frontiers in Complex Systems

TYPE Original Research
PUBLISHED 02 November 2026
DOI 10.3389/fcpxs.2026.1924415

Check for updates

OPEN ACCESS

EDITED BY
Naoki Masuda,
University of Michigan, United States

REVIEWED BY
Valeria Mazzeo,
Bruno Kessler Foundation (FBK), Italy
Antonino Ferraro,
Pegaso University, Italy

*CORRESPONDENCE
Anna Mancini,
anna.mancini@cref.it
Giulio Cimini,
giulio.cimini@roma2.infn.it

RECEIVED 30 June 2026
REVISED 05 October 2026
ACCEPTED 10 October 2026
PUBLISHED 02 November 2026

CITATION
Mancini A, Desiderio A, Palermo G, Di Clemente R and Cimini G (2026) The rise and fall of WallStreetBets: social roles and opinion leaders across the GameStop saga.
Front. Complex Syst. 4:1924415.
doi: 10.3389/fcpxs.2026.1924415

COPYRIGHT
© 2026 Mancini, Desiderio, Palermo, Di Clemente and Cimini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The rise and fall of WallStreetBets: social roles and opinion leaders across the GameStop saga

Anna Mancini[1,2]*, Antonio Desiderio[1,2,3], Giovanni Palermo[2,4], Riccardo Di Clemente[3,5,6] and Giulio Cimini[1,2]*

[1]Physics Department and INFN, University of Rome Tor Vergata, Rome, Italy, [2]Centro Ricerche Enrico Fermi, Rome, Italy, [3]ISI Foundation, Turin, Italy, [4]Physics Department, Sapienza University of Rome, Rome, Italy, [5]Complex Connections Lab, Network Science Institute, Northeastern University London, London, United Kingdom, [6]The Alan Turing Institute, London, United Kingdom

Nowadays human interactions largely take place on social networks, where online users' behavior often falls into a few general typologies or "social roles". Among these, opinion leaders are of crucial importance as they have the ability to spread an idea or opinion on a large scale across the network, with potential tangible consequences in the real world. In this work we extract and characterize the different social roles of users within the WallStreetBets community on Reddit, around the time of the GameStop short squeeze of January 2021 – when a handful of committed users led the whole community to engage in a large and risky financial operation. We discern the characteristics needed to play a particular social role and we identify the profiles of average users and relevant outliers, including candidate opinion leaders, using an iterative, semi-supervised classification algorithm. The key features of these users are making large risky investments and providing constant updates on a single stock, which attracted a large following and, in the case of GameStop, sparked the interest of the community. After the short squeeze event, we observed a substantial change in user behavior and attitude: no new opinion leaders emerged and the community became less focused on investments. Overall, this work sheds light on the users' roles and dynamics that led to the GameStop short squeeze, while also suggesting why WallStreetBets no longer wielded such large influence on financial markets, in the aftermath of this event.

KEYWORDS
collective investment, GameStop, opinion leaders, reddit, short squeeze, social roles, user behavior, WallStreetBets

Introduction

We are now in a time where the quest of understanding human behavior has entered a highly computational and quantitative era (Conte et al., 2012), thanks to the availability of data from social media and new technologies that capture interactions at the level of individual users over extended periods of time (Lazer et al., 2009). We can collect e-mail data (Gloor et al., 2017), credit card data (Di Clemente et al., 2018), records of phone calls (Vieira et al., 2010) and all kinds of digital traces an individual leaves behind when roaming in the online world (Olejnik et al., 2014). The availability of a huge amount of microscopic data allows us to deepen our understanding of human behavior by employing powerful scientific methods, such as those from statistical physics, applied to a different phenomenon: people (Castellano et al., 2009; Cimini et al., 2019). Remarkably, social media and web technologies not only provide us with such richness of data, they bring human interaction itself to a

different dimension. Interactions between millions of individuals are now possible, people can be virtually connected to everybody and everything with the touch of a finger, accessing and sharing information almost instantly with the rest of the world (Aichner et al., 2021). These factors give rise to new emergent phenomena: polarization (Bail et al., 2018), diffusion of fake news (Lazer et al., 2018), echo chambers (Cinelli et al., 2021), protests and revolutions (Scherman et al., 2015; Khondker, 2011), mass retail investment (Semenova and Winkler, 2025), just to name a few.

Among the various social networks that have been analysed in the literature, Reddit represents an ideal playground (Desiderio et al., 2025a; Balsamo et al., 2021), because it combines large-scale, user-generated text data with topic-organized communities like politics (Hofmann et al., 2022), cryptocurrencies (Wooley et al., 2019) and many more. Users are pseudonymous, which mitigates some of the direct-identification concerns present on identity-linked platforms. The interest in Reddit has risen to par with other platforms like Facebook and Twitter due to its increase in popularity caused by the GameStop (GME) short squeeze. In January 2021, users of the Reddit community WallStreetBets (WSB) coordinated in a mass effort to buy and hold GameStop shares, which drove the price to a 30-fold growth in less than a month, causing large losses for all hedge funds who were profiting by short-selling GME (Chohan, 2022; Mancini et al., 2022).

Such an unprecedented event attracted the attention of the academic community, resulting in a stream of literature that analyzed different aspects of the episode. For instance, Lucchini et al. (2022) studied the commitment of WSB users towards the short squeeze, finding how a small fraction of users may have triggered all others to follow through buying GME shares. Zheng et al. (2022) investigated the social dynamics of the event by looking at interaction networks, topic modeling and sentiment analysis. Kim et al. (2023) examined trading patterns of individual investors, finding that social sentiment enhanced the collective behavior. Haq (2022) analyzed the different behaviors of "old" users versus the behavior of those who joined following the success of the short squeeze. Mancini et al. (2022) looked for the early warnings of the event and explored an opinion dynamic model to qualitatively reproduce the sudden emergence of consensus in the community.

More recent studies have focused on the predictive value of stock-related text in online social networks (Buz and de Melo, 2024; Buz et al., 2024; Verdear et al., 2025). While the broader mechanism through which socially transmitted information can shape investment behaviour—encompassing herding, social learning, information cascades, and financial-market information diffusion—is well established in the literature (Ferraro and Sperlì, 2024; Matsumoto et al., 2025; Warkulat and Pelster, 2024). Our contribution is to develop and apply a data-driven framework that identifies and characterizes the individual actors through whom such collective dynamics emerge in a specific community, and to describe the behavioral attributes that distinguish these actors from the average user. In this sense, the framework is complementary to existing work on aggregate community dynamics: it operates at the level of individuals and their observable behavior, and it yields a behavioral characterization of what makes a user earn a prominent social role within a community organized around collective investment.

The concept of *social role*, defined as a typical behavior associated with a social position, is well known in the social sciences literature (Biddle, 1986). Some studies have already tackled the issue of extracting and studying these roles on different social networks. For example, by analyzing the editing history of users, Welser (2011) found that it is possible to identify four key social roles in the community of Wikipedia page editors. While the online questionnaire put together by Brandtzaeg and Heim (2011) informed us on the typology of Norwegian social network users based on their online habits. Moreover, the study carried out by Recuero et al. (2019) exploits network topology to identify Twitter user roles and their influence in political discussions by focusing mainly on those with high degree, while Amato et al. (2017) proposed an approach based on hypergraphs to identify user roles, introducing a new centrality measure used to detect lurkers (users who read but do not actively participate in an online community). Refer to the survey by Jin et al. (2013) for a review of the main topics of literature regarding user behavior, from the analyses of interactions on a social graph to the identification of malicious conduct.

However, these works either rely on ad-hoc techniques (such as questionnaires) and/or are based on a predetermined set of user roles to be detected, whose existence is then verified on the data available. To overcome these limitations, we develop a framework where user roles emerge "naturally" through a semi-supervised procedure which can be tuned for various contexts, starting from a set of individual features that reflect user activity and the quality of the content created. In practice, we detect the main social roles of "average" users and their temporal dynamics by informing a DBSCAN (Ester et al., 1996) algorithm with the selected features. Furthermore, by recursive application of the methodology, we can identify a handful of *outliers*: "special" users who, due to a behavior that consistently deviates from the average, possess peculiar features and social statuses.

A user role of special interest is the so called *opinion leader* (Katz, 1957): a hub of the community who is able to directly influence other members. Recently, different approaches have been employed to identify these types of users on social media. Sun et al. (2016) singled them out by analysing multiple selected features from Weibo users. Winter and Neubaum (2016) directly asked Facebook users how influential they consider themselves. Riquelme et al. (2019) defined a new centrality measure based on activity and interest of users over a specific topic to identify opinion leaders on Twitter. Haase et al. (2023) studied how financial influencers on Twitter affect or are affected by the overall sentiment towards a financial asset using Granger causality. These works are limited to social networks where a user's identity is known, and thus their influence may also be a consequence of the individual's pre-existing popularity. Given Reddit's pseudonymous accounts, individual behavioral patterns can be observed and characterized without conflation with off-platform identity, allowing us to identify well-grounded features that contribute to achieving the status of opinion leader.

In the case of WSB, where the major topic of interest is whether to invest or not in a given stock, opinion leaders can be considered as those who, through their investments, achieve a large following and sway others to invest in their same stock. By triggering large

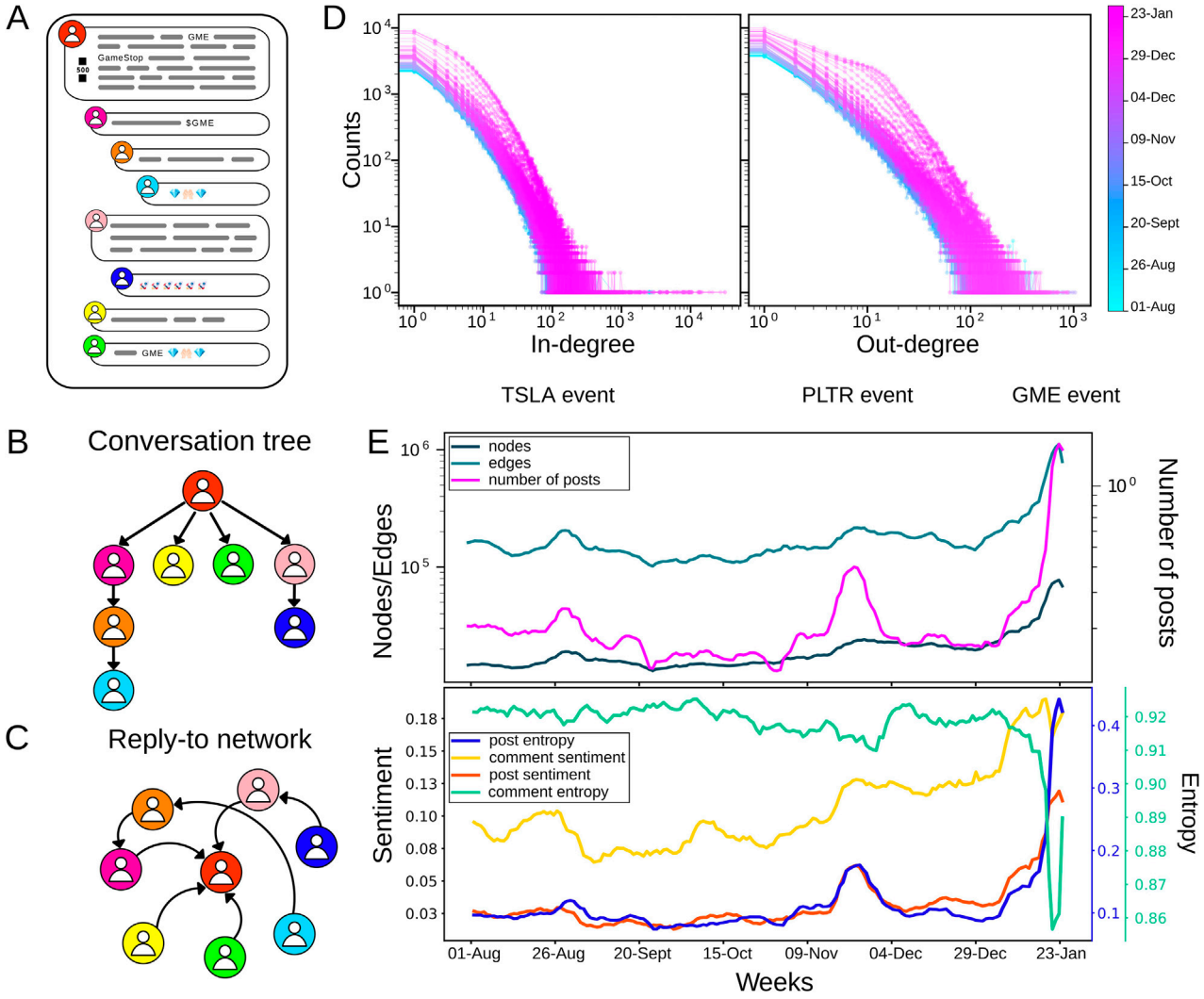


FIGURE 1
Construction of conversation networks and time evolution of user features on WSB. **(A)** Example of a Reddit conversation thread, made up of the post and its comments. **(B)** Conversation tree extracted from the thread. **(C)** User network built from the "reply-to" interactions on the thread. **(D)** In- and out-degree distributions for all weekly user networks, from 1 August 2020 to 31 January 2021. We can see the presence of some highly connected hubs in the in-degree distribution, i.e., users whose comments and posts have received many replies. **(E)** Upper panel: nodes and edges of the user network in time. The peaks of these quantities are reflected in those of the number of posts created, which correspond to some engaging events that attracted the attention of the entire community around some stocks (see the text for event description). Lower panel: sentiment and entropy of posts and comments in time. Both post and comment sentiment increase when the event occurs. Instead comment entropy diminishes since the text of comments becomes more repetitive, while post entropy peaks because of the presence of longer posts analyzing and explaining the situation.

collective investments, opinion leaders can potentially succeed in influencing the financial market.

Through our semi-supervised framework, we are able to identify user roles and in particular those who qualify as opinion leaders of WSB across the GME saga–explaining which features allowed these users to achieve their prominent status. By analyzing their activity we find that each opinion leader is linked to a particular stock investment, and we are able to determine how the "GME leader" was able to trigger massive consensus, which contributed to the success of the GME short squeeze. Finally, we detect how roles change after the events of January 2021. Indeed the notoriety that WSB achieved in the aftermath of the GME short squeeze changed the very nature of the community, making discussions less focused on finance-related topics and more on memes–suggesting the community may have lost its potential to trigger successful collective retail investments.

Taken together, these findings characterize two distinct kinds of emergent behavior on WSB. The first is the formation of behavioral roles: individual patterns of interaction on the subreddit emerge as a small number of stable feature-space clusters. The second is the concentration of collective attention, in which internal and external circumstances pull the community towards particular stock tickers, producing the coordinated engagement that characterizes events like the GME short squeeze. Distinguishing these two levels—user-level roles and community-level attention—is what allows our framework to trace how a handful of opinion leaders can, under

the right conditions, translate individual behavior into collective action.

Results

The subreddit WallStreetBets (WSB) is a community of mostly amateur traders (Boylston et al., 2021) who like to discuss high risk investments and post about their gains and losses. WSB users can post new content and comment under existing posts, resulting in tree-like conversation threads, like the one shown in Figure 1A. The community has specific rules, one of which states that any claim about an investment has to be supported by a visual proof of the latter, i.e., a screenshot, which gives us valuable information about when such an investment was made. Particular members of the community, called *moderators*, ensure that rules are observed by all users. Moderators thus have a specific social role in the community, that shapes their behavior, and we can identify them by looking at previous literature (Zheng et al., 2022) and inspecting user's comments (see Supplementary Materials S1 and S3). The time period that we consider comprises all posts and comments written by sufficiently active users on the community from 1 August 2020 to 31 July 2021 (see Methods for the precise count of comments and posts).

The rise of WallStreetBets

We now focus on the period preceding the short squeeze (August 2020 - January 2021). After a data cleaning procedure to remove bots and inactive users (see Methods) we construct conversation trees, where nodes are posts and comments and a link indicates a direct reply. From such structures we then extract user interaction networks, where users are connected according to how many direct "reply-to" actions occurred between them. This procedure allows the formation of self-links, as a user's comment on his/her own post still counts as an activity. Figure 1B shows the tree describing the thread of Figure 1A, while Figure 1C is the resulting user "reply-to" network. We obtain weekly networks by considering all "reply-to" actions occurring within a week, and shifting the 7-day moving window by 1 day each time, until covering the entire timespan of data (see Methods for further information on how these networks are created and Supplementary Material S6.2 for further considerations on the moving window applied). Figure 1D shows the out- and in-degree distributions of weekly networks, which display a power-law trend. The tail of the in-degree distribution grows of orders of magnitude around January, as a result of the growing number of new users and the appearance of highly connected hubs of the community. For the out-degree distribution we see a bump forming as we approach January, meaning in that period the majority of new users write around 20 or less comments per week.

To check how the overall behavior of users changes in time, we look at the evolution of the number of nodes and edges in the weekly user networks. In the upper panel of Figure 1E we see how both quantities increase in certain weeks–when also the number of posts peaks. Inspecting the text of posts we can see how this increase is due to the community's reaction to some event of major interest (Desiderio et al., 2025a). We can identify, in temporal order: the stock split of Tesla (TSLA) in August 2020, investor Steve Cohen disclosing a large stake in Palantir (PLTR) in November 2020 (Whiteman, 2020) and the GME frenzy due to the short squeeze in January 2021. Unsurprisingly, posts are the best proxy for the changing interests of the community, and are the primary means though which users share and spread information. Therefore post features play a crucial role in understanding behavioral changes sustained by the community.

Our main goal is to develop a semi-supervised approach to extract a reliable characterization of user profiles from all the available information we can collect on the way they interact online. Therefore we start by selecting a set of relevant features that capture a user's commenting and posting behavior, the complexity of the comments and posts written and the impact they have on the community. We thus collect features of posts, comments and "reply-to" networks for each user in each week of our dataset (refer to Methods for the complete list and description). In the lower panel of Figure 1E we show the time evolution of a subset of these quantities summarizing the textual complexity of posts and comments written, i.e., the entropy, and the tone that posts and comments convey, i.e., the sentiment, to show how the community as a whole changes in volume, tone and complexity of content posted when particular situations occur. While the sentiment and post entropy all display peaks during the highlighted events, the comment entropy has an opposite trend, since comments tend to become repetitive and plain during an engaging event (Desiderio et al., 2025a).

User profiles

The analyses reported in Figure 1 show how the community shifts altogether. Now we turn to a microscopic level where the focus is on the behavior of single users and how it changes in time. Differently from other social networks, Reddit users are mostly anonymous. It has been shown that this peculiarity results in a more dis-inhibited and honest behavior (Gagnon, 2013; Brown et al., 2018). This leads to wonder: does anonymity influence the actions and *roles* of users? Or do we find patterns common to other online social networks? Looking at the number of contributing users and of subscribers to the community, we noticed what is known as the *1% rule*: in agreement with the current literature (Buntain and Golbeck, 2014; van Mierlo, 2014), we find that only 1% of users contribute to new content on the community. Interestingly, this behavior persists even with the growth of the community around January (see Supplementary Material S5 for more details).

To identify user roles, we represent each user by her feature vector and apply the well-known DBSCAN clustering algorithm (Ester et al., 1996) to the 10-component PCA projection of the feature vectors of each week (see Methods for the full explanation of the procedure). We also apply spectral clustering (Ng et al., 2001) to test the coherence of the obtained results, and achieve an extremely good agreement between the clusters identified by both algorithms (see Supplementary Material S12). Figure 2A shows a two-dimensional t-SNE projection (van der Maaten and Hinton, 2008) of the feature space for six representative weeks, with each user colored according to the cluster to which DBSCAN assigned them (green: largest cluster; pink and orange: next two largest; grey: points classified as noise by DBSCAN). The projection is

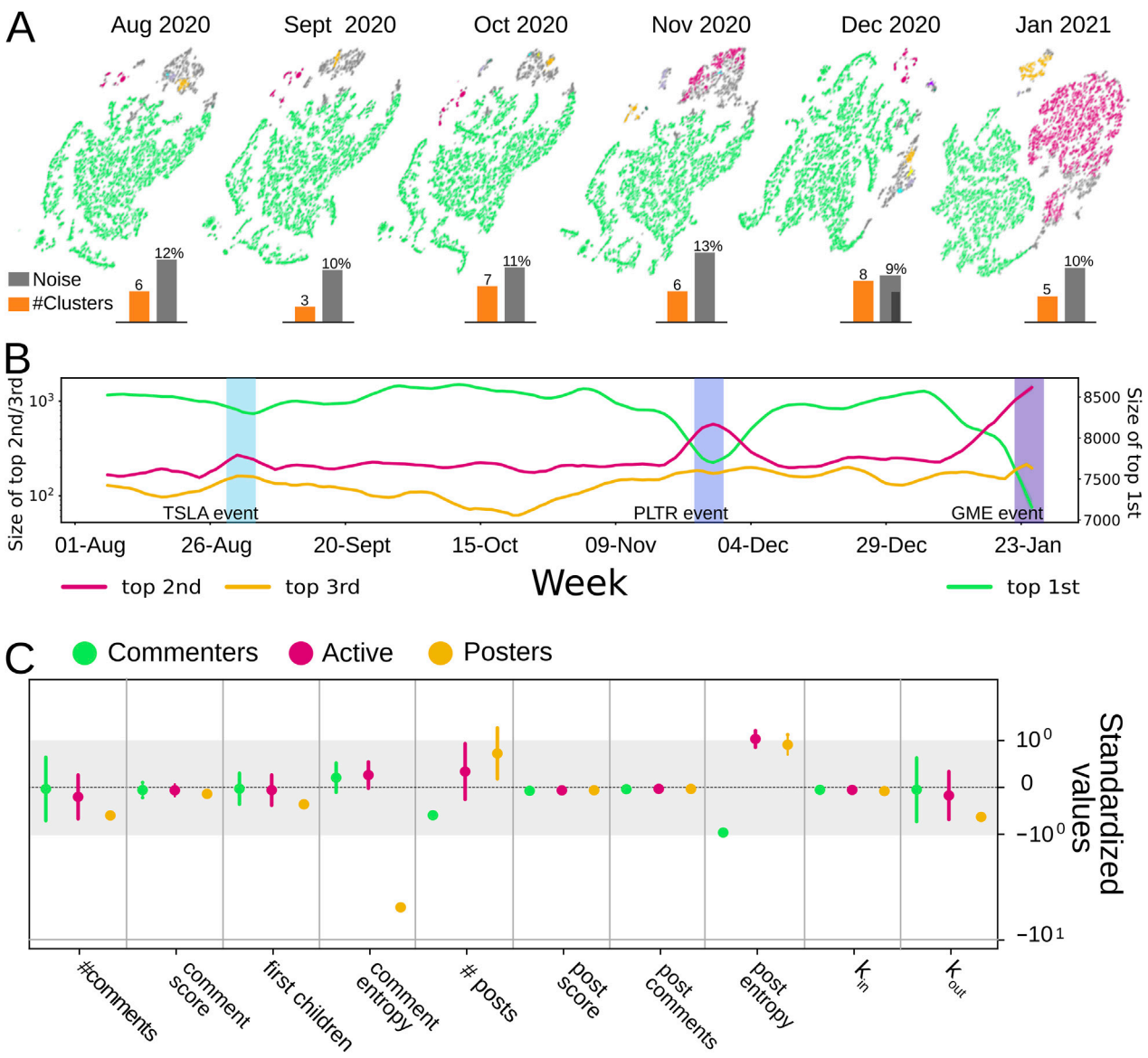


FIGURE 2
User profiles. **(A)** T-SNE projections of user clusters found by DBSCAN for the weeks of 24–30 August 2020, 24-30 September 2020, 24-30 October 2020, 24-30 November 2020, 24-30 December 2020 and 24-30 January 2021. Three largest clusters are visible (green: *commenters*, pink: *active*, orange: *posters*), while grey points belong to noise. For each realization we specify the percentage of noisepoints and the number of clusters. The weeks of November and January correspond to engaging events, and we see how clusters become fewer and well separated. **(B)** Time evolution of the sizes of the top three clusters. In the week of the event we have a "flow" of users from the top cluster (commenters) to the active cluster. In fact when an event happens that catches the interest of the community, some of those who normally comment start writing posts and consequently become active. **(C)** Average value of the feature vectors for the three biggest clusters in the week from the 24th to the 30th of January 2021. We can see how each cluster is characterized by a particular profile of the feature values, which allows us to understand the average behavior of users belonging to such cluster (for the sake of readability we show only selected features on the x axis).

used here purely for visualization, to allow qualitative inspection of the cluster structure identified by DBSCAN on the PCA-projected features. The time evolution of the clusters outlines how the top three increase in size while some of the smaller ones disappear in correspondence with events that drew the attention of the community, as can be seen from 2-D projections (November 2020: PLTR week, January 2021: GME week). In Figure 2B we track the time evolution of the size of the three largest clusters (at least an order of magnitude larger than all others). The figure suggests the occurrence of a flow of users from the largest cluster to the other two when the week of the engaging event approaches.

These three largest clusters are consistently identified by DBSCAN across all weeks, together with a set of smaller clusters and noise points. We characterize the top three clusters by examining the feature-value profiles of their members, and in order of cluster size we identify: *commenters*, users whose post features have very low values; *active users*, with high values of both comment and post features; *posters*, users whose comment features have very

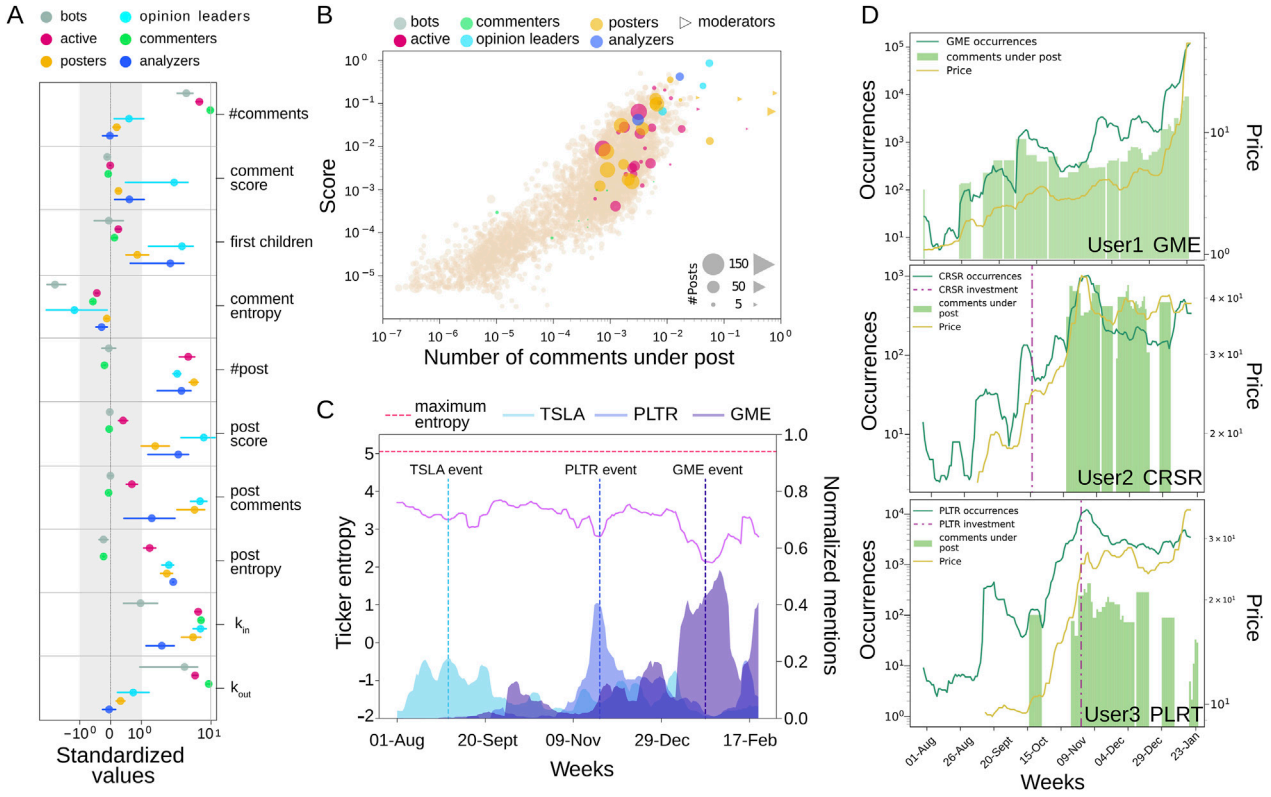


FIGURE 3
Outlier groups. **(A)** The distinct profiles of feature values for the various outlier groups (for the sake of readability not all features are shown on the y axis). **(B)** Average number of comments under posts versus average post score for all users with more than 10 posts, normalized by the total number of daily active users. A point corresponds to a user with size proportional to the number of posts written. Each group occupies a well defined region of the plot, with opinion leaders placed in the top right corner and commenters lying in the bottom left one. **(C)** Time evolution of the ticker entropy (violet solid line), which decreases each time an engaging event occurs. The minimum is reached on the week of the GameStop short squeeze. The dashed pink line is the maximum value the ticker entropy can assume. The shaded regions correspond to the normalized number of mentions for each of the three engaging tickers. **(D)** For each of the three opinion leaders we report the number of comments generated by their posts (light green bars), the number of occurrences of their chosen stock ticker in content created by the community (dark green) and the closing price of the stock (yellow). For User2 and User3 we also highlight when the first investment was made (purple dashed line), while User1 made the investment in 2019. All quantities grow after the investment, with an increase in the number of comments under their posts right after the increase of the stock price, indicating that these users become more popular after achieving considerable gains.

low values. The average feature values of the three categories can be seen in Figure 2C for the last week of January (when all three clusters have a high density of points). Interestingly, the *commenter*, or answer-person, behavior has been found in other social networks (Welser et al., 2007), as well as in other Reddit communities (Buntain and Golbeck, 2014), suggesting this as being a very general behavior of online users. In our context, when an event of interest to WSB users occurs, some of the commenters decide to change role and write posts to effectively reach more people, hence becoming active. Many users from smaller clusters (like the blue one in Figure 2A) move into one of these three categories, thus reducing the heterogeneity of behaviors. Moreover, during the week of the event, active users tend to have a higher textual entropy, both in comments and posts, compared to the other two groups. This might suggest that these users are the ones that facilitate the flow of conversations with a constant input of personal opinions and information. For an analysis focused on persistent users and their flow between different clusters see Supplementary Material S13.

Outliers

We have so far identified the average user behavior on WSB, but we know that there are users that do not fit into any of these average profiles. To investigate those individuals whose behavior diverges from the average conduct and understand what differentiates them from the rest, we shift our attention to noise points, i.e., users that do not belong to any cluster. Through an iterative clustering-removal procedure (see Methods for a detailed description and Supplementary Material S7 for robustness checks) we are able to rank users according to how far they lie from the clusters identified by DBSCAN, the largest of which represent the average user profiles. In this way we can select those individuals with a persistent diverging behavior and, by labeling each profile depending on its feature attributes, we can distinguish different roles–as reported in Figure 3A. In particular we can still identify profiles in line with the commenters, posters and active user roles found in the average user analysis, but with extreme values of their characterising features (for instance, a poster who wrote a tremendous number of posts). We also find some new interesting roles: *analyzers*, users who write posts

about what is trending on the community and share updates on market news; *bots*, those that were not removed by the initial filter; *opinion leaders*, users whose posts and comments have a great following, but who do not comment and post as frequently as some others do. In Figure 3B we plot the average number of comments under posts contributed by each user versus the average score of such posts (i.e., how many upvotes minus downvotes they receive), for all users with more than 10 posts (both quantities are normalized by the total number of daily active users). Hence each user is a point in the graph, with size proportional to the number of posts written, while the outlier groups are represented in a different color. All three opinion leaders found (light blue) lie in the upper right corner: even if they do not write too many posts, the content they post is always appreciated and generates ample discussion. Triangles instead denote the outliers who are also moderators of the community, and whose profile fits well with the "posters" role. Posters (orange) are found in the right-hand side of the plot and have the largest sizes, as their posts are more frequent and characterized by a high number of comments. Active users (pink) are similar to posters, but their sizes on average tend to be smaller since these users post less. Commenters (green) are mostly in the lower left side of the plot, as they are the ones that post less (as can be seen by their small size) and whose content does not draw much attention. Bots are not found in the plot, since they never write posts. Finally, analyzers (blue) are found in the upper right corner of the plot, highlighting how much their content is appreciated; they have medium sizes since they tend to write posts periodically to update the community. The characterization of these profiles is consistent with results obtained by an ex-post Multinomial Logistic Regression on the different user groups to identify the relevant features (see Supplementary Material S11). The latter approach also highlights that in order to single out the outliers it is not enough to select just a couple of features like following and score (while disregarding the others), as indeed a combination of multiple features is needed to characterize the different groups. In fact in Supplementary Material S10 we show that by selecting just the two most influential features for opinion leaders, score and number of comments under post, we are not able to reproduce the same results obtained with DBSCAN.

Opinion leaders

One of the main issues when trying to identify opinion leaders on social networks is topic limitation. In most cases an opinion leader is an individual with expertise in a particular subject (Richmond, 2006; Grewal et al., 2000; Flynn et al., 1994), and thus does not exert the same authority when speaking of a different matter. Trying to find the topic of interest may be a challenging task (Yang et al., 2017; Miao et al., 2016), which could result in biased or erroneous results. This problem is minimized on Reddit since it is structured into communities, each regarding a given subject, and users who comment on these communities are to some extent already knowledgeable on the matter. In addition, when trying to identify influential users, the starting point is usually to see which ones satisfy some fixed notion of *influence* (Kiss and Bichler, 2008), be it in terms of network position (Bodendorf and Kaiser, 2009), feature values (Sun et al., 2016) and so on. Here instead we want to extrapolate the concept of influence from a behavior we observe, and thus spontaneously deduce which are the actions that allow users to achieve such a position.

We focus on the posting behavior of the three candidate opinion leaders (which we refer to as User1, User2 and User3) to identify the attributes that have earned them such visibility. Firstly we noticed that their contributed posts with a large following are all linked to a large investment by the user on a particular stock–respectively on GME, CRSR (Corsair Gaming Inc) and PLTR. Figure 3D shows several quantities related to such users and their chosen stock: the number of comments under each post (light green bars), the number of mentions of the stock ticker in the community (dark green) and the closing price of the stock (yellow). The dashed-dotted line indicates when the first investment was made, except in the case of User1, since his investment dates back to 2019 (not covered by our data).

After the first investment, we observe distinct patterns for the three leaders. For User1, all three quantities grow steadily through the observation period, reflecting his sustained influence in the run-up to the GameStop event. For User2, the ticker mentions, comments received, and stock price all rise following his first investment post, indicating that his post preceded a rise in community attention and market activity above what was happening across the community at the same time. For User3, whose first post coincided with the peak of an externally-driven wave of PLTR attention (following Steve Cohen's November 2020 disclosure), the ticker mentions and comments received decline in the days after his post, even if PLTR's stock price continues to rise, indicating a very distinct route to opinion-leader status with respect to the first two users. Quantitative comparisons of these changes against community-wide baselines are reported in Supplementary Material S6. Across the three cases, three behavioral attributes recur that are consistent with the leaders' emergence as prominent figures in the community: i) Extreme action: all investments are *YOLOs*, i.e., large investment into a not-so-popular stock; such a high risk-high reward approach is highly appreciated by the community, as its name WallStreet*Bets* suggests. ii) Persistent posting behavior: once the investment is made, all leaders start updating the community on their progress (gains or losses). iii) Success: the investment has to eventually bring profit to the user, which has happened in all three cases, as can be seen from their investment updates. In this particular case study, these attributes make a single user's content increasingly visible and appreciated, which in turn draws further attention and replies—a self-reinforcing dynamic consistent with social contagion and herding mechanisms documented in retail-investment (Semenova and Winkler, 2025; Štefan et al., 2022; Smith et al., 2025) and online-collective-action settings (Desiderio et al., 2025b). In this sense, the individual behavior of a handful of users, amplified through the community's reply dynamics, provides the micro-level mechanism underlying the macro-level concentration of attention that preceded the short squeeze.

We note that the third attribute (demonstrable success) cannot be fully disentangled from the detection method: opinion leaders are identified through features tied to community engagement, which tends to grow around users whose stock choices perform well, and a

user with an unsuccessful investment would be less likely to be surfaced by our procedure.

In Supplementary Material S6 we further analyze how the community's engagement with the target stock evolved before and after each leader's activity, distinguishing empirically between two temporal patterns: cases in which the leader's activity precedes and accompanies a community-relative rise in engagement with the target stock (*attention initiation* for User1 and User2), and cases in which it coincides with the peak of engagement driven by external factors (*arrival at peak* for user User3). We emphasize throughout the paper that these characterizations are behavioral rather than causal: our data describe the ordering between individual and community activity, but do not establish that one drives the other. For studies of causality between WallStreetBets activity and the stock market, we remand the reader to the pertinent literature (Desiderio et al., 2025b; Anand and Pathak, 2022).

Out of all three opinion leaders, only one was able to completely shift the interest of the community towards his chosen stock. As Figure 3C shows, the number of ticker mentions for each "engaging" stock (the three found in the first section that produced a shift in the community's behavior), normalized by the total number of mentions, reaches a maximum in the week of the corresponding event. The ticker entropy (see Methods), which is a proxy for how many different stocks are discussed in the community on a given week, decreases in correspondence to the events of interest, but it reaches a minimum in January, when the entire community is engaged in the GME short squeeze. Overall, User1 was exceptional in his behavior, even when compared to the other two opinion leaders. In fact, since his first investment in 2019, he has been continuously posting monthly updates on his position, which then became weekly updates in 2020. Moreover at the time of first investment, GME's popularity was declining, so his 50,000 $ "YOLO" attracted much attention from the community (initially also much criticism). However, as the first plot of Figure 3D shows, all quantities increase as GME's price starts to rise - resulting in large gains on his side, allowing him to obtain a wide following from August 2020 onwards, to such a degree that moderators created a thread exclusively for GME. The other opinion leaders started investing in the stock as well. All of these events resulted in a wide consensus of the community on actively participating in the GME short squeeze (Mancini et al., 2022).

The fall of WallStreetBets

The events of the GME saga put WallStreetBets in the spotlight of the entire world, with a tremendous number of new users who joined from January onwards. Did the community change after such disruptive event and the massive influx of heterogeneous users from outside? Did new opinion leaders emerge? To find an answer we repeat the same set of analyses described above in the period after the GME short squeeze (February 2021 - July 2021).

Amongst all outliers found in this new timespan, only one truly corresponds to the previous description of an opinion leader on WSB: User1 who, even after the hype of the short squeeze died out, was still able to exert a great deal of influence in the community. All other popular users we identified lack the most important factor in becoming an influential individual on a community of traders: a successful YOLO investment (or any investment at all, to be precise). Moreover we found that the new subscribers of WSB interact and behave in a different way with respect to the *old* users of WSB (Haq, 2022). This can be seen by looking at the *labels* users attach on their posts to characterize the content. We considered the following labels: *profit*, posts about profits from an investment; *loss*, posts about losses from an investment; *dd* (*due diligence*), lengthy posts with a detailed analysis of a financial situation; *yolo*, posts about large and risky investments; *question*, posts with questions mainly concerning the world of finance; *meme*, posts with memes. Figure 4A shows, for each month, the fraction of posts with a given label. Interestingly we notice that starting from January the *meme* labelled posts increase, while the *dd* ones drop to just a small fraction. This change of wording and topics implies that the community's interest shifted from investing to amusing content: new users post and appreciate memes, while information on actual financial investments gets lost in the multitude of new content created.

Moreover, posts loose explicit mentions to tickers, as shown by Figure 4B, where we keep track of the ticker content of posts, with a highlight on some of the popular stocks on WSB such as GME, PLTR, CRSR, Rocket Companies (RKT), Nokia (NOK), AMC Theatres (AMC) and Blackberry (BB), showing how the fraction of posts involving no tickers (labelled as *other*) grows after January 2021, while posts mentioning tickers diminish (*other tickers*). The only exception to this trend are posts about meme stocks: GME, BB, AMC, NOK, which still make up a relevant fraction of the ticker content created after January. Indeed, in the wake of the enthusiasm of the GME event, users tried to initiate new short squeezes for these stocks, succeeding in their intent with AMC Theatres in May 2021 (McDowell, 2022). Moreover we can see how posts mentioning more than one meme stock (*co-occurences*) grow substantially after January, indicating again a less focused financial dialogue.

The top panel of Figure 4C shows the monthly distribution of the daily top 25 posts with highest score. The color highlights the ticker mentioned in the post, while the tone of the shaded area represents the number of posts mentioning no tickers (the darker the shade, the higher the number). Starting from January, most of the highest ranking posts are about GME, with a smaller fraction of posts mentioning other tickers and a progressively growing number of posts mentioning no tickers at all. Posts written by the three opinion leaders regularly appear amongst the top rated content, indicating that once a user reaches such position, it will likely be maintained in time. To further inspect how the behavior of users changed after the short squeeze, we display in the middle panel of Figure 4C the number of posts written in the entire time spanned by our data and the average number of comments written per user, which can be regarded as a proxy for the interest of users in the content posted in the community. We see a steady decline, denoting the reduction of engaging conversations, even if the number of posts created in the post-GME short squeeze period is higher. Finally, the bottom panel of Figure 4C presents the monthly averages of quantities describing how users interact: the number of direct replies (*first children*) of a comment diminishes, indicating once again the lack of lengthy discussions; and so does the number of jargon terms used in writing posts, which reflects a user's affinity with the community (Lucchini et al., 2022; Boylston et al., 2021). On the other hand the change in jargon concerning comments is not as

10.3389/fcpxs.2026.1924415

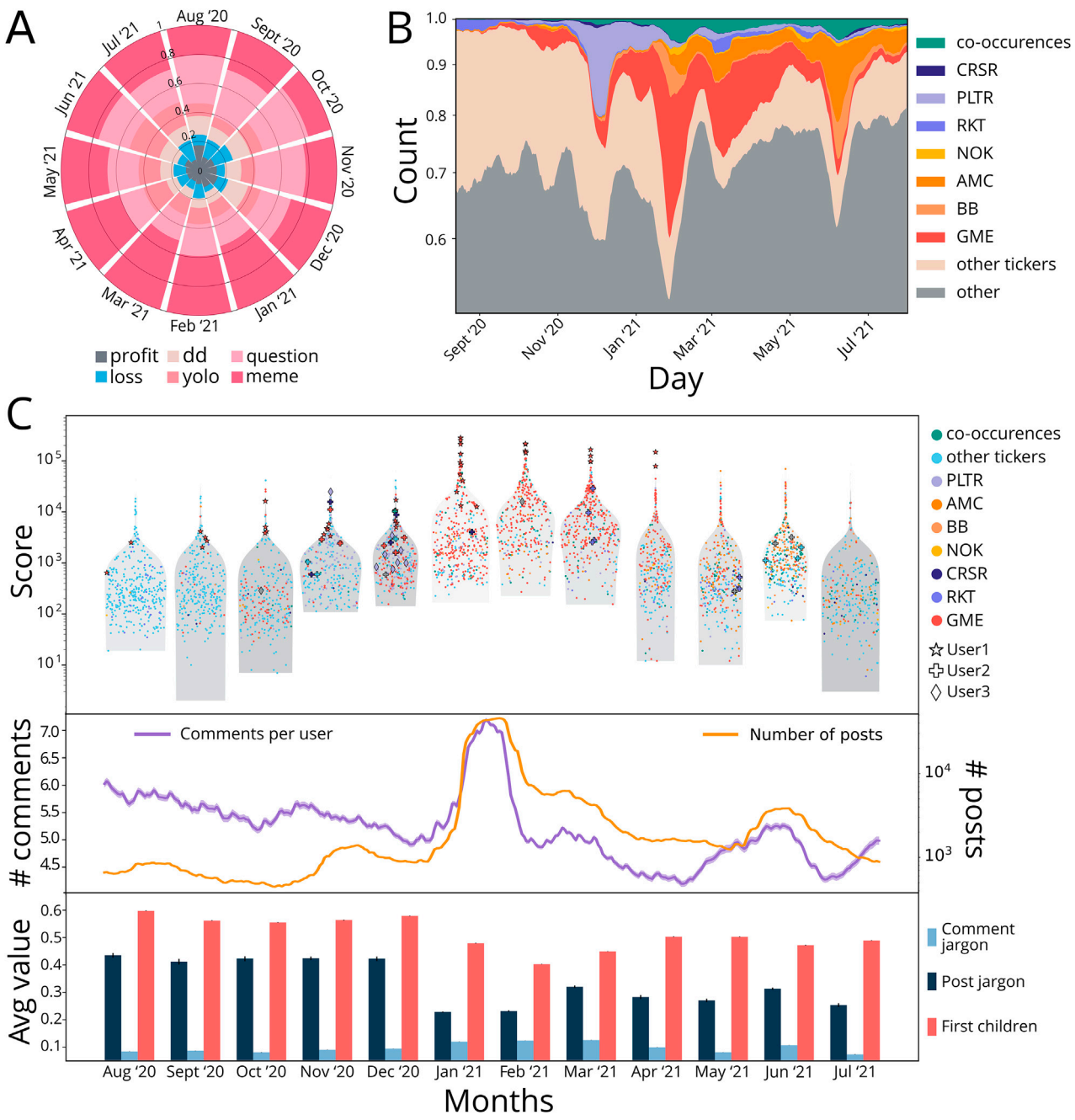


FIGURE 4
The aftermath of the short squeeze. **(A)** Fraction of posts with different labels in the various months before and after the GME event. We observe an increase in meme-labelled posts and a decrease of due diligence ones, highlighting the changing interest of the community. **(B)** Normalized daily count of posts mentioning popular tickers on WSB (GME, BB, AMC, NOK, RKT, PLTR, CRSR) and featuring multiple ticker mentions among them (*co-occurences*), mentioning other tickers (among a set of 157 hand-selected stocks, see Methods) (*other tickers*), and not mentioning any tickers at all (*others*). We see that the number of mentions for a ticker highlights the periods of maximum popularity for the corresponding stock; additionally there is a change in mention patterns after January, resulting in less focused discussions (increase in multiple ticker mentions in a single post) still linked to the effect of the short squeeze (mentions of meme stocks GME, BB, AMC, NOK). We can also see the growth of posts not mentioning any tickers at all. A 14-day moving average is applied to these figures. **(C)** Top: monthly distribution of the top 25 highest ranked posts (by score), with color indicating the ticker mentioned in the post and shaded area proportional to the number posts made in the selected month without any ticker mentions. Posts made by opinion leaders are highlighted and always lie in the upper tail of the distribution. Center: average number of comments per user made each day, with standard deviation of the mean, and number of posts (a 20-day moving average is applied). Bottom: monthly averages of the features: comment jargon, post jargon and first children, with standard deviation of the mean.

appreciable, given its already low value due to the shorter length of the text. For a discussion on how these changes affected the topological structure of the user network, see section S14 of the Supplementary Materials.

Conclusion

Understanding human behavior is a difficult challenge, but data retrieved from social media may come to our aid. The information gathered through the analysis of user roles and their evolution plays an important part in deepening our understanding of the subject. As this concept is vast and undefined, studying how individuals act on different social networks helps us in finding behavioral patterns that can be generalized to a broader environment. One of the aims of this work is to help fill in the gap towards this final goal, by examining the Reddit community of WallStreetBets and its time evolution, observing how it responds to engaging events and increasing popularity.

Reddit provides the perfect environment for these types of analyses, as on the one hand the presence of pseudonyms facilitates a looser behavior of users, on the other hand a better identification of opinion leaders is possible, since it overcomes topic limitation and prejudices that might be present when the identity of a user is known. Nonetheless, we find evidence of social roles like the "commenter" that exist also in contexts where there is no anonymity (van Mierlo, 2014; Welser et al., 2007), suggesting that some behaviors are implicit in online human interactions.

In this work, we unveiled the patterns of online behavior for average users through a semi-supervised procedure, and portrayed their time evolution. Furthermore, by looking at users whose behavior deviates considerably from average, we could recognize different important social roles. Three main candidate opinion leaders are identified, each linked to a specific stock investment. We analyze their behavior and find shared patterns that contribute to their popularity on WSB: persistent posting behavior, extreme investments (YOLOs) and successful gains. Although when considering the mass-coordination event that took place, only one of these users is truly relevant and was able to shift the focus of the community to his chosen stock: GME. We stress that these attributes emerge from a case-study analysis of three individuals identified as candidate opinion leaders in our dataset, and should be read as an illustrative characterization of the shared profile of these specific leaders rather than as a general behavioral rule for opinion-leader emergence.

After the GME short squeeze, WSB experienced a massive inflow of subscribers. In this period, we found no actual investment-related opinion leader. Moreover, by looking at the contents produced by the community, we can witness a drastic change whereby investment-oriented discussions were outnumbered by meme shares and shorter, less focused conversations.

Our framework of analysis may be applied to other online social networks as well as off-line social domains, such as a friendship network, by adapting the feature vectors with the users' traits that are most relevant to each situation. One of the limitations of this work is the difficulty in comparing the obtained results with the more canonical network centrality measures, which have been used to identify influential users and spreaders (Kitsak et al., 2010; Banerjee et al., 2013). This is because our networks are not obtained from direct social interactions or friendship ties, but are merely inferred from the users' posting and commenting activity. Further work will be needed to investigate methods to map Reddit user interactions onto friendship-like or following-like networks, thus allowing a direct comparison with other social networks by using standard tools. Another limitation of our approach is the need to hand-pick the features that are believed to be the most significant in each social context. In this direction, explainability techniques applied to machine learning approaches designed for specific tasks can help select features in a less supervised way.

Despite such limitations, this work sheds light on social dynamics in an online environment with anonymity such as Reddit. We identified user roles and investigated for the first time the factors that contribute to becoming opinion leaders on WallStreetBets, describing the features needed for users to obtain such a status. Our results pave the way for similar analyses to be carried out in other settings, may it be for the classification of users in social networks, identification of those who create and spread new information, or pinpointing the outliers, amongst which both influential and malevolent individuals could be found.

Future work and AI perspectives

Our pipeline deliberately relies on interpretable, lightweight tools–VADER for sentiment, word-frequency entropy for textual complexity, and PCA followed by DBSCAN for role extraction–chosen because each feature maps transparently onto a behavioral interpretation. A promising direction for future work is to replace or augment these descriptors with representations learned by modern AI techniques, while preserving the semi-supervised, emergent spirit of our framework. In particular, transformer-based language models (e.g., sentence-embedding or fine-tuned domain-specific encoders) could yield richer post and comment representations than lexicon-based sentiment and bag-of-words entropy, capturing irony, sarcasm and the rapidly evolving slang of WallStreetBets more faithfully than a fixed jargon list. Likewise, the weekly "reply-to" networks naturally lend themselves to graph representation learning: graph neural networks or graph autoencoders could embed each user from both their textual content and their relational position simultaneously, potentially sharpening the separation between opinion leaders, analyzers and active users that we currently recover from the feature space. Such learned embeddings could feed the same iterative clustering-and-removal procedure used here for outlier detection, so that roles still emerge from the data rather than from a predefined taxonomy. Finally, the temporal dimension–central to our characterization of how a single user triggered community-wide consensus–invites the use of temporal graph networks or sequence models to forecast, rather than merely describe, the emergence of an opinion leader, which would move the analysis from a retrospective account of the GameStop saga toward an early-warning system for incipient collective retail action. A natural caveat is that the gain in expressive power from these models comes at the cost of interpretability, so any such extension would need to be paired with explainability methods to retain the behavioral readability that motivated our feature-based design in the first place.

TABLE 1 User features.

Comment feature	Description
number of comments	Number of comments made by the user
score of comments	Average score of comments made by the user
entropy of comments	Average value of the text complexity of the comments made by the user
jargon of comments	Average number of jargon terms used in comments
sentiment of comments	Average value of the sentiment of comments made by the user
first children	Average number of comments that are direct replies to a comment by the user
Post feature	**Description**
number of posts	Number of posts made by the user
score of posts	Average score of posts made by the user
entropy of posts	Average value of the text complexity of the posts made by the user
jargon of posts	Average number of jargon terms used in posts
sentiment of posts	Average value of the sentiment of posts made by the user
post comments	Average number of comments under a given post made by the user
Network feature	**Description**
k_{out}	Out-degree of the user in the weekly network
k_{in}	In-degree of the user in the weekly network
s_{out}	Out-strength of the user in the weekly network
s_{in}	In-strength of the user in the weekly network

Methods

Dataset

The dataset was collected from Pushshift (Baumgartner et al., 2020), an API that regularly copied and saved data from Reddit and other social networks. We used Pushshift's API to query all comments and posts from the subreddit WallStreetBets from 1 August 2020 to 31 July 2021. The dataset is then divided in two periods: before the GME short squeeze (August 2020 - January 2021) and after the short squeeze (February 2021 - July 2021). We treat each dataset separately but the same procedures are applied to both.

First, all comments and posts from known Reddit bots were removed. Then, to select only users who actually contributed to the community during the considered timespan, we applied two separate filters. We consider the set of users who wrote more than two posts in the overall time period, the set of users who wrote more than 20 comments, and take their union to obtain a total of 103,597 unique users for the first period, and 112,661 for the second period (see Supplementary Material S4 for a detailed description and sensitivity analysis of these filters). The final dataset comprises only posts and comments by these users, so it contains 12,272 782 comments and 603,106 posts for the first period, and 9 809 208 comments and 224,469 posts for the second period. Data regarding screenshots of the actual investment of a user were retrieved from the posts' metadata or by manual inspection of the specific post on WSB. The data relative to stock prices was retrieved from Yahoo Finance (finance.yahoo.com).

User networks and features

The data is divided into weekly subsets by applying a 7-day moving window and shifting it by 1 day each time, so as to get a 6-day overlap between subsequent subsets. For simplicity, each subset will be referred to as a *week*. For each user the following 16 features, shown in Table 1, were computed over the given week, so in each week a single user is represented by a 16-dimensional array containing the (standardized) feature values.

The score of a post or comment is defined as the number of "upvotes" minus the number of "downvotes" it receives from users. The entropy of the posts/comments by user α is computed as

$$E_\alpha = -\sum_{i=1}^{N} p_i^\alpha \cdot ln(p_i^\alpha) \Big/ ln(N^\alpha) \quad (1)$$

where p_i^α is the number of times user α uses word i divided by the total number of words used by α in the given week, N^α. We use as normalization the maximum possible value of entropy, $log(N^\alpha)$, when the user uses every word with equal probability $1/N^\alpha$, to make Equation 1 independent from the number of comments and posts written by the user. Entropy is set to 0 when the user employs only 1 word or when the same word is used multiple times.

The root words of jargon terms selected to compute the corresponding feature are: *retard, autist, ape, degenerate, moon, bull, bear, bagholder, btfd, diamond, paper, dd, stonk, tendies, yolo, hodl* (Smythe, 2021). These words are commonly used in the community's language, far before the GME short squeeze. Thus we can say that these jargon terms have not evolved in time, as we have already shown in our previous work (Mancini et al., 2022).

Sentiment of comments and posts is evaluated using VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto and Gilbert, 2014), an algorithm that assigns a piece of text with a compound score between −1 (very negative) and +1 (very positive). VADER is sensitive to both the polarity and intensity of the text, taking into account punctuation and word shape (ALL CAPS) used to add emphasis, degree modifiers that alter intensity (boosters such as *very* and dampeners such as *kind of*), slang and acronyms. VADER is based on a lexicon of words and emojis, each with an associated score ranging from −4 to +4 according to its meaning (from negative to positive). As done in our previous study, we adapt VADER to the typical jargon and sarcasm of WSB users by adding to its lexicon particular words (see Mancini et al. (2022) for the detailed procedure). We note that the specific sentiment scores assigned to the community-specific terms added to VADER's lexicon have not been validated against a hand-labeled WSB sample; a targeted validation of this kind would strengthen future applications of lexicon-based sentiment analysis to WallStreetBets.

The weekly networks from which the network features are computed are constructed by considering the reply-to interactions, a method commonly used when dealing with online social networks (Sousa et al., 2010; Gómez et al., 2008). Given two users A and B, if A has replied to a comment or post written by B, then a directed link from A to B is added. The weight of the link represents how many times A has replied to something written by B. So the resulting networks are directed and weighted.

DBSCAN and spectral clustering

Since some of our 16 features may be correlated, we apply PCA to the matrix of users' feature vectors. This allows us to focus only on the top 10 PCA components - which on average keep around 95% of the variance of the original features (see Supplementary Material S8 for further details).

Hence both DBSCAN and spectral clustering are applied to these 10-component PCA projection of the user's feature vectors, meaning that a user is represented as a point in this 10-dimensional space. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (Ester et al., 1996) is a clustering algorithm that uses density of data points in a given region to detect clusters. It groups into same clusters points that are packed together, labeling as noise those that lie in low-density regions. It works by defining two parameters, *MinPts* and ϵ, that are used to classify points in three categories: *core points*, if at least *MinPts* points are within a distance ϵ; *reachable points*, if they lie within distance ϵ from a core point; *noise points*, all those that are not reachable from any point. Core points and reachable points are the ones that form clusters. We use $MinPts = 15$ and $\epsilon = 1$ (see Supplementary Material S9 for a robustness analysis).

We also used spectral clustering (Ng et al., 2001) to further validate the clustering obtained with DBSCAN. First, the pairwise distances between all users were computed and stored in a matrix D, then the affinity matrix was computed as $S = \exp(-D^2)$. All matrix entries whose value is below 0.5 are set to 0, and a weighted graph is built using the remaining matrix elements. The connected components of the graph are the user clusters, while all isolated nodes along with all the connected components with less than five elements are set to noise.

Both methods were implemented using the Python package Scikit-learn (Pedregosa, 2011) and give very similar results (see Supplementary Material S12). Note that to make the analysis robust against the growing size of the community, we performed the clustering on 100 different samples of 10,000 users, and consider the mean and standard deviation of all quantities obtained.

Outliers with iterative DBSCAN clustering

In addition to being a popular clustering algorithm, DBSCAN has also been used for anomaly detection (Celik et al., 2011; Alhussein et al., 2020; Dokuz et al., 2020). In fact, by focusing on points that are classified as noise, we can zoom-in on those individuals whose behavior deviates from that of the average user. To select only the most persistent outliers we follow a two-step procedure. 1) For each week we iteratively apply DBSCAN 12 times, each time we exclude all points who were assigned to clusters and only consider those labelled as noise, to which we apply again the clustering algorithm. At each iteration the parameter *MinPts* is fixed to 10, while ϵ varies from one to 12, adapting to the decreasing density of the dataset. After this first step we are left with only the furthest outliers for each week. 2) From these outliers we select those who appear for more than 15 weeks (out of the 178 total) for the first half of the dataset, and for more than 20 weeks (out of the 176 total) for users in the second half of the dataset (since we have more users we need a stronger filter). In this way we select those users who show an outlier behavior persistently in time. For a robustness check of the results with different parameter configurations see Supplementary Material S7.

Ticker entropy

Since WallStreetBets is a community of traders, where most of the conversations revolve around stocks, we define the *ticker entropy* as a measure of the heterogeneity of stocks discussed in the community. We consider the text of posts, as they contain much of the information about the preferences of the community, and select all the uppercase regular expressions of two–5 characters. From all the expressions found we manually validated those that actually corresponded to ticker symbols, and excluded all those with ambiguous meaning (FROG, YOU, etc., . . .). In the end we are left with $M = 157$ stock tickers that have been discussed in the time span considered. For each week we compute the probability p_k of ticker k by dividing the total number of mentions of k by the total number of ticker mentions in the given week. We then compute the Shannon entropy of the discrete ticker distribution of week *tau* as shown in Equation 2

$$H_\tau = -\sum_{k=1}^{M} p_k^\tau \cdot ln(p_k^\tau) \tag{2}$$

Data availability statement

All data were retrieved from Pushshift, a third-party archive of publicly available Reddit content, covering August 1, 2020 to July 31, 2021. Collection predated the 2023 changes to Reddit's API and Data Terms and the subsequent restrictions on Pushshift; at the time, Pushshift was an established resource for academic Reddit research and its use was consistent with common practice in the field. We note that Reddit has since introduced the Reddit for Researchers program as the authorized channel for academic data access, which did not exist during our collection period. The study relied exclusively on secondary analysis of anonymous, publicly available content, with no interaction with human participants and no collection or use of personally identifying information; users are represented solely as anonymous nodes and all reported results are aggregated at the level of user roles rather than individuals. The code used for the analyses and the figures of the article is available on GitHub

https://github.com/ComplexConnectionsLab/rise_and_fall_of_wallstreetbets.

Author contributions

AM: Data curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing. AD: Data curation, Formal Analysis, Methodology, Writing – original draft, Writing – review and editing. GP: Data curation, Formal Analysis, Methodology, Writing – original draft, Writing – review and editing. RD: Conceptualization, Project administration, Supervision, Validation, Writing – original draft, Writing – review and editing. GC: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing – original draft, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. GC acknowledges support from ``Deep `N Rec' Progetto di Ricerca di Ateneo of University of Rome Tor Vergata. GC acknowledges support from ``Deep `N Rec' Progetto di Ricerca di Ateneo of University of Rome Tor Vergata.

Acknowledgments

RD.C. acknowledges Sony Computer Science Laboratories Paris, for hosting him during part of the research. GC acknowledges support from "Deep 'N Rec" Progetto di Ricerca di Ateneo of University of Rome Tor Vergata.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcpxs.2026.1924415/full#supplementary-material

References

Aichner, T., Grünfelder, M., Maurer, O., and Jegeni, D. (2021). Twenty-five years of social media: a review of social media applications and definitions from 1994 to 2019. *Cyberpsychology, Behav. Soc. Netw.* 24, 215–222. doi: 10.1089/cyber.2020.0134

Alhussein, I., and Ali, A. H. (2020). "Application of dbscan to anomaly detection in airport terminals," in *2020 3rd International Conference on Engineering Technology and its Applications (IICETA)*, 112–116. doi: 10.1109/IICETA50496.2020.9318876

Amato, F., Moscato, V., Picariello, A., Piccialli, F., and Sperlì, G. (2017). Centrality in heterogeneous social networks for lurkers detection: an approach based on hypergraphs. *Concurrency Comput. Pract. Exp.* 30, e4188. doi: 10.1002/cpe.4188

Anand, A., and Pathak, J. (2022). The role of Reddit in the gamestop short squeeze. *Econ. Lett.* 211, 110249. doi: 10.1016/j.econlet.2021.110249

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., et al. (2018). "Exposure to opposing views on social media can increase political polarization,", Washington, DC: in. Proceedings of the National Academy of Sciences, 9216–9221. doi: 10.1073/pnas.1804840115

Balsamo, D., Bajardi, P., Salomone, A., and Schifanella, R. (2021). Patterns of routes of administration and drug tampering for nonmedical opioid consumption: data mining and content analysis of Reddit discussions. *J. Med. Internet Res.* 23, e21212. doi: 10.2196/21212

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science* 341, 1236498. doi: 10.1126/science.1236498

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., and Blackburn, J. (2020). *The Pushshift Reddit Dataset*.

Biddle, B. J. (1986). Recent developments in role theory. *Annu. Rev. Sociol.* 12, 67–92. doi: 10.1146/annurev.so.12.080186.000435

Bodendorf, F., and Kaiser, C. (2009). "Detecting opinion leaders and trends in online social networks," in Proceedings of the 2nd ACM Workshop on Social Web Search and Mining, *SWSM '09, 65–68* (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/1651437.1651448

Boylston, C., Palacios, B., Tassev, P., and Bruckman, A. W. (2021). *Positions or Ban.*

Brandtzaeg, P., and Heim, J. (2011). A typology of social networking sites users. *Int. J. Web Based Communities* 7, 28–51. doi: 10.1504/IJWBC.2011.038124

Brown, D. K., Ng, Y. M. M., Riedl, M. J., and Lacasa-Mas, I. (2018). "Reddit's veil of anonymity: predictors of engagement and participation in media environments with hostile reputations,", Washington, DC: Social Media + National Academy of Sciences. doi: 10.1177/2056305118810216

Buntain, C., and Golbeck, J. (2014). "Identifying social roles in Reddit using network structure," in Proceedings of the 23rd International Conference on World Wide Web, *WWW '14 Companion* (New York, NY, USA: Association for Computing Machinery), 615–620. doi: 10.1145/2567948.2579231

Buz, T., and de Melo, G. W. (2024). Assessing the collective intelligence of Reddit for investment advice. *Trans. Soc. Comput.* 7, 1–23. doi: 10.1145/3660760

Buz, T., Schneider, M., Kaffee, L.-A., and De Melo, G. (2024). "Highly regarded investors? Mining predictive value from the collective intelligence of reddit's wallstreetbets," in Proceedings of the 16th ACM Web Science Conference, *WEBSCI '24* (New York, NY, USA: Association for Computing Machinery), 320–330. doi: 10.1145/3614419.3643993

Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. *Rev. Mod. Phys.* 81, 591–646. doi: 10.1103/RevModPhys.81.591

Celik, M., Dadaser-Celik, F., and Dokuz, A. S. (2011). "Anomaly detection in temperature data using dbscan algorithm," in *2011 International Symposium on*

Innovations in Intelligent Systems and Applications, 91–95. doi: 10.1109/INISTA.2011.5946052

Chohan, U. W. (2022). Yolo capitalism. *SSRN* 432. doi: 10.2139/ssrn.3775127

Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A., and Caldarelli, G. (2019). The statistical physics of real-world networks. *Nat. Rev. Phys.* 1, 58–71. doi: 10.1038/s42254-018-0002-6

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., and Starnini, M. (2021). "The echo chamber effect on social media,", 118. Thousand Oaks, CA: Proceedings of the National Academy of Sciences SAGE Publications. doi: 10.1073/pnas.2023301118

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., et al. (2012). Manifesto of computational social science. *Eur. Phys. J. Special Top.* 214, 325–346. doi: 10.1140/epjst/e2012-01697-8

Desiderio, A., Mancini, A., Cimini, G., and Di Clemente, R. (2025a). Highly engaging events reveal semantic and temporal compression in online community discourse. *PNAS Nexus* 4, pgaf056. doi: 10.1093/pnasnexus/pgaf056

Desiderio, A., Aiello, L., Cimini, G., and Alessandretti, L. (2025b). The dynamics of the Reddit collective action leading to the gamestop short squeeze. *Npj Complex.* 2, 5. doi: 10.1038/s44260-025-00029-z

Di Clemente, R., Luengo-Oroz, M., Travizano, M., Xu, S., Vaitla, B., and González, M. C. (2018). Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nat. Commun.* 9, 3330. doi: 10.1038/s41467-018-05690-8

Dokuz, A. S., Celik, M., and Ecemis, A. (2020). Anomaly detection in bitcoin prices using dbscan algorithm. *Eur. J. Sci. Technol.* 436–443. doi: 10.31590/ejosat.araconf57

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc. Second Int. Conf. Knowl. Discov. Data Min.* 96, 226–231.

Ferraro, A., and Sperlì, G. (2024). How does user-generated content on social media affect stock predictions? A case study on gamestop. *Online Soc. Netw. Media* 43-44, 100293. doi: 10.1016/j.osnem.2024.100293

Flynn, L., Goldsmith, R., and Eastman, J. (1994). The king and summers opinion leadership scale: revision and refinement. *J. Bus. Res.* 31, 55–64. doi: 10.1016/0148-2963(94)90046-9

Gagnon, T. (2013). The disinhibition of Reddit users. Available online at: https://cah.ucf.edu/wp-content/uploads/sites/27/2019/10/KWS2_Gagnon.pdf (Accessed October 23, 2026).

Gloor, P. A., Fronzetti Colladon, A., Grippa, F., and Giacomelli, G. (2017). Forecasting managerial turnover through e-mail based social network analysis. *Comput. Hum. Behav.* 71, 343–352. doi: 10.1016/j.chb.2017.02.017

Gómez, V., Kaltenbrunner, A., and López, V. (2008). "Statistical analysis of the social network and discussion threads in slashdot," in *Proceedings of the 17th International Conference on World Wide Web* (New York, NY, USA: Association for Computing Machinery), 645–654. doi: 10.1145/1367497.1367585

Grewal, R., Mehta, R., and Kardes, F. (2000). The role of the social-identity function of attitudes in consumer innovativeness and opinion leadership. *J. Econ. Psychol.* 21, 233–252. doi: 10.1016/S0167-4870(00)00003-9

Haase, F., Rath, O., Kurka, M., and Schoder, D. (2023). *Finfluencers: Opinion Makers or Opinion Followers?* ECIS 2023 Research Papers *432*. Available online at: https://aisel.aisnet.org/ecis2023_rp/432 (Accessed October 23, 2026).

Haq, E.-U. (2022). "Short, colorful, and irreverent! a comparative analysis of new users on wallstreetbets during the gamestop short-squeeze," in *Companion Proceedings of the Web Conference 2022* (New York, NY, USA: Association for Computing Machinery), 52–61. doi: 10.1145/3487553.3524202

Hofmann, V., Schütze, H., and Pierrehumbert, J. B. (2022). The Reddit politosphere: a large-scale text and network resource of online political discourse. *Proc. Int. AAAI Conf. Web Soc. Media* 16, 1259–1267. doi: 10.1609/icwsm.v16i1.19377

Hutto, C., and Gilbert, E. (2014). Vader: a parsimonious rule-based model for sentiment analysis of social media text. *Proc. Int. AAAI Conf. Web Soc. Media* 8, 216–225. doi: 10.1609/icwsm.v8i1.14550

Jin, L., Chen, Y., Wang, T., Hui, P., and Vasilakos, A. (2013). Understanding user behavior in online social networks: a survey. *Ieee Communications Magazine* 51 (9), 144–150. doi: 10.1109/MCOM.2013.6588663

Katz, E. (1957). The two-step flow of communication: an up-to-date report on an hypothesis. *Public Opin. Q.* 21, 61–78. doi: 10.1086/266687

Khondker, H. H. (2011). Role of the new media in the arab spring. *Globalizations* 8, 675–679. doi: 10.1080/14747731.2011.621287

Kim, K., Lee, S.-Y. T., and Kauffman, R. J. (2023). Social informedness and investor sentiment in the gamestop short squeeze. *Electron. Mark.* 33, 23. doi: 10.1007/s12525-023-00632-9

Kiss, C., and Bichler, M. (2008). Identification of influencers - measuring influence in customer networks. *Decis. Support Syst.* 46, 233–253. doi: 10.1016/j.dss.2008.06.007

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., et al. (2010). Identification of influential spreaders in complex networks. *Nat. Phys.* 6, 888–893. doi: 10.1038/nphys1746

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., et al. (2009). Computational social science. *Science* 323, 721–723. doi: 10.1126/science.1167742

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., et al. (2018). The science of fake news. *Science* 359, 1094–1096. doi: 10.1126/science.aao2998

Lucchini, L., Aiello, L. M., Alessandretti, L., De Francisci Morales, G., Starnini, M., and Baronchelli, A. (2022). From Reddit to wall street: the role of committed minorities in financial collective action. *R. Soc. Open Sci.* 9, 211488. doi: 10.1098/rsos.211488

Mancini, A., Desiderio, A., Di Clemente, R., and Cimini, G. (2022). Self-induced consensus of Reddit users to characterise the gamestop short squeeze. *Sci. Rep.* 12, 13780. doi: 10.1038/s41598-022-17925-2

Matsumoto, M., Hashimoto, R., Suzuki, M., Murayama, Y., and Izumi, K. (2025). Impact of information disparity between individual investors on profits of meme stocks using an artificial market simulation approach. *J. Comput. Soc. Sci.* 8, 25. doi: 10.1007/s42001-024-00355-7

McDowell, J. (2022). The AMC Short Squeeze Explanation. Available online at: https://www.tradingsim.com/blog/the-amc-short-squeeze-explanation-tradingsim (Accessed October 23, 2026).

Miao, Q., Meng, Y., and Sun, J. (2016). "Identifying the most influential topic-sensitive opinion leaders in online review communities," in *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 330–335. doi: 10.1109/ICCCBDA.2016.7529579

Ng, A. Y., Jordan, M. I., and Weiss, Y. (2001). "On spectral clustering: analysis and an algorithm," in *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic* (Cambridge, MA, USA: MIT Press), 849–856. doi: 10.5555/2980539.2980649

Olejnik, L., Castelluccia, C., and Janc, A. (2014). On the uniqueness of web browsing history patterns. *Ann. telecommunications-annales des télécommunications* 69, 63–74. doi: 10.1007/s12243-013-0392-5

Pedregosa, F. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195

Recuero, R., Zago, G., and Soares, F. (2019). Using social network analysis and social capital to identify user roles on polarized political conversations on Twitter. *Soc. Media + Soc.* 5, 2056305119848745. doi: 10.1177/2056305119848745

Richmond, V. P. (2006). Monomorphic and polymorphic opinion leadership within a relatively closed communication System. *Hum. Commun. Res.* 6, 111–116. doi: 10.1111/j.1468-2958.1980.tb00131.x

Riquelme, F., González-Cantergiani, P., Hans, D., Villarroel, R., and Munoz, R. (2019). Identifying opinion leaders on social networks through milestones definition. *IEEE Access* 7, 75670–75677. doi: 10.1109/ACCESS.2019.2922155

Scherman, A., Arriagada, A., and Valenzuela, S. (2015). Student and environmental protests in Chile: the role of social media. *Politics* 35, 151–171. doi: 10.1111/1467-9256.12072

Semenova, V., and Winkler, J. (2025). Social contagion and asset prices: reddit's self-organized bull runs. *Quantitative Finance*, 25 (12), 1873–1904. doi: 10.1080/14697688.2025.2559970

Smith, M., Kilders, V., Kuethe, T., and Widmar, N. (2025). A time series analysis of herd investor behavior using online and social media data. *SAGE Open* 15, 21582440251375185. doi: 10.1177/21582440251375185

Smythe, C. (2021). The r/wallstreetbets glossary: a field guide to apes, stonks, tendies, and more. Available online at: https://www.businessofbusiness.com/articles/the-wallstreetbets-glossary-a-field-guide-to-apes-stonks-tendies-AMC-GME-reddit/ (Accessed October 23, 2026).

Sousa, D., Sarmento, L., and Mendes Rodrigues, E. (2010). "Characterization of the Twitter @replies network: are user ties social or topical?," in Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents, *SMUC '10, 63–70* (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/1871985.1871996

Štefan, L., Baumöhl, E., and Výrost, T. (2022). Yolo trading: riding with the herd during the gamestop episode. *Finance Res. Lett.* 46, 102359. doi: 10.1016/j.frl.2021.102359

Sun, Q., Wang, N., Zhou, Y., and Luo, Z. (2016). Identification of influential online social network users based on multi-features. *Int. J. Pattern Recognit. Artif. Intell.* 30, 1659015. doi: 10.1142/S0218001416590151

van der Maaten, L., and Hinton, G. (2008). Viualizing data using t-sne. *J. Mach. Learn. Res.* 9, 2579–2605. Available online at: https://jmlr.org/papers/v9/vandermaaten08a.html (Accessed October 23, 2026).

van Mierlo, T. (2014). The 1% rule in four digital health social networks: an observational study. *J. Med. Internet Res.* 16, e33. doi: 10.2196/jmir.2966

Verdear, D., Johnson, N. F., and Wuchty, S. (2025). Influence of Twitter social network graph topologies on traditional and meme stocks during the 2021 gamestop short squeeze. *Complexity* 2, 18. doi: 10.1038/s44260-025-00042-2

Vieira, M. R., Frias-Martinez, V., Oliver, N., and Frías-Martínez, E. (2010). "Characterizing dense urban areas from mobile phone-call data: discovery and social dynamics," in *2010 IEEE Second International Conference on Social Computing*, 241–248. doi: 10.1109/SocialCom.2010.41

Warkulat, S., and Pelster, M. (2024). Social media attention and retail investor behavior: evidence from r/wallstreetbets. *Int. Rev. Financial Analysis* 96, 103721. doi: 10.1016/j.irfa.2024.103721

Welser, H. T. (2011). "Finding social roles in wikipedia," in Proceedings of the 2011 Iconference, *Iconference '11* (New York, NY, USA: Association for Computing Machinery), 122–129. doi: 10.1145/1940761.1940778

Welser, H., Gleave, E., Fisher, D., and Smith, M. (2007). Visualizing the signatures of social roles in online discussion groups. *J. Soc. Struct.* 8, 1–31. Available online at: https://www.cmu.edu/joss/content/articles/volume8/Welser/ (Accessed October 23, 2026).

Whiteman, L. (2020). Why shares of palantir technologies soared higher in November (2020). Available online at: https://www.fool.com/investing/2020/12/01/why-shares-of-palantir-technologies-soared-higher/ (Accessed October 23, 2026).

Winter, S., and Neubaum, G. (2016). Examining characteristics of opinion leaders in social media: a motivational approach. *Soc. Media + Soc.* 2, 2056305116665858. doi: 10.1177/2056305116665858

Wooley, S., Edmonds, A., Bagavathi, A., and Krishnan, S. (2019). "Extracting cryptocurrency price movements from the Reddit network sentiment," in *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 500–505. doi: 10.1109/ICMLA.2019.00093

Yang, L., Tian, Y., Li, J., Ma, J., and Zhang, J. (2017). Identifying opinion leaders in social networks with topic limitation. *Clust. Comput.* 20, 2403–2413. doi: 10.1007/s10586-017-0732-8

Zheng, X., Tian, H., Wan, Z., Wang, X., Zeng, D. D., and Wang, F. Y. (2022). Game starts at gamestop: characterizing the collective behaviors and social dynamics in the short squeeze episode. *IEEE Trans. Comput. Soc. Syst.* 9, 45–58. doi: 10.1109/TCSS.2021.3122260