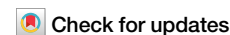# The dynamics of the Reddit collective action leading to the GameStop short squeeze

Check for updates

Antonio Desiderio [1,2,3] ✉, Luca Maria Aiello [4,5], Giulio Cimini[1,2,6] & Laura Alessandretti[3,6]

In early 2021, the stock prices of GameStop, AMC, Nokia, and BlackBerry experienced dramatic increases, triggered by short-squeeze operations that have been largely attributed to Reddit's retail investors. Here we shed light on the extent and timing of Reddit users' influence on the GameStop short squeeze. Using statistical analysis tools with high temporal resolution, we find that increasing Reddit discussions anticipated high trading volumes. This effect emerged abruptly a few weeks before the event but waned once the community gained widespread visibility through Twitter. Meanwhile, the collective investment of the community, quantified through posts of individual positions, closely mirrored the market capitalization of the stock. This evidence suggests a coordinated action of users in developing a shared financial strategy through social media—targeting GameStop first and other stocks afterward. Overall, our results provide novel insights into the role of Reddit users in the dynamics of the GameStop short squeeze.

From swinging elections to fueling social movements, instantaneous global communication enabled by online social media can shape leading global phenomena[1–5]. Financial markets represent a key area where the activity of social media users can have destabilizing effects with worldwide resonance[6]. Indeed, there is ample evidence that user activity on mainstream platforms like Twitter[7,8] and Google Search[9] not only mirrors the mood of the market but can, at times, anticipate the evolution of stock prices[7,9–13]. This goes along with the increasingly central role played by retail investors, or non-professional investors, in a sector that was once dominated by institutional funds and large corporations[14–16].

The dissemination of knowledge on social media stems directly from the dynamic interactions of users[17], who actively share knowledge and opinions on financial markets[18]. Thus, this medium has the potential to enhance the discovery of information that underlies price formation in financial markets[19]. Social communication, indeed, supports the price-discovery process[20] and improves market efficiency, when information is exogenous[21]. However, more than rational agents, social media users tend to resemble noisy traders[22], suggesting that standard models of efficient markets should account for behavioral factors, such as investors' sentiment, public mood[23], cultural traits, psychological biases, social network structure, and information asymmetries[24] as influential for price formation. What emerges from this picture is the complex and evolving interplay between social media signals and stock market trends, which enables narratives, ideas, and investment strategies to rapidly spread to a wide audience[25,26]. Notably, this relationship mainly emerges around specific events, for example, when surges in online interactions catalyze trading activity or, conversely, when large price variations spark online conversations[27]. As such, social media signals tend to capture well transient market trends and stock-specific swings[28].

In early 2021, financial markets were shaken by an unprecedented event: the stock of video-game retail company GameStop (GME) experienced a "short squeeze", with a price surge of almost 1625% within a week[29]. This financial operation was attributed to activity from users of the social media platform Reddit, particularly the subreddit WallStreetBets (WSB), and was rapidly followed by similar market rallies for other stocks: BlackBerry (BB), AMC Entertainment Holdings (AMC), and Nokia (NOK). The GME short squeeze suggested that retail investors, by sharing their investment strategies on social media, could potentially challenge the influence of large institutional funds[29]. The Reddit platform could have supported coordination by fostering community-driven discussions[26,30–32] focused on shared interests and collective goals. On the contrary, mainstream social media like Twitter are primarily used as news feeds[33], and facilitate coordination only to a limited extent through the use of hashtags[4,34].

The GME short squeeze has been extensively analyzed in the scientific literature. Several studies reported a strong correlation between the activity

[1]Physics Department and INFN, University of Rome Tor Vergata, Via della Ricerca Scientifica, 1, 00133 Rome, Italy. [2]Centro Ricerche Enrico Fermi, Via Panisperna, 89a, 00184 Rome, Italy. [3]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, 2800 Copenhagen, Denmark. [4]Computer Science Department, IT University of Copenhagen, Rued Langgaards Vej 7, 2300 Copenhagen, Denmark. [5]Pioneer Centre for AI, Øster Voldgade 3, 2100 Copenhagen, Denmark. [6]These authors jointly supervised this work: Giulio Cimini, Laura Alessandretti. ✉e-mail: antde@dtu.dk

on WSB and the evolution of the GME price throughout the event[26,30,35–38]. However, two major questions remain unanswered regarding the role of WSB users. Firstly, it is still unclear if—and from which specific moment— the activity of WSB users anticipated the actual GME price surge. Secondly, it is unclear if WSB served merely as a platform for financial discourse or as a hub to coordinate a common investment strategy. In this work, we study the temporal relation among three key signals in the period around the GME event: the stock market movements, the activity on WSB, and the broader public attention to the stocks. We employ different techniques, from standard Granger analysis and multivariate vector autoregressive models to study linear effects, to detrended cross-correlation analysis (DCCA) and convergent cross-mapping (CCM) to capture the nonlinear and complex character of the phenomenon. Our analysis provides evidence in support of the hypothesis that activity on the WSB community forecasts the price surge of GME, BB, AMC, and NOK. Further, we shed light on the temporal dynamics of the GME short squeeze, characterizing its evolution through three distinct phases of online behavior: Discussion, Action, and Visibility.

Results

Reddit activity on GME mirrors changes in GME trading volume

We characterize the discourse during the GameStop saga using the volume of GME-related activity on Reddit and on Twitter (see Supplementary Materials, Section 1), the former reflecting the conversation within the community of WSB users and the latter capturing the broader public attention to the stock. Following recent literature[11,26], we quantify these two signals by measuring the hourly occurrence of the GME ticker in the text of posts and comments on Reddit and Twitter, respectively (see Methods for more details). The GameStop saga is described by the series of events reported in Table 1. Interest in GME within the WSB community began to surge as early as December 2020[26,30]. However, despite this growing public discourse, the hourly trading volume and price of GME did not exhibit a significant upward trend until approximately 15 days before the short squeeze date of January 27th—which we associated with Elon Musk's tweet. As shown in Fig. 1a, this trend closely resembles the increase in GME-related conversation volume on social media.

This observation suggests a deeper connection between the two signals. To shed light on this point, we employ detrended cross-correlation analysis (DCCA). This method allows investigating cross-correlations between time series in the presence of non-stationarity[39], hence addressing prevalent effects observed in financial time series such as non-linearity, long-term underlying trends, and long-range dependencies[40]. Specifically, through DCCA, we gauge if fluctuations in GME-related activity are correlated with fluctuations of GME market volume across multiple time scales[11] (see Methods for further details). We apply DCCA to time series of absolute values of hourly returns for GME occurrences, stock price, and trading volume[39] from December 1st, 2020 to July 1st, 2021, computing the detrended covariances for each combination of signal pairs (see Methods). Results shown in Fig. 1b indicate that such detrended covariances follow power-law scaling, pointing to the long-term coupling between the signals (see Supplementary Materials, Section 2). Remarkably, the GME trading

volume displays larger detrended covariance with Reddit activity than with GME price, with a cross-correlation coefficient three times larger than the one observed for the other pairs of signals (see Inset Fig. 1b). This suggests that changes in trading volume are more tied to Reddit discussions, rather than to price movements.

We further employ convergent cross-mapping (CCM) to corroborate the dynamic interdependence between Reddit activity and trading volume[41]. This method quantifies the nonlinear coupling between variables by assessing whether past states of one variable x can predict future states of another variable y, without assuming model-specific dependencies. The results of this approach confirm a strong dynamic coupling between Reddit activity and trading volume (see Supplementary Materials, Section 3). Overall, we can conclude that large variations in Reddit discussions are strongly related to changes in trading volume, as more discussions correspond to a general increase in buying activity and vice versa.

The collective position of WSB users on GME amounts to at least 1% of the stock's market capitalization

In WSB, some users provide evidence of their financial investments by posting screenshots of their accounts taken from trading platforms, which attest to their positions and trading activity[30]. We analyse such screenshots to capture the actual engagement of WSB users with the GME stock[11,26], using computer vision techniques (see Methods for more details and Supplementary Materials, Section 4). We observe that the values of the individual positions on GME follow a Log-normal distribution (see the Inset of Fig. 1c). By summing the positions across all users, we obtain the posted collective investment, which provides a lower-bound estimate of the community stake in GME, as it includes only the retail investors who have publicly shared their positions. We find that this posted collective position grew significantly 15 days prior to the squeeze. This finding complements the observation that the number of WSB users posting screenshots grew in the period preceding the squeeze[30], likely indicating that we have more users investing in GME. Notably, the growth of the posted collective position closely follows the market capitalization of GME (see Fig. 1c). We estimate that the posted collective investments of Reddit users represented—on average over the time period considered—at least the 1% of the total market capitalization of GME (see Fig. 1c).

Granger analysis identifies three distinct phases of online behavior

We now test the hypothesis that changes in GME-related activity on social media *anticipated* changes in GME trading volume around the short squeeze. Specifically, we employ the Granger test to assess the chronological anticipation $x \rightarrow y$: whether a signal x has predictive power on another signal y, building a linear regression model for $y(t)$ leveraging past information on both y and x from $t - \Delta t$ up to $t - 1$. The Granger Index[42] quantifies the test result as the ratio of predictive accuracy between the model and a baseline that excludes past information derived from x (see Methods for more details). As x and y we consider the possible pairs of signals among GME

Table 1 | Major events relevant to the GameStop (GME) short squeeze and the subreddit r/WallStreetBets (WSB)

Date	Event	Description
2020-12-08	GME Q3 Earnings Reports	GameStop's Q3 earnings highlighted a significant 257% surge in e-commerce revenues.
2021-01-11	New Board of Directors	GameStop revealed a new Board of Directors, including several experts in e-commerce, signaling a new leadership direction.
2021-01-13	GME Stock Rise Begins	GME stock price started rising significantly, increasing by around 60% in a single day.
2021-01-19	Citron Tweet	Citron Research criticized GameStop's stock valuation and forecasted a price drop, belittling GME buyers on Twitter.
2021-01-22	BB Stock Surge	BlackBerry stock surged by 30%.
2021-01-27	Elon Musk Tweet	Elon Musk tweeted, "Gamestonk!" with a link to the subreddit r/WallStreetBets. GME stock surged over 135%.
2021-01-28	Robinhood Trading Restrictions	Robinhood and other brokerage platforms restricted the buying of GME, AMC, BB, and NOK stocks, leading to a temporary price crash. AMC stock rose by 300%, BlackBerry by 32%, and Nokia stocks surged by 30%.
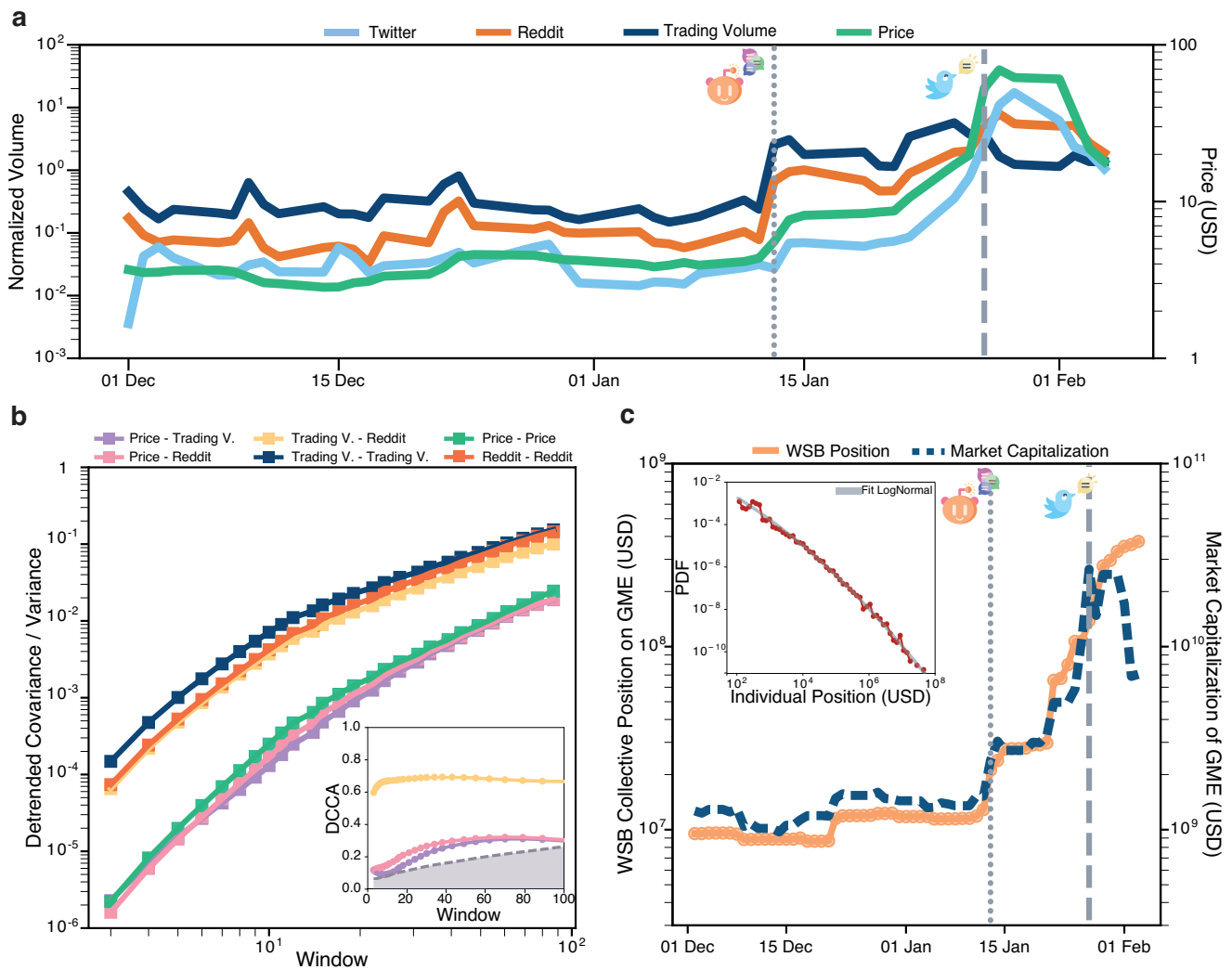
Fig. 1 | The correlation between Reddit activity and GameStop market indicators.
a Daily occurrences of the "GME" ticker in conversations on WSB (orange, y-right axis) and Twitter (lightblue, y-right axis), along with the daily trading volume of the GME stock (blue, right y-axis) and its daily closing price (green, left y-axis). The vertical gray dotted line marks January 13th, identified as the WSB-led action's start (see subsequent sections), while the dashed line marks January 27th, when Elon Musk's Tweet broadly publicized the action. To improve the chart's clarity, we applied a 5-day moving average to each signal and normalized the trading volume, Reddit, and Twitter signals by their mean (we remark these modifications are not employed in the analysis). **b** Values of detrended variance for the price (green), trading volume (blue) and Reddit activity (red) signals, and values of detrended covariance for the pairs price / trading volume (purple), price/Reddit activity (pink), trading volume/Reddit activity (yellow), as a function of the time window considered. Inset: detrended cross-correlation coefficients for paired combinations of GME signals (same color schemes of the main panel). The shaded gray area represents the 95% confidence interval for correlation values obtained from pairs of independent signals. **c** Daily value of the WSB posted collective position on GME (orange, y-right axis) and GME Market Capitalization (blue, y-left axis). Inset: Distribution of values of users' individual position on GME (red) alongside the Log-Normal fit (gray line, $\mu = 4.96$, $\sigma = 3.63$).

trading volume and GME-related discussions on Reddit and Twitter. We consider hourly values during market opening times, within a 15-day time window (corresponding to 120 points per time series – see Methods) between January 5 and February 5. This choice of window size ensures that we capture short-term patterns that are relevant to our case study[28] and, at the same time, that we have enough statistics to draw robust conclusions from the Granger analyses, even in regimes of small coupling[43]. Results are, however, robust to different choices of window length (see Supplementary Materials, Section 5). After testing the stationarity of the time series within the specified windows (see Supplementary Materials, Section 5), we apply a daily moving average (the results remain robust when we do not apply the moving average, see Supplementary Materials, Section 5) and compute the hourly log-returns of the signals.

Figure 2a shows the resulting Granger Index for $\Delta t = 1$, meaning that the model uses $x(t-1)$ to predict $y(t)$. In the plot, triangles correspond to statistically significant values (p values <0.05). Size effects are instead

captured by the coefficient of the regression of the independent variable, reported in Fig. 2b for the Reddit-to-Trading Volume direction (see Supplementary Materials, Section 5 for other values). Results highlight three distinct phases of online behavior (see Fig. 2a).

Phase 1. Before January 13th, the Granger Index and size effects are not significantly different from zero (p values >0.1) for all combinations of signals, meaning that no anticipatory effects are observed.

Phase 2. On January 13, the Granger Index capturing the anticipatory power of the Reddit activity on trading volume has a sharp transition to a very high value of 0.47 (p values <0.05). At the same time, size effects grow up to 1.5, meaning that social media activity has a large positive effect on the prediction of market trends. Simultaneously, the predictive power of the trading volume on Reddit is significant; however, the Granger Index in the Reddit-to-Trading Volume direction is roughly
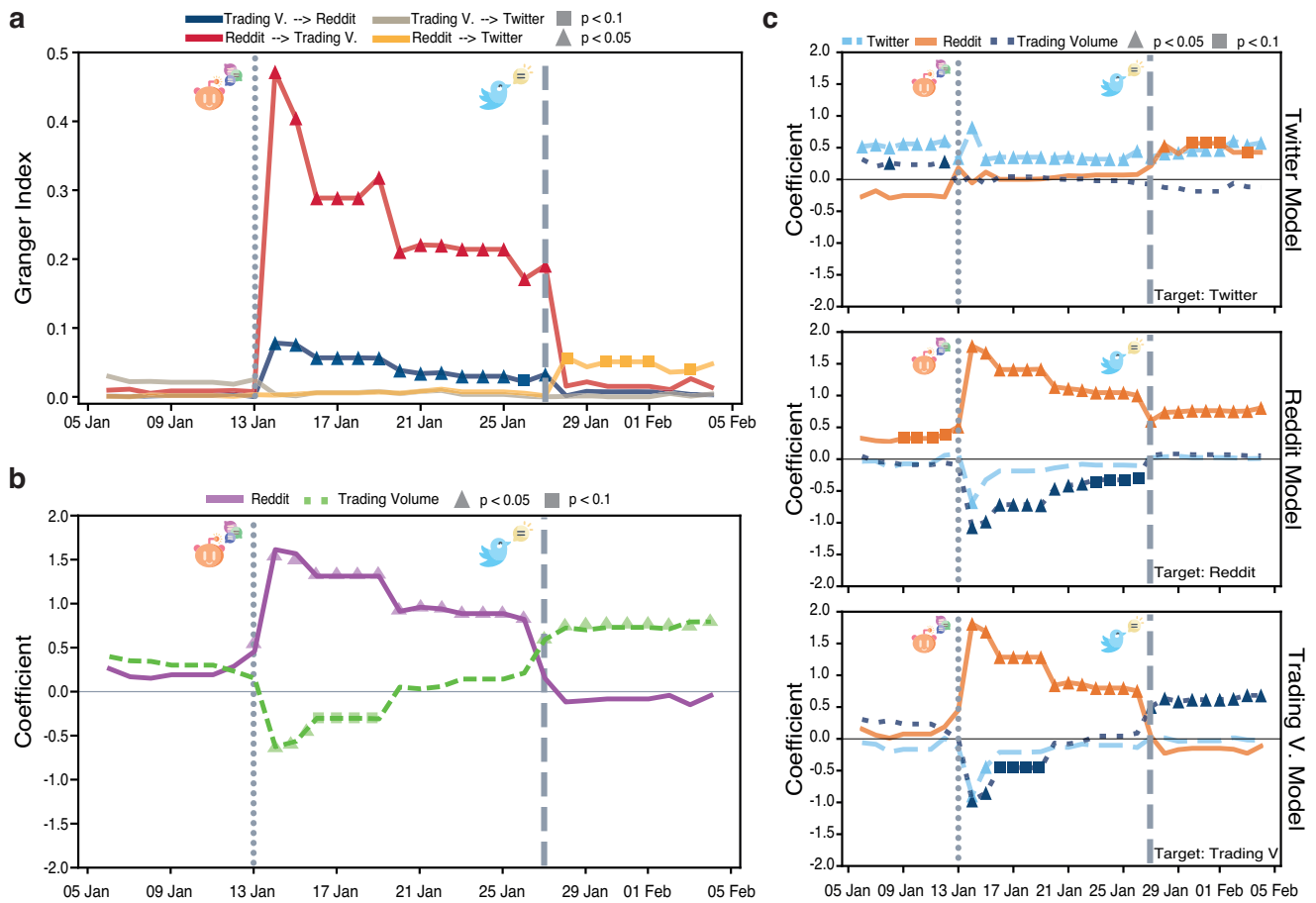
Fig. 2 | Reddit activity anticipates trading volume during the short squeeze. In all the following subplots, the vertical gray dotted line indicates the beginning of the *Action* phase (13 January 2021), whereas the vertical gray dashed line corresponds to the Tweet by Elon Musk (27 January 2021) that brought the squeeze to the public attention. Triangles correspond to *p* values <0.05 and squares to *p* values <0.1. Each point is computed considering a time series spanning the 15 preceding days. **a** Granger index capturing the predictive power of a signal on another (with a lag of 1 h) for the following pairs: Trading Volume-to-Reddit (blue), Reddit-to-Trading Volume (red), Trading Volume-to-Twitter (gray), and Reddit-to-Twitter (yellow). **b** Coefficients of the Granger model predicting Trading Volume, capturing the size effects of Reddit (magenta) and Trading Volume (green) activities. **c** Each panel shows the coefficients of a multivariate vector autoregressive model predicting Twitter (top), Reddit (middle), and Trading Volume (bottom) activities, corresponding to antecedent values of Reddit (solid orange), Twitter (dashed lightblue), and trading volume (dotted blue).

three times larger than in the opposite Trading Volume-to-Reddit direction. These indices remain significant—although steadily declining —for about 12 days. The *p* values are reported in Supplementary Materials, Section 5.

Phase 3. On January 27th, the Tweet "Gamestonk!!" by Elon Musk brought the GME short squeeze to the public attention. On this date, the anticipatory power of Reddit activity on the trading volume (and vice versa) is non significant, and size effects vanish. Instead, Reddit activity started to anticipate the activity on Twitter: once they became informed about the short squeeze, Twitter users turned to Reddit for information about GME.

The three phases identified above suggest the following sequence of events. Prior to January 13th, WSB users were discussing GME but did not play a significant role in the market. Afterward, from January 13th to 27th, their coordinated action of buying and holding GME shares began to influence the market. Finally, after January 27th, the operation gained public attention, and the market was influenced by a much broader audience and other market players—eclipsing the role of WSB users. At the same time, a larger horde of people joined the WSB community, fundamentally altering the nature of discussions[26,30]. To reflect their distinct nature, in the following, we refer to these three phases as the *Discussion*, *Action*, and *Visibility* phases, respectively.

Reinforcing effects on Reddit, Twitter, and Trading Volume

The analysis in the previous section considers only pairwise relations among signals. In order to capture how the mutual relationship between all three signals under study changes over time, we employ a multivariate vector autoregressive model. The method generalizes the Granger analysis to multiple signals: it entails fitting a linear model that predicts a target output signal y at time t, using past information from $t-1$ to $t-\Delta t$ on y itself plus one or more input feature signals $\{x_1, x_2, x_3, \dots\}$. The coefficients of the model capture how the individual input signals contribute to predicting future values of y (see Methods for more details). In particular, a positive coefficient for x_i indicates a positive reinforcing relationship of x_i on y, while a negative coefficient is an opposite and counteracting effect. We apply the method to study the hourly log-returns of Twitter activity, Reddit activity, and trading volume. As we did for the Granger analysys, we fitted the model within each 15-day period ending on a day between January 5th and February 5th, for $\Delta t = 1$ h (results for other window period lengths are reported in the Supplementary Materials, Section 5).

Figure 2c shows the coefficients that capture the predictive power of the three signals on the future values of Twitter (top plot), Reddit (mid plot), and trading volume (bottom plot). First, we observe a self-reinforcing dynamic within Reddit and Twitter throughout the whole period under study, where increased activity within the social network leads to further increases. This is evident from the consistently positive and significant coefficients of the

Reddit-to-Reddit and Twitter-to-Twitter relation. During the *Action* phase, the self-reinforcing nature of Reddit is even more pronounced.

Secondly, during the *Action* phase, the coefficient capturing the Reddit-to-Trading Volume relation is significant and positive, implying that the increase of activity on Reddit is followed—within the next hour— by an increase in trading volume. This is coherent with the size effects observed in the Granger analysis. Notably, in this phase, the magnitude of the Reddit-to-Trading Volume coefficient is similar to the magnitude of the Reddit-to-Reddit coefficient (the two can be directly compared since we are considering the log-returns of the signals).

Finally, during the *Visibility* phase, the trading volume starts following self-reinforcing dynamics (as indicated by the significant and positive coefficient for the Trading Volume-to-Trading Volume relation), while the Reddit-to-Trading Volume coefficient is non significant. In this phase, the Reddit-to-Twitter coefficient aligns in both magnitude and direction with the Twitter-to-Twitter coefficient, again highlighting that public attention was directed toward WSB.

By conducting this analysis at various lags Δt, we found that the coefficients at $\Delta t = 1$ h are larger than the others (see Supplementary Materials, Section 5), indicating the greater significance of short time scales in the entwined dynamics of social media and markets. Interestingly, the typical reply speed, defined as the average time interval between a comment and its direct responses, varies between ~2.5 h during the *Discussion* phase to ~0.5 h during the *Visibility* phase (see Supplementary Materials, Section 6). Hence, the influence of Reddit on the market occurs on a timescale closely aligned with that of the unfolding conversations.

The role of WSB users on other short squeezes
Alongside GME, also BB, AMC, and NOK experienced a short squeeze in January-February 2021. In this section, we investigate the interplay between discussions on WBS and these financial events, employing the techniques introduced in the previous sections. Our analysis focuses on the occurrences of the BB, AMC and NOK tickers on WSB posts and comments, against the trading volumes of the respective stocks.

By the end of 2020, BB, AMC, and NOK ranked below position ten among the stocks that were most discussed on WSB, while Palantir (PLTR), Tesla (TSLA), and NIO Inc. (NIO) occupied the top ranks together with GME (see Fig. 3a, top panel). Interestingly, during the GME *Action* phase, we observe a fast ramp up of BB, AMC and NOK which—by mid January 2021—became the most discussed stocks, just below GME.

The Granger analysis reveals that, during the GME *Action* phase, the Reddit-to-Trading Volume anticipatory relation is only significant for BB (see Fig. 3b). Interestingly, the Granger Index for BB is statistically significant only five days after it does so for GME. The analysis for AMC and NOK, instead, does not lead to significant conclusions (see Supplementary Materials, Section 5), possibly due to the less attention they received on Reddit (see Fig. 3a). We refer the reader to the Supplementary Materials, Section 2 for the analyses of long-range detrended cross-correlations (DCCA) and Convergent Cross-Mapping (Supplementary Materials, Section 3) between trading volume and the frequency of occurrences for BB, AMC, and NOK.

We now consider the Multivariate Autoregression analysis of BB (see Supplementary Materials, Section 5 for the analysis of NOK and AMC). As Fig. 3c shows, we find that the Reddit-to-Trading Volume anticipatory effect is reinforcing, as demonstrated by a significant and positive model coefficient. We observe, however, a different behavior compared to the GME case. At the beginning of the *Action* phase, the Trading Volume has a positive reinforcing effect on both Reddit and itself; at the same time, Trading Volume is much better predicted by itself than by Reddit. This trend reverses one week later, when Trading Volume loses all its explanatory power, while Reddit enters into a positive relationship with itself and Trading Volume.

But how do BB, AMC, and NOK gained popularity among WSB users? To study this phenomenon, we compute the overlap coefficient (size of the intersection over size of the smaller set) and Jaccard index (size of the intersection over size of the union) among the set of users who talk about

GME and those who talk about other stocks (see Fig. 3d, e). On January 13th, concurrently with the beginning of the *Action* phase, we observe a sharp increase in the overlap between users discussing GME and other stocks. After January 13th, the overlap index has a plateau at a high value close to 0.8, signaling an expansion of discussions among WSB users. Initially focused on GME, they broadened their conversations to include various other stocks. By studying the Jaccard index, we notice that up to 40% of users interested in GME focused initially also on BB; then, as the short squeeze approached, they turned to NOK and AMC. During the same period, the number of users endorsing stocks that were top-ranked before December (PLTR, TSLA, and NIO) decreased. When we compute these indices, the group of users who talk about the other stock (not GME) is always smaller, while the larger group is mostly made up of users discussing GME (see Supplementary Materials, Section 7). Additional analyses based on the raw number of users discussing the various stocks show that, prior to the start of the GME rally, users stopped discussing NIO, and during the rally started focusing on BB, NOK, and AMC (see Supplementary Materials, Section 7). Simultaneously, mega-threads dedicated to GME appear on WSB[26,44], providing additional visibility to other stocks that GME users were also talking about.

These findings suggest that WSB users engaged with GME also endorsed BB, AMC and NOK. Approaching the GME short squeeze, and concurrently with the widespread recognition of WSB on a global level, these stocks emerged as the most popular choices and ended up as short squeeze targets.

Discussion
In January 2021, the activity of WSB users anticipated financial trends. The high temporal resolution of our analysis enabled us to identify three distinct phases of online behavior and engagement with the stock market, characterized by Discussion, Action, and Visibility, respectively. First, we observed that the strong anticipatory relation between WSB users' activity and changes in GME trading volume emerged sharply on January 13th. This suggests a sudden transition between a phase characterized by *Discussion* and one characterized by financial *Action*. Second, we found that—at around the same time—the value of the collective position on GME started to closely mirror the evolution of the market capitalization of GME. We estimated that the size of the total position of WSB users amounted to at least 1% of GME market capitalization, demonstrating a tangible engagement of the WSB community in the stock market. This suggests that other market players joined the short squeeze after it was initiated by WSB users, despite having access to information about GME that was presumably not inferior to that of the WSB users[45]. After the price hike of GME and a Tweet by Elon Musk on January 27th, the short squeeze became widely known, and a broader public started to follow Reddit conversations.

Our study comes with limitations. First, our findings refer to a specific case study, the GME short squeeze, which so far represents a unique event due to its wide resonance. If and when such events will occur again, we could understand whether the patterns we have observed can be generalized to different contexts. Secondly, our analysis primarily relies on the Granger test, a standard statistical framework that relies on linear models. As such, it may oversimplify complex relationships, such as those involved in collective action between social networks and financial markets, where feedback loops and nonlinear dynamics can occur. Recognizing this limitation, we have explored nonlinear techniques; notably, an analysis based on convergent cross-mapping[41] corroborates our main results (see Supplementary Materials, Section 3). Third, due to the limitation of the available data, our analysis focused solely on Reddit and Twitter, neglecting offline interactions happening on other online platforms (e.g., Telegram or Discord), as well as the online/offline actions of institutional investors. Finally, the proxy used to estimate the size of the WSB collective position on GME, specifically from user screenshots, are inherently partial and noisy. We also have no information on the identities of the investors behind the screenshots, so we cannot exclude the potential interference by institutional investors in the financial discourse on Reddit.
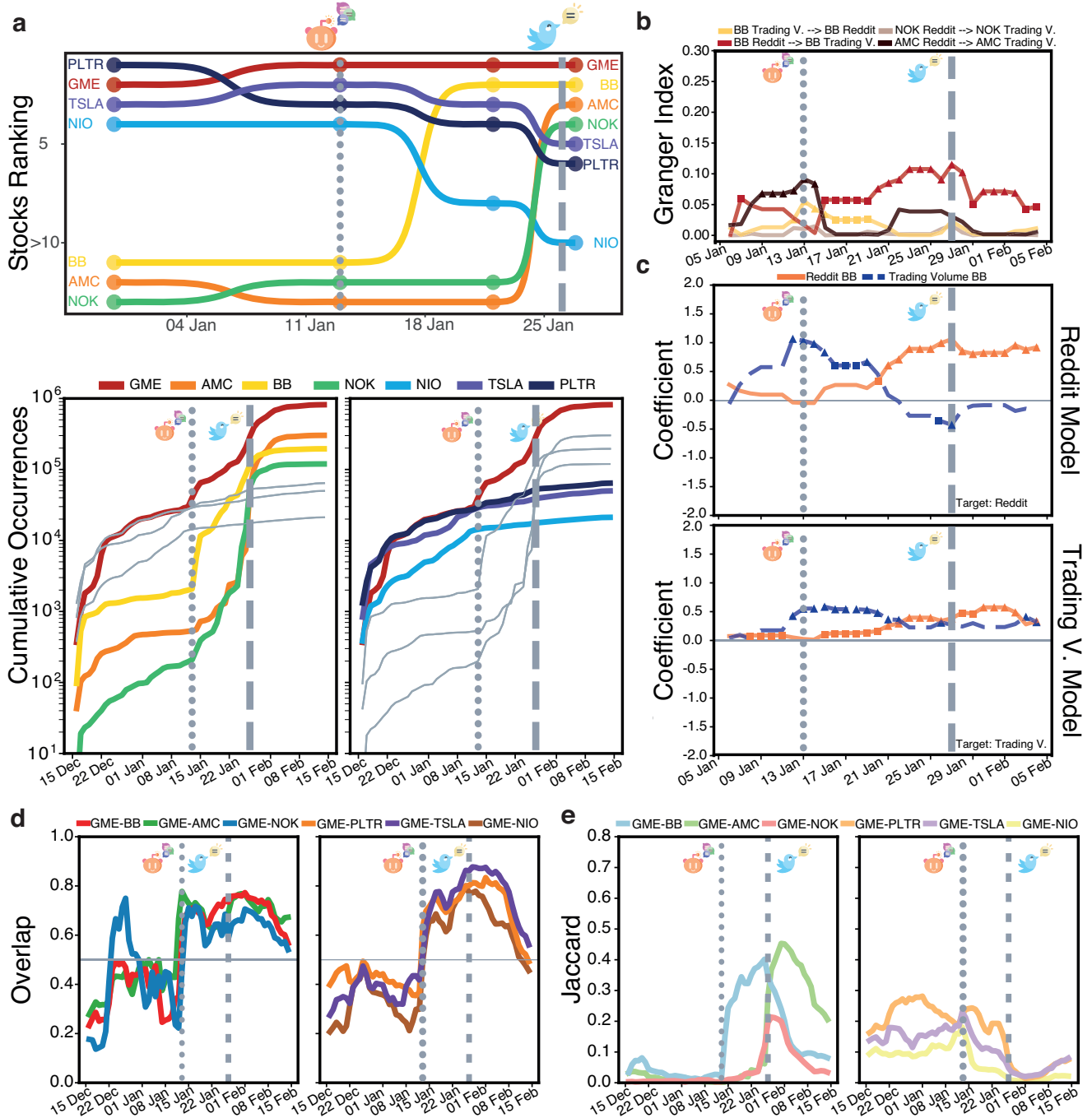
Fig. 3 | BB, AMC, NOK, and the role of GME-engaged users. In all the following subplots, the vertical gray dotted line indicates the beginning of the GME *Action* (13 January 2021), whereas the vertical gray dashed line corresponds to the Tweet by Elon Musk (27 January 2021). Triangles correspond to *p* values <0.05 and squares to *p* values <0.1. **a** (top) Rank chart of stock tickers on WSB based on their average daily occurrence within each 15-days window. **a** (bottom) Cumulative number of occurrences of a given stock on WSB. **b** Granger Index for possible combinations of BB, NOK, and AMC signals. **c** Coefficients of the Multivariate vector autoregressive model for BB-related activity on Reddit (top panel) and BB trading volume (bottom panel), related to the Reddit (solid orange) and trading volume (dashed blue) signals. **d** Overlap index for pairs of stocks, computed over a 5-day window with a one-day shift, measuring the proportion of users discussing each pair of stocks. We observe in both panels an increasing overlap after the starting of the GME rally. **e** Jaccard index for pairs of stocks, computed over a 5-day window with a one-day shift, measuring the proportion of users discussing both pairs of stocks.

Despite such limitations, we have provided extensive evidence regarding the possible influence exerted by Reddit users on the GameStop short squeeze. However, determining a genuine causal link —that Reddit users directly fueled the price rally—is not possible with the data we can access. Factors such as the true portfolio composition of each Reddit user and the potential psychological contagion within Reddit discussions add uncertainty for the attribution of direct causation.

Overall, our results add further evidence that digital platforms can catalyze collective action and influence real-world outcomes. In an era characterized by rapid technological advancements, the pace of our daily lives is accelerated to an unprecedented rate[46]. This phenomenon has also

impacted the financial landscape, where the emergence of commission-free trading platforms started a new age of dynamism and accessibility[29]. Our findings highlight the importance of investigating the dynamic interplay between online engagement and stock market movements at an hourly scale (or even finer)[7,9–11], as the coordinated action of retail traders through social media has become a no-longer-negligible market player.

Methods
Data descriptions
We gathered Reddit conversation data from Pushshift[47], an API that regularly copies the activity data of Reddit. We queried the service to retrieve information about posts and comments on WSB from December 01, 2020 to July 1, 2021. The dataset was cleaned by removing posts/comments by Reddit bots (see Supplementary Materials, Section 1) as well as by "[deleted]" users, who deleted their accounts at the time Pushshift API collected data. To obtain the occurrences of a given stock within WSB conversations, we counted how many times the ticker symbol of the stock (e.g., "NOK" for Nokia) appears as a regular expression in the raw text of a post/comment[26].

Within the WSB community, a common practice consists of posting financial investments by sharing screenshots of open positions, primarily gains, losses, or orders. These "committed" screenshots were identified and classified using computer vision techniques by Lucchini et al.[30]. Using Tesseract, an open-source optical character recognition[48], we extract textual information from these screenshots, converting image-based text into editable and searchable content. We focused on screenshots containing keywords such as 'gme,' 'gamestop,' 'investing,' and 'value,' identified through manual inspection, and subsequently narrowed down the sample to 6525 screenshots from November 1, 2020 to February 04, 2021. After parsing the structures, we isolated dollar-sign-prefixed numbers, determining the final value of the position as the highest extracted value. However, such value may represent trading volume or the price of GME taken from online trading applications, rather than an actual investment. To address this, we filtered out screenshots with values below 10^2 and above 10^6, as these are likely indicative of price or trading volume. We manually examined values above 10^6 to validate our findings. Additionally, we manually validated 1000 screenshots using log-transformed data, computing the logarithms of the ratios between true and estimated values. We determined that our procedure extracts the correct value with an accuracy of 0.85, where accuracy is based on a binary classifier where correctness is determined by applying a threshold of 0.05 to the absolute logarithmic ratios (see Supplementary Materials, Section 4). The root mean square error (RMSE) of the logarithmic ratios between true and estimated values is 0.47, and the point estimate (i.e., the average ratio for a single screenshot) amounts to -0.07. This indicates that the automated predictions slightly overestimate the actual values (on average $+17\%$). However, this difference is more pronounced at screenshots values below 10^3, with a relative logarithmic error of -0.1, while it approaches zero at higher values. Overall the automated extraction procedure leads to a total discrepancy of approximately $+0.2\%$ in the WSB posted collective position. Note that this validation was intended solely to assess the error of the employed framework in estimating the collective financial position—the incorrect estimates were not replaced. Indeed, missing posts, and possibly fake ones, represent a much larger source of error, which we cannot control.

We measured Twitter's interest in GME using Twitter API for academic research. From this service, we collected tweets with the #GME or #Gamestop hashtags posted between December 1, 2020 and February 5, 2021[30]. Data on stock close price and traded volumes were retrieved from the API service of https://polygon.io. Data on stock close price and traded volumes were retrieved from the API service of https://polygon.io, wherein the price data reflects stock split adjustments that occurred during the stock split in 2022. All signals were taken into account exclusively during market hours (9 a.m. to 5 p.m.). When analyzing the changes of a particular signal, denoted as $x(t)$, we compute its hourly logarithmic return as $r(t) = \ln[x(t+1)/x(t)]$.

Detrended cross-correlation analysis
Detrended cross-correlation analysis is a statistical technique designed to explore long-range dependencies between two time series, by studying how the fluctuations of local trends evolve across temporal scales[39,49]. The underlying idea is to decompose two signals into smaller overlapping segments using a given time window, compute the covariance between the two detrended signals within each segment, average the results obtained over the various segments, and finally repeat all the steps for windows of different lengths to assess potential scale relations. Following Podobnik et al.[49], we consider two time series $x(\tau)$ and $y(\tau)$ of equal length T and compute the integrated signals $I^{(x)}(t) = \sum_{\tau=1}^{t} x(\tau)$ of each time series. For the analysis at window length t, we divide the signals into $T - t$ consecutive segments (window t), each containing $t + 1$ values. In each segment that starts at i and ends at $i + t$, we define the local trend as the ordinate of a linear least-squares fit $S_i^{(x)}(t)$ and remove it from the integrated windowed signal $I_i^{(x)}(t)$ to compute the residuals. We define the covariance of the residuals in each segment as

$$f_{DCOV}(i, t) = \frac{1}{t-1} \sum_{\tau=i}^{t+i} (I^{(x)}(\tau) - S_i^{(x)}(\tau))(I^{(y)}(\tau) - S_i^{(y)}(\tau)), \quad (1)$$

and if $x = y$ we have the variance of the residuals $f_{DVAR}^2(i, t)$. Then, the detrended covariance $F_{DCOV}^{(x,y)}(t)$ is defined as the average covariance of the residuals over the overlapping segments at the given window t: $F_{DCOV}^{(x,y)}(t) = \frac{1}{T-t} \sum_{j=1}^{T-t} f_{DCOV}(j, t)$. In our case, the time series have temporal length $T \sim 10^3$ and the windows considered are logarithmically spaced, each containing 30 points. The detrended cross-correlations coefficient is finally defined as

$$F_{DCCA}(t) = \frac{F_{DCOV}^{(x,y)}(t)}{\sqrt{F_{DVAR}^{(x,x)}(t)} \sqrt{F_{DVAR}^{(y,y)}(t)}}. \quad (2)$$

If two series are power-law cross-correlated, then the detrended covariance as a function of the window length t follows a power law, $F_{DCOV}^{(x,y)}(t) \sim t^\lambda$ and the magnitude of F_{DCCA} indicates the correlation strength[49]. The exponent λ identifies whether this cross-correlation is short-range ($\lambda = 0.5$) or long-range ($\lambda > 0.5$). If two time series are short-range correlated or cross-correlations is absent, F_{DCCA} could be non-zero and slightly be dependent on window t due to size effects[49]. To identify statistically significant values, we compare the obtained cross-correlations against those obtained for two independent random signals (see Inset Fig. 1c, gray area)[49]. In our case, we observe values of cross-correlation always exceeding the confidence interval, and scaling exponents $\lambda > 0.5$, pointing to power-law cross-correlations (see Supplementary Materials, Section 2). However, it is crucial to interpret the fitting of the exponents as an exercise due to the limitation of short time series. We further corroborate our findings by estimating the Hurst exponent through R/S analysis (see Supplementary Materials, Section 2).

Granger test and multivariate vector autoregressive model
The Granger test is commonly used in time series analysis to determine whether the predictability of $y(t)$ in a given model decreases when the time series $x(t)$ is excluded from that model[10]. The key idea is that when both past values of x and y are incorporated rather than just past values of y, the prediction of y improves, thus indicating that x Granger-causes y. In our case, we consider a vector autoregressive model of order Δt

$$y_t = \widetilde{a}_0 + \sum_{l=1}^{\Delta t} \widetilde{a}_l y_{t-l} + \widetilde{\epsilon}_t \qquad \text{and}$$

$$y_t = \widehat{a}_0 + \sum_{l=1}^{\Delta t} \widehat{a}_l y_{t-l} + \sum_{l=1}^{\Delta t} \widehat{b}_l x_{t-l} + \widehat{\epsilon}_t, \qquad (3)$$

where the first equation is the restricted model and the second the full model. After fitting these models, the p values are computed from the

F-statistic, which reflects the difference in the residual sum of squares (*RSS*). These *p* values indicate the likelihood of obtaining results as extreme as the observed F-statistic under the assumption that the restricted model is correct. Size effects are given by the values of the regression coefficients $\hat{a}_l$ and $\hat{b}_l$, which capture the contribution of the lagged variables *y* and *x* in predicting *y*. The Granger Index, on the other hand, focuses only on the predictive accuracy of the full model in relation to the restricted model, defined as $GI = \ln(RSS_{restricted}/RSS_{full})$. To conduct the Granger test, we need first to evaluate the stationarity of signals within the specified windows. The results of the Augmented Dickey–Fuller test are reported in Supplementary Materials, Section 5.

Following the full model of the Granger test, we consider a multivariate vector autoregressive model to analyze relationships and dependencies among two or more time series simultaneously. The significance of a particular regression coefficient is established by evaluating its corresponding *p* value, computed similarly to the Granger analysis. Both for Granger and nultivariate vector autoregressive model, we analyzed the data by sliding it within 15-day windows, shifting by one day each time. The resulting length of the time series is thus 120 points (8 hourly points per day)—a reasonable statistic to perform linear regression analysis[43,50,51].

Data availability

Reddit conversation data used in this study (the same used by Mancini et al.[26] and Lucchini et al.[30]) can be retrieved from the Pushshift API at https://www.reddit.com/r/pushshift/and were collected in December 2022. Twitter data used in this analysis are the same used by Lucchini et al.[30], and were retrieved from Twitter API before April 2022. Stock market data used in this analysis can be retrieved from the Polygon.io API at https://polygon.ioand were collected in December 2022, while the GME Market Capitalization was collected in September 2023.

Code availability

The data and code to reproduce the analysis is released on GitHub. For inquiries, please contact A.D. antde@dtu.dk.

References

1. Heidemann, J., Klier, M. & Probst, F. Online social networks: a survey of a global phenomenon. *Comput. Netw.* **56**, 3866–3878 (2012).
2. Aral, S., Dellarocas, C. & Godes, D. Introduction to the special issue -social media and business transformation: a framework for research. *Inform. Syst. Res.* **24**, 3–13 (2013).
3. Halu, A., Zhao, K., Baronchelli, A. & Bianconi, G. Connect and win: the role of social networks in political elections. *Europhys. Lett.* **102**, 16002 (2013).
4. Wolfsfeld, G., Segev, E. & Sheafer, T. Social media and the Arab spring: politics comes first. *Int. J. Press Politics* **18**, 115–137 (2013).
5. Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A. & Quattrociocchi, W. Mapping social dynamics on Facebook: the Brexit debate. *Social Netw.* **50**, 6–16 (2017).
6. Bouchaud, J. -P., Farmer, J. D. & Lillo, F. *Handbooks of Financial Markets* (Wiley, 2009).
7. Bollen, J., Mao, H. & Zeng, X. Twitter mood predicts the stock market. *J. Comput. Sci.* **2**, 1–8 (2011).
8. Oliveira, N., Cortez, P. & Areal, N. The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Syst. Appl.* **73**, 125–144 (2017).
9. Moat, H. S. et al. Quantifying wikipedia usage patterns before stock market moves. *Sci. Rep.* **3**, 1801 (2013).
10. Bordino, I. et al. Web search queries can predict stock market volumes. *PLoS ONE* **7**, 40014 (2012).
11. Curme, C., Preis, T., Stanley, H. E. & Moat, H. S. Quantifying the semantics of search behavior before stock market moves. *Proc. Natl Acad. Sci. USA* **111**, 11600–11605 (2014).
12. Souma, W., Vodenska, I. & Aoyama, H. Enhanced news sentiment analysis using deep learning methods. *J. Comput. Soc. Sci.* **2**, 33–46 (2019).
13. Souza, T. T. P. & Aste, T. Predicting future stock market structure by combining social and financial network information. *Phys. A Stat. Mech. Appl.* **535**, 122343 (2019).
14. Aramonte, S. & Avalos, F. *The Rising Influence of Retail Investors*. Technical Report. BIS Quarterly Review. https://www.bis.org/publ/qtrpdf/r_qt2103v.htm (2021).
15. De Silva, T., Smith, K. & So, E. C. Losing is optional: retail option trading and earnings announcement volatility. *SSRN Electron. J.* https://doi.org/10.2139/ssrn.4050165 (2022).
16. Van Der Beck, P. & Jaunin, C. The equity market implications of the retail investment boom. *SSRN Electron. J.* https://doi.org/10.2139/ssrn.3776421 (2021).
17. González-Bailón, S., Borge-Holthoefer, J., Rivero, A. & Moreno, Y. The dynamics of protest recruitment through an online network. *Sci. Rep.* **1**, 197 (2011).
18. Lillo, F., Micciché, S., Tumminello, M., Piilo, J. & Mantegna, R. N. How news affects the trading behaviour of different categories of investors in a financial market. *Quant. Finance* **15**, 213–229 (2015).
19. Goldstein, I. Information in financial markets and its real effects. *Rev. Finance* **27**, 1–32 (2024).
20. Fan, R., Talavera, O. & Tran, V. Social media and price discovery: the case of cross-listed firms. *J. Financ. Res.* **46**, 151–167 (2023).
21. Han, B. & Yang, L. Social networks, information acquisition, and asset prices. *Manag. Sci.* **59**, 1444–1457 (2013).
22. Tetlock, P. C. Giving content to investor sentiment: the role of media in the stock market. *J. Finance* **62**, 1139–1168 (2007).
23. Bukovina, J. Social media big data and capital markets-an overview. *J. Behav. Exp. Finance* **11**, 18–26 (2016).
24. Akçay, E. & Hirshleifer, D. Social finance as cultural evolution, transmission bias, and market dynamics. *Proc. Natl Acad. Sci. USA* **118**, 2015568118 (2021).
25. RANTALA, V. How do investment ideas spread through social interaction? evidence from a ponzi scheme. *J. Finance* **74**, 2349–2389 (2019).
26. Mancini, A., Desiderio, A., Di Clemente, R. & Cimini, G. Self-induced consensus of Reddit users to characterise the GameStop short squeeze. *Sci. Rep.* **12**, 13780 (2022).
27. Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M. & Mozetič, I. The effects of Twitter sentiment on stock price returns. *PLoS ONE* **10**, 0138441 (2015).
28. Vassallo, D., Bormetti, G. & Lillo, F. A tale of two sentiment scales: disentangling short-run and long-run components in multivariate sentiment dynamics. *Quant. Finance* **22**, 2237–2255 (2022).
29. Duterme, T. Bloomberg and the GameStop saga: the fear of stock market democracy. *Econ. Soc.* https://doi.org/10.1080/03085147.2023.2189819 (2023).
30. Lucchini, L. et al. From Reddit to Wall Street: the role of committed minorities in financial collective action. *R. Soc. Open Sci.* **9**, 211488 (2022).
31. Olson, R. S. & Neal, Z. P. Navigating the massive world of reddit: using backbone networks to map user interests in social media. *PeerJ Comput. Sci.* **1**, 4 (2015).
32. Centola, D., Becker, J., Brackbill, D. & Baronchelli, A. Experimental evidence for tipping points in social convention. *Science* **360**, 1116–1119 (2018).
33. Center, P. R. The behaviors and attitudes of U.S. adults on Twitter. https://www.pewresearch.org/internet/2021/11/15/the-behaviors-and-attitudes-of-u-s-adults-on-twitter/ (2021).

34. Borge-Holthoefer, J. et al. The dynamics of information-driven coordination phenomena: a transfer entropy analysis. *Sci. Adv.* **2**, 1501158 (2016).

35. Sornette, D., Lera, S. C., Lin, J. & Wu, K. Non-normal interactions create socio-economic bubbles. *Commun. Phys.* **6**, 261 (2023).

36. Lyócsa, Š., Baumöhl, E. & Výrost, T. YOLO trading: riding with the herd during the GameStop episode. *Finance Res. Lett.* **46**, 102359 (2022).

37. Long, S. C., Lucey, B., Xie, Y. & Yarovaya, L. "I just like the stock": the role of Reddit sentiment in the GameStop share rally. *Financ. Rev.* **58**, 19–37 (2023).

38. Suchanek, M. Social interactions in short squeeze scenarios. *Int. Rev. Econ. Finance* **91**, 898–919 (2024).

39. Podobnik, B., Horvatic, D., Petersen, A. M. & Stanley, H. E. Cross-correlations between volume change and price change. *Proc. Natl Acad. Sci.* **106**, 22079–22084 (2009).

40. Mantegna, R. N. & Stanley, H. E. *Introduction to Econophysics: Correlations and Complexity in Finance* 1st edn. (Cambridge Univ. Press, 1999).

41. Sugihara, G. et al. Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).

42. Geweke, J. F. Measures of conditional linear dependence and feedback between time series. *J. Am. Stat. Assoc.* **79**, 907–915 (1984).

43. Ramos, A. & Macau, E. Minimum sample size for reliable causal inference using transfer entropy. *Entropy* **19**, 150 (2017).

44. Pei, S., Muchnik, L., Andrade, J. S., Zheng, Z. & Makse, H. A. Searching for superspreaders of information in real-world social media. *Sci. Rep.* **4**, 5547 (2014).

45. Pelaez, A., R. Winston, E. & Samuel, J. David and Goliath revisited: How small investors are changing the landscape of financial markets. In *50th Anniversary Northeast Decision Sciences Institute Annual Conference Proceedings* 287–292 (Northeast Decision Sciences Institute, 2021).

46. Lorenz-Spreen, P., Mønsted, B. M., Hövel, P. & Lehmann, S. Accelerating dynamics of collective attention. *Nat. Commun.* **10**, 1759 (2019).

47. Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. & Blackburn, J. The Pushshift Reddit dataset. *Proc. Int. AAAI Conf. Web Social Media* **14**, 830–839 (2020).

48. Smith, R. An overview of the tesseract ocr engine. In *2007 9th International Conference on Document Analysis and Recognition* 629–633 (IEEE Computer Society, 2007).

49. Podobnik, B., Jiang, Z.-Q., Zhou, W.-X. & Stanley, H. E. Statistical tests for power-law cross-correlated processes. *Phys. Rev. E* **84**, 066118 (2011).

50. Austin, P. C. & Steyerberg, E. W. The number of subjects per variable required in linear regression analyses. *J. Clin. Epidemiol.* **68**, 627–636 (2015).

51. Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. *Time Series Analysis* 1st edn. (Wiley, 2008).

Author contributions

A.D. and L.M.A. gathered the data. A.D. performed the analysis and made the figures. L.A. and G.C. designed the analysis and supervised the project. All authors discussed the results and contributed to the final manuscript.

Competing interests

The authors declare no competing interests

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44260-025-00029-z.

Correspondence and requests for materials should be addressed to Antonio Desiderio.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.