

# Twitter Financial Community Modeling using Agent Based Simulation

Steve Y. Yang, Anqi Liu and Sheung Yin Kevin Mo

Financial Engineering Program  
Stevens Institute of Technology  
Hoboken, New Jersey, USA

**Abstract**— With the empirical evidence that Twitter influences the financial market, there is a need for a bottom-up approach focusing on individual Twitter users and their message propagation among a selected Twitter community with regard to the financial market. This paper presents an agent-based simulation framework to model the Twitter network growth and message propagation mechanism in the Twitter financial community. Using the data collected through the Twitter API, the model generates a dynamic community network with message propagation rates by different agent types. The model successfully validates against the empirical characteristics of the Twitter financial community in terms of network demographics and aggregated message propagation pattern. Simulation of the 2013 Associated Press hoax incident demonstrates that removing critical nodes of the network (users with top centrality) dampens the message propagation process linearly and critical node of the highest betweenness centrality has the optimal effect in reducing the spread of the malicious message to lesser ratio of the community.

**Keywords**—*Agent-based model, Twitter, Information diffusion, Financial market.*

## I. INTRODUCTION

Can Twitter influence the financial market? The question has been pursued by academia and industry in recent years with the popularity of social media. On April 2 2013, the Securities and Exchange Commission revised its disclosure rules of public companies that they could reveal corporate news on social media platforms and major financial data vendor Bloomberg soon announced afterwards that they would integrate real-time Twitter feeds into their data terminals that are currently serving a wide investment community [1] [2]. Although Twitter serves as an excellent platform in facilitating news propagation, it also draws major criticism regarding the potential threats towards the financial market. The 2013 Associated Press hoax incident was an example of how Twitter could directly influence the market by propagating falsehood news among the investment community. Major regulatory agencies, such as the Commodity Futures Trading Commission, have expressed heightened concerns over the effect of social media on domains such as high-frequency trading and how the threat could be mitigated by the use of safeguards in place to protect the financial market [3].

Empirical evidence has shown that Twitter can predict the financial market by tracking statistics on tweeting activity and message sentiment. These approaches enable the top-down perspective of understanding how selective Twitter indicators can lead to identifiable effects to the financial market. The motivation of this study is to take the bottom-up approach toward modeling the individual behavior of Twitter users and how their static and dynamic interaction form emerging

patterns related to the financial system. Through an effort to establish closer linkage between the financial market and the Twitter universe, Yang et al. proposes a Twitter financial community whose users' interest align with the financial market [4]. The underlying intuition is that the financial community contains more refined and precise information related to the financial market, and therefore its influence to market movement can be better captured [4].

The objective of this paper is to establish a framework for modeling the population growth and message propagation mechanism of the Twitter financial community using agent-based simulation. We first create classes of agents to represent different types of social media participants and define simulation topology related to agent behaviors based on empirical statistics drawn from the Twitter financial community. The agent-based model is then validated against the empirical characteristics of the Twitter financial community in terms of network demographics and aggregated message propagation pattern. The model allows the replication of extreme events in the community such as the 2013 Associated Press hoax incident. Based on intuition that not all nodes in the community network have equal significance, we seek to investigate the community impact by implementing different policies. This paper seeks to address three research questions related to the propagation of message under extreme events and its associated impacts towards the financial market:

1. What is the optimal strategy in dampening the message propagation process under extreme events?
2. Do certain critical nodes, measured by degree, betweenness centrality and closeness centrality, play a larger role in the message propagation mechanism under extreme events?
3. Can the Associated Press hoax incident be replicated in terms of the widespread message propagation and its impact in the financial market?

## II. BACKGROUND

### A. Literature Review of Social Media Network

Social media network is characterized by its complex relationships among multiple users and their dynamic behaviors. With mobile and web-based technologies, social media creates highly interactive platforms for individuals and communities to share and distribute information [5]. Social media platforms are composed by users, links and groups [6]. A fundamental property of social media is that data is driven by heterogeneous users, whose intention cannot often be distinguished clearly. Agichtein et al. pinpoints the need to extract high-quality content “with an accuracy close to that of

humans” [7]. Analyzing the structure of social networks offers an opportunity for measuring aggregated network phenomenon. Network structure can convey information about the dynamics of network and how the network serves as a conduit for information propagation [8].

On social media networks, people post news and links to share and discuss news stories and their opinions in real time [8]. The spread of information through online social network is similar to the propagation of innovation through the network [9]. In the study of diffusion of innovation in social network, it is assumed that at a certain time point, some nodes adopt the innovation and others do not [9]. Two fundamental models depict process of a node adopting innovation are Threshold models and Cascade models. Cascade models means when a social contact of a node  $u$  adopts, then  $u$  adopts with some probability [9]. Some studies of information propagation through a social network are from the standpoint of game-theory, in which one increases possibility of adopting if enough of their friends have also adopted [9].

### B. Literature Review of Agent-Based Modeling

Agent-based modeling (ABM) is an approach to describe systems comprised of autonomous, interacting agents [10]. As decision-makers in the systems, agents live in an environment and adapt their behaviors based on learning experiences [10]. Compared with statistical and mathematical models, agent-based models are more capable in replicating real social phenomena as interactions among agents and information diffusion within the network [11] [12].

The magnitude and complexity of social media data presents a computational challenge for researchers to derive meaningful hypothesis and outcome. Twitter reached 200 million monthly active users as of December 2012 [13]. With increasing usage of social media, these factors add to the challenges for using traditional mathematical and statistical framework to process and analyze data. As an alternative to the large and complex network system, simple agent-based models can incorporate key characteristics of large social networks, such as the differing sizes of personal networks and high clustering [4]. An agent-based model can represent social network user by nodes and their relationship by links [4]. Such framework is effective at simulating social phenomena that are complex and dynamic [14]. In addition, it has the benefits to describe the overall system in aggregate term and “capturing emergent phenomena resulting from the interactions of individual entities” [15]. It suits well with systems constituting entities with measurable behaviors such as Internet users [15].

## III. METHODOLOGY

### A. Motivation and Objective

The agent-based model aims to simulate the network growth and message propagation mechanism in the Twitter financial community. We first define the domain of interest, the Twitter Financial Community, which is a subset of users whose interest align with the financial market. Second, the simulation is constructed through initiating the agent types and defining the simulation topology. The resultant simulation outputs are then validated against empirical characteristics of the Twitter Financial Community.

### B. Simulation Specification

#### 1) Model Assumptions:

- *Twitter Financial Community*: A key assumption of the simulation program is that it models after the Twitter Financial Community, which is a group of users with interest aligned with the financial market. Yang et al. detail the specification of such community and highlights its predictive property through sentiment towards the financial market [4]. The rationale of its adoption is that the financial community contains more refined and precise financial information and therefore its linkage with the financial market can be better captured. The study utilizes empirical facts about this community for its network growth and message propagation characteristics.

- *Empirical Dataset*: Through Twitter API, a large Twitter dataset was collected from August 27<sup>th</sup> 2013 to November 5<sup>th</sup> 2013. This dataset tracks all the financial community users we selected, and it consists of the users’ profile information, structural relation among the community members, and the users’ tweet and retweet messages. We split this dataset into two parts by time. The first part, from August 27<sup>th</sup> 2013 to October 1<sup>st</sup> 2013, is used for estimating the model input parameters and the statistical distribution parameters of the simulation model, while the second part, from October 2<sup>nd</sup> 2013 to November 5<sup>th</sup> 2013, is used for testing and validating the network community and its message propagation characteristics. Overall, we keep our data collection process consistent and maintain high quality control throughout the collection process.

- *Simulation tick as 5-minute interval*: We assume each tick in the simulation model as equivalent to a 5-minute interval. This facilitates the translation of empirical data inputs in terms of tweeting and retweeting rate among users in the community. In addition, it helps interpreting the resultant output in real time scale.

#### 2) Model Parameters

##### **Network Growth**

GrowthRate: Population growth rate for Community

PopulationRatio<sub>B</sub>: Population ratio for Broadcasters

PopulationRatio<sub>A</sub>: Population ratio for Acquaintances

PopulationRatio<sub>O</sub>: Population ratio for Odd users

##### **Relationship**

FollowerRatio<sub>B</sub>: % of community followers for Broadcasters

FollowerRatio<sub>A</sub>: % of community followers for Acquaintances

FollowerRatio<sub>O</sub>: % of community followers for Odd users

Follower%<sub>B</sub>: % of followers that are Broadcasters

Follower%<sub>A</sub>: % of followers that are Acquaintances

Follower%<sub>O</sub>: % of followers that are Odd users

##### **Message Propagation Rate**

TweetingRate<sub>B</sub>: Tweeting rate for Broadcaster per tick

TweetingRate<sub>A</sub>: Tweeting rate for Acquaintances per tick

TweetingRate<sub>O</sub>: Tweeting rate for Odd users per tick

Retweeting%<sub>B</sub>: % of retweeting message from Broadcasters

Retweeting%<sub>A</sub>: % of retweeting message from Acquaintances

Retweeting%<sub>O</sub>: % of retweeting message from Odd users

### C. Simulation Setup

#### 1) Agents: Twitter Users

Three types of agents are adopted in the agent-based simulation: broadcaster, acquaintance and odd user. The classification is based on the Twitter user study by Krishnamurthy [16]. Each agent has a number of followers and friends who are defined as the users he follows. The ratio of the number of followers and the number of friends can be computed to distinguish the type of agent. The empirical data of the financial community indicates that there are 11,063 broadcasters, 130,802 acquaintances and 12,338 odd users [4]. We use this proportion to determine the type of users in our agent-based simulation.

- **Broadcaster:** The Twitter users who have a higher proportion of followers than people they follow.
- **Acquaintance:** The Twitter users who demonstrate reciprocity in their relationships.
- **Odd users/Small clusters:** The Twitter users who have more friends than their followers.

#### 2) Topology: Twitter Mechanism

- *Network Growth:*

The simulation contains two stages of network growth. In the initial stage, the network begins with a few nodes and a tick is equivalent to 1 month. Using the creation date of the community user profile, the growth process can be identified in constructing the financial community. To replicate the growth from April 2009, we assume that the community population growth can be denoted by  $N_T = N_0 e^{\int_0^T r(t) dt}$ , where  $N_t$  and  $r(t)$  represent the number of community population and growth rate at tick  $t$  respectively. The community growth rate exhibits a time-decaying trend since April 2009, in which we successfully fit the model growth rate with a power function of  $r(t) = 0.313t^{-0.8126}$  by taking the least squares method (see Figure 1).

In the latter growth stage, the network continues to increase with the same growth rate in the initial stage but the simulation starts running in 5-minute interval as a tick. We use Poisson distribution to describe the arrival rate of new users in the community. Based on the monthly growth rate, the average number of new users per tick is set as the parameter  $\lambda$  for Poisson distribution. The arrival rate in September 2009 and September 2012 are tested to prove that Poisson distribution matches very well (see Figure 2).

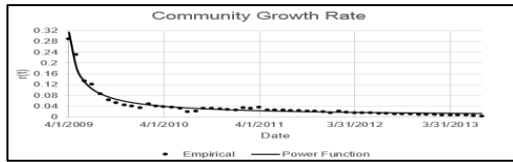


Figure 1: Community Growth Rate

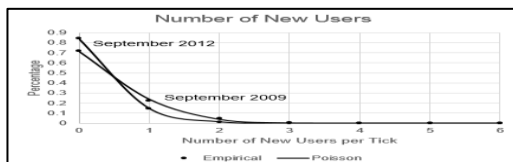


Figure 2: Poisson Distribution of Arrival Rates

#### • Relation Composition

The composition of the network is formed by pairwise relationships between a Twitter user and his followers, which determines the structure of the community network. We define follower ratio as the number of followers over the total number of community users. The empirical data indicates that broadcasters have the highest follower ratio, followed by acquaintances and the least by odd users (see Table 1). The agent-based model replicates the community relation composition by matching each user type's follower ratio with Beta distribution (see Figure 3). The mean of Beta distribution for each user type's follower ratio is consistent with the empirical data (see Table 1). In addition, we use the empirical data to parameterize the composition of followers in the agent-based model. The data suggests that majority of the followers are acquaintances followed by odd users (see Table 2).

Mean of Follower Ratio	Empirical	Beta Distribution
Broadcasters	0.002755	0.002386
Acquaintances	0.000121	0.000119
Odd Users	0.000105	0.000117

Table 1: Follower Ratio Statistics

Follower Composition	Broadcasters	Acquaintances	Odd Users
	0.12	0.63	0.25

Table 2: Follower Composition

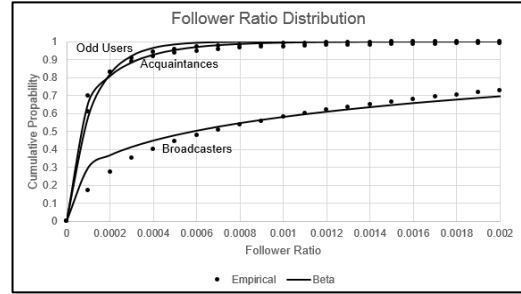


Figure 3: Beta Distribution of Follower Ratio

#### • Message Propagation Mechanism:

Tweeting and retweeting are the two processes for message propagation in the Twitter network. By tweeting a message, Twitter users can directly send information to their followers. Tweeting rate is the number of messages initialized per tick that is specific to the given user type. From the empirical observation of the message data, broadcasters have a larger tweeting rate than acquaintances and odd users (see Table 3). From the training data collected from August 27<sup>th</sup> 2013 to October 1<sup>st</sup> 2013, tweeting rate for each group follows a Gamma distribution after fitting by the least squares method (see Figure 4). The mean and standard deviation of the tweeting rate from the Gamma distributions match well with those of the empirical data (see Table 3).

Tweeting Rate Statistics	Mean		Standard Deviation	
	Empirical	Gamma	Empirical	Gamma
Broadcasters	14.57	14.21	9.02	9.58
Acquaintances	3.09	3.60	2.67	2.66
Odd Users	4.73	5.25	3.69	3.82

Table 3: Tweeting Rate Statistics

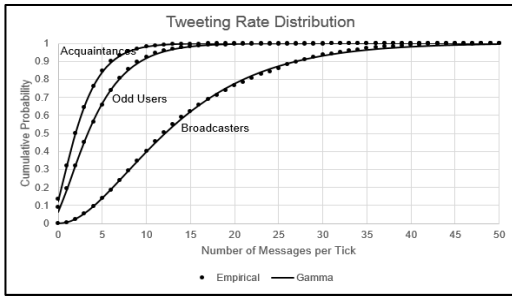


Figure 4: Gamma Distribution of Tweeting Rate

Acquaintances and odd users are more likely to send out less than five messages per tick (see Figure 4). On the other hand, broadcasters have a higher likelihood in sending out more messages per tick. In addition, broadcasters have a flatter tweeting rate cumulative distribution than acquaintances and odd users, indicating a potential fat-tail for sending a large number of messages per tick. For instance, the empirical data suggests that broadcasters have 2% probability to send out more than 40 messages per tick.

Retweeting is the key mechanism in Twitter network for message propagation. When a message is initiated by a user, it has a chance to be retweeted by its followers. Retweeting probability function describes the probability of a message being retweeted on each tick after it is initialized. Two factors of the retweeting probability function are the type of initiator who sends the message and time. From empirical observation, messages from broadcasters have higher likelihood to be retweeted than those from acquaintances and odd users. In addition, retweeting probability are often observed to follow a time-decaying distribution. The retweeting probability is the highest at the tick after the messages are initialized, consistent across all three user types. Based on the time-decaying pattern, we use the least squares method to fit the retweeting probability with a power function (see Figure 5).

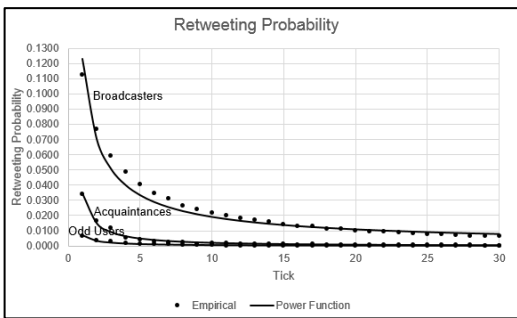


Figure 5: Power Function of Retweeting Probability

#### D. Model Dynamics

- Model Dynamics (tick = 0, 1, 2, ..., T)

The model simulation starts with a small network with 100 nodes. With the growth rate and relation composition parameters, the model generates new nodes and links starting from tick 1 in the initial growth phase. As discussed in the network simulation topology, the first growth phase reflects the monthly growth process treating 1 tick as a month. After the network achieves a stable state, the model configures each tick as a 5-minute interval and starts the message propagation

process. In every tick, tweeting rate of each group follows the pre-determined Gamma distribution while the retweeting process follows the retweeting probability function. When a message is tweeted or retweeted by a node, it is programmed in the model to spread to the adjacent followers.

- Model Outputs for Analysis

We archive all information regarding the profile, structural relation composition and message data of the community network. Every node's identification number, tick of creation, number of followers and friends, and number of tweet and retweet messages are logged into the Twitter profile log. The relation log captures the pairwise relationship between the Twitter user and his followers. Lastly, the message data records the specific message identification number, initiator user identification number, tick of initialization and last active status, retweet ratio, spread ratio, and survival tick. The archive of this information allows us to conduct model validation and scenario analysis.

#### IV. SIMULATION VALIDATION

It is important to validate the agent-based model against the empirical characteristics of the Twitter financial community with respect to the network demographics and aggregated message propagation pattern.

##### A. Community Network Characteristics

The agent-based simulation models the growth of the community population at a similar rate as the empirical data (see Figure 6 and Figure 7). The tick in the model corresponds to one month in real time. Another important validation for the network characteristics is to examine the summary statistics in terms of the average degree and number of nodes and links. By running the simulation 100 times, the network statistics is consistent with the empirical findings shown in Table 4.

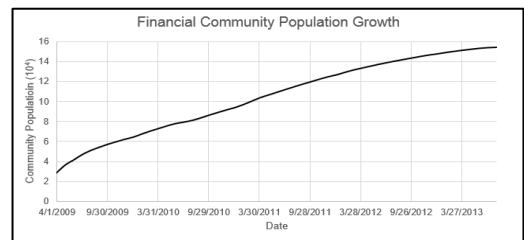


Figure 6: Empirical Community Growth

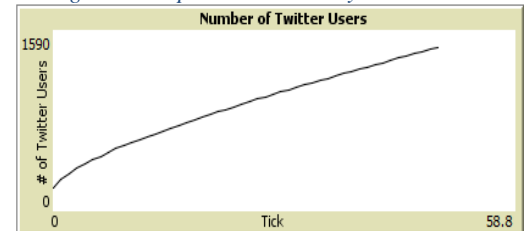


Figure 7: Agent-Based Simulation Community Growth

Network Statistics	Simulation	Empirical Findings
Number of Nodes	14,810	154,203
Number of Links	111,502	10,384,479
Average Degree (Adjusted for scaling)	21.5	35.71

Table 4: Network Statistics Validation

**B. Message Propagation Validation**

**1) Tweeting Rate and Retweeting Rate**

By using empirical data from October 2<sup>nd</sup> 2013 to November 5<sup>th</sup> 2013, we validate the tweeting rate for all three types of users in the simulation at a 5-minute interval.

Tweeting Rate	Broadcasters	Acquaintances	Odd Users
Empirical (10/02-11/05)	15.42	3.52	5.37
Simulation	14.58	3.60	5.12

Table 5: Network Statistics Validation

Retweeting rate is the number of retweets per tick for a message. For each message, its retweeting rate is counted every tick after it is initialized. Empirical data indicates the average retweeting rate is low for all types of users which our simulated outputs have similar results (see Figure 8 and 9). Furthermore, the model replicates the time decaying feature of retweeting rate with a sharp decline at tick 2 (see Figure 9).

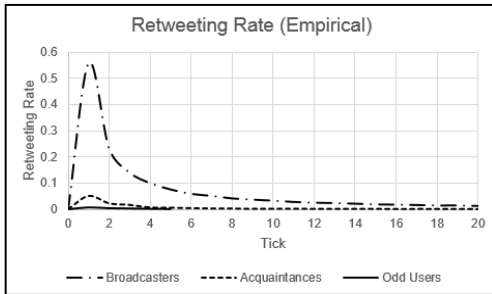


Figure 8: Retweeting Rate in the Empirical Community

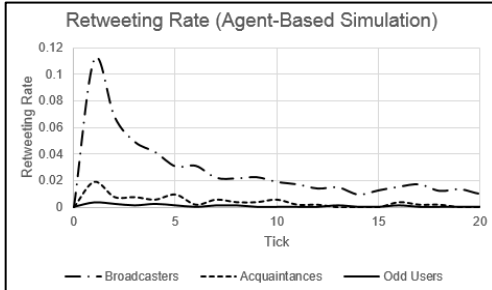


Figure 9: Retweeting Rate in the ABS Community

Number of retweets for each message is another measure in validating the message retweeting process. Over 90% of messages from acquaintances and odd users have no retweets because of the empirical low retweeting probability (see Figure 10). The simulation has higher frequency for 1 and 2 retweets for messages from broadcasters, but the overall histogram matches with empirical observation across all three types of user in the community network (see Figure 11).

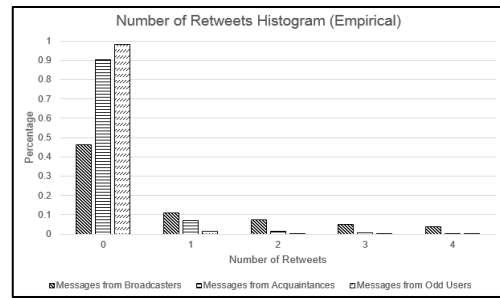


Figure 10: Retweets Histogram in the Empirical Community

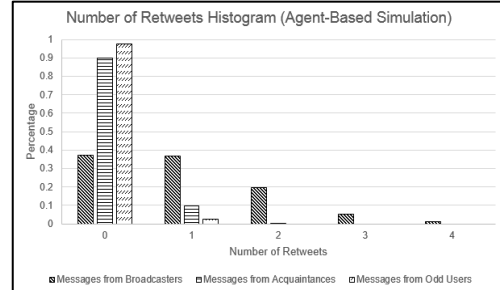


Figure 11: Retweets Histogram in the ABS Community

**2) Message Survival Time Distribution**

Survival time measures the time interval of a message being active in the community. The activeness of the message indicates whether it is being retweeted among the community. Therefore, the survival time is defined as the duration from the message initiation to the last time it gets retweeted. A survival time of 1 tick means no retweeting activity for the message after its initialization. From the empirical data, the percentage of messages drops as survival time increases (see Figure 12). The number of messages with survival time over 10 ticks converges to 0 for all type of users. The agent-based simulation is able to replicate the overall distribution related to message survival time (see Figure 13).

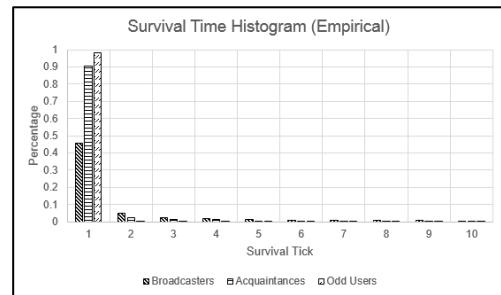


Figure 12: Survival Time Histogram in the Empirical Community

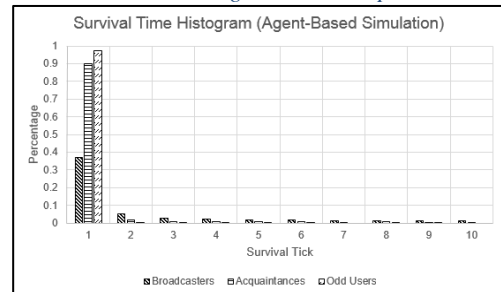


Figure 13: Survival Time Histogram in the ABS Community



## V. CASE STUDY: 2013 ASSOCIATED PRESS HOAX INCIDENT

### A. Background and Objective

On April 23 2013, the Associated Press Twitter account was hacked and a malicious message was sent stating that the White house had been attacked and President Obama was injured at 1:07:50pm ET. Immediately after, thousands of users retweeted the malicious message and this phenomenon triggered a strong wave of selling pressure to the financial market. Among the most notable observations include over 1% drop for both Dow Jones and S&P 500 Indices and intensive investigations from traditional financial news media such as Wall Street Journal and Bloomberg.

The Twitter agent-based model aims to replicate this incident by invoking a similar trigger and validating its empirical message propagation statistics. With the successful replication of the Associated Press Hoax, multiple scenario tests are conducted to evaluate the qualitative influence of key network metrics towards the event formation.

### B. Replication of the Associated Press Hoax

To replicate the Associated Press Hoax requires the fundamental understanding of the trigger's characteristics. The Associated Press Twitter account was one of the largest broadcasters with more than 3 million followers. The malicious message reflected a large credibility gap between the longstanding reputation of the Associated Press and the fictitious nature of its message. This gap contributed to the intense retweeting activities within the first 15 minutes and the majority of the Associated Press retweets were collected from our Twitter dataset. The model captures the average spread ratio of the malicious message in the community, the retweet ratio of the number of retweets over all messages and the survival time when the message is last retweeted. The empirical data has shown similar distribution in all three measures for the Associated Press Hoax event.

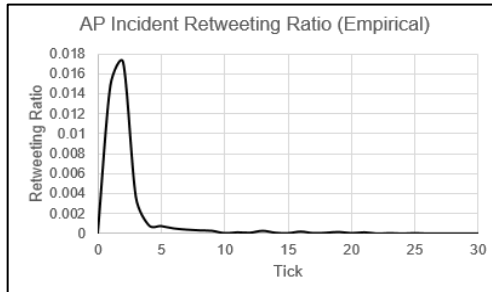


Figure 14: Retweet Ratio Histogram in the Empirical data

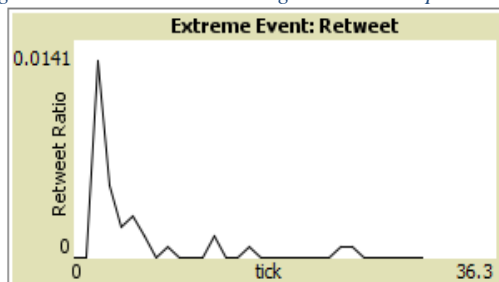


Figure 15: Retweet Ratio Histogram in the ABS Community

### 1) Retweeting Ratio

Retweeting ratio measures the percentage of community users who retweeted the malicious message. The agent-based simulation replicates the die-down pattern of retweeting ratio. The simulation also shows that the retweeting ratio reaching the peak at 1.4% is consistent with the empirical finding (see Figure 14 and 15).

### 2) Survival Time

Based on the retweet ratio, the survival time can be identified as the last time tick that the Associated Press malicious message is active. From the empirical data, the survival time of the message was at tick 20.2 (110 minutes) which is equivalent to 2:52pm Eastern Time. Through running multiple simulations, the average survival time of AP hacked message is 20 ticks. Since our time tick is equivalent to 5 minutes, the result is consistent with the empirical occurrence.

### C. Critical Node Empirical Findings

There exist users who play a central role in the connectedness of the network. These users, known as critical nodes, are situated at the most critical locations of the community network and therefore bear a large weight in the network dynamic properties such as connectedness and message propagation pattern. Analyzing these nodes is essential for understanding the financial community because they represent the most influential nodes in the network in terms of facilitating the message propagation process. Through social network analysis, we identify these critical nodes by applying standard centrality measures: degree distribution, betweenness centrality and closeness centrality. The critical nodes were extracted from the empirical Twitter financial community established by Yang et al., tracks the top 100 users with the highest value by each of the three centrality measures [4]. The description of each centrality measure is detailed below:

1. **Degree centrality** measures the number of connection a specific node has in the network.
2. **Betweenness centrality** captures the number of shortest paths from all vertices to other nodes in the network that passes through the specific node.
3. **Closeness centrality** measures how central a node is in a network by the inverse of farness of the node.

As part of the investigation of the 2013 Associated Press hoax incident, we were able to collect the list of users who retweeted the malicious message and their followers through Twitter API. We are interested to determine empirically the percentage of Twitter users identified as critical nodes who received the malicious message. Our results show that critical nodes by betweenness and degree centrality have large representation involved in the incident with a minimum of 78% participation (see Table 6). The most representative group is the critical node with the highest betweenness centrality, reaching over 88% involvement. Critical node by degree centrality is also well represented with the high participation rate. The underlying intuition for this empirical exercise is that critical nodes contribute significantly to the

message propagation of the malicious message. By reaching successfully to the critical nodes, the message can propagate more effectively to a wider community.

Critical Nodes	% Involved in AP Hoax Incident
Top Degree	78%
Top Betweenness Centrality	88%
Top Closeness Centrality	27%

Table 6: Critical Nodes along the AP Malicious Message

#### D. Scenario Testing

With the empirical evidence that majority of the critical nodes were exposed to the malicious message during the Associated Press hoax incident, the scenario analysis aims to determine the degree of impact triggered by the removal of specific group of critical nodes. Our hypothesis is that not all nodes have the same significance in the community network and removing certain groups of critical nodes would dampen the message propagation pattern and the overall community spread ratio in different degree of impact. We establish three scenarios of removing each three group of critical nodes in incremental scale from 0%-30% and each scenario is run for a hundred times to ensure consistency of the resultant outputs.

All three scenarios result in a linear decaying process of the spread ratio as more critical nodes are removed. However, we show that removing different groups of critical nodes leads to different degree of impact in community spread (see Figure 16). Top betweenness centrality nodes have the steepest slope, indicating the fastest decline of the community spread ratio. Top degree nodes have the next highest level of impact and top closeness centrality nodes has the least influence on spread ratio. This result is consistent with the empirical observation that betweenness and degree centrality serve large representation to those who were involved in the incident.

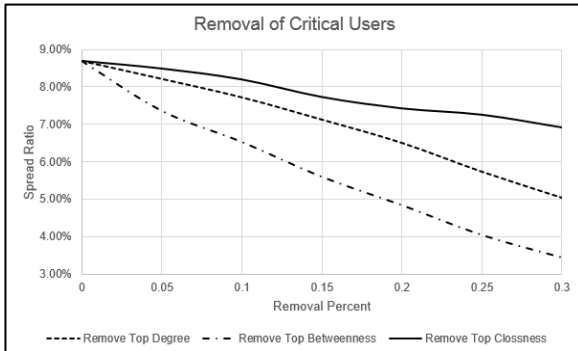


Figure 16: Scenario Testing of Spread Ratio

#### E. Simulation of the 2013 Associated Press Hoax Incident

Another interest of this study is to replicate the Associated Press hoax incident through connecting the social media and financial market simulation. Paddrik et al. investigated the flash crash incident using a zero-intelligence agent-based market simulation and successfully validated against the price and volatility movement during the Flash Crash [17]. In this study, we connect the two simulations through introducing some informed high frequency traders who rely on social media information. In the artificial market simulation, we simulate the price movement of the S&P 500

ETF during the 2013 Associated Press hoax incident. We successfully replicate the hoax incident with the artificial market simulation that describes the financial market crash (see Figure 17 and 18). When the majority of HFT agents first received the trigger signal from the abnormal state in the Twitter financial community, they began to place consecutive orders to sell immediately. The action triggered a violent wave of selling orders based on market interactions and resulted in a major market crash. After a certain time, the market participants realized that the event was fictitious and the market started to recover to the initial level. The HFT flash crash investigation by Paddrik et al. follows a similar conclusion [17], indicating that HFT plays a major role in contributing to the market flash crash.



Figure 17: Empirical Market Crash of AP Hoax

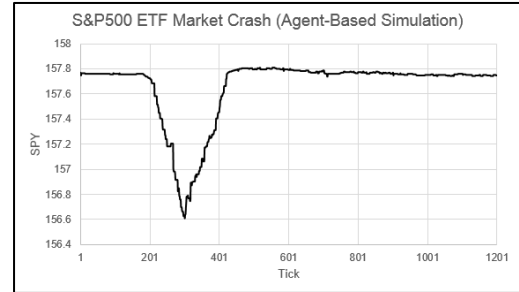


Figure 18: Agent-Based Simulation Market Crash of AP Hoax

## VI. DISCUSSION

This study establishes a framework for simulating the Twitter financial community. Our results reflect closely with the empirical characteristics in terms of Twitters' behaviors and aggregated network dynamics.

The scenario analysis shows that not all users have the same influence in terms of the message propagation. We investigate among the network critical nodes and determine that the removal of nodes with the highest betweenness centrality has the optimal effectiveness in reducing the overall spread of message during extreme events. We also show that the removal of critical nodes does not have significant difference with retweeting ratio and survival tick. This finding has significant implication toward policy making, specifically in fraud detection practice for preventing potential manipulative strategies of the financial market through social media. Policies such as the removal of network critical nodes can be implemented during extreme events such as the 2013 Associated Press hoax incident.

## VII. CONCLUSION

In this paper, we demonstrate that modeling the Twitter financial community is feasible through using agent-based simulation. The use of empirical statistics drawn from Twitter API has allowed the model to be closely matched in terms of the network demographics and aggregated message propagation pattern. The main contribution of this paper rests on the ability to replicate extreme events such as the 2013 Associated Press hoax incident in terms of its impact to the social community network and the financial market. We show that the widespread phenomenon of the malicious message triggers the financial market agents, who are informed about the negative shock, to sell a large volume of shares. Moreover, scenario analysis is conducted regarding implementing preventive policies that could dampen the message propagation process. We show that removing critical nodes of the network, especially the nodes with the highest betweenness centrality, has the optimal effect in reducing the spread of the malicious message to lesser ratio of the community.

For future research effort, this work can incorporate message sentiment as an additional factor for the message propagation function. With appropriate sentiment analysis, the distribution of sentiment on each message can be built using empirical statistics among messages collected from the Twitter financial community. This can further enhance the accuracy of determining which message would have a higher likelihood to be propagated based on its sentiment contrast to the sentiment norm within the community. Another extension is to conduct a larger scale simulation study in determining the optimal strategy in manipulative the financial market through actions in the Twitter financial community. Known strategies such as direct hacking to send malicious message and impersonation of a credible source have been identified as means to manipulative movement in the financial market. This research work can generate significant insights in terms of policy-making so that regulators can mitigate the risk of social media threat towards the financial market.

## REFERENCES

- [1] U.S. Securities and Exchange Commission, "SEC Says Social Media OK for Company Announcements if Investors Are Alerted," [online] 2 April 2013, Available at: <http://www.sec.gov/News/PressRelease>.

- [2] L. Indvik, "Tweets Coming to Bloomberg Terminals," [online] 4 April 2013, Available at: <http://mashable.com/2013/04/04/bloomberg-terminals-twitter>.
- [3] G. Vatu, "Twitter's Influence on the Financial Markets Concerns Traders," [online] 29 April 2013, Available at: <http://news.softpedia.com/news/Twitter-s-Influence-on-the-Financial-Markets-Concerns-Traders-349307.shtml>.
- [4] S. Yang, S. Y. K. Mo, and X. Zhu, "An Empirical Study of the Twitter Financial Community." Working paper.
- [5] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B.S. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media." *Business Horizons* 54, no. 3 (2011): 241-251.
- [6] A. Teutle, "Twitter: Network properties analysis." In *Electronics, Communications and Computer (CONIELECOMP), 2010 20th International Conference on*, pp. 180-186. IEEE, 2010.
- [7] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, "Finding high-quality content in social media," *Proceedings of the international conference on Web search and web data mining*, 2008, pp. 183-194.
- [8] K. Lerman and R. Ghosh, "Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks." *ICWSM 10 (2010)*: 90-97.
- [9] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace." In *Proceedings of the 13th international conference on World Wide Web*, pp. 491-501. ACM, 2004.
- [10] C. M. Macal and M. J. North, "Agent-based modeling and simulation." In *Winter Simulation Conference*, pp. 86-98. Winter Simulation Conference, 2009.
- [11] M. W. Macy and R. Willer, "From factors to actors: Computational sociology and agent-based modeling." *Annual review of sociology* (2002): 143-166.
- [12] N. Gilbert and T. Pietro, "How to build and use agent-based models in social science." *Mind & Society* 1, no. 1 (2000): 57-72.
- [13] Twitter, "There are more than 200M monthly active twitter users," [online], 18 December 2013, Available at: <https://twitter.com/twitter/status/281051652235087872>.
- [14] P. Davidsson, "Agent based social simulation: A computer science view." *Journal of artificial societies and social simulation* 5, no. 1 (2002).
- [15] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 7280-7287, 2002.
- [16] B. Krishnamurthy, P. Gill and M. Arlitt, "A Few Chirps About Twitter," in *WOSN'08*, Seattle, Washington, 2008.
- [17] M. Paddrik, R. Hayes, A. Todd, S. Yang, W. Scherer, and P. Beling, "An Agent Based Model of the E-Mini S&P 500 and the Flash Crash," in *IEEE Computational Intelligence for Financial Engineering & Economics*, 2012, pp 1-8.