Twitter: Identifying Spam Accounts

By: Erik Hencier, Logan Margulis, Jennifer Shriver

Mentor: Hani Dawoud

Table of Contents

Abstract

Spam on the web has many different characteristics and is a serious problem in terms of network slow downs and malicious intent. In particular, spam accounts on Twitter can inhibit the free flow of accurate information, slow-down or even disrupt service. Spam can also deceive users into clicking their malicious-intent links. Our initial goal was to better understand Twitter spam accounts, in hopes that we can discover a foolproof way to detect spam accounts. To begin this process we researched scholarly journals for background information on Twitter and attempts to identify spam accounts. This gave us a base to start our initial outline of our research. As a team, we determined three common characteristics we thought would identify a spam account: the number of tweets per day, the number of links per tweet, and the number of mentions per tweet. From the research that we previously read we thought that spam accounts would show a significantly higher volume of data compared to real accounts. From there we collected several different real accounts and potential spam accounts to compare these characteristics. Using Java code in NetBeans IDE platform we crawled through the information provided by the accounts and Twitter API to obtain data we could analyze. Using this data we created graphs to represent the data in a readable format to show our results. While we felt that we had collected a significant amount of data and were confident in the characteristics that we chose to identify spam accounts, our results showed that it is difficult to determine a spam account based only on these characteristics. Further, we discuss other options and techniques we could use if we were to continue this research.

Introduction

Our group consisted of Logan Margulis, Erik Hencier, Jennifer Shriver, and our mentor Hani Dawoud. We had a diverse group of undergraduate students, one of which has already been in the workforce so it brought a lot of different perspectives during our research process. Our entire group was dedicated to the team and our cause which helped us progress faster and more effectively to complete the project. Hani was a great help with the Java code being used to interpret the Twitter API, as we did not have any experience coding in this language. It would have been a huge learning curve that we really did not have as much time for throughout the semester. That being said we did all the research ourselves and learned a great deal from the entire process.

Statement of Research Problem

Background and Related Work

To fully understand the Twittersphere in terms of spam, their malicious behaviors, as well as personal accounts and other types of non-personal accounts, we scavenged academic articles related to this subject. We felt it was necessary to get a firm background from the beginning before we jumped to any conclusions. However, part of our gain from our research project was from practical application and trial-error type strategies. In other words, one can learn through trial-and-error better than simply reading academic articles.

The first article we read to become introduced to the concept of identifying spam accounts on Twitter was “Don’t Follow Me: Spam Detection on Twitter” by Alex Hai Wang. In this article, Wang, created a social graph based on the relationships developed on Twitter. He characterized those relationships as a follower, a friend, a mutual friend, and a stranger. Wang felt examining the relationships of Twitter users, as well as identifying possible characteristics to determine a spam account would ultimately help him single out spam accounts. Wang found his results to be somewhat inconclusive as well. Initially, he thought that all spam account would be following a large amount of users, while having little or no followers themselves. This was not the result. Based on the conclusion of Wang’s research, we thought that we could try to build upon his idea in our own project.

In “Detecting Spammers on Twitter” by Fabricio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida, the research was data intensive. We did not have resources like the team from the Universidade Federal de Minas Gerais Brazil. They had gathered a large dataset of Twitter accounts that included over 54 millions users, 1.9 billion links, and nearly 1.8 billion tweets. However, we did find something to add to our research project. The team had trouble with finding potential spam accounts on Twitter. The Brazilian team’s project found potential spam accounts on Twitter via the top 3 global trends happening during the time. Using this strategy, the team had no problem finding spam accounts for analysis. This being said, Twitter’s own spam detecting algorithms are sufficient and we had to analyze these potential spam accounts quickly otherwise Twitter would remove them from api and its website database. The Brazilian team’s efforts were approximately 70% effective in classifying spam accounts and 96% effective in classifying non-spammers.

In “Mutually Reinforcing Spam Detection on Twitter and Web” by Nikita Spirin, the team gained a unique insight that we added to our own research project. Spirin focus in large part on the analysis of the URLs shared by user accounts on Twitter. She called this analysis “spammicity” of users that share these links. Assigning accounts this “spammicity” score derived by her web spam detection algorithms helped her better understand spam on twitter. She proposed a new set of URL features for study on different Twitter user accounts. As she did, however not in as much great detail considering our beginner status as researchers, we added the analysis of the number of links per tweet as a characteristic.

Research Methodology

Our research project is based on Twitter and identifying spam accounts within its platform. Spam accounts have been known to clog the system with mass tweeting, as well as following and unfollowing other users. Twitter is the main micro blog site on the Internet today and any disruption on Twitter will affect the efficiency in which information is spread. Therefore, we feel that there needs to be a stronger way or algorithm to detect spam accounts that destroy the main purpose of Twitter. To try and identify spam accounts, we chose three properties that would show the differences between spam accounts and real accounts. The properties we chose were: the number of tweets in a day, the number of mentions in a tweet and the number of links in a tweet. In this process, we were somewhat limited in the characteristics we could use due to our knowledge of Java and coding techniques. Within the properties themselves, we designated certain time periods in which we would record information to keep consistent research across our entire group. This way there could be a solid conclusion using all of the evidence we had collected.

We chose several real accounts and potential spam accounts to compare these characteristics. We used different types of real accounts; such as business accounts, personal accounts and entertainment or news accounts to try to cover the whole spectrum of accounts on Twitter. We developed our own basic methodology in order to hunt down spam accounts to do research. We used our own knowledge as Twitter users to determine if accounts were spam. We also used rules and definitions given by Twitter to determine a spam account. For example, common tactics listed by Twitter that are used by a spam account include: posting harmful links and unrelated links, including links to phishing or malware sites; aggressive following behavior, including mass following and unfollowing; abusing @ functions to post unwanted messages; repeatedly posting trending accounts to attract attention; and repeatedly posting duplicate updates.This methodology allowed us to have a good base for our analysis and to try to stay on the same page throughout the team.

We used three different Java based web crawler programs to pull data from our collected accounts, as well as Twitter API. The Net Beans IDE was the most useful application as it allowed us to integrate our Java code as well as all of the packages that were needed to interpret the API. The first program we used pulled the follower identification numbers from the chosen accounts. This identification number is different that the common username on Twitter. It consists of a string of numbers, uniquely identifying that user. What this program told us is how many followers each chosen account had. This was part of our initial research to find spam accounts, as typically spam accounts have significantly less followers than the amount they follow. The second program we used pulled the most recent 200 tweets from each account we chose. We did this every two days for ten days in intervals to obtain the maximum data from each account. The data included within these files included the date and time of the tweet, ID/string, text, source, HREF, location, name, protected status, follower count and friends count, to name a few. From there we used a Json file reader to convert all this information into a format we could easily read and understand. The third program we used specifically determined the number of tweets per day, number of mentions, and the number of links in a tweet for each account. This final program data is what we based our final results upon and from which we created our representation graphs.

One limitation we encountered was that we had a limited amount of queries per IP address that we could input per 24 hours. This meant that we had to time our search correctly to achieve the maximum results for our final analysis. We each kept a schedule to achieve our goal at the correct with the limitations that we were working around. Another limitation we found was that when we did find a spam account we only had a small time frame to collect the data we needed from it before it was taken off of Twitter. Twitter does have its' own spam reporting system, which is used to take accounts completely off Twitter that have been identified by other users as spam. While this did not throw off our analysis, it did limit us to a certain extent.

Analysis

Once we collected all of our data, we created several graphs to represent the data we collected. In the creation of these graphs, we had to make sure the intervals in collecting data were correct to show strong results in our final presentation. As a group, we decided these graphs would define our semester long project and show the effort we had put in to our project as a whole. While the text and methodology would still be part of our final project, we felt that the graphs really personified our comparison of the different accounts and clarified how exactly we went about our research process. The clearer our process was defined, the easier it would be to gather other researchers in the future to help us with our cause and possibly create a stronger spam algorithm. To come up with the graph representations we gathered all the data we collected. We grouped it together by the three characteristics we chose and decided to show a percentage or probability of the set of data. We felt that this would clearly show the data collected for each account. Our goal was to have conclusive results when we started the project but obviously as shown by our conclusion, this was not the the case.

Results

We pulled data from over 8,000 tweets and used 22 different Twitter accounts. What we found is that there is not much variance between the real accounts and the potential spam accounts that we chose, in terms of the three characteristics we chose. As shown in figure 1.1, there was variation in the number of tweets per day throughout the whole spectrum of accounts. However, when you compared the different types of accounts we looked at within the graph there was little variation. For instance, if you look at the real account IndianaUniv and the spam account IAmEvilTebow, they both have an average of

25 tweets per day. There are other accounts that have similar results to these two within the graph. Similarly, in graph 1.2, there is even less variation in the percentages of of mentions of all the accounts shown. We expected to see higher amount of mentions specifically by the spam accounts. In graph 1.3 we see that there are two accounts with very high url counts in their tweets, however, the rest of the accounts only have an average of about 30% of links in their tweets. Looking at each of these graphs, we can tell that it is hard to determine which account is a real account and a spam account. This is also true when we compare all three characteristics individually for each account. This ultimately tells us that there is little difference between the real accounts and spam accounts we chose. We thought that there would be significantly more tweets per day, mentions in a tweet, and links in a tweet from spam accounts, but that was not the case. We found that there are several business accounts and personal accounts that identify with these characteristics as well, making our initial hypothesis inconclusive.

Conclusions

In conclusion, we found that it is difficult to determine and identify a spam account based only on the three characteristics we chose. Again, these characteristics were the number of tweets per day, the number of mentions in a tweet, and the number of links in a tweet. Looking back at the research we have conducted, we feel like there were some areas we could have improved upon. For instance, maybe the spam accounts that we found were not actually spam accounts. Spam accounts are becoming more and more sophisticated to look like real accounts. It is possible that the reason why we did not get conclusive results is because we did not have any spam accounts. Also, maybe we should have narrowed down the type of real account that we would be comparing to the spam account. We looked at business accounts as well as personal accounts for this research. What we found is that several of the business accounts showed a high amount of the spam characteristics we identified. That tells us right there that it will be hard to determine what accounts are real or spam without a large sample size. Additionally, if we did not have the time constraint of only one semester to complete the project, we could have done more extensive research before choosing a set of characteristics to compare. This would have given us better insight into what kind of characteristics to chose or to conduct our research in a complete different direction. We did have a few limitations throughout the course of our research, including a limited query search through Twitter API and deletion of spam accounts we were researching. We feel like there are limitations to every research project and we felt like we adapted to those limitations and came up with a good plan to get the best results we could. Even though our final outcome did not confirm our thoughts from the beginning of the project, we learned a tremendous amount about from the research we conducted.

Recommendations for Future Work

Throughout the semester, we as a group, enjoyed working with the Twitter API as it was something that we could all relate to as daily Twitter users. Near the end of the semester, we felt that we had accomplished a base in terms of characteristics we were using and our common research goal. That said, we agreed that are additional processes we can go through and properties we can use in addition to our prior research to create stronger evidence as well as come to a distinct conclusion. If we did not have the time constraint of only one semester, there would definitely be other techniques that we would want to explore. By using other analytical techniques we could create more professional and conclusive analytical research which we could pass on to others and that he/she would fully be able to comprehend.

Initially, we felt that three properties we had discussed were enough to come to a conclusion with significant evidence. As mentioned earlier these were: the number of tweets in a day, the number of mentions in a tweet, and the number of links in a tweet. What we soon realized was that although these displayed evidence as we added new data week by week, it just was not enough. As a group, we would have like to been able to add more properties such as common followers and common followees. This would create the opportunity to interpret a network of spammers and allow us to attempt to find a common source, such as a corporate hacking ring. With that information, we may have actually been able to create a strong algorithm to begin the elimination the spam accounts via the networks in which they were created from.

It would have also been interesting to find a way to determine what kind of site was contained in the link posted by an account. As we discussed earlier, a common trait of a spam account was to post links to malicious sites. With the limited sources that we were allotted, the only way to know what kind of site a link contained, was to click on it. If there was a way to create an algorithm or Java program to analyze the contents of the link posted, we think that could help identify spam accounts.

Near the end of the project timeline, we realized that we should have tripled the number of accounts that we were pursuing. Unfortunately, this would not have been possible during the semester due to the small size of our research group. Thus, if we had more members of the group to communicate with, we could have create a stronger statistical backing which may have brought us in a straighter path to our conclusion. Having a large sample size is essential in having strong evidence to support our claim of an issue with a heavy amount of spam on Twitter. Therefore this would allow us to create a spam detection algorithm to show how significant of an issue this is to the developers of Twitter in their daily traffic.

Citations

Alex Hai Wang “DON’T FOLLOW ME: SPAM DETECTION IN TWITTER” College of Information Sciences and Technology, The Pennsylvania State University
http://test.scripts.psu.edu/students/h/x/hxw164/files/SECRYPT2010_Wang.pdf

Fabr´ıcio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virg´ılio Almeida “Detecting Spammers on Twitter” Universidade Federal de Minas Gerais, Belo Horizonte, BZ
http://ceas.cc/2010/papers/Paper%2021.pdf

Nikita Spirin. “Mutually Reinforcing Spam Detection on Twitter and Web.” University of Illinois at Urbana-Champaign, IL.
https://agora.cs.illinois.edu/download/attachments/30422540/ReportCS512.pdf

Appendix

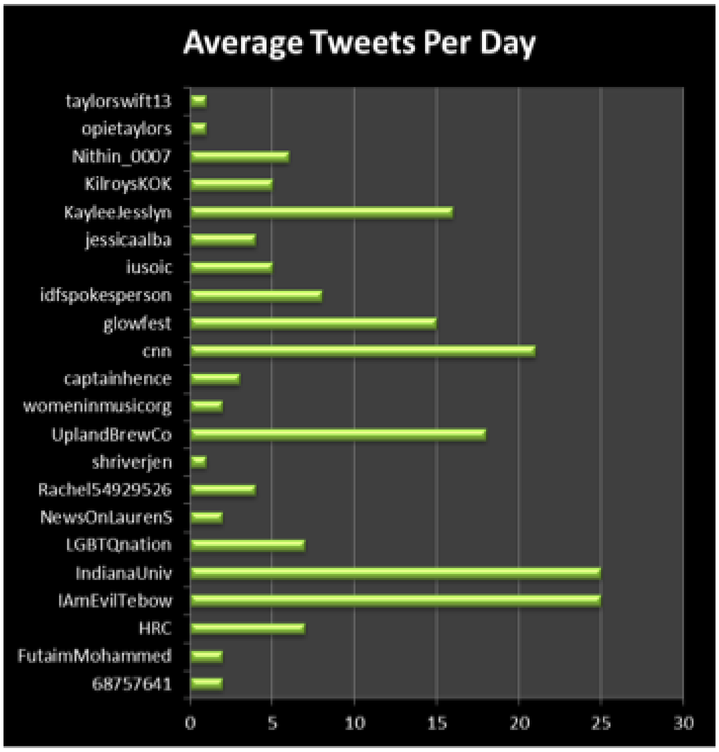

Figure 1.1

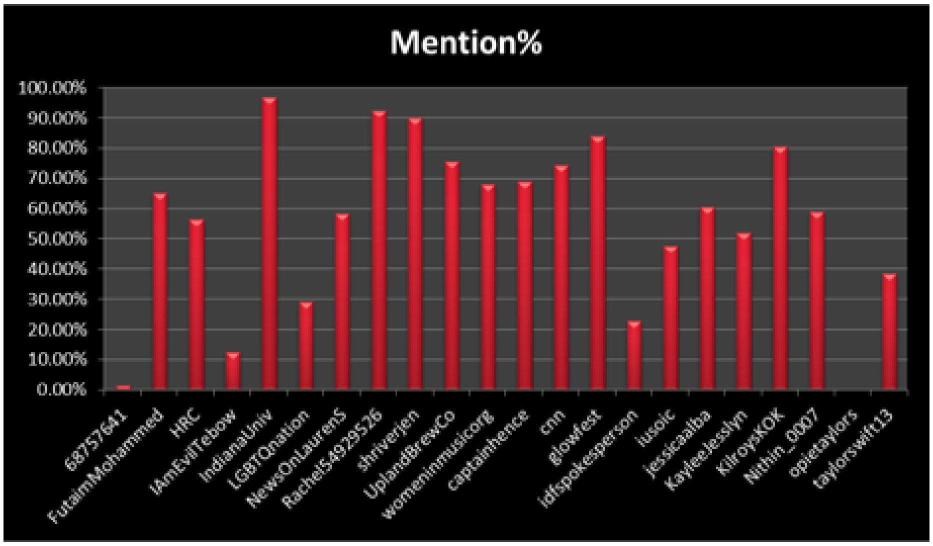

Figure 1.2

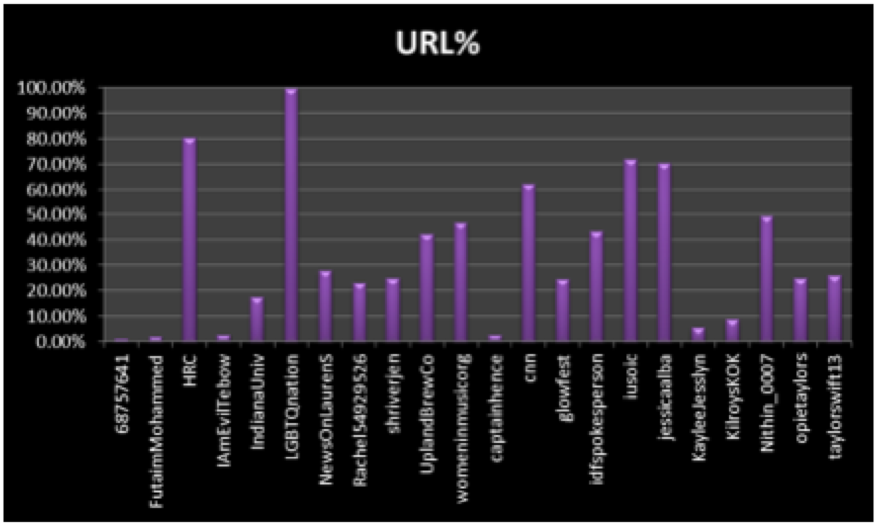
URL%
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
68757641
FutaimMohammed
HRC
IAmEvilTebow
IndianaUniv
LGBTQnation
NewsOnLaurenS
Rachel54929526
shriverjen
UplandBrewCo
womeninmusicorg
captainhence
cnn
glowfest
idfspokesperson
iusoic
jessicaalba
KayleeJesslyn
KilroysKOK
Nithin_0007
opietaylors
taylorswift13

Figure 1.3