Misbehavior and Account Suspension in an Online Financial Communication Platform

Taro Tsuchiya
ttsuchiy@cs.cmu.com
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Alejandro Cuevas
acuevasv@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Thomas Magelinski
tmagelin@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Nicolas Christin
nicolasc@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

ABSTRACT

The expanding accessibility and appeal of investing have attracted millions of new retail investors. As such, investment discussion boards became the *de facto* communities where traders create, disseminate, and discuss investing ideas. These communities, which can provide useful information to support investors, have anecdotally also attracted a wide range of misbehavior – toxicity, spam/fraud, and reputation manipulation. This paper is the first comprehensive analysis of online misbehavior in the context of investment communities. We study TradingView, the largest online communication platform for financial trading. We collect 2.76M user profiles with their corresponding social graphs, 4.2M historical article posts, and 5.3M comments, including information on nearly 4 000 suspended accounts and 17 000 removed comments. Price fluctuations seem to drive abuse across the platform and certain types of assets, such as "meme" stocks, attract disproportionate misbehavior. Suspended user accounts tend to form more closely-knit communities than those formed by non-suspended accounts; and paying accounts are less likely to be suspended than free accounts even when posting similar levels of content violating platform policies. We conclude by offering guidelines on how to adapt content moderation efforts to fit the particularities of online investment communities.

CCS CONCEPTS

• **General and reference** → *Measurement*; • **Security and privacy** → **Social aspects of security and privacy**.

KEYWORDS

Online abuse, Finance, OSN, Trust and safety, Toxicity detection

ACM Reference Format:
Taro Tsuchiya, Alejandro Cuevas, Thomas Magelinski, and Nicolas Christin. 2023. Misbehavior and Account Suspension in an Online Financial Communication Platform. In *Proceedings of the ACM Web Conference 2023 (WWW '23), April 30–May 04, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3543507.3583385

WWW '23, April 30–May 04, 2023, Austin, TX, USA

ACM ISBN 978-1-4503-9416-1/23/04.
https://doi.org/10.1145/3543507.3583385

1 INTRODUCTION

Over the past decade, individual trading behaviors have experienced a marked change. Cryptocurrency and related financial instruments have become important actors in finance [48]. Individual investors have increasingly relied on social media and other online outlets to 1) show off their profits in hopes of becoming "financial influencers," and 2) ask for financial advice from high-performing traders (i.e., social trading [17]). A similar movement has emerged in the traditional stock market, as discussed in online forums such as the `r/wallstreetbets` [43] (WSB) "sub-Reddit." For instance, in 2021, users of WSB self-organized to purchase large numbers of shares from GameStop (GME) in an alleged attempt to "short squeeze" hedge funds. These activities have been facilitated by the rise of user-friendly financial services/apps such as Robinhood, lowering the barrier-to-entry to start trading. However, this increased democratization has been at the expense of a similar increase in online misbehavior. New investors have become the target of various types of attacks, including spam, fraud, and misleading financial advice.

According to cases reported to Federal Trade Commission, from Jan. 2021 to Mar. 2022, fraud starting from ads/messages on social media has reached USD 1.1 Bn, with 40% paid through cryptocurrencies [20]. Despite increasing calls for regulating such malicious behavior, little is known about the types of misbehavior, their prevalence, and potential mitigations.

TradingView[1] is an online platform where traders analyze price charts, post ideas about particular assets, and create trading strategies. It is reportedly the largest trading communication website, with 30M monthly users [51]. The website also presents social features and encourages investors to interact with each other; for instance, users can disclose their personal information, social media handles, and follow other accounts.

Using snowball sampling starting from the official TradingView account, we found information about 2.76M users on the platform and collected more detailed information for over 206 000 active users, including their follower/followee information, historical public articles, and comments they posted. TradingView data present a number of interesting features that allow for an in-depth study of investor behavior. First, the asset symbol is attached to each

[1] https://www.tradingview.com

article, so that we can accurately infer investors' financial interests. Second, these data contain complete information about suspended accounts and comments removed by the moderators, giving us an unprecedented opportunity to analyze online malicious behavior and investigate how people get suspended.

The contributions of the paper are as follows.

(1) Our study is the first to characterize the world's largest financial communication platform by leveraging users' profiles, activity, and financial interests.
(2) We manually classify the types of misbehavior for a sample of removed comments and identify that spam (36%) and toxicity (31%) are the major reasons for removal.
(3) We find that there are certain types of assets that are most often targeted by bad actors and that the platform seems to get more abusive along with market fluctuations.
(4) We show that the suspended accounts tend to form denser communities than regular users, and disproportionately interact with other suspended accounts.
(5) We demonstrate that the number of removed comments, the number of moderated articles, and the difference in registration date, are correlated with account suspension likelihood. Those violations do not equally lead to suspension between free and pro users (those with paid subscriptions).

2 BACKGROUND

We next review relevant related literature and provide the necessary background on TradingView.

2.1 Online communities in finance

The rise of "meme stocks," cryptocurrencies, and ever-decreasing barriers to trading have caused online financial communities to gain immense traction. These online communities have changed the way traders communicate online, and as a consequence, how traders get information and make trading decisions.

The cryptocurrency craze has also made traders a target of various types of scams/online misbehavior, because 1) there is no central authority that monitors the malicious transaction [38], 2) a good level of anonymity favors scammers/criminals [12, 47], 3) the huge price volatility [63] lures naive investors, and 4) a huge price increase incentivizes the attackers with higher profits.

Cryptocurrency traders are particularly active on social media. Pump and dump schemes are one example of cryptocurrency traders using social media to manipulate the market [22, 27, 33, 35, 40, 55, 59]. By coordinating a large number of trades on an asset, the price can be manipulated. This coordination can take place online using social media applications such as Telegram and Discord. Another example that illuminates the use of social media is the "meme coin." For instance, Dogecoin was originally created in 2013 as a form of "community value" [39]. The coin rapidly appreciated in price when Elon Musk made numerous comments about it on Twitter. The relationship between social media and the financial markets is not limited to cryptocurrencies. A Reddit community named `/r/wallstreetbets` which began in 2019 (and currently boasts over 12M users [43]), gained popularity for aggressive trading strategies and bets on "meme stocks." This community was at the center of one of the largest coordinated trading efforts organized online when traders attempted a "short squeeze" on GME. This campaign seemed to increase the overall toxicity in Reddit in Jan. 2021 [29].

2.2 Social networks in finance

Although professional (informed) traders often manage funds on behalf of more inexperienced traders, social websites have recently taken this practice to another level, by allowing users to automatically copy other (presumably successful) traders' strategies. This process is called social trading [17, 57]. Social trading gives us more transparency on the "signal providers" (i.e., those who provide trading strategies), partly because 1) their trading strategy is often public, and 2) they tend to disclose their personal information to be seen as reliable to cover the lack of face-to-face communication. Wohlgemuth et al. [57] confirm that both financial metrics and social metrics (the existence of the profile pictures, names, and social activities) are important to build trust in social trading. At the same time, the low barrier to entry increases the number of charlatans [17] or scam/unauthenticated accounts.

There has been extensive research regarding the impact of social networks in financial markets such as the effect of neighbors/word-of-mouth on stock participation [6], the transaction graphs on the level of volatility in the market [4], the economic impact of the reputation system for portfolio management [24], and the recent work on the social aspects of cryptocurrencies [5, 41]. However, less is known about the social network of malicious actors in finance.

2.3 Platform security

A decade of research on online misbehavior focused on spam [50, 60], Sybil attacks in online platforms [2, 7, 9, 56, 62], manipulation of the reputation/ranking systems [23], financial scams and Ponzi schemes [36, 37, 54], and account suspension [3, 7, 10, 11, 30, 50].

Recently, non-profit-centric online misbehavior has garnered attention, in particular, harassment, hate speech, and toxicity (see details in Section 5.1), but in particular, measurement and user study [29, 44, 49], toxicity detection models [15, 26, 28], and data annotation issues [42, 45]. Social bots [14, 16, 19, 58, 61], or automated accounts posing as humans, are increasingly deployed, leading to an increase in the spread of mis- and disinformation [18, 34, 46] and the lifespan of malicious accounts over the last decade [10].

Compared to normal users, suspended accounts are known to have different social network structures (e.g., retweet structure [30]), attributes (e.g., political ideology), and behaviors (e.g., levels of toxicity and spam [11]). Suspended/Sybil accounts are generally observed to be clustered together, which enables them to boost their credibility [8, 30, 60]. However, Yang et al. [62] have found that the majority are not tightly coupled. We keep this tension in mind in our analysis of suspended accounts in Sections 5 and 6.

2.4 TradingView

TradingView is the largest online communication platform focusing on financial investments. The site provides real-time/historical financial information/news and allows users to technically analyze the data. Users can write opinions about each asset or post trading strategies based on TradingView's programming language, Pine, so

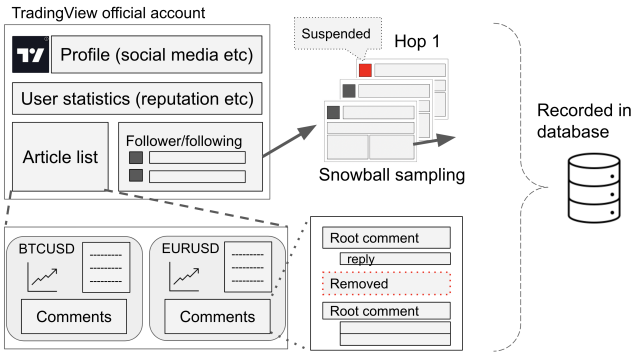

Figure 1: The overall data collection architecture (user profiles through snowball sampling, articles, and comments)

that other users can easily duplicate these strategies. The site does not charge any fee for registration and only requires a valid email address, but offers options to get better access to available resources (i.e., data/technical tools) through "pro" accounts. The subscription costs USD 14.95, 29.95, and 59.95 per month for pro, pro+, and pro premium, respectively. While TradingView host 20 sites to cover major languages, this paper will focus on English. Each user has a page with their activity statistics; TradingView internally calculates the reputation score, based on the number of "likes" received, the reputation of users who "liked" posts, the number of followers, etc.[2] Tradingview differs from brokerage-oriented trading sites (e.g., eToro) [17] in that it focuses on user communication, not on the direct execution of trading itself.

3 DATA COLLECTION

We performed account-based snowball sampling, starting with seed users and recursively expanding through their following and follower relationships [21]. Importantly, for the next hop, we restrict the users who have posted a public article at least once. We define those users as "(socially) active" to infer their financial interests and to facilitate more efficient data collection. Figure 1 describes our data collection architecture. In details, the process is as follows: (1) Set TradingView's official account[3] to be the collection seed; (2) Collect followees and followers[4] using TradingView's APIs (i.e., the APIs that are being used when rendering TradingView's web page), and store their basic information; (3) From the newly collected (active) accounts, consider all previously unseen accounts to be new seed accounts; (4) Return to step 2 twice; (5) For each user, collect all article posts and comments. The dataset collection procedure is robust to the number of collection hops and the selection of seed accounts. (See details in Appendix A.1 about the experiments we have done to explore the possible bias that exists in our dataset.)

We ran steps 1–4 between July 20th and July 28th, 2022 and obtained 2 756 809 users (205 842 active; 2 550 967 inactive). User data include user plan (free/pro users), number of followers, published articles, reputation score, and social media handles (if listed). For active users, the data also include their registration date and the user data for all of their followers and followees.

Article data collection took place from July 29th to August 1st, 2022, and resulted in 4 181 673 article posts. The articles collected were published between Sep. 5th, 2011, and Aug. 8th, 2022.[5] All articles include the asset symbol being discussed, which can be used to infer the financial interests of the user. Comment data collection took place from Aug. 2nd to Aug. 8th, 2022, resulting in 5 273 351 comments, posted between Sep. 6th, 2011 and Aug. 5th, 2022.[6]

To combat malicious actors, TradingView employs a set of house rules [52] enforced by moderators,[7] a mix of volunteers and TradingView staff. A user who violates the house rules receives a temporary or permanent suspension. Suspended users can still access the platform resources covered by pro subscription but cannot interact socially (posting/following/commenting etc.). TradingView claims that suspension decisions are solely based on social activities and are identical regardless of user status (free/pro).[8] The TradingView API returns a "`permanently_suspended`" JSON field, which we use to infer that status. We find 3 981 permanently suspended accounts, which accounts for 1.93% of active users.

While users cannot delete or edit their articles and comments, content moderators can suppress those that violate their policies. The number of moderated articles is obtained for each user by subtracting the number of articles displayed in the following-follower list from the number shown on their profile page. Moderated comments are identified as those which can be obtained through an API query but are not rendered on the HTML page. We find 16 735 comments, or 0.32% of the total, have been removed. 0.58% of posted articles contain at least one removed comment.

4 OVERALL PLATFORM CHARACTERISTICS

We have 206K active users and 2.55M inactive users in total. We define the top users who have the top 5% of the reputation score among active users, corresponding to 0.05% of the total users we observe. We summarize user status (free/pro) in Appendix Table 5.[9] 88% of the users we observe are free users. The ratio of pro/pro+/pro-premium gets higher for active and top users: the more one pays, the more one will engage on the platform. The ratio of suspended accounts for pro users and top users are 0.23% and 3.4%, respectively — to be compared with suspension rates on Twitter ranging from 1.6% [11], to 9.5% [30], depending on the time/domain of the study.

One of the objectives of using TradingView is to gain social status and become a financial influencer. Top users often develop investment consulting services on their external websites by quoting TradingView's reputation score or gain more followers on other platforms by disclosing their social media handles. We summarize user profiles in Appendix Table 5. Among all the users (2.76M), we have 140K (5%) Twitter accounts, 3.7K (0.13%) websites (often a personal website or Telegram/Discord channel), 10K (0.3%) YouTube accounts, and so forth. Active users, in particular top traders, are

[2]https://www.tradingview.com/support/solutions/43000482545-how-s-my-user-reputation-calculated/
[3]https://www.tradingview.com/u/TradingView
[4]We excluded (labeled) brokers due to lack of social features
[5]To prevent users from rewriting their trading history, articles cannot be edited or deleted 15 minutes after they have been posted.
[6]For calculating the likelihood of suspension in Section 6, we only used the data until Jul. 28th, 2022 to be consistent with user data.
[7]https://www.tradingview.com/moderators/
[8]https://www.tradingview.com/support/solutions/43000591357-how-bans-work/
[9]We ignored users on a trial subscription.

more likely to have an in-depth profile by disclosing many social media handles, changing default profile pictures, and writing self-introduction and locations. This finding is consistent with previous work [57] in social trading: disclosing personal information seems to play an essential part in gaining trust, compensating for the lack of face-to-face communication. For those disclosing their Twitter accounts, activity, and popularity on TradingView and Twitter are positively correlated ($r = 0.209$ for the number of followers, $r = 0.095$ between the number of TradingView posts and Tweets)[10].

We now focus on active users' activities. The distribution of the number of followers, the number of articles, and the reputation score for each user all follow long-tailed distributions, seen in other social media platforms [1]. More than 80% of all users have less than 13 followers, 17 charts, and a reputation score of 90. Thus, the platform would be vulnerable to reputation manipulation attacks [23], a practice whereby an account posts simple positive comments to increase its profile statistics. This strategy can also manipulate the popularity of articles. Most articles rarely attract much attention. More than 90% of the articles have less than 11 likes and 9 comments. We observed evidence of reputation manipulation and further discuss in Section 5.3. The cumulative distribution function (CDF) of the above-mentioned variables are in Appendix Figures 9–13.

To illustrate platform evolution over time, Figure 2 describes the number of newly registered *active* accounts per month based on their registration date. The color breakdown denotes the average financial interest of users who entered in the same month.[11] Drastic increases around Dec. 2017 (first spike) and Feb. 2021 (third spike) correspond to historical Bitcoin price spikes, evidenced by the domination of cryptocurrency (Blue) around those times. However, we did not observe an increase in the number around Jun. 2019 or Oct. 2021, when the price of Bitcoin surged again, possibly because most of the investors interested at that time had already joined the platform during the large spikes. Another large spike was seen around March 2020 (the second spike), which can be explained by the start of the Covid-19 pandemic, which may have attracted people to engage in financial trading in their extra spare time [32]. Around this time, "Forex" and "Stocks" seem to show more dominance compared to the first and third hikes. The breakdown of each market (the three small windows on the left side of Figure 2 illustrates these observations. Notably, the stock market seems to be affected by both cryptocurrency bubbles and the pandemic.

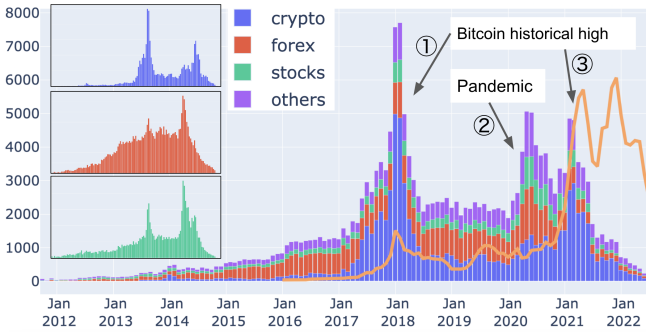

Figure 2: The number of new active users registered (only for active users) per month (Sep. 2011-) along with their financial interests. The three sub-figures on the left side break down each market (The scale of the y-axis is different). The yellow line is the Bitcoin price in USD (1/10 scale)

5 ONLINE MISBEHAVIOR

We next look at online misbehavior, by examining removed comments, the financial assets they pertain to, and by characterizing the network relationships among suspended users.

5.1 Removed comments

To characterize the types of online abuse across removed comments on TradingView, we randomly subsample a set of 500 comments and employ a qualitative coding approach (see details in Appendix A.2) with pre-defined labels. We only select the root (top-level) comments in our subsampling, given that sometimes moderators may remove an abusive top-level comment with all of its replies along with it despite them being benign. Two of the authors independently coded 500 comments (based on content, without context) using the predefined set of categories from Kumar et al's analysis of Reddit toxicity [29]. Additionally, we define a Spam category that captures comments which contain URLs or invitations to other social media platforms (e.g., Whatsapp, Facebook, Telegram etc.). Lastly, an "Undefined" category denotes comments for which the coders could not identify the reason for removal.

After the first pass, coders computed their inter-rater agreement using Cohen's Kappa [13] $\kappa = 0.734$, which indicates substantial agreement. The coders then met to discuss, resolve differences, and ultimately agreed on the following breakdown: 30.8% toxic comments, 35.8% spam, and 33.4% undefined. The breakdown of toxic comments are: Insult (25.2%), Identity Attack (1.8%), Call to Leave (0.2%), Threat (0.6%), Sexual Aggression (0.4%), and Identity Misrepresentation (0.0%). Most of the toxic comments seem to originate as an attack on trading ideas and ability, with some variation on the type of insult (call to leave/suspend or attacking someone's identity based on user profiles). For spam comments, we observed invites to external trading websites or social media. These comments often start as related to the article but quickly proceed to exhort users to visit a URL. A few include blockchain addresses, which seem to be related to phishing attempts. With regards to the Undefined category, the coders found many comments which seemed to be related to reputation manipulation. However, since the coders could not conclusively rule, without context, whether these comments were harvesting reputation, they were labeled as Undefined. Also, the coders also labeled as Undefined comments in a different language, comments that contained political or religious references, and anything else that did not conclusively fall in any of the other categories. See Appendix Table 4 for details.

Recent work [29] has also measured the toxicity of online communities by using state-of-the-art toxicity detection models, such as Google's Perspective API [25]. To understand the usefulness of toxicity detection models in online financial boards at different levels of granularity, we selected 4 sets of comments: 1) a random sample,

[10]The log-transformed attributes were used to account for heavy-tailed distributions.
[11]If a user has written 80% of articles for cryptocurrency and 20% for forex, we added 0.8 and 0.2 for each category, which eventually sums up to 1

Table 1: The use of URL shorteners for removed/normal comments (R: removed, N: normal)

Domain	# of shorteners		Proportion		Likelihood ratio
	R	N	R	N	
is.gd	71	2	0.42%	0.00%	11150.87
bit.ly	88	218	0.53%	0.00%	126.80
tinyurl	18	45	0.11%	0.00%	125.64
goo.gl	16	277	0.10%	0.01%	18.14
invst.ly	3	207	0.02%	0.00%	4.55
all	196	734	1.17%	0.01%	83.88

2) the set of moderator-removed comments, 3) the comments we manually labeled as toxic split into whether the commenter was a pro or 4) free. The mean toxicity was 0.05, 0.25, 0.55, and 0.63 respectively. The detailed results are shown in Appendix Table 6. We find that toxicity on TradingView, as per Google's Perspective API is relatively low. For reference, Kumar et al. employed a threshold of at least 0.8/0.9 to consider comments as toxic in their study of Reddit [29]. The result could be interpreted in two ways. The first interpretation is that TradingView's community is not as toxic as other online boards as our coders perceived. If so, employing automated toxicity detection methods may require calibration using lower thresholds to catch insulting comments. Another possibility is that the current toxic detection algorithm might fail to detect some of the toxic comments in finance and needs to be fine-tuned to this domain. Also, free users appear to be slightly more toxic than pro users, but the difference is unclear given the small sample size.

To make our analysis more quantitative, we compare the number of URLs for removed and non-removed comments. 405 816 out of 5 273 351 (8.0%) comments contain at least one URL for normal comments, 1 949 out of 16 735 (12.0%) comments for removed ones, which is slightly higher than normal comments. To make the difference clearer, we also look at the use of URL shorteners. Following Thomas et al. [50]'s method, we calculate the likelihood ratio $p1/p2$ where $p1 = p(shortener|removed)$ and $p2 = p(shortener|normal)$ for each URL shortener service manually identified in our dataset, as shown in Table 1. We observe that the removed comments are more likely to rely on URL shorteners. The likelihood ratio ($p1/p2$) is $83.88 \gg 1$ overall. Particularly, *is.gd* often shows up in removed comments but rarely appears in regular comments, leading to an extremely high likelihood ratio. The result is consistent with the study [50], conducted a decade ago.

5.2 Misbehavior and financial market

We quantify abuse on the platform through the number of removed comments. We are interested in examining what types of assets have often been the targets of online abuse, based on the symbols that have been attached to all the articles and thus comments. Table 2 describes the ratio of removed comments aggregated by 1) all comments, 2) all root comments, and 3) articles (that contain removed comments), to reflect the prevalence of abuse. We exclude minor financial assets with less than 500 articles in total. Some

Table 2: The top 10 assets based on the ratio of the removed comments, the removed root comments, and the articles that contain removed comments

All comments		Root comments		Articles	
VETUSD	3.28%	ALICEUSDT	2.84%	AMC	2.84%
BSVUSD	2.75%	AAVEUSDT	2.29%	BLX	1.86%
ALICEUSDT	1.79%	ZILUSDT	1.41%	BSVUSD	1.72%
AAVEUSDT	1.43%	BTCUSDTPERP	0.99%	XBT	1.60%
AMC	1.08%	AMC	0.96%	BTCUSD	1.49%
XBT	0.95%	BSVUSD	0.84%	GME	1.42%
BA	0.85%	XBT	0.70%	MRNA	1.39%
ZILUSDT	0.85%	ETHUSDT	0.58%	XRPUSD	1.27%
USDZAR	0.85%	BLX	0.52%	LUNAUSDT	1.16%
AUDCHF	0.72%	FB	0.51%	TRXBTC	1.15%
Avg.	0.32%	Avg.	0.29%	Avg.	0.58%

assets attract a disproportionately large number of malicious activities. For instance, nearly 2.84% of articles on "AMC" contain at least one removed comment, which is five times larger than the average. In general, those highly abusive assets seem to be often featured on social media, including meme stocks ("AMC", "GME"), DeFi/NFT related coins ("AAVE", "ALICEUSDT"), the LUNA stable coin meltdown ("LUNA"), Bitcoin-related ("BTCUSD", "BTCUSDT", "XBT", "BSVUSD") and vaccine-related ("MRNA"). Monitoring assets that trend as "meme assets" could lead to a more efficient content moderation process. Though it is possible that moderators intentionally pick assets that are popular, we observe that many moderated comments are also reported by the users.

Furthermore, we investigate how the abuse of the platform changes over time, since 2016.[12] In Figure 3, the first row is the ratio of articles with removed comments, and the second row is the total number of articles with removed comments. Though the ratio seems to be relatively stable over time, there are spikes around (1) Sep. 23rd to Oct. 21st, 2018 (2) Jul. 4th to Sep. 12th, 2021 (3) Apr. 17th to Jun. 5th, 2022. We manually investigate each spike to come up with explanations. For the first spike, Ripple ("XRP") had a relatively high removal rate; it experienced a huge price turmoil, which may have triggered some of the abusive behavior, but the root cause is still unclear. For the second spike, "AMC" has a remarkably high toxic rate (6.85%), which corresponds to the second price spike from Aug. to Sep. 2021. The third spike can be attributed to the crash in Terra's LUNA/UST, where most cryptocurrencies crashed; "BTCUSDT" and "LUNAUSDT" got particularly abusive. The price fluctuation seems to have some association with the abuse of the platform. The more detailed list of abusive assets around each spike is in Appendix Figure 7.

5.3 Network of suspended accounts

We compare the behavior of suspended accounts to that of non-suspended accounts, informing some of our features in the proceeding suspended account prediction.

First, we look at the suspended accounts' *registration* date to infer when malicious users join the platform, as inspired by Ribeiro

[12]We restrict the time range given the low activity before 2016.

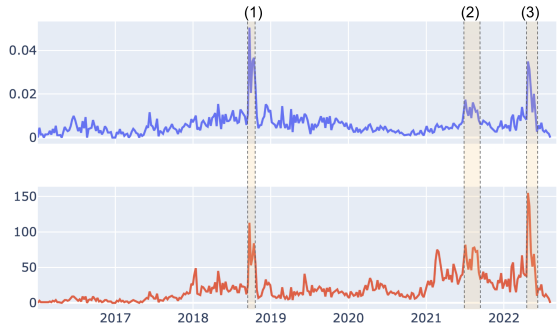

Figure 3: The ratio of removed comments per week (Blue: the ratio of the articles with removed comments, Red: # of articles with removed comments). (1)-(3) corresponds to the spike explained in Appendix Table 7

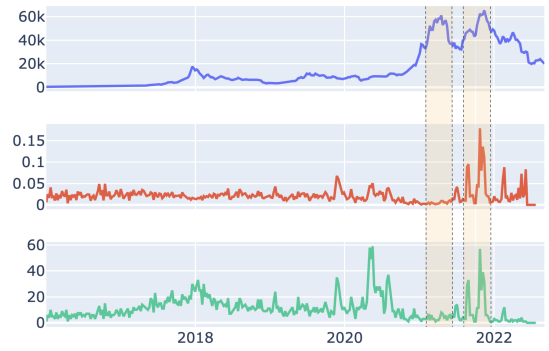

Figure 4: The ratio of the suspended accounts for newly registered accounts (Blue: The price of Bitcoin, Red: The ratio of suspended accounts, Green: # of new suspended accounts.)

et al. [44]. Figure 4 illustrates the ratio of newly registered accounts that will end up being suspended, aligned with the price of Bitcoin. While the ratio was small in the first price spike in early 2021 (first shaded area), we observe a drastic increase in the second price spike (second shaded area). As attack cost/benefits highly depend on the asset price, monetary incentives to attack were stronger immediately after Bitcoin reached a historical high. However, we cannot rule out that those spikes may be impacted by a number of fake accounts that are created together.

Next, we quantify an account's following strategy by calculating the average number of followees' followers (i.e., how popular the set of accounts you follow is). We call this "following-quality" [60]. The following-quality of suspended accounts is lower than that of non-suspended accounts. Normal users tend to follow influencers or highly reputable users with many followers. Suspended users, on the other hand, tend to follow more accounts with few followers. Figure 5 (left) illustrates this behavioral distinction.

After that, we quantify how tightly-knit accounts' following communities are, through their ego-network density. An account's ego network is constructed by adding all the account's neighbors (followers and followees), to a network and then drawing all of the connections between these accounts using *their* followers and followees. The density of the ego network is the number of connections divided by the number of possible connections. An ego-network density of 1 would indicate that all of an account's followers follow each other, forming the most tightly-knit community possible. The distribution of ego-network densities for suspended and non-suspended accounts, excluding users who have less than 5 connections in total, is shown in Figure 5 (right). While the majority of suspended accounts have sparse ego networks just like non-suspended accounts, the upper quartile of suspended accounts forms much denser following communities than those of non-suspended accounts.

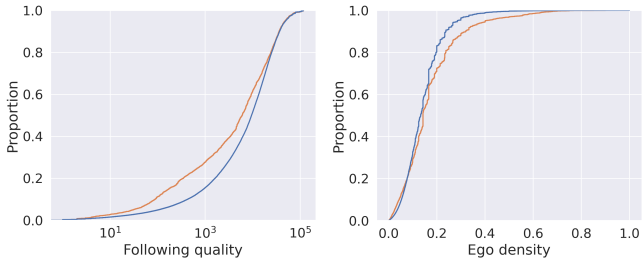

Figure 5: CDF of following-quality and ego density (Blue: normal users, Orange: suspended users)

Furthermore, we investigate the possibility of suspended accounts working together, or in clusters. We first test whether some groups of accounts are created within a short time period of each other, with the purpose of coordinating. We do so by measuring the difference in registration time between pairs of users interacting through comments. The distributions of these interactions are broken down by interaction type (suspended to suspended, suspended to non-suspended, and non-suspended to non-suspended), and are shown in Figure 6. Interactions between accounts registered within less than 6 months of one another disproportionately include suspended accounts. The effect is stronger for those registered within 1 to 3 months of one another. The results are affected by the act of suspension itself; cutting off the lifespan of an account shrinks the maximum potential difference in registration. However, the fact that the results hold for suspended-non-suspended interactions hedges against this effect.

Finally, we show interaction rates between accounts of different categories in Figure 7. For suspended users, 10% of comments are directed towards other suspended users. This is three times the rate seen by non-suspended users. To further investigate this high level of interaction between suspended users, we look into their network of interactions in Figure 7. Nodes represent suspended users, and edges comments between them. Isolate nodes, removed for readability, make up 47.7% of the suspended users. We discover a giant component, where many of the suspended users are interacting, within which tight account clusters exist. We observe a similar behavior in the following network, depicted in Appendix Figure 14. Without knowing the specific reasons for removal, we cannot conclusively determine the activity of each cluster. However, this

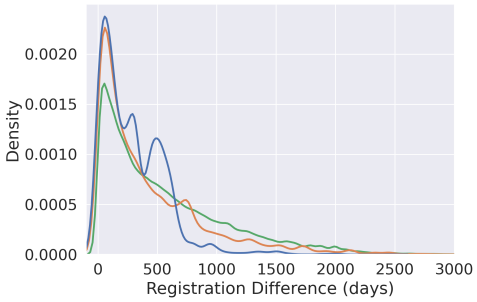

Figure 6: Probability density function (PDF) of comment interactions by the registration time difference (Blue: Suspended-Suspended, Orange: Suspended-Non Suspended, Green: Non Suspended-Non Suspended)

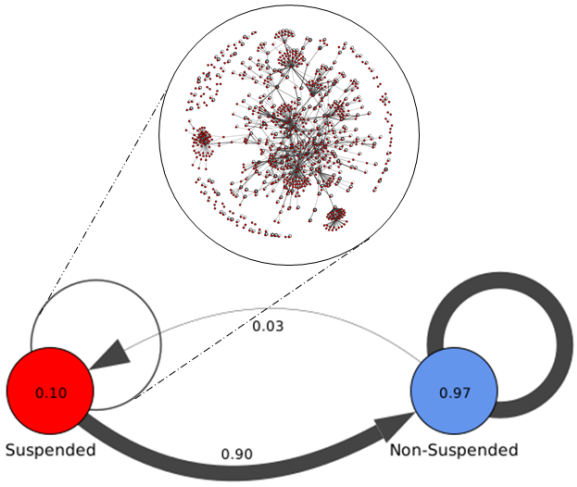

Figure 7: The probabilities of comment-based interaction between suspended and non-suspended accounts (the bottom). The interaction network of suspended users among themselves (top). Isolates have been removed.

behavior is consistent with that of users coordinating to increase their reputation or the popularity of their articles.

Lastly, while the rate of interaction is high between suspended users, 90% of their comments are targeting non-suspended users. While suspended users may be interacting with each other to further their goals, the vast majority of their content targets the general TradingView population.

6 THE MECHANISM OF SUSPENSION

We now examine the relationship between suspension and the number of committed violations. We calculate the probability of getting caught based on the number of removed comments, the number of moderated articles, and the number of follower/followee who have registered within one day of that account, which is a suspicious sign, as described in Section 5.3. We refer to this variable as the number of fake accounts for simplicity, despite some discrepancies between these concepts. In Figure 8, the x-axis is the number of violations, and the y-axis is the ratio of suspended accounts for those with *more than* x violations. The observed upward trend suggests that the more violations an account commits, the more likely they are to be suspended. Benign variables such as the number of articles or comments do not show this trend.

Figure 8: The ratio of suspension for those with more than x violations (Blue: all active accounts, Green: top 5% of reputable users, Red: all pro-users, Pink: pro+/premium users)

In addition, we examine if the probability changes when we narrow our scope on the top users (as defined in Section 4). In Figure 8, the probability of top users (Green) is lower despite having the exact same number of malicious/suspicious activities. The effect seems to be more pronounced for the pro-users. The pro users – pro, pro+, and pro premium (Red) seem to be significantly less likely to be suspended despite the same level of violation. We further restrict the users (Pink) who have pro+ and premium users who pay the higher level of subscription fees, but the ratio does not seem to be significantly different from the group of all the pro users. Also, the change in the ratio appears to increase more rapidly for free users.

Changes in account status after suspension can affect our analysis. For example, suspended pro users have the option to 1) continue using the trading tools/resources as a pro user; 2) stop paying and become a free user; or 3) delete the account. To investigate the impact on our analysis, we continuously monitor user data on the platform and get 306 accounts for which we can confidently identify the time of suspension. We look at status changes before and 30 days after the suspension (because subscriptions start at monthly intervals). Most free users stay free, while some pro users become free users. Specifically, 297 free accounts stayed as free accounts, and only 2 free accounts became pro users after suspension. 8 pro accounts turned into free accounts while 9 pro accounts stayed as a pro. Of those 306 suspended accounts, 17 (5.6%) eventually deleted their accounts. The result implies that the Red/Pink lines in Figure 8 should shift upwards given that the number of suspended pro accounts is underestimated.

Table 3: The result of the logistic regression

variables		coef.	std err	z score	p-value
β_0	constant	-4.1073	0.055	-74.836	0.000
β_1	# of removed comments	0.0353	0.009	3.822	0.000
β_2	# of moderated articles	0.0183	0.001	13.362	0.000
β_3	# of fake accounts (proxy)	0.5051	0.018	28.131	0.000
β_4	All pro users	-2.3388	0.103	-22.726	0.000
β_5	Tier 1 (Top -20%)	0.5484	0.055	9.905	0.000
β_6	Tier 2 (Top 20-40%)	0.3394	0.057	5.978	0.000
β_7	Tier 3 (Top 40-60%)	0.0585	0.058	1.004	0.316
β_8	Tier 4 (Top 60-80%)	-0.3730	0.064	-5.831	0.000
β_9	Registered 2017-2018	0.1943	0.047	4.145	0.000
β_{10}	Registered 2018-2019	0.3217	0.051	6.325	0.000
β_{11}	Registered 2019-2020	0.2483	0.049	5.084	0.000
β_{12}	Registered 2021-	-0.0566	0.064	-0.888	0.374
β_{13}	# of removed comments * pro	-0.0105	0.013	-0.805	0.421
β_{14}	# of moderated articles * pro	-0.0145	0.002	-8.931	0.000
β_{15}	# of fake accounts * pro	-0.4050	0.050	-8.173	0.000

To quantitatively corroborate our argument and remove confounding factors, we construct a logistic regression to predict suspension (4K out of 200K) based on those violations, and the status of the users (free/pro). The model examines 1) if each violation is associated with the account ban and 2) if the effect is ubiquitous across pro/free users by adding interaction variables. For confounding variables, Tier 1-5 (based on reputation scores) and registration year are added to incorporate the level of user activities and partially address the time variance. The baselines of these binary variables are those who are in tier 5 (top 80-100%) and registered before 2017.

The list of variables and the estimated coefficients/significance are summarized in Table 3.[13] We use 1% as a threshold to determine the statistical significance. When we have interaction variables, we have different slopes for free/pro users. For instance, for the number of removed comments, β_1 is the slope for free users while $\beta_1 + \beta_{13}$ is for pro users. For free users, all types of violations are significantly positive ($\beta_{1,2,3}$), indicating that they are good indicators of suspension. For interaction variables, we observe that $\beta_{14,15}$ are negatively significant, meaning that those violations are less associated with pro than free users, although we could not find any evidence for the number of removed comments. This result quantitatively supports that, depending on user status, different amounts of violations lead to suspension, as hinted earlier. For the control variables, in comparison with the baseline (Tier 5), we observe that Tier 1-2 have a positive significant coefficient, which indicates that more activity could lead to a higher rate of suspension, but Tier 4 is significantly negative. It could either imply 1) the effect of activity/popularity may not be linear, or 2) Tier 5 contains more fake accounts, pushing up the suspension rate. The registration year is also significant, except for 2021, showing that the likelihood of suspension differs based on the conditions of the market or the level of content moderation (see Figure 4).

One possible explanation for the observed difference between free/pro is that the moderators may have unconsciously changed the content moderation policy toward paying users. However, one limitation of our analysis is that we do not take into account the severity and prevalence of violations. For instance, one scam comment and one inappropriate word may not equally lead to suspension (i.e., severity). It can also be argued that free users tend to have more severe types of violations (e.g., sending scam comments), leading to more suspensions (i.e., prevalence). Our model does not address this issue since we only look at the number of violations but not at the content of the violations, which remains a future work.

Although Vaidya et al. [53] previously investigated the effect of verified badges on Twitter, none of the literature has investigated the effect of subscription (the existence of payment to the platform) on suspension. This analysis could stimulate discussions of how content moderation should take place when there are different classes of users, and possibly sheds light on a new direction of new research (e.g., pro accounts in Twitch, YouTube, or GitHub, etc.).

Lastly, we implement a machine learning model (Balanced Random Forest) to better predict suspension with more than 30 derived features (e.g., account profiles, financial interests, articles/comments, social networks). However, our model does not seem very effective in the moderation process given the limitations of our data. Appendix A.3 contains more details.

7 POSSIBLE DEFENSES AND LIMITATIONS

Based on our observations, we propose two general strategies to make the platform healthier. First, moderators could pay closer attention to financial assets that have received significant attention on social media platforms since abuse levels might be increasing with social media popularity. Second, one way of easily enhancing the fairness of content moderation policy is to hide the status of the users, which, on the other hand, may decrease the level of satisfaction of being a pro user. For specific types of misbehavior,
(1) *Toxic comments*: The current state-of-the-art method does not capture toxic comments well in the platform. ML methods require calibration of thresholds or fine-tuning to financial text.
(2) *Spam/Fraud*: Automated detection of spam/fraud is especially difficult when spam is manually crafted (by humans) and tailored to article contents. However, greater attention to URL shorteners, particularly from the services we outline, would be beneficial.
(3) *Reputation manipulation*: Account network and interaction metrics, such as ego density, following quality, and difference in registration date, could be useful features in identifying fake accounts.

8 CONCLUSION

While the emergence of cryptocurrencies, `r/wallstreetbets`, and the Covid-19 pandemic expanded the online financial community, toxicity, spam/fraud, and reputation manipulation can have dire consequences for vulnerable retail investors with little experience. Progress has been made on these problems in other domains, however, the nuances of these issues in the financial domain are less well-studied. We conducted the first in-depth study of TradingView, the largest financial communication platform, documented the characteristics of users and platform, and analyzed prevalent misbehaviors. While our results on suspended account behavior and overall misbehavior trends are consistent with previous studies, the effects we observe from account status and from movements in financial markets seem to be unique to the platform, indicating a need to specifically tailor defenses to the online financial sector.

[13]The estimation can be unstable since we are estimating a skewed distribution (only 2% is suspended) with many independent variables. We used two independent libraries (python "statsmodels" and R "glm") to cross-validate results.

ACKNOWLEDGMENTS

We are grateful for the feedback from the anonymous reviewers and from the cryptocurrency research group at Carnegie Mellon University. This research was partially supported by Carnegie Mellon CyLab's Secure Blockchain Initiative, Nakajima Foundation, and ONR (N00014-21-1-2229).

REFERENCES

[1] Yong-Yeol Ahn, Seungyeop Han, Haewoon Kwak, Sue Moon, and Hawoong Jeong. 2007. Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th international conference on World Wide Web*. 835–844.

[2] Muhammad Al-Qurishi, Majed Alrubaian, Sk Md Mizanur Rahman, Atif Alamri, and Mohammad Mehedi Hassan. 2018. A prediction system of Sybil attack in social network using deep-regression model. *Future Generation Computer Systems* 87 (2018), 743–753.

[3] Abdullah Almaatouq, Ahmad Alabdulkareem, Mariam Nouh, Erez Shmueli, Mansour Alsaleh, Vivek K Singh, Abdulrahman Alarifi, Anas Alfaris, and Alex Pentland. 2014. Twitter: who gets caught? observed trends in social micro-blogging spam. In *Proceedings of the 2014 ACM conference on Web science*. 33–41.

[4] Wayne E Baker. 1984. The social structure of a national securities market. *American journal of sociology* 89, 4 (1984), 775–811.

[5] Gianluca Bonifazi, Enrico Corradini, Domenico Ursino, and Luca Virgili. 2021. A Social Network Analysis–based approach to investigate user behaviour during a cryptocurrency speculative bubble. *Journal of Information Science* (2021), 01655515211047428.

[6] Jeffrey R Brown, Zoran Ivković, Paul A Smith, and Scott Weisbenner. 2008. Neighbors matter: Causal community effects and stock market participation. *The Journal of Finance* 63, 3 (2008), 1509–1531.

[7] Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro. 2012. Aiding the detection of fake accounts in large scale social online services. In *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*. 197–210.

[8] Qiang Cao, Xiaowei Yang, Jieqi Yu, and Christopher Palow. 2014. Uncovering large groups of active malicious accounts in online social networks. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. 477–488.

[9] Meta Transparency Center. 2022. Inauthentic behavior. https://transparency.fb.com/policies/community-standards/inauthentic-behavior/. Accessed Sep. 29th, 2022.

[10] Farhan Asif Chowdhury, Lawrence Allen, Mohammad Yousuf, and Abdullah Mueen. 2020. On Twitter purge: a retrospective analysis of suspended users. In *Companion proceedings of the web conference 2020*. 371–378.

[11] Farhan Asif Chowdhury, Dheeman Saha, Md Rashidul Hasan, Koustuv Saha, and Abdullah Mueen. 2021. Examining factors associated with twitter account suspension following the 2020 us presidential election. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 607–612.

[12] Nicolas Christin. 2013. Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd international conference on World Wide Web*. 213–224.

[13] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.

[14] Stefano Cresci. 2020. A decade of social bot detection. *Commun. ACM* 63, 10 (2020), 72–83.

[15] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. In *Proceedings of the international AAAI conference on web and social media*, Vol. 11. 512–515.

[16] Clayton Allen Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. Botornot: A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web*. 273–274.

[17] Philipp Doering, Sascha Neumann, and Stephan Paul. 2015. A primer on social trading networks–institutional aspects and empirical evidenc. In *EFMA annual meetings*.

[18] Don Fallis. 2015. What is disinformation? *Library trends* 63, 3 (2015), 401–426.

[19] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.

[20] Emma Fletcher. 2022. Reports show scammers cashing in on crypto craze. https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2022/06/reports-show-scammers-cashing-crypto-craze#crypto1

[21] Leo A Goodman. 1961. Snowball sampling. *The annals of mathematical statistics* (1961), 148–170.

[22] JT Hamrick, Farhang Rouhi, Arghya Mukherjee, Amir Feder, Neil Gandal, Tyler Moore, and Marie Vasek. 2018. The economics of cryptocurrency pump and dump schemes. *Available at SSRN 3310307* (2018).

[23] Kevin Hoffman, David Zage, and Cristina Nita-Rotaru. 2009. A survey of attack and defense techniques for reputation systems. *ACM Computing Surveys (CSUR)* 42, 1 (2009), 1–31.

[24] Steven Huddart. 1999. Reputation and performance fee effects on portfolio choice by investment advisers. *Journal of financial Markets* 2, 3 (1999), 227–271.

[25] Google Jigsaw. Accessed: Oct. 12th, 2022. Perspective: Using machine learning to reduce toxicity online. https://perspectiveapi.com/.

[26] Mika Juuti, Tommi Gröndahl, Adrian Flanagan, and N. Asokan. 2020. A little goes a long way: Improving toxic language classification despite data scarcity. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, 2991–3009. https://doi.org/10.18653/v1/2020.findings-emnlp.269

[27] Josh Kamps and Bennett Kleinberg. 2018. To the moon: defining and detecting cryptocurrency pump-and-dumps. *Crime Science* 7, 1 (2018), 1–18.

[28] Hyunwoo Kim, Youngjae Yu, Liwei Jiang, Ximing Lu, Daniel Khashabi, Gunhee Kim, Yejin Choi, and Maarten Sap. 2022. ProsocialDialog: A Prosocial Backbone for Conversational Agents. *arXiv preprint arXiv:2205.12688* (2022).

[29] Deepak Kumar, Jeff Hancock, Kurt Thomas, and Zakir Durumeric. 2022. Understanding Longitudinal Behaviors of Toxic Accounts on Reddit. *arXiv preprint arXiv:2209.02533* (2022).

[30] Huyen Le, GR Boynton, Zubair Shafiq, and Padmini Srinivasan. 2019. A postmortem of suspended Twitter accounts in the 2016 US presidential election. In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 258–265.

[31] Guillaume Lemaître, Fernando Nogueira, and Christos K Aridas. 2017. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *The Journal of Machine Learning Research* 18, 1 (2017), 559–563.

[32] Ivan Levingston. 2021. TradingView's $3 Billion Valuation Fed by Retail Investing Boom. https://www.bloomberg.com/news/articles/2021-10-14/tradingview-s-3-billion-valuation-fed-by-retail-investing-boom. Accessed Jan. 31st, 2022.

[33] Tao Li, Donghwa Shin, and Baolian Wang. 2019. Cryptocurrency pump-and-dump schemes. *Available at SSRN 3267041* (2019).

[34] Alice E Marwick and Rebecca Lewis. 2017. Media manipulation and disinformation online. (2017).

[35] Mehrnoosh Mirtaheri, Sami Abu-El-Haija, Fred Morstatter, Greg Ver Steeg, and Aram Galstyan. 2021. Identifying and analyzing cryptocurrency manipulations in social media. *IEEE Transactions on Computational Social Systems* 8, 3 (2021), 607–617.

[36] Tyler Moore and Nicolas Christin. 2013. Beware the middleman: Empirical analysis of Bitcoin-exchange risk. In *International conference on financial cryptography and data security*. Springer, 25–33.

[37] Tyler Moore, Jie Han, and Richard Clayton. 2012. The postmodern Ponzi scheme: Empirical analysis of high-yield investment programs. In *International Conference on financial cryptography and data security*. Springer, 41–56.

[38] Satoshi Nakamoto. 2008. *Bitcoin: A peer-to-peer electronic cash system*. Technical Report.

[39] Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder. 2016. *Bitcoin and cryptocurrency technologies: a comprehensive introduction*. Princeton University Press.

[40] Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti, Stefano Cresci, Maurizio Tesconi, and Emilio Ferrara. 2020. Charting the landscape of online cryptocurrency manipulation. *IEEE Access* 8 (2020), 113230–113245.

[41] Han Woo Park and LEE Youngjoo. 2019. How Are Twitter Activities Related to Top Cryptocurrencies' Performance? Evidence from Social Media Network and Sentiment Analysis. *Drustvena Istrazivanja* 28, 3 (2019).

[42] John Pavlopoulos, Jeffrey Sorensen, Lucas Dixon, Nithum Thain, and Ion Androutsopoulos. 2020. Toxicity detection: Does context really matter? *arXiv preprint arXiv:2006.00998* (2020).

[43] Reddit. Accessed 2022-10-12. /r/wallstreetbets. https://www.reddit.com/r/wallstreetbets/.

[44] Manoel Horta Ribeiro, Pedro H Calais, Yuri A Santos, Virgílio AF Almeida, and Wagner Meira Jr. 2018. Characterizing and detecting hateful users on twitter. In *Twelfth international AAAI conference on web and social media*.

[45] Maarten Sap, Swabha Swayamdipta, Laura Vianna, Xuhui Zhou, Yejin Choi, and Noah A Smith. 2021. Annotators with attitudes: How annotator beliefs and identities bias toxic language detection. *arXiv preprint arXiv:2111.07997* (2021).

[46] Jieun Shin, Lian Jian, Kevin Driscoll, and François Bar. 2018. The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior* 83 (2018), 278–287.

[47] Kyle Soska and Nicolas Christin. 2015. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In *24th USENIX Security Symposium (USENIX Security 15)*. 33–48.

[48] Kyle Soska, Jin-Dong Dong, Alex Khodaverdian, Ariel Zetlin-Jones, Bryan Routledge, and Nicolas Christin. 2021. Towards understanding cryptocurrency derivatives: A case study of BitMEX. In *Proceedings of the 30th Web Conference (WWW'21). Ljubljana, Slovenia (online)*.

[49] Kurt Thomas, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, Zakir Durumeric, Patrick Gage Kelley, Deepak Kumar, et al. 2021. Sok: Hate, harassment, and the changing landscape of online abuse. In *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 247–267.

[50] Kurt Thomas, Chris Grier, Dawn Song, and Vern Paxson. 2011. Suspended accounts in retrospect: an analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. 243–258.

[51] TradingView. 2022. Advertise on TradingView. https://www.tradingview.com/advertising-info/ Accessed Oct. 6th, 2022.

[52] TradingView. 2022. Our House rules. https://www.tradingview.com/support/solutions/43000591638-our-house-rules/. Accessed Sep. 28th, 2022.

[53] Tavish Vaidya, Daniel Votipka, Michelle L Mazurek, and Micah Sherr. 2019. Does being verified make you more credible? Account verification's effect on tweet credibility. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[54] Marie Vasek and Tyler Moore. 2018. Analyzing the Bitcoin Ponzi scheme ecosystem. In *International Conference on Financial Cryptography and Data Security*. Springer, 101–112.

[55] Friedhelm Victor and Tanja Hagemann. 2019. Cryptocurrency pump and dump schemes: Quantification and detection. In *2019 International Conference on Data Mining Workshops (ICDMW)*. IEEE, 244–251.

[56] Bimal Viswanath, Ansley Post, Krishna P Gummadi, and Alan Mislove. 2010. An analysis of social network-based sybil defenses. *ACM SIGCOMM Computer Communication Review* 40, 4 (2010), 363–374.

[57] Veit Wohlgemuth, Elisabeth SC Berger, and Matthias Wenzel. 2016. More than just financial performance: Trusting investors in social trading. *Journal of Business Research* 69, 11 (2016), 4970–4974.

[58] Samuel C Woolley. 2016. Automating power: Social bot interference in global politics. *First Monday* (2016).

[59] Jiahua Xu and Benjamin Livshits. 2019. The anatomy of a cryptocurrency pump-and-dump scheme. In *28th USENIX Security Symposium)*. 1609–1625.

[60] Chao Yang, Robert Harkreader, Jialong Zhang, Seungwon Shin, and Guofei Gu. 2012. Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In *Proceedings of the 21st international conference on World Wide Web*. 71–80.

[61] Kai-Cheng Yang, Onur Varol, Pik-Mai Hui, and Filippo Menczer. 2020. Scalable and generalizable social bot detection through data selection. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 1096–1103.

[62] Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y Zhao, and Yafei Dai. 2014. Uncovering social network sybils in the wild. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 8, 1 (2014), 1–29.

[63] David Yermack. 2015. Is Bitcoin a real currency? An economic appraisal. In *Handbook of digital currency*. Elsevier, 31–43.

A APPENDICES

A.1 Robustness and bias of the data collection

we have performed experiments to check the validity of our data collection. First, we looked at how the number of newly discovered users is diminishingly decreasing over hops. There are 7 841, 109 065, and 89 141 active users for hop 1, 2, 3, respectively. We identified that the possible users we can collect at hop 4 would be 4 075 active users (the followers/followees of step 3 who have not been explored), which is significantly smaller than hop 3 (89 141), indicating that even hop size 3 provides a good sample coverage for active users in TradingView. Second, because the result of snowball sampling is known to rely heavily on the initial/seed users, we have tried the same sampling with a different set of seeds. On behalf of TradingView's official account, we choose the users called "Pine wizards" who have contributed to the development of TradingView's programming language Pine and have been recognized by the platform. We picked all 20 Pine wizards as a seed and collected 209K users (from Aug. 9th to Aug. 18th, 2022). Of those, 98.1% of the users are covered by the original snowball sampling, which validates the choice of our initial seed to some extent. Snowball sampling is the only effective method for collecting TradingView user data at scale since the only other mechanism available to the authors for user discovery is keyword-based searches, which would result in more biased (by keyword selection), and likely less complete coverage. Note that since our data collection tends to focus on users who are active (follow other accounts or make posts), the dataset does not represent the entire population of the platform (30M users), and miss inactive, small and socially isolated communities. However, this would match our intention that we are only interested in the users actively engaging in social trading whom we can make an inference upon, rather than the read-only users. A final remark is that even though it took several days to complete the data collection, we regard the state of the users to be the same (i.e., the dataset is not time-variant). There would be a maximum of 200 hours of time discrepancy between users who are collected at the beginning and the end.

A.2 Qualitative analysis

Table 4 is the breakdown of a random sample (n=500) of comments removed by moderators. The coders noted two main sources of disagreement; 1) coder 1 considered comments which were sarcastic/mocking to be Insults, while Coder 2 considered them to be Undefined, 2) Coder 2 labeled comments which called for a ban or removal to be Call to Leave, while Coder 1 noted them as Insults.

Table 4: The breakdown of removed comments

Category	Coder 1	Coder 2	Final Agreement
Insult	31.6%	23.8%	25.2%
Identity Attack	2.4%	3.5%	1.8%
Call to Leave	0.2%	3.1%	2.8%
Threat	0.4%	0.6%	0.6%
Sexual Aggression	0.4%	0.2%	0.4%
Identity Misrepresentation	0.0%	0.0%	0.0%
Spam	35.6%	35.2%	35.8%
Undefined	29.4%	33.6%	33.4%

A.3 Machine learning to predict suspension

Given the fact that suspended users only accounts are 2% of total active users, we deploy a Balanced Random Forest [31] which changes the sampling method to effectively learn more about a minor class (i.e., suspended accounts). The hyper-parameters are tuned with five times cross-validation in grid-search. The train/test set is split with 80% and 20%, respectively. Because we do not want to miss catching any bad actors, we prioritize having high recall over precision. While we achieved 89% of recall (i.e., we identify nearly 90% of suspended accounts (698 out of 784) as suspended), we have a low precision (around 11%, many false positives). We confirm that, from our data, it would be difficult to maintain the recall precision high enough to implement in practice. The possible explanations for the low performance of our model are as follows; 1) our data does not cover all types of misbehavior; for instance, we did not collect any live-streaming data/chats. Despite that, we think, it would not be a major reason for suspension for many users, 2) there would exist malicious actors who are supposed to be banned but have not been discovered by the platform, increasing the false positives, and 3) we did not take into account the content (texts) of articles/comments, which would be important to understand the gravity of violations.

A.4 Supplementary figures/tables

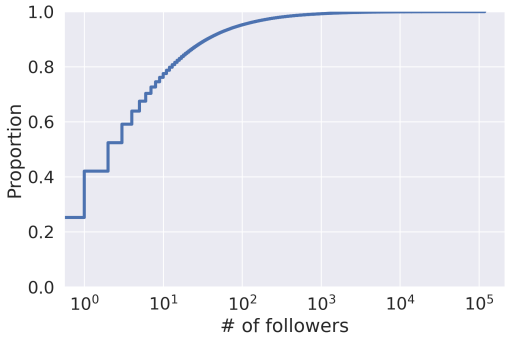

Figure 9: CDF of the number of followers, for active users.

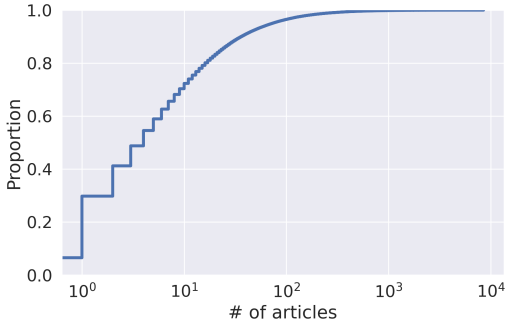

Figure 10: CDF of the number of articles, for active users.

Table 5: The user plan, social media statistics, and basic profiles for each category (all users, active users, and top users)

	free	pro	pro+	premium	Twitter	Website	YouTube	Facebook	Instagram	Default pic	Location	Self-bio	# of users
all users	87.58%	5.75%	3.73%	2.62%	5.08%	0.13%	0.34%	0.03%	0.09%	76.80%	4.22%	2.89%	2756809
active users	78.09%	9.44%	6.21%	6.04%	22.40%	1.23%	2.32%	0.26%	0.63%	50.24%	15.59%	15.73%	205842
top users	68.50%	10.50%	6.98%	13.93%	41.45%	7.55%	7.38%	1.00%	2.20%	17.16%	39.89%	43.86%	10293

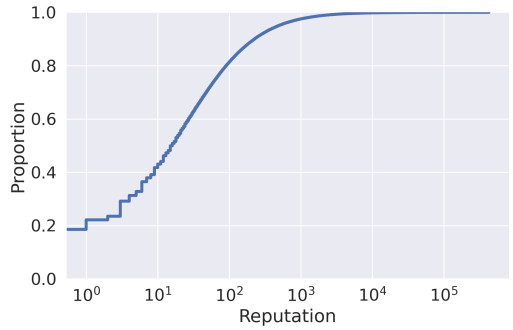

Figure 11: CDF of the reputation scores, for active users.

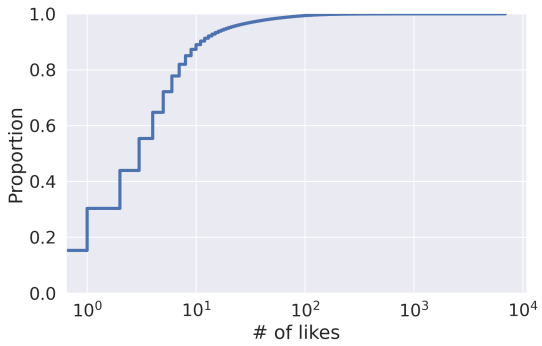

Figure 12: CDF of the number of likes, for each article.

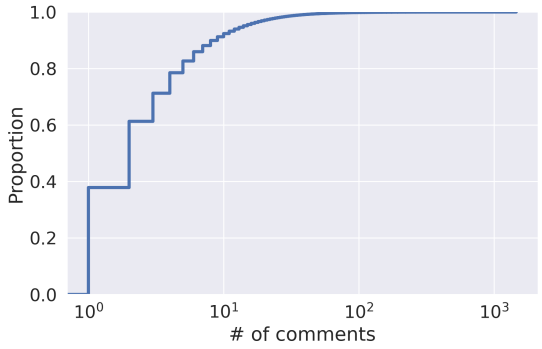

Figure 13: CDF of the number of comments, for articles with at least one comment.

Table 6: Toxicity statistics for various samples of comments using Google's Perspective API.

Comment Group	N=	Mean	Std.
Random Sample of All Comments	46 393	0.05	0.09
comments Removed by Moderators	16 233	0.25	0.27
Toxic comments (by pro users)	35	0.55	0.27
Toxic comments (by free users)	118	0.63	0.26

Table 7: The top 5 assets based on the number of articles with removed comments (the ratio in the blanket) for each spike in Figure 3: (1) Sep. 23rd - Oct. 21st, 2018, (2) Jul. 4th - Sep. 12th, 2021, (3) Apr. 17th - Jun. 5th, 2022.

(1)		(2)		(3)	
BTCUSD	74 (5.00%)	BTCUSD	165 (2.80%)	BTCUSDT	100 (4.10%)
XRPUSD	24 (5.91%)	BTCUSDT	73 (1.74%)	BTCUSD	90 (3.58%)
ETHUSD	10 (3.22%)	XAUUSD	54 (1.90%)	XAUUSD	43 (2.50%)
GBPJPY	7 (4.83%)	AMC	22 (6.85%)	EURUSD	20 (1.83%)
XRPBTC	5 (3.52%)	EURUSD	21 (1.12%)	LUNAUSDT	17 (4.05%)

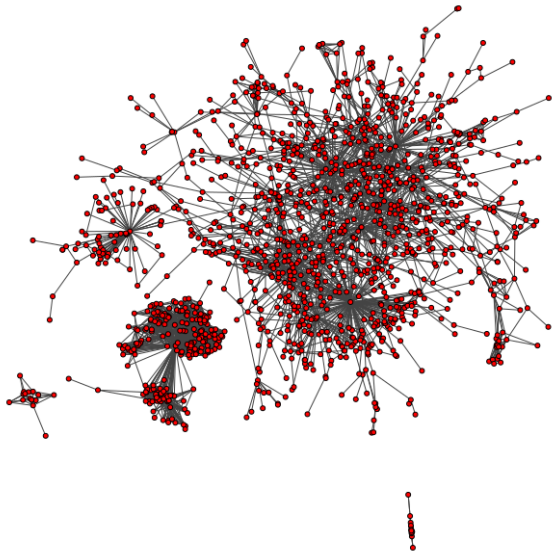

Figure 14: The following network of suspended accounts. Components of 5 accounts or fewer have been removed.