

An Evaluation of the Effect of Spam on Twitter Trending Topics

Grant Stafford

Department of Computer Science
Pomona College
Claremont, California 91711
Email: grant.d.stafford@gmail.com

Louis Lei Yu

Department of Mathematics and Computer Science
Gustavus Adolphus College
Saint Peter, Minnesota 56082
Email: lyu@gustavus.edu

Abstract—In this paper we investigate to what extent the trending topics in Twitter, a popular social network, are manipulated by spammers. Researchers have developed various models for spam detection in social media, but there has been little analysis on the effects of spam on Twitter’s trending topics. We gathered over 9 million tweets in Twitter’s hourly trending topics over a 7 day period and extracted tweet features identified by previous research as relevant to spam detection. Hand-labeling a random sample of 1500 tweets allowed us to train a moderately accurate naive Bayes classifier for tweet classification. Our findings suggest that spammers do not drive the trending topics in Twitter, but may opportunistically target certain topics for their messages.

I. INTRODUCTION

In recent years, social media services have grown to become important mediums of communication. The distributed nature and the massive scale of these services have created an environment whose patterns of content generation and consumption are not yet well understood. Researchers have searched for ways to identify characteristics that will cause topics to trend in social networks, while the competition for attention in this environment has inevitably led to the emergence of spam. We look at the relationship between trending topics and spam in Twitter, one of the largest online social networks. Although previous research has investigated trending topics in Twitter and spam in social networks, spam in the Twitter trending topics has not been investigated. The incidence of spam in the Twitter trending topics is of special interest due to their potential high visibility to users worldwide.

In order to investigate how trending topics are affected by spam, we applied a four step approach. First, we gathered data on over 9 million tweets relating to trending topics over a 7 day period. Second, we extracted features from the raw data and hand-labeled a sample of 1500 tweets to train a classifier. Third, we applied the classifier to filter the data. Finally, we evaluated the effects of spam removal on the trending topics by comparing pre-filter and post-filter results.

We found that the frequency of spam among tweets containing trending topics did not differ significantly from those in Twitter overall. Using a chi-squared goodness of fit test to evaluate the hypothesis that all topics are affected equally, we found that topics varied greatly in the percentage of spam they contained. Regardless of this, re-ranking topics by popularity after the filter was applied only rarely changed the pre-filter rankings. Analysis of these results suggests that spammers

do not change which topics will trend in Twitter, but may opportunistically target some topics over others.

The rest of the paper is organized as follows: Section II gives an overview of related work. Section III outlines our data collection strategy and data analysis methods. Section IV gives the result of our analysis. Finally, Section V gives some conclusion and discussion.

II. BACKGROUND AND RELATED WORK

A. Twitter

Twitter is a social networking and micro-blogging service founded in 2006. Users post updates called *tweets* containing up to 140 characters of text and HTTP links. The tweets posted by a user are shared on the newsfeeds of the user’s followers. Twitter users often use hashtags to identify the topic of their messages. For example a message containing the hashtag “#baseball” would be related to baseball.

Hashtags or keywords that appear the most frequently in tweets at a given time appear in Twitter’s list of trending topics. The list is then displayed on the sidebar of users’ newsfeeds. The trending topics are valuable for informing users of current trends, but their visibility could make them a potential target for spammers seeking user traffic.

B. Spam Detection in Twitter

Before we start our investigation, it is important to have a clear definition of Twitter spam. Although there are many vulgar and banal messages in Twitter, they do not meet our definition unless they either contain a URL to a website unrelated to the topic of the tweet, or are retweets in which legitimate links are changed to illegitimate ones, obfuscated by URL shorteners. This criterion was effectively employed to detect Twitter spam in [1].

Spam detection in social networks is a relatively recent area of research. Most of the research in this area follows the same general method of detection: 1) use empirical study to select some structural or textual features to examine; 2) use classification and machine learning techniques with these features to find patterns across users and messages; 3) evaluate whether models based on the patterns are effective in detecting unwanted behavior.

There are two main papers that motivate our research on Twitter. The first is by Benevenuto et al. [1]. In their work, the

authors examined spam detection in Twitter by first collecting a large dataset of more than 54 million users, 1.9 billion links, and 1.8 billion tweets. After exploring various content and behavior attributes, they developed a SVM classifier with high precision, and were able to detect spammers with 70% precision and non-spammers with 96% precision. As an insightful follow up, the authors used χ^2 statistics to evaluate the importance of attributes used in their model. They gave the following ranking of the top 10 attributes: 1) fraction of tweets with URLs; 2) age of the user account; 3) average number of URLs per tweet; 4) fraction of followers per followee; 5) fraction of tweets the user had replied; 6) number of tweets the user replied; 7) number of tweets the user received a reply; 8) number of followees; 9) number of followers; 10) average number of hashtags per tweet.

The second paper with direct application to spam detection in Twitter is by Wang in 2010 [2]. Wang motivated his research with the statistic that an estimated 3% of the messages in Twitter are spam. The dataset used by Wang was smaller than the dataset used by Benevenuto et al. [1], covering a 3 week period and gathering information from 25,847 users, 500 thousand tweets, and 49 million follower/friend relationships. Wang developed decision tree, neural network, SVM, and naive Bayesian models using the following features: number of friends, number of followers, reputation (based on ratio of followers to followees), number of pairwise duplications, number of mentions and replies, number of links, and number of hashtags. After testing the models on a set of 500 manually classified user accounts he found that the naive Bayes classifier performed the best, with an F-measure of 0.917. He concluded that reputation, the percentage of tweets with links, and the reply signs are the best features for spam detection in Twitter.

C. Spam Detection in Other Online Social Networks

Although our research concentrates on Twitter trending topics, there is a large body of work on spam detection in other online social networks which can provide useful insights (e.g. [3]). One relevant paper to our research is by Chen et al. [4]. The authors examined comments on Chinese news websites and used reply, activity, and semantic features to develop an SVM classifier with 95% accuracy at detecting paid posters. Another approach taken to detect spammers is by Lee et al. [5]. The authors created special honeypot user accounts on MySpace and Twitter and recorded the features of users who interact with these accounts. They then used these features to develop a classifier with high precision and was able to find spammers previously ignored by other classifiers.

In social bookmarking sites, detecting spammers is not quite as difficult. Markines et al. [6] used 6 features: tag spam, tag blur, document structure, number of ads, plagiarism, and valid links to develop a classifier with 98% accuracy.

Yu et al. [7] analyzed the growth and persistence of trends in Sina Weibo, a popular Chinese microblogging network, and observed that the effect of retweets in Sina Weibo is much higher than that in Twitter. Upon closer inspection, the authors observed that a large percentage of trends in Sina Weibo are due to the continuous retweets of a small number of spam accounts. These accounts are set up to artificially inflate certain posts, causing them to shoot up into the Sina Weibo trending list, which are in turn displayed to users.

D. Trends in Twitter

One of the most extensive investigations into trending topics in Twitter was by Asur et al. [8]. The authors examined the growth and persistence of trending topics in Twitter. Topics were observed to follow a log-normal distribution in popularity. Accordingly most topics faded from popularity relatively quickly, while a few topics lasted for long periods of time. An interesting result was that news topics from traditional media sources proved to be some of the most popular and long lasting trending topics in Twitter, suggesting that Twitter amplifies the general trends in society.

Cha et al. [9] explored user influences on Twitter trends and also found several interesting results. First, users with many followers are not necessarily effective in generating mentions or retweets. Second, the most influential users can influence the popularity of more than one topic. Third, influence does not arise spontaneously, but is instead the result of focused effort, often concentrating on one topic.

III. EXPERIMENT

A. Gathering Tweets

To obtain tweets for labeling, we constructed a program to interact with Twitter's public API. Once hourly, our program found the top 10 trending topics worldwide for the "en" language code and opened a connection filtered on those topics to receive a stream of data. For the next hour we gathered as many of the tweets and associated metadata as allowed by the Twitter API. The Twitter streaming API will cap the amount of tweets sent at 1% of the overall Twitter traffic, sending a random sampling of tweets and the amount of overflow when the limit is exceeded. This limit made more aggressive updating and collection methods infeasible for us.

Our program ran from February 1 to February 7, 2013, gathering over 9 million tweets across 801 distinct trending topics. Brief observation revealed that not all tweets were in English. We expected this, as filtering on the "en" language code merely restricts the character set of the tweets. Twitter API options for filtering the stream by geographic region were insufficient for our purposes, as such options only include tweets from users who have opted to report their geographic locations with their messages. It seemed likely to us that spammers would be under-represented if such options were used. Ultimately, our conclusion would not differ due to the inclusion of a small amount of non-English tweets.

B. Labeling Tweets

Once the data was gathered, our next task was to develop a collection of tweets labeled into spam and non-spam categories which could be used to train our classifier. For such a collection to be useful, it had to contain adequate numbers of spam and non-spam tweets; include tweets from a range of times and topics, and be as unbiased as possible given these constraints.

To construct such a collection for manual labeling, we developed a second program to randomly sample from the gathered tweets. We ensured that over 40 spam examples and examples from each of the 170 hour-long observation periods were included in this "gold standard" dataset. The result (see table I) contained nearly 1500 labeled spam and non-spam

Non-Spam Instances	Spam Instances
1453	42

TABLE I. LABELED COLLECTION OVERVIEW

examples across many distinct topics. As a concrete example of the type of data we analyzed, we give some non-spam messages in the labeled dataset:

- #30FactsAboutMe I like eggs
- And suddenly the MA thesis is fact not fiction. YAY!!! Scientists identify remains as those of King Richard III <http://t.co/kVGIV7to>
- RT @Jsprech3: I wanna go to ihop #nationalpancake-day

These stand in contrast to the examples of spam messages in the same dataset, which are clearly off topic and are commercial in intent:

- ILL SHOW YOU HOW TO EARN \$900+ DAILY FROM HOME! <http://t.co/zEKBe5Yd> #10ConfesionesDeMi
- Richard III - Justin Bieber caught NAKED in Miami with an Girl [PICS] <http://t.co/UqOh2T4e> #SuperBowl
- IHOP #Answer4Everything <http://t.co/OLS7RpWL> COOL VIDEO TELLS A METHOD TO EARN \$700+ DAILY!

C. Analysis Methods

There were two distinct phases to our analysis. The first was selecting attributes to use and validating the effectiveness of our classifier via information retrieval metrics. The second was evaluating the impact of spam filtering on the Twitter trending topics via statistical tests.

1) *Attribute Selection and Evaluation*: Previous work by Benvenuto et. al. [1] identified the following attributes as being useful for detecting spam in Twitter, with the exception of topic rank: 1) rank of topic; 2) URLs per word; 3) total number of words; 4) number of numeric characters; 5) number of total characters; 6) number of URLs; 7) number of hashtags; 8) number of mentions; 9) number of retweets; 10) whether the tweet was a reply. Although these attributes had been employed effectively by previous researchers, we wanted to determine which of the attributes were the most relevant to our task and dataset. Thus, we applied the chi-squared attribute selection method available in the Weka machine learning software to our training data and ranked the effectiveness of various attributes. The chi-squared attribute selection method determines which attributes vary the most significantly between classes by means of computing the chi-squared statistic with respect to each class.

The relevant attributes were extracted from the raw data and each tweet was represented as a vector of its attributes for use in our classifiers. To gain intuition into the effects of the attributes we compared the distribution of the attributes between the spam and non-spam classes. In particular, the mean values of the attributes was of interest.

2) *Classification*: To classify the majority of the dataset, we employed supervised machine learning algorithms. These algorithms are first trained on the labeled data to develop classification models that are then applied to unlabeled data to predict which tweets are in the spam class and which are in the non-spam class.

Previous research by Wang [2] suggested that a naive Bayesian classifier was best for spam classification. As the name suggests, naive Bayes classifiers apply the well-known Bayes theorem from probability:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

In our case, Y is the event that a given tweet belongs to a given class and X is the d dimensional feature vector corresponding to the tweet. The naive Bayes model makes the strong independence assumption that the attributes are all independent, allowing us to directly multiply the conditional probabilities for each attribute:

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^d (X_i|Y)}{P(X)}$$

The denominator $P(X)$ is the same for both spam and non-spam classes so we can discard it for the purposes of our classification. We computed the following:

$$P(\text{Spam}) \prod_{i=1}^d (X_i|\text{Spam}), \text{ and}$$

$$P(\text{NonSpam}) \prod_{i=1}^d (X_i|\text{NonSpam})$$

and classified the given tweet as the class with higher probability. Our training dataset was used to determine the conditional probabilities. For discrete attributes, such as whether the tweet was a reply, the frequency of the attribute among the class was used as the probability. For continuous attributes, such as the number of URLs per word, the conditional probability was found by comparing the tweet's value of the attribute to a normal distribution with mean and standard deviation matching the attribute for the class.

3) *Classifier Evaluation*: To evaluate the effectiveness of the classifier we employed standard information retrieval metrics: recall, precision, Micro-F1, and Macro-F1. Recall (r) for a given class is the number of instances correctly classified (true positives) divided by the total number of instances of that class (true positive and false negatives). Precision (p) for a given class is the number of instances correctly classified (true positives) divided by the total number of instances predicted to be in that class (true and false positives).

The $F1$ metric is the harmonic mean of recall and precision, that is, $F1 = 2pr/(p+r)$. Micro-F1 calculates the precision and recall values for all classes before computing F1. This measures the effectiveness of the classifier on a per-user basis. Macro-F1 computes F1 values for each class and then averages results over each class, measuring the effectiveness of the classifier on a per-class basis.

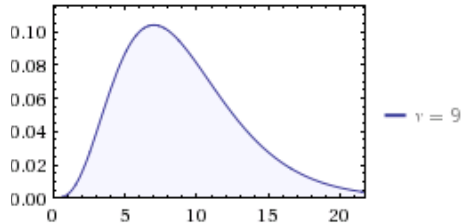


Fig. 1. A Chi-Squared Distribution With 9 Degrees of Freedom

Deriving these metrics for our classifier was accomplished by 10 fold cross validation. In 10 fold cross validation, the data in the training set is partitioned into 10 sets of equal size. Until each set has been tested, the classifier is first trained on all of the other sets and then tested on the remaining set. The results of these 10 tests are averaged to obtain the information retrieval metrics for the classifier.

4) *Spam Impact Evaluation*: To gain a thorough understanding of the impact of spam in the Twitter trending topics we evaluated our data from several different angles. To evaluate the overall impact we simply examined the proportion of tweets in our sample which were spam and compared it to previous research. To evaluate the impact by topic we analyzed whether the incidence of spam was spread equally across the trending topics by means of a chi-squared goodness of fit test.

The chi-squared goodness of fit test establishes whether or not an observed distribution of frequencies differs from an expected theoretical distribution. It is found by first computing the χ^2 statistic:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

where for a given entry i , O_i is the observed frequency, E_i is the expected theoretical frequency, and n is the number of entries. In our case the entries were the observed counts of tweets for each topic after the classifier was applied to filter out spam. The expected theoretical counts for each topic were the amount of tweets for the topic multiplied by the overall percentage of spam tweets for that hour. Once the test statistic is found, a chi-squared distribution with the appropriate number of degrees of freedom is consulted to determine the probability that the observed distribution would arise assuming the theoretical distribution is true. Figure 1 shows a chi-squared distribution with 9 degrees of freedom.

In addition to evaluating whether topics varied in their vulnerability to spam, we wanted to determine whether there was evidence of spammers manipulating topics enough to change the topic rankings. Thus, we re-ranked the topics after the spam filter had been applied and compared the results to the original rankings. For each of our hour periods under study we counted the number of ranking differences to find out by what percentage the rankings differed.

IV. RESULTS

Following the methods outlined above we were able to obtain useful results for both evaluating the effectiveness of our classifier and answering our questions regarding Twitter's trending topics.

Attribute	χ^2 Statistic
URLs per word	116
URLs	111
Number of hashtags	71
Numeric characters	17
Rank of topic	12
Whether tweet was a reply	3
Hashtags per word	0
Number of mentions	0
Number of retweets	0
Total number of words	0
Total number of characters	0

TABLE II. CHI-SQUARED RANKING FILTER

Attribute	Non-Spam Mean	Spam Mean
URLs per word	0.0077	0.0476
URLs	0.0847	0.5714
Number of hashtags	0.8671	1.0238
Numeric characters	1.3896	3.2177
Rank of topic	4.2638	6.1429

TABLE III. ATTRIBUTE DISTRIBUTION BY CLASS

A. Attribute Evaluation

The results of the chi-squared attribute test are shown in table II. As expected, some attributes were significantly more important in detecting spam than others. In light of our background research, it was not very surprising that the presence of URLs was a key attribute, since most spammers display messages with the hope of attracting users to follow a link. The fact that the number of words and the number of characters provided essentially no predictive power was also not surprising given the diversity of both spam and non-spam messages in Twitter.

Table III gives a summary of how the distribution of attributes differed between spam and non-spam tweets in our training dataset. It is interesting to observe how the distribution between the classes varied by attribute. As expected, spam messages had URLs with much higher frequency, and perhaps more numeric characters as a result of that in combination with monetary values. Spammers used hashtags more often than regular users, perhaps as a way to ensure their messages would be grouped with the trending topics. Interestingly, spam messages targeted topics with a lower mean ranking than non-spam messages. It is unclear why this is the case.

B. Classifier Evaluation

Table IV shows the confusion matrix obtained from running our naive Bayes classifier on the training dataset. Table V gives the information retrieval metrics for the classifier. The Micro-F1 measure was 0.929 and the Macro-F1 measure was 0.596. Over 90% of the instances are classified correctly.

These results are not outstanding but must be compared to a baseline for perspective. For our baseline classifier we classified all tweets as non-spam. This baseline classifier had

		Predicted	
		Spam	Non-Spam
True	Spam	1327	125
	Non-Spam	19	23

TABLE IV. CONFUSION MATRIX

		Metric		
		Precision	Recall	F1
Class	Non-Spam	0.986	0.914	0.949
	Spam	0.155	0.548	0.242

TABLE V. INFORMATION RETRIEVAL METRICS FOR NAIVE BAYES

a Micro-F1 measure of 0.958 and Macro-F1 measure of 0.479. We note that the Micro-F1 measure is 3% smaller for the naive Bayes model than this baseline. This is offset by the fact that the Macro-F1 measure is 24% larger for naive Bayes. Since our goal is to investigate the effects of spam, the increased spam recall of the naive Bayes model offsets its unimpressive precision.

C. Spam Overall

Previous research [2] has estimated that 3% of tweets in Twitter are spam. Our hand-labeled collection contained about 2.8% spam messages, which agrees with the assessment. When the classifier was applied to the rest of the dataset, we found an average of 9.0% spam. Because our classifier classified 9.9% of the training dataset as spam, this is a reasonable value and suggests that the likely proportion of spam in Twitter trending topics is not much different than 3%.

D. Spam Variance Among Topics

The results of our chi-squared goodness of fit test strongly suggested that topics are not affected in uniform proportions by spam. Every one of the 170 tests was significant at the 5% level. The average value of the chi-squared statistics was 7008, suggesting that the probability of spam being represented equally across topics was infinitesimally small. Though there is surely noise from our imperfect classifier, the probability that the observed distributions matched the theoretical distribution would be less than one in a million even if this average chi-squared statistic were off by a factor of 100.

E. Manipulation of Topic Rankings

Comparing the rankings of the topics before the spam filter was applied to the rankings of the topics after the spam filter was applied revealed some interesting results. In 47% (81 of the 170) time periods investigated there was no change in the rankings of the topics before and after the filter was applied. On average there were 1.66 topics that differed from their previous ranking after the filter was applied. Since any change in rankings after the filter was applied necessarily moves 2 topics out of position, this number suggests that rankings were not greatly affected by the presence of spam. Figure 2 gives

a histogram of the number of ranking differences across the time periods we studied. Reinforcing the average found above, large numbers of ranking changes occur infrequently.

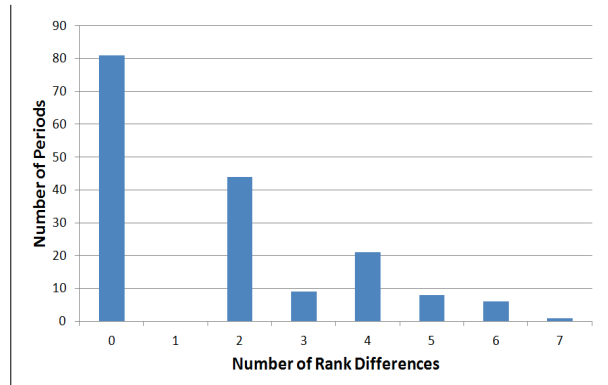


Fig. 2. Histogram of Ranking Differences

If spammers target some topics with greater frequency than others, how can the rankings of topics remain so consistent once the spam filter is applied? The answer lies in the relative popularities of the trending topics. The difference in the quantity of tweets for topics with different rankings is typically very large. Figure 3 shows the average number of tweets for topics with a particular rank, pre-filtered. The fitted exponential trend line has an R^2 value of 0.9694 suggesting a clear power law relationship with an exponent of -0.741 between topic rank and tweet volume. It is a mathematical fact that if one topic has over twice the number of tweets as the next highest in rank, 50% of it can be filtered out and it will still maintain its ranking. This observation makes it clear why spammers cannot easily manipulate the trending topics. They are fighting against the natural power law distribution of topic popularity. **Spammers don't drive topics in Twitter, but they do attempt to piggyback on their visibility.**

F. Spam in the Trending Topics

We shown in Section IV-D that spammers do not spread their messages uniformly across the trending topics. We conducted further experiments to verify a few hypotheses regarding how spammers might target trending topics.

First, we hypothesized that spammers might target topics that rank higher. For the purpose of our analysis, we define "Spam Incidence" as the percentage of tweets our classifier labeled as spam divided by the total number of tweets. Figure 4 shows the spam incidence for topics by rank, where the incidence was calculated by summing all tweets whose topic had a particular rank. Using a linear regression, it appears that there is a mild correlation between spam and the topics ranked lower in the top 10.

Next we hypothesized that spammers might target topics that stays in the trending topics list longer, since that the longer a topic stays, the more likely it will catch spammers' attention. Figure 5 illustrates the spam incidence for topics that appear in the trending topics list for a particular period of time. Using linear regression, it appears that there is no real correlation between spam and the longevity of a trending topic.

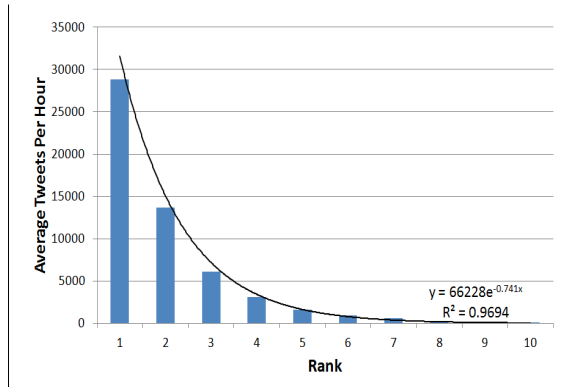


Fig. 3. Histogram of Average Hourly Tweets per Rank Position

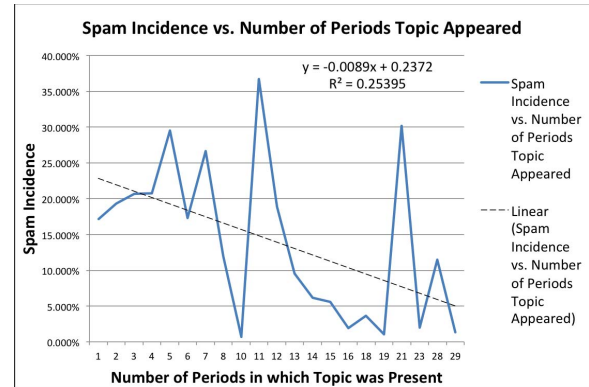


Fig. 5. Spam Incidence and the Longevity of Topics

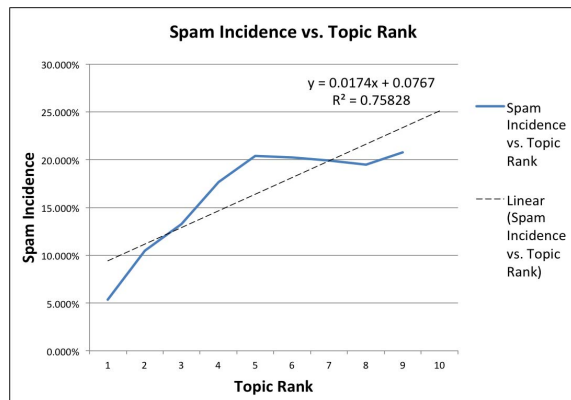


Fig. 4. Spam Incidence for Topics by Rank

V. CONCLUSION AND DISCUSSION

Social networks such as Twitter are becoming increasingly important for users, businesses, researchers and the public at large. In Twitter, the trending topics are highly visible to users. We investigated whether such important topics were the victims of manipulation by spammers. Various models for spam detection in social media have been developed by researchers, but none have been tailored towards investigating Twitter’s trending topics.

Over a 7 day period, we employed Twitter’s streaming API to gather over 9 million tweets on hourly trending topics. Processing the raw data from the API, we extracted tweet features relevant to spam detection as identified by previous research. We trained a naive Bayes classifier for tweet classification on a hand-labeled random sample of nearly 1500 tweets and verified its effectiveness via 10-fold cross validation. Filtering the trending topic data with this classifier, we obtained results on the prevalence of spam overall, between topics, and the effect of spam on topic ranks.

Spam frequency in the trending topics overall seemed to correspond with previous results suggesting a 3% spam rate in Twitter messages. Comparing the observed spam frequencies across topics by means of a chi-squared goodness of fit test, we found that topics varied greatly in the proportion of spam they contained. Re-ranking of topics after applying our spam filter changed existing rankings very little. This was a

result seemingly in contrast with our previous finding. The puzzle was solved by observing the power law distribution of topic popularity. We conclude that spammers do not drive the trending topics in Twitter, but instead opportunistically target topics with desirable qualities.

Our results bode well for the health of the Twitter social network and offer some direction for improvements. Social scientists can take confidence in that the trending topics in Twitter reflect the overall spirit of the social network and have the potential to aid predictive models. Furthermore, users of Twitter can take precautions regarding spam when certain topics are involved.

REFERENCES

- [1] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, “Detecting spammers on twitter,” in *Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS)*, vol. 6. National Academy Press, 2010.
- [2] A. Wang, “Detecting spam bots in online social networking sites: a machine learning approach,” *Data and Applications Security and Privacy XXIV*, pp. 335–342, 2010.
- [3] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, “The socialbot network: when bots socialize for fame and money,” in *Proceedings of the 27th Annual Computer Security Applications Conference*. ACM, 2011, pp. 93–102.
- [4] C. Chen, K. Wu, V. Srinivasan, and X. Zhang, “Battling the internet water army: Detection of hidden paid posters,” *arXiv preprint arXiv:1111.4297*, 2011.
- [5] K. Lee, J. Caverlee, and S. Webb, “Uncovering social spammers: social honeypots+ machine learning,” in *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2010, pp. 435–442.
- [6] B. Markines, C. Cattuto, and F. Menczer, “Social spam detection,” in *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web*. ACM, 2009, pp. 41–48.
- [7] L. Yu, S. Asur, and B. Huberman, “Artificial inflation: The real story of trends and trend-setters in sina weibo,” in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)*. IEEE, 2012, pp. 514–519.
- [8] S. Asur, B. A. Huberman, G. Szabo, and C. Wang, “Trends in social media: Persistence and decay,” in *5th International AAAI Conference on Weblogs and Social Media*. AAAI, 2011, pp. 434–437.
- [9] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, “Measuring user influence in twitter: The million follower fallacy,” in *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*. AAAI, 2010.