

Short, Colorful, and Irreverent! A Comparative Analysis of New Users on WallstreetBets During the Gamestop Short-squeeze

Ehsan-Ul Haq

euhaq@connect.ust.hk

The Hong Kong University of Science and
Technology
Hong Kong SAR

Tristan Braud

braudt@ust.hk

The Hong Kong University of Science and
Technology
Hong Kong SAR

Lik-Hang Lee

likhang.lee@kaist.ac.kr

Korea Advanced Institute of Science and
Technology
South Korea

Anish K. Vallapuram

akvallapuram@connect.ust.hk

The Hong Kong University of Science and
Technology
Hong Kong SAR

Yue Yu

yyubv@connect.ust.hk

The Hong Kong University of Science and
Technology
Hong Kong SAR

Gareth Tyson

gtyson@ust.hk

The Hong Kong University of Science and
Technology
Hong Kong SAR

Pan Hui*

panhui@ust.hk

The Hong Kong University of Science and
Technology
Hong Kong SAR

ABSTRACT

WallStreetBets (WSB) is a Reddit community that primarily discusses high-risk options and stock trading. In January 2021, it attracted worldwide attention as one of the epicentres of a significant short squeeze on US markets. Following this event, the number of users and their activity increased exponentially. In this paper, we study the changes caused in the WSB community by such an increase in activity. We perform a comparative analysis between long-term users and newcomers and examine their respective writing styles, topics, and susceptibility to community feedback. We report a significant difference in the post length and the number of emojis between the regular and new users joining WSB. Newer users' activity also closely follows the affected companies' stock prices. Finally, although community feedback affects the choices of topics for all users, new users are less prone to select their subsequent message topics based on past community feedback.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in collaborative and social computing*; • **Information systems** → *Social networks*.

KEYWORDS

Reddit, WallStreet, Short Squeeze, Trading

*Pan Hui has second affiliation with University of Helsinki, Finland

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9130-6/22/04...\$15.00
<https://doi.org/10.1145/3487553.3524202>

ACM Reference Format:

Ehsan-Ul Haq, Tristan Braud, Lik-Hang Lee, Anish K. Vallapuram, Yue Yu, Gareth Tyson, and Pan Hui. 2022. Short, Colorful, and Irreverent! A Comparative Analysis of New Users on WallstreetBets During the Gamestop Short-squeeze. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3487553.3524202>

1 INTRODUCTION

In January 2021, a short squeeze¹ of the stock of GameStop (GME) occurred, causing significant financial losses for short-sellers. This short-squeeze was closely monitored by the WallStreetBets² (WSB) community on Reddit. American investor Keith Gill, a member of the community with the username *DeepFuckingValue*, had been sharing his long positions on the GME company since 2019, transforming an initial investment of US\$53,000 into close to US\$50 million in January 2021. Keith started posting daily updates of his position on the forum, bringing considerable exposure to the situation. As the epicentre of this event, the WSB community grew exponentially in terms of users, posts and comments during January 2021 [7]. In January 2021, the WSB subreddit grew from 1.5 million to 6 million users. The number of posts in January 2021 was twice as high as all prior posts in 2020. This raises the question of its impact on the community.

A high influx of new members may disrupt the community's communication norms, and such changes can affect the members' affinity towards the community [23, 25]. Bursts of activity often lead to duplicate messages [15] and chaotic discussions with no concentrated effort of leading information cascades [6]. For instance, one recent study shows that WSB activity in January alternated between "digital protests" against Wall Street and speculative trading tactics with no clear directive line [20]. The GME short-squeeze was a significant event and one of the most extensive occurrences of

¹https://en.wikipedia.org/wiki/Short_squeeze

²<https://www.reddit.com/r/wallstreetbets/>

an online community significantly disrupting the stock market. As such, the study of the WSB community may present unique implications on how online activity can affect offline events. Conversely, we posit that the sudden increase in community size provides us with a singular opportunity to study how communication norms evolve in an online community following an offline event. This paper explores the user writing and engagement characteristics concerning this event. This analysis evaluates how new users behave in the WSB subreddit and their reception by the existing community. The results from this study can offer valuable insight for downstream research problems, such as feature engineering for machine learning tasks [32, 39]. Additionally, characterising users on their traits and what factors can affect their communication choices plays an essential role in marketing or designing better algorithms to help retain users [16].

To the best of our knowledge, no prior work has characterised the impact and nature of such rapid community shifts in reaction to offline events. We argue that WSB offers a perfect case study for this exploration, particularly given the significant financial impact of the event itself. Note that such an increase in community size is also different from user influx caused by *defaulting* the subreddit [28].

To understand the shifts in community norms during the short-squeeze, we classify users based on their assumed subscription date to WSB. We separate older users from those who joined during the short squeeze and comparatively analyse their activity. We first explore the differences in the communication style of each user group and aim to evaluate new users' communication based on text length, emoji usage, and community feedback. We expect the new users to be primarily attracted to the short squeeze and for most discussions to revolve around the price of the companies' stocks. To confirm this assumption, we evaluate the correlation of the users' mentions of company names to the stock prices. Finally, previous research shows that community feedback plays a significant factor in authors' choice of topics [1]. We later revisit this hypothesis, looking at users' activity and the community itself. We compare the effect of such feedback in two different periods and based on the users' categorisation depending on their joining and activity. In summary, we ask the following research questions:

- **RQ 1:** How does the writing style differ between long-term users and users that became active after January 2021?
- **RQ 2:** Which users have a higher correlation with the stock market volatility based on their activity and writing styles.
- **RQ 3:** What is the influence of community feedback (comments) and topical trends on the users choice of topics?

We discover a significant difference between the writing style of users who joined in January 2021 (new users) and those users who were active in the previous year. New users (who most likely joined to follow the online trends and participate in the campaigns for holding the stock prices) wrote shorter messages with more emojis than senior users. We also find that community feedback has a much lower influence on new users' topical choices. This triggers a shift in community behaviour, driven by this influx. Although we notice significant differences in writing styles, community topics and behaviours between user groups, we note that newer users do adapt to the community to some extent. For example, we observe

them adopting the characteristic “lingo” and emojis of the WSB community.

2 RELATED WORK

Discussion-oriented online communities, including Reddit, are widely studied and can offer valuable insight into users, *e.g.* concerning demographics and social events [18, 30]. Users' behaviour can also be profiled based on their participation and engagement with the rest of the community members [14, 19].

Our study focuses on the impact of sudden community growth, member characterization and their correlation with offline events. Several prior works focus on the outcomes of increases in community size. Often, an increase in community size is associated with friction between new and old users, where old users may consider new users as a reason for community norm disturbance [28]. Additionally, the influx of new users is associated with reduced content quality and users who are less susceptible to community norms [17, 26]. Similarly, the linguistic patterns of new users can be linked to moderation practices in the community. Cheng et al. show that users who deviate from the general community linguistics while posting a high number of comments may get banned [9].

Others have focused on understanding how user behaviours are associated with feedback and engagement from the rest of the community members [25, 27]. This feedback has been measured using various metrics, including the number of upvotes, downvotes, and comments on a given post [2]. This, in turn, has led to work on trying to predict future user behaviour based on observed community feedback within communities [1, 12].

These predictions rely on several features and observations. For example, Adelani et al. show that community feedback affects the topical choice of about one-third of active users on Reddit [1]. Similarly, the scores on a user's previous posts can often predict the sentiment of future posts [12]. These observations have also led to a better understanding of strategies to improve future community engagements. For instance, posting about trending topics can help users to gain more followers [34]. In this paper, we build upon the work from Adelani et al. [1] to show that new users are more susceptible to community feedback than regular users on WSB. Compared to Adelani et al., who studied all users indifferently of their provenance, we categorise WSB users based on their potential joining time-period and the differences in their activity.

In this paper, we focus on WSB as an exemplar of a growing community with considerable effect on the real-life market, as evidenced by [10]. A small set of recent studies analysed the WSB subreddit. One study associates several social factors (*e.g.* COVID-19 and inequality) with the WSB activity that led to the short-squeeze [8, 10]. Several studies also correlate WSB users' activity and stocks prices [37]. A recent work reports that WSB posts containing buy signals from one firm can affect the return value of options [7]. Similarly, [36] correlates user activity on WSB with price fluctuations. Social media combined with the search engine data can also provide market insights [29], whereas another work observes consistent speculative activity patterns for WSB users who are involved in stock trading [20].

In contrast to these prior efforts on WSB, we focus on using WSB as an exemplar for a community with a sudden and significant influx

triggered by external events. In particular, we look at how new users tried to adapt to the community norms and the community response to them. Thus, our study focuses on characterising different user groups based on their activity and the correlation with the financial markets.

3 WALLSTREETBETS

WallStreetBets is a subreddit that takes a less serious approach to stocks and options trading, self-described as “*Like Achan found a Bloomberg Terminal*”. WSB users also differentiate themselves through specific, often politically incorrect terminologies, mixing text and emojis. This community is considered to be one of the epicentres of the 2021 GME short squeeze [4]. It received worldwide attention, leading to exponential community growth in January 2021. In only a month, the number of users quadrupled, and the number of posts exceeded the entire 2020 count. In the rest of this paper, we discuss the impact of such sudden community growth on users’ communication patterns in the WSB subreddit.

Dataset: We collect all posts on the WSB subreddit between the 1st January 2020 and the 30th January 2021. Due to the large number of posts, we use the pushshift API.³ Pushshift collects content generated on Reddit, considerably simplifying the data collection process. In total, we collect 300,292 posts from the year 2020 and 378,548 posts during January 2021. These posts are made from 107,542 and 218,279 users in 2020 and 2021, respectively. 8,286 and 7,452 posts were deleted in 2020 and 2021, respectively. Note Pushshift keeps records of the content even after it is deleted. Only content that was automatically deleted (primarily for violating community guidelines) appears deleted within the query results, explaining the low number of deleted posts in our dataset.

Data Pre-Processing: We later perform a comparative analysis between the two periods. Thus, we divide the data into two sets: the first set contains all posts from 2020, before the GME short squeeze, and the second set contains the content posted during January 2021, i.e., when the short squeeze happened, and the WSB community came into the spotlight. This time split allows us to study the differences in content generation patterns. However, some users are active in both periods. Therefore, we further divide the data based on user activity. We do not have the users’ profile information, and we do not know when the users subscribed to Reddit, let alone the date they joined WSB (which is not accessible to the public). Thus, we assume that users who appeared in 2021 but have no posts on WSB in the preceding year are new users who joined during the short squeeze in January 2021. This categorisation helps us differentiate the users who might have joined WSB because of offline activity in the stock market from earlier users in WSB. Note that our dataset contains many users whose posts only appear in January 2021, supporting this hypothesis.

Finally, we divide users into four groups. (1) The 2020 users (who only posted in 2020 and not in 2021), and (2) The 2021 users (who only posted in January 2021). The continuing users (who posted in both years) are then divided into two groups: (3) continuing-2020 (with posts exclusive to 2020), and (4) continuing-2021 (with posts exclusive to 2021). The last two groups give us additional comparative space to characterize the community communication

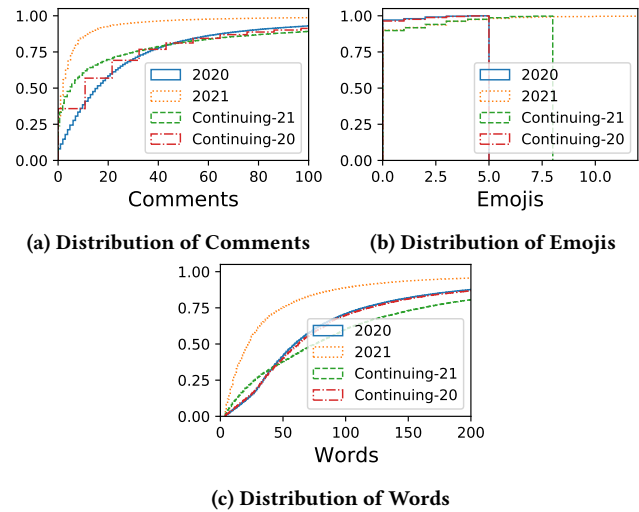


Figure 1: Distribution of the writing characteristics for three groups of users. (a) comments, (b) number of emojis in text, and (c) number of words in the post. Users active in 2020 (2020 and continuing user groups) have longer posts, most posts but use less emojis.

norms. In the rest of this paper, we often group the 2020 group and the continuing group under the name “senior users” for ease of readability and comparison when the behaviour is generalizable. In total, 87,622 users belong to the 2020 user group, 218,279 users belong to the 2021 user group, and 19,920 have contributed to the community in 2020 and 2021.

4 METHODOLOGY & RESULTS

In the following sections, we first compare the authors’ writing patterns between the two time periods (RQ1) before studying the correlation between offline stock prices and online activity for new users as compared to senior users (RQ2). Finally, we perform topic analysis to evaluate the effect of community feedback depending on the time period and the user categories (RQ3).

4.1 RQ1: Author Writing Patterns

RQ1 shows the differences between users’ writing styles depending on when they joined the WSB community, alongside the ensuing response from the community. This research question helps us confirm that the user groups, as defined in the previous section, display statistically significant differences, exhibiting a shift in users and content in January 2021.

We start by comparing the writing styles of new users (2021 group) with senior users (2020 users and continuing users based on their activity in the two periods) as well as the respective responses of the community through the following parameters:

- **Use of Emojis:** WSB members use certain emojis within their text. In January 2021, such usage of emojis (e.g., ‘to the moon’ – 🚀🌕) became more pervasive among WSB users. To study the community shift, we compare the volume of emojis used among user groups.

³<https://pushshift.io>

- **Post length:** The length of a message can highlight the effort expended by the author to write posts, and it impacts the readers' ability to comprehend the message.
- **Number of comments:** The number of comments is one of the standard engagement markers.

The two first properties have been used in various user classification tasks [14], and help us confirm whether users who joined during the GME short squeeze exhibit a different behaviour to long-term members of the community. The number of comments represents the community's response to the new influx of users during January 2021. Previous research has shown that such characteristics can also highlight anti-social behavior [9]. We use Kruskal-Wallis(KW) test, followed by Dunn's test, to show the statistical difference between each group [13]. In addition, we use the Bonferroni adjustment to control the simultaneous test problem.

These tests confirm that *the users who joined in 2021 have a different writing style and response than the users active in 2020*. Users' activity periods have a significant effect on the number of emojis (KW Test: $X^2(3) = 2683, p < 0.0001$), comments (KW Test: $X^2(3) = 28150, p < 0.0001$), and number of words (KW Test: $X^2(3) = 18663, p < 0.001$) for users in the four groups. Within each group, we further consider two subgroups: (1) the users who are among the top 25% based on their volume of posts, and (2) the users who have more than $\mu + \sigma$ posts where μ is the mean number of posts and σ the standard deviation of posts from all users. We report the test statistics for all groups in Table 1. We further perform the pairwise comparison across the groups and show Dunn's test statistics in Table 4 (Appendix). For all comparisons, the p-value is below 0.001, confirming the significant differences in the communication styles and community feedback between users present in 2020 and new users in 2021. The distributions in Figure 1 and the pairwise comparisons show that newer users (2021) favour shorter messages with more emojis.

We note that emojis were part of the WSB writing styles before the short-squeeze (more discussion in Section 5). We assume that higher usage of emojis and shorter messages may provide a quick workaround to post, considering that WSB involves chaotic discussions and new users wanted to raise their voice during this short-squeeze event [6, 20]. Another observation is the deviation in the behaviour of the continuing users from the community norms in 2020. The continuing-2021 users used more emojis in 2021 as compared to 2020. Additionally, continuing-2021 users wrote longer posts than the new users in 2021 (albeit shorter than continuing-2020 norms). This shows that the time period of the analysis is of significance when observing the user characterization. The difference for the new users is intuitively based on the sudden surge in community size and posts. However, the significant change in norms for continuing users is an interesting observation that can be explored further for causal analysis in future studies.

4.2 RQ2: Author groups and stock prices

RQ2 aims to establish whether a user group is more sensitive to the volatility of stock prices in their choices of topics. We thus study the correlation between price volatility and the number of mentions of the company name or ticker. Previous studies have suggested that volume-based measures (mentions of the company

	Comments	Emojis Text	Words
overall	28150.3	2683	18663.2
top 25	23355.1	6881.8	16213.8
$> \mu + \sigma$	9504.4	4170.9	7284.1

Table 1: Kruskal-Wallis Results ($\chi^2(3)$) for the writing style differences between the three user groups. All results are statistically significant with $* = p < 0.001$.**

name) have a positive correlation with stock prices [36]. Contrary to these studies, we focus on volatility. Volatility can capture the risk associated with financial assets during a given period [31]. As high-risk trading strategies characterise the WSB community, we expect that users are more eager to react to price variations rather than the absolute price itself.

We define volatility as the standard deviation of the logarithmic return of the stock prices over a particular time period, $\sigma_P = \sigma_{daily} \sqrt{P}$, where P is the number of days, and σ_{daily} is the standard deviation of logs of returns. We perform the volatility analysis over five-day periods (one week in the stock market).

We then perform Pearson correlation analysis to see which user group's mentions of a given company have a stronger correlation with the stock price volatility. Although users mention many companies in our dataset, we only consider the three most discussed and speculated companies during short squeeze, namely GameStop (GME), BlackBerry (BB), and AMC Entertainment (AMC). We use the Yahoo Finance API⁴ to collect the stock price data starting from 1st January 2020 until 30th January 2021.

We first test if the daily volume of posts correlates with the stock volatility. For this, we first extract all *mentions* of a given company. We include stock market tickers for the company, the company name, along with all sensible variants as *mention*. For instance *GME*, *GameStop*, and *Game Stop* will all be considered as mentions for GameStop. To understand the importance of different users, we separate the mentions based on which user group posted them.

Table 5 (Appendix) presents the correlation between the volume and number of mentions for the three stocks. New users exhibit a stronger correlation with the stock volatility, confirming that new users likely subscribed to WSB because of the offline stock squeeze. Their discussions primarily follow the variations of the studied stock prices. That said, there are interesting observations on senior users. The continuing users (continuing-2020) did not have a significant or have very low correlation during 2020. But the rest of the community collectively (2020 users) have a relatively medium to high correlation with the stock volatility. However, the correlation for continuing users increased significantly during 2021, albeit lower than new users. Possible reasons for this could be a significant change in emoji use coupled with a higher frequency of the message during 2021 for the continuing-2021 users. We then run the above analysis for posts that contain at least one emoji. WSB users use a wide variety of emojis to express sentiment related to stock prices movements. This analysis examines whether posts containing emojis have a higher correlation with volatility. 2021-users (new users) tend to use more emojis than the other senior

⁴<https://rapidapi.com/apidojo/api/yahoo-finance1>

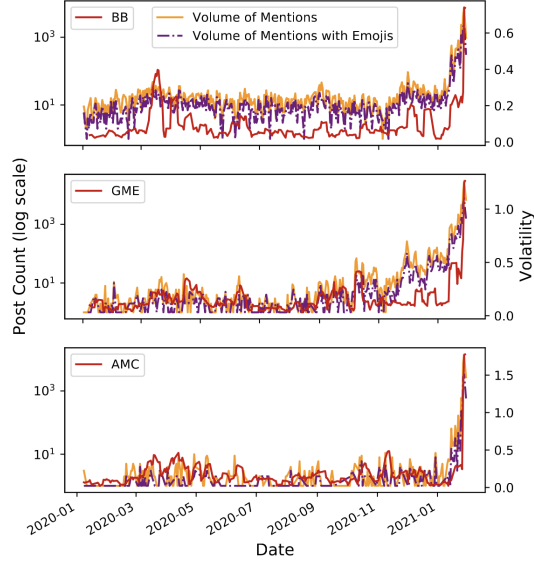


Figure 2: Daily volume of overall posts (yellow solid line) and posts containing emojis (purple dashed lines) for three companies and the respective company’s 5-days period stock volatility (red solid line). Mentions of the company name closely follows the volatility.

(See Table 6) (Appendix). However, the volume of the mentions that contains emojis has less correlation to stock price volatility than the volume of all mentions.

These results also highlight that continuing users became part of the large-scale information cascade. The qualitative analysis of their messages can highlight whether continuing users aligned with the new users in the discourse. Nonetheless, the senior users changed their communications significantly (RQ1), and their higher correlation with stock volatility further supports this change compared to 2020.

4.3 RQ3: Community Feedback and Topic Continuation

Finally, we analyse and compare users’ susceptibility to continue writing on a given topic in two consecutive posts based on the feedback received on their previous posts in WSB. If a user writes two successive posts on the same topic, the user is said to continue the topic. We want to remind readers that community feedback refers to the number of comments and scores on a Reddit post. In our work, we only use the number of comments users receive on their posts as the primary indicator of community feedback.

4.3.1 Overview of Methodology. We employ the framework proposed by [1]. This framework uses a regression model to predict whether users will revisit a topic based on previous posts’ feedback. We narrow down the model’s scope to fit the smaller WSB community and study the two time periods separately. The model relies on three features: user topic preference, user feedback likelihood, and topic trend at a given time t . Eventually, we use the results from this model to measure the effect size of the feedback feature in each

user group. That is, the influence of community feedback on the choice of subsequent topics. The effect size of feedback can help us understand which user group is more susceptible to community feedback.

4.3.2 Feature Extraction. This section presents the three features and how we extract them from the data.

Topic Modeling: We use the Latent Dirichlet Allocation (LDA) model to find the topics [5]. The number of posts in January 2021 is similar to the number of posts for the entirety of 2020. We divide the data into two time periods: 2020 and 2021. We then apply LDA to each year’s data separately. We assume that such separation does not affect our users’ grouping as the users active in 2020 and 2021 will have their posts considered in both periods. We evaluate the quality of the topic modelling using the coherence score [33]. We experiment with LDA models with topics ranging from 5 to 30. The optimal number of topics is 16 and 14 for 2020 and 2021, respectively. This evaluation confirms our decision to separate the data into two subsets: the quality of the topics using the whole dataset is lower than considering two separate subsets.

Topic Preference: The topic preference (TP) refers to the author’s preference for a topic at the time of the post. Formally, it is the probability of a user i to post on a topic k . Laplacian Smoothing is applied to this probability. Hence, the probability can be denoted as

$$P_i(k) = \frac{N_i(k) + 1}{N_i + K}$$

where $N_i(k)$ and N_i are the number of posts on topic k and total number of posts from the user i , respectively and K is the number of possible topics. We then apply the inverse of the sigmoid function (σ^{-1}) to $P_i(k)$, thus a topic preference for a user i is then defined as

$$X_i(k) = \sigma^{-1}(P_i(k))$$

Topic Trend: Topic trend (TT) refers to the most popular topic at any particular time in the community. We use a daily time window for trending topics for the sake of simplicity. The topic with the most posts on a given day is considered a trending topic. There are 362 time-bins in 2020 and 30 bins in 2021; for each day, each topic’s posts are counted and normalised. Based on the day of the post, the trending topic k (from K topics as calculated by LDA in the first step) is given to the regression model.

Feedback for a User: In the feedback feature, we model the number of comments received by a user on a post. Feedback is modelled in three different ways: (a) feedback rate $r_i = N_c/\Delta t$ where N_c is the number of comments received for a post by user i and Δt is the time since the previous post; (b) the log of feedback rate $\log(r_i)$; and (c) the cumulative probability of the feedback rate $P(R_i < r_i)$ that the user receives on one post compared to all of the users’ posts.

4.3.3 Topic Continuation Model. We now look at predictive analysis of whether a user will continue the topic in the next post based on the features mentioned above. The topic continuation label for a given post is decided by checking the previous post’s topic. If both posts have the same topic, we consider the topic continuation label as *true*.

	Moderately Active	Highly Active
2020	0.881 ± 0.037	0.963 ± 0.058
2021	0.919 ± 0.018	0.989 ± 0.012
Cont. 2020	0.877 ± 0.083	
Cont. 2021	0.828 ± 0.096	

Table 2: Mean and standard deviation of cross-validated test accuracy for different user groups. Model achieves higher accuracy for new and highly active users.

We train several models using different combinations of the above features to perform feature importance analysis. We use cross-validation with 20 splits. We report the average accuracy scores and standard deviations for each model in Table 7 (Appendix). All models trained with the feedback features are combined with the topic trend (TT) and topic preference (TP) features. Using the TT and TP features together results in a significant increase compared to using either individually.

The model with all features performs the best, suggesting users are susceptible to posting on topics for which they receive higher feedback. The results for both periods (2020 and 2021) are consistent. However, this behaviour is more predictable in 2021 (during the short squeeze) than in 2020, with higher accuracy for 2021. In other words, community feedback is a better predictor of the topics in 2021 compared to 2020. The performance improvements using all three features is shown through Cohen’s d effect size. We compare the test scores from the model with $\log(N_c/\Delta t)$ feedback and the model without feedback. The effect sizes are 0.42 and 0.16 for 2020 and 2021, respectively. Although community feedback is a better predictor of topic choices in 2021, users in 2020 were more susceptible to community feedback than users who posted in 2021. Such a result suggests new users being less receptive to the community feedback, diverting the community from its original purpose.

4.3.4 User Characterisation. After establishing the model features and the effect size of the feedback for the time split data, we understand the effect of such features on the different user groups based on their activity level. We measure the activity of users as the average time delay between posts. A user is considered highly active if their activity is in the top 25 percentile. The remaining users are considered moderately active users. We train the logistic regression model for each group separately. We use the best performing configuration in Table 7 with cross-validation of 20 splits and report the test accuracy scores in Table 2.

The model presents a higher accuracy to predict the topic continuation for users who joined in 2021, confirming the results from the previous section. Moreover, the model displays a high accuracy for highly active users in all user groups. To further explore the effect size of community feedback, we perform Cohen’s test across all group sizes and show the results in Table 3. The effect size varies based on the activity of users within a group. The effect size is also different across all groups. We observe several patterns. Highly active users from 2021 and 2020 are more susceptible to community feedback than moderately active users. Continuing users (the regular users) remain the least susceptible to community feedback. Even during the short squeeze event, their susceptibility remained lower. Regular users seem to care less about the community’s feedback

	2020 (Mod.)	2020 (Act.)	2021 (Mod.)	2021 (Act.)	Cont-2020
2020 (Act.)	0.81	0.00			
2021 (Mod.)	0.911	-0.30	0.00		
2021 (Act.)	2.62	0.757	2.55	0.00	
Cont-2020	-0.13	-0.59	-0.49	-1.10	0.00
Cont-2021	-1.45	-1.83	-1.99	-2.80	-0.85

Table 3: Cohen’s effect size comparison of feedback susceptibility on topic continuations on different combinations of user groups with different levels of activeness: moderate (Mod.) and active (Act.). Community feedback has a higher effect size for the higher active user groups.

on their topic choice when compared to the rest of the users. On the other hand, the newer users, particularly highly active new users, cared more for the community feedback. This also suggests that newer users tried to adjust with the community preference compared to the continuing users. Even though the continuing users significantly changed their communication patterns (RQ1), their susceptibility to the feedback changed only by the margin of 0.85 during this period.

5 FINDINGS AND IMPLICATIONS

We summarize our findings as follows:

RQ 1: Significant differences exist among the writing styles of users groups. New users (2021) write shorter posts with more emojis as compared to the posts by users in 2020. Continuing users changed their patterns to shorter posts with more emojis during short squeeze, albeit still writing longer posts than newer users. Continuing users, in general, have a higher number of comments.

RQ 2: There is a positive correlation between the number of posts mentioning the three companies with the volatility of the stock prices. Such correlation is higher for new users (2021) than senior users (2020 and continuing), as new users joined the subreddit following the explosion in the volatility of these companies.

RQ 3: The model predicts users’ next conversational topic with high accuracy. Community feedback is a better predictor for topic choice in 2021. Highly active users tend to be more sensitive to community feedback than moderately active users. However, the continuing users remain less susceptible to feedback during 2021 and 2020.

Emojis usage and community evolution: These research questions reflect the significance of user behaviours in WSB and hence their topical choices. To further analyse and understand WSB users’ collective action, we delve into frequent expressions and emojis in messages. The WSB users developed a self-deprecating norm derived from *meme* and emojis cultures, i.e., humorous ideation with a broad cultural reference propagating through social media. These norms lead to the appearance of a characteristic lingo, mixing original expressions and emojis to carry the message. These sentences involve market-related sentiment (‘to the moon’ or 🚀🌕, ‘diamond hands’ or 💎👉, ‘paper hands’ or 📄👉, and ‘gay bears’ or 🐻👉), but also expressions shedding light on how users perceive the community in its whole (‘wife’s boyfriend’, ‘tendies’ or 🍷, ‘apes

(👊) *together strong*'), and others purely related to popular memes on Reddit as a whole (*'it ain't much, but it's honest work'*). Often, these phrases are replaced or reinforced by emojis. For instance, 🏳️‍🌈🐻 stands for 'gay bears', referring to the uncooperative stock selling or shorting, and hence the fall of the stock prices, while 🚀🌕 corresponds to the price skyrocketing 'to the moon'.⁵ Similarly, 💎👉, or 'diamond hands' implies that the users will hold the stock regardless of the investment risk.⁶

Figure 3a further depicts the trends of the top-5 expressions in 13 months, while Figure 3c represents the top-5 emojis. These two figures show that the community had embraced these expressions and the corresponding emojis long before the 2021 increase in community size. New users joining the community in January 2021 rapidly adopted such an emoji-popular communication media, as we demonstrated in RQ1. It is important to note that communication among users aligns with a shared goal of stimulating the stock prices to work against short-selling by professional traders from Wall Street. The expressions and emojis serve as a symbol conveying such a shared goal. For instance, the top-rated expressions in WSB 'to the moon' and 'diamond hands', respectively represented by the emojis 'rocket' and 'gem_stone' encourage participants in WSB to buy in and hold the stock. Meanwhile, expressions such as 'paper hands' and the related emojis 'bear' and 'rainbow' discourage counterproductive actions through selling the stocks. As a result, the stock prices reached their 52-week peaks on 27th Jan 2021 (BB: \$ 25.10 USD; GME: \$ 347.51 USD; AMC: \$19.90 USD).

In Figure 3b and 3d, we present the normalized proportion of the top-5 expressions and emojis over each day. Such a representation allows us to evaluate the most popular expressions during a given period and shed light on the market sentiment. As expected, the bearish market in 2020, primarily due to high uncertainty from the COVID-19 pandemic, makes popular use of emojis such as 'bear' and 'rainbow' representing *selling the stocks*. In contrast, the emojis representing 'buy-in' or 'holding the stock' (i.e., 'rocket', 'gem_stone', and 'raising_hands') start to surface in late 2020. During Jan 2021, the users' sentiments in WSB were reflected by the encouragement of 'buy-in' or 'holding the stock', albeit a shared goal supports such positive sentiments. Finally, it is interesting to note that usage of the expression 'tendies' significantly drops by the end of 2020. This expression was often used as a synonym for monetary gains. Therefore, we would expect to explode during January 2021, where many users made or looked forward to monetary profit. A potential explanation for discrepancy could be the influx of new users who did not entirely adopt the community codes they are joining. In addition, a significant change in the use of emojis by continuing users occurred during January 2021, with communication norms in 2021 being closer to the new users. However, we do not have a causal analysis for this behaviour whether new users were following this change in the regular users' behaviour or is vice-versa. Causal analysis for similarity in behaviour remains a potential question for further study. Our works quantify such differences, and future work based on cross-interaction between

different user types can explain the causal for reported community change.

Community changes: concerted effort or individual voices?

Despite disrupting the discourse in the community (RQ 1 and RQ 3) and evident offline correlation (RQ 2), the new users' characteristics do not fully conform to anti-social behaviour. Such deviation in communication norms is associated with potential trolling and anti-social behaviour in previous research [9] notably when the regular users of the community shifted their communication closer to the new users than in 2020. Our results show that even though the use of emojis and words is different for new members, their posts still cannot attain a higher number of comments as compared to the continuing users (Figure 1). The new users, who were the primary source of the posts surge during the short squeeze event, could thus not get greater attention from other community members. In other words, new users could not succeed in generating longer discussions threads compared to the continuing users. This phenomenon confirms the finding from [20], showing that new users have less cohesive campaigns and their posts are primarily the result of individual voices.

This behavioural difference between the users can affect the community maintainability. Previous research on sudden change highlights that effective community moderation in the sudden surge can help to maintain the order and sustain the growth in the community [23]. Although we do not study moderation efforts by senior users, our data statistics show that continuing users have also increased their activity during January, posting over 8,000 messages in January 2021 compared to 3,548 messages in December 2020. Further content analysis can highlight how senior users interact with new users during this period.

Sensitivity of users to community feedback: Lastly, we reviewed the community feedback and examined its relationship with the user's topical choice. Such a relationship can improve the prediction of topical choices for all user groups. Indeed, the behaviours of participants in WSB aligns with a theory of social science [21]: *"positive experiences lead to ingrained memory in the participants' minds, and the participants have the tendency of repeating the events to obtain such positive experience"*. In this sense, participants in WSB get positive experiences by gradually achieving the goal of securing the stock prices at surging trends and thus sabotaging the profitability of professional traders. It is noteworthy that, different from the above prior work, the participants in WSB show different effect sizes. Community feedback displays a more noticeable effect size for highly active users. On the other hand, new users are significantly more affected by community feedback in their topic selection than senior users. Although we do not conduct causal analysis between the user behaviours in WSB and the user goal(s), one possible explanation of such a different effect size is as follows.

New users entered the community during the hype of stock price volatility. As a result, their interests are primarily related to the volatility of a few companies. Hence, that would have resulted in fewer topics discussed during 2021 (also evident from our topic modelling results from Results Section regarding RQ3). Their higher susceptibility to the feedback can also show their efforts to reinforce their sense of belonging to the community. On the other hand, more senior users have already integrated into the community's and thus care less about their topical choice based on community feedback.

⁵https://www.reddit.com/r/wallstreetbets/comments/kmnj9d/reloaded_today_20k_shares_1950_diversification_is/

⁶https://www.reddit.com/r/wallstreetbets/comments/lmkzmd/weve_come_this_far_retards_all_paper_hands_are/

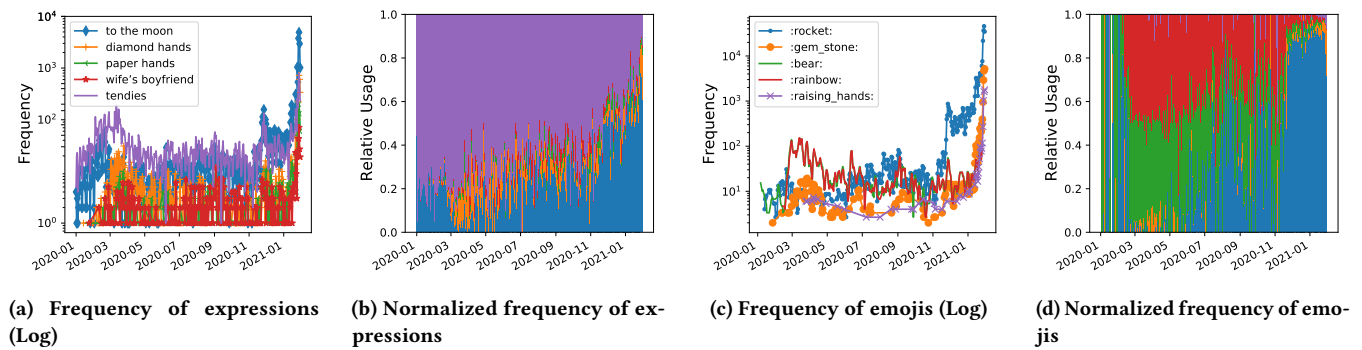


Figure 3: Absolute frequency (a,c), and normalized frequency (b,d) of the top-5 expressions and emojis. Both expressions and emojis give a view into the overall market sentiment, with bearish expressions and emojis from March 2020 until late 2020, and a resurgence of calls to buy and hold in late 2020-early 2021.

Our results on the feedback susceptibility pinpoint a theoretical question: What is the optimum balance in choosing a topic? On one end, users may choose a topic independently. On the other end, users might get influenced by community feedback. Additionally, the users from three categories might have varied information needs and may perform different search queries to see related content. As such, the reciprocity of the engagement might be diversified based on their needs [38]. Such behaviour needs to be studied in relation with users [22], who are considered to have influence power on the peer users. In other words, *how would the influencers react to the community feedback* remains a question for future studies. Our analysis highlights another practical implication, especially when a high level of conversation rate exists. Users can be overwhelmed by large volumes of messages and comments, albeit we omitted the potential exposure of the contents [3]. The exposure of a post can be impacted by information overload. For instance, some posts get less exposure in the community due to a higher volume of posts being generated together with users' browsing time regarded as a limited resource. These factors limit the posts a user can see. Thus, a limited exposure of posts could limit the possibility of receiving user feedback. Our findings reinforce the existing knowledge that users' topical choice depends on the feedback mentioned above. Accordingly, the user's topical choice could impact the users' affinity towards the community [25]. This offers an opportunity for the research community to examine how Reddit users tackle the situation mentioned above to improve engagement.

Limitations: This paper focuses only on empirical measurements of user feedback and the volume of posts and comments within different user groups. State-of-the-art natural language processing models could provide an insightful analysis of the user's comments and uncover latent behaviours. For instance, we may check the existence of stock price manipulation through co-uses of stock names in the same discussion thread [11]. Moreover, this study did not quantify user interaction such as reciprocity of engagement between different user groups. Inter-group and intra-group engagement is necessary to support the causal analysis of user feedback among user groups and the user's topical choices. As such, we can understand how new users and existing users interact with each other and the effect of such interaction on their topical

choices. Interaction reciprocity is associated with users' perceived commitment to engage with each other and can vary based on the users' activity in the community [35, 38]. In terms of stock price prediction, advanced predictive methods such as LSTM or other deep learning methods may produce better accuracy [24]. However, our primary purpose was to present the basis of different users' characterisation within the WSB subreddit.

6 CONCLUSION

In this paper, we presented a comparative analysis of writing style, correlation with the stock prices, and the effect of community feedback on users' topics. We have shown that new users wrote shorter messages and used more emojis than the community norms during 2020. The continuing users wrote, on average, longer posts in 2020 and 2021 than the other two groups. However, regular users significantly changed their communications during January 2021, with more emojis and shorter posts than their previous norms. In addition, the correlation of continuing users with stocks volatility increased significantly during 2021, suggesting that continuing users were part of the discussions around the short-squeeze. We also showed that community feedback significantly affects the topic choices for WSB users, particularly for highly active users. Interestingly, users who joined in 2021 are more susceptible to community feedback than 2020 users, and continuing users are least susceptible to the feedback. Finally, we found that, although the communication styles are significantly different between new users and older users, new users adopted a specific language style consisting of specific emojis that characterises the WSB community. In our future work, we will include the comments in the predictive analysis and link the user profile information to characterise the users efficiently.

ACKNOWLEDGMENTS

This research has been supported in part by project 16214817 from the Research Grants Council of Hong Kong, and the 5GEAR project (Grant No. 318927) and the FIT project (Grant No. 325570) funded by the Academy of Finland.

REFERENCES

- [1] David Ifeoluwa Adelani, Ryota Kobayashi, Ingmar Weber, and Przemyslaw A Grabowicz. 2020. Estimating community feedback effect on topic choice in social media with predictive modeling. *EPJ Data Science* 9, 1 (2020), 25.
- [2] Kholoud Khalil Aldous, Jisun An, and Bernard J Jansen. 2019. View, like, comment, post: Analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. In *Proceedings of AAAI ICWSM 2019*. 47–57.
- [3] Jisun An, Daniele Quercia, and Jon Crowcroft. 2013. Fragmented social media: a look into selective exposure to political news. In *Proceedings of ACM WWW 2013*. 51–52.
- [4] Abhinav Anand and Jalaj Pathak. 2021. WallStreetBets against wall street: The Role of reddit in the GameStop short squeeze. *IIM Bangalore Research Paper* 644 (2021).
- [5] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* (March 2003), 993–1022.
- [6] Christian Boylston, Beatriz Palacios, Plamen Tassev, and Amy Bruckman. 2021. WallStreetBets: Positions or Ban. *arXiv preprint arXiv:2101.12110* (2021).
- [7] Daniel Bradley, Jan Hanousek Jr, Russell Jame, and Zicheng Xiao. 2021. Place Your Bets? The Market Consequences of Investment Advice on Reddit's Wallstreetbets. *Social Science Research Network* (2021).
- [8] Tolga Buz and Gerard de Melo. 2021. Should You Take Investment Advice From WallStreetBets? A Data-Driven Approach. *arXiv preprint arXiv:2105.02728* (2021).
- [9] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Antisocial behavior in online discussion communities. In *Ninth International AAAI ICWSM*.
- [10] Usman W Chohan. 2021. Too Big to Fail, Too Small to Win: The Counter-Hegemony of WallStreetBets. *Too Small to Win: The Counter-Hegemony of Wall-StreetBets (May 18, 2021)* (2021).
- [11] Stefano Cresci, Fabrizio Lillo, Daniele Regoli, Serena Tardelli, and Maurizio Tesconi. 2018. \$FAKE: Evidence of spam and bot activity in stock microblogs on Twitter. In *Proceedings of AAAI ICWSM 2018*.
- [12] Jenny L Davis and Timothy Graham. 2021. Emotional consequences and attention rewards: the social effects of ratings on Reddit. *Information, Communication & Society* (2021), 1–18.
- [13] Alexis Dinno. 2015. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *The Stata Journal* 15, 1 (2015), 292–300.
- [14] Christopher Ifeanyi Eke, Azah Anir Norman, Liyana Shuib, and Henry Friday Nweke. 2019. A survey of user profiling: State-of-the-art, challenges, and solutions. *IEEE Access* 7 (2019), 144907–144924.
- [15] Eric Gilbert. 2013. Widespread underprovision on reddit. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 803–808.
- [16] Wei Gong, Ee-Peng Lim, and Feida Zhu. 2015. Characterizing silent users in social media communities. In *Proceedings of the International AAAI ICWSM*.
- [17] Andreea Gorbatai. 2011. Aligning collective production with demand: Evidence from Wikipedia. *Available at SSRN* (2011).
- [18] Ehsan Ul Haq, Tristan Braud, and Pan Hui. 2020. Community Matters more than Anonymity: Analysis of User Interactions on the Quora Q&A Platform. In *Proceedings of IEEE/ACM ASONAM 2020*. IEEE, 94–98.
- [19] Ehsan Ul Haq, Tristan Braud, Young D Kwon, and Pan Hui. 2020. A survey on computational politics. *IEEE Access* 8 (2020), 197379–197406.
- [20] Tim Hasso, Daniel Müller, Matthias Pelster, and Sonja Warkulat. 2021. Who participated in the GameStop frenzy? Evidence from brokerage accounts. (2021).
- [21] Yuheng Hu, Shelly Farnham, and Kartik Talamadupula. 2015. Predicting user engagement on twitter with real-world events. In *Proceedings of the International AAAI ICWSM*, Vol. 9.
- [22] Chung-Wha 'Chloe' Ki and Youn-Kyung Kim. 2019. The mechanism by which social media influencers persuade consumers: The role of consumers' desire to mimic. *Psychology & Marketing* (2019).
- [23] Charles Kiene, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. Surviving an "Eternal September" How an Online Community Managed a Surge of Newcomers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1152–1156.
- [24] Prahlad Koratamaddi, Karan Wadhvani, Mridul Gupta, and Sriram G Sanjeevi. 2021. Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. *Engineering Science and Technology, an International Journal* (2021).
- [25] Robert E Kraut, John M Levine, Marisol Martinez Escobar, and Amaç Herdağdelen. 2020. What Makes People Feel Close to Online Groups? The Roles of Group Attributes and Group Types. In *Proceedings of AAAI ICWSM 2020*.
- [26] Robert E Kraut and Paul Resnick. 2012. *Building successful online communities: Evidence-based social design*. MIT Press.
- [27] John M Levine and Richard L Moreland. 1994. Group socialization: Theory and research. *European review of social psychology* 5, 1 (1994), 305–336.
- [28] Zhiyuan Lin, Niloufar Salehi, Bowen Yao, Yiqi Chen, and Michael S Bernstein. 2017. Better when it was smaller? community content and behavior after massive growth. In *Eleventh International AAAI Conference on Web and Social Media*.
- [29] Ștefan Lyócsa, Eduard Baumöhl, and Tomáš Vřrost. 2021. YOLO trading: Riding with the herd during the GameStop episode. (2021).
- [30] Edward Newell, David Jurgens, Haji Saleem, Hardik Vala, Jad Sassine, Cairtin Armstrong, and Derek Ruths. 2016. User migration in online social networks: A case study on reddit during a period of community unrest. In *Proceedings of AAAI ICWSM 2016*.
- [31] Nuno Oliveira, Paulo Cortez, and Nelson Areal. 2017. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications* (2017).
- [32] Rebekah Overdorf and Rachel Greenstadt. 2016. Blogs, twitter feeds, and reddit comments: Cross-domain authorship attribution. *Proceedings on Privacy Enhancing Technologies* 2016, 3 (2016).
- [33] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the Space of Topic Coherence Measures. In *Proceedings of ACM WSDM 2015*.
- [34] Luca Rossi and Matteo Magnani. 2012. Conversation practices and network structure in Twitter. In *Proceedings of AAAI ICWSM 2012*.
- [35] Sally A Shumaker and Arlene Brownell. 1984. Toward a theory of social support: Closing conceptual gaps. *Journal of social issues* 40, 4 (1984), 11–36.
- [36] Angel Tengulov, Franklin Allen, Eric Nowak, and Matteo Pirovano. 2021. Squeezing Shorts Through Social News Platforms. *Available at SSRN 3823151* (2021).
- [37] Joey Wong. 2021. The Impact of Subreddit Comments on Daily Return and Volume. (2021).
- [38] Xue Yang, Guoxin Li, and Songshan Sam Huang. 2017. Perceived online community support, member relations, and commitment: Differences between posters and lurkers. *Information & management* 54, 2 (2017), 154–165.
- [39] Jiangchuan Zheng, Siyuan Liu, and Lionel M Ni. 2014. User characterization from geographic topic analysis in online social media. In *2014 IEEE/ACM ASONAM*.

A APPENDIX

	Groups	Comments	Emojis	Words
Exclusive	con_new/2020	38.1 ***	-29.9 ***	-6 ***
	cont_old/2020	-11.6 ***	-4.8 ***	-2.9 ***
	cont_old/cont_new	-42.8 ***	25.2 ***	3.9 ***
	2021/2020	151.1 ***	-45.2 ***	123.2 ***
	2021/con_new	47.8 ***	3.6 ***	75 ***
	2021/con_old	137.9 ***	-33.6 ***	106.4 ***
Top 25	con_new/2020	46.9 ***	-45.8 ***	-24.3 ***
	cont_old/2020	-9.1 ***	-4.6 ***	2.3
	cont_old/cont_new	-51.4 ***	42.1 ***	25.2 ***
	2021/2020	139.7 ***	-71.4 ***	117.1 ***
	2021/con_new	69.8 ***	-18.5 ***	108.8 ***
	2021/con_old	142.3 ***	-67 ***	113.1 ***
> $\mu + \sigma$	con_new/2020	30.5 ***	-34.6 ***	-24.8 ***
	cont_old/2020	-10.4	-0.99	1.8
	cont_old/cont_new	-35.8 ***	33.8 ***	25.7 ***
	2021/2020	88 ***	-54.9 ***	77.3 ***
	2021/con_new	51.6 ***	-21.2 ***	80.6 ***
	2021/con_old	92.6 ***	-54.2 ***	76.07 ***

Table 4: Dunn’s test to identify the differences across the group of users. (* = $p < 0.001$) Use of emojis in titles and post length is significantly different among user groups. In $\mu + \sigma$ group the difference between the 2020 and continuing-2020 users is not statistically significant.**

	Users	GME	AMC	BB
2020	All	0.59 ***	0.61 ***	0.42 ***
	Top 25	0.50 ***	0.46 ***	0.36 ***
	Mean SD	0.42 ***	0.28 ***	0.31 ***
2021	All	0.75 ***	0.78 ***	0.65 ***
	Top 25	0.77 ***	0.79 ***	0.63 ***
	Mean SD	0.78 ***	0.80 ***	0.61 ***
Continuing-2020	All	0.13 ***	0.01	0.15 ***
	Top 25	0.07	0.08	0.05
	Mean SD	0.01	0.04	0.16 ***
Continuing-2021	All	0.57 ***	0.63 ***	0.41 ***
	Top 25	0.47 ***	0.49 ***	0.35 ***
	Mean SD	0.41 ***	0.24 ***	0.29 ***

Table 5: Volatility correlation with the mentions. All results are statistically significant with * = $p < 0.001$. New users display higher correlation of mentions with stock prices.**

	Users	GME	AMC	BB
2020	All	0.56 ***	0.62 ***	0.39 ***
	Top25	0.48 ***	0.49 ***	0.35 ***
	Mean SD	0.41 ***	0.33 ***	0.33 ***
2021	All	0.76 ***	0.77 ***	0.60 ***
	Top25	0.77 ***	0.78 ***	0.58 ***
	Mean SD	0.77 ***	0.78 ***	0.57 ***
Continuing-2020	All	0.18 ***	0.12	0.28 ***
	Top 25	0.10	0.07	0.21 ***
	Mean SD	0.04	0.08	0.22 ***
Continuing-2021	All	0.61 ***	0.63 ***	0.43 ***
	Top 25	0.50 ***	0.49 ***	0.38 ***
	Mean SD	0.44 ***	0.37 ***	0.30 ***

Table 6: Volatility correlation with mentions for posts containing emojis. All results are statistically significant with * = $p < 0.001$. New users display higher correlation of mentions with stock prices.**

Primer on WSB

Reddit is a discussion-based online social platform that allows users to discuss a multitude of topics on communities known as subreddits [19]. These communities are focused on specific niches, which help Reddit to attract many users [30]. WallStreetBets (WSB) is a subreddit community that focuses on (high-risk) stock market and options trading discussions. Contrary to other stock market-related communities, WallStreetBets takes a less serious approach to stocks and options trading, self-described as “Like *Achan found a Bloomberg Terminal*”. As a result, content shared on this community combines typical discussions, analysis and strategies with more unconventional slogans and memes, such as “YOLO” – posts in which the user invested a significant sum into a stock based on little more than intuition – and “Loss”, where users discuss large percentage losses in their investment. The subreddit rules state that gain or losses posts shall present more than \$2,500/\$10,000, while YOLO posts must be a minimum of \$10,000/\$25,000 for options and stocks, respectively.

In parallel to such topics, WSB users also differentiate themselves through specific, often politically incorrect terminologies, mixing text and emojis. This community is considered to be one of the epicentres of the 2021 GME short squeeze [4]. It received worldwide attention, leading to exponential community growth in January 2021. In only a month, the number of users quadrupled, and the number of posts exceeded the entire 2020 count. In the rest of this paper, we discuss the impact of such sudden community growth on users’ communication patterns in the WSB subreddit.

Features	2020	2021
random	0.625 ± 0.051	0.632 ± 0.023
TT	0.577 ± 0.040	0.623 ± 0.020
TP	0.696 ± 0.048	0.714 ± 0.022
TT,TP	0.576 ± 0.0053	0.675 ± 0.023
Feedback		
r_i	0.915 ± 0.033	0.921 ± 0.020
$\log(r_i)$	0.921 ± 0.025	0.924 ± 0.015
$P(R_i < r_i)$	0.913 ± 0.029	0.923 ± 0.014

Table 7: Accuracies of logistic regression models trained with different combination of features. The features are abbreviated as follows: TT: Topic Trend, TP: Topic Preference. Use of all features increase the accuracy, and accuracy is higher for new user group.