Wisdom of the Crowds or Ignorance of the Masses?

A Data-Driven Guide to WallStreetBets

Valentina Semenova, Dragos Gorduza, William Wildi, Xiaowen Dong, and Stefan Zohren

Valentina Semenova
is a PhD student in mathematics at Oxford University and a member of the Institute for New Economic Thinking and Oxford-Man Institute of Quantitative Finance in Oxford, UK.
semenova@maths.ox.ac.uk

Dragos Gorduza
is a PhD student in engineering at Oxford University and a member of the Oxford-Man Institute of Quantitative Finance in Oxford, UK.
dragos.gorduza@st-annes.ox.ac.uk

William Wildi
is an MSc student in engineering at Oxford University and a member of the Oxford-Man Institute of Quantitative Finance in Oxford, UK.
bill.wildi@oxford-dp.com

Xiaowen Dong
is an associate professor at Oxford University and a member of the Oxford-Man Institute of Quantitative Finance in Oxford, UK.
xiaowen.dong@eng.ox.ac.uk

Stefan Zohren
is an associate professor at Oxford University and a research fellow in financial machine learning at the Oxford-Man Institute of Quantitative Finance in Oxford, UK.
stefan.zohren@eng.ox.ac.uk

KEY FINDINGS

- The article uncovers text-based and user pattern-based asset clusters within the WallStreetBets (WSB) forum by applying large language models and network techniques to WSB submission and comment data.
- The authors explore the reaction of WSB users to market moves through an analysis of cumulative abnormal returns preceding and following submissions. They observe that users are generally reactive to price and that this pattern is particularly distinctive in meme stocks.
- The article quantifies the impact that WSB activity has had on the market through documenting whether posts on the forum Granger-cause asset returns and through studying the predictive signals within due diligence posts.
- In addition to the analysis, the article presents the dataset of hand-annotated due diligence posts and posts labeled with a sentiment classifier, as well as an interactive dashboard, to promote further exploration and research.

ABSTRACT

A trite yet fundamental question in economics is: What causes large asset price fluctuations? A 10-fold rise in the price of GameStop equity, between January 22, 2021, and January 28, 2021, demonstrated that herding behavior among retail investors is an important contributing factor. This article presents a data-driven guide to the forum that started the hype: WallStreetBets (WSB). The article's initial experiments decompose the forum using a large language topic model and network tools. The topic model describes the evolution of the forum over time and shows the persistence of certain topics (such as the market/S&P 500 discussion) and the sporadic interest in others, such as COVID-19 or crude oil. The authors use network analysis to decompose the landscape of retail investors into clusters based on their posting and discussion habits; several large, correlated asset discussion clusters emerge, surrounded by smaller, niche ones. A second set of experiments assesses the impact that WSB discussions have had on the market. The authors show that forum activity has a Granger causal relationship with the returns of several assets, some of which are now commonly classified as *meme stocks*, while others have gone under the radar. The article extracts a set of short-term trade signals from posts and long-term (monthly and weekly) trade signals from forum dynamics and considers their predictive power at different time horizons. In addition to the analysis, the article presents the dataset, as well as an interactive dashboard, in order to promote further research.

Social media has changed the structure of our society. As many as 4.9 billion people, or 61% of the world population, are active social media users, each just a few clicks away from the next viral phenomenon. People turn to their accounts for everything from news to product suggestions. Now, a growing audience turns to social media for promising stock market gambles. Even though investor discussion forums have existed for decades, the r/WallStreetBets (WSB) subreddit was arguably the first to reach an unprecedented retail following. Since its creation in 2012, the forum grew exponentially in membership, attracting followers not only through lucrative trade ideas but also through the promise that coordination among smaller retail traders could unseat investment giants. Their aspirations became a reality in January 2021 when the forum's cherished stock, GameStop (GME), experienced a 22-fold rise in asset price, while Melvin Capital, an investment fund with a large short position in GME, experienced a 30% decline in its value.

A single trader with $1,000 in her bank account couldn't move the markets. Seven hundred fifty thousand individuals on the other hand, each with $1,000 in their bank accounts, could have bought all of the floating shares of GameStop at its early January price of $17.25 per share. However, how did they overcome the colossal coordination challenge? What was the structure of the discussion on WSB, and how did the conversation evolve? In this article, we dive under the hood with the goal of shedding light on the dynamics of the forum that started the hype.

Even though the literature has explored specific phenomena within the WSB forum, we believe that our work is the first to take a broad approach.[1] Our goal with this article is to thoroughly characterize the trading signals and behaviors on WSB. We focus on the following questions:

- What are the primary topics of interest to WSB users, and how do they relate to each other in the forum discussion landscape?
- What assets are forum participants discussing, and can their conversations be used as indicators of market moves?
- Are WSB users trend followers or predictors of asset returns?

We answer these questions by applying several machine learning tools to the dataset of WSB submissions and comments:

1. We extract text features from the forum data, such as sentiments and topics using large language models, as well as tickers mentioned within posts.[2]
2. We estimate the relationship between assets using a network approach applied to both our extracted topics and user submissions.
3. We approximate the interaction of asset returns and user sentiment by estimating the cumulative abnormal returns in the days preceding and following submissions.
4. We study whether WSB discussions forecast market returns by estimating a Granger causal relationship between the two.

The text features within our data allow us to uncover several important characteristics. Our topic model shows how the interests of the forum have evolved over time. Certain topics, such as the discussion of Tesla and Elon Musk, retain the interest of WSB users. Others, such as oil and COVID-19, briefly grab the forum's attention but are promptly replaced by other discussion themes. Our sentiment model allows

[1] See Boylston et al. (2021); Semenova and Winkler (2021); Buz and Melo (2021); Witts, Tortosa-Ausina, and Arribas (2021); and Chacon, Morillon, and Wang (2022).

[2] See Araci (2019) and Angelov (2020).

us to extract a perspective taken by users (bullish, bearish, or neutral) about the future movement of the asset mentioned in a submission. This sentiment measure, in turn, allows us to understand the predictive power of the forum and user reactions to market moves.

The network approach allows us to cluster assets based on (1) whether they are frequently mentioned together while discussing the same topic (Topic Network) and (2) whether the same group of users is interested in both assets (Submission Network). We observe that distinct clusters of similar assets emerge. Posts within the large asset clusters generally tend to incorrectly forecast future asset prices. However, smaller, niche clusters have better market prediction performance, perhaps indicating that close-knit groups of well-informed retail investors may have market insights.

We analyze the seven-day cumulative abnormal returns (CAR) in the 14 days preceding and following a post about an asset. The CAR is the asset log return, which is distinct from simultaneous market returns, and is typically used to model sudden price shocks and event detection in financial markets. We observe that, overall, WSB appears to be reactive to, rather than predictive of, CAR in an asset. The average CAR appears to run up gradually immediately before a post and subsequently experiences a sharp decline. The pattern is particularly distinctive in *meme stocks*, in which high asset returns potentially reinforce the hype. In stocks with a broader following, such as Microsoft or Apple, CAR appears flat around WSB posts.

In a final exercise, we study a Granger causal relationship between asset returns and sentiments in posts. A time series A is said to Granger-cause B if it can be shown that the lagged values of A provide statistically significant information about future values of B. Our analysis highlights that a Granger causal relationship exists between the sentiments on WSB of many *meme* assets and their returns—the result is not surprising because conventional wisdom is that the WSB discussions heavily influenced *meme* stocks. The Granger causal analysis allows us to look at the comparative degree of influence between different tickers and to identify new assets that may have been influenced by WSB but have gone under the radar, such as Snapchat.

Beyond the immediate quantitative insights, an additional goal of the article is to provide tools and data for researchers and practitioners. We hope that this research will contribute to a broader effort within the financial industry to take advantage of and evaluate trade-offs in nonstationary and quickly evolving signals. We make the dataset used within the article and manual annotations, as well as code database, available upon request (complying with the data provider's dissemination policies). Furthermore, in an effort to encourage a more interactive data experience for practitioners, we publish our insights in a dashboard, allowing users to explore various aspects of the dataset and evaluate opportunities for market returns.[3]

RELEVANT LITERATURE

We believe that this article may be of interest to those looking into social media in finance and the WSB forum specifically, as well as to those studying new, noisy signals within the finance community.

Our article takes inspiration from an emerging literature looking at the relationship between social media activity and stock market prices. It is well understood that new information about a company (such as from a news report) can affect its stock price, as well as the stock price of similar or competing companies.[4] Social media

[3] https://sites.google.com/view/wsbtrialsite.
[4] See Wan et al. (2021).

data provide information beyond news: shedding light on how individuals engage with financial information, which can reflect the sentiment of the market more broadly. Many researchers have exploited this to find associations between publicly expressed sentiment on Twitter/StockTwits and stock market prices.[5] Social media has thereby provided rich new text data for quantitative finance researchers to exploit and is now a key input (along with news articles and company reports) for text analysis in finance.[6] We contribute to this literature both through our analysis of a specific social media dataset, the WSB forum, and through providing annotated data for future research. A more targeted set of papers focuses on specific characteristics of the forum. We add to this literature by taking a broader approach to analyzing the forum.

Data

Data collection. Submission and comment data from WSB are collected through the Pushshift Reddit database, which provides an archive of all posts and comments on Reddit.[7] We collect data from the inception of the forum in April 2012 up to June 24, 2022. Posts on WSB typically refer to companies by their stock symbol, and we use the work of Semenova and Winkler (2021) to obtain a list of stock symbols (from Yahoo Finance and Compustat) and extract these from the submission text. Stock price data are sourced from AlphaVantage.

Due diligence posts. Due diligence (DD) posts have attracted attention in the literature as higher-quality posts from WSB, which contain meaningful stock-level insights. To obtain the subset of DD reports in the forum, we first filter to posts that have the DD flair. We supplement the posts containing the DD flair with posts that contain the DD acronym. We manually review every single potential DD post (i.e., flaired posts that are not removed by moderators or contain the text DD), and we remove any post that does not appear to be a valid DD—that is, does not contain market insight. The DD posts are sufficiently small in number that we can manually label each post as bullish/positive or bearish/negative (neutral posts are removed). We find 77% of posts are bullish, and 23% are bearish.

Sentiment classification. For the forum as a whole there are far too many posts for manual classification. Previous papers have either relied on counting keywords, or they have used off-the-shelf rule-based sentiment classifiers, resulting in certain misclassification issues, discussed in our supplementary materials posted to our dashboard site.[8] To resolve these issues, we first hand code a random sample of 4,000 WSB posts, categorizing them as bullish, bearish, or neutral, depending on what the author's attitude is on future price increases. Of the posts in the sample 41% are bullish, 37% are neutral, and 22% are bearish.[9] We then take a pretrained bidirectional encoder representations from transformers (BERT) model called a language model based on BERT for financial natural language processing tasks (FinBERT), which we fine-tune using our manually labeled sample.[10] Our final model had a test accuracy of 69.2%, which we consider sufficient due to the challenging task of predicting future-facing sentiment. Further details on data sourcing, cleaning, and annotation are presented in supplementary materials posted to our dashboard site.[11]

[5] See Azar and Lo (2016); Pagolu et al. (2016); Sul, Dennis, and Yuan (2017); Agrawal et al. (2018); and Duz Tan and Tas (2021).

[6] See Gao et al. (2021) for a review.

[7] See Baumgartner et al. (2020).

[8] https://sites.google.com/view/wsbtrialsite.

[9] Because bearish posts tend to be a minority on the forum, bearish posts were up-sampled to reduce class imbalance.

[10] See Devlin et al. (2019) and Araci (2019).

[11] https://sites.google.com/view/wsbtrialsite.

WSB FORUM DYNAMICS

In this section, we characterize the discussion landscape of the WSB forum. We consider how assets are related on WSB in terms of topic and user interest. Additionally, we explore the relationship between WSB and markets through studying the returns preceding and following posts. The section that follows discusses predicting market movements using data from WSB.

The Topic Landscape of WSB

Topic models are often employed to map out the distribution of information in a textual corpus.[12] In economics and finance, they are used to uncover the latent discursive directions that individuals can follow when expressing an opinion, which allows researchers to quickly detect salient points in the discourse and use them as signals for downstream analysis.[13] However, prior work leverages older topic models that do not make full use of the new modeling capacities offered by large language models (LLMs).[14] We use the BERT LLM Topic package to perform topic modeling on WSB, which accounts for nonlinear semantic similarities between texts that go beyond word-level co-occurrences (a limitation of traditional topic models).[15] This model gives us a mapping from each post in the sample to a learned set of topics. Each topic has a unique set of representative terms associated with it, and posts are mapped to topics based on how well the text within the post matches that of a particular topic.

Exhibit 1 highlights our key findings from employing the BERT topic model: Exhibit 1, Panel A presents some of the key topics of discussion on WSB, while Exhibit 1, Panel B displays their prevalence on the forum over time. Each post within WSB is given a probability vector whose length is equal to the number of extracted topics, identifying the extent to which the post is associated with different topics. For example, a post that discusses pharmaceutical companies and how they will profit from the COVID-19 vaccine will be given a higher probability of being associated with topics 35 and 36.

Exhibit 1, Panel B is constructed by first splitting our dataset into months and then calculating the fraction of the discourse dedicated to that topic within that month for the 10 most popular topics. The span of topics covered here is a subset of the overall topics identified by the model, which serve to illustrate the key areas of attention for the forum across time. We note that several topics and their prevalence follow news, such as coronavirus (topic 35), which appears in March 2021, and vaccine (topic 36), which are more prevalent during the introductions of the first COVID-19 vaccines in end of the year 2021. The topic model also tracks the emerging focus of the forum on so called *meme stocks* such as Tesla (topic 0), Palantir (topic 3), and Blackberry (topic 16). Contrary to other transient topics, we detect a persistent interest of the forum in SPY (S&P 500) as noted by the prevalence of topic 1, which represents attention toward the overall state of the market.

How Are Assets Related to Each Other on WSB?

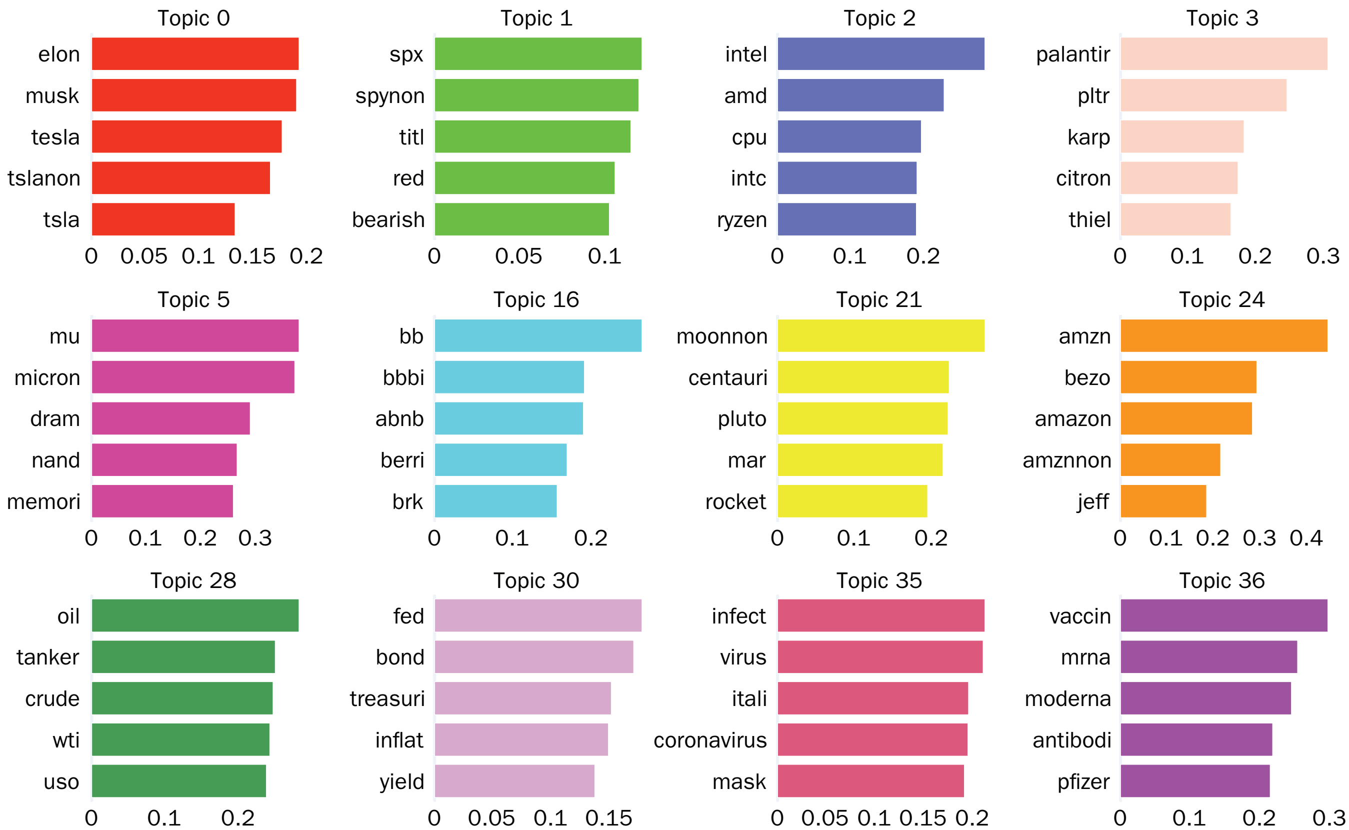
The WSB forum is uniquely suited to study the relationships between assets, as they are perceived and discussed together by retail investors. We choose two methods to map the relationships between assets and to create a ticker-to-ticker

---

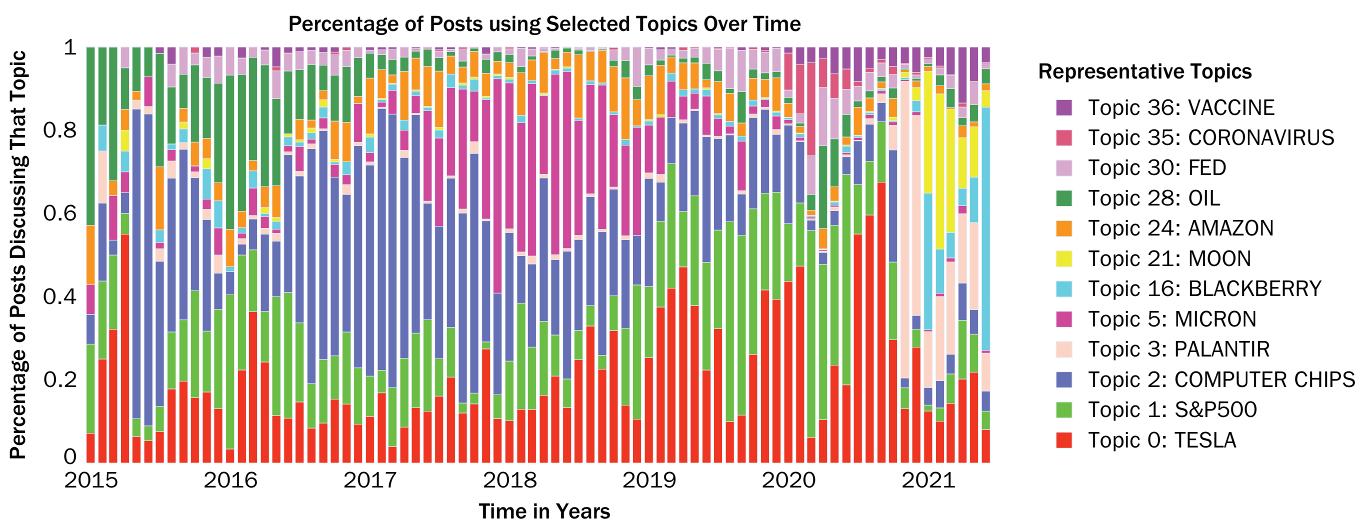
[12] See Blei, Ng, and Jordan (2003) and Angelov (2020).
[13] See Gentzkow, Kelly, and Taddy (2019).
[14] See Schou et al. (2022).
[15] See Angelov (2020).

EXHIBIT 1
The WSB Discussion

Panel A: Sample of Representative Discussions Found on WSB



Panel B: Topic Distribution over Time



NOTES: Sample of Representative Topics Discussed on WSB; this exhibit displays topics and the frequency with which the most popular words appear within the topics (ranked by the categorical term frequency—inverse document frequency score, described in Angelov 2020). We present several representative topics from WSB in Panel A, as well as their relative importance within the forum in Panel B.

network structure: the Topic Network approach and the Submission Network approach. The Topic Network uses the frequency with which assets are discussed within the same topics to create "inter-asset" connections, indicating that investors perceive the assets to be influenced by the same general financial trends. The Submission Network tracks which groups of investors are interested in, and create submissions about, the same assets. We describe both approaches below.

Topic Network. In the Topic Network approach, we link assets based on the frequency with which they are mentioned in the same set of topics. The intuition behind the approach rests on the idea that assets that are discussed within the same financial topics are likely related to each other through some underlying fundamental relationship.

We explain our intuition through a simplified example. Let us consider two assets: (1) interest rate swaps and (2) bonds. We extract a topic from our topic model about the federal reserve (Fed), and we observe that both bonds and interest rate swaps are frequently brought up in posts that are labeled with the Fed topic. This is unsurprising, given the fact that Fed decisions would strongly affect the valuations of both bonds and interest rate swaps. In our Topic Network exercise, we would place a connection between bonds and interest rate swaps because they are linked by the Fed topic. This intuition can be extended to better understand the meaning behind the tickers linked through our topic model: They are tickers that are frequently brought up under the same topics, indicating that they are linked through some underlying economic discussion theme.

Exhibit 2, Panel A provides a toy example. TSLA and SPY are connected because they are both mentioned in the topic discussing the overall stability of the market (market topic); SPY and APPL are connected because they are both discussed in conjunction with the Fed topic.

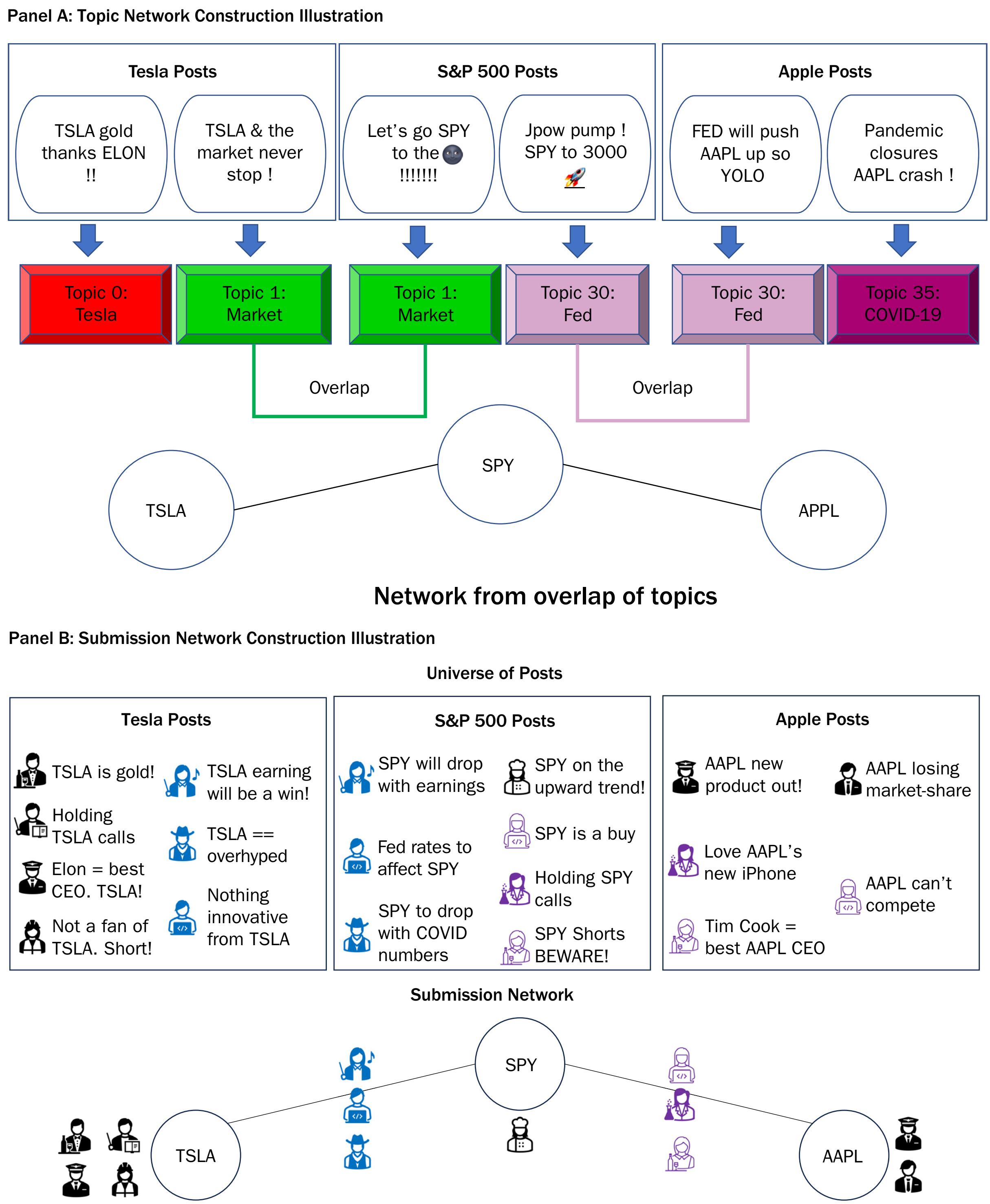
In practice, we also add a thresholding step when constructing links between tickers based on overlapping topics. We select the top 50 topics, ranked by frequency of appearance in posts. This means that we only count a topic if it is represented often enough across our dataset. We then count all the topics that correspond to a given ticker and we keep the topics that are present in 20% or more posts related to that ticker—this is done to filter out less important conversations about assets. The filtering leads us to drop tickers whose topic distribution is very diffuse, meaning that discussions about it do not have consistency. As a prime example, this leads us to drop GME from our topic network because the topic distribution in GME posts is too dispersed.

Submission Network. The community structure within WSB offers a different perspective on how assets are related within the forum. To construct our submission network, we look at the overlap of users who create posts about two assets. Two assets are linked if there is a sufficiently large fraction of users discussing both assets simultaneously.

Exhibit 2, Panel B provides a simplified example. We observe a subset of users creating posts about TSLA and SPY simultaneously and a different group of users writing about AAPL and SPY. However, we do not observe the same users creating posts about TSLA and AAPL. Therefore, we create two links: TSLA/SPY and AAPL/SPY.

In practice, our network construction exercise is slightly more complicated. We create weighted links between assets, normalized by the total number of users mentioning each asset. This means that if asset A is discussed by 10 users and all of them also post about asset B, we would create a link with weight one from asset A to B. However, if 10,000 users post about asset B, we would not create a link back from B to A because only a negligible fraction of posters about B care about A. Therefore, the total weight of the link between A and B would be 1 from the first created link. In practice, our threshold for including a link in our weight calculation

EXHIBIT 2
Network Construction



NOTE: We demonstrate how links are identified between assets in the Topic Network and Submission Network.

is 20%—at least 20% of users posting about one ticker must be posting about the other ticker. Relating this back to Exhibit 2, Panel B, we observe that three out of seven SPY posters also post about AAPL (weight 3/7) and that three out of five AAPL posters post about SPY (weight 3/5)—the total weight for the link between AAPL and SPY would, therefore, be 3/5 + 3/7. We filter our submission network to contain tickers that are mentioned at least 150 times on the forum.

Asset Network Structures

Topic Network: Results. Exhibit 3, Panel A presents our Topic Network—links indicate which two tickers are likely to be mentioned together within the same economic discussion topics. A list of tickers and their associated clusters is presented in Exhibit 4. We observe that larger and frequently discussed companies, such as AMC, have more cohesive topics that are not referenced in many other companies' posts. Therefore, the posts of several tickers are isolated from the main connected component of the topic network. These include: AAPL (cluster 1), FB (cluster 6), AMC (cluster 9), and BABA (cluster 16). We interpret the isolation of these network components as the result of a systematic use of a limited and distinct set of topics when discussing these stocks. A post about AAPL will mobilize only topics similar to other AAPL discussions: focusing the attention of readers more narrowly on discussion points unique to Apple, Inc. as opposed to the wider market. Contrary to that, smaller companies tend to have a more central position in this network and are often connected to stocks with higher market caps. This is visible by the more spread-out cluster 19, which encompasses a variety of both smaller and larger companies as well as exchange-traded fund (ETFs), likely connected through broader discussions about the economic climate.

Additionally, we find smaller clusters of assets related to a narrow set of economic themes within the market. A prime example of this is cluster 15, including INO, OCGN, and PFE. These firms are likely linked by the same topics of drug discovery, US Food and Drug Administration approvals, and the COVID-19 vaccine, in a way that is distinct to this particular group of assets.

We use our Topic Network as a distance mapping between firms that takes into account their discursive environments. We interpret tickers belonging to the same cluster as an indication of greater similarity within their discussion topics. Consequently, belonging to the same cluster could be an indicator of a more correlated information set driving returns for all assets within the cluster.

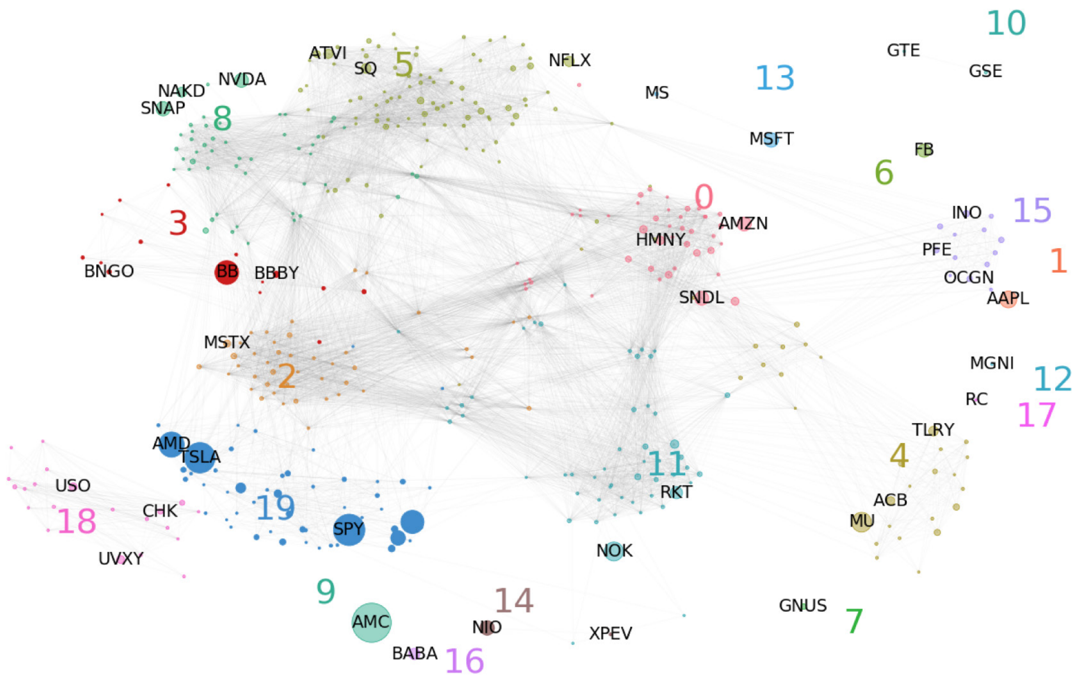
Submission Network: Results. Exhibit 3, Panel B presents the results of a clustering exercise on the Submission Network. We observe that similar assets appear to be mentioned within the detected clusters—implying that individuals self-select into discussions about certain asset categories. Said differently, a distinct group of investors is interested in marijuana stocks to those discussing the cruise line industry. This is perhaps most pronounced in the small cluster of gold ETFs containing JNUG, NUGT, and JDST—cluster 8. Several other smaller clusters, dominated by a niche sector, are also visible in clusters 0, 3, 8, and 9. We observe a large cluster around the SPY ETF and TSLA (cluster 4), as well as a large *meme* stock cluster around GME (cluster 5). We observe a distinct cluster with some large, popular tech stocks including AAPL, FB, MU, AMZN, and AMD (cluster 2). In our supplementary materials, we explore the correlation among assets in different clusters, as well as returns to WSB advice within the different clusters.

Returns and correlations across clusters. In our Submission Network, we observe that distinct groups of investors are interested in different types of tickers: Some users come to WSB in order to discuss pharmaceutical companies, whereas others are interested in trading natural resource ETFs or airline stocks. For this reason, tickers clustered through the Submission Network tend to have returns that are

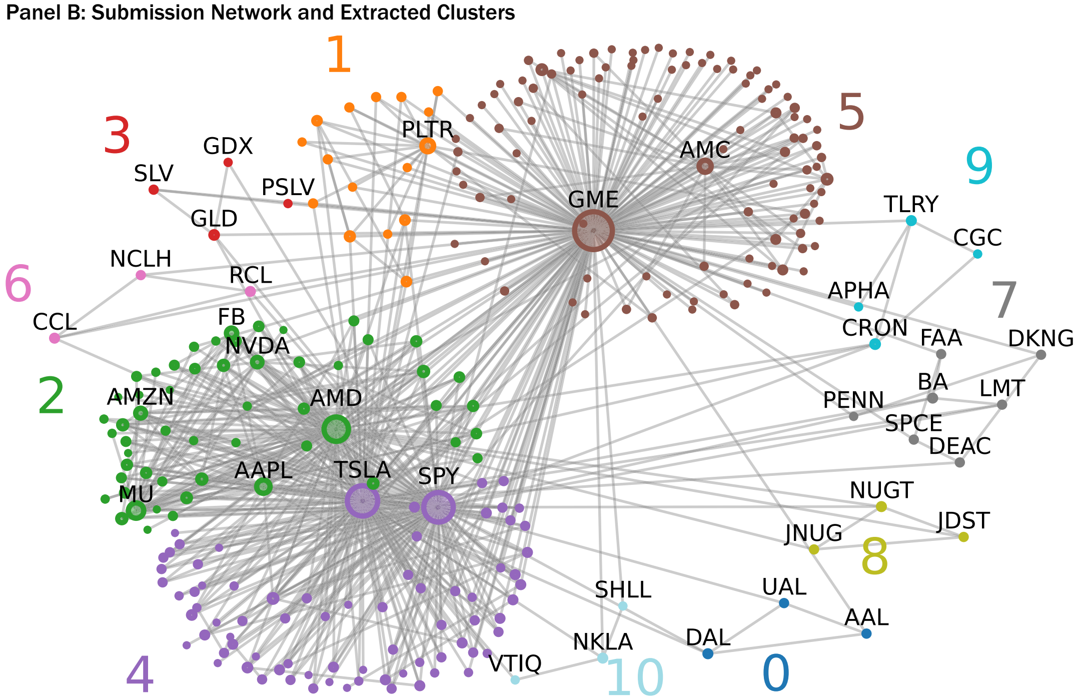
EXHIBIT 3

Extracted Network from WSB

Panel A: Topic Network and Extracted Clusters



Panel B: Submission Network and Extracted Clusters



NOTES: We illustrate the ticker networks extracted using both the Topic Network and Submission Network approaches. Clusters of closely connected tickers are extracted using the Leiden algorithm from Traag, Waltman, and Van Eck (2019). Tickers within the same cluster are shaded with the same color. In Panel A, the size of ticker nodes is determined by the number of posts mentioning the ticker. In Panel B, nodes are scaled by the number of connections they have.

EXHIBIT 4

Subset of Clusters Extracted from the Topic Network

Cluster	Popular Tickers within Cluster	Posts	Total No. of Tickers
0	AMZN, SNDL, HMNY, MVIS, WISH, CLNE, CCL, DKNG, SUNE	12,373	41
1	AAPL	2,530	1
2	MSTX, AVXL, MNKD, OPK, AMDA, AUPH, SENS, AMRN, ADMP	3,312	44
3	BB, BBBY, BNGO, BRK, BBY, DB, BLNK, ABNB, BBW, BTT	9,898	12
4	MU, TLRY, ACB, CGC, APHA, CRON, AMAT, MO, OGI, HEXO	6,557	17
5	NFLX, SQ, ATVI, PLUG, WMT, SBUX, LULU, CRM, CMG, MCD	10,045	72
6	FB	1,580	1
8	SNAP, NVDA, NAKD, NAK, FSLY, TTD, DBX, KR, TWLO, GNC	5,763	35
9	AMC	19,686	1
11	NOK, RKT, WKHS, ZOM, QS, SRNE, TRCH, ASO, LFIN	11,349	33
13	MSFT, MS	1,849	2
14	NIO, XPEV	2,160	2
15	INO, OCGN, PFE, NVAX, MRNA, CVS, JNJ, AZN, VXRT, TEVA	2,474	13
16	BABA	1,175	1
18	USO, UVXY, CHK, WTI, XOM, SVXY, NAT, RIG, UCO, BP	1,001	12
19	SPY, TSLA, AMD, PLTR, SPCE, NKLA, QQQ, PRPL, GM, ROKU	28,438	30

highly correlated—clustered assets appear to have an average return correlation of greater than 0.2, whereas assets across our entire dataset appear to exhibit an average correlation of 0.16. The most highly correlated asset returns appear to be within clusters 0, 6, and 9: The returns of the assets within these clusters exhibit correlations of 0.80, 0.84, and 0.60, respectively. The average next-day returns of investing according to the extracted sentiments within submissions is statistically significant and negative across most clusters. The most negative average returns are exhibited within clusters 0 and 5; average returns of submissions in cluster 7, on the other hand, are statistically significant and positive. This demonstrates that most investors on WSB lack insights into future market moves. Notably, investors into the hype cluster tend to lose the most. Niche groups of investors may have worthwhile market insights, however, as demonstrated by returns for cluster 7. Our supplementary materials present returns and correlations across clusters.

A similar pattern can be observed from the Topic Network clusters. Investing according to the sentiments of submissions within most topic clusters results in a statistically significant, negative average next day return. However, similarly to our observations from the Submission Network, certain topic clusters exhibit positive daily returns, on average: cluster 1 containing AAPL, 13 containing MSFT and MS, 14 containing NIO and XPEV (two electric car makers), and 16 containing BABA. We notice that smaller clusters with fewer assets perform better than much larger clusters, again demonstrating that the forum as a whole may lack market insights, however, specific topics and groups of investors may have promising insights.

Returns Preceding and Following Posts

What are the characteristics of assets returns before and after they are mentioned on WSB? Is there any evidence that WSB users can predict returns, or are they trend followers just like other retail investors?

Framework. We consider how asset prices are changing shortly before and after posts on the forum. To do this, we follow the methodology of Wan et al. (2021), which analyzes market movements around news events by looking at changes in abnormal return (AR). AR is derived from the capital asset pricing model (CAPM),[16] which describes returns for company *i* at time *t* as follows:

$$r_{i,t} = \alpha_t + \beta_t r_{m,t} + \epsilon_{i,t} \tag{1}$$

Here, $r_{i,t}$ is the log return in the price of stock *i* on day *t* compared with the previous day, where $r_{i,t} = \log\left(\frac{p_{i,t}}{p_{i,t-1}}\right)$ and $p_{i,t}$ is the adjusted close stock price. $r_{m,t}$ is the return of the market (in our case, we use the S&P 500), hence β captures stock price moves that are driven by movements in the wider market. $\alpha_{i,t}$ captures stock over/underperformance relative to the market. $\epsilon_{i,t}$ is a stochastic error term, often referred to as abnormal return (AR).

AR tends to have high magnitude in the presence of sudden price shocks (for example, a news story about a particular company), and hence it is often used for event detection in financial markets. In our case, it is useful for assessing if WSB sentiment is able to provide indication about a future price shock. We fit the CAPM model to each stock and day in our data, using a moving 180-day window. Following Wan et al. (2021), we calculate the seven-day CAR in stock *i* preceding day *t*:

$$CAR_{i,t} = \sum_{t-6}^{t} \epsilon_{i,t} \tag{2}$$

Once we round the timestamp of a WSB submission to the nearest future market close time, we can match the CAR for each company/time to the WSB data, along with a time series of how the CAR changed over the 14 trading days preceding and following a WSB post.

CAR results. Exhibit 5 shows how the average seven-day CAR changes 14 trading days before and after a submission. Exhibit 5, Panel A shows the CAR plots for all WSB posts, while Exhibit 5, Panel B excludes GME and AMC because the short squeeze events on these stocks created exceptionally high abnormal returns. We observe the following: First, CAR tends to be rising up to the submission date and rapidly declines afterward. This suggests that WSB activity tends to follow significant price changes in the market, rather than providing a leading indicator of price changes. Second, the shapes of the curves are quite similar, regardless of sentiment breakdown—it is primarily the magnitude of the CAR that differs. This suggests that high CAR is associated with more posts of any sentiment, although the effect is more pronounced for bullish posts.

One might attempt to infer from the sharp fall in CAR after posts that shorting stocks that are popular on WSB might be a profitable trading strategy. This inference is, however, mistaken for two reasons. First, negative CAR does not necessarily imply negative returns. Second, and more importantly, the CAR plots suggest that WSB sentiment peaks when CAR peaks; however, predicting when sentiment will peak is challenging. If sentiment is already high and AR increases, then sentiment is likely to continue to increase. In our supplementary materials posted to our dashboard site, we consider a breakdown of the CAR plots for the 20 most popular stocks on WSB and show that the pattern presented in Exhibit 5, Panel A is most prevalent for *meme* tickers, such as GME, AMC, and BB.[17] In stocks that have a broad following outside of the WSB forum, such as AAPL or MSFT, the CAR appears flat before and after submissions.

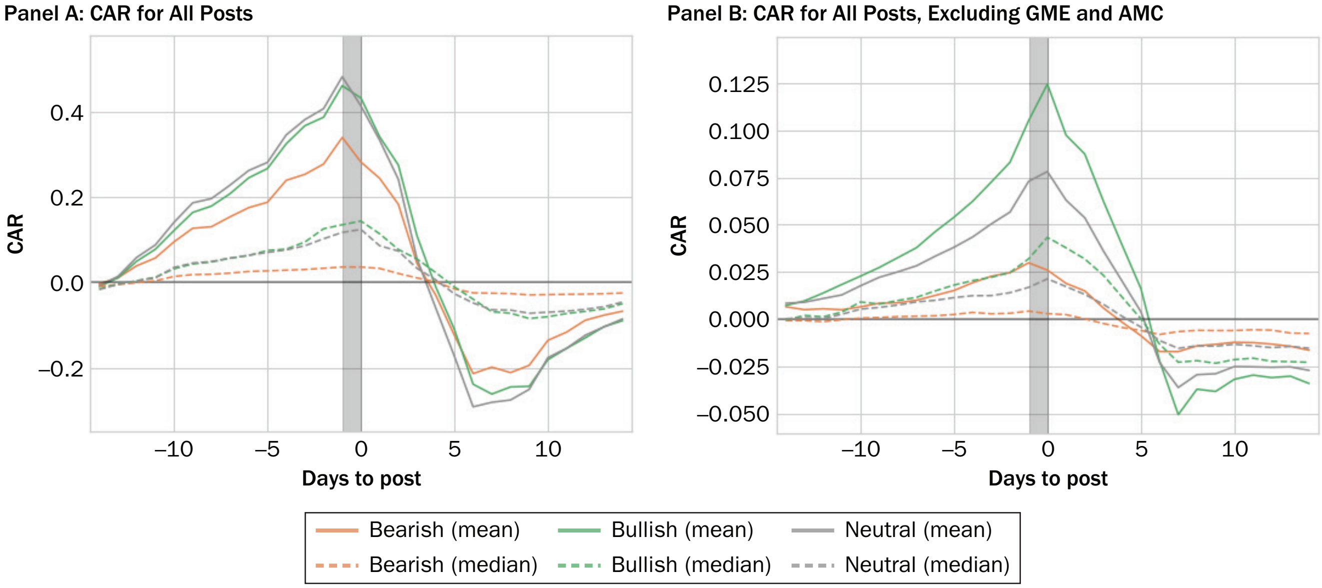
---

[16] See Fama and French (2004).

[17] https://sites.google.com/view/wsbtrialsite.

EXHIBIT 5

Average Seven-Day Cumulative Abnormal Return 14 Trading Days before and after Post Submission, Grouped by Post Sentiment



NOTES: The black bar shows the window of time when posts were written. Panel A uses all posts in the data, whereas Panel B excludes GME and AMC.

Is Information on WSB Predictive of Returns?

Initial exploration. We look at the returns of several portfolios constructed from WSB data. We consider the average next-day log returns for investing into all WSB posts mentioning a single ticker according to our sentiment classifier, as well as the next day log-returns for investing into the flaired DD and labeled DD posts. The results are presented in the top half of Exhibit 6. We observe that investing in all submissions on WSB results in consistently losing money, with an average next-day log return of –0.0282 and a P-value indicating that the returns are statistically significantly different from zero. The DD posts, on the other hand, which are specifically selected by moderators, and our hand-labeled posts both appear to have positive returns; however, the signal is weak, perhaps indicating that DD content must be further analyzed for quality (a description of DD posts is available in the "Data" section). We perform a similar exercise after splitting the data by year and observe that the forum receives negative or nearly zero returns across all years. In the year 2021, when the infamous GameStop short squeeze occurred, the average next-day log return for WSB posts was around –0.044, perhaps indicating that even though some forum participants made money on the incident, the subsequent hype and low-quality discussions drove a losing strategy for the forum as a whole.

We construct several control portfolios of interest, presented in the bottom half of Exhibit 6. First, we look at the average, daily returns of the stocks discussed on WSB across all time—the returns appear to be very close to zero, on average, but with very heavy tails (as indicated by the kurtosis) and a positive skew. The heavy tails (high kurtosis measure) indicate that the stocks mentioned within WSB frequently experience returns that are far from the mean—extreme return values. The high positive skew, on the other hand, implies that the stocks experience more extremely high returns than low returns. This is consistent with earlier observations that WSB users are swayed by large cumulative abnormal returns of stocks; however, as noted

EXHIBIT 6
Distribution of Log Returns

	μ	σ	Skew	Kurtosis	P-Value	No. of Posts
WSB Returns						
All Submissions	–0.0282	0.28	0.10	6	0.00	207,155
Flaired DD	0.0033	0.16	0.09	10	0.26	2,909
Labeled DD	0.0037	0.14	3.77	46	0.22	2,220
Control Portfolio Returns						
Stock Returns	0.0000	0.04	4.80	1,903	0.90	10,572,619
Previous	0.0242	0.34	0.50	4	0.00	207,155
Random	0.0003	0.05	1.92	109	0.01	206,631

NOTES: We present some summary statistics for the next-day log returns for a portfolio invested according to the sentiment of all submissions on WSB (all submissions), flaired DD posts (flaired DD), and hand-labeled DD posts (labeled DD). We compare this to the daily log returns of all stocks mentioned on WSB (stock returns), a randomly selected sample of stocks returns (random), and the log returns on the day submissions are made (previous). The exhibit presents several summary statistics, as well as the P-value for the mean of the distribution of the daily log returns to be equal to zero (null hypothesis).

in the previous section, the users tend to be trend followers and the forum sentiment peaks when the asset experiences a price reversion.

We consider the returns of assets on the trading day that a submission is made on WSB (or on the day before if a submission is made outside of trading hours)—labeled as "previous" in Exhibit 6. The high, average positive log returns indicate that the WSB forum is likely reactive to news and is consistent with the earlier CAR analysis. We also construct a randomly selected portfolio, where we invest in a given stock proportionally to the number of times it is mentioned in WSB in non-neutral posts but select days at random. The results indicate a very small, positive return. We anticipate this to be driven by the large returns preceding WSB posts, rather than an inherent ability of WSB users to choose lucrative assets. The supplementary materials consider returns of portfolios built from the clusters within our Topic Network and Submission Network.[18]

Granger Causal Relationships on WSB

We look at whether sentiment is useful for forecasting future returns by conducting a Granger causality test.[19] A time series A is said to Granger-cause B if it can be shown that the lagged values of A provide statistically significant information about future values of B, even when lagged values of B are also included in the forecasting exercise. In other words, time series A must provide additional information about the future values of time series B, beyond B's autoregressive terms.

Our goal is to test whether the sentiments expressed about an asset on WSB have a Granger causal relationship with the future returns of that asset. Our result is displayed as the Wald test statistic, which allows us to estimate a P-value—our confidence that WSB sentiments about an asset are related to the asset's future returns.

Sentiment time series. To get a time series of sentiment, we first need a sentiment score for each company and date. Let $s_{i,t}^{(j)}$ represent the net sentiment of post j, which is calculated by taking the (softmax) output of our sentiment classifier and calculating the difference between the bullish and bearish outputs (hence $-1 \le s_{i,t}^{(j)} \le 1$).

[18] https://sites.google.com/view/wsbtrialsite.
[19] See Granger (1969).

We can then say the net sentiment of each stock *i* on day *t*, denoted $S_{i,t}$, is the sum of $s_{i,t}^{(j)}$ for all posts for that company and day. $S_{i,t}$, however, has a tendency to increase over time because the forum has grown in popularity, and most posts are bullish. To resolve this, we normalize $S_{i,t}$ by dividing it by the average number of posts on the forum over the past seven days (n_t)—we denote the normalized sentiment $\hat{S}_{i,t}$. We then calculate the difference in normalized sentiment to get a value for our sentiment time series $\Delta\hat{S}_{i,t}$, where

$$\Delta\hat{S}_{i,t} = \hat{S}_{i,t} - \hat{S}_{i,t-1} = \frac{S_{i,t}}{n_t} - \frac{S_{i,t-1}}{n_{t-1}}. \tag{3}$$

Granger causality results. We check whether our normalized sentiment measure $\Delta\hat{S}_{i,t}$ is Granger causal of daily stock price returns. We assess whether lagged values of WSB sentiment are useful for forecasting future returns by observing the P-value in our Granger causality test in Exhibit 7. WSB sentiments are lagged between 1 and 10 days with stock return. We use time series for the top 22 most popular stocks on the forum,[20] starting from either of January 1, 2016, or the first time a stock was mentioned on the forum. The removal of pre-2016 data is motivated by the small size of the forum before this point, which can result in volatile changes in our sentiment measure.

For all 22 stocks, our sentiment measure and the daily log return time series appear stationary, which we check with an augmented Dickey–Fuller test. We run Granger causality tests at lags of 1, 2, 5, and 10 trading days. The results are shown in Exhibit 7. For most stocks, WSB sentiment is not useful for forecasting returns. This is not true, however, for the *meme* stocks GME, AMC, NOK, and BB, but this fits with the conventional wisdom that retail traders on WSB drove up the prices of these shares in early 2021.

We do get statistically significant results ($0.01 < p < 0.05$) for AMZN at a 10-day lag, PLTR at a 10-day lag, SNDL at a 2-day lag, and SPCE at a 1- and 10-day lag. These results are unlikely to be meaningful, however—given we are running a total of 88 tests, we would expect four or five results in this range. Given that sentiment is not Granger causal of returns at other lags for these stocks, this is most likely to be a spurious result.

The most interesting statistically significant results come from SNAP, where the test statistic is strongly significant at one and two days ($p < 0.001$), statistically significant at five days ($p < 0.05$), and almost significant at 10 days ($p = 0.055$). This is also consistent with the individual stock CAR results in our supplementary materials.[21] SNAP is not conventionally considered a *meme* stock, and it was considerably more discussed prior to 2021.

Additional Results

In an additional analysis, we attempt to extract long-term and short-term metrics from WSB, which we consider as potential trade signals and test their predictive power. For our short-term analysis, we focus on extracting signals from our hand-annotated sample of DD posts. We fit 3,744 models predicting returns at time horizons varying from next day to 24 weeks into the future using various DD post properties. Our models provide some indication of the fact that bullish sentiment in DD posts is predictive of returns at various time horizons. Among other variables that we considered, the existence of a URL (indicative of external references outside of

[20] For consistency with our CAR plots, we include GE and DIS, in addition to RKT and PLTR.
[21] https://sites.google.com/view/wsbtrialsite.

EXHIBIT 7

Results of Granger Causality Tests

	Obs.	1	2	5	10
AAPL	1,338	0.140	0.047	0.439	1.177
		(0.708)	(0.954)	(0.821)	(0.302)
AMC	1,184	9.833	36.189	21.110	10.551
		(0.002)**	(0.000)***	(0.000)***	(0.000)***
AMD	1,336	0.801	2.182	1.446	1.147
		(0.371)	(0.113)	(0.205)	(0.323)
AMZN	1,334	0.224	0.183	1.735	2.033
		(0.636)	(0.833)	(0.123)	(0.027)*
BABA	1,339	1.055	1.329	1.260	0.887
		(0.305)	(0.265)	(0.279)	(0.545)
BB	1,212	31.792	14.402	16.236	12.083
		(0.000)***	(0.000)***	(0.000)***	(0.000)***
DIS	1,326	0.410	0.760	0.731	0.950
		(0.522)	(0.468)	(0.6)	(0.486)
FB	1,338	0.303	0.147	0.620	0.815
		(0.582)	(0.863)	(0.685)	(0.614)
GE	1,301	0.329	0.901	1.118	1.037
		(0.566)	(0.406)	(0.349)	(0.409)
GME	1,284	28.820	16.922	16.206	16.749
		(0.000)***	(0.000)***	(0.000)***	(0.000)***
MSFT	1,325	2.313	1.176	0.928	0.816
		(0.129)	(0.309)	(0.462)	(0.614)
MU	1,295	0.120	0.128	0.098	0.920
		(0.729)	(0.88)	(0.992)	(0.514)
MVIS	261	0.306	1.658	1.322	0.632
		(0.58)	(0.192)	(0.255)	(0.786)
NIO	637	0.169	0.645	0.240	0.704
		(0.682)	(0.525)	(0.945)	(0.721)
NOK	1,319	10.168	5.884	3.476	2.003
		(0.001)***	(0.003)**	(0.004)**	(0.03)*
NVDA	1,336	0.093	0.168	0.494	0.617
		(0.761)	(0.845)	(0.781)	(0.801)
PLTR	130	0.202	0.839	1.088	2.065
		(0.654)	(0.434)	(0.371)	(0.034)*
RKT	167	0.168	0.529	0.283	0.287
		(0.682)	(0.59)	(0.922)	(0.983)
SNAP	1,022	13.782	6.907	2.815	1.811
		(0.000)***	(0.001)***	(0.016)*	(0.055)
SNDL	327	0.312	3.500	1.810	1.212
		(0.577)	(0.031)*	(0.111)	(0.283)
SPCE	391	3.901	1.169	1.298	2.017
		(0.049)*	(0.312)	(0.264)	(0.031)*
TSLA	1,334	0.293	0.369	0.627	0.393
		(0.588)	(0.691)	(0.679)	(0.95)

NOTES: Granger causality tests for log returns for the top 22 most popular stocks on WSB. The main numbers show the Wald test statistic, with P-values in parentheses. ***Significant at 0.1% level; **significant at 1% level; *significant at 5% level.

the forum) is more likely to have a positive effect for models looking at time horizons greater than eight weeks, and especially at time horizons greater than 15 weeks. For our long-term analysis, we consider monthly and weekly changes on the WSB forum versus returns, trading volumes, and volatility; we observe that changes in normalized sentiment in an asset in a given month are negatively correlated with returns in the following month. Our full results are presented in the supplementary materials.[22] Overall, our findings suggest that WSB metrics do not have strong, predictive power for market movements but can be used in conjunction with other factors as potential weak signals in portfolio construction.

CONCLUSION

What went on under the hood of the most infamous investor forum, WSB? This article takes a data-driven look. We break down the discussion in several different ways: through topic modeling, by sentiment analysis, and through a network perspective, used to map the relationships between assets. Our network approach allows us to extract groups, or clusters, of assets that possess similar characteristics on the WSB forum. We conclude that returns for assets within large clusters (those containing many assets) are generally predicted poorly by the forum. However, niche groups of investors may have potential insights into the markets, as indicated by positive returns to WSB submissions within smaller asset clusters.

Our analysis of CAR indicates that WSB investors are, on average, reactive to market news—they follow the hype and express bullish sentiments precisely at the time when assets reach their peak and, subsequently, begin a CAR price reversal. The pattern is particularly distinct in *meme* stocks, and less pronounced in stocks with a broad following outside of WSB. A final analysis of the average sentiment expressed within the forum indicates that a Granger causal relationship exists between changes in average sentiment expressed about an asset and future asset returns in several popular assets on WSB.

The scope of our study leaves several promising directions for future research. A key area for investigation lies in the appropriate way to analyze a nonstationary dataset, such as WSB, and its predictive value across different time periods. Additionally, this analysis presents several methods for extracting signal from text and network interactions. As unstructured datasets are becoming more and more common, it may be valuable to investigate and standardize additional methods that can be useful for forecasting.

ACKNOWLEDGMENTS

We are grateful to Sandy Pentland and Doyne Farmer for their valuable feedback and advice on the project. We thank Baillie Gifford and the Institute for New Economic Thinking at the Oxford Martin School for funding our work and the Oxford-Man Institute of Quantitative Finance for their support. This article is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 870245. We would like to acknowledge the equal contribution of Valentina Semenova and Dragos Gorduza to this article.

[22] https://sites.google.com/view/wsbtrialsite.

REFERENCES

Agrawal, S., P. D. Azar, A. W. Lo, and T. Singh. 2018. “Momentum, Mean-Reversion, and Social Media: Evidence from StockTwits and Twitter.” *The Journal of Portfolio Management* 44 (7): 85–95.

Angelov, D. 2020. “Top2vec: Distributed Representations of Topics.” *arXiv* 2008.09470.

Araci, D. 2019. “FinBERT: Financial Sentiment Analysis with Pre-trained Language Models.” *arXiv* 1908.10063.

Azar, P. D., and A. W. Lo. 2016. “The Wisdom of Twitter Crowds: Predicting Stock Market Reactions to FOMC Meetings via Twitter Feeds.” *The Journal of Portfolio Management* 42 (5): 123–134.

Baumgartner, J., S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. 2020. “The Pushshift Reddit Dataset.” *Proceedings of the International AAAI Conference on Web and Social Media* 14: 830–839.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3: 993–1022.

Boylston, C., B. Palacios, P. Tassev, and A. Bruckman. 2021. “WallStreetBets: Positions or Ban.” *arXiv* 2101.12110.

Buz, T., and G. de Melo. 2021. “Should You Take Investment Advice from WallStreetBets? A Data-Driven Approach.” *arXiv* 2105.02728.

Chacon, R. G., T. G. Morillon, and R. Wang. 2022. “Will the Reddit Rebellion Take You to the Moon? Evidence from WallStreetBets.” *Financial Markets and Portfolio Management* 37: 1–25.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186. Minneapolis, MN: Association for Computational Linguistics.

Duz Tan, S., and O. Tas. 2021. “Social Media Sentiment in International Stock Returns and Trading Activity.” *Journal of Behavioral Finance* 22 (2): 221–234.

Fama, E. F., and K. R. French. 2004. “The Capital Asset Pricing Model: Theory and Evidence.” *Journal of Economic Perspectives* 18 (3): 25–46.

Gao, R., Z. Zhang, Z. Shi, D. Xu, W. Zhang, and D. Zhu. 2021. “A Review of Natural Language Processing for Financial Technology.” In *International Symposium on Artificial Intelligence and Robotics 2021*, pp. 262–277. SPIE, Fukuoka, Japan.

Gentzkow, M., B. Kelly, and M. Taddy. 2019. “Text as Data.” *Journal of Economic Literature* 57 (3): 535–74.

Granger, C. W. J. 1969. “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” *Econometrica: Journal of the Econometric Society* 37 (3): 424–438.

Pagolu, V. S., K. N. Reddy, G. Panda, and B. Majhi. 2016. “Sentiment Analysis of Twitter Data for Predicting Stock Market Movements.” In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pp. 1345–1350, IEEE, Paralakhemundi, India.

Schou, P. K., E. Bucher, M. Waldkirch, and E. Grünwald. 2022. “We Did Start the Fire: r/WallStreetBets, ‘Flash Movements’ and the GameStop Short-Squeeze.” In *Academy of Management Proceedings, 2022*, p. 14028. New York: Academy of Management Briarcliff Manor.

Semenova, V., and J. Winkler. 2021. “Reddit’s Self-Organised Bull Runs: Social Contagion and Asset Prices.” Working Paper No. 2021-04, Institute for New Economic Thinking at the Oxford Martin School.

Sul, H. K., A. R. Dennis, and L. Yuan. 2017. “Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns.” *Decision Sciences* 48 (3): 454–488.

Traag, V. A., L. Waltman, and N. J. Van Eck. 2019. “From Louvain to Leiden: Guaranteeing Well-Connected Communities.” *Scientific Reports* 9 (1): 5233.

Wan, X., J. Yang, S. Marinov, J.-P. Calliess, S. Zohren, and X. Dong. 2021. “Sentiment Correlation in Financial News Networks and Associated Market Movements.” *Scientific Reports* 11 (1): 3062.

Witts, D. W., E. Tortosa-Ausina, and I. Arribas. 2021. “The Irrational Market: Considering the Effect of the Online Community Wall Street Bets on Financial Market Variables.” Working paper, Universitat Jaume I.