

# Game Starts at GameStop: Characterizing the Collective Behaviors and Social Dynamics in the Short Squeeze Episode

Xiaolong Zheng<sup>ID</sup>, Member, IEEE, Hu Tian, Zhe Wan, Xiao Wang, Member, IEEE,

Daniel Dajun Zeng, Fellow, IEEE, and Fei-Yue Wang<sup>ID</sup>, Fellow, IEEE

**Abstract**—In January 2021, the users of subreddit r/wallstreetbets (WSB) triggered an unprecedented short squeeze by driving up GameStop's stock price to an unimaginable high point. During the event, a large number of users participated in the discussion about GameStop and coordinated trading behavior on r/WSB to push the stock price higher. In this article, we investigate the characteristics of the collective behaviors and social dynamics from the evolutions of topological structure, discussed topics, and user sentiment polarity (SP) by constructing dynamic interaction networks, modeling the topic, and analyzing the user sentiment. We find that the topological structure of the interaction network evolves toward a more efficient direction, the discussed topics change more centralized, and the user sentiment tends to be more positive and divergent. And we reveal that part of GameStop's stock price is explained by the social media activity, popularity of the dominant topic, topic cohesiveness, SP of users, and sentiment divergence between interacted users on r/WSB. Our work quantitatively characterizes the interaction networks and user behavior during the GameStop short squeeze and provides an example to analyze the event which synchronously evolves in the physical space and cyberspace. It not only contributes to the analysis of social system behavior and structure but also provides valuable insights into the financial practice and policy decision-making.

**Index Terms**—Dynamic interaction network, financial market, GameStop, short squeeze, social network analysis.

## I. INTRODUCTION

IN JANUARY 2021, the biggest short squeeze of the decade took place on the New York Stock Exchange, causing major financial consequences for certain hedge funds and large losses for short sellers [1]. Due to the competition

Manuscript received May 29, 2021; revised September 28, 2021; accepted October 8, 2021. This work was supported in part by the Ministry of Science and Technology of China under Grant 2020AAA0108401 and in part by the Natural Science Foundation of China under Grant 71602184 and Grant 71621002. (Xiaolong Zheng and Hu Tian contributed equally to this work.) (Corresponding author: Fei-Yue Wang.)

Xiaolong Zheng, Hu Tian, Xiao Wang, Daniel Dajun Zeng, and Fei-Yue Wang are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100190, China (e-mail: xiaolong.zheng@ia.ac.cn; tianhu2018@ia.ac.cn; xiao.wang@ia.ac.cn; dajun.zeng@ia.ac.cn; feiyue.wang@ia.ac.cn).

Zhe Wan is with the Belt and Road School, Beijing Normal University, Beijing 100875, China (e-mail: wanze@bnu.edu.cn).

Digital Object Identifier 10.1109/TCSS.2021.3122260

from digital distribution services, as well as the lockdown restrictions during the 2020 coronavirus pandemic, the public float of the American video game retailer GameStop (GME) was sold short by over 140% [2]. The users of the subreddit r/wallstreetbets (WSB), an Internet forum on the social news website Reddit, initially discussed their intentions to revolt against the short sellers and successfully triggered this short squeeze against the hedge funds by combining the buying power of retail investors to push GameStop's stock price to an unimaginable point. As of January 27, 2021, the short squeeze resulted in a 1700% increase in GameStop's close price, reaching U.S. \$347.51 from U.S. \$17.25 at the ending of the last month [3]. At the end of January, the unusually high price made Melvin Capital that heavily shorted GameStop lose 53% of its investments [4].

The GameStop short squeeze is the first time the retail investors coordinated with each other via social media platforms and fulfilled their plan of squeezing short to change the trading landscape [5]–[7]. It reveals the potential power that a growing force of retail investors can wield in stock markets when equipped with social media. The traditional financial transaction mode has changed with the rapid development of the Internet and social networks [8], [9]. The game of retail investors banding together to impact the financial market has just started. We must pay great attention to this emerging social phenomenon and deeply investigate its internally collective behavior and social dynamics mechanism.

The GameStop short squeeze highlights the significant role of social media in the propagation of investment ideas and the coordination of collective behaviors [10], [11] and produces many research questions about analyzing the behavioral characteristics of participants on social media during the GameStop frenzy [12]. In fact, during the event, a large number of users participated in the discussion about GameStop and coordinated their trading behaviors on r/WSB to push the surge of stock price while forming dynamic interaction networks. Characterizing the social networks under the GameStop short squeeze background with the quantitative technologies not only contributes to the analysis of social system behavior but also provides valuable insights into the financial practice and policy decision-making.

Several preprints exist to analyze the GameStop short squeeze from various viewpoints [13]–[17]. The generating

mechanism of a short squeeze and its impediment to market quality were the first to be explored [13], [14] from the perspective of financial theory. Vasileiou [15] next confirmed that the GameStop short squeeze leads to market abnormality and antileverage effect, which provides a new way to econometrically show a violation of the efficient market hypothesis (EMH). Angel [16] proposed that the short squeeze reveals several loopholes in the system of the U.S. equity market that need to be plugged. From the viewpoint of social media, Lyócsa *et al.* [38] utilized the level of activity on r/WSB and the search volume of Google to explain the variation of the stock price. Angel [16] presented a dynamics model that unifies naive investors, fanatical investors, rational short-term investors, and long-term investors in a social network to study the observed phenomena in the events of GameStop. Long *et al.* [17] found that the comment sentiments extracted from r/WSB influence GameStop's intraday returns.

In summary, the existing works offer noticeable insights on the GameStop short squeeze from both financial and social media perspectives. However, it is empirically unknown to the researchers how the stock market is interconnected with social media activity (SMA) and the potential characteristics of collective behaviors and social dynamics during the short squeeze frenzy. There are few large-scale empirical investigations on participants' behaviors in the social networks, which forms an unprecedented force to push the stock prices soaring. Fortunately, the subreddit r/WSB containing mass social history records is a natural experimental site to explore the behavioral characteristics of users in the formed dynamic interaction networks.

In this article, by constructing and analyzing the dynamic interaction networks between users, we investigate the critical characteristics of user behavior on r/WSB during the short squeeze and how they are likely to influence the information spreading and the volatility of the financial market. We first explore the evolutional dynamics characteristics of topological structure of dynamic interaction networks by analyzing several network statistics. Then, we obtain the basic perceptions for the user behavior in the dynamic interaction networks by executing quantitative analysis from two aspects: 1) the topic modeling and 2) sentiment analysis. In particular, for the discussed topics on r/WSB, we propose a user topic embedding method to investigate the topic evolution from the individual level to the community level and compare the difference between influential users and common users. For the investment sentiment, we explore the evolutional characteristics of the sentiment polarity (SP) and measure the sentiment homophily and divergence. Lastly, to study the relationship between the market and social media, we explain GameStop's price trend by modeling the relationship between price response and five explanatory variables derived from the topic and sentiment characteristics.

<sup>1</sup><https://www.reuters.com/article/us-gamestop-ryan-cohen-idUSKBN29G1EP>

<sup>2</sup><https://twitter.com/CitronResearch>

<sup>3</sup><https://twitter.com/elonmusk>

<sup>4</sup><https://www.cnbc.com/2021/02/01/gamestop-shares-reddit-trader-frenzy-continues-into-february.html>

The main contributions of this article are twofold.

- 1) We quantitatively depicted the evolutional characteristics of interaction networks and user behavior during the GameStop short squeeze and provide a case study to analyze the event which synchronously evolves in the physical space and cyberspace. The findings in this article may provide insight to regulate financial behaviors.
- 2) We constructed five factors from the evolutions of dynamic networks and topic and sentiment characteristics to explain the association between GameStop's price trend and the social behavior of retail investors. It also provides a new and parsimonious way to analyze and predict the stock price from the unstructured social media data.

## II. DATA

In this section, we first split the short squeeze event into five phases according to some remarkable events that significantly promote the GME short squeeze development. Then, we present a brief description of the WSB dataset. Next, we report the construction of dynamic user interaction networks. Last, we introduce the sentimental polarity dictionary used in this work.

### A. Timeline

To explore the interaction between social media and the financial market, we acquired the stock price data of GameStop from [Bartchart](#).<sup>5</sup> To untangle the development of the short squeeze, we divide the event into five phases: prehistory, incubation, development, climax, and fade, according to some important signs that significantly promote the evolution of the GME short squeeze. The detailed information is shown in Table I. We carefully reviewed the process of the short squeeze event and thought that GameStop Corp's announcement that Ryan Cohen and his two partners will be the directors on GameStop's board is the prolog of the event because GameStop's stock price continually rises and finally surges after the announcement. During its development, the battle between [Citron Research](#) and the retailers of r/WSB accelerating the rise of GameStop's price can be seen as a demarcation point from incubation to development. That Elon Musk tweeted "GameStonk!!" sharply pushed up the stock price and made the short squeeze into climax. After February 1, 2021, the continual decline of stock price means that the price frenzy fades. The overall development process can be seen in Fig. 1.

### B. WSB Dataset

To perform an empirical analysis about the behavioral characteristics of participants, we collected more than 9 million unique pieces of content, including posts and comments from 8, 2020 to February 3, 2021 from r/WSB using the Pushshift API.<sup>6</sup> Hereafter, we call the collection WSB dataset.

<sup>5</sup><https://www.barchart.com/>

<sup>6</sup>Pushshift API is an important tool to obtain big data from various social media companies such as Twitter and Reddit. Its main page is <https://pushshift.io/>.

TABLE I  
FIVE PHASES OF THE GAMESTOP SHORT SQUEEZE

Phase	Remarkable event	Date range
Prehistory	No remarkable event.	12/08/20-01/11/21
Incubation	Announcement of board refreshment to accelerate transformation <sup>1</sup> .	01/12/21-01/18/21
Development	Comment from Citron Research predicting the value of the stock would decrease <sup>2</sup> .	01/19/21-01/25/21
Climax	Elon Musk tweeted "Gamestonk!!". <sup>3</sup>	01/26/21-01/31/21
Fade	Continual of decline of GameStop's price <sup>4</sup> .	02/01/21-02/03/21

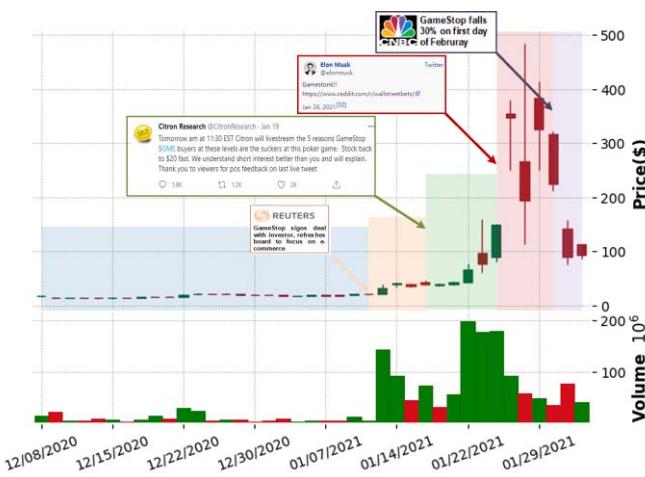


Fig. 1. Diagram of the development of the short squeeze.

TABLE II  
DETAILS OF WSB DATASET IN DIFFERENT PHASES

Phase	#Post	#Comment	#User	Date range
Prehistory	34350	1885554	107698	12/08/20-01/11/21
Incubation	15724	653585	65146	01/12/21-01/18/21
Development	23829	819950	100071	01/19/21-01/25/21
Climax	335362	3124902	516775	01/26/21-01/31/21
Fade	372949	3333469	592021	02/01/21-02/03/21

The collected WSB dataset contains 6 330 271 posts and 8 666 673 comments which are submitted to the subreddit r/WSB from the period December 8, 2020 to February 3, 2021. Due to some technical reasons, the Pushshift sever did not ingest the intraday data from Reddit on January 25, 2021 and January 26, 2021; thus, our WSB dataset excludes these days for the sake of simplicity. The detailed statistical descriptions are reported in Table II and Fig. 2.

### C. Dynamic Interaction Networks

Given a set of posts crawled on r/WSB for someday, we retrieved all their comments to extract the social interaction between users. To accurately match the trading period of New York Stock Exchange, we regarded the data from 4:00 to 4:00 P.M. in the next day as a complete day. A user will be added into the network when the user first submits content, i.e., a post or comment on r/WSB at time  $t$ . If a user commented on

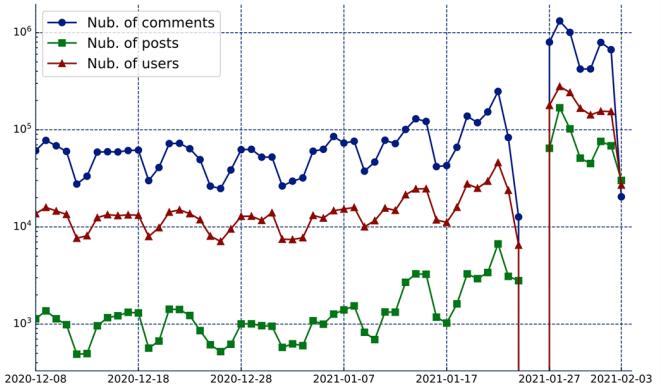


Fig. 2. Evolving trends of the number of comments, posts, and users who participate in social activity. The data are missing on January 25, 2021 and January 26, 2021.

another one's post or commented at time  $t$ , we think there exists an edge between them after  $t$ . In this way, we can construct a user interaction network for each day.

### D. SP Dictionary

Investor's sentiment has been considered as an important aspect to analyze the economic and financial phenomena [18]–[20]. To investigate the investment sentiment of users on r/WSB, it is necessary to use an appropriate social media-orientated SP dictionary rather than a general sentiment dictionary. In this article, we utilize the NTUSD-Fin market sentiment dictionary [21] to calculate the users' sentiment polarities for the market and GameStop. NTUSD-Fin is a market sentiment dictionary based on more than 330K labeled posts crawled from financial social media and contains 8331 words, 112 hashtags, and 115 emojis. A positive word score in NTUSD-Fin represents a bullish market SP.

## III. METHOD

### A. LDA Topic Model

As a classical generative statistical model for discovering the hidden semantic structures in a text body, latent Dirichlet allocation (LDA) is heavily used in financial text mining to extract the meaningful topics in a collection of documents [22]. LDA postulates that documents are produced from a mixture of topics and those topics are words' probability distribution. In this article, we consider all statements of a user before time  $t$  on a particular day as a document and carry out LDA topic modeling to obtain the user's topic. The idea that organizing the weight of each topic word into a vector to represent the user's statement comes from the vector space perspective [23]. We describe the topic weight vector  $v_i$  as a user  $i$ 's topic embedding where each element represents the proportion of a particular topic in a document. Furthermore, a given topic refers to a weighted sum of a series of words that have related meaning and appear in the document more frequently. Thus, we can get the representation for a single topic  $i$  by extracting top- $n$  words' weights  $c_i = [w_j]_{j=1}^n$ . The LDA is implemented using the Python library Gensim which is developed for topic

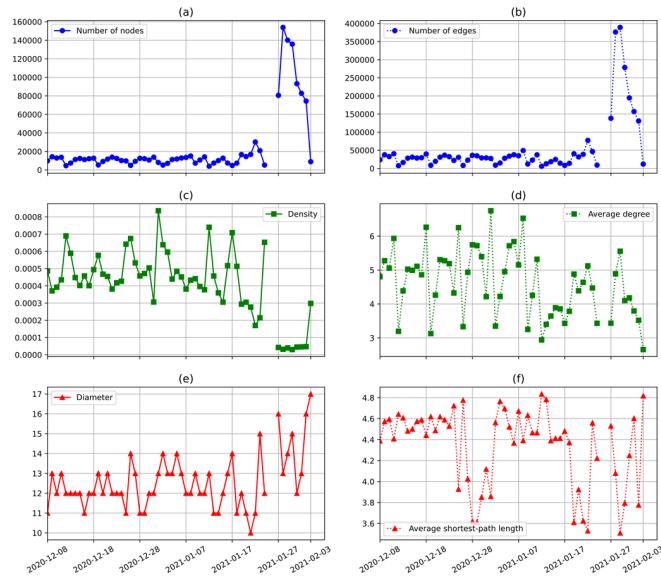


Fig. 3. (a) Daily number of nodes on WSB dataset. (b) Daily number of edges on WSB dataset. (c) Density of user interaction network on each day. (d) Average degree of user interaction network on each day. (e) Diameter of user interaction network on each day. (f) Average shortest-path length of user interaction network on each day.

modeling, document indexing, and similarity retrieval with large corpora.<sup>7</sup>

### B. SP Identification

Due to the high cost of manually labeling the SP for the large-scale social media data, it is reasonable to use a lexicon and rule-based SP identification method. In this work, we adopt Valence Aware Dictionary and sEntiment Reasoner (VADER) [24] as the SP classifier, which is specifically attuned to sentiments expressed in social media. The used sentiment lexicon is reported in Section II-D. VADER generates several sentiment scores. We choose the compound score that is the most useful metric for polarity identification as the final sentiment score. The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between  $-1$  (most extreme negative) and  $+1$  (most extreme positive).

## IV. TOPOLOGICAL CHARACTERISTICS OF INTERACTION NETWORKS

In this section, we mainly concern the topological structure and statistical description of our constructed user interaction networks. We first present the changes of structural indicators of the network over time, including the number of nodes and edges, density, average degree, diameter, and average shortest-path length. Then, several nontrivial topological features examined, including the assortativity, clustering, and centrality. We finally verify whether these interaction networks are scale free and small world.

To describe the dynamic response of social forum r/WSB to the GameStop squeeze, we build the user interaction network

<sup>7</sup><https://radimrehurek.com/gensim/>

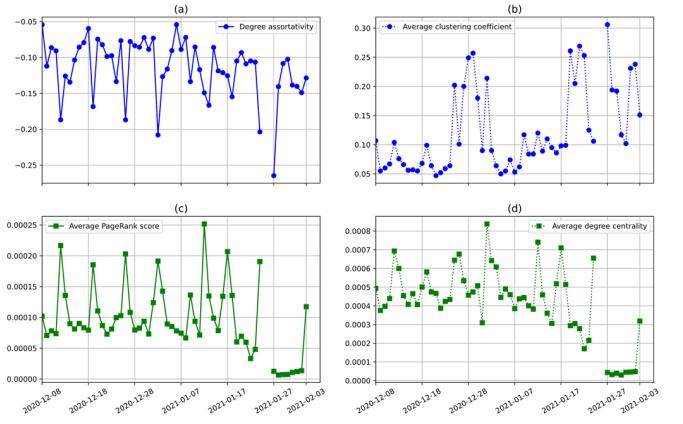


Fig. 4. (a) Degree assortativity of user interaction network on each day. (b) Average clustering coefficient of user interaction network on each day. (c) Average PageRank score of user interaction network on each day. (d) Average degree centrality of user interaction network on each day.

each day. We present the basic topological features, including the number of nodes and edges, density, average degree, diameter, and average shortest-path length in Fig. 3. From the shape perspective, the curves of the number of nodes and edges are similar to GameStop's close price curve. We can see that the number of nodes fluctuates smoothly in the prehistory and incubation phases, rises slightly in the development phase, leaps upward in the climax phase, and drops dramatically during the fade. A significant decrease in the density and the average degree during the GameStop short squeeze is entirely predictable because of the expansion of the network scale. On the contrary, the diameters of user interaction networks rise with the expansion of the network scale.

Another interesting phenomenon is that the networks' average shortest-path lengths do not scale to the massive sizes with the huge expansion of nodes. Instead, they slightly rise and decline in the development, climax, and fade phases compared with the dramatic changes of nodes and edges during the corresponding periods. It suggests that the user-generated interaction networks have low average shortest-path lengths. It has been proven that a network with a low average shortest-path length can provide efficient information diffusion [25]. Thus, the small average shortest-path length of the interaction network makes up the fundamental condition for the possibility that retailers efficiently communicate and coordinate on r/WSB to resist the shorting of institutional investors.

In addition to the direct structural description, we further calculated several nontrivial topological characteristics including the degree assortativity, average clustering coefficient, average PageRank score, and average degree centrality. The degree assortativity measures the tendency for a network's nodes to attach to others that have similar degrees values [26]. Fig. 4(a) shows that the correlations between nodes of similar degree are negative and the values become more negative with the increase of the number of the nodes. In the context of the r/WSB, it means that more users tend to be connected with users with low degree values with the development of GameStop short squeeze. The reason behind is that more common users/low degree users build interactions with those influential users/high degree users by commenting on influen-

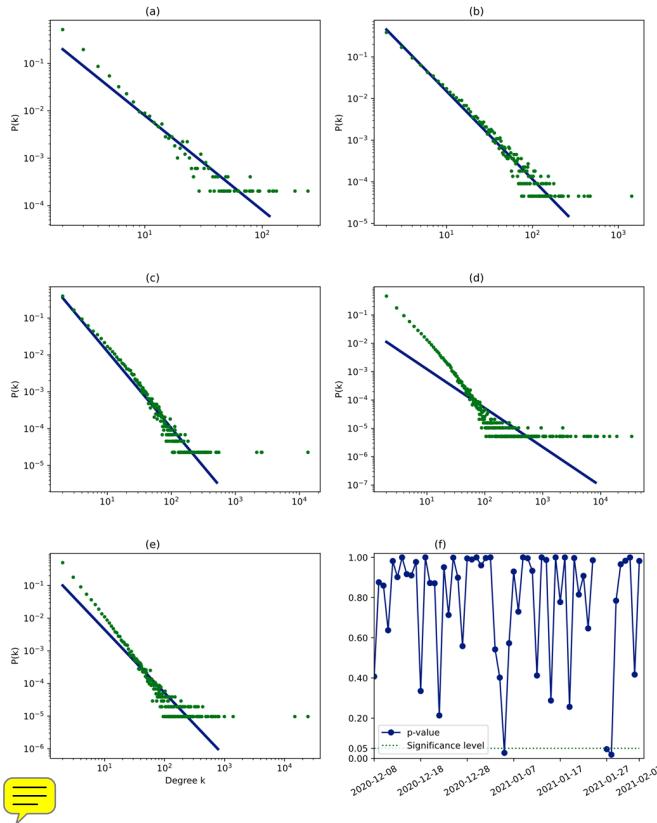


Fig. 5. Degree distribution of the interaction network on some example days. The chosen dates cover all five phases defined in Table I. (a) December 26, 2020. (b) January 15, 2021. (c) January 22, 2021. (d) January 28, 2021. (e) February 1, 2021. (f) Kolmogorov–Smirnov test for the power-law distribution. In this article, we set the significance level as 0.05.

tial users' posts. This reason is also supported by Fig. 4(c) and (d), in which the increase of low degree users leads to the gradual decline of average PageRank score, and average degree centrality with the coming of the climax.

The clustering coefficient represents the degree to which nodes in a network tend to cluster together [25]. The average of daily average clustering coefficient is 0.122, which is slightly smaller than those social media for finding and creating friends such as Twitter and Facebook but is bigger than Yelp, Orkut, and Flickr [27], [28]. Moreover, we find that the average clustering coefficient tends to rise with fluctuation with the evolution of the GameStop event. Considering the smaller average shortest-path length shown in Fig. 3(f) during the last three phases, it is reasonable to think that the interaction network evolves toward the small-world network. That is, with the development of the GameStop event, most users can be reached from every other user by a small number of hops in the user interaction networks on r/WSB. Such topological property is conducive to the transmission of short squeeze intention to revolt against the short sellers.

Furthermore, we verify whether the user interaction networks are scale-free networks. For the constructed interaction network, we implement the Kolmogorov–Smirnov test that compares the fit distribution with the real network data [29]. We select the median dates for each phase to show the degree distribution and fit distribution in Fig. 5. The  $p$ -value of the

TABLE III  
TOP-10 INFLUENTIAL USERS ON R/WALLSTREETS

User	User category	Karma
u/grebfar	Moderator	118920
u/OPTION_IS_UNPOPULAR	Moderator	1517473
u/DeepFuckingValue	-	3695026
u/MotorizedDoucheCanoe	-	201530
u/ITradeBaconFutures	Moderator	NULL
u/wallstreetboyfriend	-	108698
u/SignedUpWhilePooping	Moderator	22981
u/GoBeaversOSU	Moderator	160522
u/theycallmeryan	-	252396
u/theycallme1	Moderator	109675

Kolmogorov–Smirnov test is reported in Fig. 5(f). The test results indicate that the degree distributions follow the power law at the 0.05 significance level on most days, while the power law is invalid on January 27, 2021 and January 28, 2021 during the climax phase. Despite the data missing due to the server ban on January 27, 2021,<sup>8</sup> a potential reason is that a large number of discussions outbreak on r/WSB, thus producing many high degree users on January 28, which makes the statistical test fail. The emergence of the scale-free property means that there exist some hubs in the user interaction networks [30]. We selected the top-10 users who have the highest PageRank score on r/WSB during the GameStop event. As shown in Table III, six of them are moderators on r/WSB. The other four common users have a large number of Karma which reflects how much the users' contributions mean to the community. The emergence of the scale-free property and the category of influential users demonstrate a simple but important finding that most common users prefer to connect to the posts whose authors have considerable influence or certain popularity on r/WSB.

Overall, the users' collective behaviors tend to form the interaction networks on r/WSB, which have the topological structural superiority to transport and boost the diffusion of the investment information. From the evolution perspective, the topological structure can dynamically change to respond to the financial market during the different development phases of the GameStop short squeeze. The interaction network evolves to a small network for efficient information transmission while keeping robust scale-free property with the surge of the stock price.

## V. TOPIC ANALYSIS OF INTERACTION NETWORKS

The evolution of the topic is a side of users' collective behaviors. This section examines the topic characteristics of the user-generated contents on r/WSB during the GameStop short squeeze. We mainly focus on the topic evolution from the individual level to the community level and compare the difference between influential users' concerned topics and common users' concerned topics.

<sup>8</sup><https://www.theverge.com/2021/1/27/22253251/discord-bans-the-r-wall-streets-server>

TABLE IV  
TOP-10 MOST FREQUENT KEYWORDS IN EACH PHASE

Phase	Top-10 most common words
Prehistory	Call, buy, stock, money, share, going, one, day, go, year
Incubation	Gme, share, buy, call, stock, day, money, go, short, market
Development	Gme, share, buy, short, stock, go, call, bb, money, sell
Climax	Gme, share, buy, hold, stock, short, fuck, people, amc, robinhood
Fade	Gme, share, buy, stock, short, people, hold, money, price, sell

### A. Individual Topic Evolution Analysis



Except for the evolution of topological structure, we must know what the users really discuss, how their concerned topics evolve during the GameStop short squeeze, and whether their trending topics are correlated with GameStop's stock price.

We exploited word count to extract the top-10 most frequent words in each evolutional phase. All extracted keywords are shown in Table IV. As we expected, the terms about financial investment such as "stock," "money," and "buy" appear in all phases since r/WSB is a social forum where participants discuss stock and options trading. Meanwhile, the term "gme," the GameStop Corporation's ticker symbol, does not appear in the prehistory phase, indicating that few users focus on the GameStop's stock. With the price of GameStop rising slowly in the incubation phase, the term "gme" emerges and the terms "buy" and "call" are the first two verbs about investment operation, which means that GameStop is the trending topic about investing. Similar hot topics appear in the development phase. It is noteworthy that the term "bb," the BlackBerry Ltd.'s ticker symbol, emerges in the development phase, which satisfies the fact that bb is a company in a similar position to GameStop and a short squeeze about BlackBerry also is triggered during the GameStop short squeeze. The term "amc" appeared in the Climax phase is another company in a similar position to "gme" and "bb." The term "hold" first appears in the climax phase because the war between long buyers and short sellers went to a critical moment in which the price reached a high point and some retailers started to sell shares. In that time, users on r/WSB called retailers to hold GameStop's share and push the price to the moon. Moreover, brokerage service Robinhood halted the buying of GameStop, which aroused much online discussion on r/WSB. Thus, the term "robinhood" appears in the climax phase. During the fade phase, GameStop's price plummeted and, therefore, the terms "buy," "short," "hold," and "sell" are the keywords.

To better investigate the evolutional characteristics of topics, we define an LDA-based topic embedding for every user according to the user's all posts and comments under the given phase. As the introduction about LDA-based topic embedding from Section III, each element of the embedding represents a topic weight. Furthermore, we can introduce a dominant topic label for each user  $i$  by choosing the index of the maximum value of  $v_i$ , i.e.,  $l_i = \text{argmax}_j v_i = [v_{i1}, \dots, v_{ij}, \dots, v_{i10}]$ . The dominant topic label stands for the most likely tendency to some topic. Using t-Distributed Stochastic Neighbor Embedding (t-SNE) [31], [32], an unsupervised, non-linear technique primarily used for data exploration and visualizing high-

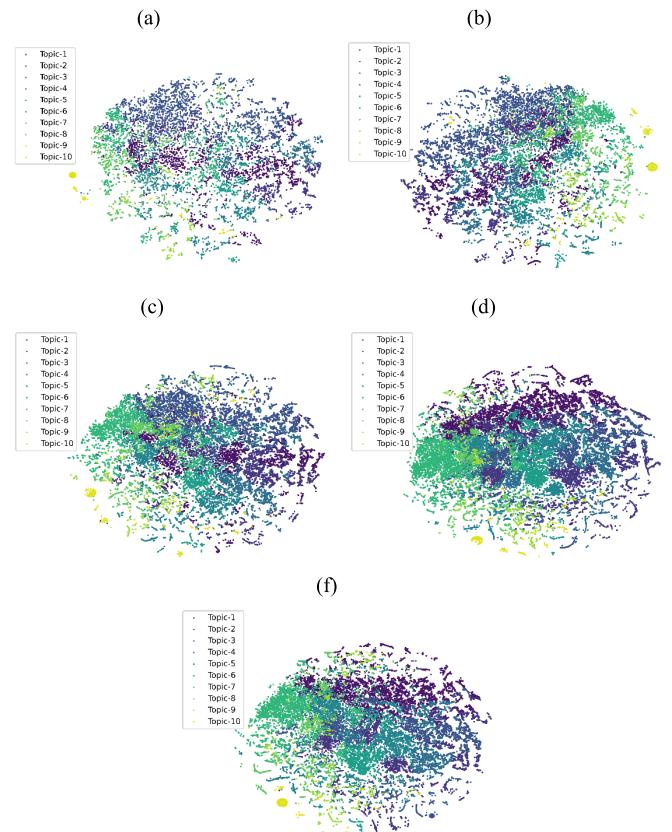


Fig. 6. Evolution of the users' topic embeddings in 2-D space. (a) Prehistory. (b) Incubation. (c) Development. (d) Climax. (e) Fade.

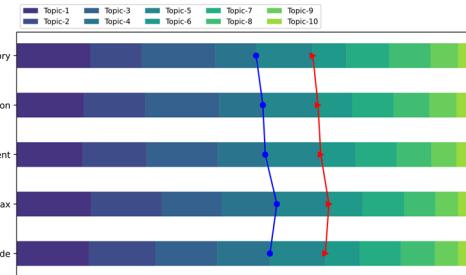


Fig. 7. Evolution of the proportion of users who focus on the same topic.

dimensional data, we reduce the topic embedding into 2-D space for every user in each phase. The results of dimension reduction are shown in Fig. 6. It is clear that the users with the same topic roughly form a cluster and topics 8, 9, and 10 take a smaller proportion than topics 1, 2, and 3 in all phases.

To more clearly show the change of each topic, we calculated the proportion of users who focus on the same topic in each phase. From Fig. 7, we can observe that the proportions of the first four and first five topics gradually increase from prehistory to climax and the proportions of the last four and last five are compressed. We further extracted the top-10 topic words and only contained the word whose weight is greater than or equal to 0.01 for each topic. The topic words are reported in Table V. By jointly analyzing Fig. 7 and Table V, we discover that the proportions of topic 4 and topic 5 rise significantly in the climax and fade phases, respectively.

TABLE V  
TOP-10 TOPIC WORDS WHOSE WEIGHT ARE GREATER THAN OR EQUAL TO 0.01

Topic	Topic words
1	Money, lol, robinhood, account, make
2	Gme, still, amc, hold, moon, go
3	Share, buy, call, week, today, day, 1, 100, 10, sell
4	Buy, hold, sell, gme, still, stop, don't, can't
5	Us, people, wsb, know, guy, one, need, please, thanks, help
6	↗, yes, retard, tendies <sup>9</sup> , ⇠, stay, wtf, green,
7	Short, price, market, squeeze, share, gme, short, fund, hedge, sell
8	Like, stock, it's, gamestop, right, really
9	Fuck, shit, wait, guy, ill, hope, love, lmao
10	BB, let's, power, pltr, amp

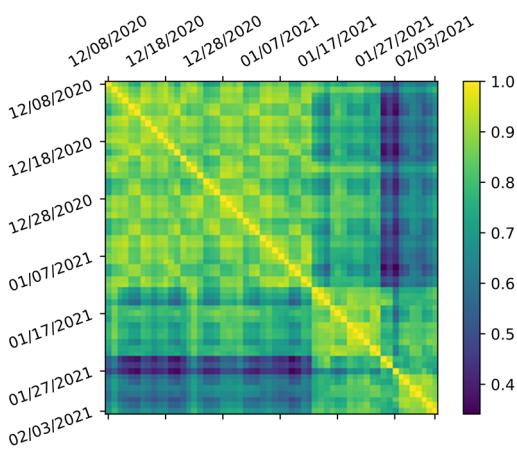


Fig. 8. Heat map of the Pearson correlation coefficient of daily topics. Please note that the data on January 25 and 26 are missing. We used the linear insertion method to complete the values.

And topic 4 refers to GameStop's share-hold call and the topic 5 evolves to an eagerness to people holding the share together. Such topic tone change reflects the transformation of users' investment belief with GameStop's price roller-coaster ride. Our findings are in good agreement with the facts that happened in the stock market and social forum.

To deeply understand the topic evolution, we investigate the correlation between daily topics. For a day  $t$ , we define its daily topic  $\mathbf{d}_t = (1/n_t) \sum_i \mathbf{v}_i$ , where  $n_t$  and  $\mathbf{v}_i$  are the number of the users and the topic embedding of user  $i$  on day  $t$ , respectively. Then, we calculated the Pearson correlation coefficient between any paired daily topics and obtained the heat map in Fig. 8. We can observe a clear dividing line on January 12, which divides the heat map into two distinct blocks. Such result supports Table IV's finding that there is a significant transformation about the hot topics from the prehistory phase to the incubation phase. The topics are highly similar inside the prehistory phase, while there are three smaller subblocks inside the GameStop episode, which correspond to the incubation, development, and fade phases, respectively. Another finding is that consecutive two days usually have higher topic similarity than discontinuous days.

<sup>9</sup>Tendies means gains earned from an investment. The term is usually used by amateur investors day trading on Robinhood.

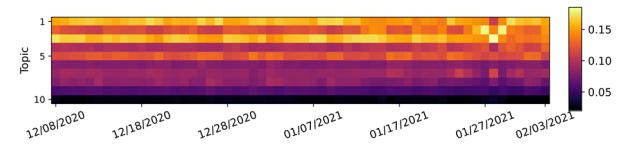


Fig. 9. Heat map of topic embedding in each daily interaction network. Please note that the data on January 25 and 26 are missing. We used the linear insertion method to complete the values.

### B. Community Topic Evolution Analysis

Individual topic evolution analysis obtains an overall view of the topics on r/WSB. However, more extensive discussion about some topics usually takes place at the community level. Individual-level average topic embedding amplifies the personal statement and introduces noise. Thus, we construct a community-level LDA-based topic embedding to describe the topic distribution on r/WSB. For a community  $i$  at time  $t$ , the community-level topic  $c_{ti}$  is defined as the average of community users, i.e.,  $c_{ti} = (1/n_{ti}) \sum_j \mathbf{v}_{tj}$ , where  $n_{ti}$  is the number of users in community  $i$  and  $\mathbf{v}_{tj}$  is the topic embedding of user  $i$  at time  $t$ . Then, the final topic embedding can be defined as  $\tilde{\mathbf{d}}_t = (1/n_t) \sum_i n_{ti} c_{ti}$ . By weighting the community topic, the final topic distribution makes a tradeoff between individual topics and community topic.

In this article, we use the Louvain method [33] to detect the communities in each daily interaction network. We visualized the final topic embedding in Fig. 9. We can find that topic 1 and topic 3 continually have high proportions before January 22. Topics 2 and 3 are the dominant topics during the climax phase. And topics 1–4 are the main topics during the fade phase.

Though we can observe an overall daily topic distribution, we still need to obtain the degree of topic dispersion in the user interaction network because the more dispersed the discussed topics are, the less information the topic distribution contains. Thus, we propose an index based on the community level topic to measure the topic cohesiveness (TC). We define the TC as the Kullback–Leibler (KL) divergence from 10-topic uniform distribution to  $\tilde{\mathbf{d}}_t$ . The higher TC represents that the estimated topic distribution  $\tilde{\mathbf{d}}_t$  is more valuable. The daily TC is exhibited in Fig. 10. The cohesiveness first has a significant increase at the start of the incubation phase then falls to the normal level. From the development phase, the TC starts to rise rapidly again. The TC reveals the evolution of information quantity of topic distribution with the development of the GameStop event. The TC can also be defined at the finer granularity of time to indicate the consistency of users' concerns on the social network.

### C. Comparison Analysis

In Section IV, we have known the importance of the influential users in the construction of the interaction networks. It inspires us to compare the influential users and common users in the topic aspect.

We chose the first two thousand users with the highest PageRank score as the influential users and others as the

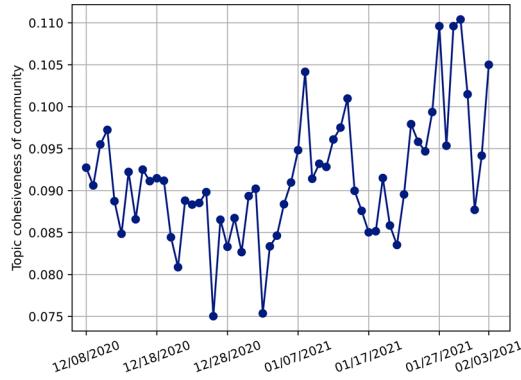


Fig. 10. TC of user interaction network on each day. Please note that the data on January 25 and 26 are missing. We used the linear insertion method to complete the values.

TABLE VI  
TOP-10 MOST FREQUENT WORDS IN EACH PHASE

Phase	Top-10 most frequent words of influential users	Top-10 most frequent words of common users
Prehistory	Vote, gme, click, call, share, buy, post, stock, money, go	Call, buy, day, money, go, one, share, time, market, fuck
Incubation	Gme, share, buy, short, sell, stock, hold, squeeze, day, go	Gme, share, buy, call, day, go, stock, market, fuck, money
Development	Gme, share, buy, short, go, stock, money, fuck, go, sell	Gme, share, buy, bb, call, stock, go, short, sell, hold
Climax	Gme, share, buy, short, hold, sell, stock, fuck, people, squeeze, market	Gme, buy, hold, stock, amc, share, sell, fuck, short, robinhood
Fade	Gme, share, short, buy, people, stock, money, price, hold, sell	Gme, buy, stock, share, hold, people, short, money, sell, hold

common users. We first extracted the high-frequency words of the influential users and common users in each phase. The top-10 most common words in each phase are summarized in Table VI. Comparing their most common words, we find that their high-frequency words in their posts and comments are similar in all phases except the prehistory phase. In prehistory, the term “gme” appears in influential users’ high-frequency words but not in common users’, which indicates that the influential users followed the GameStop earlier than the common users.

Though the common words are similar, their concerned topics still exist differences. Fig. 11 shows the proportion of their dominant topic in each phase. The most significant difference between their topic distributions is that influential users discussed more topic 7 from the incubation phase to the fade phase. Two types of uses have similar topic distributions in other topics. We also find that influential users hardly concern the topics 9 and 10 during the GameStop short squeeze. That is, the influential users express few obscenities and other stocks except GME.

To better compare the topic characteristic between two types of users, we visualized their topic embedding in the 2-D space in Fig. 12. As shown in all subfigures in Fig. 12, the most influential users’ 2-D points appear close together in all phases, while the common users’ 2-D points scatter. Such phenomenon demonstrates that influential users’ topics are more centralized than common users’ topics.

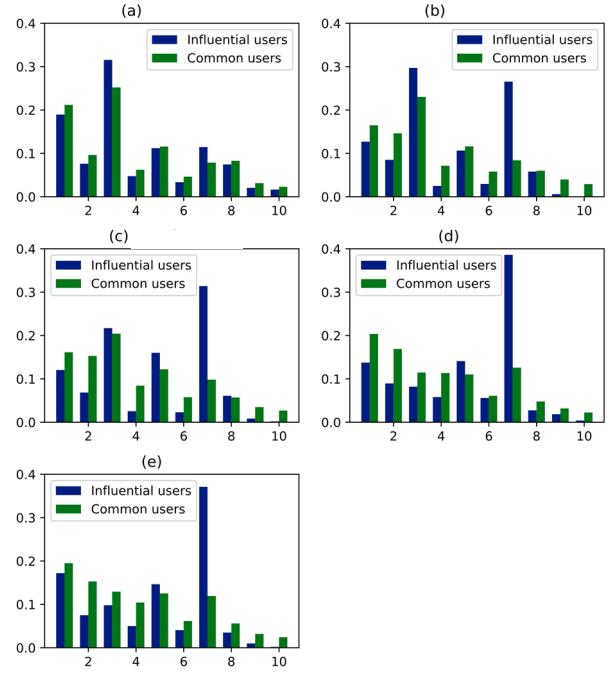


Fig. 11. Proportion of dominant topic of the influential users and common users in each phase. (a) Prehistory. (b) Incubation. (c) Development. (d) Climax. (e) Fade.

Another important problem is how the topics of influential users and common users are interrelated to each other during the GameStop short squeeze. The correlation between the two types of users is shown in Fig. 13. There are six edges in the successive two-time steps. Among them, we think that the mutual correlations between influential users and common users at time  $t$  are equal. We still use the Pearson correlation coefficient to measure the relationships. From Fig. 13, we can see that those correlations 2 and 5 continually have high values which indicate that the correlation between the same type of users is high. But after January 5, the correlations 1, 3, 4, and 6 were high, revealing that the daily topic changes were negligible since January 5. Furthermore, we calculated the Pearson correlation coefficient for both influential users and common users. The heat maps are shown in Fig. 14. For the common users, we can observe two distinct blocks with a clear dividing line on January 12, similar to Fig. 8. In contrast, there were more distinct blocks for the influential users before January 12, which reflects the quick transfers of influential users’ topics.

## VI. SENTIMENT ANALYSIS OF INTERACTION NETWORKS

Analyzing the user’s SP for investment is an essential aspect of understanding the influence of investors’ collective behaviors on the financial market. In this section, we examine the sentiment evolution of users on r/WSB during the GameStop short squeeze and measure the sentiment homophily and divergence in our interaction networks.

### A. Sentiment Evolution Analysis

Investment sentiment on r/WSB is a user’s anticipation about the price development in a market. In this article, the

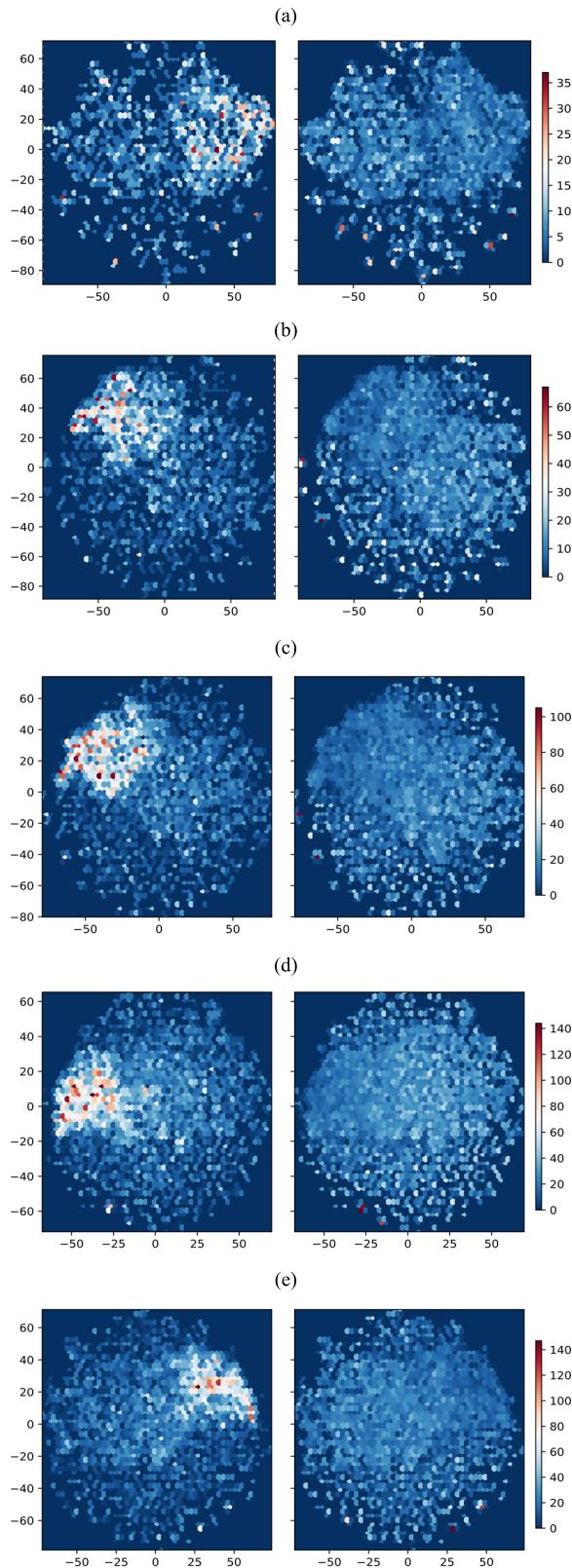


Fig. 12. Two types of users' topic embeddings in the 2-D space. The left subfigure represents the influential users' scatter and the right is the common users' scatter. (a) Prehistory. (b) Incubation. (c) Development. (d) Climax. (e) Fade.

VADER and SP dictionary which are introduced in Sections II and III are applied to identify the user's SP. The polarity score is normalized to be between  $-1$  (most extreme negative)

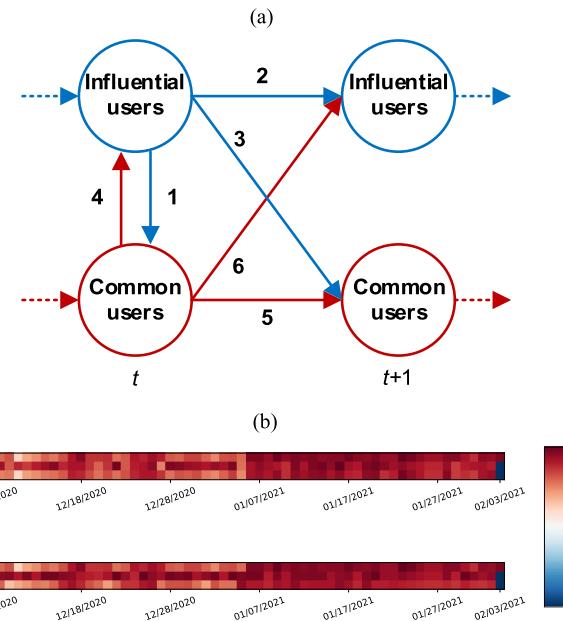


Fig. 13. Topic correlation of the influential users and common users. (a) Six types of topic correlations. (b) Correlation value on each day.

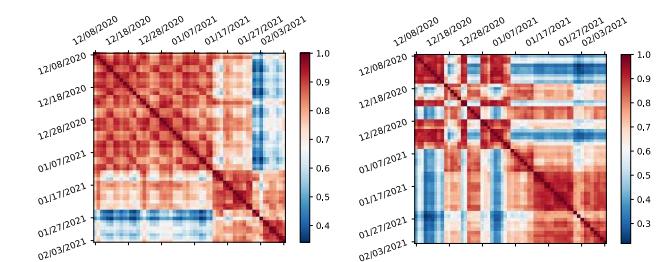


Fig. 14. Heat maps of the Pearson correlation coefficient of daily topics. The left subfigure is the common users' similarity matrix, and the right subfigure is the influential users' similarity matrix. Please note that the data on January 25 and 26 are missing. We used the linear insertion method to complete the values.

TABLE VII  
DETAILED STATISTICAL DESCRIPTION OF USERS' SP

Phase	#Negative	#Positive	Mean	Std.	Kr.	Skew.
Prehistory	34649	21739	-0.01	0.12	17.61	2.53
Incubation	26399	22568	0.02	0.17	9.20	2.42
Development	37226	37012	0.03	0.18	8.13	2.31
Climax	153642	144521	0.02	0.17	9.74	2.51
Fade	168028	149093	0.03	0.18	6.89	2.02

and  $+1$  (most extreme positive). Table VII gives the detailed statistical description of users' SP score in each phase. Overall, the number of users with negative SP is higher than the number of users with positive SP for all phases. The average polarity score is close to zero, i.e., the whole sentiment is almost neutral. The large kurtosis means that many users' polarity scores are close to zero. The positive skewness indicates that the tail of the polarity distribution is on the right.

We regard the users who have positive polarity scores as the positive users. Reversely, we regard the users who have negative polarity scores as the negative users. In that case, we can obtain the daily user number with different polarities.

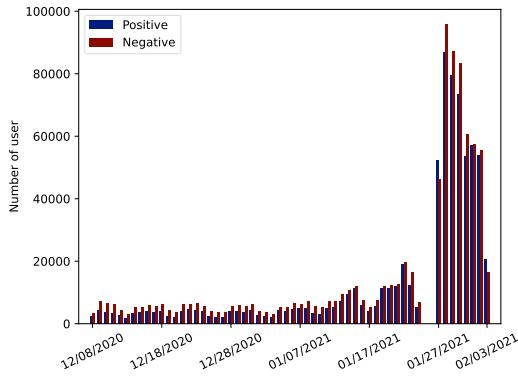


Fig. 15. Daily number of users with different sentiment polarities. The data on January 25 and 26 are missing.

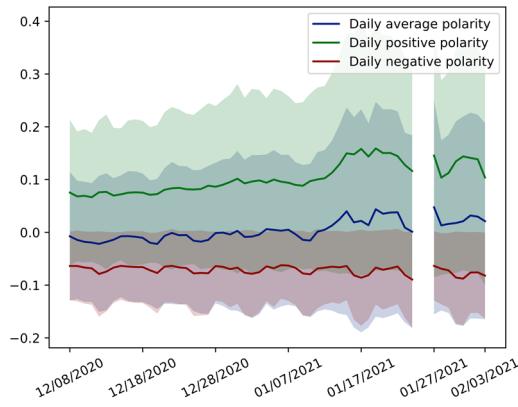


Fig. 16. Daily average, positive, and negative sentiment curves. The data on January 25 and 26 are missing.

The daily number of users with different sentiment polarities is shown in Fig. 15. As our statistical result indicates that users with negative investment polarity are more than those with positive investment polarity on most days, we plotted the daily average polarity score curves of all users, negative users, and positive users in Fig. 16. The average polarity curve of all users slowly rises but is still below zero in the prehistory phase. After January 12, the cure of positive users has significantly increased, which leads to a positive average polarity score. However, the average score of negative users is steady in the prehistory phase and slowly decreases after January 12. Overall, although the proportion of users with positive polarity does not increase, their positive polarity strength has significantly improved during the GameStop short squeeze.

Moreover, we compared the sentiment polarities of influential users and common users in Fig. 17. In fact, their sentiment polarities do not have a significant difference before January 18, i.e., the end of the incubation phase. After that, common users show higher positive polarity scores than those of influential users, which demonstrates that common users have more positive investment leaning for GameStop's stock price in the development, climax, and fade phase. We also count the daily number of communities with negative polarity scores and positive polarity scores. The results are shown in Fig. 18. We can find that the number of negative communities is larger than the number of positive communities on most days. In comparison, the number of positive communities is

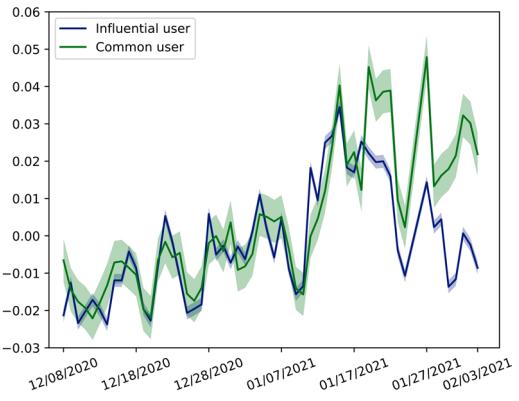


Fig. 17. Daily score curves of influential users and common users. Please note that the data on January 25 and 26 are missing. We used the linear insertion method to complete the values.

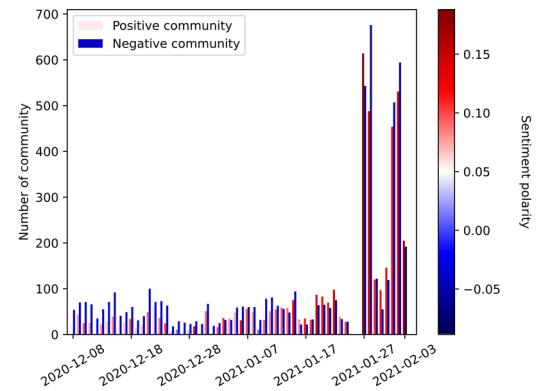


Fig. 18. Daily number of communities with negative polarity scores and positive polarity scores.

larger than the number of negative communities from January 16 to January 24.

Identifying the investment SP can help to extract meaningful sentimental topic words. We implemented LDA in the contents with positive scores to obtain the positive topic words. In the same way, the negative topic words are also extracted from the contents with negative scores. These words are listed in Table VIII, where 82 words only appear in the contents with positive scores, 67 words only appear in the contents with negative scores, and 55 words appear both in positive and negative contents. We can see some words and emoticons with strong positive investment tendency appear in the positive topic word list such as the mars 🚀, moon 🌙, and diamond 💎. Similarly, some words and emoticons with strong negative investment tendency also appear in the negative topic word list such as lose, transfer, 🐾, and 😢. Some topic words have dual polarity despite their meanings such as buy, call, sell, and short.

### B. Evolution Analysis of Sentiment Homophily and Divergence

In the interaction networks, an interesting problem is whether a user will tend to interact with similar people who have similar SP during the GameStop short squeeze. In this section, we investigate the sentiment homophily and divergence in the interaction networks to reveal the potential user behavior under the short squeeze background.

TABLE VIII  
SENTIMENTAL TOPIC WORDS EXTRACTED FROM POSITIVE USER'S  
CONTENT AND NEGATIVE USER'S CONTENT

Positive topic words (82)	Ik, attack, autist, better, bro, brother, come, crash, damn, delete, diamond, elon, end, enjoy, financial, forever, friday, gang, getting, gonna, good, got, great, happy, jack, jan, january, jesus, jmia, job, king, let's, line, lol, long, loss, love, man, mars, melvin, moon, needs, next, nice, nok, oh, paper, pltr, post, retard, rocket, ship, sold, stay, stop, strong, take, target, tendency, thank, thank, till, together, tsla, two, wait, went, we've, won't, work, yes, ♦, ♪, ♠, ♣, ♤, ♦, ♧, ♨, ♩, ♪, ♦.
Dual polarity topic words (55)	Amc, back, bb, buy, call, cover, day, dip, even, fuck, fund, gamestop, get, gme, go, guy, hand, hedge, hold, interest, know, last, like, limit, make, market, money, one, option, people, price, really, robinhood, sec, sell, share, shit, short, silver, squeeze, stock, street, support, think, time, today, tomorrow, trading, us, wall, want, way, week, would, year.
Negative topic words (67)	Account, actually, also, app, ass, bank, big, broker, business, cap, cash, citadel, click, company, could, earning, ever, every, fidelity, fraud, future, gain, game, hour, investor, lose, manipulation, many, margin, media, minute, monday, morning, much, new, open, order, pay, position, put, real, retail, rh, rich, rule, see, shut, stimulus, sub, sure, system, tech, term, tesla, that, trade, transfer, use, value, volume, world, wsb, year, yesterday, 📈, 🚫, 🚫.

Notes: We only contained the words whose weights are greater or equal to 0.01 for each topic.

TABLE IX  
OLS REGRESSION RESULTS

Variable	Coef.	S.E.	t-statistic	P >  t	[0.025	0.975]
<i>SMA</i>	0.4558	0.022	21.028	0.000	0.413	0.498
<i>DTP</i>	0.1065	0.035	3.032	0.003	0.037	0.175
<i>TC</i>	0.4015	0.023	17.094	0.000	0.355	0.448
<i>SP</i>	0.1950	0.021	9.235	0.000	0.153	0.237
<i>SD</i>	-0.2848	0.021	-13.367	0.000	-0.327	-0.243
R-squared (uncentered)				0.870		
Adj. R-squared (uncentered)				0.869		
F-statistic				549.5		
Prob (F-statistic)				5.83E-179		
Log-Likelihood				303.83		
AIC				-597.7		
BIC				-577.5		

Notes:  $R^2$  is computed without centering (uncentered) since the model does not contain a constant. Standard Errors assume that the covariance matrix of the errors is correctly specified.

A simple method to describe the sentiment homophily is comparing the focal user  $i$ 's SP  $s_i$  with his neighbor users set  $\mathcal{N}_i$ 's SP. The neighbor users set  $\mathcal{N}_i$ 's SP is defined as the average polarity scores of individual users from  $\mathcal{N}_i$ , i.e.,  $\tilde{s}_i = (1/|\mathcal{N}_i|) \sum_{j \in \mathcal{N}_i} s_j$ , where  $\tilde{s}_i$  is the SP of  $\mathcal{N}_i$  and  $|\mathcal{N}_i|$  is the number of users in  $\mathcal{N}_i$ . We made a 2-D hexagonal binning plot of all points  $(s_i, \tilde{s}_i)$  in Fig. 19 and the value of the hexagon is determined by the number of points in the hexagon. As shown in Fig. 19, quadrant I has more points than quadrant IV in all phases, i.e., a user with a positive polarity score tends to interact with other users with positive polarity scores. Moreover, with the developing of the GameStop short squeeze, such a phenomenon is more obvious. On the contrary, there is no significant negative sentiment homophily because quadrant II and quadrant III have the roughly same size points in all phases. Considering a large number of users sticking around score zero,  $x$ -axis,  $y$ -axis, and origin have many points.

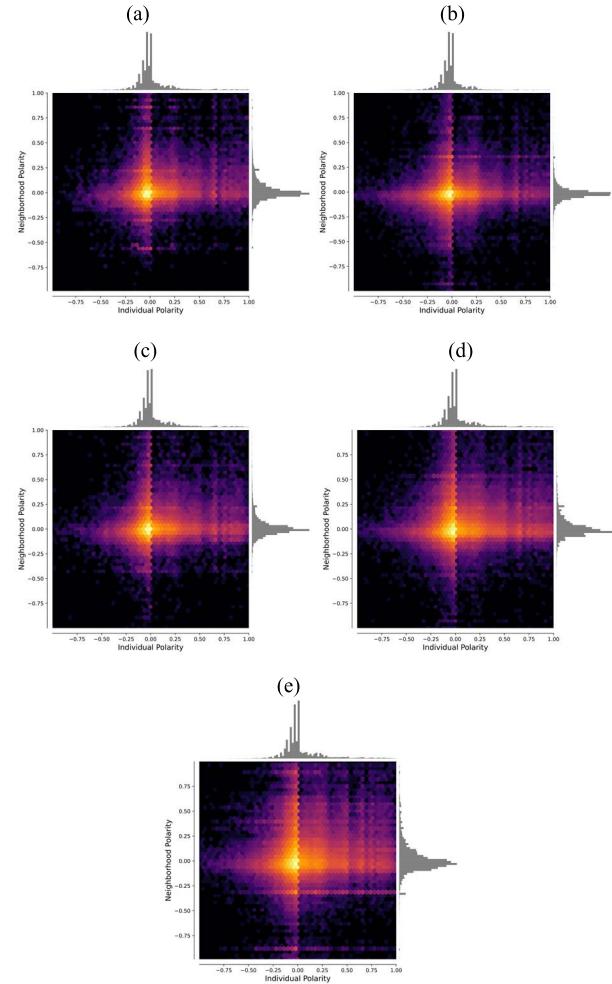


Fig. 19. Hexagonal binning plot of all points  $(s_i, \tilde{s}_i)$  in each phase. The  $x$ -coordinate represents the focal user  $i$ 's sentiment score and the  $y$ -coordinate represents the average polarity score of  $i$ 's neighbors. (a) Prehistory. (b) Incubation. (c) Development. (d) Climax. (e) Fade.

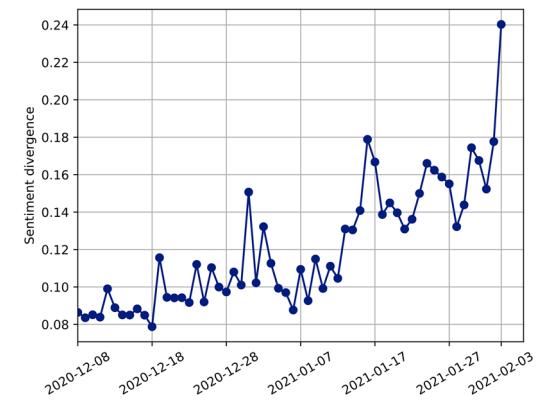


Fig. 20. Daily SD. Please note that the data on January 25 and 26 are missing. We used the linear insertion method to complete the values.

Except for the sentiment homophily, we further explore the sentiment divergence (SD). For the sentiment point  $(s_i, \tilde{s}_i)$ , we define the SD of a network at time  $t$  as  $sd_t = (1/n_t) \sum_i |s_i - \tilde{s}_i|$ . Then, we plotted the SD in Fig. 20. From Fig. 20, we observe that the SD grows wavy and the rising rate has an obvious improvement after January 7, which indicates that the SD between the adjacent users in the interaction



nets escalates with the development of the GameStop short squeeze.



## VII. EXPLANATION OF GAMESTOP'S PRICE FROM INTERACTION NETWORKS

In Sections IV–VI, we investigated the characteristics of the collective behaviors and social dynamics in interaction networks constructed during the evolution of the GameStop event on r/WSB. These characteristics cover the evolutions of the topological structure, user's topic, and sentiment. A clear fact is that some of the characteristics have considerable relationships with GameStop's stock price. The important influence on social media for investment has been proven [34], [35], which inspires us whether these attributes can help us to explain the GameStop's stock price.

According to the previous analysis, we construct five explanatory variables from three perspectives.

- 1) SMA measures the user's activity level on r/WSB. From the perspective of the topological structure, we use the number of edges of the interaction network [see Fig. 3(b)] at time  $t$  to indicate the SMA at time  $t$ .
- 2) Dominant topic popularity (DTP) measures the degree of concern about the topic which is discussed by most people at time  $t$ . DTP at time  $t$  is defined as the proportion of the users who focus on dominant topic at time  $t$ . The definition of DTP is similar to the operation in Fig. 7.
- 3) TC measures the degree of topic dispersion, defined in Section V-B. Generally, the more dispersed the discussed topics are, the less information the topic distribution contains.
- 4) SP is the average value of the all users on r/WSB at time  $t$ , which depicts the sentiment trend of the whole social forum.
- 5) SD measures the degree of sentiment dispersion in the interaction network at time  $t$ , which has been defined in Section VI-B.

Finally, we exploit the linear regression to model the relationship between the stock price response  $y$  and the five explanatory variables

$$y_t = \beta_1 \text{SMA}_{t-1} + \beta_2 \text{DTP}_{t-1} + \beta_3 \text{TC}_{t-1} + \beta_4 \text{SP}_{t-1} + \beta_5 \text{SD}_{t-1} + \epsilon_t \quad (1)$$

where  $\beta_i, i = 1, \dots, 5$  are the estimated regression parameters and  $\epsilon_t$  is the error term. These parameters are estimated via ordinary least squares (OLS). The OLS regression results are reported in Table IX.

Considering the developing process, we collected a 15-min stock price from January 12 to February 3. The five explanatory variables are recalculated at the 15-min level. The response variable and explanatory variables are shown in Fig. 21. We report the estimated coefficients, standard errors, and corresponding model characteristics.  $R^2$  and F-statistic show that our model adequately describes GameStop's stock price in our sample period. The coefficients of SMA, DTP, TC, and SP are positive and significant. Therefore, the more the interaction of the users, the more popularity of the dominant topic, the higher cohesiveness of the discussed topics, and the

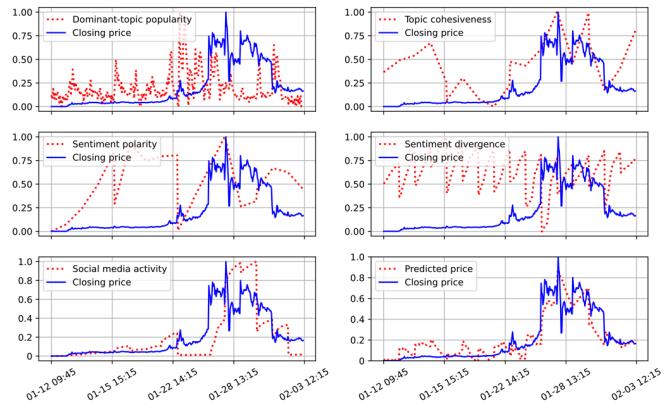


Fig. 21. Five explanatory variables, GameStop's stock price, and the predicted stock price.

more positive investment sentiment, the higher the next 15-min stock price is. Interestingly, the coefficient of SD is negative and significant, which indicates that the more SD among users on r/WSB, the lower the next 15-min stock price is. Our results suggest that part of GameStop's stock price is driven by the SMA, popularity of the dominant topic, TC, SP of users, and SD between interacted users on r/WSB. We think it is a new and parsimonious way to analyze and predict the stock price from the unstructured social media data.

## VIII. CONCLUSION

In January 2021, the users from subreddit r/WSB triggered an unprecedented short squeeze by driving up GameStop's stock price to an unbelievable high point. The GameStop short squeeze highlights the significant role of social media in the propagation of investment ideas and the coordination of behaviors, as well as produces a research question about analyzing the collective behaviors and social dynamics characteristics of participants on the social media during the GameStop frenzy. The game of retail investors banding together to impact the financial market has just started. We must pay great attention to this emerging social phenomenon and deeply investigate its internally collective behavior and social dynamics mechanism.

In this article, we quantitatively investigate the characteristics of the dynamic interaction networks from the evolutions of the topological structure, discussed topics, and user's SP by exploiting network construction technologies, topic modeling, and sentiment analysis.

- 1) For the topological structure, we found that the users' collective behaviors tend to form the interaction networks on r/WSB, which have the topological structural superiority to transport and boost the diffusion of the investment information. And the interaction network evolves toward a small network while keeping robust scale-free property with the surge of the stock price.
- 2) For the evolution of the discussed topic, we observed that the topic distribution produces a significant change whether in individual level and community with the rise of the GameStop event and becomes more centralized, and the degree of topic dispersion of influential users is larger than that of common users.
- 3) For the evolution of the sentiment, we discovered that the collective SP slowly increases toward the positive

direction with the development of the GameStop short squeeze, but more users tend to hold negative polarity on most days, and the sentiment homophily and divergence become more and more significant.

- 4) Finally, we suggested that part of GameStop's stock price is explained by the SMA, popularity of the dominant topic, TC, SP of users, and SD between interacted users on r/WSB by analyzing the relationships between price and social media.

Quantitatively characterizing the interaction networks and participants' behaviors during the GameStop short squeeze contributes to the analysis of social system behavior and structure and provides valuable insights into financial practice and policy decision-making. These findings in this article may provide evidence to regulate financial behaviors. Based on these findings, we also provide a new and parsimonious way to analyze and predict the stock price from the unstructured social media data.

However, there are still many problems about the GameStop short squeeze to explore. We only consider the participants from r/WSB in this article, while other social media such as Twitter and mainstream news reports also influence the development of the hot event [36]. How to jointly model the influence from multiple social media platforms is a challenge [37] to social computing.

#### ACKNOWLEDGMENT

The authors thank Xingwei Zhang and Yuejiao Wang for precious insights for the development of this article. They are grateful to Rongrong Jing and Gang Zhou for inspiring the data analysis and result visualization.

#### REFERENCES

- [1] B. Pisani. (Jan. 26, 2021). *Short Sellers are Down 91 Billion in January as GameStop Leads Squeeze in Stocks They Bet Against*. CNBC. Accessed: May 28, 2021. [Online]. Available: <https://www.cnbc.com/2021/01/26/short-sellers-are-down-91-billion-in-january-as-gamestop-leads-squeeze-in-stocks-they-bet-against.html>
- [2] J. Ponciano. *Meme Stock Saga Officially Over? GameStop Short Interest Plunged 70% Amid 20 Billion Loss*. Forbes. Accessed: May 28, 2021. [Online]. Available: <https://www.forbes.com/sites/jonathanponciano/2021/02/10/meme-stock-saga-officially-over-gamestop-short-interest-plunged-70-amid-20-billion-loss/>
- [3] (Jan. 28, 2021). *GameStop Skyrockets as Retail Investors Force Short Squeeze*. CNBC. Accessed: May 28, 2021. [Online]. Available: <https://www.cnbc.com/video/2021/01/27/gamestop-skyrockets-as-retail-investors-force-short-squeeze.html>
- [4] *Melvin Capital Lost 53% in January, Hurt by GameStop and Other Bets—WSJ*. Accessed: May 28, 2021. [Online]. Available: <https://www.wsj.com/articles/melvin-capital-lost-53-in-january-hurt-by-gamestop-and-other-bets-11612103117>
- [5] H. Gao, G. Barbier, and R. Goolsby, "Harnessing the crowdsourcing power of social media for disaster relief," *IEEE Intell. Syst.*, vol. 26, no. 3, pp. 10–14, Jun. 2011.
- [6] X. Zheng, D. Zeng, and F.-Y. Wang, "Social balance in signed networks," *Inf. Syst. Frontiers*, vol. 17, no. 5, pp. 1077–1095, 2015.
- [7] X. Zheng, Y. Zhong, D. Zeng, and F.-Y. Wang, "Social influence and spread dynamics in social networks," *Frontiers Comput. Sci.*, vol. 6, no. 5, pp. 611–620, Sep. 2012.
- [8] J. S. Juris, "Reflections on# occupy everywhere: Social media, public space, and emerging logics of aggregation," *Amer. Ethnol.*, vol. 39, no. 2, pp. 259–279, 2012.
- [9] X. Zheng *et al.*, "Donald J. Trump's presidency in cyberspace: A case study of social perception and social influence in digital oligarchy era," *IEEE Trans. Comput. Social Syst.*, vol. 8, no. 2, pp. 279–293, Apr. 2021.
- [10] J. Mattke, C. Maier, L. Reis, and T. Weitzel, "Herd behavior in social media: The role of Facebook likes, strength of ties, and expertise," *Inf. Manage.*, vol. 57, no. 8, Dec. 2020, Art. no. 103370, doi: [10.1016/j.im.2020.103370](https://doi.org/10.1016/j.im.2020.103370).
- [11] D. S. Scharfstein and J. C. Stein, "Herd behavior and investment," *Amer. Econ. Rev.*, vol. 80, no. 3, pp. 465–479, Jun. 1990.
- [12] S. Stieglitz and L. Dang-Xuan, "Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior," *J. Manage. Inf. Syst.*, vol. 29, no. 4, pp. 217–248, Apr. 2013.
- [13] L. H. Pedersen, "Game on: Social networks and markets," 2021. [Online]. Available: <https://ssrn.com/abstract=3794616>
- [14] A. Tengulov, F. Allen, E. Nowak, and M. Pirovano, "Squeezing shorts through social news platforms," Swiss Finance Inst., Zürich, Switzerland, Tech. Rep. TR-0195 (21-31)-3, 2021.
- [15] E. Vasileiou, "Does the short squeeze lead to market abnormality and anti-leverage effect? Evidence from the GameStop case," Apr. 2021. [Online]. Available: <https://ssrn.com/abstract=3831619>, doi: [10.2139/ssrn.3831619](https://doi.org/10.2139/ssrn.3831619).
- [16] J. J. Angel, "Gamestonk: What happened and what to do about it," 2021.
- [17] C. Long, B. M. Lucey, and L. Yarovaya, "'I just like the stock' versus 'fear and loathing on main street': The role of Reddit sentiment in the GameStop short squeeze," 2021. [Online]. Available: <https://ssrn.com/abstract=3822315>, doi: [10.2139/ssrn.3822315](https://doi.org/10.2139/ssrn.3822315).
- [18] M. P. Baker and J. A. Wurgler, "Investor sentiment in the stock market," *SSRN Electron. J.*, vol. 21, no. 2, pp. 129–152, 2007.
- [19] S. Liu, "Investor sentiment and stock market liquidity," *J. Behav. Finance*, vol. 16, no. 1, pp. 51–67, Jan. 2015.
- [20] M. Zouaoui, G. Nouyrigat, and F. Beer, "How does investor sentiment affect stock market crises? Evidence from panel data," *Financial Rev.*, vol. 46, no. 4, pp. 723–747, Nov. 2011.
- [21] C.-C. Chen, H.-H. Huang, and H.-H. Chen, "NTUSD-Fin: A market sentiment dictionary for financial social media data applications," in *Proc. 1st Financial Narrative Process. Workshop (FNP)*, 2018, pp. 37–43.
- [22] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
- [23] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, 1975.
- [24] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 8, no. 1, 2014, pp. 216–225.
- [25] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, nos. 66–84, pp. 440–442, 1998.
- [26] M. E. J. Newman, "Mixing patterns in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 67, no. 2, Feb. 2003, Art. no. 026126.
- [27] K. Crain, K. Heh, and J. Winston. (2016). *An Analysis of the Elite Users on Yelp*. [Online]. Available: <https://snap.stanford.edu/class/cs224w-2014/projects2014/cs224w-32-final.pdf>
- [28] S. J. Hardiman and L. Katzir, "Estimating clustering coefficients and size of social networks via random walk," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 539–550.
- [29] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, 2009.
- [30] H. Tian, X. Zheng, and D. D. Zeng, "Analyzing the dynamic sectoral influence in Chinese and American stock markets," *Phys. A, Stat. Mech. Appl.*, vol. 536, Dec. 2019, Art. no. 120922.
- [31] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
- [32] M. Wattenberg, F. Viégas, and I. Johnson, "How to use t-SNE effectively," *Distill*, vol. 1, no. 10, p. e2, Oct. 2016.
- [33] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, vol. 2008, no. 10, Oct. 2008, Art. no. P10008.
- [34] B. Lal, E. Ismagilova, Y. K. Dwivedi, and S. Kwayu, "Return on investment in social media marketing: Literature review and suggestions for future research," in *Digital and Social Media Marketing*. New York, NY, USA: Springer, 2020, pp. 3–17.
- [35] M. Clark and J. Melancon, "The influence of social media investment on relational outcomes: A relationship marketing perspective," *Int. J. Marketing Stud.*, vol. 5, no. 4, p. 132, Jul. 2013.
- [36] Y.-X. Peng *et al.*, "Cross-media analysis and reasoning: Advances and directions," *Frontiers Inf. Technol.*, vol. 18, no. 5, pp. 44–57, 2017.
- [37] A. Buyalskaya, M. Gallo, and C. F. Camerer, "The golden age of social science," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 5, pp. 1–11, 2021.

- [38] Š. Lyócsa, E. Baumohl, and T. Výrost, “YOLO trading: Riding with the herd during the GameStop episode,” to be published.



**Xiaolong Zheng** (Member, IEEE) received the B.S. degree from China Jiliang University, Hangzhou, China, in 2003, the M.S. degree from Beijing Jiaotong University, Beijing, China, in 2006, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 2009.

He is currently a Professor with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences. He has authored more than 120 peer-to-peer academic articles in journals/magazines. His research interests include social computing, big data analytics, knowledge graphs, financial technologies, and complex system intelligence.

Dr. Zheng has served as a Program Co-Chair of six international conferences and a Technical Program Committee Member for more than 40 international conferences including IJCAI, AAAI, and WWW.



**Hu Tian** received the B.S. degree in mathematics from the School of Mathematics and Physics, Chinese University of Geosciences, Wuhan, China, in 2018. He is currently pursuing the Ph.D. degree with the Institute of Automation, Chinese Academy of Sciences, Beijing, China.

His research interests include machine learning and data mining, particularly in machine learning techniques and applications such as graph neural networks, adversarial learning, FinTech, and social computing.



**Zhe Wan** is currently a Professor with the Belt and Road School, Beijing Normal University, Beijing, China. She used to be the Chief Economist of the International Cooperation Centre of National Development and Reform Commission, Beijing, and China National Gold Group Corporation, Beijing. She has authored *State Financial Strategy*, which was recommended in the reading listing of China's state organs reading activities in economics books. She has coauthored *State Financial Safety*, which won the 4th Financial Publication Jinyang Award in 2015 and Hexun Chinese Business & Finance Publication Award (Financial). She has authored a series of articles and research articles on financial strategy, green economy, “The Belt and Road,” and so on and on significant academic publications and journals. She is specialized in the research of international and domestic economic matters, appointed as a special contributor or a commentator/columnist to a number of most significant media agencies, including Xinhua News Agency-affiliated Liaowang Institute, China Financial Weekly, People’s Daily’s Economic Focus, Xikake\_island, Global People, China News Weekly, China News Service, The Paper, Global Finance, and the Financial Times. Her research fields include macro-economy, financial strategy, commodity price, international affairs, overseas investment, and market strategy research.

Ms. Wan is the Vice Chairman of IFAC Social, Economic, and Financial Technical Committee, a member of Global Future Council, and the Standing Director of China Urban Financial Society. She is the Financial Commentator of CCTV.



**Xiao Wang** (Member, IEEE) received the bachelor’s degree in network engineering from Dalian University of Technology, Liaoning, China, in 2011, and the Ph.D. degree in social computing from the University of Chinese Academy of Sciences, Beijing, China, in 2016.

She is currently an Associate Professor with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing. She has authored or coauthored more than a dozen of SCI/EI articles and translated three technical books (English to Chinese). Her research interests include social transportation, cyber movement organizations, artificial intelligence, and social network analysis.

Dr. Wang served as a Peer Reviewer, with a good reputation, of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE/CAA JOURNAL OF AUTOMATICA SINICA, and ACM Transactions of Intelligent Systems and Technology.



**Daniel Dajun Zeng** (Fellow, IEEE) received the B.S. degree in economics and operations research from the University of Science and Technology of China, Hefei, China, in 1990, and the M.S. and Ph.D. degrees in industrial administration from Carnegie Mellon University, Pittsburgh, PA, USA, in 1994 and 1998, respectively.

He is currently a Research Fellow with the Institute of Automation, Chinese Academy of Sciences, Beijing, China. He has authored more than 300 peer-reviewed articles. His research interests include intelligence and security informatics, infectious disease informatics, social computing, recommender systems, software agents, and applied operations research and game theory.

Dr. Zeng currently serves as the Editor-in-Chief of ACM Transactions on MIS.



**Fei-Yue Wang** (Fellow, IEEE) received the Ph.D. degree in computer and systems engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990.

He joined The University of Arizona, Tucson, AZ, USA, in 1990, and became a Professor and the Director of the Robotics and Automation Laboratory and the Program in Advanced Research for Complex Systems. In 1999, he founded the Intelligent Control and Systems Engineering Center, Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, under the support of the Outstanding Overseas Chinese Talents Program from the State Planning Council and 100 Talent Program from CAS. In 2002, he joined the Laboratory of Complex Systems and Intelligence Science, CAS, as the Director, where he was the Vice President of Research, Education, and Academic Exchanges with the Institute of Automation from 2006 to 2010. In 2011, he was named the State Specially Appointed Expert and the Director of the State Key Laboratory for Management and Control of Complex Systems, Beijing. His current research interests include methods and applications for parallel systems, social computing, parallel intelligence, and knowledge automation.