Bot or Not

Husna Siddiqui, Elizabeth Healy, Aspen Olmsted
Department of Computer Science
College of Charleston
Charleston, SC 29401
siddiquih@g.cofc.edu, healye@g.cofc.edu, olmsteda@cofc.edu

Abstract—In recent years, Twitter, a social networking website, has been affected by a steady rise in spam on its network. Hijacking of social media accounts has become a modern-day danger. Motivations for this can range from attempts in identity theft to simply skewing the perception of an audience. In this paper, we extend our previous work, Engineering Your Social Network to Detect Fraudulent Profiles, by doing an investigation of spam bots on Twitter. We propose an algorithm that will distinguish a spam bot, from a genuine user account by using a JavaScript testing framework that consumes Twitter's REST API. We ran a dataset of 700 Twitter accounts through our algorithm and identified that roughly 11% of the dataset were bots.

Keywords-spam; social network;bot; API; Twitter;

I. INTRODUCTION

Social media is changing the world. Networking websites such as Twitter, Facebook, and Instagram have become some of the most popular activities. For many, it is an important aspect of daily life. These websites influence the news that users consume on a day to day basis. Spam bots are automated accounts that mimic real users. In some cases, these bots can be harmless. For example, there is a prime numbers bot that systematically tweets prime numbers. However, some bots can be malicious and are intended to tamper with commonly tracked statistics by posting automated and false information.

In this paper, we extend our previous results [1] to determine the validity of a user account on Facebook. Here we will investigate the detection of bots on Twitter. To do this, we use Twitter's REST API, application program interface, to collect and analyze user data. The organization of this paper is as follows: Section II will describe the works that are related to our research area. Section III further explains the motivation for our research. Section IV explains our implementation method and data gathering techniques. Section V gives an overview of our results. Finally, Section VI will give a conclusion of our work.

II. LITERATURE REVIEW

The most important document we will consider is Twitter's developer API. The API will allow us to integrate with Twitter and programmatically access Twitter data, so we can gather a database of records to run through our algorithm.

In our previous work [1], we proposed an improvement to the calculation of a validity score for a Facebook user account to determine if the profile is real or fake. We used empirical analysis, to define the attributes that impact a user's validity score. Each of these attributes was given a weighted score, which was then aggregated to determine a validity score.

In his paper, Detecting Spam Bots in Online Social Networking Sites, Wang extracts Twitter data to discover spam bots. Wang analyzes three graph based features which are the number of friends, the number of followers, and the follower ratio. He also looks at three content-based features, which are the number of duplicate tweets, the number of HTTP links, and the number of replies from the user's 20 most recent tweets. [2] This data is run through various classification methods to determine if an account is a spam bot or not. We will use findings from Wang's research as a basis of our knowledge when identifying which features are most significant in identifying a spam bot.

In Securing e-Loyalty Currencies, Olmsted analyzed user activity on various social networks to determine the validity of a user's social network rewards. To validate the authenticity of a user account, particularly for Twitter, the following account data was used: the presence of an account, the number of people following this account, number of accounts this account follows and the number of times the account has tweeted a microblog. [3]Each attribute has a maximum cap to ensure that one feature does not contribute too much weight to the validity of a user.

Freitas, Benevenuto, Ghosh and Veloso designed an experiment to understand infiltration strategies of socialbots on Twitter. They created 120 socialbot accounts with different characteristics and strategies to investigate what constitutes a successful bot. [4] Their study reveals findings that are key for the detection and counter measurement approaches for bots on Twitter.

III. KNOWLEDGE GAP

Twitter's appeal is in the use of "tweets." A tweet is a status message that is limited to 140 characters. This warrants that a tweet is concise and can be easily scanned. If you see a tweet that you find interesting, you can "like" the tweet to let the user know that you enjoy their content. Additionally, you can "follow" the corresponding account to view and get updates on their account activity. Twitter also has functionality that allows users to "retweet" a tweet which instantly shares and spreads information. Popular topics on Twitter include politics and news, sports, fashion, and pop culture. Users, can skim through tweets and read the trending

topics to get a quick update on what is happening around the world.

TABLE 2. Sample Range of Attribute values and Scores

Twitter Attribute	User Range	Score
ratio	0, <=0.2, >= 0.3	10, 5 0
default profile image	False or true	0, 15
keywords in tweets	Zero, 1-2, >= 3	0, 3, 15
url patterns in statuses	Zero, 1-2, >= 3	0, 3, 15
Statuses-count (frequency)	0 -3.0, 3.1-5.0, 5.1-10.0, > 10.0	0, 5, 10,15
verified	False or true	0, 10

Currently, Twitter has 319 million monthly active users. [5] Based on research from the University of Southern California and Indiana University, up to 15% of these are bots. This means that roughly 48 million accounts are bots, not humans. We strive to analyze Twitter data to expose these bots and uncover commonalities that may relate them.

Spam bots can be used for malicious acts such as the spread of fake news, cyber-stalking, spread of malware and clickjacking. We propose a way to detect bots and educate users so that they are aware of these issues.

IV. Research and Impementation

Our first step was to identify characteristics of a successful Twitter bot. By utilizing knowledge from the research community, we determined the following attributes to be the most significant: ratio of number of followers and number of following, profile image, keywords in tweets, url patterns in tweets, whether it is a verified account and the status count. Each of these attributes provides a different weight towards the calculation of a score to uncover a spam bot. TABLE 1 gives sample extended rules that are used to assign a score. Each attribute has a specific maximum score. This max score is based on the predictive significance of each metric. For example, the verification of an account contributes a smaller weight than the ratio of followers versus following.

The next step is to gather a dataset by using the Twitter API. We use the followers/list.json method to acquire details on 700 random twitter accounts. This endpoint returns a JSON object that contains fields such as username, id, verification-status, followers, friends, status count and profile image. We extract relevant values from the JSON object. After we gather our dataset, we build an array of the usernames that were returned. We use these usernames to make an API call to the search/tweets.json endpoint which requires the username and a query string that contains the keywords we want to search for. We target the following keywords in a user's status to flag it is a bot: offer, free, click, prize, debt, deal, credit, and sex. The last API call hits the search/tweets.json endpoint, and this time we look for url patterns in a user's tweet.

After we record the results and have fully iterated through the array of users, we run a script to iterate through the array to assign a score based on each attribute value. TABLE 2 details the process of assigning a spam bot score. If an attribute falls within a range, the user is given a score for that attribute. Finally, the user is assigned a total score by adding up the points for each attribute. An overall score of 40 or more signifies that the account is a spam bot.

TABLE 1. Sample User Attribute Points

Sample Attribute	Score	Max
ratio	10	10
default profile image	5	15
keywords in tweets	15	15
url patterns in statuses	2	15
Statuses-count	15	15
verified	15	10
Total Points	77	80

V. Results

Fig. 1 shows the results of our implementation. We ran a sample database of 700 Twitter accounts through our algorithm and identified 79 accounts as bots, which is roughly 11% of our dataset.

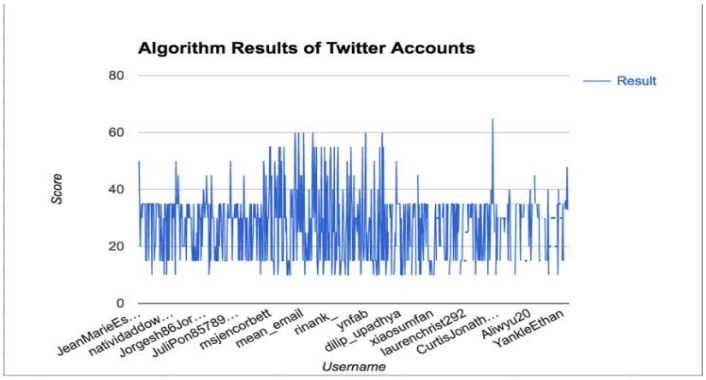

Figure 1. Implementation Results

VI. Conclusion

In this paper, we propose a method to identify spam bots on Twitter. Our solution is based on empirical analysis and a JavaScript testing framework that uses Twitter's REST API. We identified roughly 11% of our dataset to be spam bots.

References

[1] A. O. Z. D. Husna Siddiqui, "Engineering your social network to detect fraudulent profiles," in *Information Society (i-Society)*.

[2] A. H. Wang, "Detecting Spam Bots in Online Social Networking Sites: A Machine Learning Approach.".

[3] A. Olmsted, "Securing e-Loyalty Currencies," *Jounal of Internet Technology and Secured Transactions.*

[4] F. B. S. G. a. A. V. Carlos Freitas, "Reverse Engineering Socialbot Infiltration Strategies in Twitter".

[5] M. Newberg, "CNBC," 20 03 2017. [Online]. Available: http://www.cnbc.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.html. [Accessed 20 04 2017].