Original Research Article

BIG DATA & SOCIETY

Big Data & Society
July–December 2016: 1–13

DOI: 10.1177/2053951716674236
bds.sagepub.com

SAGE

Profile characteristics of fake Twitter accounts

Supraja Gurajala, Joshua S White, Brian Hudson, Brian R Voter and Jeanna N Matthews

Abstract

In online social networks, the audience size commanded by an organization or an individual is a critical measure of that entity's popularity and this measure has important economic and/or political implications. Such efforts to measure popularity of users or exploit knowledge about their audience are complicated by the presence of fake profiles on these networks. In this study, analysis of 62 million publicly available Twitter user profiles was conducted and a strategy to identify automatically generated fake profiles was established. Using a combination of a pattern-matching algorithm on screen-names and an analysis of update times, a reasonable number (~0.1% of total users) of highly reliable fake user accounts were identified. Analysis of profile creation times and URLs of these fake accounts revealed their distinct behavior relative to a ground truth data set. The characteristics of friends and followers of users in the two data sets further revealed the very different nature of the two groups. The ratio of number of followers-to-friends for ground truth users was ~1, consistent with past observations, while the fake profiles had a median ratio ~30, indicating that the fake users we identified were primarily focused on gathering friends. An analysis of the temporal evolution of accounts over 2 years showed that the friends-to-followers ratio increased over time for fake profiles while they decreased for ground truth users. Our results, thus, suggest that a profile-based approach can be used for identifying a core set of fake online social network users in a time-efficient manner.

Keywords

Twitter, fake profiles, online social networks, online social networks, detection, friends, followers

Introduction[1]

Online social networks (OSNs) have become the preferred means of communication among a diverse set of users, including: individuals, companies, and families divided over several continents. Some of the most popular OSNs include: Facebook, Twitter, GooglePlus, Instagram, LinkedIn, Weibo, and RenRen. Of the different OSNs, Twitter has become popular with users such as young adults, governments, commercial enterprises, and politicians as a way of instantaneously connecting with their audience and directly conveying their message (Parmelee and Bichard, 2011). Tweets-based information and communication have begun to play an important role in diverse areas such as political uprisings (Syeed et al., 2014), digital epidemiology (Salathé et al., 2013), and scientific communication (Bennett, 2014). These diverse advances are facilitated by the large size of the platform's audience and the ability to effectively and directly reach their targeted audience. Another advantage for users on OSN platforms is the ability to understand the characteristics of their audience, such as age, location, etc., thus, allowing users to tailor their products or messages as needed (Mainwaring, 2011).

The continued success of Twitter (and other OSNs) as a platform for large-scale communication and the expansion of efforts to mine their data for new and novel applications related to public health, economic

Department of Computer Science, Clarkson University, USA

Corresponding author:
Supraja Gurajala, Department of Computer Science, Clarkson University, Potsdam, NY 13699, USA.
Email: gurajas@clarkson.edu

development, scientific dissemination, etc. necessitates confidence in the authenticity of OSN users. The presence of fake profiles (or Sybils/Socialbots) generated by cyber-opportunists (or cyber-criminals) that are nearly indistinguishable from real profiles complicates authentication of user accounts (Douceur, 2002). Some fake accounts are created to mimic a specific person's account while others are created simply as a general account to serve as a fake follower.

Sybils are automatic/semi-automatic profiles created to mimic human profiles. Fake profiles (or their operators) send requests to "follow" or "friend" OSN users and these requests are often accepted (80% probability when there are several common "friends") by unsuspecting users (Yang et al., 2014). Sometimes, a fake profile is created to essentially duplicate a user's online presence. Such attacks, called identity clone attacks, are devised to collect personal information and direct online fraud. Upon gaining acceptance of a few legitimate users, they then begin to gather "friends" or "followers" because of their ability to accurately fake the accounts of legitimate users.

Even when there is no attempt to impersonate a real person, fake accounts can still cause problems. Many are created to serve as "followers for hire" and are used for inflating follower numbers for other accounts. The presence of fake followers can: affect the popularity rating of individuals or organizations measured based on follower number count (Kwak et al., 2010); alter the characteristics of the audience (Stringhini et al., 2013); or create a legitimacy problem for individuals/organizations (Parmelee and Bichard, 2011).

As a result of these problems, OSNs such as Twitter are interested in mechanisms for identifying and eliminating fake accounts on a near real-time basis. In this work, we present a detection technique that was developed to identify automatically generated Twitter accounts based on analysis of publicly available Twitter user profile information. The profile-based detection allows for fast analysis of the large Twitter user database and allows for identifying possible fake accounts even before they accumulate a "Tweet" history. The characteristics of the fake profile set are then compared and contrasted against a ground truth data set. This paper is organized into the following sections: Related work on spam detection in Twitter; Acquisition of Twitter data; Design and Methodology; Analysis of Results; and Conclusions.

Related work

There have been several detection strategies developed to tackle the problem of Spam in social networks. These techniques have largely relied on using a graph theory approach to characterize social graph properties of Sybil accounts (Danezis and Mittal, 2009). In response, "spammers" have worked to integrate Sybils into authentic user communities by creating accounts with full profiles and background information similar to authentic users (Yang et al., 2011). Such techniques have complicated detection efforts, requiring continued development of new spam-recognition approaches.

Machine learning techniques and honeypot harvesting approaches have been used to classify Twitter accounts as legitimate or not. Honeypot techniques help identify spammers by attracting them to their sites and embedding themselves into their networks with the goal of harvesting information from them (Lee et al., 2010). Machine learning techniques use spammer profile information, obtained via approaches such as honeypot harvesting, to train themselves to understand spammer behavior and thus aiding in the development of detection techniques (Stringhini et al., 2010). Other Twitter-specific approaches to identify spammers and fake profiles include: detection based on tweet-content (e.g. "number of hashtags per word of each tweet" (Benevenuto et al., 2010); use of tweet/tweeter characteristics such as "reputation score", "number of duplicate Tweets", and "number of URLs" (Wang, 2010); and comparison of tweet links (URLs) to publicly blacklisted URLs/domains (Grier et al., 2010).

As the creation and maintenance of fake accounts is generally done automatically, profile-pattern detection provides a fast way of detecting spam without detailed "Tweet" analysis (Thomas et al., 2013). For example, Benevenuto et al. (2010) used profile information such as "number of followers and number of followings" to identify fake profiles. Thomas et al. (2013) used a multivariable pattern-recognition approach based on user-profile-name, screen-name, and email parameters. They used fake accounts purchased from an underground market and determined that there was strong and consistent correlation between the three parameters for all fake accounts. Such a pattern recognition approach could potentially be used as a pre-identifier of fake accounts prior to a more detailed authentication effort.

Most recent approaches to detect fake accounts in Twitter and other OSNs have focused on detection of clustered fake accounts based on their activity patterns. For example, Jiang et al. (2014, 2016) used a method, called CatchSync, to detect suspicious behavior in Twitter based on synchronized and abnormal user activity. They were able to show that their parameter-free approach resulted in high detection efficiency of fake accounts in Twitter. Similarly, Clark et al. (2016) used a classification scheme based on natural language trained on organic users to then identify messages from

automated accounts and detect fake accounts. Such activity-based techniques can identify fake accounts after they establish their tweet-history. A different approach to identify fake accounts is based on analysis of a combination of features including tweets and user profiles for early and efficient identification of fake profiles (Xiao et al., 2015). For example, El Azab et al. (2015) showed that fake accounts on Twitter could be identified with high efficiency based on an established minimum set of factors, including number of followers, availability of geo-information, used a hashtag in a tweet, etc. Xiao et al. (2015) used a supervised machine-learning pipeline to compare text frequencies in features such as name, email address, etc. to classify LinkedIn accounts as either malicious or legitimate.

In this study, we further examine the possibility of identifying fake accounts merely based on profile information to enable early classification of clustered fake accounts in Twitter. We use a user profile-pattern detection based approach with the inclusion of user activity time stamp information, to develop a new process for detection of fake profiles with high reliability. The details of our schema and the analysis of the fake profile set enabled with this approach are presented below.

Design and implementation

For analysis of user profiles, the first step was to obtain publicly available Twitter user profile information. Utilizing social web-crawling, we gathered profiles of ~62 million users and used map-reducing techniques and pattern recognition approaches to detect fake profiles. The details of the Twitter user profile acquisition approach and analysis techniques are discussed below.

Acquisition of Twitter user profiles

Twitter provides access to user and social graph data through the Twitter REST API v1.1.2. (Twitter Inc., n.d.). For non-protected users, the majority of the users' profile information is publicly available and accessible through this API. In order for Twitter to maintain the reliability of its services and control costs, they employ rate-limiting steps, restricting the number of calls that can be made to the API per rate limit window, currently defined as a 15-minute interval. Version 1.1 of the API rate limits an application based on the number of users that have authorized the application, and therefore granted it an access token.

Crawling the social graph

As recently as 2010, it was feasible to crawl over the entire Twitter user ID space. This was at a time when the maximum user ID was estimated to be less than 800 million (Gabielkov and Legout, 2012). With the maximum Twitter user ID significantly larger than 1 billion now, an exhaustive search over the Twitter user ID space has become challenging. Furthermore, the sparsity of the Twitter user ID space also complicates the search process. It has been estimated that spam account creation accounts for as high as 93% of all new accounts, 68% of which are detected and automatically suspended or deleted by Twitter (Dawson, 2013). This would mean that in an exhaustive search, approximately 68% of user IDs (roughly 1.4B) for whom we would request information are accounts that no longer exist or are unavailable due to suspension. This would waste both our and Twitter's resources. With user IDs now exceeding 2 billion, the sparsity of the user ID space, and the rate limits imposed by Twitter with the rollout of the Twitter REST API v1.1, an exhaustive search over the entire user ID space is no longer feasible.

To overcome these issues, our approach performs a breadth first search (BFS) over a given set of seed users. As the social graph is crawled, previously unknown Twitter user IDs obtained from the list of the user's followers are pursued and eventually the user profiles for these IDs are acquired. This ensures that all user profile requests we make to Twitter include only valid Twitter user IDs. Effectively, this adds each previously unknown follower of a user as a seed user for the next iteration of the search. We used a multi-account approach with application access to crawl the Twitter social graph and gathered ~62 million Twitter user profiles, within a three-month period in late 2013.

Methodology

Our crawler obtained 33 different attributes for each Twitter profile, and these attributes are listed in the online supplementary material. Our schema then analyzed patterns among combinations of these attributes to identify a highly reliable core set of fake profiles, which provided the basis for identifying key distinguishing characteristics of fake accounts based on their publicly available profile information. To limit the parameter space of our analysis, we initially investigated the database semi-manually to determine the primary attributes that differed among most users. From this analysis, it was established that several of the 33 attributes were largely unused (or left as default) by most users. The key attributes that were either user-selected or varied with account-usage were: *id*, *followers_count*, *friends_count*, *verified*, *created_at*, *description*, *location*, *updated*, *profile_image_url*, and *screen_name*. Using this reduced attribute set, we used the analysis approach described below for identification of a reliable fake profile set.

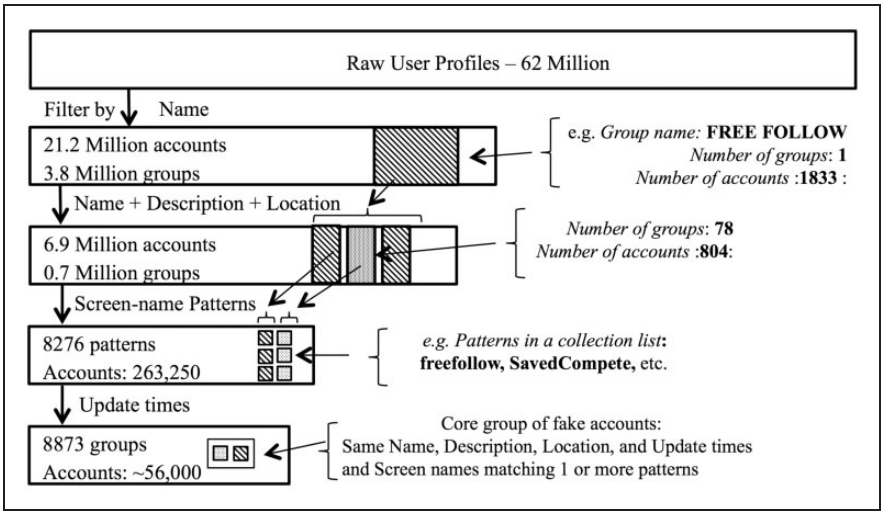

Figure 1. Schematic diagram of the algorithm for identification of fake profile groups (Gurajala et al., 2015).

A schematic diagram outlining our algorithm for detection of the core fake group is shown in Figure 1. Starting with our 62 million user-profile database, we obtained groups (containing at least two accounts) with the same user profile information of name, description, and location. This filtering process resulted in generating 724,494 groups containing a total of 6,958,523 accounts. To further refine the 724,494 groups, their screen names were analyzed and near-identical ones were identified using a pattern-recognition algorithm procedure described below.

To identify screen name patterns in the 724,494 groups, a Shannon entropy-based analysis of the screen names in each group was conducted. For this, first, one of the screen names in a group was selected as a base screen name and its Shannon entropy was determined. Second, the next screen name in the group was concatenated with the selected base name and the Shannon entropy of the concatenated string was calculated. If the entropy of the concatenated string was greater than that of the base name by a threshold value (0.1), then the concatenated screen name was added to a collection list. This entropy comparison with the selected base screen name was repeated with all screen names in the group. All screen names that were not accumulated in the collection list associated with the first screen name were then re-grouped and analyzed with the above described pattern recognition procedure to generate other collection lists. This procedure was repeated until all screen names were either placed in a collection list or identified as not being a part of any collection. This procedure results in the division of the 724,494 groups into several collection lists with cohesive screen name entropy distributions.

A regular expression pattern (more than four characters long) search was then conducted within each collection list to obtain any pattern(s) that might exist in their screen names. The screen names associated with a pattern formed a ''pattern list''. From a manual inspection of the pattern lists, it was determined that this procedure was able to identify and group mass fake profiles with screen names that were seemingly generated automatically (e.g. freefollow1, freefollow3, etc.). The procedure was even able to group profiles with a common string well-hidden within the screen name, with substantial lengths of prefixes and suffixes to them. We did notice, however, that the entropies of some pattern lists were not tightly bound, suggesting that the entropy filtration procedure needed to be revised. This was accomplished by analyzing the broadness of the Shannon entropy distributions of the screen names in each pattern list, which was quantified by the normalized standard deviation of their entropies σ as shown in equation (1).

$$\bar{\sigma} = \frac{\left[\sum_i^N (x_i - \bar{x})^2\right]^{\frac{1}{2}}}{\bar{x}} \quad (1)$$

where N represents the total number of screen names in a pattern list, x_i is the Shannon entropy of a ith screen name, and $\bar{x}$ is the mean Shannon entropy of a pattern list.

A refined pattern list was generated by eliminating all pattern lists with relative standard deviations greater than a critical value (0.03; established empirically during this study). An example of two pattern lists with varying distributions of their Shannon entropies is shown in Figure 2. Of the two pattern lists, one with a narrower relative Shannon entropy distribution (Figure 2a; manually confirmed to be composed of patterned screen names) was retained, while the other (Figure 2b;

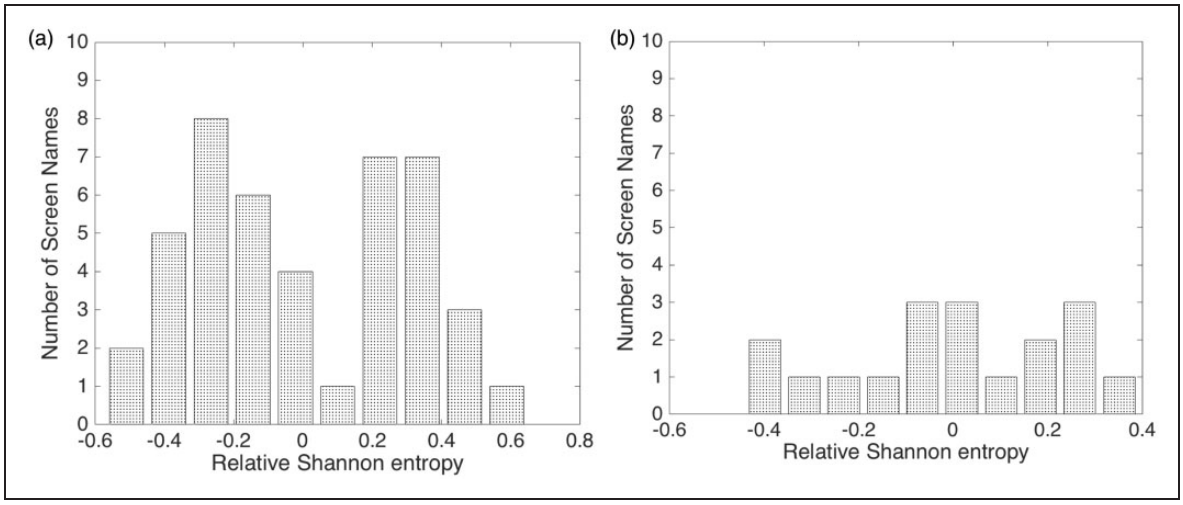

Figure 2. Entropy distribution of a group with (a) low and (b) high normalized standard deviations of Shannon entropies. Note that the values plotted in the *x*-axis are Shannon entropy values relative to the mean entropy value in each list.

screen names that were not obviously patterned) was eliminated. This step helped ensure a high likelihood that the obtained group of accounts was mostly composed of fake accounts.

The procedure, thus far, identifies closely associated accounts that are very likely to be automatically generated. This procedure, however, will have false positives associated with highly popular names (e.g. we identified a list with the names ''Chopra'' many of which were manually inspected to be genuine). As a final filter, the accounts were examined to determine their distribution of update times. The update times of a pattern list with genuine accounts are likely to be uncorrelated and hence have a broad distribution, while the fake accounts that are automatically updated from a single operator will likely have closely related update times. In Figure 3, update time distributions of two manually evaluated pattern lists are shown. The update times of a false positive list (''Chopra'') was seen to be uniformly distributed while the distribution for a fake accounts' list was relatively non-uniform. Pattern lists with broad update time distributions were then eliminated from our collection.

The above procedure results in the generation of a fake profile set that contains 8873 groups with ~56,000 accounts that have identical name, description, location, and a narrow range of update times. In addition, the accounts in each group had screen names with matching patterns. Further investigation of the fake profile set revealed that all the identified accounts were updated not just within a narrow time distribution but as groups, i.e. with identical update times. While this list is not very large in size, the highly refined constraints applied to produce this fake profile set ensure a high likelihood of reducing false positives. This was confirmed with a manual inspection of the activity of some randomly chosen accounts in the fake profile set and no false positives were identified during this inspection. The characteristics of this highly reliable set of fake profiles were then analyzed to determine their profile-based distinguishing features.

Results

To determine the characteristics of this set of fake accounts, the generation of a ground-truth data set was required. The ground-truth data set was obtained from a random sample of our Twitter user profile database, with the assumption that majority of Twitter accounts are authentic, and hence their random collection would represent largely authentic users. For consistency with the identified fake group data sets, different ground-truth group data sets were generated such that they were of similar size and from a similar timeline as each of the fake group data sets. Analysis of the updated times, creation times, and profile URLs of the two data sets was then conducted to understand the relative characteristics of the two sets.

Update times

A comparison of the update times of all 8873 groups of fake profiles and of the generated truth data set is shown in Figure 4. The truth data set was randomly divided into a similar number of groups and group sizes to closely match the fake profile set. The update times of the truth data set (Figure 4a) were seen to be almost uniformly distributed over the entire day, with no obvious time bias. The update times of the fake profile set (Figure 4b), as observed earlier, were non-uniformly

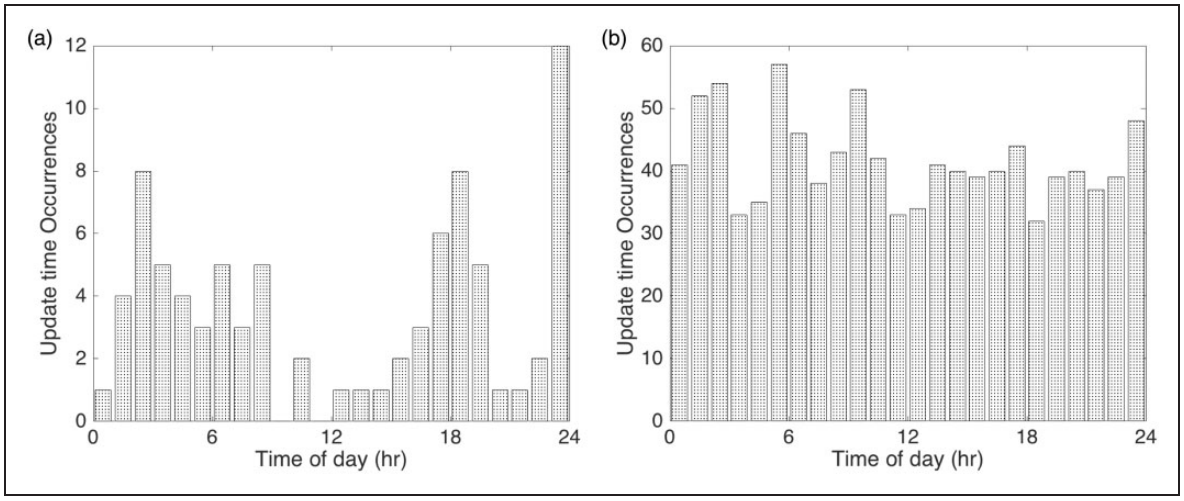

Figure 3. Histogram of update times of two accounts with screen name patterns: (a) Chopra and (b) Freefollow.

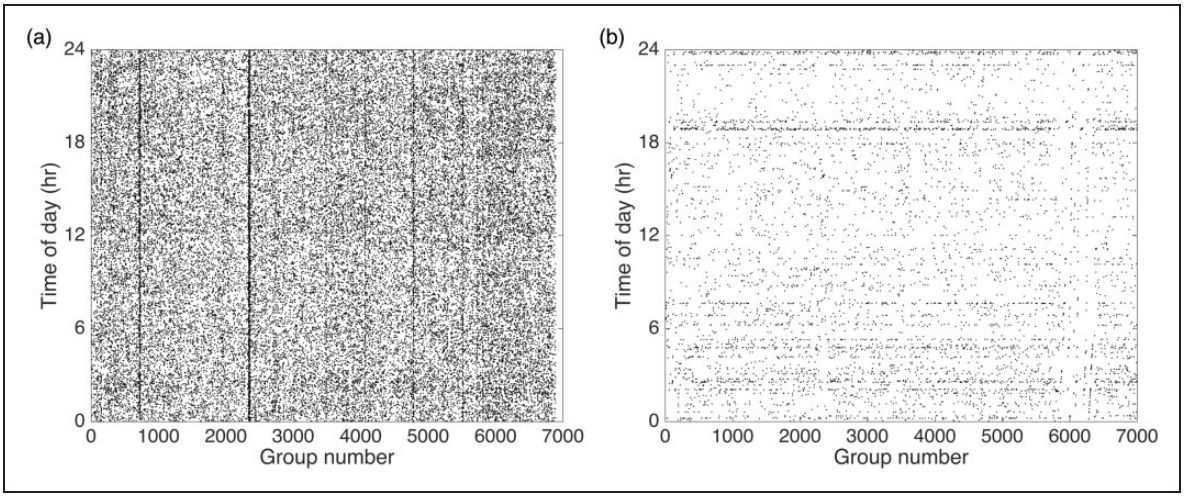

Figure 4. Comparison of update times for (a) ground truth and (b) fake profile data sets. The update time of an account in a group is indicated as dark point in the figure. For the fake profiles, as multiple accounts were updated at the same time, the indicated dark points overlap and hence seem less numerous in number.

distributed with significant time periods in a day when there was no update activity. Analysis of our data set showed that the maximum number of profiles updated at a given time was 100 and this limit was achieved by three groups in the fake profile set.

The distribution of the days of the week when the groups in the fake profile set and the ground truth dataset were updated is shown in Figure 5. For the ground truth data, the frequency distribution reveals that these users preferentially updated on Sundays and Mondays (UTC time). A decreasing number of users updated as the week progressed. Considering that the data is in UTC time (and could not be converted to local time, as location information was not always available), there is some uncertainty in the actual update days. The distribution of update days for the fake profiles was seen to have a highly non-uniform distribution, with a bias for update during the later part of the week.

Creation times

The difference in the update time distributions of the two data sets provides confirmation of the distinct nature of our generated fake profile set. It is, however, not entirely surprising that the two data sets have different update characteristics, given that the update time was a factor in filtering the data set. A better measure of the difference between the two data sets is the distribution of creation times.

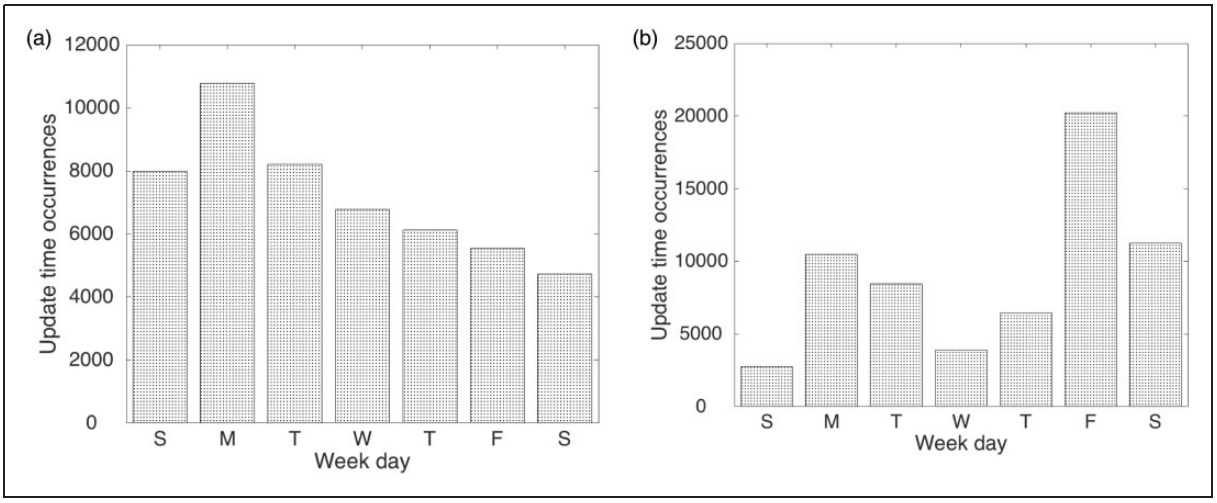

Figure 5. Occurrence frequency of update days for (a) ground truth and (b) fake profiles.

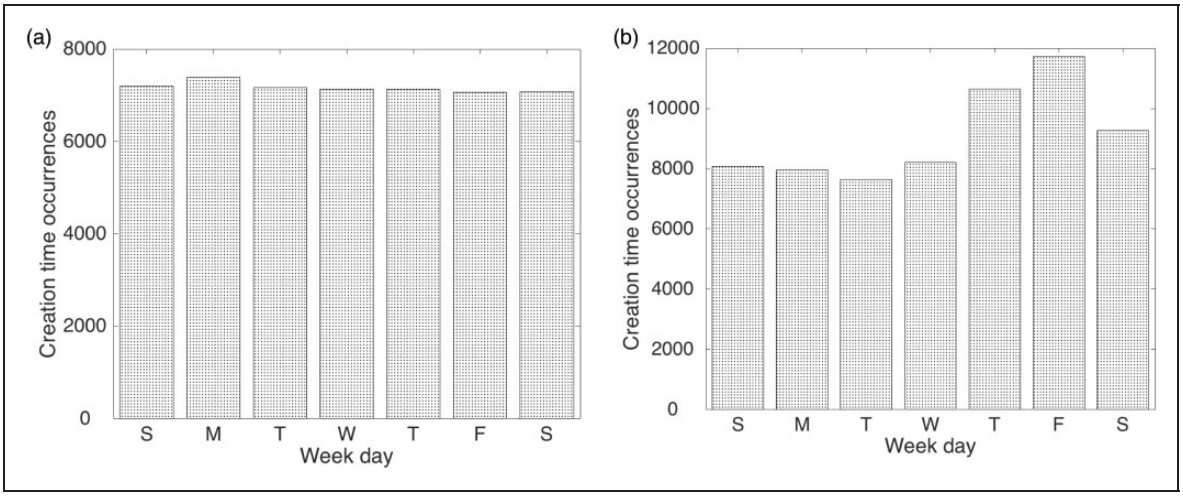

Figure 6. Occurrence frequency of creation days for (a) ground truth and (b) fake profile data sets.

The distribution of days of the week when profiles in the two data sets were created is shown in Figure 6. For the ground truth data, the distribution is nearly uniform (Figure 6a), with no preference for any particular day of the week. This suggests that a typical legitimate user would create a Twitter profile any day of the week. For the fake profile set, the creation days are biased towards the later part of the week (Figure 6b). While it is not obvious why this bias exists, it could feasibly relate to a possible automated element in the creation of these fake profiles.

The creation times for the different groups in the two data sets are shown in Figure 7. The ground truth profile creation times were largely distributed uniformly during the day, with some reduction in the number of created accounts during the 5–10 hour time period (Figure 7a). The distribution of fake profile creation times (Figure 7b) was, however, seen to be very different from the ground truth data set. The creation times were significantly more non-uniformly distributed during the day than the ground truth data set. It could be concluded that the accounts in the fake profile set were, as opposed to the previously stated inference, created in batches.

The creation rate of fake profiles was investigated by first sorting the creation times of the two sets and then calculating the interval times between consecutive profile creations in each data set. The creation times were then binned in geometrically spaced intervals and the resultant distributions are shown in Figure 8. The median time intervals for the identified fake profiles were seen to be at least one order of magnitude less

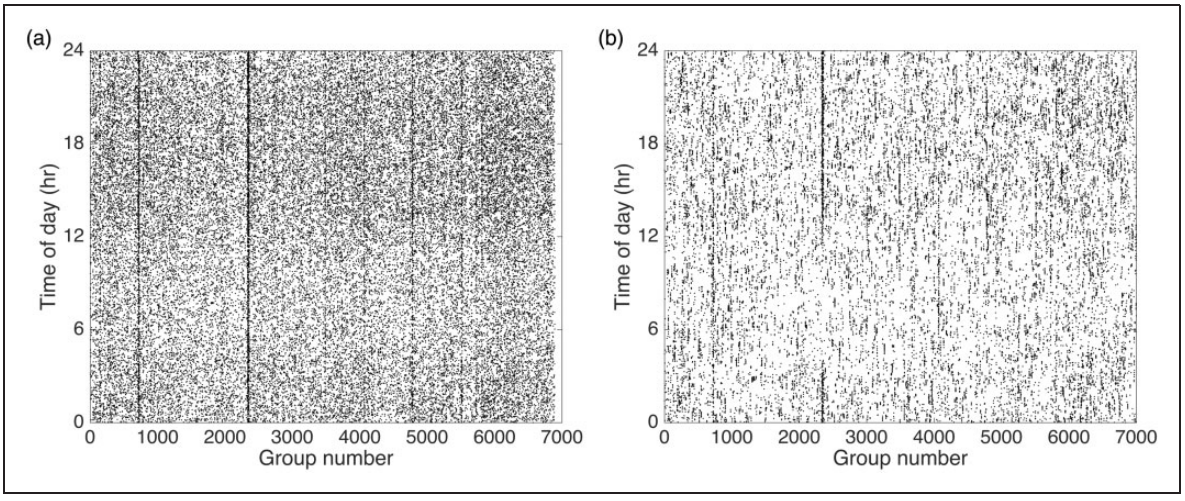

Figure 7. Comparison of creation times for (a) ground truth and (b) fake profile data sets. The creation time of an account in a group is indicated as dark point in the figure. For the fake profiles, as multiple accounts were created as close clusters, the dark points often closely placed and hence seem less numerous in number.

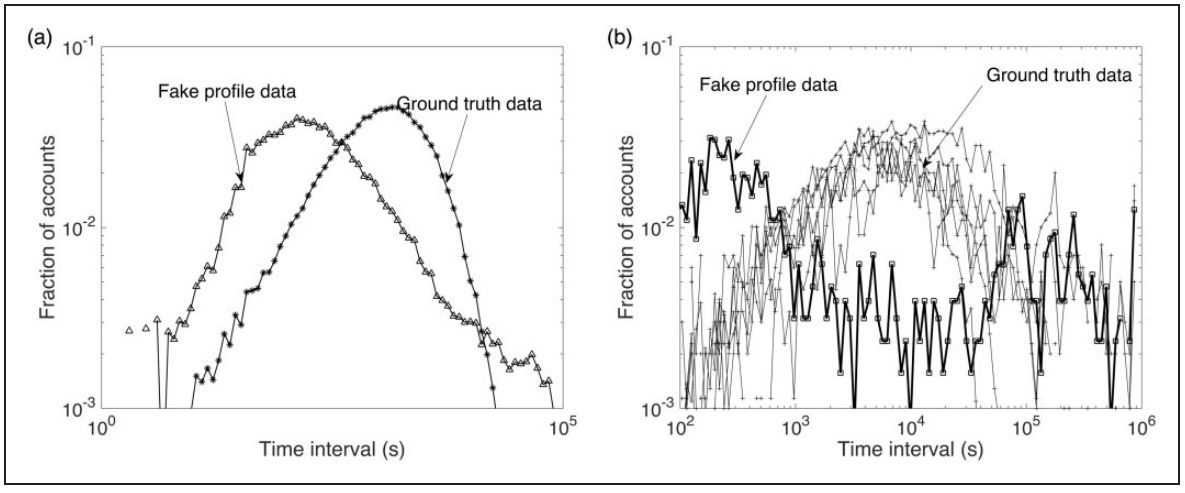

Figure 8. The fraction of accounts created in a selected time interval for the fake profile and ground truth data sets for two cases: (a) entire data set of the fake profiles and a ground truth data set of similar size; (b) limited data set of manually confirmed fake profiles compared against different groups of ground truth data sets.

than that of the ground truth profiles (Figure 8a). The faster creation times of the fake profiles are consistent with our earlier observations of batch-creation of these accounts. The shortest creation time difference between two account creations in the fake profiles group was generally ~20–40 seconds, with some groups exhibiting even faster generation rates (3–5 seconds).

In Figure 8(b), the creation time interval distribution of nine large groups in the fake profile list was compared against ground truth groups of similar sizes. The ground truth data for this comparison was obtained using a variety of aggregation approaches, including random collection and using profiles with matching popular first and/or last names. The ground truth creation time distributions were seen to be mostly invariant relative to each other, and had median interval times significantly larger than that of the fake profiles. A two-sample t-test analysis confirmed that, independent of the aggregation approach used, the creation time distributions of the ground truth data were distinct from that of the fake profiles (p-value less than 5e−4). The median interval times of the ground truth data were seen to be significantly larger than that of the fake profiles. Another interesting observation that can be made relates to the similar distributions that arise from comparing the data sets at long creation time intervals. This could suggest that the profiles contributing to the tail of the fake profile distribution represent

the false positive fraction of the fake profile set, and these were ~1% of the total number of profiles.

URL analysis

The profile_image_url attribute allows users to upload an image to personalize their account. To determine the diversity of the URLs in the fake and the ground truth data sets, the Shannon entropy values of the URLs were obtained for the different groups in the two data sets. The normalized standard deviation (equation (1)) of the entropies in each group was then determined. The complementary cumulative distribution function (CCDF) of the normalized standard deviations of Shannon entropies was calculated as shown in equation (2).

$$CCDF(x) = 1 - \frac{\sum_{x_i < x} E(x_i)}{\sum_{x_i < \infty} E(x_i)} \quad (2)$$

where E is the number of groups with a selected normalized Shannon entropy (x_i). For the fake profiles, the CCDF (Figure 9) shows that the URLs are not very dissimilar compared to the URLs of the ground truth data. A large fraction of profiles in the fake group were actually seen to have similar or the same URL, resulting is very small values of normalized standard deviation of entropies.

For fake accounts, even when their URLs were very different, the images were seen to be the same. For one of the groups in the fake profile set containing 659 accounts, a collage of images from distinct URLs is shown in Figure 10. The number of distinct images for the 659 accounts was just 14. Thus, using image analysis of profile URLs, we could further refine our groups within the fake profile set.

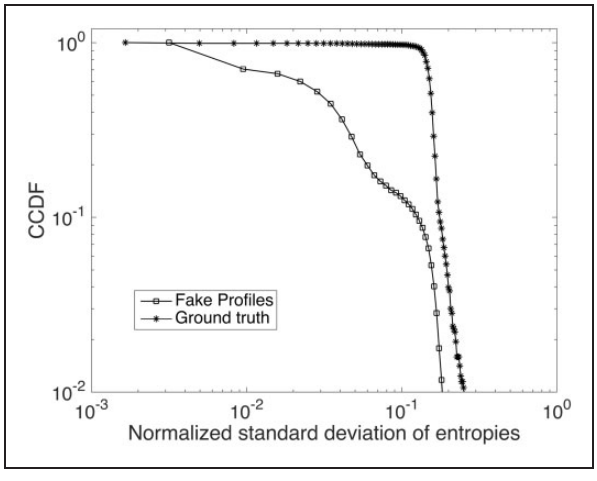

Figure 9. Complementary cumulative distribution function of the normalized standard deviation of URL entropies.

Characteristics of friends and followers

To further examine the difference in the nature of two data sets—fake and ground truth—the characteristics of distribution of friends and followers of users in the two data sets were analyzed. The friends and followers information for the two data sets was obtained at two different times—October 2013 and October 2015—to understand their temporal evolution. Considering that the users in the two data sets have a range of creation dates, and hence a range of time periods to build up friends and followers, for an effective comparison, we normalized the number of friends and followers by the number of days since a user's account was created.

In Figure 11, the cumulative distribution functions (CDFs) of the normalized number of followers are shown for the fake and ground truth accounts determined at two instances (October 2013 and 2015). The users in the ground truth data set are seen on an average to have accumulated higher followers per day than the fake users, consistent with past observations for fake Twitter users (e.g. Stringhini et al., 2013). The low growth in the number of users in the fake data set suggests that the users we identified as fake were not focused on building a reputation, but on providing a follower service for others. Another observation from our analysis is that the number of followers per day was growing for the ground truth data but was stagnant for the fake users. This is consistent with the expectation that genuine users will accumulate more followers over time. Another difference is in the shape of the distribution curve of the number of followers for the two data sets. For the fake users, the distribution is narrower (i.e. the CDF is sharper) than it is for the ground truth users, suggesting that the fake users were more similar in their follower characteristics than the ground truth users.

The cumulative distribution of the number of friends for the two sets (Figure 12) shows that the fake users have a larger number of friends than the ground truth users. These results are in contrast to that observed for the number of followers. Also, both data sets witnessed an increase in the rate of friend accumulation with time. With increasing time, the distribution of friends per day for fake users is seen to broaden, while it is largely unchanged for ground truth users. The broadening of the fake users' friends' distribution is driven by the increasing efficiency of users at the upper end of distribution in gathering friends relative to those at the lower end of the distribution. Note that, Twitter's rules limit the number of accounts that a user can follow and this limits the rate at which a user can build up their number of friends.

A commonly used measure to compare the characteristics of fake and real users is the ratio of the number of friends to followers. In Figure 13, this ratio is shown

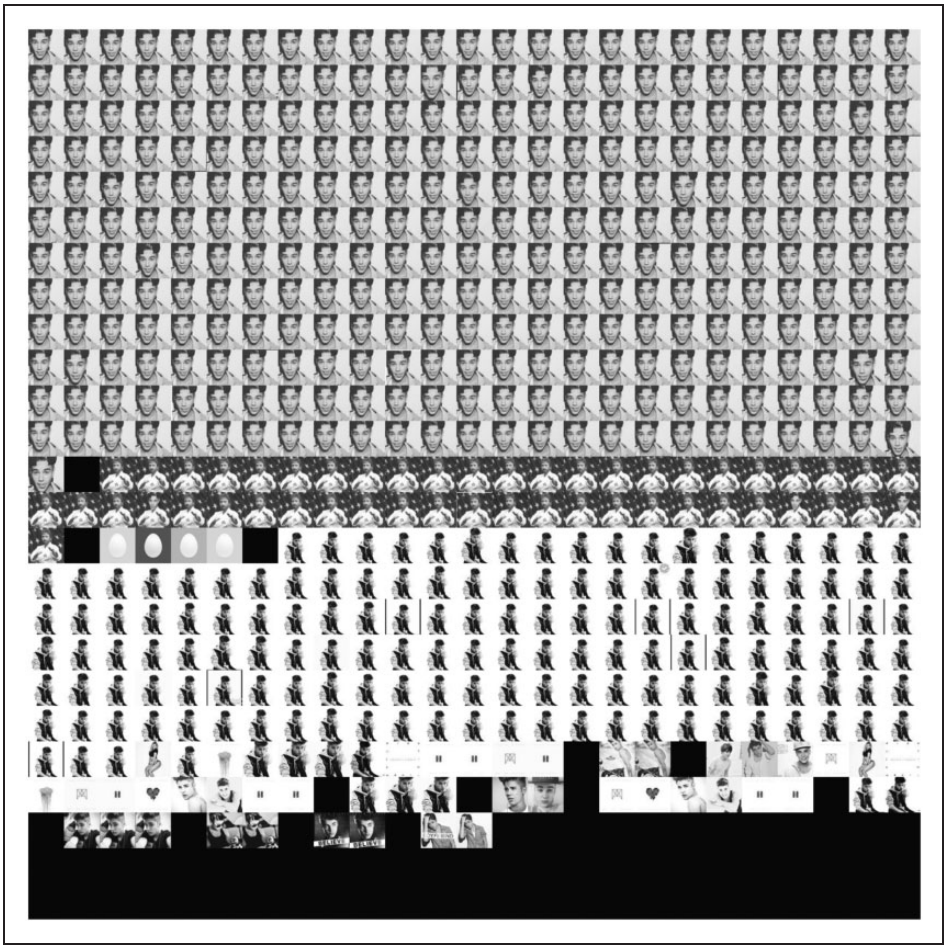

Figure 10. Collage of images from distinct URLs for a group of 659 accounts within the fake profile set (Gurajala et al., 2015).

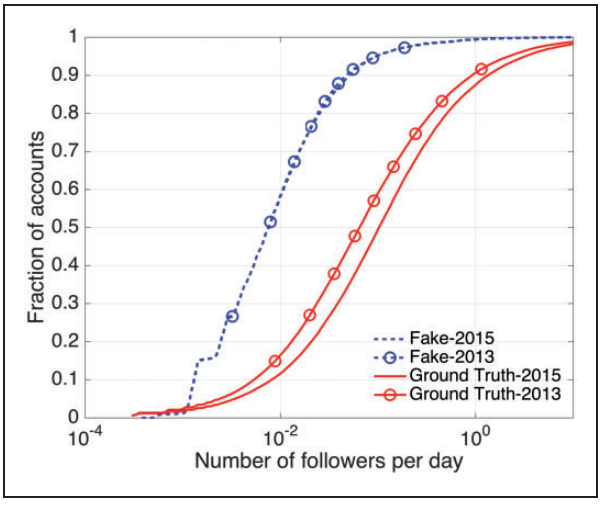

Figure 11. Cumulative distribution of normalized number of followers for the fake and ground truth data sets examined at years 2013 and 2015.

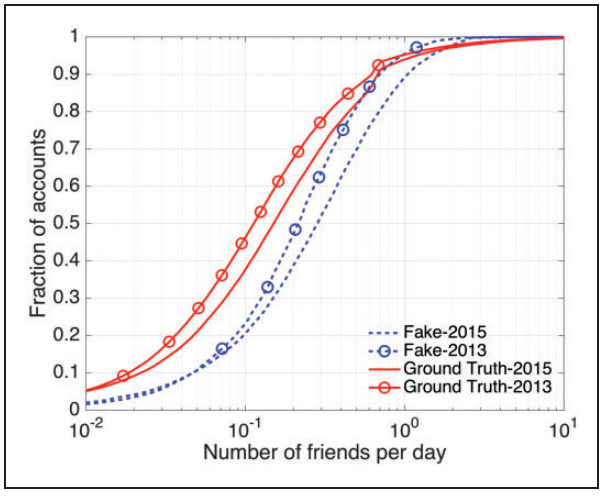

Figure 12. Cumulative distribution of the normalized number of friends for the fake and ground truth data sets examined at years 2013 and 2015.

for the two data sets on October 2013 and 2015. The CDF ratio plots reveal the very different characteristics of the two data sets. The ground truth users have a median friends-to-followers ratio of ~1, consistent with findings of other researchers (e.g. Stringhini et al., 2013). The fake users were seen to have much higher friends-to-followers ratio, indicating that the fake users we identified were likely primarily geared to accumulate friends. With increasing time, the ratio for the ground truth users increases slightly but decreases for the fake users. The temporal trend in the ratio distribution for the fake users indicates both their success in gathering friends and their comparative inability to gain followers.

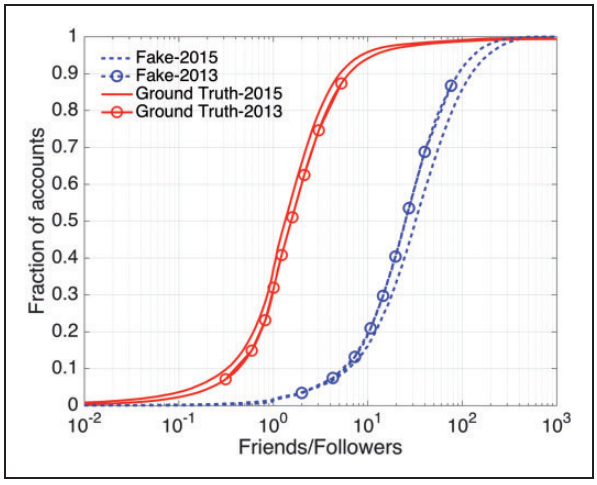

Figure 13. Cumulative distribution of ratio of friends to followers for the fake and ground truth data sets examined at years 2013 and 2015.

Tweet analysis

To further understand the nature of our identified fake profile set, we examined the tweets of users in a few randomly chosen fake profile groups identified in this study. Twitter provides content and date of the last tweet sent out by a user as part of their profile information. We obtained the last tweets of users in the different fake profile groups (as of January 2016) and a sample set of tweets are listed in Table 1. Also included in Table 1 are the number of accounts tweeting the same text (listed under "Sample Tweet" column), the range of update times for these tweets (described in the "Update Info" column), and the date of the last tweet (listed in the "Update Info" column).

Analysis of last-tweets of our fake profile set shows that groups of accounts are not just updated near-simultaneously, but the users in the group also often tweet the same text. Also, in general, the fake users were not seen to be very active in sending out tweets. These tweet-based observations provide some validation of the clustered nature of our fake profile users and the semi-automatic nature of the management of these accounts. A common theme emerging from the tweet snapshot in Table 1 is the consistent effort of all these accounts to obtain followers, possibly to lower their friends-to-follower ratio.

Conclusions

To identify and understand the characteristics of fake accounts in Twitter, we used a crawler to gather a large user profile database of 62 million user accounts. A highly reliable clustered fake profile set was generated by grouping user accounts based on: matched multiple-profile-attributes; patterns in their screen names; and an update-time distribution filter. A subset of the

Table 1. Sample last tweets obtained (in January 2016) for a select set of fake profile users that were manually examined. The common tweets and their similar update times point to the clustered nature of the accounts that we identified.

Number of accounts	Sample Tweet	Update info
32	Who wants to trade free follows? I have 38! Tweet...	all accounts updated within 3 minutes of each other; Aug 2013
27	TrADE FOLLOWS? I HAVE 60 x	all updated within 1 second of each other; Sept 2014
64	TRADE FREE FOLLOWS I HAVE 182, TWEET @ ME	all updated within 3 seconds of each other; Dec 2014
105	"."	All updated at the exact same time; Dec 2015
111	"MANI BEAR! FOLLOW @The1Dgangasta..."	all updated at the exact same time; Jan 2014
183	@Harry_Styles Follow me! Step 2: Give yourself a pat on the back because you just made me the happiest girl alive	all updated within 5 seconds of each other; Sept. 2013

accounts identified as fake by our algorithm were manually inspected and verified as all being fake (based on their Tweet activity). Analysis of the characteristics of the fake profile set revealed that these fake profiles were almost always created in batches and over intervals of less than 40 seconds. These accounts were created preferentially on some weekdays and during select times of the day, suggesting a manual element in their generation. The creation time characteristics of our identified fake profile set were very different from that of a ground truth data set of similar size. The URLs of the fake profile set were seen to have lower diversity than the ground truth images and even the distinct URLs of these profiles corresponded to just a few different images. An analysis of the friends and followers of the two data sets revealed the very different nature of the users in the two groups. Compared to the ground truth users, fake users were seen to accumulate friends faster and followers slower. This suggests that our approach primarily identifies fake users that were generated to create followers for others. From an examination of the characteristics of the friends and followers of the two groups at two times (October 2013 and 2015), the temporal evolution of the profiles was studied. The ground truth and fake users were seen to have a very different evolution with time. The fake users were on an average becoming more efficient in gathering friends with time, but less efficient with respect to followers. A further confirmation that the fake users we identified were indeed fake was obtained when we noticed that ~10% of the accounts identified as fake in 2013 were inactive/suspended by 2015.

Our activity-based profile-pattern detection scheme provides a means to identify potential spammers without detailed analysis of their ''Tweets''. One limitation of our approach is that it only identifies a relatively small percentage of fake accounts. But the low numbers of false positives in the obtained fake profiles make this set an ideal seed database for use with social graph techniques for efficient and fast spam detection.

Our fake account detection approach is distinctive from the other published approaches as it is based on analysis of user-profile features rather than tweet histories to detect fake accounts. While analysis of tweet history allows for a more accurate detection of fake accounts, the proposed profile-based fake detection approach allows for early detection of at least a subset of accounts which can shutdown or may be flagged for further monitoring. From our analysis of the 2016 Twitter user database, we determined that most of the fake users (~93%) identified in this study are still active, suggesting that our proposed approach can be complementary to existing fake profile/spammer identification techniques.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplementary material

The supplementary files are available at http://bds.sagepub.com/content/3/2.

Note

1. This work is based on an earlier work: Fake Twitter accounts: profile characteristics obtained using an activity-based pattern detection approach. In: *Proceedings of the 2015 international conference on social media & society*. © ACM 2015, http://dx.doi.org/10.1145/2789187.2789206

References

Benevenuto F, Magno G, Rodrigues T, et al. (2010) Detecting spammers on twitter. In: *Collaboration, electronic messaging, anti-abuse and spam conference*, p.12.

Bennett E (2014) Social media as a tool for improving research and teaching. *Frontiers in Ecology and the Environment* 12(5): 259.

Danezis G and Mittal P (2009) SybilInfer: Detecting sybil nodes using social networks. In: *NDSS*.

Clark, EM, Williams JR, Jones CA., et al. (2016) Sifting robotic from organic text: a natural language approach for detecting automation on Twitter. *Journal of Computational Science* 16: 1–7.

Dawson S (2013) Is Twitter telling the truth about their ''active user'' stats? Available at: http://eggsbacon.co.nz/twitter.

Douceur JR (2002) The sybil attack. *Peer-to-Peer Systems*. Berlin Heidelberg: Springer, pp. 251–260.

Gabielkov M and Legout A (2012) The complete picture of the Twitter social graph. In: *CoNEXT Student '12 Proceedings of ACM conference on CoNEXT student workshop*, New York, NY, USA, pp.19–20.

Grier C, Thomas K, Paxson V, et al. (2010) @ spam: The underground on 140 characters or less. In: *Proceedings of the 17th ACM conference on computer and communications security*, pp.27–37.

Gurajala S, White JS, Hudson B, et al. Fake Twitter accounts: profile characteristics obtained using an activity-based pattern detection approach. In: *Proceedings of the 2015 International Conference on Social Media & Society*, July 2015, p. 9. ACM.

Jiang M, Cui P, Beutel, A, et al. Detecting suspicious following behavior in multimillion-node social networks. In: *Proceedings of the 23rd International Conference on World Wide Web*, April 2014, pp. 305–306. ACM.

Jiang M, Cui P, Beutel A, et al. (2016) Catching synchronized behaviors in large networks: A graph mining approach.

ACM Transactions on Knowledge Discovery from Data (TKDD). Vancouver.

Kwak H, Lee C, Park H, et al. (2010) What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on World Wide Web*, pp.591–600.

Lee K, Caverlee J and Webb S (2010) Uncovering social spammers: Social honeypots + machine learning. In: *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*, pp.435–442.

Mainwaring S (2011) *We first: How brands and consumers use social media to build a better world*. New York: Palgrave Macmillan.

Parmelee JH and Bichard SL (2011) *Politics and the Twitter Revolution: How Tweets Influence the Relationship Between Political Leaders and the Public*. Lanham, MD: Lexington Books.

Salathé M, Freifeld CC, Mekaru SR, et al. (2013) Influenza A (H7N9) and the importance of digital epidemiology. *New England Journal of Medicine* 369(5): 401–404.

Stringhini G, Kruegel C and Vigna G (2010) Detecting spammers on social networks. In: *Proceedings of the 26th annual computer security applications conference*, pp.1–9.

Stringhini G, Wang G, Egele M, et al. (2013) Follow the green: Growth and dynamics in twitter follower markets. In: *Proceedings of the 2013 conference on Internet measurement*, pp.163–176.

Syeed N, Zafar R, Asaad R, et al. (2014) Arab women rising: 35 Entrepreneurs making a difference in the Arab World. *Knowledge@ Wharton*.

Thomas K, McCoy D, Grier C, et al. (2013) Trafficking fraudulent accounts: The role of the underground market in Twitter spam and abuse. In: *USENIX Security*, pp.195–210.

Twitter Inc. (n.d.) The Twitter. Available at: https://dev.twitter.com/rest/public.

Wang AH (2010) Don't follow me: Spam detection in twitter. In: *Proceedings of the international conference on Security and Cryptography (SECRYPT)*, pp.1–10.

Xiao C, Freeman DM and Hwa T. Detecting clusters of fake accounts in online social networks. In: *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security*, October 2015, pp. 91–101. ACM.

Yang Z, Wilson C, Wang X, et al. (2014) Uncovering social network Sybils in the wild. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 8(1): 2 Available at: http://arxiv.org/abs/1106.5321.

This article is part of a special theme on Social Media & Society 2014. To see a full list of all articles in this special theme, please click here: http://bds.sagepub.com/content/social-media-society.