2020

The Effect of Incentive Hierarchy System of Social Media In the Delivery of Quality Information

Peng Xie
California State University East Bay, peng.xie@csueastbay.edu

Peng Xie
pengxie1987@gmail.com

Follow this and additional works at: https://scholarworks.lib.csusb.edu/jitim

Part of the Management Information Systems Commons

Recommended Citation
Available at: https://scholarworks.lib.csusb.edu/jitim/vol28/iss4/1

This Article is brought to you for free and open access by CSUSB ScholarWorks. It has been accepted for inclusion in Journal of International Technology and Information Management by an authorized editor of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.
The Effect of Incentive Hierarchy System of Social Media In the Delivery of Quality Information

Cover Page Footnote
*Corresponding Author. Affiliation: College of Business and Economics, California State University, East Bay. Address: 25800 Carlos Bee Blvd, Hayward. Email: peng.xie@csueastbay.edu. Phone: (510) 674-6724.

This article is available in Journal of International Technology and Information Management: https://scholarworks.lib.csusb.edu/jitim/vol28/iss4/1
The Effect of Incentive Hierarchy System of Social Media in the Delivery of Quality Information

Peng Xie
(California State University East Bay)

ABSTRACT

This paper investigates the effect of social media incentive hierarchy system on the quality of the shared information and the spillover effect using data from 67 cryptocurrency markets along with the corresponding social media discussions. We show empirical evidence that high-rank social media users displaying high-level badges earned from the social media tend to provide low-quality information due to reduced incentives after obtaining the badge and increased tendency to engage in less informative socialization activities. In contrast, low-rank social media users with low-level badges tend to provide high-quality information. However, messages shared by high-rank social media users spill over to other cryptocurrency markets more easily because of higher visibility in the online community.

Keywords: social media, incentive hierarchy system, spillover effect, cryptocurrency, text analysis, panel data

INTRODUCTION

Incentive hierarchies are common practice in online gaming as a way to motive user activities. Users are awarded badges by achieving various goals. In recent years, many social media platforms also implemented incentive hierarchy systems to gamify the user experience in order to encourage participation and contribution. The fundamental idea is to help users internalize the benefits of content sharing in a “free-riding” environment where all information shared is available to everyone (Goes et al., 2016).

In most cases, the incentive hierarchy systems allow users to accumulate points for contributing new content or engaging in other types of social interactions. Badges are awarded when the points accumulated reach a threshold (Goes et al., 2016). However, do users with high-level badges always share high-quality information? This is the first question we try to answer.

In this research, we mainly focus on social media that allows the users to communicate and exchange opinions on cryptocurrency investments. The quality
of the shared messages is measured by the association between the message sentiment and the future cryptocurrency returns.

Social media users incur time cost and effort cost to share private information with others, and they also forfeit their information advantage by publicizing their private information. So the users must be motivated in some way to share. Wasko and Faraj (2005) summarized the socialization-related motivations that incentivize people to communicate online with others: (1) reputation: the approval, respect, and status gained when engaging in social interaction (Blau, 2017); (2) enjoying helping: the good feelings and enjoyment when helping others (Kollock, 1999); (3) individual’s structural centrality increase willingness to contribute (Wasko and Faraj, 2005); (4) commitment: the perceived duty and obligation to engage in interactions (Coleman, 1994); (5) reciprocity: the perceived moral obligation to pay back to peers and the network (Wasko and Faraj, 2000).

Besides the socialization-related motivations, economic-related motivations also play an important role. On message boards dedicated to investment opinion discussions, informed traders benefit from constructive feedback, complementary information, and confidence while communicating with their peers (Gray and Kern, 2011). Sometimes even with high-quality information, informed traders don’t necessarily have the financial resource to correct the price discrepancy and realize the profits. So they have the incentive to share their private information to create a trading momentum. Together with their peers, they might move the market to the desired direction to realize the profits (Tumarkin and Whitelaw 2001).

Social media users are motivated by both the socialization-related factors and the economic-related factors to justify the cost associated with the sharing activities. We argue that high-quality information is shared when the users are primarily motivated by economic-related factors rather than by the socialization-related factors. For active high-rank users with many connections within the online community, the cost associated with online sharing is more easily compensated by socialization-related motivations compared to low-rank users.

In comparison, low-rank users are comparatively less active in peer communications and their activities are unlikely motivated by socialization-related factors but by economic-related factors. Therefore we expect that these users holding low-level badges to share more informative and value-relevant content.

Our prediction can also be explained by the drive-reduction theory (Dewey, 2007), which states that the motivation drops immediately after the goal is reached. In most cases, the badges are permanently offered by the social media incentive hierarchy.
system when they are obtained. Users will lose the incentives after receiving the badges because there is less or no room to improve.

The social media hierarchy system also has implications on the spillover effects. It has long been established in the finance literature that the information outlets from intra-industry competitors will influence each other (Helwege and Zhang, 2015, Lang and Stulz, 1992, Otchere, 2007, Goins and Gruca, 2008, Dajcman et al., 2012, Hameed et al., 2015). The cryptocurrency industry thrived around the end of 2012 when many other cryptocurrencies besides the Bitcoin started to emerge. Due to the decentralized nature of the cryptocurrency industry, there is no earning releases, no firm announcements, no professional financial analysts, no quarterly or annual financial statements (Xie et al. 2019). With limited official information sources, social media becomes vital in transmitting related information. As a result, most of the information becomes public. So we expect that the cryptocurrency markets experience strong spillover effect through social media.

We predict that information from high-rank users is more likely to induce the spillover effect than the information from low-rank users because the high-rank users enjoy higher visibility and recognition in the online community. The badges advertise one’s achievements and past accomplishments and are easily treated as symbol of experience and tenure in the field (Antin and Churchill, 2011). So the posts from these users will have wider exposure to the public than their peers with low-level badges.

We collected social media messages from a leading cryptocurrency message board called Bitcointalk.org, along with the badge information, from February 2015 to February 2017. The dataset contains a discussion about the industry leader Bitcoin and 66 of its major competitors (usually referred to as Altcoins). The sample consists of more than 190,000 Bitcoin-related discussion messages and more than 620,000 discussion messages for each of the 66 Altcoins. The price data for Bitcoin and all Altcoins during the same period are also collected. We first set up a baseline analysis to check the predictive power of the collective discussion sentiments. Then we verify if the low-rank users share information with higher quality. Finally, we test if the spillover effect exists in the cryptocurrency industry.

The paper is organized as follows. Section 2 introduces the related literature and develops hypotheses. Section 3 describes the data collection process and variable operationalization methods. Section 4 describes the empirical models used to test our predictions and presents the results. Section 5 concludes the study.
LITERATURE REVIEW

This study is based on the recent finding that social media contents provide valuable insights into future return predictions in the financial markets. Actually, researchers have long been aware that traditional financial reports and editorial media can predict stock market returns (Davis et al., 2012, Loughran and McDonald, 2011, Tetlock, 2007, Tetlock et al., 2008, Solomon, 2012). Studies also show that the discussions on many online message boards demonstrate predictive powers for future price movement, even though social media discussions are unregulated and there is no guarantee for the information quality (Tumarkin and Whitelaw, 2001, Das and Chen, 2007, Chen et al., 2014).

However, there are also studies that did not find support for this predictive power. Dewally (2003) used buy and sell recommendations from an online discussion group to predict the stock market returns but failed to establish the relationship. Antweiler and Frank (2004) studied the effects of messages posted on Yahoo! Finance and found only mild influence. It is inevitable that there is a huge amount of noise information and off-topic discussions on the social media platform. We aim to establish a relationship between the quality of social media content and the user characteristics represented by the social media hierarchies.

The social media hierarchy is a representation of the user’s past achievements and level of participation in online activities such as posting and commenting. Intuitively, others will look more favorably upon someone who has undertaken a series of activities that earn him or her a certain badge. But does this necessarily imply superior information quality? To answer the question, it is important to dissect the motivation to share from the standpoint of a high badge user and of a low badge user.

To share information or communicate with peers, social media users have to incur time and effort costs. And by posting it, they give up their information advantage for publicizing private information. So initially, sharing information seems to benefit everyone else but sharers. Obviously, these costs have to be justified. Wasko (2005) drew from prior research on collective action and summarized the socialization-related motivations for online sharing (reputation, enjoying helping, centrality in community, tenure in the field, commitment to the community, and reciprocity.

Besides the socialization-related motivation, the finance literature also documented economic-related motivations for online sharing. Message board viewers’ reading and trading can have price impact and expedite the convergence of market prices to
what the sharer perceived to be fair. Because informed investors may not have the financial power to reap all the value conveyed in their private information, they have to stimulate other investors to move the market to the desired direction (Tumarkin and Whitelaw 2001). Informed traders also benefit from constructive feedback, complementary information, and confidence in trading while communicating with their peers (Gray and Kern, 2011).

Based on the theories, we argue that online sharing activities are mainly motivated by socialization-related factors and economic-related factors. Low-rank users are less likely motivated by socialization-related factors judging from their infrequent online activities. As a result, their sharing behaviors must be driven by economic-related factors, which suggests the superior quality of their posts. In contrast, high-rank social media users engage in social media activities not only for economic-related reasons but also for the purpose of socialization. The result is a higher probability of irrelevant and off-topic messages such as greetings. So we expect better information quality from low-rank social media users.

Our prediction is also supported by the Drive-Reduction Theory (Dewey, 2007). The motivation drops after the goal is reached. In most cases, the social media incentive hierarchy ranks are permanently offered when they are obtained. And then the users will lose the incentives to keep sharing quality content. Conversely, social media users who value and respect high-level badges but are currently at a low rank must have a stronger incentive to share quality information. So it is expected that social media users with low-level badges tend to share higher quality information than social media users with high-level badges. To measure the quality of social media discussion messages, I observe the association between the social media users’ sentiment and the future market movements. More details are

In light of the explanations centered around sharing motivations and the driven reduction theory, we propose our first hypothesis:

**Hypothesis 1**: The quality of social media discussion messages is negatively associated with the author’s incentive hierarchy.

Though we have shown that messages from high-rank social media users are less informative, they exert a more significant influence among peers in the online communities. Firstly, high-rank users are expected to have more connections with other users due to their active participation in online social interactions. Social network theories suggested that the number of social connections plays an important role in speeding up the information diffusion (Brown and Reingen, 1987). In online communities, weak ties play an important role in the dissemination
of novel information due to their sheer quantity (Bakshy et al., 2012). In Bitcointalk.org, most ties are weak ties. So messages written by high-rank users who have established many social ties will diffuse faster among the social network than messages written by low-rank users. Furthermore, the badge awarded to a user communicates that the user's past accomplishments and experience and other users can use it to infer the trustworthiness and reliability of the content (Antin and Churchill, 2011). For a given message, if it is posted by a user displaying a high-level badge, it will become more attractive and draw more attention. Based on the arguments above, we propose our second hypothesis:

**Hypothesis 2:** The spillover effect of the social media discussion messages is positively associated with the author’s incentive hierarchy.

This research mainly contributes to the social media incentive hierarchy literature by studying how the incentive hierarchy system shapes users’ motivation to contribute in the online communities and suggesting methods to infer the information quality and the spillover effect based on the user ranks obtained from the social media. This study also contributes to the spillover effect literature by studying the information spillover through social media. Most of the related studies focus on a single event at a time (such as bankruptcy). Such events include bankruptcy (Ferris et al., 1997, Helwege and Zhang, 2015, Lang and Stulz, 1992), IPO announcements (Hsu et al., 2010), new product introductions (Chen et al., 2005), merger announcements (Akhigbe and Martin, 2000), dividend-related announcements (Laux et al., 1998, Slovin et al., 1999), privatization announcements (Otchere, 2007), layoff announcements (Goins and Gruca, 2008), stock split announcements (Tawatnuntachai and D'Mello, 2002), going-concern audit opinions (Elliott et al., 2006), and stock price surprises (Akhigbe et al., 2015), etc. However, in recent years, besides the major shocks that rarely happen, a comprehensive mixture of business information is transmitted through social media at a much higher frequency. It is necessary to extend the related literature to include the information spillover through the social media platforms.

We also directly contribute to the cryptocurrency literature. Two streams of studies exist in this area. First, the technical aspects of cryptocurrency are investigated. Examples include mining (Li et al. 2019), blockchain (Hawlitschek et al. 2018; Saberi et al. 2019; Francisco and Swanson 2018), smart contract (Gatteschi et al. 2018), and security issues (Gao et al. 2018; Conti et al. 2018; Kim and Lee 2018). Our paper falls into the other category where cryptocurrency market dynamics are studied. Omane-Adjepong and Alagidede (2019) examined the volatility spillover among different cryptocurrencies and found that the diversification benefits for only short term investment. Mills and Nower (2019) used an online survey to show
that cryptocurrency investment is usually associated with a tendency to gamble. Antonakakis et al. (2019) studied the co-movement of the cryptocurrency and found that market volatility increases with market co-movement. Caporale et al. (2018) examined the correlation between the past cryptocurrency market values and the future cryptocurrency market values and found a positive correlation. They claimed that such correlation presents evidence of market inefficiency. Bouri et al. (2018) focused on the co-explosivity (co-occurrence of price spikes) of the cryptocurrency market and found that the co-explosivity exists regardless of the market maturity. Our paper contributes to this literature by examining the interaction between the online community and the cryptocurrency market.

This research also provides practical implications. We demonstrate that social media incentive hierarchy systems can be used to sort out the valuable investment advice within an enormous amount of social data generated each day. Since there is no guarantee for the quality of the information shared on social media due to its unregulated nature, our insights will help investors narrow down the search for high-quality social media contents, reduce the information acquisition cost, and improve the quality of the investment decision.

DATA

Bitcoin

This section presents a brief introduction to the Bitcoin market and the related data used in the study. Bitcoin is a decentralized peer-to-peer electronic payment platform. It is a web-based system that enables users to transfer values across the globe quickly and anonymously without the need for third-party verifications.

Bitcoin has seen significant growth since it was created. The market capitalization is valued at around 186 billion US dollars at the time of writing. An increasing number of businesses have accepted Bitcoin as a payment method including many industry-leading corporations such as Microsoft, Expedia, Newegg, Tesla, Home Depot, etc.

We collected Bitcoin price data from Poloniex.com. Poloniex is a major “foreign exchange” between Bitcoin and many other fiat currencies. Though it is not the largest Bitcoin-USD exchange, it runs many Bitcoin-Altcoin markets (“Altcoin” is usually used to refer other non-Bitcoin cryptocurrencies) and provides public access to the historical price information. Similar to foreign exchange markets, these markets are active 24 hours a day, and seven days a week.
The Bitcoin prices used in the analyses are the 24:00 o'clock price each day (the daily close price). All timestamps are based on GMT (Greenwich Mean Time). The day $t$ Bitcoin return is calculated as $(P_t - P_{t-1})/P_{t-1}$, where $P_t$ is the Bitcoin close price on day $t$. The data spans from 2015/2/19 to 2017/2/17. The date 2015/2/19 is chosen as the start date because it is the earliest trading data on Poloniex. Panel A of Table 1 presents the descriptive statistics on Bitcoin-related variables. The aggregated daily sentiment is calculated as the total number of negative words divided by the total number of words in all Bitcoin-related posts within a particular day. We will explain the sentiment calculation in more detail in section 3.4. The # Post is the number of Bitcoin-related posts within a particular day.

**Table 1. Descriptive statistics**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
<th>Min</th>
<th>Std. Dev</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Daily Bitcoin Return</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily Sentiment</td>
<td>0.014</td>
<td>0.014</td>
<td>0.024</td>
<td>0.007</td>
<td>0.003</td>
<td>730</td>
</tr>
<tr>
<td>Daily Return</td>
<td>0.36%</td>
<td>0.293%</td>
<td>18.66%</td>
<td>-31.89%</td>
<td>3.30%</td>
<td>730</td>
</tr>
<tr>
<td>Close Price</td>
<td>467.092</td>
<td>421.782</td>
<td>1136</td>
<td>178.719</td>
<td>218.656</td>
<td>730</td>
</tr>
<tr>
<td># Post</td>
<td>256.893</td>
<td>234</td>
<td>1211</td>
<td>4</td>
<td>136.761</td>
<td>730</td>
</tr>
<tr>
<td><strong>Panel B: Altcoin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily Sentiment</td>
<td>0.012</td>
<td>0.008</td>
<td>1</td>
<td>0</td>
<td>0.040</td>
<td>33,083</td>
</tr>
<tr>
<td>Daily Return</td>
<td>1.40%</td>
<td>-0.11%</td>
<td>2684.06%</td>
<td>-99.99%</td>
<td>24.07%</td>
<td>20,882</td>
</tr>
<tr>
<td># Post</td>
<td>18.837</td>
<td>5</td>
<td>2,160</td>
<td>0</td>
<td>54.052</td>
<td>33,083</td>
</tr>
<tr>
<td># Author</td>
<td>8.913</td>
<td>4</td>
<td>446</td>
<td>0</td>
<td>16.020</td>
<td>33,083</td>
</tr>
<tr>
<td><strong>Panel C: Bitcoin Thread-Day Return</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thread-Day Sentiment</td>
<td>0.013</td>
<td>0.010</td>
<td>1</td>
<td>0</td>
<td>0.016</td>
<td>28,194</td>
</tr>
<tr>
<td># Post</td>
<td>6.882</td>
<td>4</td>
<td>232</td>
<td>1</td>
<td>9.251</td>
<td>28,194</td>
</tr>
</tbody>
</table>
Altcoin

The term “Altcoin” stands for “alternative to Bitcoin” and describes any cryptocurrency that is not a Bitcoin. Most Altcoins share similar technology as Bitcoin, but they usually have a different monetary policy such as currency issuance rules, transaction confirmation methods, and mining methods, etc. They can be treated as intra-industry competitors to Bitcoin because of the technical similarity.

Bitcoin, the earliest cryptocurrency in the market, was created in 2009. Starting in 2014, the development of Altcoins flourished. A large number of Altcoins suddenly emerged. While many of them soon went out of the market due to extremely inactive trading, many of them survived and grew rapidly in market capitalization and attracted significant public attention. Though there were thousands of Altcoins in active trading, attention is limited to those major competitors listed in the Poloniex exchange. Similarly, the data spans from 2015/2/19 to 2017/2/17. All timestamps are based on GMT (Greenwich Mean Time). Panel B of Table 1 presents the descriptive statistics on Altcoin-related variables. The discussion about each Altcoin in Bitcointalk.org is arranged within a thread. The Altcoin aggregated daily sentiment is calculated as the total number of negative words divided by the total number of words in all Altcoin-related posts within a particular day. The # Posts and # Authors are the number of Altcoin posts and the number of distinct posting users within a particular day.

The average daily sentiment measured by the percentage of negative words is around 1% for both Bitcoin-related discussions and Altcoin-related discussions. It may appear very small at first, but this observation is echoed by other related studies. For example, Chen et al. (2014) reported that the average negative word percentage in all Seeking Alpha comments to be 1.2%, very close to our observation.

Social Media Discussion

Social media discussion data is downloaded for the sentiment calculation. Both Bitcoin-related discussions and Altcoin-related discussions are downloaded from Bitcointalk.org. Bitcointalk.org is a leading message board for cryptocurrency investors to share thoughts on various topics. By the time of this writing, Bitcointalk.org has accumulated 2,650,061 registered users and reached an average daily page view of 1,346,940. It receives on average 7,367 posts each day.

There are 248 discussion boards on Bitcointalk.org. Most of them are dedicated to Bitcoin-related discussions, but not all of them are directly related to Bitcoin price
discovery. To avoid noise information in the analysis, we use the messages from the “Speculation” discussion board, which is only second to the largest general Bitcoin discussion boards in terms of posting volume. Besides Bitcoin-related discussions, Bitcointalk.org also provided places for Altcoin discussions. The largest and the most popular board in terms of post volume is the “Altcoin Announcement” discussion board. It may seem ironic at first that the most popular discussion board on Bitcointalk.org is about Altcoin. This is due to the large number of Altcoins being discussed on the forum. Within the “Altcoin Announcement” board, new Altcoins are announced with a new thread, and the title of the thread follows a fixed format that can be used to identify the Altcoin uniquely (like a ticker symbol in the stock market). All discussions about that Altcoin is posted under that thread. The discussion threads for the 66 actively traded Altcoins listed on Poloniex are located and over 600,000 messages posted for the 66 Altcoins are downloaded.

**Extracting Social Media Discussion Sentiment**

This study follows the literature and quantifies the sentiment expressed in the communications by calculating the percentage of negative words in the messages (Chen *et al.*, 2014, Loughran and McDonald, 2011, Tetlock, 2007, Tetlock *et al.*, 2008). In early studies, General Inquirer’s Harvard-IV-4 classification dictionary (Harvard-IV-4 TagNeg) is used to identify the occurrence of negative words. However, Loughran and McDonald (2011) argued that the Harvard-IV-4 TagNeg substantially misclassifies words when gauging tones in financial applications and created a new lexicon containing words that typically have negative implications in a financial context. This study adopts this lexicon developed by (Loughran and McDonald, 2011) in the study to identify negative words. The sentiment of a discussion network in a day is calculated as the ratio of the total number of negative words to the total number of words in all related posts.

We did not consider the percentage of positive words because there are far fewer positive words in the positive lexicon designed by Loughran and McDonald (2011). Many posts will be assigned a sentiment of zero if we use the percent of positive word to measure the sentiment.

**Incentive Hierarchy System in Bitcointalk.org**

Bitcointalk.org employs a simple activity-based incentive hierarchy system. The purpose of introducing this system is to encourage user activity. Similar incentive hierarchy systems have been deployed in many other social media platforms. For Bitcointalk.org users, the formula used to calculate their activity points is shown as follows:
activity = \min (time \times 14, \text{number of posts}) \quad (1)

The parameter time is the number of two-week periods when the user is active since registration. From the formula, we know that to get high activity points, the user must be (1) posting many messages, and (2) remain active for a long period of time. Though the method to calculate the user points differs on different sites, the basic principle is mostly the same.

Based on the activity scores, the users are awarded eight badges of different levels by Bitcointalk.org. They are Brand New, Newbie, Jr. Member, Member, Full Member, Sr. Member, Hero Member, and legendary (from the lowest level to the highest level). Figure 1 illustrates the discussions on this message board. The badges of users are highlighted in red boxes. In this study, a user is recognized as a high-rank user if he or she possesses the Full Member badge or better. Otherwise, the user is recognized as a low-rank user.

Figure 1. Illustration of user badges on Bitcointalk.org

Please note that the badges in the data are observed at the end of the data collection period, and it is not the badge the users were holding at the time when they posted
the message. The badge at the time of the post is reverse engineered with the formula provided on Bitcointalk.org (Equation (1)) to locate those users who must be a high-rank user or low-rank user at the time of post.

However, the badges the users were holding at the end of the data collection period may represent the user’s natural willingness to engage in online social activities. So as a robustness check, we also group high-rank and low-rank users based on their badge at the end of the data collection period.

**EMPIRICAL ANALYSIS**

*Incentive Hierarchy and Prediction Accuracy, Evidence from the Altcoin Markets*

Though Altcoin and Bitcoin are very similar technologies, we decide to test our predictions in the Altcoin market and the Bitcoin market separately for the following consideration. The Bitcoin discussion board selected (the speculation discussion board) is expected to contain the most relevant information for the Bitcoin price movement, while many other Bitcoin discussion boards are less relevant or completely off-topic (such as the technical support board and the project development board). However, the Altcoin-related discussions are not categorized into different discussion boards. All the discussions are pooled together in one thread. Therefore, the overall prediction accuracy in the Altcoin markets is expected to be lower.

This section focuses on next-day price prediction for the 66 Altcoins in the sample. A fixed-effect linear model with each Altcoin as a cross-section is used to test hypothesis 1. The \( t+1 \) return is regressed on the sentiment measures and other control variables during time \( t \). The analysis is conducted using the following model:

\[
R_{i,t+1} = \alpha + \beta_1 HSentiment_{i,t} + \beta_2 LSentiment_{i,t} + \delta X_{i,t} + \alpha t + \eta_{i,t}
\]

(2)

The dependent variable \( R_{i,t+1} \) is the time \( t+1 \) return for altcoin \( i \), \( HSentiment_{i,t} \) is the daily aggregate sentiment extracted from social media discussions posted by high-rank users for Altcoin \( i \) at time \( t \). \( LSentiment_{i,t} \) is the daily aggregate sentiment extracted from social media discussions posted by low-rank users for Altcoin \( i \) at time \( t \).
The eight levels of badges awarded by Bitcontalk.org are Brand New, Newbie, Jr. Member, Member, Full Member, Sr. Member, Hero Member and legendary (from low-level badge to high-level badge). The high-rank user group threshold is Full Member or Above (120 activity points or more). The coefficient estimates for $\beta_1$ and $\beta_2$ reflect the effect of high-rank user messages and low-rank user messages on the next-day return respectively. The time dummy $\alpha_t$ (week dummy) controls for the differences in the returns in different time periods. The Altcoin dummy $i$ controls for the Altcoin-specific fixed effect. X contains the returns for Altcoin $i$ and Bitcoin at time $t$ ($ALTR_{i,t}$ and $BTCR_{i,t}$), the one-day lagged returns for Altcoin $i$ and Bitcoin ($ALTR_{i,t-1}$ and $BTCR_{i,t-1}$), the two-day lagged returns for Altcoin $i$ and Bitcoin ($ALTR_{i,t-2}$ and $BTCR_{i,t-2}$), the logarithm of the time $t$ post count for Altcoin $i$ $Log(AltPostCount)_{i,t}$, the logarithm of the time $t$ author count for Altcoin $i$ $Log(AltAuthorCount)_{i,t}$ (the author count is the number of distinct users who participated in the discussion at time $t$), and weekly market capitalization share $MarketCapShare_{i,t}$ ranging from 0 to 1. $Log(AltPostCount)_{i,t}$ and $Log(AltAuthorCount)_{i,t}$ are used to control for the popularity of the discussion.

Hausman test is conducted to verify the choice of the fixed-effect model, however, the Hausman test result doesn’t reject the use of the random effect model. Therefore, the random effect model is also used as one of the robustness checks. The estimation result of Equation (1) is shown in Table 2.

The first column of Table 2 shows the prediction accuracy of the combined sentiments (sentiments from both high-level users and low-level users). The coefficient estimate for CombinedSentiment$_{i,t}$ is not statistically significant, indicating noisy overall information. In Column (2) to Column (5) of Table 2, the badge at the time of the post is used to categorize users into high-rank or low-rank user groups. The coefficient estimates of HSentiment$_{i,t}$ is not statistically significant (Column 2), meaning that the high-rank users fail to offer value-relevant information for future return prediction. However, the coefficient estimates of LSentiment$_{i,t}$ is consistently negative and statistically significant at the 5% level, meaning that the higher the percentage of the negative words in low-rank users’ messages, the lower the next-day return. More specifically, if there are 1% more negative words in the posts from the low-rank users regarding Altcoin $i$, the next-day Altcoin $i$ return will be around 0.19% lower.

Column (5) to Column (7) in Table 2 tells a similar story when the badge observed at the end of the data collection period is used to categorize users into high-rank or low-rank users. In Column (8) of Table 2, a random effect model with clustered standard errors (error terms are clustered over Altcoins) is used as a robustness
check. The coefficient estimate on $L_{Sentiment_{i,t}}$ is -0.192, significant at a 1% level, very consistent with the fixed-effect model results.

In contrast, the coefficient estimates on $H_{Sentiment_{i,t}}$ are not statistically significant across all model specifications. These results support our first prediction that the low-rank social media users provide a better prediction for the future price movement.
### Table 2. Predictive power of social media users with different ranks-Altcoin

<table>
<thead>
<tr>
<th>Model</th>
<th>$R_{i,t+1}$ (1)</th>
<th>$R_{i,t+1}$ (2)</th>
<th>$R_{i,t+1}$ (3)</th>
<th>$R_{i,t+1}$ (4)</th>
<th>$R_{i,t+1}$ (5)</th>
<th>$R_{i,t+1}$ (6)</th>
<th>$R_{i,t+1}$ (7)</th>
<th>$R_{i,t+1}$ (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users Grouped Based on</td>
<td>CombinedSentiment$_{i,t}$</td>
<td>0.006 (0.13)</td>
<td>0.008 (0.08)</td>
<td>0.014 (0.14)</td>
<td>0.165 (1.40)</td>
<td>0.177 (1.50)</td>
<td>0.0003</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>HSentiment$_{i,t}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.190** (-2.04)</td>
<td>-0.190** (-2.04)</td>
<td>-0.149** (-1.99)</td>
<td>-0.155** (-2.06)</td>
<td>0.192*** (2.88)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSentiment$_{i,t}$</td>
<td>-0.037*** (-5.44)</td>
<td>-0.026** (-2.49)</td>
<td>-0.026** (-2.55)</td>
<td>-0.027*** (-3.14)</td>
<td>-0.031** (-3.21)</td>
<td>-0.028*** (-3.17)</td>
<td>-0.020</td>
</tr>
<tr>
<td></td>
<td>ALTR$_{i,t}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALTR$_{i,t-1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALTR$_{i,t-2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BTCR$_{i,t}$</td>
<td>-0.303*** (-5.42)</td>
<td>-0.327*** (-4.99)</td>
<td>-0.330*** (-5.03)</td>
<td>-0.312*** (-5.84)</td>
<td>-0.315*** (-5.90)</td>
<td>-0.314*** (-5.88)</td>
<td>-0.335*** (-7.25)</td>
</tr>
<tr>
<td></td>
<td>BTCR$_{i,t-1}$</td>
<td>-0.169*** (-2.98)</td>
<td>-0.204*** (-3.11)</td>
<td>-0.205*** (-3.14)</td>
<td>-0.192*** (-3.60)</td>
<td>-0.192*** (-3.61)</td>
<td>-0.193*** (-3.64)</td>
<td>-0.198*** (-3.94)</td>
</tr>
<tr>
<td></td>
<td>BTCR$_{i,t-2}$</td>
<td>-0.141*** (-2.49)</td>
<td>-0.189*** (-2.93)</td>
<td>-0.189*** (-2.94)</td>
<td>-0.167*** (-3.14)</td>
<td>-0.166*** (-3.11)</td>
<td>-0.167*** (-3.14)</td>
<td>-0.193*** (-3.10)</td>
</tr>
<tr>
<td></td>
<td>Log(PostCount$_{i,t}$)</td>
<td>0.012** (1.82)</td>
<td>-0.002 (-0.33)</td>
<td>-0.002 (-0.28)</td>
<td>0.0001 (0.02)</td>
<td>0.001 (0.11)</td>
<td>0.0004 (0.06)</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Log(AuthorCount$_{i,t}$)</td>
<td>-0.014* (-1.70)</td>
<td>0.0008 (0.08)</td>
<td>0.0004 (0.05)</td>
<td>0.0005 (0.05)</td>
<td>-0.003 (-0.40)</td>
<td>-0.004 (-0.45)</td>
<td>-0.003 (-0.43)</td>
</tr>
<tr>
<td></td>
<td>MarketCapShare$_{i,t}$</td>
<td>-0.317 -0.719*** (-1.09)</td>
<td>-0.715*** (-2.85)</td>
<td>-0.715*** (-2.83)</td>
<td>-0.668*** (-3.00)</td>
<td>-0.659*** (-2.96)</td>
<td>-0.666*** (-3.00)</td>
<td>-0.666*** (-0.63)</td>
</tr>
<tr>
<td></td>
<td>WeekDummy</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

Note: *** p<0.01, ** p<0.05, * p<0.1.
**Incentive Hierarchy and Prediction Accuracy, Evidence from the Bitcoin Markets**

As mentioned before at the start of section 4.1, the discussions from Altcoin-related threads are not categorized into different topics, which means that we may have included discussions from irrelevant topics. In contrast, the Bitcoin-related discussions are categorized into different discussion boards, and the message board we selected (speculation discussion board) should contain more relevant information for Bitcoin pricing. In the following section, we show further evidence for our hypothesis 1 using Bitcoin-related social media discussions.

Bitcoin-related discussions from 3,372 different threads in the speculation board from 2015/2/19 to 2017/2/17 are collected. On Bitcointalk.org, every registered user is allowed to start a new discussion thread and wait for others to join the discussion (post on this thread). Different from the first analysis in section 4.1, this analysis treats each discussion thread as a cross-section when constructing the panel dataset (the unit of observation is the collection of all messages in a particular thread \( i \) within a given day \( t \)). We switch to this panel specification because now we are only dealing with the Bitcoin, using cryptocurrency types as the panel variable is no longer possible. Following this method, a panel dataset of 3,372 cross-sections and 57,063 individual observations is generated. This analysis compares high_rank users to low-rank users in terms of prediction accuracy using the following model specification:

\[
R_{t+1} = \alpha + \alpha_t + \beta_1 HSentiment_{it} + \beta_2 LSentiment_{it} + \delta X + \eta_{it}
\]  

(3)

In Equation (3), \( i \) is the thread index. \( HSentiment_{it} \) is the aggregate sentiment extracted from the daily discussions posted by high-rank users at time \( t \) in thread \( i \). \( LSentiment_{it} \) is the aggregate daily discussion sentiment from messages posted by low-rank users at time \( t \) in thread \( i \). The time dummy \( \alpha_t \) (weekly dummy) controls for the differences in the returns in different time periods. \( X \) contains the intraday return \( R_t \), the one-day lagged return \( R_{t-1} \), the two-day lagged return \( R_{t-2} \), and the logarithm of thread-day post count \( \log(PostCount_t) \).

A random-effects model is chosen over a fixed-effects model because the unobserved disturbance for each cross-section (thread) is more likely to be random rather than fixed across different time periods. First, the group of people participating in the discussion on a particular thread keeps changing every day. This leads to changes in their collective wisdom as well. Second, the focuses of the same thread also change over time. As new information emerges, discussions also evolve...
and move from one topic to another. As a result, the unobserved impact of the thread on the dependent variable (future price movement) is not constant over time, and it is not appropriate to represent these unobserved disturbances with fixed effects.

In addition, we also limit our attention only to large enough daily discussions because a thread receiving very few posts during a certain day implies uninterested or obsolete discussion topics. Only large enough daily discussions within a certain thread (with the number of posts greater than 10, 15, or 20) are considered. We use three different thresholds for robustness checks.

The estimation results are presented in Table 3. The coefficient estimates for $HS_{Sentiment,t}$ are not statistically significant across all model specifications. This is consistent with the results in the previous section 4.1 that high-level badge users fail to provide value-relevant information for the future price prediction. In contrast, the coefficient estimates for $LS_{Sentiment,t}$ are negative and statistically significant at least at a 5% level in all six model specifications, meaning that a high percentage of negative words (the lower the sentiment) in social media discussions predicts lower next-day Bitcoin returns. The predictive power (captured by the negative coefficient estimates for $LS_{Sentiment,t}$) increases with more posts. This observation is consistent with the argument that larger daily discussions contain more value-relevant information.

Both the evidence from the Altcoin market (section 4.1) and the evidence from the Bitcoin market (section 4.2) point to superior predictive power from low-rank users, providing consistent support for our hypothesis 1.

**Table 3. Predictive power of social media users with different incentive hierarchy rank-bitcoin**

<table>
<thead>
<tr>
<th>Users Grouped Based on</th>
<th>$R_{t+1}$ $(1)$</th>
<th>$R_{t+1}$ $(2)$</th>
<th>$R_{t+1}$ $(3)$</th>
<th>$R_{t+1}$ $(4)$</th>
<th>$R_{t+1}$ $(5)$</th>
<th>$R_{t+1}$ $(6)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PostCount</td>
<td>&gt; 15</td>
<td>&gt; 20</td>
<td>&gt; 25</td>
<td>&gt; 15</td>
<td>&gt; 20</td>
<td>&gt; 25</td>
</tr>
<tr>
<td>$HS_{Sentiment,t}$</td>
<td>0.010</td>
<td>0.016</td>
<td>-0.103</td>
<td>0.002</td>
<td>-0.004</td>
<td>-0.136</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.15)</td>
<td>(-0.61)</td>
<td>(0.02)</td>
<td>(-0.02)</td>
<td>(-0.59)</td>
</tr>
<tr>
<td>$LS_{Sentiment,t}$</td>
<td>-0.088**</td>
<td>-0.145***</td>
<td>-0.206***</td>
<td>-0.083**</td>
<td>-0.142***</td>
<td>-0.203***</td>
</tr>
<tr>
<td></td>
<td>(-2.13)</td>
<td>(-3.76)</td>
<td>(-4.31)</td>
<td>(-1.96)</td>
<td>(-3.62)</td>
<td>(-4.18)</td>
</tr>
<tr>
<td>$BTC_{R,t}$</td>
<td>-0.295***</td>
<td>-0.268***</td>
<td>-0.261***</td>
<td>-0.296***</td>
<td>-0.265***</td>
<td>-0.255***</td>
</tr>
<tr>
<td></td>
<td>(-9.09)</td>
<td>(-6.14)</td>
<td>(-3.92)</td>
<td>(-9.27)</td>
<td>(-6.09)</td>
<td>(-3.80)</td>
</tr>
<tr>
<td>$BTC_{R_{t-1}}$</td>
<td>-0.326***</td>
<td>-0.344***</td>
<td>-0.337***</td>
<td>-0.323***</td>
<td>-0.337***</td>
<td>-0.327***</td>
</tr>
<tr>
<td></td>
<td>(-12.51)</td>
<td>(-10.17)</td>
<td>(-6.45)</td>
<td>(-12.62)</td>
<td>(-9.99)</td>
<td>(-6.21)</td>
</tr>
</tbody>
</table>
The Effect of Incentive Hierarchy System of Social Media in the Delivery of Quality Information

P. Xie

The implication of Incentive Hierarchy on the Spillover Effect

This section investigates the implications of the incentive hierarchy system in the social media spillover effects. The finance literature has well documented the phenomenon that new information about a focal firm can spill over to its intra-industry rivals. In this section, we examine if this phenomenon extends to the information transmitted through social media. And we test our hypothesis 2 that if the spillover effect is mainly caused by the high-rank users due to their greater visibility in the online community.

Specifically, we empirically study if the information contained in the Bitcoin-related discussions spills over to the Altcoin markets and how the spillover effect differs across user groups. The analysis is organized around the following model specification:

\[
R_{i,t+1} = \alpha + \beta_1 AltHSentiment_{i,t} + \beta_2 AltLSentiment_{i,t} + \beta_3 BtcHSentiment_t + \beta_4 BtcLSentiment_t + \delta X + \epsilon_t
\]  

(4)

The dependent variable \( R_{i,t+1} \) is the next-day return for Altcoin \( i \), \( AltHSentiment_{i,t} \) is the aggregated sentiment from high-rank users writing for Altcoin \( i \) during time \( t \). \( AltLSentiment_{i,t} \) is the aggregated sentiment from low-rank users writing for Altcoin \( i \) during time \( t \). Similarly, \( BtcHSentiment_t \) is the time \( t \) aggregated sentiment from high-rank users in the Bitcoin discussion board, and \( BtcLSentiment_t \) is the time \( t \) aggregated sentiment from low-rank users in the Bitcoin discussion board.

If our predictions are correct, the coefficient estimate for \( \beta_3 \) should be statistically significant. A negative \( \beta_3 \) indicates a stronger contagion effect, meaning that when bad news strikes the Bitcoin market (more negative words about Bitcoin), Altcoin prices will also decrease. While a positive \( \beta_3 \) indicates stronger competition effect, meaning that the Altcoin prices will increase after their major competitor Bitcoin suffers from bad news.
Table 4 presents the results. In Column (1) to Column (3) of Table 4, the social media users’ badge at the time of the post is used to categorize them into the high-rank user and the low-rank user. Again, users with the Full Member badge and above are recognized as high-rank. In Column (1) of Table 4, the coefficient estimate of $BtcHSentiment_t$ is negative and statistically significant at the 5% level, meaning that the Bitcoin-related social media messages posted by high-rank users exert spillover effects. Specifically, when the percentage of negative words in all Bitcoin-related discussions posted by high-rank users increases by 1% during day $t$, the $t+1$ Altcoin return will increase by 1.39% on average. This is evidence of the competition effect in the cryptocurrency market. In contrast, the coefficient estimate of $BtcLSentiment_t$ in Column (1) is not statistically significant, meaning that the Bitcoin-related social media messages posted by low-rank users do not exert spillover effects.
Table 4. Spillover effect from the Bitcoin market to the Altcoin market

<table>
<thead>
<tr>
<th>Badge Used</th>
<th>( R_{t+1} ) (1)</th>
<th>( R_{t+1} ) (2)</th>
<th>( R_{t+1} ) (3)</th>
<th>( R_{t+1} ) (4)</th>
<th>( R_{t+1} ) (5)</th>
<th>( R_{t+1} ) (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Social Media Activity</td>
<td>No</td>
<td># Authors Above Median</td>
<td># Posts Above Median</td>
<td>No</td>
<td># Authors Above Median</td>
<td># Posts Above Median</td>
</tr>
<tr>
<td>AltHSentiment(_{i,t})</td>
<td>0.176 (1.49)</td>
<td>-0.002 (-0.01)</td>
<td>-0.230 (-1.03)</td>
<td>0.177 (1.49)</td>
<td>-0.001 (-0.00)</td>
<td>-0.232 (-0.04)</td>
</tr>
<tr>
<td>AltLSentiment(_{i,t})</td>
<td>-0.153** (-2.03)</td>
<td>-0.156* (-1.69)</td>
<td>-0.157* (-1.72)</td>
<td>-0.154** (-2.05)</td>
<td>-0.157* (-1.70)</td>
<td>-0.155* (-1.71)</td>
</tr>
<tr>
<td>AltHSentiment(_{i,t})</td>
<td>1.397** (2.15)</td>
<td>1.693** (2.25)</td>
<td>1.608** (1.71)</td>
<td>1.425* (2.81)</td>
<td>2.591*** (2.95)</td>
<td>2.296*** (2.95)</td>
</tr>
<tr>
<td>AltLSentiment(_{i,t})</td>
<td>-0.214 (-0.31)</td>
<td>0.898 (1.19)</td>
<td>0.687 (0.93)</td>
<td>-0.123 (-0.48)</td>
<td>-0.080 (0.377)</td>
<td>0.377 (0.64)</td>
</tr>
<tr>
<td>ALTR(_{i,t})</td>
<td>-0.027*** (-3.16)</td>
<td>-0.026*** (-2.95)</td>
<td>-0.032*** (-3.77)</td>
<td>-0.028*** (-3.16)</td>
<td>-0.026*** (-2.96)</td>
<td>-0.032*** (-3.75)</td>
</tr>
<tr>
<td>ALTR(_{i,t-1})</td>
<td>-0.018*** (-4.01)</td>
<td>-0.016*** (-3.81)</td>
<td>-0.017*** (-3.97)</td>
<td>-0.018*** (-3.99)</td>
<td>-0.016*** (-3.79)</td>
<td>-0.017*** (-3.96)</td>
</tr>
<tr>
<td>ALTR(_{i,t-2})</td>
<td>-0.013*** (-2.86)</td>
<td>-0.013*** (-2.92)</td>
<td>-0.013*** (-2.86)</td>
<td>-0.014*** (-2.89)</td>
<td>-0.013*** (-2.95)</td>
<td>-0.013*** (-2.86)</td>
</tr>
<tr>
<td>BtCR(_{i,t})</td>
<td>-0.309*** (-5.78)</td>
<td>-0.309*** (-5.29)</td>
<td>-0.254*** (-4.40)</td>
<td>-0.312*** (-5.84)</td>
<td>-0.312*** (-5.34)</td>
<td>-0.254*** (-4.40)</td>
</tr>
<tr>
<td>BtCR(_{i,t-1})</td>
<td>-0.186*** (-3.49)</td>
<td>-0.154*** (-2.65)</td>
<td>-0.139** (-2.39)</td>
<td>-0.187*** (-3.52)</td>
<td>-0.153*** (-2.63)</td>
<td>-0.138*** (-2.38)</td>
</tr>
<tr>
<td>BtCR(_{i,t-2})</td>
<td>-0.164*** (-3.10)</td>
<td>-0.184*** (-3.19)</td>
<td>-0.170*** (-2.99)</td>
<td>-0.166*** (-3.12)</td>
<td>-0.185*** (-3.21)</td>
<td>-0.171*** (-2.99)</td>
</tr>
<tr>
<td>Log(PostCount(_{i,t}))</td>
<td>0.0006 (-0.09)</td>
<td>0.0009 (-0.13)</td>
<td>0.0018 (0.30)</td>
<td>0.0004 (0.08)</td>
<td>-0.0009 (0.14)</td>
<td>0.002 (0.32)</td>
</tr>
<tr>
<td>Log(AuthorCount(_{i,t}))</td>
<td>-0.004 (-0.45)</td>
<td>-0.0002 (-0.03)</td>
<td>-0.0060 (-0.73)</td>
<td>-0.003 (-0.43)</td>
<td>-0.002 (-0.02)</td>
<td>-0.006 (-0.76)</td>
</tr>
<tr>
<td>MarketCapShare(_{i,t})</td>
<td>-0.667*** (-3.00)</td>
<td>-0.717*** (-3.36)</td>
<td>-0.618*** (-2.87)</td>
<td>-0.667*** (-3.00)</td>
<td>-0.720*** (-3.37)</td>
<td>-0.618*** (-2.87)</td>
</tr>
<tr>
<td>WeekDummy</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

Note: *** p<0.01, ** p<0.05, * p<0.1.

In Column (2) and Column (3) of Table 4, the study checks the spillover effect especially when there is a surge in Altcoin social media activities (when the number of authors or the number of posts is greater than the median). Larger \( \beta_3 \) coefficient estimates are observed, which indicates stronger spillover effects. This implies an increased reliance on information spillover from Bitcoin when there is a need for more information.
The same results hold in Column (4) to Column (6) when the study uses social media users’ badge at the end of the data collection period to categorize them into the high-rank user group and the low-rank user group. These results support our hypothesis 2.

CONCLUSION

This paper revisits the literature on social media’s role in the financial market and extends it to the context of cryptocurrency. Different from the traditional stock market setting, these cryptocurrency markets are very speculative due to the lack of fundamental information. Therefore, social media would play a more important role in these markets. By analyzing social media users’ motivation to share private information and drawing from the drive reduction theory, we demonstrate that low-rank users are the primary source of value-relevant information on social media. Empirical evidence in both the Bitcoin market and the Altcoin markets are provided to support our predictions.

Though high-rank users are shown to be less informative, we claim that they exert stronger spillover effects due to their high visibility within the online community. We observed competition effects within the cryptocurrency industry. The bad news shared on the Bitcoin-related message board will spill over to the Altcoin markets, and drives up the Altcoin prices.

According to Cogent Research, One-third of investors are using social media like Facebook, LinkedIn, and company blogs for personal finance and investing (PF&I) purposes. However, there is no guarantee for the quality of the information shared on social media. Our study offers insights for these investors utilizing social media to make trading decisions and suggests a way to potentially filter out low-quality information on social media platforms.

It is worth noting that the superior predictive power from the low-rank users cannot be driven by their superior amount. Though Bitcointalk.org is a major message board for cryptocurrency investment, the average number of posts per day is only around 8000 at the time of the writing. However, there are over 350,000 daily Bitcoin transactions. Even if the 8000 authors are all low-rank users, and even if they all trade during a particular day, their trading is only a small fraction of all trades. The predictive power should come from the information embedded in the messages, but not the trading behaviors of the authors.
Finally, we acknowledge a few limitations in this study. First, the social media we selected for data collection uses an activity-based incentive hierarchy system, and the badges are permanently awarded when the milestones are reached. But nowadays there are other types of incentive hierarchy systems that award badges based on various user behaviors such as the number of followers and the number of likes. In some cases, the badges can be lost if the user does not maintain active participation. Future research may look into these alternative incentive hierarchy systems and investigate the differences. Second, we use the next-day return prediction accuracy to measure the quality of the social media discussion messages, but this method requires social media dedicated to discussing investment opinions. Future research may design other information quality measures and check the robustness of our results.

REFERENCES


Kollock, P. (1999), The economies of online cooperation. Communities in cyberspace, 220.


