstrate that across different platforms, parameters, and metrics, there is near zero correlation between discussion sentiment and short-term price movements of a stock. Regardless of how a stock performed, sentiment in discussion boards is almost uniformly positive and upbeat.

To confirm the lack of value in board sentiment, we run detailed trading simulations to quantify the predictive quality of discussion board sentiment. We find that trading based on discussion board sentiment significantly underperforms market indices. While we cannot know to what extent discussion boards influence active (or passive) users, the general lack of diversity of opinions hurts the quality of the advice overall.

*Third*, we show that while board sentiment does not predict future stock performance, board activity as measured by the number of posts, does correlate with future stock trading volume and volatility. Importantly this correlation was strongest for stocks with the most attention by board members. Together with the lack of effect for board sentiment, this suggests that activity in the discussion boards signals misguided market activity and potentially is a source of market noise. These results are broadly consistent with financial economics literature with studies done on smaller samples (see (Das 2014) for a summary).

*Finally*, to better understand *why* these boards are poor sources of information, we explore their echo chamber-like qualities and show they are insular and actively reject disliked viewpoints. We use unsupervised learning techniques to identify clusters of users with similar communication patterns. We find that discussion threads on most boards might be controlled by a very small group of users, and negative sentiment drops significantly as threads are taken over by large groups of “reacting users and followers” whose posts reinforce each other and the overall sense of positivity in these boards. Using a small dictionary of keywords expressing personal animosity, we find that posts with negative stock sentiment tend to evoke hostile responses from the community, perhaps explaining why users who post the most negative content tend to be outside of the dominant user groups.

## Noise Traders and Echo Chambers

We begin with an introduction to the phenomenon of noise trading in the context of finance theory, and discuss the possible role that echo chambers play in financial markets.

*Noise Traders.* In classical finance theory, rational investors should only trade in ways consistent with the objective probabilities of the state of the economy. Increasingly, the financial literature recognizes that the average investor is not the perfectly rational investor of theory (Kogan et al. 2006). In this paper we investigate how the social mechanism of message boards may introduce noise and hence con-

tribute to the lack of useful economic information (trading signals) and potentially result in irrational or noise trading.

We thus refer to the trading actions of individuals based on noise as noise trading, consistent with literature that defines noise traders as traders having imperfect or irrational expectations (Black 1986; Shiller, Fischer, and Friedman 1984; Kyle 1985), some of which may arise from behavioral biases (Benartzi and Thaler 2001; Kahneman and Tversky 1979). The role noise traders play is a source of debate, where some (Black 1986) argue that they provide beneficial liquidity, others argue that they can exert influence and drive the prices away from economic fundamentals and render the market inefficient (Kogan et al. 2006). One of the insights from the efficient markets hypothesis (Malkiel and Fama 1970) is that new information should be unforecastable. Thus if message boards truly were places to collect valuable information, they should uncover opposing views with equal frequency. However, as we show, these boards appear to be influenced by social constructs that generate heavily biased information.

*Messages Boards as Echo Chambers.* Message boards and other online forums support discourse at scale, in both range of topics discussed and the number of people involved. Since participation is largely self-selected and often in the context of a social network, the emergence of echo chambers is a distinct possibility. Here we consider an echo chamber to be a social discourse environment in which ideas tend to be mutually reinforced by a like-minded group of people (Sustein 2001). For instance, in a study of blogs, agreement in comments outweighed disagreement by a ratio of three to one (Gilbert, Bergstrom, and Karahalios 2009). Further, it has been shown that people are more actively engaged in their online communities if they encounter less disagreement (Grevet, Terveen, and Gilbert 2014). More recently, Facebook has been examined for echo chamber effects, to mixed results. The Facebook newsfeed is slightly more aligned with an individual’s own ideology, but a notable minority of friends and news feed content represents opposing views (Bakshy, Messing, and Adamic 2015).

Given evidence of echo chambers online, what are the consequences? The primary concern is that a group of people becomes insulated from diverse perspectives. There is evidence from the political sphere, that online political forums can resemble echo chambers, in part because the massive number of online information sources allows people to gravitate to politically opinion-confirming Internet news (Sustein 2001; Garrett 2009). In addition to selective information exposure, there are network effects where online networks of those with more extreme views tend to be more homophilous (Boutyline and Willer 2016).

*Investments and Echo Chambers.* Tying this together, in a financial investment setting, the consequences of an echo chamber would be that investors would irrationally follow the investment advice and behavior of others in financial discussion boards insulated from external market news. In other words, echo chambers in these online forums could give rise to the aforementioned noise traders. For individuals, this could lead to direct financial losses through poor

Message Boards	# of Boards	# of Messages	# of Users	Data Since
iHub	6,839	5,329,547	56,421	03/25/2000
Yahoo	6,135	28,845,907	1,494,769	02/13/1996

Table 1: Basic statistics of crawled Yahoo and iHub boards.

Dataset	Exchange	# of Boards	# of Boards with Price
iHub	NASDAQ	2018	2014(99.8%)
	NYSE	1765	1757(99.5%)
	AMEX	214	213(99.5%)
	Other	2842	1471(51.8%)
Yahoo	NASDAQ	2608	2526(96.9%)
	NYSE	2404	2401(99.9%)
	AMEX	293	292(99.7%)

Table 2: Breakdown of both datasets in stock exchanges and stock price coverage. For Yahoo, we only consider boards with complete post history.

investments. For the market as a whole, this can lead to inefficiencies if enough people make poor investment decisions.

### Data Collection and Preliminary Analysis

Investors Hub (iHub) and Yahoo Message Boards are the most popular online financial forums for individual investors. These financial forums are built as platforms for individual investors to converse and share insights related to stock picks and trading strategies. Content is usually organized as individual message boards, each associated with a specific stock or fund.

*iHub.* iHub has been online since early 2000, and claims to have 562M registered users, and 122M messages on 25,000 boards<sup>1</sup>. Each iHub board is moderated for content, and registered users can start new message threads, reply to messages, and *follow* users to receive automatic updates.

In March 2016, we crawled all posts under iHub boards associated with stocks listed in US markets. Our dataset has a total of 5,329,547 messages posted by 56,421 users across 6,839 boards since March 25, 2000 (Table 1), with detailed timestamp associated to each message. Each iHub board is associated with a ticker symbol, which is an abbreviation used to uniquely identify a particular stock. Among them, 60% are valid stock symbols listed in the top 3 US stock exchanges in the U.S.: New York Exchange (NYSE), NASDAQ, and American Stock Exchange (AMEX)<sup>2</sup>. The remaining 40% of symbols include names of non-listed companies, delisted stocks, stocks on international exchanges, small stocks sold on Over-The-Counter markets (OTC), stock funds and ETFs, stock options, or user-created names. Table 2 shows the number of boards associated with stocks in each major exchange.

*Yahoo Message Boards.* Yahoo Message Boards are part of Yahoo! Finance, the most popular financial news website. Started in 1994, Yahoo! Finance claims 70M unique visitors each month. Yahoo Finance provides a webpage for each stock or fund, which links to its own message board. Message boards are created by default for each ticker symbol,

<sup>1</sup><http://investorshub.advfn.com/boards/about.aspx>.

<sup>2</sup>For stocks listed on multiple exchanges, we only count it once.

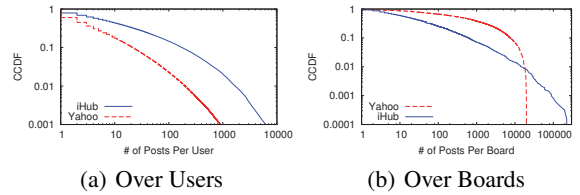


Figure 1: Distribution of posts over users and over message boards.

even for non-tradable, non-stock entities like corporate debt. Users can post messages and replies on boards, and all messages in Yahoo! Message boards are public.

We used a list of publicly traded stock symbols to look up message boards for US-based stocks. We downloaded a total of 6,135 message boards, 28,845,907 messages by 1,494,769 users since Feb. 13, 1996 (Table 1). Of these boards, we retrieved all posts from 5,305 (86%) of them. The remaining boards had more content than what Yahoo Finance displays (a maximum of 1002 display pages). For consistency, we restrict analysis and simulation in the rest of this paper to the 5,305 fully crawled message boards. Board coverage from Yahoo boards is also listed in Table 2.

Finally, we used the Yahoo! Finance open API<sup>3</sup> to crawl historical stock prices for each stock symbol, including daily closing price and daily volume (# of shares traded). We then adjusted closing prices and volume to account for stock splits and dividends in order to finalize a set of complete daily price histories for all stocks.

### Basic Analysis

*Distribution of Posts.* We plot the distribution of posts over users and boards in Figure 1. The distribution shown is using the Complementary Cumulative Distribution Function (CCDF). Both distributions are highly skewed. Roughly 80% of users on iHub and 95% of users on Yahoo post less than 50 posts, and a significant number of users do not post any messages. We see similar trends across message boards, where 80% of iHub boards and 40% of Yahoo boards have less than 200 posts over their lifetime. Relatively speaking, iHub’s user population, while smaller, tends to generate more messages than Yahoo users. This is likely because discussion board is iHub’s primary function, while message boards are only a small component of Yahoo Finance.

*Distribution of Market Cap.* *Market cap.*, short for market capitalization, is the total dollar market value of all of a company’s outstanding shares. It effectively captures the size of a company, and is a critical part of any fundamental method for computing the valuation of a company and its shares. Market cap is also heavily associated with a stock’s risk/reward trade-off. Generally speaking, large cap companies tend to be lower risk and give lower returns compared to companies with smaller market cap.

We use a variant of the categories defined in (Wayman 2016) and divide companies into 5 different categories us-

<sup>3</sup><http://finance.yahoo.com/>

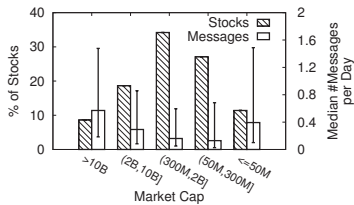


Figure 2: Distribution of Yahoo boards and average messages by market cap.

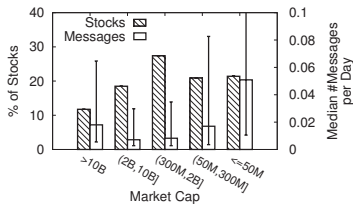


Figure 3: Distribution of iHub boards and average messages by market cap.

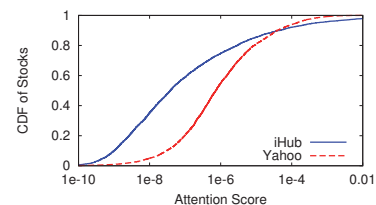


Figure 4: Distribution of stock attention scores for both datasets.

ing their market cap on April 5, 2016, as reported by Yahoo! Finance. Figure 2 uses a dual Y-axis graph to show the distribution of companies with Yahoo message boards sorted by market cap, and the median posts per day (along with 25<sup>th</sup> and 75<sup>th</sup> percentiles). Clearly, most of Yahoo boards belong to mid-cap sized companies, but boards with the heaviest post traffic are in the highly speculative, nano cap ( $\leq 50M$ ) companies and the large cap companies. In contrast, as shown in Figure 3, iHub’s message boards are more evenly distributed by company size, but iHub’s heaviest traffic also comes from the small set of speculative, nano cap ( $\leq 50M$ ) companies.

*User Attention.* Finally, we propose a new metric called *attention score* to capture the level of user attention received by a stock. We define it as the number of total posts of a stock divided by its market cap. The goal is to normalize for the impact of proportionally larger number of investors in the user population who own a given large cap stocks, e.g. Apple. A higher score means that users are more active in talking about the stock. We plot the Cumulative Distribution Function (CDF) for message boards in both datasets in Figure 4, and note that while most boards in Yahoo have generally higher attention scores, the most active message boards at Yahoo are less active than those at iHub (Figure 1(b)). This is likely due to the extremely high activity on microcap stocks in iHub (market cap  $< \$50M$  in Figure 3).

## Sentiment Extraction

To measure the value of user posts on guiding investment strategies, we perform sentiment analysis on online discussion boards in the context of stock market, which explores interactions between media content and stock market activity. Previous work focused primarily on extracting content from news media (Zhang and Skiena 2010), general social media (Gilbert and Karahalios 2010; Bollen, Mao, and Zeng 2011) and finance-specific crowdsourced services (Wang et al. 2015).

Our first step is to develop reliable tools to interpret the sentiment of posts. The Yahoo data has a subset of posts with sentiment naturally labeled by users themselves, while iHub data does not. Thus we apply a *classification based* approach to Yahoo and a *keyword dictionary based* approach to iHub. Our validation results show these tools achieve an accuracy of 72.1% for Yahoo and 81.0% for iHub, which are on par with or significantly better than popular sentiment analysis tools (Chen et al. 2014; Gonçalves et al. 2013).

## Classification Approach (for Yahoo)

A classification approach trains a machine learning classifier on labeled posts, where labels are sentiment of posts, and features are extracted based on content of posts. The classifier is then applied to unlabeled posts to recover their sentiment. Roughly 5% of Yahoo messages already have sentiment labeled by users themselves, as one of the following: “Strong Buy”, “Strong Sell”, “Buy”, “Sell” and “Hold”. We regard both “Buy” and “Hold” as positive sentiment, while “Sell” as negative. Holding a stock is positive, because it means the owner is still expecting the stock price go up in the future. Out of these 5% of messages, we have 17.2% negative posts and 82.8% positive posts.

To generate features, we follow a supervised machine learning method from (Wang et al. 2015), where unigrams are regarded as features. Unigrams of a post are the numbers of occurrences of each unique word in the post. We also remove stop words, stock symbols and urls from messages, and exclude infrequent unigrams that occur less than 700 times<sup>4</sup> over all messages. We apply Naive Bayes, Random Forest, Supported Vector Machine (SVM), Decision Trees and Logistic Regression from the standard machine learning package Scikit-learn (Pedregosa and others 2011).

We run 10-fold cross validation on the labeled data, and find that SVM performs best. To deal with strong data imbalance (fewer negative posts than positive posts), we apply undersampling to build our training dataset, keeping all negative posts while varying the number of positive posts (He and Garcia 2009). The best results appear when # of positive posts: # of negative posts = 2.0, with an overall accuracy of 72.1%, 75% precision and 86% recall for positive predictions, 61% precision and 43% recall for negative predictions.

## Keyword Based Approach (for iHub)

A keyword-based approach requires a dictionary of positive and negative keywords, with the resulting sentiment of a post extracted by counting the number of positive or negative keywords in the post.

The dictionary is generated by analyzing content of a small number of labeled posts. To generate such labeled posts for iHub, we randomly sampled 4,500 (about 0.1%) posts across all boards, and asked three native speakers in English to manually label them with one of the following: “Buy”, “Sell” and “Hold”. Like Yahoo, we regard both

<sup>4</sup>We tried thresholds from 100 to 2000 and found similar results.

“Buy” and “Hold” as positive sentiment while “Sell” as negative. In total, we have 96% positives and 4% negatives for these 4500 posts. Next, we extract keywords from the newly labeled posts and build a sentiment dictionary to identify positive and negative keywords (Wang et al. 2015). Specifically, we extract unigrams from labeled posts, separate unigrams of positive posts from those of negative posts, and rank them in descending order of frequency. Then for the top ranked unigrams in the positive and negative lists, we use Chi-square statistics (Lowry 2014) to select the most distinguishing words as the sentiment dictionary. Finally, we add negation, which usually reverses the sentiment of words (e.g., “bullish” is positive while “not bullish” is negative), by searching for negation words and reversing the sentiment of affected words (Pang, Lee, and Vaithyanathan 2002).

The resulting dictionary is then used to calculate the sentiment of a post. We label each sentence of the post to be either positive or negative, depending on whether there are more positive or negative words in the sentence. The overall post is considered positive if positive sentences outnumber negative sentences and vice versa.

For validation, we randomly sampled another 500 posts and manually labeled their sentiment. We get an overall 81.0% accuracy, with 98.8% precision and 81.3% recall for positive posts, and 13.9% precision and 75.3% recall for negative posts. Although the precision for negative posts is low, recall is high. This means that our approach is often aggressive in finding negative posts. In the context of our analysis, more false positives in detecting negative sentiment means we could over-report negative sentiment. This means our results are likely understated, and insulation from negative sentiment is likely even stronger than we report in the next section.

## Echo Chambers

In this section we present evidence of echo chambers in the discussion boards. To do so, we bring together four sets of analyses. While our analyses are largely correlational, they do allow us to show how social phenomena in online communities can align with concretely measurable outcomes in a distinct ‘social computation engine’ of significant import, the stock market. The four analyses are as follows:

- *Poor User Generated Information*: sentiment in the discussion boards does not correlate with future stock price movements, and in fact stock trading strategies based on the sentiment in these boards underperform the overall market.
- *Failing to Incorporate New Information*: sentiment in the discussion boards does not correlate with recent past stock price movements.
- *Resistance to Viewpoint Diversity*: user clustering reveals that the majority of sentiment in the discussion boards may reflect the viewpoints of a very small percent of users, and that many users resist negative viewpoints.
- *Mistimed Activity Adds Noise*: posting activity correlates positively with future stock trading volume and volatility.

## Poor User Generated Information

A central tenet of an echo chamber is that it is likely to generate information misaligned with that from outside the echo chamber. This may be due to its insular nature, social structure of its membership, or resistance to alternate viewpoints, all attributes we address below. For now, in this section we present evidence that the output of these investment discussion boards is poor with respect to the very thing these discussion are trying to predict: future stock price movements. First, we apply our two sentiment analysis methods over board posts and correlate the resulting sentiment time series with future stock price moves. As we will see, these correlations are effectively zero. For further evidence we ran trading simulations based on investment board sentiment to show that such strategies underperform simple buy-and-hold market index strategies.

### *Sentiment and future stock performance not correlated*

We measure the correlation between the daily sentiment score and future price change for each stock on every exchange day. Messages on non-exchange days are combined into the most recent exchange day. We skip days when there is no post for a given stock, and all analyses are restricted to stocks with posts on at least 5 days, though changing this 5 day minimum threshold yields similar results.

*Sentiment score* is a daily measure, calculated as the ratio of positive posts over all posts for a given stock on a given day. Higher values indicating more positive daily sentiment. A potential limitation of this metric is that for some days the score is only based on a single post when there is only one post at that day. This may make the results less robust.

*Price change* is a daily measure, calculated as  $price(X + T) - price(X)$ , where  $price(X)$  is the adjusted closing price for a stock on day  $X$ , and  $T$  is a time window during which we consider the price movement. We choose  $T = 5$  days for reporting results, though other values yield similar results.

*Correlations* are calculated by Pearson Correlation Coefficient (Pearson 1895), which measures the linear correlation between two variables. It is calculated as the covariance of the two variables divided by the product of their standard deviation, which produces a value in range  $[-1, 1]$ . A value of 1 means total positive correlation between the two variables, -1 means total negative correlation while 0 means no correlation.

We show the cumulative distribution function (CDF) of correlation coefficients between board post sentiment and future stock price moves in Figure 5. The majority of stocks have a correlation coefficient around 0 (median  $\leq 0.003$ ). These results indicate that broadly there is almost no correlation between the sentiment around a stock in the discussion boards and future performance of that stock. However, it is possible that these summary correlations are obscuring informative signals in discussion board sentiment that is only detectable under certain conditions. Thus, we explore a variety of nuances to better understand the lack of correlation between sentiment and future stock performance.

*Users are always optimistic no matter whether stock prices go up or down.*



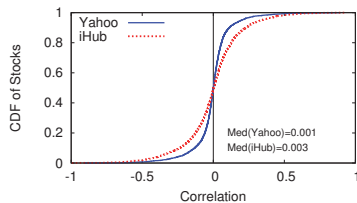


Figure 5: Correlation evaluation between sentiment and stock price movements.

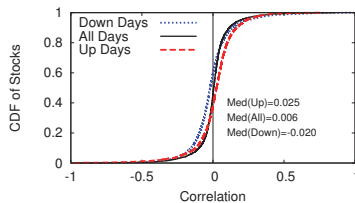


Figure 6: Pearson Correlation between sentiments and stock price movements for up/down/all days (Yahoo).

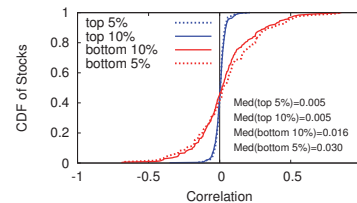


Figure 7: Correlation evaluation for stocks with top/bottom attention scores (Yahoo).

For each stock, we split its exchange days into “up days”, where the stock price change is positive, and “down days” where stock price change is negative, and calculate their correlation separately. The result is shown in In Figure 6. Compared to all days, which show almost no correlation, the curve of up days does have a higher correlation coefficient (median 0.025), suggesting a still very small, but non-zero positive correlation. On the other hand, the curve of down days has a median value of -0.020, indicating a very small negative correlation. While these are essentially null results, there is a small signal that when the market is bullish, investors’ expectations for stock price performance are better met, while in a bearish market investors’ sentiment is more irrational since they are still positive regardless how stock actually behaves.

#### *People pay more attention to higher risk stocks.*

Here we test whether the correlations are different between stocks that receive more attentions and stocks that receive less attention. As previously mentioned, we use the attention score to quantify board users’ level of engagement with each stock.

Figure 7 shows that stocks receiving the highest attention have almost no correlation between board sentiment and future price movement, while stocks with the lowest attention level have small but positive correlation. Inspection of high attention stocks shows that many of them are highly volatile and thus despite being inherently more unpredictable appear to attract the interest by people irrationally focused, presumably, on the high return rather than the high risk.

#### *Trading Simulations Underperform Market*

As a further test, we developed a trading strategy where a user makes investment decisions based on stock-specific board post sentiment. Specifically, each user takes a pool of money and splits it evenly on stocks whose message boards she has engaged with. For each stock, we consider two levels of ownership: a total (100%) position vs. a 50% position. If the sentiment of the previous day is positive (sentiment score > 0.5), the user increase her position by 50% or does nothing if it is already 100%. Correspondingly, for negative sentiment (sentiment score < 0.5) the user decreases her position by 50% or does nothing if it is already 0%. For days with no posts, we do not take any trading action.

We evaluate this sentiment-based trading strategy in a simulation study, based on the past  $N$  years of historical

stock data<sup>5</sup>. At the end of the simulation period we calculate the net return for each user, and compare this net return to the net return of simple S&P 500 market index.

We find that the sentiment-based trading strategy underperforms the market. Figure 8 shows the CDF of net returns of our trading strategy with different  $N$  values. The big dot on each line shows the return of S&P 500 index for the corresponding  $N$  value. We see that at least 50% of users underperform the market index. In particular, for shorter simulation periods ( $N = 2, 4, 6$  years), 60-70% of users lose money compared to the market index. This means that in general, message sentiment is not helpful, and can even be harmful as a stock trading strategy. We observe very similar trends in iHub to those based on Yahoo data, or even worse where 60-80% of users have less return than the S&P 500. Individual investors are better off buying the entire market than trying to time price movements of individual stocks using discussion boards.

### **Failing to Incorporate New Information**

In the previous section we showed that these investment boards produce information that in fact is non-informative and generally will hurt investors. Here we investigate the possibility that these boards are insulated from outside information. Such insulation is both an attribute of echo chambers and a possible explanation for why these discussion board communities produce bad information. In the context of investment discussion boards, such isolation would be reflected in the inability of the community members to incorporate important new information about stocks. To test this, we examine whether the sentiment of board posts reflects changes in stock prices of the recent past. A lack of correlation with subsequent board sentiment implies that board discussion is not adequately capturing these changes. Also, given that the in the previous section we documented the lack of forward looking correlation between board sentiment and stock price movement, a lack of backward-looking correlation rules out the explanation that board sentiment simply reflects commentary on past events.

To address this, we consider past price change ( $price(X) - price(X - T)$ ) instead of future price change ( $price(X + T) - price(X)$ ). We denote this new correlation as “Past” and the original as “Future”.

<sup>5</sup>We omit stocks that do not have full price history within the past  $N$  years.

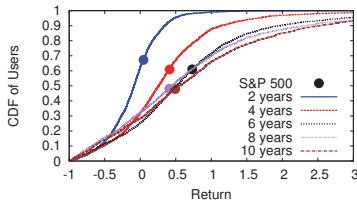


Figure 8: Distribution of net return using sentiment-based trading, over different simulation periods  $N$  (Yahoo).

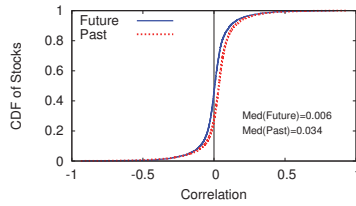


Figure 9: Pearson Correlation sentiment vs. future/past stock performance (“Future” vs. “Past”)

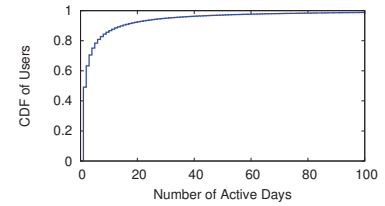


Figure 10: Distribution of active days over users in Yahoo message boards.

Figure 9<sup>6</sup> shows very similar curves for “Future” (blue line) and “Past” (red dashed line) except that “Past” has a slightly more positive correlation (with a median correlation of 0.034). Thus, while people’s sentiment might reflect past stock movement marginally better than as a forecast for future movement, they still have very little correlation with stock performance, either past or future.

### Resistance to Viewpoint Diversity

A third characteristic of an echo chamber is that only a few viewpoints are represented and that the group generally is resistant to opinion change. This is particularly damaging to the ability to make successfully predictive commentary about the stock market given that the market is an aggregation of many thousands of people. The isolated opinions of a few are unlike to align with the average of many. In this section we identify the *roles* of board members, which capture commonalities in behavior patterns (Arazy et al. 2015). Examples of roles repeatedly mentioned in the literature include *lurkers* (Sun, Rau, and Ma 2014), *leaders* (Zhu, Kraut, and Kittur 2012) and *chatters* (Nolker and Zhou 2005). Our results show that the bulk of sentiment is likely to be produced by a very small minority, and importantly that the majority of users that react to new information are highly resistant to negative sentiment.

We start by filtering out *lurkers*, who are inactive users that contribute very little content. We then do clustering on the remaining users, whom we call *contributors*. We adopt an unsupervised learning approach called *similarity graph* (Wang et al. 2016) to identify different user groups. We then study the flow of messages between these user groups and how it affects overall sentiment. In the following, we report our findings on Yahoo message boards, omitting similar results for iHub due to space limitations.

#### Identifying Lurkers

As reported in many online community studies (Sun, Rau, and Ma 2014), the bulk of users are inactive, often referred to as lurkers. As contributors are the focus of this analysis, to exclude lurkers we set a threshold on *active days* to quantify the activity level of a user, which is the total count of days that she posts messages. We show the CDF of active days per user in Figure 10, and observe a highly skewed distribution.

<sup>6</sup>We also compare results between different strategies for up/down days and attention scores, and observe similar trends.

We use a threshold of 2 weeks to select contributors<sup>7</sup>, and find that these 10% of users contribute 74.5% of total posts.

#### Clustering via Similarity Graph

To identify behavior groups among contributors we apply the aforementioned *similarity graph* to find natural clusters among users. In a similarity graph, each user is represented by a node, and every pair of users are connected by a weighted edge, which represents the distance between them. User distance is calculated by comparing a set of behavioral features between the pair of users. Thus in the similarity graph, similar users are connected by short-distance edges. After building the similarity graph, we can extract natural user clusters by applying a clustering algorithm on the similarity graph.

The key step in building a similarity graph is to choose appropriate features that are representative in user behaviors. We calculate the following 3 types of features to cover the most important aspects:

- *Posting behaviors*: the count of posts a user initiated and replied to per active day respectively.
- *Interaction with others*: the time difference between a user’s post and the post she replies to, *i.e.*, how fast she responds to others.
- *Influence level*: the average length (*i.e.*, count of total posts in a thread) of message threads initiated by a user, where a larger value means she leads longer discussions and thus is more influential.

After calculating these features for each user, we apply Euclidean distance as the distance metric for edge weight in the similarity graph. We identify different user behavior groups by applying Divisive Hierarchical Clustering (Kaufman and Rousseeuw 2009). Divisive Hierarchical Clustering begins with all users in a single cluster, and continuously splits clusters. We terminate the splitting process when a further split does not improve the clustering quality, which is measured by *modularity* (Blondel et al. 2008), a popular metric that compares the distance of edges inside and outside clusters.

#### User Behavior Groups

By applying the above clustering algorithm on contributors, we obtain 5 main clusters, with 0.03% users left as

<sup>7</sup>We also used other thresholds such as 1 month, and omit similar results here.

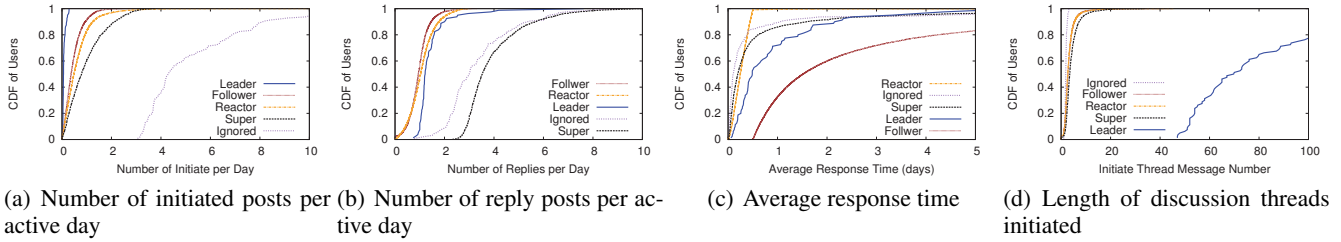


Figure 11: Feature distribution in different user groups.

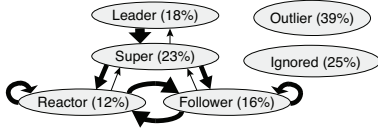


Figure 12: Flow of messages with negative sentiment.

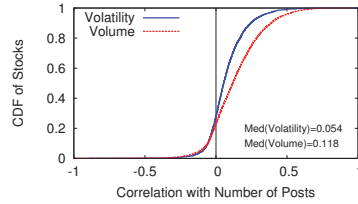


Figure 13: Correlation between number of posts with trading volume/volatility (Yahoo).

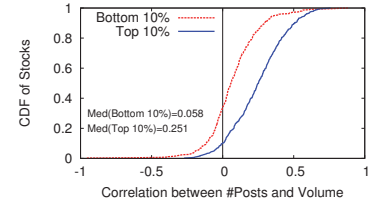


Figure 14: Correlation between number of posts and volume for stocks with top/bottom attention scores (Yahoo).

outliers. To interpret the roles of these user groups, we measure the distribution of each feature among different clusters, shown in Figure 11.

- **Leaders (0.09%).** These are generally discussion leaders in message threads. They do not initiate or reply to messages frequently, but when they initiate a message thread, they attract many replies (*i.e.* the blue curve in Figure 11(d)).
- **Super users (1.66%).** These users are active in both initiating posts and replying to others (the black curve in Figure 11(a) and Figure 11(b)).
- **Ignored (0.14%).** These have extremely high frequency of initiating posts (purple line in Figure 11(a)), but get very few replies (Figure 11(d)). They tend to participate in more message boards, and write short messages (*e.g.*, “START BUYING!! GO CGO!”). They are likely spam accounts trying to draw attention to certain stocks, and are generally ignored by others.
- **Followers (61.8%) and Reactors (36.3%).** These two user clusters make up the bulk of active users. What separates the two groups is the interaction pattern with others. Followers have the longest average time to respond (brown curve in Figure 11(c)), *i.e.* they tend to reply very late to posted messages. In contrast, Reactors respond to messages with the shortest average delay (orange curve in Figure 11(c)), indicating that they actively participate in discussions.

#### Sentiment and Message Flow

Having identified natural roles in message boards, we want to understand how they affect the flow of messages and overall sentiment. To do so, we look at our messages from the Yahoo message board dataset that come with embedded sentiment (either bullish or bearish). Note that this provides

us with a ground truth dataset of sentiment in these messages about the stocks to which they refer.

We illustrate the possible directions of message flow between different groups in Figure 12, and note the ratio of messages authored by each group that have negative (bearish) sentiment. There are a few things worth noting. First, messages from the small “leader” and “super” groups have more negative content than their followers. More specifically, the fastest reacting “reactor” group registers a significant drop in negative sentiment from near 20% to only 12%, meaning there is a significant dilution or reversal of negative opinion within the highly active group. It’s also worth noting that the most negative sentiment is generated by those in outlier clusters, users who do not fit into the normal communication patterns of the message boards.

Finally, we find a higher level of animosity or antagonism in responses to those who post negative messages. To do so, we randomly select 200 of our ground truth sample of threads from Yahoo, and manually label all followup messages (~650 posts) as “antagonistic” or “neutral/friendly”. From messages marked “antagonistic”, we build a dictionary of the strongest negative keywords that capture animosity, including: attack, blame, blind, bad, bet, blowhard, fool, failure, dad (I’m your daddy), disgruntled, ignorant, incorrect, homework (do your homework), misunderstand, newbie, ugly, sick, suck, scam, stupid, piddling, peabrain, pitiful, poor, wary, worst.

We use this “animosity” dictionary to capture the level of personal animosity in responses to both negative and positive posts. We compare our 200 threads starting with negative sentiment against 200 threads with positive sentiment, and look at the frequency of occurrence of our dictionary



words in the responses. The differences are dramatic. When we count the frequency with which our negative words occur, they account for 0.25% of all words in responses to positive posts, but 0.7% of all words in responses to negative posts. Similarly, when we look at the portion of responses that contain at least one word from our animosity dictionary, 7.8% of responses to positive posts match, compared to 19.2% for responses to negative posts. These results highlight the antagonistic response from the community to negative sentiment posts in message boards, and may help to explain why there is so little room for dissenting opinions in these forums.

### **Mistimed Activity Adds Noise**

In the previous subsections we showed that investment board discussions neither captured important information from the recent past nor generated information of value going forward. Further, the bulk of the discussion is initiated by a small subset of users and many community members actively resist negative sentiment. In this section we examine possible effects of users taking action based on this poor information. To be clear, we do not actually know what, if any, action users took. Instead we present evidence of the most likely outcome of people acting en masse on poor information: increased activity and variance in the market. Specifically, we show that board posting activity correlates with future trading volumes and volatility.

The trading volume of a stock on a particular day is the number of shares exchanged during that day. Volatility of a stock describes the extent to which the price changes or fluctuates. To calculate volatility, we use the standard deviation of price change during a fixed-length time range. To describe user activity, we measure the *number of posts*, which is simply how many messages appear on the discussion board.

We again use the Pearson correlation coefficient to describe the relationship between user behavior and market status. As volatility calculation requires a time range, for each stock we correlate the number of posts from day  $X - W + 1$  to day  $X$  with volume or volatility from day  $X + 1$  to day  $X + W$ . The  $W$ -day window sweeps through the whole history of each stock. We use  $W = 5$  to report our results, but other window lengths produce similar results.

Figure 13 shows positive correlations between number of posts and volume (median  $r = .11$ , mean  $r = .14$ ) and volatility (median  $r = .05$ , mean  $r = .08$ ). Furthermore, when we break out stocks by their attention score, we find that the correlation is stronger in higher-attention stocks (e.g., the median correlation between post activity and volume is .25), as shown in Figure 14. These results suggest that user activity in these investment boards are somewhat correlated with market activity (volume and volatility), indicating a possible source of noise and inefficiency in the stock market.

## **Discussion**

The preceding results provide evidence of the ineffectiveness of two extremely large online communities. These investment boards have the near-singular goal of helping people discuss and make sense of the stock market, and yet sentiments expressed in board discussions fail to correlate with

either past or future stock price movements. If nothing else, one would imagine these boards would be vehicles for people to vent about falling stock prices, but we simply do not see that. Instead, we find that discussions are often initiated by a very small number of community members, and discussions that fail to conform to an overly positive outlook are often rejected by the community.

Methodologically, a unique feature of investment discussion boards is that they can be aligned with extremely concrete outcomes in a related but distinct environment. Here, we found non-trivial correlations between investment board activity and near-future noise in the stock market. As mentioned, we are clear that we have no way of knowing what market actions our investment board members actually took. Circumstantially, the size of the boards in terms of numbers of people is large and our data samples cover substantial lengths of time and numbers of stocks, so it is not inconceivable that these boards are a source of noise in the market.

The financial market is a social computation engine, and such inefficiencies in this engine can have real world consequences. Even a small number of noise traders may adversely affect markets informationally and allocationally with significant price impact, by changing the prices of assets in extreme states of the economy (Kogan et al. 2006). We can easily imagine that such extreme states and price moves are much more likely in a heightened emotional state, when investors are more likely to seek conformity and herd. Thus it can create a feedback loop that is impervious to actual information. By increasing the volatility (risk), it subsequently increases the rate of return (reward) or the cost of capital any asset must offer. Based on our results where the most negative impacts are felt on small stocks, these echo chambers might have significant impact on the viability of small firms in the real economy by increasing their financing cost.

Given the fairly clear negative effects of these boards, both individually and for the market as a whole, why would people engage or pay any attention at all to these discussions? The picture painted of the bulk of active members is one of irrational investors stubbornly positive about long shot bets. These bets appear to reflect a skewed risk/reward assessment, where members are overly focused on ‘big payoff’ small cap stocks that in reality are also more likely to yield big losses. In itself this is damaging to these people’s personal investments, as the market simply is more balanced. The stock market is extraordinarily complex. Thus it is natural that investors would look to others for guidance, akin to a social learning process. However, as we have shown here, this community of other investors is a poor group to model.

Ideally, these discussion communities are facilitating efficiency rather than noise. To leverage rather than waste the considerable amount of human attention and information sharing that investment discussion communities support, we suggest designing for three online community characteristics in particular. First and foremost, support diversity of opinion, which almost by definition will increase market efficiency. While there is no direct evidence correlating stronger opinion diversity with outperformance, lack of di-

iversity in this case is like extracting “wisdom” from a biased crowd. Second, ensure that new information can be easily incorporated. This could potentially be done through technological intervention. For instance, news and financial data for each company could be aggregated and kept current within the discussion forum. Finally, as with any task, feedback is critical. In this case, performance based on sentiment can be tallied and reflected back for both individuals and the community as a whole.

## Conclusion

Our work studies the presence of echo chambers in investment discussion boards, and their potential contribution to noise trading behavior by participating users. Using large datasets drawn from Yahoo and iHub messages boards, we show that discussion board activity and post sentiment show negligible correlation with future market returns, and trading simulations based on board sentiment are poor investment strategies. We examine the social structure of communities inside discussion boards, and observe that the large majority of sentiment is led by a small core of users. Our results suggest that discussion boards might play a significant role in adding to noise trading by individual investors. We hope that by shedding light on the quality of these discussion boards, we encourage the adoption of forums more encouraging of divergent opinions and open exchange of ideas.

*Acknowledgments.* The authors wish to thank the anonymous reviewers for their comments. This project was supported by NSF grants CNS-1527939 and IIS-1321083. 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.

## References

- Arazy, O.; Ortega, F.; Nov, O.; Yeo, L.; and Balila, A. 2015. Functional roles and career paths in wikipedia. In *Proc. of CSCW*.
- Bakshy, E.; Messing, S.; and Adamic, L. A. 2015. Exposure to ideologically diverse news and opinion on facebook. *Science* 7(6239):1130–1132.
- Benartzi, S., and Thaler, R. H. 2001. Naïve diversification strategies in defined contribution saving plans. *American Economic Review* 91:79–98.
- Black, F. 1986. Noise. *The journal of finance* 41(3):529–543.
- Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008(10).
- Bloomfield, R.; O’hara, M.; and Saar, G. 2009. How noise trading affects markets: An experimental analysis. *Review of Financial Studies* 22(6):2275–2302.
- Bollen, J.; Mao, H.; and Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2(1):1–8.
- Boutyline, A., and Willer, R. 2016. The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology*.
- Chen, H.; De, P.; Hu, Y. J.; and Hwang, B.-H. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies* 27(5):1367–1403.
- Das, S. R. 2014. Text and context: Language analytics in finance. *Foundations and Trends (R) in Finance* 8(3):145–261.
- Garrett, R. K. 2009. Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of Computer-Mediated Communication* 14(2):265–285.
- Gilbert, E., and Karahalios, K. 2010. Widespread worry and the stock market. In *Proc. of ICWSM*.
- Gilbert, E.; Bergstrom, T.; and Karahalios, K. 2009. Blogs are echo chambers: Blogs are echo chambers. In *Proc. of HICCS*.
- Gonçalves, P.; Araújo, M.; Benevenuto, F.; and Cha, M. 2013. Comparing and combining sentiment analysis methods. In *Proc. of COSN*.
- Grevet, C.; Terveen, L. G.; and Gilbert, E. 2014. Managing political differences in social media. In *Proc. of CSCW*.
- He, H., and Garcia, E. A. 2009. Learning from imbalanced data. In *Proc. of TKDE*.
- Kahneman, D., and Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–291.
- Kaufman, L., and Rousseeuw, P. J. 2009. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.
- Kogan, L.; Ross, S.; Wang, J.; and Westerfield, M. 2006. The survival and price impact of irrational traders. *Journal of Finance* 61(1):195–229.
- Kyle, A. S. 1985. Continuous auctions and insider trading. *Econometrica* 53:1315–1336.
- Lowry, R. 2014. Concepts and applications of inferential statistics.
- Malkiel, B. G., and Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25(2):383–417.
- Nolker, R. D., and Zhou, L. 2005. Social computing and weighting to identify member roles in online communities. In *Proc. of WI*.
- Pang, B.; Lee, L.; and Vaithyanathan, S. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proc. of EMNLP*.
- Pearson, K. 1895. Note on regression and inheritance in the case of two parents. *Proc. of the Royal Society of London*.
- Pedregosa, F., et al. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.
- Shiller, R. J.; Fischer, S.; and Friedman, B. M. 1984. Stock prices and social dynamics. *Brookings Papers on Economic Activity* 1984(2):457–498.
- Sun, N.; Rau, P. P.-L.; and Ma, L. 2014. Understanding lurkers in online communities: A literature review. *Computers in Human Behavior* 38:110–117.
- Sustein, C. 2001. *Republic.com*. Princeton University Press.
- Wang, G.; Wang, T.; Wang, B.; Sambasivan, D.; Zhang, Z.; Zheng, H.; and Zhao, B. Y. 2015. Crowds on wall street: Extracting value from collaborative investing platforms. In *Proc. of CSCW*.
- Wang, G.; Zhang, X.; Tang, S.; Zheng, H.; and Zhao, B. Y. 2016. Unsupervised clickstream clustering for user behavior analysis. In *Proc. of CHI*.
- Wayman, R. 2016. Understanding small- and big-cap stocks. *Investopedia.com*.
- Yardi, S., and Boyd, D. 2010. Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of Science, Technology, and Society* 30(5):316–327.
- Zhang, W., and Skiena, S. 2010. Trading strategies to exploit blog and news sentiment. In *Proc. of ICWSM*.
- Zhu, H.; Kraut, R.; and Kittur, A. 2012. Effectiveness of shared leadership in online communities. In *Proc. of CSCW*.