

Self-induced consensus of Reddit users to characterise the GameStop short squeeze

Keywords: network science, sentiment analysis, Reddit, voter model, collective coordination

Extended Abstract

The short squeeze of GameStop (GME) shares in mid-January 2021, primarily orchestrated by retail investors on the Reddit r/wallstreetbets (WSB) community, caused major losses for short seller hedge funds and a drastic surge of the stock price. Such an unprecedented event in finance represents a paramount example of a collective coordination action on online social media, resulting in consensus formation on a large scale.

In this work we characterize the structure and time evolution of Reddit conversation data (Figure 1ABC), from September 01, 2019 to February 01, 2021, showing that the occurrence of GME-related comments and their sentiment grew much before the short squeeze actually took place. These early signs of the collective action can be associated to a self-reinforcing, increasing level of commitment and engagement of individual users.

Since WSB is a community of traders, the occurrence of stock tickers in the text of posts and comments represents a first indicator of what is a popular conversation topic. Figure 1E shows the Z-scores for GME occurrences, compared to the average Z-score of all tickers. We clearly see how the peaks given by significant Z-scores correspond to all major events of the GameStop saga. Further these peaks become more frequent in time, signaling that the community's interest towards GameStop has grown substantially until January, when GME monopolizes the conversation on WSB. The same plot reports two other quantities related to the market trend of GME: the Z-score of the trading volume and the Close share price. It is noteworthy that peaks related to GME occurrences in WSB conversations and trading volume of the stock mostly coincide, pointing to a strong relation between the two variables.

One variable that we focused on is *sentiment*, in fact it has been often pointed out as a predictor of market movements. We thus perform text sentiment analysis using VADER (*Valence Aware Dictionary and sEntiment Reasoner*) [2]. We adapt the VADER lexicon to the peculiar jargon and sarcasm used by WSB members. Figure 1D shows an intensive sentiment indicator, *i.e.*, the average sentiment of all daily posts/comments that mention GME. We see that the signal is initially quite noisy due to the low number of GME-related comments until mid-October, then as early as the beginning of December it starts to grow significantly (both with respect to its previous trend and to the mean sentiment of all comments), far before the short squeeze of January. We can interpret this empirical evidence as a growing engagement and commitment towards the GME operation, which happens also in the absence of profits [3] and represents an early sign of consensus formation in the community.

In light of the results obtained, we work out a model in which engagement can influence collective behavior and foster the emergence of consensus or cooperation. We build on one of the most popular theoretical frameworks of opinion dynamics, the *voter model*, following an approach formally similar to [1]. Figure 2A shows a graphical representation of the update rule, at each time step a user is selected at random: when user i is selected for the update, with probability $1 - \lambda$ she copies the state of a random neighbor j (*i.e.*, each neighbor is selected

with probability $\frac{1}{k_i}$, where k_i is the *degree* or number of neighbors of i). Instead, with probability λ she follows a global field given by a random variable $e(t) = \pm 1$, where the probability $P_1[e(t) = +1] = P_1[e(t)] = \frac{e^{m(t)}}{1+e^{m(t)}}$. When $m(t) = 0$ we have $P_1[e(t)] = \frac{1}{2}$: the global field acts randomly on each user and is equivalent to a white noise term. Instead $m(t) \rightarrow +1$ leads to $P_1[e(t)] \rightarrow 1$ and analogously $m(t) \rightarrow -1$ to $P_1[e(t)] \rightarrow 0$: the global field is increasingly able to align users with the average opinion. $c \geq 1$ is a control parameter that can be associated with the level of engagement: the higher the value of c , the less consensus is required for users to align with $m(t)$.

We can understand the behavior of the model for different values of the c through its analytical mean-field solution. Figure 2B shows different equilibrium values $|m^*|$ of the dynamics as a function of c , and we can see how the system exhibits a classic second order phase transition for $c = e^2$. Above this threshold value the dynamics reaches a stable equilibrium point $|m^*| \neq 0$ that becomes closer to full consensus as c grows.

We proceed on studying how the model behaves on user-user interaction networks extracted from WSB conversation data. We focus on the four months from October 2020 to January 2021 that are the most relevant for the GameStop saga and build a user network for each month. Model simulations on these networks show that the degree heterogeneity of real user interactions leads to the emergence of a non-negligible level of consensus also for very small values of engagement c . January shows a different behavior with respect to the previous months due to the presence of more numerous and more connected super-hubs that ease the formation of an initial consensus. However magnetization alone does not allow for an appropriate comparison between networks of different sizes, since it represents the average opinion and is therefore an intensive variable. Equally important is the extension of the consensus in terms of number of users. Indeed the success of the short squeeze required a large number of investors who bought and held GME shares. Another factor to take into account is the notable growth in the number of WSB users in correspondence with the short squeeze. All together these observations suggest to consider an extensive order parameter, namely the number of individuals in the consensus state within a population that grows with the level of consensus reached: $M(t) = m(t) N_0 e^{qm(t)}$. This extensive magnetization has a much sharper transition than the intensive one, as shown in Figure 2C where again c represents the growth of activity and commitment on WSB, properly describing a sudden and large-scale formation of consensus that is qualitatively similar to the abrupt growth of GME price which ultimately represents the best proxy for the success of the short squeeze. The results presented here can hopefully shed light on the increasingly important phenomenon of self-organized collective actions on social networks.

References

- [1] G. De Marzo *et al.*, Emergence of polarization in a voter model with personalized information, *Phys. Rev. Res.*, 2:043117, Oct 2020.
- [2] C. Hutto and E. Gilbert, VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text, *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):216–225, May 2014.
- [3] L. Lucchini *et al.*, From reddit to wall street: the role of committed minorities in financial collective action. *Royal Society Open Science*, 9(4):211488, 2022

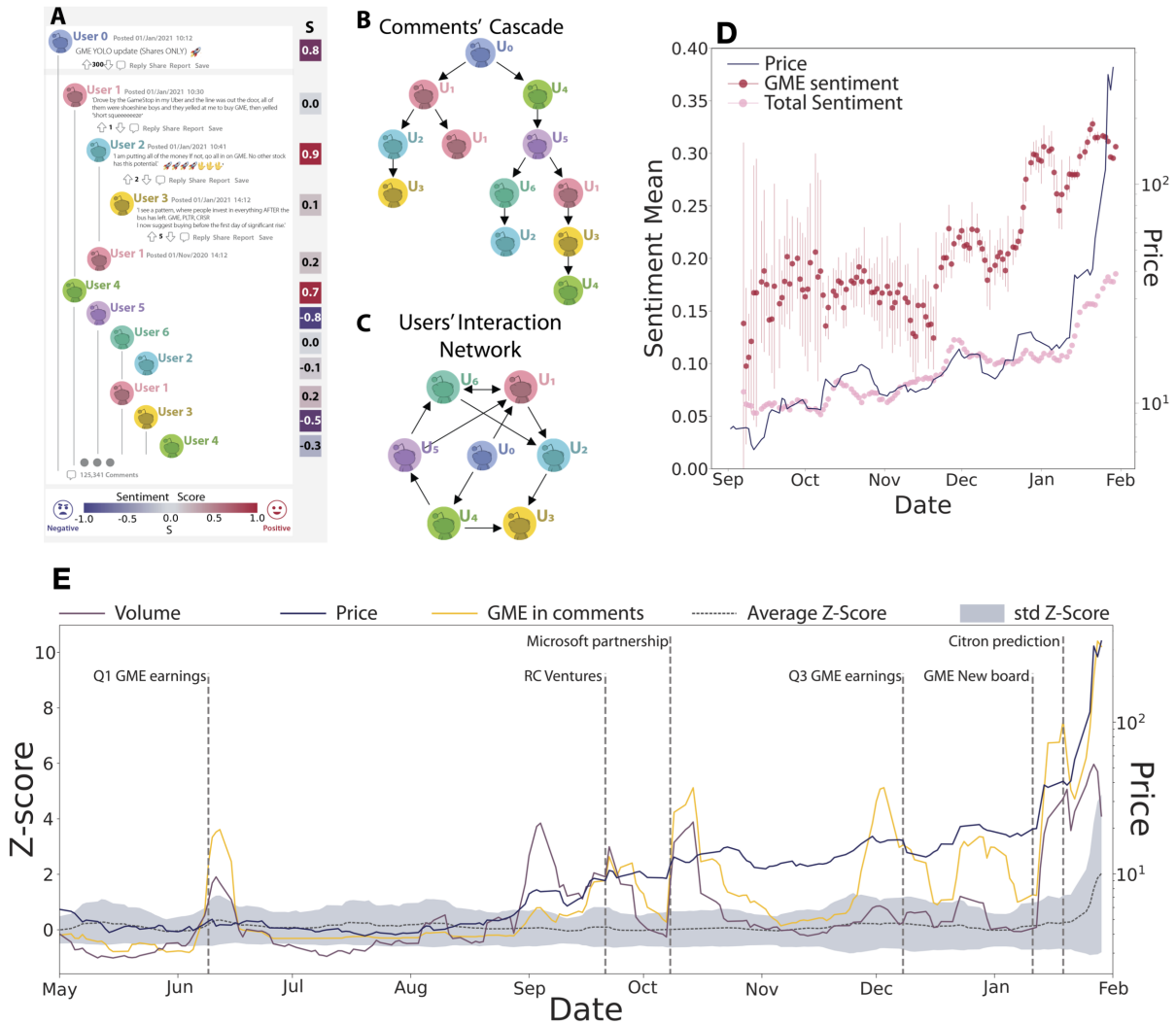


Figure 1: **Reddit conversations and sentiment.** A) Reddit users can write posts to initiate a conversation and can reply to posts or other comments. A sentiment score between -1 and +1 can be computed for each post and comment. B) The conversation on the social network can be represented by a forest of *trees*, where each post is the root of a tree and subsequent comments represent its branches. C) Trees can be used to extract a user-user network, where a link from user i to user j represents the number of times i commented on posts/comments by user j . D) Average sentiment (and standard deviation of the mean) of comments containing 'GME', with respect to the same quantity computed on all comments. GME Close price is also reported for illustrative purposes. E) Z-score for the occurrences of 'GME' in WSB conversations, compared to the mean Z-score for the occurrence of all stock tickers. GME peaks correspond to major events in the GameStop saga. We also report the Z-score of GME trading volumes and the Close share price.

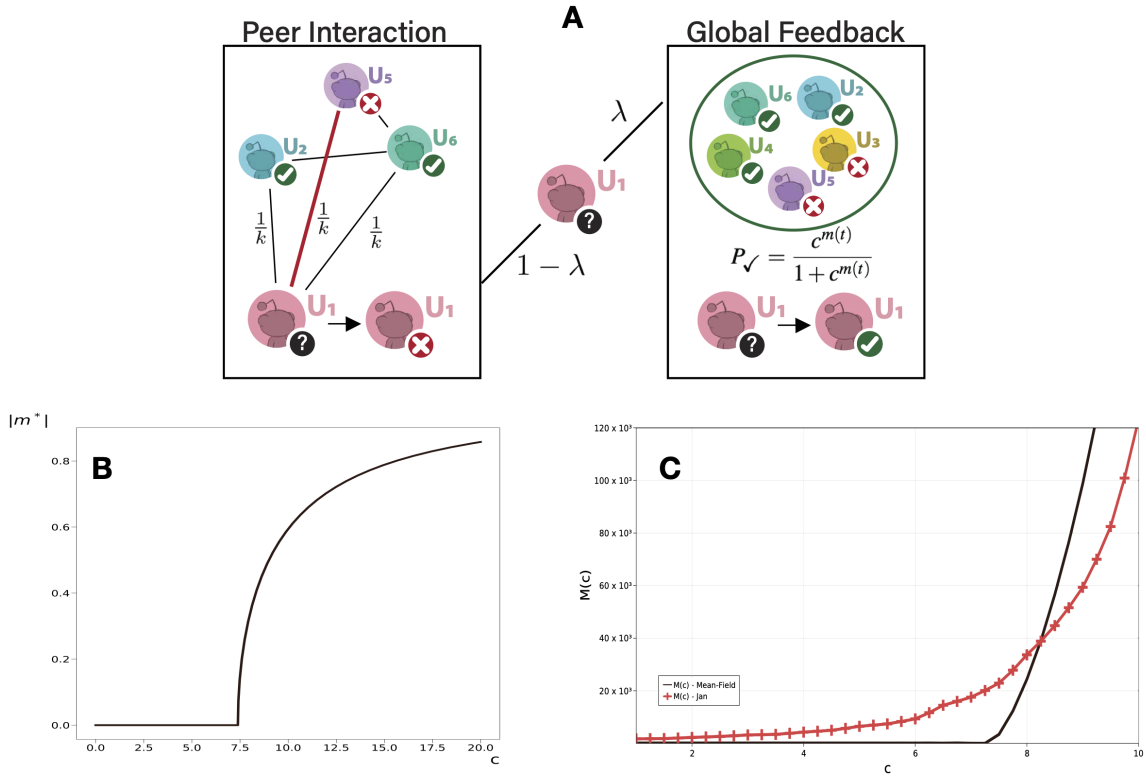


Figure 2: Voter model with self-induced global feedback. A) Schematic representation of the update rule of the model. At each time step, a user takes on the opinion of either a randomly chosen neighbor or is influenced by the current level of consensus in the community. B) Stable equilibrium points $|m^*|$ of the dynamics as a function of c . C) Phase transition of the extensive order parameter of the model (total magnetization for a community that grows as an exponential of m) according to the mean-field solution and to numerical simulations on the January user network.