IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 April 2023, accepted 24 May 2023, date of publication 2 June 2023, date of current version 8 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3282580

RESEARCH ARTICLE

Twitter Newcomers: Uncovering the Behavior and Fate of New Accounts Through Early Detection and Monitoring

GUGLIELMO COLA, MICHELE MAZZA, AND MAURIZIO TESCONI
Institute of Informatics and Telematics (IIT), National Research Council (CNR), 56124 Pisa, Italy

Corresponding author: Guglielmo Cola (guglielmo.cola@iit.cnr.it)

This work was supported by the European Union—Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019—Integrating Activities for Advanced Communities", Grant Agreement n.871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu).

ABSTRACT There is a significant body of literature concerning the analysis of Twitter accounts, yet the behavior of newly created accounts remains relatively unexplored. In this study, we introduce a novel approach to detect Twitter accounts right after registration and explore their behavioral patterns. In a two-week period in April 2020, our technique identified over 500,000 accounts before they even started interacting with the platform. Each account was monitored for 21 days by sampling profile information and timelines at scheduled intervals, retrieving over 8 million tweets. An additional sample of profile information was collected approximately two years after creation, in May 2022. One of the key findings of our study is the lack of sustained and genuine engagement from new accounts. Indeed, a large proportion of them (almost 25%) were suspended by Twitter in the first 21 days, and the evaluation conducted after two years reveals that only a tiny fraction of the remaining enabled accounts seem to be active and genuine users (3.8% of the initial sample). Additionally, despite the early suspensions enforced by Twitter, it turns out that some short-lived accounts still managed to have a substantial impact on the total volume of content and interactions from new accounts. Overall, our findings may have important implications for understanding the dynamics of new accounts' behavior as well as Twitter's suspension policy prior to the recent change in ownership. This could stimulate further research to evaluate the impact of the ongoing changes introduced by the new administration.

INDEX TERMS Ephemeral accounts, fake accounts, social bots, social media analysis, suspended accounts, user engagement.

I. INTRODUCTION

Twitter has emerged as one of the most popular online social networks, with over 300 million active users [1]. Popularity inevitably attracted malicious actors, who targeted the platform with activities like spamming [2], phishing [3], distributing malware [4] and diffusing false information [5], [6]. Recently, Twitter's ability to deal with these challenges has been under scrutiny, especially in light of the dispute with business magnate Elon Musk, who then acquired Twitter itself in October 2022. In particular, Musk accused the board of directors of lying about the real impact of bots, as the company reported that "false or spam accounts" represent less than 5% of its monetizable daily active users, i.e., daily users who accessed the platform. In contrast, Musk claimed that, according to a study he had commissioned, false accounts represented 33% of visible accounts during the first week of July 2022, and 10% of monetizable daily active users [7].

The presence of coordinated and inauthentic accounts spreading misinformation has also raised concern among the scientific community, as Twitter has become an essential source of news for many people. For instance, it has been shown that Twitter played a relevant role in the debate surrounding events like general elections [5], or health-related

The associate editor coordinating the review of this manuscript and approving it for publication was Vijay Mago.

emergencies [8], [9]. Researchers have devised a variety of reactive approaches to identify malicious accounts, typically based on features like specific patterns in screen names [10], [11], a high rate of tweets containing URLs [12], [13], or an unusual increase in the number of followings or followers [14].

Supposedly, bad actors have two main ways to deal with Twitter's suspension techniques: *advanced* and *ephemeral* accounts [11]. The first approach consists in preparing sophisticated accounts that are able to elude Twitter filters: such advanced accounts generally require a relatively high operation cost. In contrast, it is hypothesized that a much simpler and cost-effective approach consists in continuously creating a large number of new accounts that rapidly fulfill their malicious task until they are suspended by Twitter. The impact of such ''ephemeral'' accounts on the platform is not fully understood, as the API does not explicitly allow researchers to detect and monitor new accounts since their creation time. Indeed, newly created accounts can only be studied after they share content on the platform.

In our study, we used an innovative technique to detect Twitter accounts right after registration and before any other interaction with the platform. This allowed us to detect over 500,000 new accounts created in the second half of April 2020, and then monitor profile information and tweeting activity during their first 21 days on Twitter. For a broader temporal view, we expanded the analysis on deactivations and suspensions by re-sampling profile information about two years after the accounts were created, in May 2022. This dataset offers an unprecedented opportunity to study the behavior of new accounts as well as to evaluate the measures adopted by Twitter to limit malicious behaviors.

The analysis provided in this paper aims to answer the following research questions:

- **RQ1** What is the prevalence of ephemeral (i.e., short-lived) and potentially malicious accounts among new Twitter users?
- **RQ2** What are the trends in deactivations and suspensions during the first 21 days?
- **RQ3** How do ephemeral accounts use the platform and what is their overall impact compared to normal accounts within the first 21 days?

Notable findings include that about one in four accounts do not remain enabled beyond their first 21 days on the platform, and only 3.8% of new accounts keep using the platform with human-like behavior after two years. We believe that our study and the resulting dataset could spark further research aimed at evaluating how the policies enforced by the new Twitter ownership, including the changes in the access to the API, might have affected the prevalence and impact of ephemeral accounts.

The paper is organized as follows. Section II briefly reports the most relevant work from the state of the art. Section III presents the innovative technique devised to spot new account IDs seconds after registration. Section IV describes the dataset and shows all the analyses we performed to answer the research questions mentioned above. Finally, we draw our main conclusions in Section V.

II. RELATED WORK

Most of the Twitter-related studies turn out to be tweet-based [11], [15]. This approach implies collecting tweets using queries related to a particular topic or sampling random tweets in real time through the API. Thus, only accounts that tweeted at least once and that matched the chosen criteria could be identified and studied. Our work differs from these as it is user-based: accounts are collected first and then, through monitoring, the tweets shared by them are captured.

Only a few works described a user-based approach. Pioneering user-based research was carried out by [16]: the authors collected 537 million Twitter accounts (the entire Twitter social graph as of July 2012) through a distributed crawler. Accounts were identified by their ID, which at the time consisted of a 32-bit integer allocated sequentially [17]. Of the gathered accounts, only 50% shared at least one tweet, while 40% were not followed by anybody and 25% did not follow anybody. The authors of [18] discovered a botnet of more than 350,000 social bots that tweeted random quotes from Star Wars novels; their initial dataset of 6 million random accounts was created by choosing a uniform 1% sample in the ID space (2^{32} possible values).

As mentioned above, Twitter account IDs used to be simple incremental numbers on 32 bits, so it was relatively easy to automatically generate new valid IDs. However, since 2016 Twitter has adopted *Snowflake* to generate new IDs: this new approach does not rely on simple sequential numbers and extends the space of possible IDs to 2^{64} (more detail in Section III). Some useful hints on how Snowflake-based IDs could be guessed are presented as a corollary contribution in [19], which nevertheless based its main analysis on a dataset created with an approach similar to [18] combined with a tweet-based method.

Using account IDs is not the only means for identifying Twitter accounts. In [20], the authors first collected 20 seed Twitter accounts from the public timeline, then they gathered the seed accounts' followers and followings. The latter step was repeated until they built a collection of nearly 500,000 accounts, which were used to study spammers. A similar strategy is implemented in [21]. To perform a retrospective analysis of accounts suspended from Twitter, the authors collected the followers of the top 100 most-followed Twitter accounts. By doing so, they were able to construct a dataset of approximately 560 million accounts. A different approach is presented in [22], where accounts listed for sale by an underground merchant were detected and monitored, identifying 23,579 fake accounts that shared at least one tweet in 2020.

To the best of our knowledge, our paper is the first work presenting a study on a large number of new Twitter accounts that have been detected right after registration and before they produced any visible content or interaction. This allowed us

TABLE 1. Structure of the 64-bit Twitter ID, starting from the most significant bit.

Field	Bits	Description
Reserved	1	The most significant bit is set to zero.
Timestamp	41	Milliseconds since the 4^{th} of November 2010 at 01:42:54 UTC (can be found as the current UNIX epoch minus 1288834974657).
Worker	10	Unique identifier for the worker thread that generated the ID.
Sequence number	12	Used by a worker for incremental labeling when contents are generated at the same timestamp.

to include in the analysis not only the accounts that tweeted, but also those that engaged in other interactions such as likes, followers, and followings, as well as the accounts that were completely inactive.

III. DATA COLLECTION METHOD

In 2010, Twitter announced a new internal service called *Snowflake* to manage the creation of time-ordered IDs for user accounts and tweets in a distributed environment.[1] In the same year, an initial implementation of this service was released in a public repository.[2] However, support for this repository was discontinued in 2014. The adoption of Snowflake for user IDs only became effective in early 2016.[3]

In order to perform reverse engineering of Snowflake IDs, we referred to the blog post from 2010, the initial source code, and incorporated insights from previous literature [19], [23]. Snowflake Twitter IDs are structured as presented in Table 1. The use of a timestamp in the most significant bits ensures that IDs are time-ordered with millisecond precision; the worker part identifies a specific thread or process, namely an ''object'' capable of generating a new ID; sequence numbers are supposedly adopted by workers to create different IDs with the same timestamp. This design ensures that different ''workers'' operating in distributed data centers can generate unique account IDs without the need for any centralized coordination.

Starting from these specifications of Snowflake, we devised a technique to spot new account IDs right after their creation. Each Twitter API developer credential (v1.1) allowed us to test 100 different IDs per second by using the ''users/lookup'' endpoint. Our technique exploited 10 credentials in parallel in order to test 1,000 possible new IDs per second. These 1,000 IDs were based on 250 possible timestamps and 4 different worker IDs. Timestamps were chosen randomly among the 1,000 unique timestamps available per second (millisecond precision), whereas the four worker IDs were chosen dynamically by exploiting a dedicated algorithm. This algorithm continuously keeps an updated list of the four most recently active workers. Finally, the sequence number part of the tried IDs was set to zero because this is by far the most common value according to our experiments and as suggested by [19].

[1] blog.twitter.com/engineering/en_us/a/2010/announcing-snowflake.
[2] github.com/twitter-archive/snowflake/releases/tag/snowflake-2010.
[3] twittercommunity.com/t/migration-of-twitter-core-entities-to-64-bit-ids/56881.

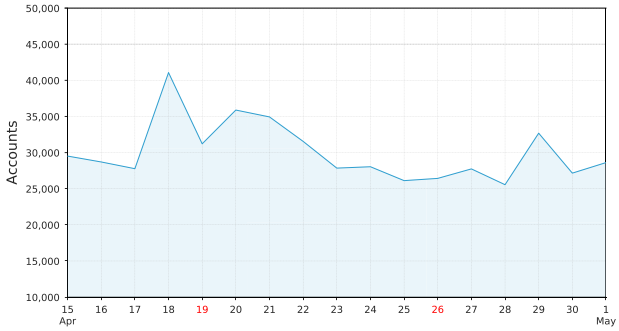


FIGURE 1. New accounts detection trend (Sundays in red).

The ''users/lookup'' endpoint, for each valid ID, returns the main profile information (''user object''): this information was stored in an internal database. Each detected account was then monitored for 21 days by two dedicated scripts, which were executed in parallel. At scheduled intervals, the first script retrieved new samples of the user object by using again the ''user/lookup'' endpoint, while the second script leveraged the ''statuses/user_timeline'' endpoint to retrieve the tweets from the user's timeline.

If an account is deactivated or suspended, the ''users/lookup'' endpoint does not return any information. In this case, the same script automatically calls the ''users/show/ID'' endpoint to discriminate whether the account has been deactivated or suspended (deactivation is a willing action taken by the user, whereas suspension is enforced by Twitter). This information is then stored in the database and the profile information is kept being refreshed according to the usual schedule, in case the account is reactivated. A more detailed view of the parallel subtasks involved in detecting and monitoring new IDs is shown in Table 2.

To enable a broader temporal view, we also collected a sample of the user objects approximately 2 years after creation. We used the ''users/lookup'' and ''users/show'' endpoints to refresh account information, which allowed us to update the statistics related to the number of deactivated, suspended, and enabled accounts, as well as the total number of tweets (''statuses_count'') shared by enabled accounts.

IV. RESULTS AND DISCUSSION

Between April 15 and May 1, 2020, our technique identified 510,841 new accounts. On average, accounts were detected 16 seconds after the creation date reported in the user object. The maximum delay with respect to creation was just 34 seconds. Figure 1 depicts the number of accounts found on each day. On average, 30,049 accounts were found daily, with a standard deviation of 4,112. There is no apparent correlation

TABLE 2. Subtasks in the Twitter ID detection and monitoring technique.

Subtask	API Endpoint(s)	# Threads	Description
ID detection	users/lookup	10	Tests 1,000 potential new account IDs per second by using the users/lookup endpoint with a combination of 250 different timestamps, the four most active worker IDs, and a sequence number of 0.
Find Active Workers	users/lookup	2	Estimates the most recently used "workers" to generate new account IDs. This is achieved by constantly trying to find new account IDs with all the worker IDs that may be used by Twitter at different points in time (IDs ranging from 320 to 382 according to the literature [19] and our experiments). When a new account ID is found, the respective worker ID is moved to the top of the list of worker IDs. The first four worker IDs in the list are used by the ID detection subtask. This way, ID detection can be more effective, as it considers solely the most active workers.
Sample User Objects	users/lookup; users/show/ID	12	For each new account ID, retrieves updated samples of the main profile information (user objects) at regular intervals through the users/lookup endpoint. Sampling is done according to the following schedule: every 10 minutes in the first hour following registration; every hour for the rest of day 1; every 2 hours for the rest of week 1; every 4 hours up to the end of the 21-day monitoring period. When the user object is not returned by the users/lookup endpoint, the users/show/ID endpoint is used to determine if the account has been deactivated or suspended.
Sample User Timelines	statuses/user_timeline	16	For each new account ID, retrieves updated samples of the user timeline (tweet objects) through the statuses/user_timeline endpoint. Sampling is done according to the same schedule used for user objects.

TABLE 3. Total activity by monitored accounts in their first 21 days.

Interaction	Count	Accounts	% of total
New tweets	1,757,697	126,447	24.8%
Replies	2,732,646	98,872	19.4%
Retweets	3,242,951	72,134	14.1%
Quotes	284,039	26,866	5.3%
All tweets	8,000,093	179,535	35.1%
Likes	14,326,650	182,887	35.8%
Followers	3,266,163	157,367	30.8%
Followings	9,932,960	284,644	55.7%

with specific weekdays. The IDs of detected accounts and tweets are publicly available.[4]

As outlined in Section III, our analysis examined only a subset of the possible IDs: i) 25% of the possible timestamps (250 milliseconds checked per second); ii) the four most active ''workers''; iii) sequence number equal to zero. Therefore, we can only find a conservative estimate of the number of new Twitter accounts per day by multiplying the daily mean value times four, resulting in about 120,000 new Twitter accounts per day and about 3.6 million new accounts per month.

Table 3 summarizes the platform activity for different types of interactions. For each interaction, it is shown: the total count produced by the monitored accounts in their first 21 days; the number and percentage of accounts involved in such interactions. Over 8 million tweets were produced by approximately 179,000 accounts, corresponding to 35.1% of the accounts in our dataset. Retweets were the most common tweet type (40.5% of all tweets), followed by replies (34.2%) and new tweets (22%). The prevalence of retweets is not surprising, as retweeting involves just a couple of clicks/taps on already-existing content [24]. The number of unique accounts per tweet type offers a different perspective: about one in four accounts posted at least one original tweet, whereas less than 20% posted at least one retweet or reply. Only 5.3% of accounts used quotes, while 65% did not tweet at all. Likes represent an even simpler form of interaction and were used by 35.8% of accounts. Notably, 52.1% of accounts were inactive in terms of both likes and tweets of any kind. Another way of interacting with the platform involves being followed by other accounts (followers) or following other accounts (followings). Table 3 reveals that 30.8% of accounts achieved at least one follower, while 55.7% followed at least another account. The imbalance between the number of followers and followings can be attributed to the fact that gaining followers requires more effort than simply following other accounts. Furthermore, Twitter itself facilitates new followings by suggesting potentially interesting accounts during the registration procedure.

In the following, we describe our analyses to answer the three research questions highlighted in the Introduction. Subsection IV-A addresses RQ1 and, to a lesser extent, RQ2 by presenting the prevalence of deactivated and suspended accounts during the first 21 days and the number of potential social bots after two years. Subsection IV-B provides a more detailed analysis of RQ2 by showing the timing of deactivations and suspensions of new accounts. Finally, Section IV-C answers RQ3 by describing the impact of early deactivated and suspended accounts on the platform.

A. PREVALENCE OF POTENTIALLY MALICIOUS ACCOUNTS AND SOCIAL BOTS

There are two ways a Twitter account may become unable to interact with the platform: deactivation and suspension.

[4]https://data.d4science.net/Yr5d

TABLE 4. Deactivated and suspended accounts after 21 days.

Status	Accounts	
Enabled	363,733	71.2%
Deactivated	21,905	4.3%
Suspended	125,203	24.5%

Deactivation occurs when users willingly deactivate their accounts. After this decision, users have up to 30 days to access the account and re-enable it, otherwise, the account is permanently deleted from the platform. The Twitter API can reveal if an account has been deactivated, but it does not allow us to discriminate between deactivated and permanently deleted accounts.

Suspension is executed by Twitter itself when an account somehow violates its policy. According to Twitter, suspension mostly occurs because accounts ''are spammy, or just plain fake, and they introduce security risks for Twitter and for everyone using Twitter''. Other motivations are the suspicion that an account might have been hacked, or ''abusive tweets or behavior''. Accounts marked as suspended by the Twitter API cannot be reached on Twitter, as their URL returns a ''page not found'' error. Users may have the opportunity to appeal against it, but reactivation after suspension is an extremely rare event. On some occasions accounts may be ''temporarily limited'' by Twitter: these accounts can still be reached by other users, though the latter are warned that the account's interaction with the platform has been restricted due to suspicious activity. The Twitter API only allows us to detect when an account is deactivated or suspended, whereas temporary restrictions are not reported.

The first result related to deactivations and suspensions in our dataset is shown in Table 4, which shows the total figure of deactivated and suspended accounts at the end of the 21-day interval relative to account creation. Notably, about one in four accounts were suspended in the first 21 days (24.5%), while 4.3% of accounts were deactivated by their users. Hence, only 71.2% of accounts were still able to interact with the platform after only 21 days since their creation.

From the screening of accounts' status after two years (May 2022) it turns out that 20,985 more accounts have been deactivated (+95.8%), whereas another 16,673 accounts have been suspended by Twitter (+13,3%). Compared to what we observed in the first 21 days, the number of deactivations has nearly doubled, whereas the proportional increase in suspensions was far lower. Seemingly, Twitter's suspension policy tends to be applied mostly in the first few weeks of an account's existence.

After this overview of new deactivations and suspensions, we deepened our investigation of the behavior of the remaining enabled accounts in this two-year window, from May 2020 to May 2022. More specifically, we used the ''Botometer'' classifier [25] to estimate the presence of bots among enabled accounts. Botometer returns a probability estimate $\in [0, 1]$ for each account: values near one denote a likely social bot, whereas lower values indicate a higher probability of being a human. To identify potential social bots, we set the threshold to 0.76, as done in other works [26], [27]. Botometer requires that the analyzed account has shared at least 20 tweets to provide a reliable score. Considering that the age of our accounts is approximately two years, 20 tweets correspond to less than one tweet per month.

Two years after creation, only 56,525 accounts (11.1% of the initial sample) are relevant to social bot detection, being enabled and having shared at least 20 tweets in two years. Moreover, Botometer-based analysis reveals that most of these accounts (around 65%) are potential social bots.

A recap of the findings presented in this subsection is shown in Figure 2. From our initial sample of 510,841 accounts created between April 15 and May 1, 2020, only 363,733 (71.2%) survived the first 21 days, while 326,075 (63.8%) were still enabled after two years. Botometer-based analysis of the 56,525 enabled accounts with more than 20 tweets after two years reveals that only 19,350 accounts (3.8% of the initial sample) are likely to be genuine users.

B. DAILY TRENDS IN DEACTIVATIONS AND SUSPENSIONS DURING THE FIRST 21 DAYS

In this subsection, we delve into the timing of deactivations and suspensions during the first 21 days after registration. Hereafter, we refer to the three groups of accounts described in Table 4 as Enabled, Deactivated, and Suspended. These groups are based on the account status after 21 days. Our dataset does not show suspicious bursts of potentially malicious accounts created on a specific day, but rather a stable production of such accounts. Figure 3 shows when the accounts from the three groups were detected, which corresponds to their creation date (log scale was used to better visualize the number of accounts in the Deactivated group, which is roughly ten times smaller with respect to the other two groups). It can be observed that the creation of accounts in all groups was relatively stable throughout the monitored interval: the average daily creation value was 21,396 $\pm$ 3,284 for Enabled, 1,289 $\pm$ 119 for Deactivated, and 7,365 $\pm$ 1,043 for Suspended. Also, there is a strong correlation between the three sets of values, indicating that more deactivated/suspended accounts were found on days when our technique was able to spot a higher total number of newly created accounts.

A different perspective is provided in Figure 4, which depicts the number of deactivated and suspended accounts on any given day relative to account creation, from day 1 (first day on the platform, right after creation) to day 21 (last day in the monitored interval). Concerning deactivations, most of them (58.5%) occurred during the first day: a possible explanation is that these users briefly tried the platform before deactivating their new account. Suspensions show an even more peculiar pattern: most suspensions occurred around the 16^{th} day relative to creation, when over 100,000 accounts

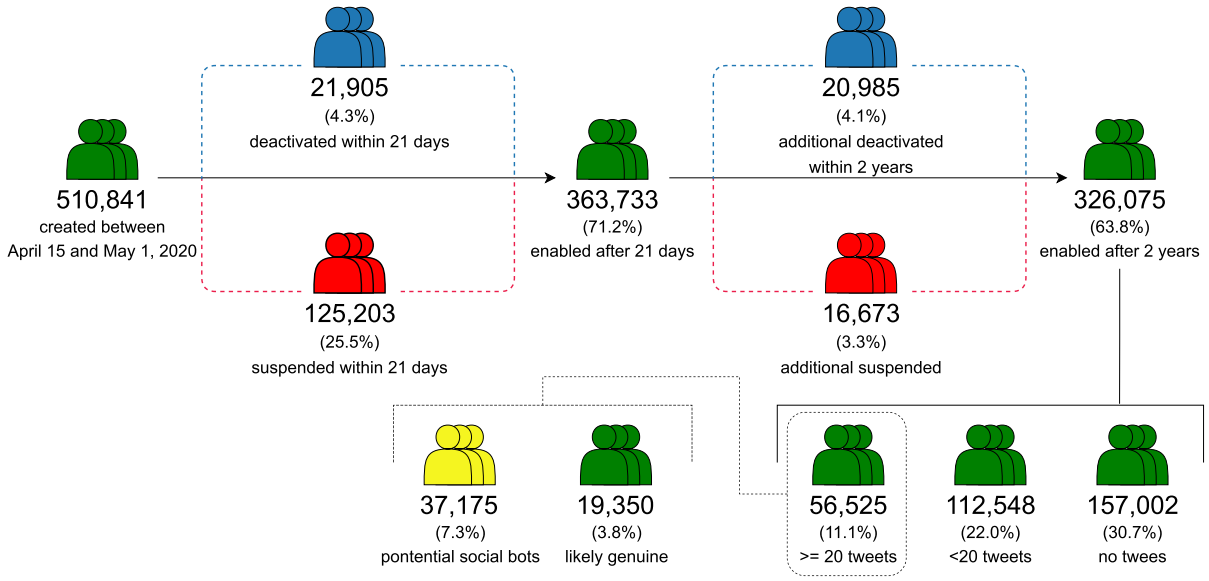


FIGURE 2. Temporal view of the fate of new Twitter accounts: deactivations and suspensions in the first 21 days, subsequent events within 2 years, and identification of genuine active users and social bots.

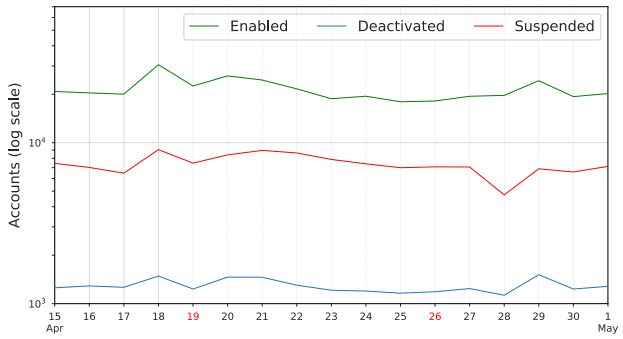


FIGURE 3. New accounts detection trend for each group.

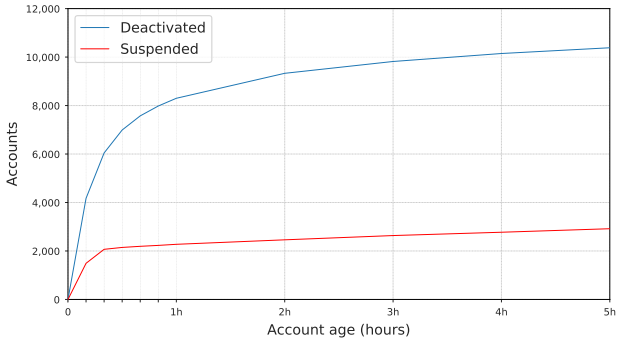


FIGURE 5. Deactivations and suspensions in the first 5 hours.

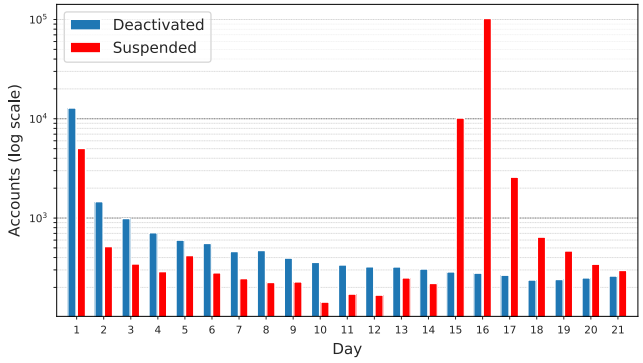


FIGURE 4. Deactivations and suspensions in the first 21 days.

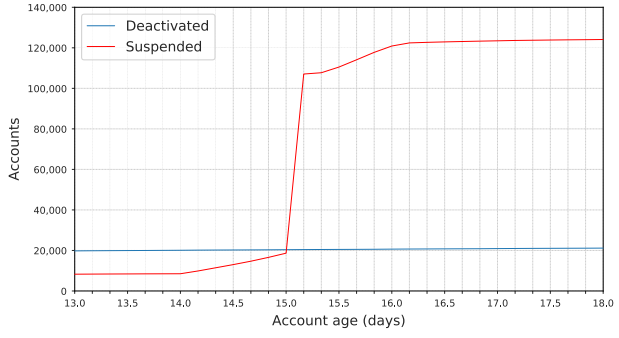
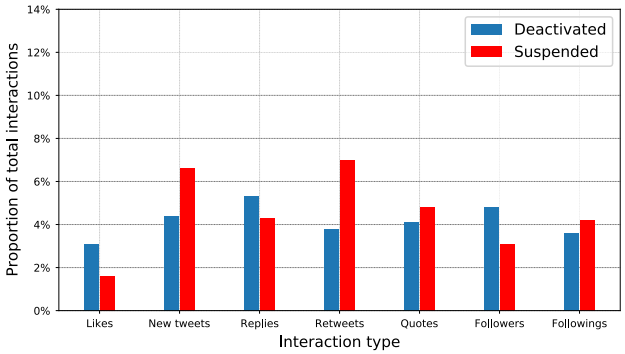


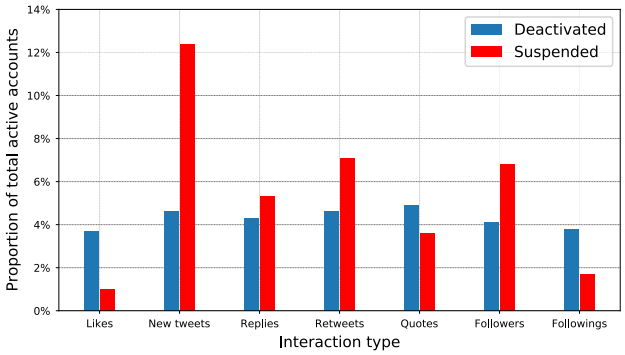
FIGURE 6. Deactivations and suspensions around the 16^{th} day.

from our dataset were suspended by Twitter. More precisely, 8.1% of suspensions occurred on the 15^{th} day, 81.7% on the 16^{th}, and 2.1% on the 17^{th}. Hence, these three days (relative to account creation) determined 91.8% of the total number of suspensions in our dataset. Another day with a high rate of suspensions was the first relative to creation, when 4% of the total suspensions occurred. On the remaining days, the average rate of daily suspensions was substantially lower (0.2%).

More details on the timing of deactivations and suspensions are provided in Figure 5, which depicts the first 5 hours



(a) Volumes produced by Deactivated and Suspended accounts per each interaction type (percentage of the interactions from new accounts)



(b) Active accounts among Deactivated and Suspended per each interaction type (percentage of the total active accounts per interaction type)

FIGURE 7. Comparison of volumes and active accounts among Deactivated and Suspended for each interaction type.

after creation, and in Figure 6, which shows what happened around the 16^{th} day. Figure 5 reveals that the majority of first-day deactivations and suspensions occurred within the first hour. Specifically, within the first 20 minutes, Twitter enforced 2,068 suspensions, accounting for over 40% of day 1 suspensions. Regarding the anomaly of the high rate of suspensions around day 16, Figure 6 shows that, indeed, there is a visible rise in the suspension rate starting from day 15 and up to the beginning of day 17. However, most suspensions occurred in the first four hours of day 16, when Twitter suspended 88,408 accounts, representing approximately 70% of the overall suspensions in our dataset.

C. BEHAVIOR OF DEACTIVATED AND SUSPENDED ACCOUNTS

In this subsection, we analyze the behavior of new accounts during the first 21 days, with a particular focus on the different groups of accounts based on their status at the end of the monitored period. This way, we aim to show whether ephemeral accounts (Deactivated and Suspended) managed to have a substantial impact in terms of interactions when compared to Enabled accounts.

Figure 7a illustrates the percentage of interactions generated by Deactivated and Suspended accounts for each possible interaction type, relative to the total interactions produced by new accounts. Likewise, for each interaction type, Figure 7b illustrates the percentage of accounts that belong to Deactivated or Suspended with respect to the total active accounts for that interaction.

TABLE 5. Subgroups identified among Deactivated and Suspended.

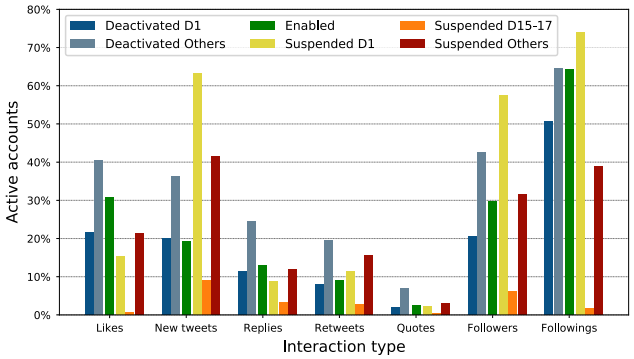
Status	Accounts	
Deactivated D1	8,644	39.5%
Deactivated Others	9,089	41.5%
Suspended D1	3,524	2.8%
Suspended D15-17	114,955	91.8%
Suspended Others	5,233	4.2%

Some interesting results can be observed. In terms of total volumes, even though Enabled produced the large majority of interactions, Suspended and Deactivated combined were responsible for over 10% of the total tweets from new accounts on the platform, almost 5% of likes, and around 8% of achieved followers and new followings. In terms of accounts, there is a high proportion of Suspended involved in producing new tweets: 12.4% of the total number of accounts that tweeted new content were later suspended. If we consider all the tweet types combined, only 84.1% of tweeters belong to the Enabled group. Finally, it is worth mentioning that suspended accounts were particularly active in content amplification (7.0% of total retweets and 7.1% of total retweeters) as well as in achieving followers from other accounts (6.8% of the accounts that obtained at least one follower belong to Suspended).

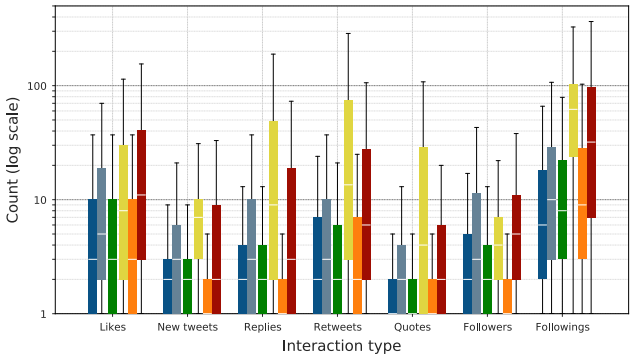
It is worth noting that the total count values presented here (interactions and active accounts) may underestimate the true impact of Deactivated and Suspended. This is because our monitoring technique relies on a predefined sampling schedule, hence interactions that were performed shortly before an account was suspended or deactivated might have been missed. Also, it should be considered that Deactivated and Suspended produced these volumes in less time, as they were not active for the whole 21-day interval.

DEACTIVATED AND SUSPENDED SUBGROUPS

In order to provide more detail on the behavior of Deactivated and Suspended, we identified specific subgroups based on the timing of deactivations/suspensions as presented in the previous subsection. For Deactivated, we define two subgroups: *Deactivated D1*, the accounts deactivated on their day 1 and that were enabled for at least 10 minutes; *Deactivated Others*, the accounts deactivated after their first day and within the monitored interval of 21 days. For Suspended, we define three groups: *Suspended D1*, the accounts suspended on their day 1 and that were enabled for at least 10 minutes; *Suspended D15-17*, the accounts suspended between day 15 and day 17; *Suspended Others*, the accounts suspended on the remaining days relative to account creation (2-14 or 18-21). The number of accounts in each subgroup and the percentage of such

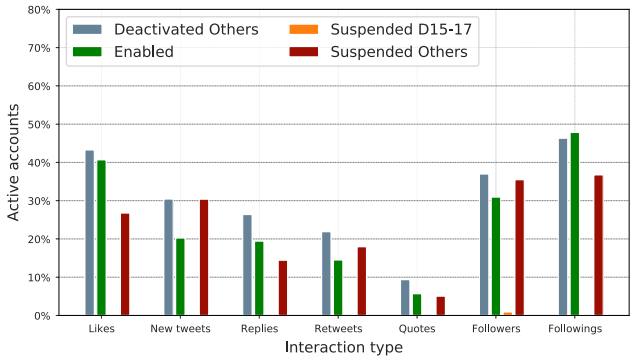


(a) Day 1 active accounts per group and interaction (percentage of each group total accounts)

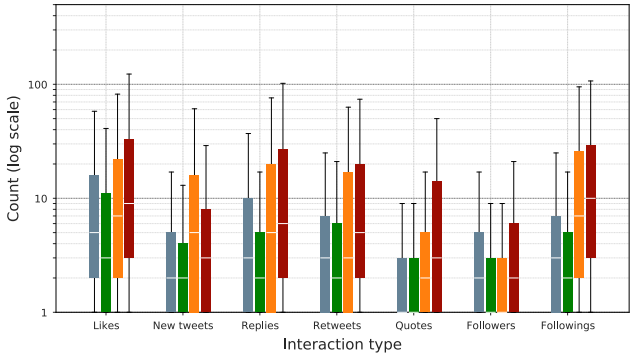


(b) Day 1 interaction count boxplot

FIGURE 8. Day 1 analysis of the interactions performed by the different groups of accounts, in terms of active accounts (%) and number of interactions among active accounts.



(a) Active accounts after day 1, per group and interaction (percentage of each group total accounts)



(b) Daily interaction count boxplot (interactions after day 1)

FIGURE 9. Activity performed by the different groups of accounts after day 1, in terms active accounts (percentage of accounts that made at least one interaction after day 1, with respect to the total number of accounts in the group) and number of daily interactions among active accounts.

accounts with respect to the total number of accounts in the group are shown in Table 5.

The decision to exclude from these analyses the accounts deactivated within ten minutes (4,172 accounts, 19.0% of Deactivated) or suspended within ten minutes (1,491 accounts, 1.2% of Suspended) stems from the inability to detect any kind of interaction for such accounts, due to the sampling interval of our monitoring technique (one sample of the users' timeline and one user object every ten minutes in the first hour after account creation).

Day 1 analysis may reveal early signs of abnormal behavior soon after creation. The two plots in Figure 8 depict the day 1 activity of the different subgroups of Deactivated and Suspended accounts compared to the group of Enabled accounts (green bars). More precisely, Figure 8a shows the percentage of active accounts for each subgroup and for each interaction type, while the boxplots in Figure 8b show the distribution of the number of interactions performed on day 1 by active accounts (whiskers $= 3 * IRQ$). Interaction count is shown in log scale for better visibility of lower values. The behavior after day 1 is analyzed in a similar manner in Figure 9. This time, Figure 9a shows the percentage of active accounts that performed at least one interaction *after* day 1, while Figure 9b shows the distribution of the number of daily interactions from active accounts, considering any day from day 2 to 21.

Let us first discuss the behavior of Deactivated subgroups. In Figure 8, Deactivated D1 (dark blue bars) show a behavioral pattern relatively similar to Enabled, both in terms of active accounts and distribution of interactions. This suggests that most of Deactivated D1 were indeed users that briefly tried the platform before deactivating their account during day 1. On the other hand, Deactivated Others (light blue bars) appear to be more productive with respect to Enabled for all kinds of interactions. This trend is confirmed in the analysis after day 1 shown in Figure 9.

Regarding the accounts suspended on day 1 (Suspended D1, yellow bars), Figure 8a shows that a high percentage of them was involved in producing new tweets (over 60%, compared to only 20% of Enabled), achieving followers (over 50%), and following other accounts (over 70%). Moreover, Suspended D1 were substantially more prolific than Enabled accounts for all interactions, as shown in Figure 8b. For instance, the median values for new tweets and new followings were 7 and 62, respectively, compared to just 2 and 8 for Enabled. It should also be highlighted that Suspended D1 accounts had less than 24 hours to interact with the platform and that some of these interactions might have been missed by our technique due to the sampling schedule. Overall, we may suppose that spammy behavior was among the main causes for day 1 suspensions.

The results related to Suspended D15-17 reveal a peculiar activity pattern. Considering the analysis of day 1 in Figure 8a

(orange bars), it can be seen that this subgroup was the least active, in proportion to the total number of accounts in the subgroup, for all kinds of activities. For example, the two activities involving the highest percentage of accounts were new tweets (9.1% of active accounts) and achieving followers (6.1%). Still, as Suspended D15-17 alone includes almost 92% of all the suspended accounts, its number of total active accounts on day 1 (20,419) is about two times the sum of the accounts in the other two suspended subgroups. Therefore, it is not surprising that these accounts managed to have a substantial impact on the total number of active accounts on the platform. In particular, 11.6% of the accounts that produced at least one new tweet on day 1 belonged to Suspended D15-17. In terms of interaction count, the distribution is similar to that of Enabled. Again, due to the large number of accounts in this subgroup, the active accounts in Suspended D15-17 were responsible for the greatest share of interactions from suspended accounts on day 1, with the only exceptions being likes (35.1% of suspended volume) and followings (30.9%). Notably, Suspended D15-17 produced 7.4% of the total number of day 1 new tweets from any account, as well as 12.9% of retweets.

The behavior of Suspended D15-17 after day 1 (Figure 9) is even more revealing: only 1,389 accounts (1.2% of the subgroup total) were active in producing tweets, likes, or in following other accounts. The interaction that involved the greatest number of accounts in the subgroup was achieving new followers (1,052 accounts, 0.9%). Overall, it is plausible to suppose that most of these accounts might have been ''frozen'' by Twitter soon after creation and thus were totally inactive after day one. As illustrated in Figure 6, Suspended D15-17 were finally suspended in bulk around their 16^{th} day on the platform.

For the remaining suspended accounts (Suspended Others, red bars), which represent 4.2% of the total number of suspensions occurred, we can report that the percentage of active accounts for all activities is not particularly high for any interaction with respect to Enabled, the only exception being new tweets on day 1 (about 40% of Suspended Others). In terms of interaction count per active accounts, the distributions show a slightly more productive behavior with respect to Enabled.

Overall, it can be observed that the vast majority of suspended accounts (77.7%) were completely inactive according to our monitoring technique. Either Twitter managed to block these accounts right after they started interacting (and before our monitoring technique was able to capture such interactions), or it took specific measures that actually prevented these accounts from interacting with the platform. Nevertheless, detected interactions from suspended accounts still represent a relevant portion of the total number of interactions from new accounts.

V. CONCLUSION

In this study we have provided insights into the behavior of new accounts on Twitter, starting from an unprecedented dataset of over 500,000 accounts. Thanks to an innovative approach, these accounts were monitored since their registration on the platform.

The first research question (RQ1), inspired by the recent Musk-Twitter dispute, concerned the proportion of ephemeral and potentially malicious accounts among new Twitter accounts. Our analysis of deactivations and suspensions in the first 21 days confirms the relevance of such accounts. Indeed, almost 25% of accounts were suspended in the first three weeks, whereas another 4.3% was deliberately deactivated. For a broader temporal view, we screened the accounts' status two years after creation. The results suggest a lack of sustained engagement with the platform: only 11.1% of the initial set of accounts were still active and produced more than 20 tweets over two years. Moreover, Botometer-based classification revealed that most of these accounts are potential social bots.

The second research question (RQ2) led us to evaluate the timing of deactivations and suspensions during the first 21 days. The timing of suspensions is particularly interesting, as it provides insights into how Twitter enforces its policy against accounts recognized as malicious. A peculiar pattern emerged: about 70% of all the suspensions were enforced on the first hours of day 16, relative to account creation. Also, 90% of the total suspended accounts over the 21-day period stopped interacting with the platform less than 24 hours after registration, as if they were restricted. These results highlight Twitter's attempt to tackle malicious accounts as soon as possible.

The third and last research question (RQ3) concerned the behavior of ephemeral accounts and their impact on the platform. Twitter's early intervention led to the suspension of a large proportion of totally inactive accounts: our method did not detect any interaction for 77% of the new accounts suspended within 21 days from registration. Either Twitter blocked these accounts before they started interacting with the platform or right after the first tweet, which was eliminated before we could detect it according to our sampling schedule. Despite this, ephemeral accounts had a non-negligible impact in terms of content production and social interactions. For instance, considering the total volumes among new accounts, suspended accounts represented 12.4% of the authors of new tweets, while the combined total of deactivated and suspended accounts produced over 10% of tweets and almost 8% of social interactions.

Overall our results suggest that, at least when applied to new accounts, Twitter's reports on the prevalence of malicious accounts might have been too optimistic. As a conservative figure, we estimated that more than 120,000 new accounts are created daily, of which only 3.8% show a high probability of being genuine and long-term active users. Undoubtedly, the high volume of daily account creation poses a significant challenge, in which both users' safety and revenues from advertisers are at stake. It is hoped that Twitter's new ownership will follow through on its promise to reduce the impact of spam and fake accounts on the platform. As a future research direction, it would be interesting to conduct a

similar study on new data, in order to compare the behavior and suspension rate of new accounts with those presented in this paper. This would allow for an evaluation of how the policy changes introduced by the new ownership, including the removal of free API usage, have affected the proliferation of malicious accounts.

REFERENCES

[1] A. Karami, M. Lundy, F. Webb, and Y. K. Dwivedi, "Twitter and research: A systematic literature review through text mining," *IEEE Access*, vol. 8, pp. 67698–67717, 2020.
[2] K. Thomas, D. McCoy, C. Grier, A. Kolcz, and V. Paxson, "Trafficking fraudulent accounts: The role of the underground market in Twitter spam and abuse," in *Proc. 22nd USENIX Conf. Secur.*, 2013, pp. 195–210.
[3] M. Shafahi, L. Kempers, and H. Afsarmanesh, "Phishing through social bots on Twitter," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 3703–3712.
[4] K. Thomas, F. Li, C. Grier, and V. Paxson, "Consequences of connectivity: Characterizing account hijacking on Twitter," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, New York, NY, USA, Nov. 2014, pp. 489–500.
[5] S. Zannettou, T. Caulfield, W. Setzer, M. Sirivianos, G. Stringhini, and J. Blackburn, "Who let the trolls out? Towards understanding state-sponsored trolls," in *Proc. 10th ACM Conf. Web Sci.*, New York, NY, USA, 2019, pp. 353–362.
[6] S. Zannettou, T. Caulfield, E. De Cristofaro, M. Sirivianos, G. Stringhini, and J. Blackburn, "Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web," in *Proc. World Wide Web Conf.*, New York, NY, USA, 2019, pp. 218–226.
[7] C. Duffy. (2022). *Elon Musk Cited This Tool in His Bot Dispute With Twitter, Its Creator Has Thoughts*. [Online]. Available: https://edition.cnn.com/2022/08/09/tech/elon-musk-twitter-botometer/index.html
[8] R. Gallotti, F. Valle, N. Castaldo, P. Sacco, and M. De Domenico, "Assessing the risks of 'infodemics' in response to COVID-19 epidemics," *Nature Hum. Behav.*, vol. 4, no. 12, pp. 1285–1293, Dec. 2020.
[9] P. Zola, G. Cola, A. Martella, and M. Tesconi, "Italian top actors during the COVID-19 infodemic on Twitter," *Int. J. Web Based Communities*, vol. 18, no. 2, pp. 150–172, 2022.
[10] D. Pacheco, P.-M. Hui, C. Torres-Lugo, B. T. Truong, A. Flammini, and F. Menczer, "Uncovering coordinated networks on social media: Methods and case studies," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 15, no. 1, pp. 455–466, May 2021.
[11] S. Lee and J. Kim, "Early filtering of ephemeral malicious accounts on Twitter," *Comput. Commun.*, vol. 54, pp. 48–57, Dec. 2014.
[12] F. Giglietto, N. Righetti, L. Rossi, and G. Marino, "It takes a village to manipulate the media: Coordinated link sharing behavior during 2018 and 2019 Italian elections," *Inf., Commun. Soc.*, vol. 23, no. 6, pp. 867–891, May 2020.
[13] M. Mazza, M. Avvenuti, S. Cresci, and M. Tesconi, "Investigating the difference between trolls, social bots, and humans on Twitter," *Comput. Commun.*, vol. 196, pp. 23–36, Dec. 2022.
[14] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Fame for sale: Efficient detection of fake Twitter followers," *Decis. Support Syst.*, vol. 80, pp. 56–71, Dec. 2015.
[15] K. Thomas, C. Grier, D. Song, and V. Paxson, "Suspended accounts in retrospect: An analysis of Twitter spam," in *Proc. ACM SIGCOMM Conf. Internet Meas. Conf.*, New York, NY, USA, Nov. 2011, pp. 243–258.
[16] M. Gabielkov and A. Legout, "The complete picture of the Twitter social graph," in *Proc. ACM Conf. CoNEXT Student Workshop*, New York, NY, USA, Dec. 2012, pp. 19–20.
[17] A. Roomann-Kurrik. (2013). *Moving to 64-Bit Twitter User IDs*. [Online]. Available: https://blog.twitter.com/developer/en_us/a/2013/64-bit-twitter-user-idpocalypse
[18] J. Echeverria and S. Zhou, "Discovery, retrieval, and analysis of the 'Star Wars' botnet in Twitter," in *Proc. 2017 IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining 2017*, New York, NY, USA, 2017, pp. 1–8.
[19] J. Wright and O. Anise, "Don't@ Me: Hunting Twitter bots at scale," Duo Security, Inc., 2018. [Online]. Available: https://duo.com/blog/dont-me-hunting-twitter-bots-at-scale
[20] C. Yang, R. C. Harkreader, and G. Gu, "Die free or live hard? Empirical evaluation and new design for fighting evolving Twitter spammers," in *Recent Advances in Intrusion Detection*, R. Sommer, D. Balzarotti, and G. Maier, Eds. Berlin, Germany: Springer, 2011, pp. 318–337.
[21] F. A. Chowdhury, L. Allen, M. Yousuf, and A. Mueen, "On Twitter purge: A retrospective analysis of suspended users," in *Proc. Companion Web Conf.*, New York, NY, USA, Apr. 2020, pp. 371–378.
[22] M. Mazza, G. Cola, and M. Tesconi, "Ready-to-(ab)use: From fake account trafficking to coordinated inauthentic behavior on Twitter," *Online Social Netw. Media*, vol. 31, Sep. 2022, Art. no. 100224.
[23] D. Kergl, R. Roedler, and S. Seeber, "On the endogenesis of Twitter's spritzer and gardenhose sample streams," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2014, pp. 357–364.
[24] P. Zola, G. Cola, M. Mazza, and M. Tesconi, "Interaction strength analysis to model retweet cascade graphs," *Appl. Sci.*, vol. 10, no. 23, p. 8394, Nov. 2020.
[25] M. Sayyadiharikandeh, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, "Detection of novel social bots by ensembles of specialized classifiers," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manag.*, New York, NY, USA, Oct. 2020, pp. 2725–2732.
[26] T. R. Keller and U. Klinger, "Social bots in election campaigns: Theoretical, empirical, and methodological implications," *Political Commun.*, vol. 36, no. 1, pp. 171–189, Jan. 2019.
[27] A. Rauchfleisch and J. Kaiser, "The false positive problem of automatic bot detection in social science research," *PLoS ONE*, vol. 15, no. 10, pp. 1–20, 2020.

GUGLIELMO COLA received the Ph.D. degree in computer engineering from the Leonardo da Vinci Doctoral School, University of Pisa, in 2015. He is currently a Researcher of computer science with the Cyber Intelligence Laboratory, Institute of Informatics and Telematics (IIT), National Research Council (CNR), Pisa. Since 2020, he has been leading the Social Media Observatory of the EU H2020 Research Project SoBigData++. His current research interests include social network analysis and the use of wearable sensors for pervasive healthcare.

MICHELE MAZZA is currently pursuing the Ph.D. degree with the Cyber Intelligence Laboratory, Institute of Informatics and Telematics (IIT), National Research Council (CNR), Pisa. His current research interests include social media manipulation where coordinated behaviors are involved, the dynamics of information dissemination, social bots detection, and fake accounts characterization. He is a member of IIT, CNR.

MAURIZIO TESCONI received the Ph.D. degree. He is currently a Researcher of computer science and leads the Cyber Intelligence Laboratory, Institute of Informatics and Telematics, CNR. He teaches the master's courses in cyber intelligence and cyber security. He has published more than 100 articles on social networks, machine learning, and data science in international journals and conferences. His research interests include artificial intelligence, big data, web mining, social network analysis, and visual analytics within the context of open-source intelligence. He is a member of the European Laboratory on Big Data Analytics and Social Mining.

Open Access funding provided by 'Consiglio Nazionale delle Ricerche-CARI-CARE-ITALY' within the CRUI CARE Agreement