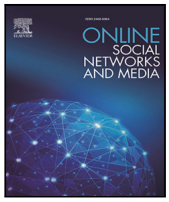
Online Social Networks and Media 37–38 (2023) 100263

Contents lists available at ScienceDirect

Online Social Networks and Media

journal homepage: www.elsevier.com/locate/osnem



Should we agree to disagree about Twitter's bot problem?

Onur Varol

Faculty of Engineering and Natural Sciences, Sabanci University, Turkiye
Center of Excellence in Data Analytics, Sabanci University, Turkiye

ARTICLE INFO

Keywords:
Social media
Twitter
Social bots
Online manipulation

ABSTRACT

Bots, simply defined as accounts controlled by automation, can be used as a weapon for online manipulation and pose a threat to the health of platforms. Researchers have studied online platforms to detect, estimate, and characterize bot accounts. Concerns about the prevalence of bots were raised following Elon Musk's bid to acquire Twitter. In this work, we want to stress that crucial questions need to be answered in order to make a proper estimation and compare different methodologies and definitions based on behaviors and activities; otherwise the real questions concerning the health of online platforms will be confounded by disagreements about definitions and models. We argue how assumptions on bot-likely behavior, the detection approach, and the population inspected can affect the estimation of the percentage of bots on Twitter. Finally, we emphasize the responsibility of platforms to be vigilant, transparent, and unbiased in dealing with threats that may affect their users.

1. Introduction

Social networking platforms provide the ability to communicate through a medium that hosts millions of accounts. Some of these accounts are partially or fully controlled by software to automate content creation and distribution, network structure, and the presentation of online personas [1,2]. The existence of such accounts can be beneficial for certain use cases to foster efficient communication [3,4], effect segregation of user, and eliminate platform biases [5,6]. However, current research focuses primarily on the impact of bots when used by malicious organizations to manipulate public opinion [7–11]. Impact of automated and spam accounts on online platforms can reach wide range of topics concerning public health to politics. To quantify the scale of the problem, bot detection techniques needs to be applied large scale data or statistical methods needs to be employed for estimating automated accounts and their activities. Bot population estimation has implications not only for the field of online manipulation, but also for the valuation of companies, as this can be an important metric for investors. The discussion about the legal battle between Twitter and Musk revolves around Twitter bots and their proliferation on the network. In Fig. 1, we present a timeline of events and show how some of the financial and online metrics change in relation to these events. The first notable event is Musk's announcement of a 9.2% stake in Twitter on April 4. He later made an offer to buy Twitter, offering $54.20 per share. Since the offer and Musk's official filing with SEC, we have observed a series of conversations between Elon Musk and former and current Twitter employees. Musk claimed that Twitter's reports about the bot population on the platform are misleading, and he refused to proceed with the transaction. After several months of debates, Musk purchased Twitter on October 27, 2022. After Musk took over, there has been a change in the way the company is run and its general attitude toward academic research.

We can observe how the current debate affects both stock prices and online activity. For example, Musk's follower count increased much faster in May 2022, when he was seeking funding to buy Twitter. His tweets can also affect stock prices and investor behavior. His concern about bots also coincides with the date that Twitter shares lost significant value. Significant changes in stocks and crypto markets also occur around key events related to the deal, as seen in Fig. 1. Once the Twitter deal was closed by Musk's purchase of Twitter, there was a significant increase in Doge coin price.

In this analysis, we compared Elon Musk's online activities with cryptocurrencies, because at the time of Twitter v. Musk case there were ongoing lawsuits against Musk on cryptomarket manipulation. We believe these two cases are related and recently unfolding events supports our suspicion. On April 3rd, 2023, Elon Musk post a tweet[1] about the change of logo on the platform to Doge coin. Right after this change Dogecoin's value rapidly increased from $0.079 to $0.094, the highest value the currency in the past 6 months. Elon Musk's actions on Twitter, unfortunutely, limits academics to further investigate online cryptomarket manipulations, while further promoting cryptobots to use Twitter. Several researchers investigating post-Musk activities on

E-mail address: onur.varol@sabanciuniv.edu.

[1] https://twitter.com/elonmusk/status/1642962756906418176

https://doi.org/10.1016/j.osnem.2023.100263
Received 4 March 2023; Received in revised form 27 June 2023; Accepted 21 July 2023
Available online 7 August 2023

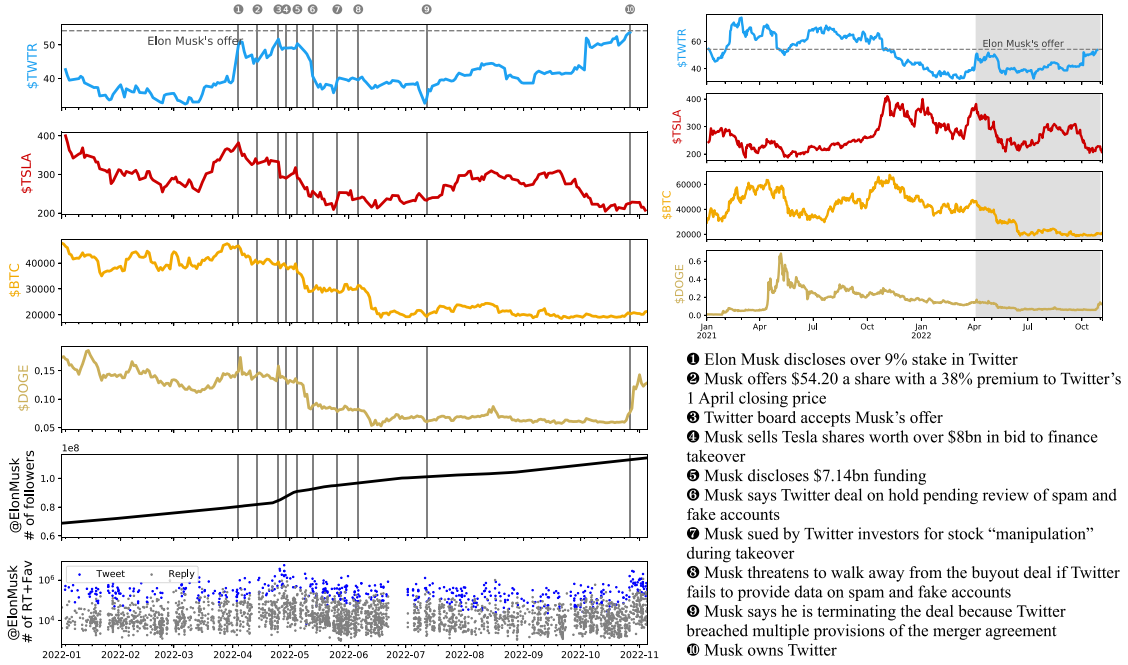


Fig. 1. Timeline of Twitter v. Musk in relation to financial and online metrics. We present how stock prices for Twitter ($TWTR), Tesla ($TSLA), and two cryptocurrencies Bitcoin and DogeCoin ($BTC, $DOGE) changes. Follower counts and engagement metrics for Elon Musk's Twitter account (@ElonMusk) also presented for the last year.

Twitter and how hate speech and bot activities have changed since the takeover [12–14].

Elon Musk's online popularity is reflected in the engagement of his content. Analyzing his recent tweets (retweets, replies and quotes from the last 3,200 tweets excluded), we found that the average engagement is over 240k. His 10 most popular pieces of content before the deal closed past year are all related to Twitter's acquisition. Some of his tweets are jokes about acquisitions. For instance on April 28, he tweeted, "Next I'm buying Coca-Cola to put cocaine back in[2]" which resulted in more than 5.5 million engagements. His second and third tweets with the highest engagement rates were about free speech and reached more than 6.5 million engagements.[3,4]

To make a fair assessment about Elon Musk's claim about bots, we must first critically address three crucial questions: (i) What is a bot? (ii) How many accounts exist? and (iii) How to identify potential risks?

2. Related work

Research on bots stems from the risks that exist, especially in sensitive areas such as politics and health. Researchers have developed machine learning systems to detect these automated entities [15–18] and study their role in information spread [7,8,19]. Our 2017 research is still the largest analysis of bot prevalence on Twitter [20]. We analyzed over 14 million active accounts that use English as their primary language. We identified these accounts based on two criteria: (i) more than 200 tweets in total and (ii) more than 90 tweets in the last three months. By collecting information required by a social bot detection system called Botometer, we estimated 9%–15% of accounts that exhibited bot-like behavior. Research conducted in 2017 is still the largest analysis of bot prevalence on Twitter [20]. By collecting information required by a social bot detection system called Botometer, they estimated 9%–15% of accounts that exhibited bot-like behavior.

One of the most important issues is to find an appropriate and commonly agreed definition for bot accounts. "What is a bot account?" has been an important research question. To make comparisons across datasets and platforms, we must first need to establish on definitions. Social media companies, researchers, and journalists may have overlapping but slightly different definitions for bot accounts. These slight differences can lead to disagreements about the extent of the bot problem on these platforms.

2.1. How Twitter defines bots?

Since Twitter is at the center of all debates about bots, their definition of bot is important. The company has stated in the past that it is actively fighting online manipulation, coordinated activity, and malicious bot accounts.[5] They are also proactive in initially vetting new accounts, and recently Twitter CEO Parag Agrawal stated that Twitter suspends over half a million accounts a day before they impact the platform.[6] These examples highlight the company's efforts to create a healthier platform for all users.

Although Twitter struggles with malicious automated accounts, they also help developers create automated accounts by providing an interface for marketing and other engagement activities.[7] Twitter has also taken steps to collect data on self-identified bots. The activities and behaviors of these accounts can be classified as automated/bot activities by detection tools and recognized as bot accounts by humans. Twitter's own reporting also focus on behavior of accounts as they are being used to spam or not. There can be legitimate accounts owned by humans that are used for spam.

Twitter's efforts to identify accounts that *cause harm* might overshadow accounts that are compromised by 3rd-party applications, created as dormant account to be used in coordinated activities, and used as deceptive in the context that is not prioritized by the company.

[2] https://twitter.com/elonmusk/status/1519480761749016577
[3] https://twitter.com/elonmusk/status/1518623997054918657
[4] https://twitter.com/elonmusk/status/1518677066325053441
[5] https://help.twitter.com/en/resources/addressing-misleading-info
[6] https://twitter.com/paraga/status/1526237583058952192
[7] https://developer.twitter.com/en/docs/tutorials/how-to-create-a-twitter-bot-with-twitter-api-v2

2.2. How bot detection systems define bots?

Bot detection systems focus on detecting irregularities in account behavior. One of the most popular tools, Botometer, defines bots as a social media account that is at least partially controlled by software [1,17,20,21]. This system extracts signals from profile, content, temporal, and network information to provide continuous score of bot-likeliness [20,22]. In the latest version [17], researchers have developed a system that evaluates each account for different scenarios such as spammers, fake followers, financial, etc. to cover different bot behaviors in the system.

Another popular tool, BotSentinel, defines inauthentic accounts as "nefarious individuals pretending to be something they are not to deceive their followers and audience, or automated accounts (bots) developed to behave as humanly as possible with the intent to deceive".[8] In this system, automation is part of the equation, and they also focus on account intent. They track "disruptive" accounts that frequently harass other accounts and use offensive language. Similarly, "problematic" accounts that frequently target other accounts and frequently use malicious tactics to harass their targets are evaluated by this system.

A recent study compares the predictions of two bot detection systems and an account labeling heuristic and shows limited overlap between these approaches, while also showing significant dependence on the dataset in estimating the prevalence of bot accounts [23]. Differences between systems are partly due to different definitions of automated accounts and quality of available datasets. Machine learning approaches developed for model generalizability addresses criticism about false positive problem of detection systems [17]. Recent work by Cresci et al. addresses various misconception about social bot detection research [24].

Research activities and tools developed for bot detection focus on behavioral features to isolate organic behaviors from others. Current approaches detect account-level anomalies; however, recent efforts have also focused on identifying coordinated activity as bot-like behavior has become increasingly difficult to detect in recent years [15,25].

2.3. How journalists talk about bots?

Especially after the current events, journalists have turned to researchers for an expert opinion on Twitter bots. Because journalists play an important role in public understanding of the issue, how they convey the consequences of bots and define automated activity matters.

One of the most comprehensive research on bots conducted by PEW research center [26,27]. The researchers analyzed over 1.2 million links shared on Twitter and identified the most popular links they examined. They found that bots were responsible for 66% of tweeted links to sites focused on news and current events and 89% of links to popular news aggregation sites [26]. In another study, a survey of 4,581 U.S. adults was conducted to determine public perceptions and concerns about bots. About two-thirds of this representative sample have heard of social media bots, and of those, 66% believe social media bots have a mostly negative impact on how well they are informed about current events, while only 11% believe they have a mostly positive impact [27].

Journalists were among the first professions to adopt the use of social media for their work [28]. They enrolled for Twitter's verification platform, use Periscope for live streaming, and share the breaking news on the platform while engaging with their followers as digital journalism rapidly grows and changes journalism practices [19,29,30].

Journalists' selection of newsworthy material influences how automated accounts are presented. Aside from technology news, bots are usually discussed in a political context, focusing on the negative aspects of their use. In this context, bots usually spread misinformation and manipulate public opinion, so journalists tend to portray bots as malicious automated accounts.

Overall, it should be noted that most of these definitions are slightly biased towards bots being malicious entities that are controlled for nefarious activities. However, bots can also be used to share useful information and news [31,32]; as well as collaborate with accounts to provide useful services [3,33–36].

3. Dataset and methods

In this work, we analyzed a publicly available dataset of Twitter from the Internet Web Archive.[9] We analyzed tweets from a year before the acquisition to capture deletion statistics. We selected an 80-day interval starting in June 2021 and processed over 243 million tweets and accounts posting those tweets. Collections from Internet Web Archieve contains API responses that follows standards of Twitter API v1, which captures tweet meta-data such as time, text, and tweet entities as well as user profile that contains account creation time, friend and follower counts, and meta-data about profile descriptions etc.

We processed tweets collected through the Twitter API and we count (i) the number of unique tweeting accounts, (ii) the number of unique interacted accounts (mentioned, retweeted, quoted, and replied), and (iii) the deletion statistics for those accounts. Since this work conducted prior to Twitter's aggressive activities towards limiting academic access, we could capture detailed information about accounts and their activities, which otherwise would not be possible. To collect user profile information and deletion status, we used Twitter API's `users/lookup` endpoint.

To answer the question "What is a bot?" we should focus on automated activities and coordination between accounts. Since the boundaries between automated and organic accounts are not clear, we should consider the bot-likeliness as a spectrum. There are various approach proposed in the literature and one of the most widely-used solution is called Botometer [37]. Since its inception Botometer released four different versions to adapt changes due to Twitter API and improve the system performance with additional datasets and machine learning models.

In this work, we use light-weighted model of Botometer, called BotometerLite [37]. This model requires profile metadata and the time of the tweet for analysis. Since these features are readily available from streaming data, it is a scalable alternative to the Botometer model. We collected over 243 million tweets during our observation period, and there were over 53 million unique accounts that posted this content. Since an account can post more than one tweet during the observation period, we may calculated bot scores for these accounts at different time. In our dataset, the number of observations for a user follows a heavy-tailed distribution, where more than 10 million accounts have a single observation, while there are hundreds of accounts with over 100 tweets. We analyzed bot scores of accounts with more than 30 tweets and presented how bot scores and their variance are distributed in Fig. 2. Bot detection systems internal validity can be interpreted with variance of bot scores. We observe standard deviation of bot scores rarely exceeds 0.1 and mostly close to zero, making this system a reliable tool to study accounts overtime.

[8] https://help.botsentinel.com/support/solutions/articles/64000108232-what-are-inauthentic-accounts-

[9] https://archive.org/search.php?query=collection%3Atwitterstream&sort=-publicdate

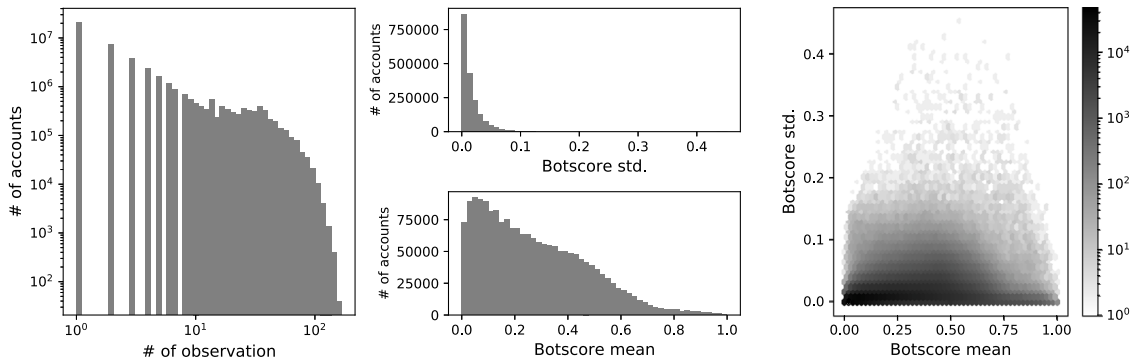


Fig. 2. Analysis of bot scores captured by individual tweets. Observation frequency for accounts follows heavy-tailed distributions (left). Accounts with more than 30 tweets during the observation period are analyzed for their bot scores and variability of scores (middle). Accounts with more confident classification results show a lower variance of predictions (right).

4. Estimating number of accounts on Twitter

One question we want to answer is "How many accounts exists on Twitter?" The answer to this question may change depending on the methodology used to identify and count each follower. Researchers who have API access can capture unique users using two main mechanisms:

Tweet activity: Twitter offers API endpoints to provide public tweets. People with developer accounts can stream about 1% of all public tweets. Accounts with an elevated access can retrieve nearly 10% of public tweets. Using this streamed data, one can analyze all tweets and collect unique users to count the number of social media accounts. This approach can be problematic because (i) there may be accounts that never post or are in a dormant state, (ii) Twitter may delete or suspend some accounts since the last observed tweet, and (iii) the activities of dormant accounts can be controlled by 3rd-party applications on the platform.

Network connectivity: Even though it is tedious, one can collect the friends and followers of each account to exhaustively discover underlying social network. Since there may be some disconnected components, the initial seed selection for collection is strategically important. Since deleted and suspended accounts also lose their social ties, this approach can provide an accurate picture of accounts that have at least one friend or follower. There may be coordinated bot armies that never tweet and only follow each other. These accounts are virtually invisible to all methods available to users of the Twitter API. These dormant accounts, which can be activated for coordinated action at critical times such as elections or wars, pose a major risk that can fuel misinformation. We argue that platforms should be more considerate in removing dormant accounts, as it is difficult to estimate the problems they cause.

These two approaches rely on the Twitter API, which is offered in various forms, such as a developer API, a data stream in the form of Gardenhose (~10%), and a recently introduced Academic access. Even a heroic effort of over 80 researchers can only capture a single day of almost complete activity when they combine their application keys [38]. The limitations of these streams have been discussed in recent literature [39–41]. Research efforts that aim to count the number of users on platforms should be aware of these limitations and the sampling strategies engineered by Twitter. Samples collected via APIs are not statistically correct "random samples" but filtered tweets based on tweet IDs.

Twitter has records of every account ever created, making it the only source that can accurately answer the question, "How many accounts are there on Twitter?" Twitter's quarterly reports to SEC (U.S. Securities and Exchange Commission) provide information on monetizable daily active usage (mDAU). Twitter defines mDAU as individuals, organizations, or other accounts that have logged in or otherwise been authenticated and accessed Twitter through its website on a given day. In its latest filing, Twitter reports 237.8 million average mDAU as of June 30, 2022 [42]. However, we could not find information on the total number of registered accounts on Twitter, whether or not they are monthly active during the current observation period. Twitter's mDAU metric is useful for evaluating active accounts within a given time period, but ignores dormant accounts that may tweet at any time in the future through automation.

Estimating the prevalence of bots in terms of percentiles requires not only the number of bot accounts on the platform, but also the total number of users. In Fig. 3, we illustrate a space of social media accounts and how they can be observed through different collection methods. Above, we discussed the limitations of estimating the total number of accounts using different approaches. Twitter also only considers "active" accounts and acknowledges a potential underestimate in its reporting, as it reports in SEC filing as "our estimation of false or spam accounts may not accurately represent the actual number of such accounts, and the actual number of false or spam accounts could be higher than we have estimated [42]". In this analysis, Twitter estimates that 5% of mDAU accounts are bot accounts; however, this analysis does not take into account accounts that are not active at the time of the analysis but pose a potential threat to the health of the platform.

To exhaustively account for the number of unique users on Twitter, we analyzed the daily tweets from the streams. This included not only the tweeting accounts, but also all other accounts that participated in these tweets by retweeting them, replying to them, or being mentioned in these tweets. In Fig. 4(top), we show the cumulative number of unique accounts and also how many of them are still active in 2022. On average, there are two millions active accounts daily and about 500,000 accounts posting on these days are no longer active in the Twitter stream.

Our analysis of streaming data can also answer the question of how many days to observe the Twitter stream to capture as many users as possible. We tracked the cumulative number of unique users and calculated the derivative of the change for each day in Fig. 4(bottom). In calculating the robust estimate for the number of novel accounts observed on each day, we identified about 350,000 accounts in the stream, with the first 15 days having significantly more observations because user activity rates vary and some users tweet less frequently than others. We can suggest that by observing Twitter stream for 15 days, one can capture significant portion of the active accounts on Twitter.

To offer estimates for bot-likely population, we utilize light-weighted model of Botometer, called BotometerLite, which can works with historical data and scalable since it only requires the profile meta-data [37]. In Fig. 5, we compared different assumptions to build populations and we estimated bot scores and compared them available estimates for bot prevalence.

When we analyzed the dataset, 11.3% of accounts were deleted within a year. We calculated bot scores for active accounts using

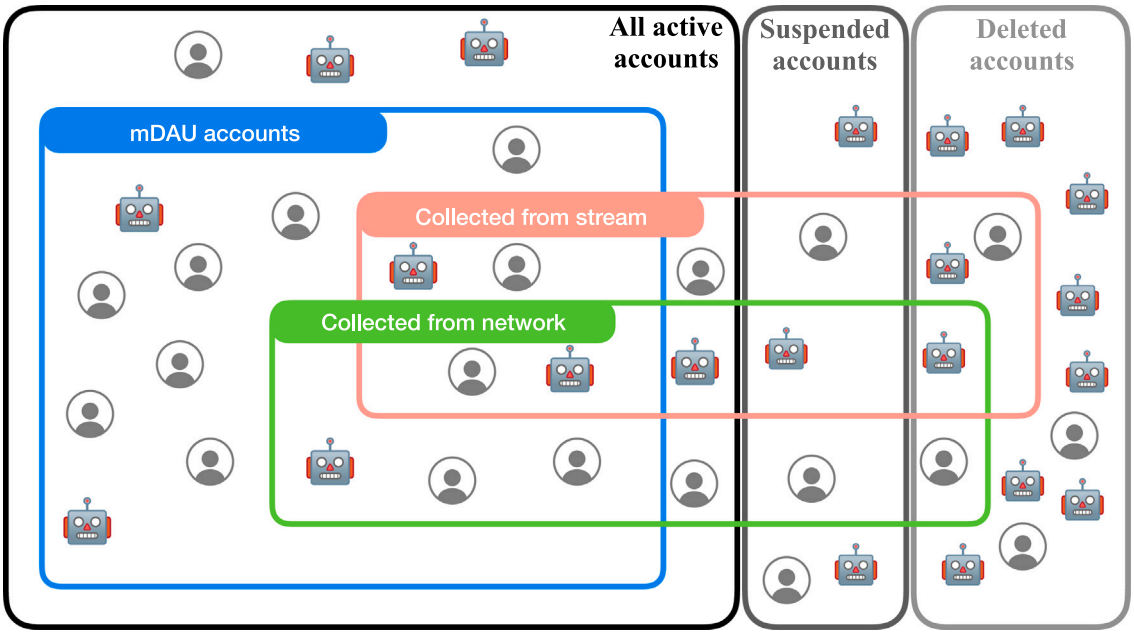


Fig. 3. Properly estimating population of Twitter users require agreement on definitions of relevant accounts and data collection strategies.

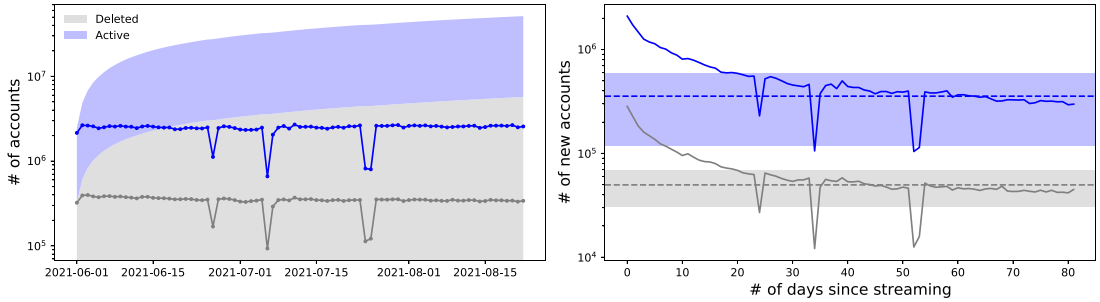


Fig. 4. Analysis of tweet stream for gathering unique number of accounts that are currently active or deleted. On average there are 200,000 unique accounts that are still active (upper). By observing data streams for more than 15 days, number of additional new accounts saturates (lower).

current profile information. These active accounts have 8.6% bot-likely population. However, this could be an underestimate since some of the bot accounts were detected and suspended by Twitter in that period. If we assume that all deleted accounts are bots, this results in a higher and likely overestimated bot prevalence of about 20%. To address this issue, we used historical data at the time of the tweet's creation to compute bot scores, and this approach resulted in an estimate of 16.5% of daily active accounts being bot-likely.

Twitter's mDAU population covers not only active tweeting users but also passive accounts that logins to platform for solely consuming content. This large pool of accounts might lead to lower estimate of 5% indicated by Twitter. Twitter's definition of automated accounts also focus on spam and this may not capture more sophisticated automated behavior. *Varol et al.* estimated bot prevalence in the platform ranging between 9% and 15%. This estimation based on a large-scale study conducted on English speaking active users and Botometer V3 model was used in that paper. Our fundings aligns with the observations in this work and the deletion statistics in the past one year.

An alternative interpretation of this result may focus on deletion statistics, as the difference attributable to deleted accounts is 11% of accounts, which is also double Twitter's own estimate. Since inactive accounts remain on the platform and are not deleted by either account owners or Twitter, these deletion statistics may be the result of Twitter's operations due to user agreement.

5. Discussions

Bot detection has been an active research area since 2010 [2]. Considering the developments in deep learning technologies, we can observe profile pictures created by deepfake technologies [43,44] and realistic messages through conversational bots [45,46]. These technologies lead to the development of novel automated behaviors and enrich the ways of interacting with the platform. Current approaches focus on models that can identify different types of bots and consider the bot detection problem as a binary classification between human and bot accounts [17,37].

These developments compel the arms race between bot detection and bot developers. As the boundaries between automated behavior and organic activity become more blurred, we should find ways to detect user intent and identify early warning signals of coordinated behaviors. There are still gray areas that require further research and policy recommendations for social media companies. Although these questions are not trivial to answer, they pose a risk if malicious organizations exploit them. Recent research also studies to what extend AI technologies can be used on social media campaigns [47,48]

How to treat a suspicious but inactive account? Although Twitter's mDAU reaches 230 million, some accounts on the platform may be inactive. Accounts with no posts or social network do not provide sufficient information to determine their intentions or potential for harm. However, there are companies that sell aged accounts, and these accounts can be repurposed for various campaigns. The right to be

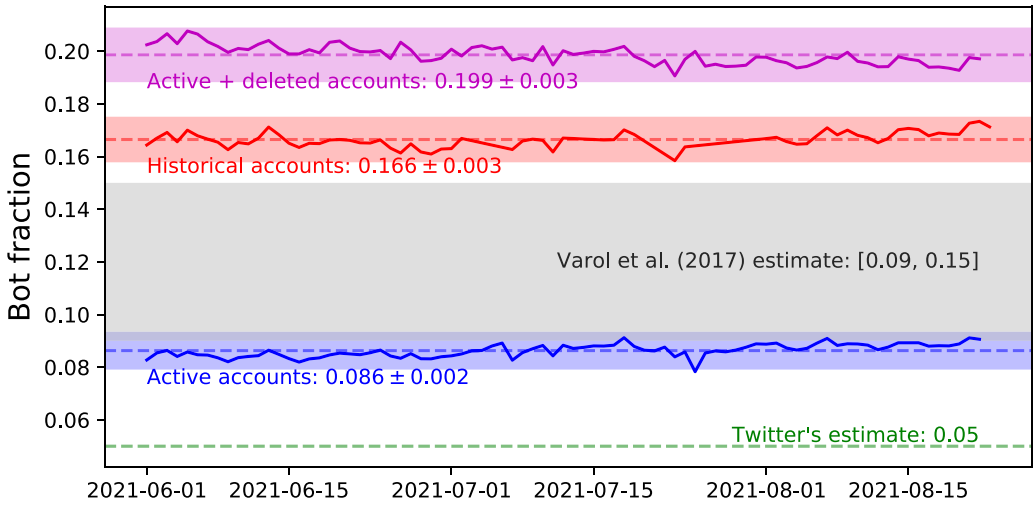


Fig. 5. Estimating fraction of bots accounts over time. Active accounts (blue) consider only set off accounts still available on the platform. We can also use profile information of deleted accounts and conduct historical analysis (red) or we can make assumption that may lead overestimation by considering all deleted accounts being bot (pink). There are also two other estimates by Twitter for spam accounts (green) and Varol et al. (2017). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

forgotten laws protect such accounts, as they periodically delete all posts to remove traces of their automated activities. The platforms nudge accounts and periodically suspend inactive accounts, but they rarely check to see if a particular account changes ownership. These accounts, for which there is no clear evidence that it is a bot account, are considered human.

How to regulate 3rd-party apps? Twitter and many other social media platforms offer developers the ability to create applications that can automate tasks such as content creation, social interactions and private messaging. These applications manages permissions for thousands of users, and once compromised, they can be a tool for coordinated activity that controls content creation and dissemination. In 2017, there was a coordinated attack in which political messages were posted from thousands of accounts, some of which are popular accounts such as Forbes and BBC America.[10] It is alleged that the Twitter Counter application was compromised and all accounts that provide their permissions to this application were used for the attack.

Should Twitter act in benign looking bot armies? Intentions are important in Twitter's definitions of being a bot accounts. The health of the platform takes precedence when it comes to suspending accounts from the platform, but the existence of coordinated bot armies still poses a problem because they can be used for coordinated campaigns. The researchers identified bot armies that simply post quotes from Star Wars [49]. They found that most of these accounts were active and systematically created on a large scale.

Twitter collects valuable information about how accounts access the platform, such as IP addresses and timestamps, as well as their latent interaction with content, such as impressions and clicks. These can be used to identify automated behavior, as manipulating them is more challenging than controlling their API. That said, not all bot-like behaviors are problematic, and Twitter supports developers in creating their own bots to automate certain activities.

6. Conclusion

The activities of bots are an important problem for maintaining the health of the platform. Social media companies use various mechanisms to address concerns about the use of automation for spamming or malicious intent. These concerns were also raised when Elon Musk made bidding to acquire Twitter. The debate between Musk and Twitter is not an easy one to resolve, because the reported numbers for bot accounts, the methods used for bot detection, and the user populations studied vary quite significantly.

We motivate this work to address important questions about platform-wide bot analysis and the challenges of estimating bot prevalence. "What is a bot?" is the first important question, as the use of automation and the intentions of accounts are difficult to disentangle. If only the bot accounts with bad intentions are considered, there is a risk to the health of the platform as the behavior of the accounts may change or they may be repurposed. Establishing what counts as a bot account will allow them to be identified and counted; however, there is more than one way to access user data on Twitter, which raises the second important question, "Which accounts should be analyzed?" Historical collection data from the stream may include deleted accounts and is not representative of dormant accounts, while network collection and mentioned accounts may capture them. Finally, there are gray areas where the platform can be proactive. We have to ask, "How to identify potential risks?" because vulnerabilities can be exploited and have consequences for platform.

Depending on the topic or type of social conversation, we can find different assessments of bots in the literature. For example, politics is one of the most sensitive topics, as the use of bots can manipulate political debates and the credibility of politicians [50–55]. The use of bots was also investigated for the current public health threat that spread misinformation about vaccines during the COVID -19 pandemic [56–58]. Because bot prevalence can depend on context and risk of manipulation, it is important to compare it in the same context. The BotAmp[11] application uses BotometerLite to provide a scalable tool for comparing different sets of online conversations [37].

Despite technology companies' efforts to fight misinformation and the use of automation to manipulate public opinion, users should also take responsibility and prioritize learning media literacy. Community-driven efforts are valuable to engage users in the process. Twitter announced Birdwatch, a community-driven approach to misinformation, as a pilot for users from the United States in early 2021.[12] This platform allows approved users to flag suspicious content and provide additional notes. Although this system is a great initiative to combat

[10] https://www.theguardian.com/technology/2017/mar/15/twitter-turkey-accounts-hack-tweet-swastikas-pro-erdogan

[11] https://botometer.osome.iu.edu/botamp/

[12] https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation

misinformation, researchers point out pitfalls and opportunities to improve this community-driven mechanism [59–61]. A similar approach can be introduced for account-level annotations, flagging bots, trolls, sockpuppets, etc. This would be a step toward creating more transparent data requested by researchers [62]. Community-driven efforts to label accounts can be used to initiate challenges to detect social bots [63].

Despite the best intentions of academic researchers, platforms are reluctant to share data with researchers. When Elon Musk bid for Twitter, he claimed the reason for his interest was to make Twitter a secure and bot-free platform. However, his initial actions were aimed at limiting research efforts to detect misinformation and automated accounts, and incentivizing paying users to use automation.

As a concluding remark, we believe this work may inspire future research questions for investigating platforms and their decision makers. It is also important to remember platforms with lack of transparency may become more vulnerable to internal and external efforts to manipulate online discourse. Providing data access to researchers can benefit user experience and health of the platforms.

CRediT authorship contribution statement

Onur Varol: Conceptualization, Data curation, Methodology, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Onur Varol reports financial support was provided by Scientific and Technological Research Council of Turkey.

Data availability

Data will be made available on request.

Acknowledgments

I thank the Indiana University OSoMe team for fruitful conversations over the years and Aziz Simsir for his guidance in learning details about acquisitions. This work was supported in part by the TUBITAK Grant (121C220).

References

[1] E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, The rise of social bots, Commun. ACM 59 (7) (2016) 96–104.
[2] S. Cresci, A decade of social bot detection, Commun. ACM 63 (10) (2020) 72–83.
[3] B. Mønsted, P. Sapieżyński, E. Ferrara, S. Lehmann, Evidence of complex contagion of information in social media: An experiment using Twitter bots, PLoS One 12 (9) (2017) e0184148.
[4] G. Pennycook, J. McPhetres, Y. Zhang, J.G. Lu, D.G. Rand, Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention, Psychol. Sci. 31 (7) (2020) 770–780.
[5] X. Wang, O. Varol, T. Eliassi-Rad, Information access equality on generative models of complex networks, Appl. Netw. Sci. 7 (1) (2022) 1–20.
[6] D. Freelon, M. Bossetta, C. Wells, J. Lukito, Y. Xia, K. Adams, Black trolls matter: Racial and ideological asymmetries in social media disinformation, Soc. Sci. Comput. Rev. 40 (3) (2022) 560–578.
[7] C. Shao, G.L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, F. Menczer, The spread of low-credibility content by social bots, Nat. Commun. 9 (1) (2018) 1–9.
[8] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, science 359 (6380) (2018) 1146–1151.
[9] D.M. Lazer, M.A. Baum, Y. Benkler, A.J. Berinsky, K.M. Greenhill, F. Menczer, M.J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, et al., The science of fake news, Science 359 (6380) (2018) 1094–1096.
[10] O. Varol, I. Uluturk, Deception strategies and threats for online discussions, First Monday (2018).
[11] K. Starbird, Disinformation's spread: Bots, trolls and all of us, Nature 571 (7766) (2019) 449–450.
[12] B. Benton, J.-A. Choi, Y. Luo, K. Green, Hate Speech Spikes on Twitter After Elon Musk Acquires the Platform, School of Communication and Media, Montclair State University, 2022.
[13] R. Ray, J. Anyanwu, Why is Elon Musks Twitter takeover increasing hate speech? 2022.
[14] D. Hickey, M. Schmitz, D. Fessler, P.E. Smaldino, G. Muric, K. Burghardt, Auditing Elon Musk's impact on hate speech and bots, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 17, 2023, pp. 1133–1137.
[15] N. Chavoshi, H. Hamooni, A. Mueen, Debot: Twitter bot detection via warped correlation, in: Icdm, vol. 18, 2016, pp. 28–65.
[16] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, M. Tesconi, DNA-inspired online behavioral modeling and its application to spambot detection, IEEE Intell. Syst. 31 (5) (2016) 58–64.
[17] M. Sayyadiharikandeh, O. Varol, K.-C. Yang, A. Flammini, F. Menczer, Detection of novel social bots by ensembles of specialized classifiers, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 2725–2732.
[18] L.H.X. Ng, K.M. Carley, BotBuster: Multi-platform bot detection using a mixture of experts, 2022, arXiv preprint arXiv:2207.13658.
[19] O. Varol, I. Uluturk, Journalists on Twitter: Self-branding, audiences, and involvement of bots, J. Comput. Soc. Sci. 3 (1) (2020) 83–101.
[20] O. Varol, E. Ferrara, C. Davis, F. Menczer, A. Flammini, Online human-bot interactions: Detection, estimation, and characterization, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 11, (no. 1) 2017, pp. 280–289.
[21] C.A. Davis, O. Varol, E. Ferrara, A. Flammini, F. Menczer, Botornot: A system to evaluate social bots, in: Proceedings of the 25th International Conference Companion on World Wide Web, 2016, pp. 273–274.
[22] O. Varol, C.A. Davis, F. Menczer, A. Flammini, Feature engineering for social bot detection, in: Feature Engineering for Machine Learning and Data Analytics, CRC Press, 2018, pp. 311–334.
[23] F. Martini, P. Samula, T.R. Keller, U. Klinger, Bot, or not? Comparing three methods for detecting social bots in five political discourses, Big Data Soc. 8 (2) (2021) 20539517211033566.
[24] S. Cresci, R. Di Pietro, A. Spognardi, M. Tesconi, M. Petrocchi, Demystifying misconceptions in social bots research, 2023, arXiv preprint arXiv:2303.17251.
[25] M. Mazza, S. Cresci, M. Avvenuti, W. Quattrociocchi, M. Tesconi, Rtbust: Exploiting temporal patterns for botnet detection on Twitter, in: Proceedings of the 10th ACM Conference on Web Science, 2019, pp. 183–192.
[26] S. Wojcik, S. Messing, A.W. Smith, L. Rainie, P. Hitlin, Bots in the Twittersphere, 2018.
[27] G. Stocking, N. Sumida, Social media bots draw public's attention and concern, 2018.
[28] B. Mullin, Report: Journalists are largest, most active verified group on Twitter, Poynter Inst., May 26 (2015).
[29] C. Brems, M. Temmerman, T. Graham, M. Broersma, Personal branding on Twitter: How employed and freelance journalists stage themselves on social media, Digit. J. 5 (4) (2017) 443–459.
[30] L. Molyneux, A. Holton, S.C. Lewis, How journalists engage in branding on Twitter: Individual, organizational, and institutional levels, Inf., Commun. Soc. 21 (10) (2018) 1386–1401.
[31] S. Haustein, T.D. Bowman, K. Holmberg, A. Tsou, C.R. Sugimoto, V. Larivière, Tweets as impact indicators: Examining the implications of automated "bot" accounts on T witter, J. Assoc. Inf. Sci. Technol. 67 (1) (2016) 232–238.
[32] T. Lokot, N. Diakopoulos, News bots: Automating news and information dissemination on Twitter, Digit. J. 4 (6) (2016) 682–699.
[33] A. Smith, S. Colton, The@ artbhot text-to-image Twitter bot, in: Proceedings of the International Conference on Computational Creativity, 2022.
[34] F. Brachten, M. Mirbabaie, S. Stieglitz, O. Berger, S. Bludau, K. Schrickel, Threat or opportunity?-examining social bots in social media crisis communication, 2018, arXiv preprint arXiv:1810.09159.
[35] S. Deshpande, J. Warren, Self-harm detection for mental health chatbots, in: MIE, 2021, pp. 48–52.
[36] S. Savage, A. Monroy-Hernandez, T. Höllerer, Botivist: Calling volunteers to action using online bots, in: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, 2016, pp. 813–822.
[37] K.-C. Yang, O. Varol, P.-M. Hui, F. Menczer, Scalable and generalizable social bot detection through data selection, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (no. 01) 2020, pp. 1096–1103.
[38] J. Pfeffer, D. Matter, K. Jaidka, O. Varol, A. Mashhadi, J. Lasser, D. Assenmacher, S. Wu, D. Yang, C. Brantner, et al., Just another day on Twitter: A complete 24 hours of Twitter data, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 17, 2023, pp. 1073–1081.
[39] F. Morstatter, J. Pfeffer, H. Liu, K. Carley, Is the sample good enough? Comparing data from Twitter's streaming api with Twitter's firehose, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 7, (no. 1) 2013, pp. 400–408.
[40] J. Pfeffer, K. Mayer, F. Morstatter, Tampering with Twitter's sample API, EPJ Data Sci. 7 (1) (2018) 50.

[41] J. Pfeffer, A. Mooseder, L. Hammer, O. Stritzel, D. Garcia, This sample seems to be good enough! Assessing coverage and temporal reliability of Twitter's academic API, 2022, arXiv preprint arXiv:2204.02290.
[42] Form 10-Q Twitter, Inc., 2022, https://web.archive.org/web/20220821120344/https://sec.report/Document/0001418091-22-000147/. (Accessed 21 August 2022).
[43] T. Fagni, F. Falchi, M. Gambini, A. Martella, M. Tesconi, TweepFake: About detecting deepfake tweets, PLoS One 16 (5) (2021) e0251415.
[44] K. Narayan, H. Agarwal, S. Mittal, K. Thakral, S. Kundu, M. Vatsa, R. Singh, DeSI: Deepfake source identifier for social media, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2858–2867.
[45] Y.M. Çetinkaya, İ.H. Toroslu, H. Davulcu, Developing a Twitter bot that can join a discussion using state-of-the-art architectures, Soc. Netw. Anal. Min. 10 (1) (2020) 1–21.
[46] S.-S. Jeong, Y.-S. Seo, Improving response capability of chatbot using Twitter, J. Ambient Intell. Humaniz. Comput. (2019) 1–14.
[47] D. Assenmacher, L. Clever, L. Frischlich, T. Quandt, H. Trautmann, C. Grimme, Demystifying social bots: On the intelligence of automated social media actors, Soc. Media+ Soc. 6 (3) (2020) 2056305120939264.
[48] T.R. Keller, U. Klinger, Social bots in election campaigns: Theoretical, empirical, and methodological implications, Political Commun. 36 (1) (2019) 171–189.
[49] J. Echeverria, S. Zhou, Discovery, retrieval, and analysis of the'star wars' botnet in Twitter, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, 2017, pp. 1–8.
[50] I. Alsmadi, M.J. O'Brien, How many bots in Russian troll tweets? Inf. Process. Manage. 57 (6) (2020) 102303.
[51] A. Bessi, E. Ferrara, Social bots distort the 2016 US presidential election online discussion, First Monday 21 (11–7) (2016).
[52] J. Uyheng, K.M. Carley, Bot impacts on public sentiment and community structures: Comparative analysis of three elections in the Asia-Pacific, in: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Springer, 2020, pp. 12–22.
[53] M. Stella, M. Cristoforetti, M. De Domenico, Influence of augmented humans in online interactions during voting events, PLoS One 14 (5) (2019) e0214210.
[54] S. Rossi, M. Rossi, B. Upreti, Y. Liu, Detecting political bots on Twitter during the 2019 finnish parliamentary election, in: Proceedings of the 53rd Hawaii International Conference on System Sciences, 2020.
[55] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, M. Tesconi, Fame for sale: Efficient detection of fake Twitter followers, Decis. Support Syst. 80 (2015) 56–71.
[56] X. Teng, Y.-R. Lin, W.-T. Chung, A. Li, A. Kovashka, Characterizing user susceptibility to COVID-19 misinformation on Twitter, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 16, 2022, pp. 1005–1016.
[57] K.-C. Yang, C. Torres-Lugo, F. Menczer, Prevalence of low-credibility information on Twitter during the covid-19 outbreak, 2020, arXiv preprint arXiv:2004.14484.
[58] E. Ferrara, # Covid-19 on Twitter: Bots, conspiracies, and social media activism, 2020, arXiv preprint arXiv:2004.09531.
[59] N. Pröllochs, Community-based fact-checking on Twitter's Birdwatch platform, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 16, 2022, pp. 794–805.
[60] J. Allen, C. Martel, D.G. Rand, Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program, in: CHI Conference on Human Factors in Computing Systems, 2022, pp. 1–19.
[61] T. Yasseri, F. Menczer, Can the Wikipedia moderation model rescue the social marketplace of ideas? 2021.
[62] I.V. Pasquetto, B. Swire-Thompson, M.A. Amazeen, F. Benevenuto, N.M. Brashier, R.M. Bond, L.C. Bozarth, C. Budak, U.K. Ecker, L.K. Fazio, et al., Tackling misinformation: What researchers could do with social media data, Harvard Kennedy School Misinform. Rev. (2020).
[63] V.S. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, The DARPA Twitter bot challenge, Computer 49 (6) (2016) 38–46.