Understanding and Combating Link Farming in the Twitter Social Network

Saptarshi Ghosh
IIT Kharagpur, India

Bimal Viswanath
MPI-SWS, Germany

Farshad Kooti
MPI-SWS, Germany

Naveen K. Sharma
IIT Kharagpur, India

Gautam Korlam
IIT Kharagpur, India

Fabricio Benevenuto
UFOP, Brazil

Niloy Ganguly
IIT Kharagpur, India

Krishna P. Gummadi
MPI-SWS, Germany

ABSTRACT

Recently, Twitter has emerged as a popular platform for discovering real-time information on the Web, such as news stories and people's reaction to them. Like the Web, Twitter has become a target for *link farming*, where users, especially spammers, try to acquire large numbers of follower links in the social network. Acquiring followers not only increases the size of a user's direct audience, but also contributes to the perceived influence of the user, which in turn impacts the ranking of the user's tweets by search engines.

In this paper, we first investigate link farming in the Twitter network and then explore mechanisms to discourage the activity. To this end, we conducted a detailed analysis of links acquired by over 40,000 spammer accounts suspended by Twitter. We find that link farming is wide spread and that a majority of spammers' links are farmed from a small fraction of Twitter users, the `social capitalists`, who are themselves seeking to amass social capital and links by following back anyone who follows them. Our findings shed light on the social dynamics that are at the root of the link farming problem in Twitter network and they have important implications for future designs of link spam defenses. In particular, we show that a simple user ranking scheme that penalizes users for connecting to spammers can effectively address the problem by disincentivizing users from linking with other users simply to gain influence.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services; C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*

Keywords

Twitter, spam, link farming, Pagerank, Collusionrank

1. INTRODUCTION

Recently, the Twitter social network has emerged as a popular platform for discovering real-time information on the Web, such as current events, news stories, and people's opinion about them. Traditional media, celebrities, and marketers are increasingly using Twitter to directly reach audiences in the millions. Furthermore, millions of individual users are sharing the information they discover over Twitter, making it an important source of breaking news during emergencies like revolutions and disasters [17, 23]. Recent estimates suggest that 200 million active Twitter users post 150 million tweets (messages) containing more than 23 million URLs (links to web pages) daily [3, 28].

As the information shared over Twitter grows rapidly, search is increasingly being used to find interesting trending topics and recent news [26]. As Twitter becomes more accessible via search, it has also started to attract the attention of spammers, who strive to get their tweets appear near the top of the search results. Search engines rank tweets based not only on the content of the tweet, but also on how influential the user who posted the tweet is [25]. The exact details of the influence metric depend on the search engine, but all of them critically depend on the user's connectivity in the social graph. The more followers a user has, the more likely his tweets are to be ranked highly. So spammers attempt to enhance their influence score (and thus the ranking of their tweets) by acquiring links in the social network. Similar to the Web, where some websites exchange reciprocal links with other sites to improve their ranking by search engines, spammers try to infiltrate the Twitter network by building social relationships – they follow other users and try to get others to follow them. We refer to this process of reciprocal exchange of links between unrelated users to gain influence in the network as *link farming*.

While link farming in the Web graph has been well studied and understood [8, 14, 30], there is no existing work on link farming in the Twitter social network to the best of our knowledge. Furthermore, unlike the Web, where a link from web page A to web-page B implies that B's content is relevant to A's content, the meaning of (or the reason for establishing) a social link between two users is unknown to anyone but the users themselves. This makes it considerably harder to detect and analyze link farming activities in Twitter.

In this paper, we investigate the vulnerability of the Twitter social network to link farming. Specifically, we focus

WWW 2012, April 16–20, 2012, Lyon, France
ACM 978-1-4503-1229-5/12/04.

on better understanding the users who establish links to spammers and the potential reasons for their behavior. To this end, we gathered data of 41,352 spammer accounts suspended by Twitter and conduct a detailed analysis of the users who connect to them.[1] We also used extensive data from a previous measurement study that included a complete snapshot of the Twitter network and the complete history of tweets posted by all users as of August 2009 [9].

Our analysis reveals surprising social dynamics that drive link farming in Twitter. When we started the study, we expected that spammers in Twitter would be farming links in two ways: first, by targeting (following) lay Twitter users with very few followers, who then reciprocate out of social etiquette [12], and second, from other spammers and fake accounts (Sybils) that have been explicitly created for the purpose of farming links. In sharp contrast to our expectation, we found that a majority of farmed links come from a small number (100,000 or so) of legitimate, popular, and highly active Twitter users. This is very different from the Web, where popular pages would rarely point to spam pages.

We conjecture that the Twitter users engaging in link farming are `social capitalists`, whose goal is to amass social capital and promote their legitimate content on Twitter. Examples of social capitalists range from popular bloggers of social media and Internet technologies to celebrities like Britney Spears and from politicians like Barack Obama to businesses like JetBlue airways. We show that social capitalists tend to reciprocate (i.e., follow back) to anyone who connects to them, to increase their social capital. Unfortunately, spammers exploit this behavior of capitalists to farm links in the Twitter network and promote spam content.

Finally, we explore mechanisms to deter link farming in the future. We argue that any effective solution to fight link spam should take into consideration the current incentive structure of the Twitter network that encourages social capitalists to collude with other social capitalists, including users who they might not know. Inspired by ideas from spam-defense strategies proposed for the Web graph, we propose a ranking system, called *Collusionrank* that penalizes users (by lowering their influence scores) for connecting to spammers. Collusionrank disincentivizes users from colluding with people that are unknown to them, who might potentially be spammers. We show that, even when only a small fraction (1%) of all spammers are identified, Collusionrank successfully lowers the influence scores of the other spammers in the network.

2. RELATED WORK

In this section, we briefly discuss related literature on link-farming in the Web and on spam in Twitter.

2.1 Link-farming in the Web

Link-farming has been widely studied in the context of the Web. Bharat *et al.* [7] were possibly the first to show that iterative ranking algorithm such as HITS can be dominated by 'mutually reinforcing relationships' between webpages of two hosts. Lempel *et al.* [21] showed that pages within a tightly-knit community (TKC) get high scores in the HITS algorithm (known as the 'TKC effect'). Link farms attempt to exploit this effect to get high rankings. Gyongyi *et al.* studied the structure of link farms and how the pages in the farms can be interconnected to optimize rankings [14].

Several solutions to counter link farming have also been proposed; these solutions fall under two main categories – techniques which leverage only the properties of the link structure, and those which use the *content* of the web-pages along with the link structure.

Link-based statistics were used by Becchetti *et al.* [5] to build a classifier for automatic detection of Web-spam. Gyongyi *et al.* proposed the Trustrank algorithm [15] to combat web spam, where the basic assumption is that good pages usually link to other good pages; hence they start by assigning high scores to a set of known good pages, and then propagate the good ranks in a way similar to Pagerank. Some algorithms that are *inversions* of Trustrank have also been proposed to identify bad (spam) pages. An algorithm known as BadRank is believed to be used by a commercial search engine to identify and blacklist pages which link to spam pages [24]. Wu *et al.* [30] proposed an algorithm that initially identifies a set of bad pages based on the common link set between incoming and outgoing links of webpages, and then expands this set by marking a page as bad if it links to more than a certain number of other bad pages. Wu *et al.* also proposed methods of combining the trust and distrust scores of pages to demote spam pages in the Web [31]. In Section 5, we show that a similar approach is effective in the Twitter social network as well, in demoting not only spammers but also the link-farmers who frequently follow spammers.

Examples of methods to counter link-farms, that utilize both the link structure and webpage content, include [10] which uses the Document Object Model along with the hyperlinks to beat nepotistic 'clique attacks'. Davison identified nepotistic links on the web by using a rule-based classifier on features based on the content and link structure of the pages [11]. Castillo *et al.* [8] also combined link-based and content-based features to build a spam classifier using the idea that hosts that are well-linked together are likely to have the same class label (spam or non-spam).

Unlike the Web, the semantics of link structure in social networks is very different – links are not between web pages, but between users. This suggests that the factors that can lead to link farming in a social network could be different from that in the Web, and thus requires a fresh look.

2.2 Spam in Twitter

There have been a number of recent studies of Twitter spam, most of which proposed machine learning algorithms for detecting spammers. Lee *et al.* [19] created social honeypots to identify spammers on MySpace and Twitter and proposed classification algorithms to distinguish between spammers and legitimate users. Benevenuto *et al.* [6] approached the problem of detecting trending-topic spammers – users who include unrelated URLs with trending words in tweets, in order to make the tweets appear in the results of searches or meme-tracking tools. They manually labeled a collection of users as spammers or non-spammers and identified properties that are able to distinguish between the two classes of users through a machine learning approach.

There have also been studies on the tools for automatic dissemination of spam in Twitter. Zhang *et al.* [33] proposed an approach to identify automatic tweeting in Twitter and showed that 16% of the active Twitter accounts they evalu-

[1]We make anonymized data about spammer-nodes available to the research community at: twitter.mpi-sws.org/spam

ated exhibit a high degree of automation. Grier *et al.* presented an extensive study of tweets containing blacklisted URLs which were obfuscated using *bit.ly* and found that spam URLs in Twitter get much higher click-through than URLs in spam emails [13]. They also identify spam campaigns by grouping two spam-accounts in the same campaign if they tweet the same spam URL. Recently, Thomas *et al.* [27] studied the tools used by spam-accounts that were identified from among the accounts that were suspended by Twitter. We use a similar strategy of identifying spam-accounts from among the suspended accounts.

Overall, the existing studies of Twitter spam attempt to understand the tools used by spammers, or to design techniques for identifying spammers. In comparison, research on Web spam has reached the 'next' stage beyond basic spam detection, such as understanding how spammers are actually establishing links in the Web in order to deceive search-engines [5] and unearthing the support structures (such as specific ISPs) for Web spammers [22]. Our study takes the first important step to reach a similar 'next' level in the study of Twitter spam, by unveiling the 'support structures' for spammers in Twitter. More specifically, this work shows the existence of a set of users in Twitter – the social capitalists – who are *not* spammers themselves, but who engage in link farming, and thereby unwittingly help spammers in farming links. To the best of our knowledge, no prior work has investigated link farming either in Twitter or in any other online social network.

3. LINK FARMING IN TWITTER

Our strategy to study link farming in Twitter relies on analyzing how spammers acquire links in the Twitter network. Given their desire to promote unwanted (spam) messages, it is intuitive that spammers rely on link farming strategies to acquire follower links. Also, a number of recent studies report an increasing number of spammers infiltrating the Twitter network [20].

3.1 Methodology

To study how spammers farm links in the Twitter network, we need (i) a large and representative sample of spammers and (ii) the connectivity and activity of both spammers and the users who connect to spammers. To the best of our knowledge, no such collection is publicly available. We describe below how we gathered these data.

The Twitter dataset used in this work contains extensive data from a previous measurement study that included a complete snapshot of the Twitter network and the complete history of tweets posted by all users as of August 2009 [9]. More specifically, the dataset contains **54,981,152** user accounts connected to each other by **1,963,263,821** social links. The dataset also contains all tweets ever posted by the collected users, which consists of **1,755,925,520** tweets. About 8% of all users had set their accounts to private, which implies that only their followers could view their tweets. We ignore these users in our analysis. For a detailed description of this dataset we refer the user to reference [9].

3.1.1 Identifying spammers in Twitter

To identify a large set of spammers in the dataset, we rely on Twitter's official policy of *suspending* accounts that it deems to have participated in malicious activity [4]. We can verify whether a user account has been suspended by simply attempting to crawl the user's profile page – if the user is suspended, the crawl would lead to the webpage `http://twitter.com/suspended`. We re-crawled the profile page of each user in the above dataset in February 2011, and found that as many as 379,340 accounts had been suspended in the interval from August 2009 to February 2011.

Although the primary cause for suspension of Twitter accounts is spam-activity, Twitter's policy page states that accounts that are inactive for more than 6 months may also be suspended [4]. Hence, we can confirm that a suspended user is a spammer only if we can explicitly detect some malicious activity. We examined the URLs posted by an account using two of the most popular URL shortening services - *bit.ly* and *tinyurl*, looking for the presence of blacklisted URLs. Fetching a blacklisted shortened URL on these services leads to an interstitial warning page [1]. We fetched all the *bit.ly* or *tinyurl* URLs that were posted by each of the 379,340 suspended accounts and found that **41,352** suspended accounts had posted at least one shortened URL blacklisted by either of these two shortening services. We consider these 41,352 user-accounts as spammers.

It should be noted that our goal here is *not* to exhaustively identify *all* spammers in the Twitter social network, but rather select a large and accurate (i.e., with very few false positives) sample of spammers that would provide us with sufficient data to study how spammers farm social links in Twitter. We believe our methodology achieves our goals.

Also, a study by Grier *et. al.* [13] suggested that part of the spam activity in Twitter might originate from user-accounts that have been *compromised* (e.g., by phishing attacks or password-guessing), but the same work also said that most of such compromises have a short timespan, after which the original owner reclaims the account by contacting Twitter. We verified that the accounts we identified as spammers did in fact remain suspended over a period of several months, thereby reducing the chance that our spam-accounts are compromised accounts and affirming that they are dedicated spam-accounts (i.e., the accounts created explicitly by spammers). Further, in a more recent study [27], Grier *et. al.* themselves verified that a very large majority of the suspended accounts in Twitter are in fact dedicated spam-accounts.

3.2 Terminology

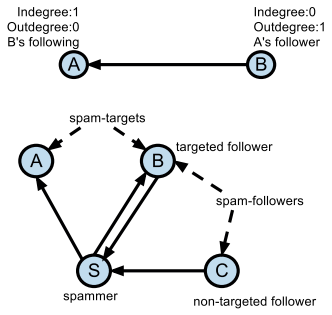


Figure 1: Terminology for the spammer's social neighborhood

Figure 1 introduces the terminology we use in the rest of the paper. If node (user) B follows node A, we refer to B

as A's follower and A as B's following. We represent the relationship in the network graph with a directed edge from B to A, which increases A's indegree and B's outdegree by 1. We refer to the Twitter nodes (users) followed by a spammer as *spam-targets* and the nodes that follow a spammer as *spam-followers*. Spam-targets who also follow the spammer are called *targeted followers*, while nodes which follow a spammer without being targeted by the spammer are called *non-targeted followers*.

3.3 Spammers farm links at a large-scale

We begin by analyzing the nodes following and followed by the 41,352 spammers identified in the previous section. We investigate the extent to which their follower links are farmed, by looking for evidence of link reciprocation in acquiring followers. We compute how many of the spam-followers connect to spammers with and without being targeted by spammers and the number of follower links they contribute.

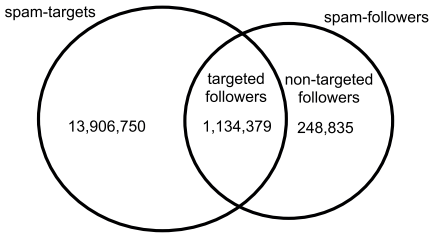


Figure 2: Number of spam-targets, spam-followers and their overlap. 82% of spam-followers overlap with the spam-targets.

Figure 2 shows a Venn diagram of the total number of unique nodes targeted by and following all spammers in our data set. We make two observations here. First, over 15 million Twitter nodes, which account for over 27% of the entire Twitter network (as of August 2009), have been targeted by at least one of the 41,352 spammers we identified. This statistics highlights the scale at which spammers try to farm links in Twitter: even a small number of spammers (less than 0.08% of all users) target and affect a large fraction of the Twitter network.

Second, looking at the breakdown of spam-followers into targeted and non-targeted followers, we find that a majority 82% (or 1,134,379 out of 1,134,379 + 248,835) of spam-followers have also been targeted by spammers. Thus, link reciprocation appears to play an important role in spam-followers' decision to connect to the spammers. This hypothesis is further strengthened by the fact that the 248,835 non-targeted followers account for only a minority 9% of all links to spammers, whereas the 1,134,379 targeted followers contribute 7,739,591 (91%) of all links acquired by spammers.[2] Thus, spammers get most of their followers and follow-links from among the nodes they target, suggesting that spammers rely on link farming to acquire most of their followers. Consequently, in the rest of the paper, we focus our attention on only the targeted spam-followers.

[2]On further investigation of the top 10,000 non-targeted spam-followers (based on number of links to spammers), we found that 9,725 of them were suspended by Twitter. This suggests that this set of users are mostly Sybil accounts or hired helpers of spammers.

		Median	Mean	90thperc.
# Followers	spammer	35	234	197
	random	7	36	36

Table 1: Follower-count statistics for spammers and a random sample of Twitter users. Spammers have an order of magnitude larger number of followers.

3.4 Link farming makes spammers influential

We now investigate whether spammers succeed in gaining influence in the network through link farming. Recollect that several social media analytics companies and search engines estimate the importance of tweets based on their estimate of the importance or influence of the user posting the tweet. So the primary motivation for spammers to farm links is to accumulate social capital and thereby, increase the chance that their spam tweets, tagged with hashtags related to the popular trending topics of the day [6,32], would show up higher in search results.

To estimate user influence, search engines use metrics based on the network structure. For example, a simple metric is the follower count of the user, while a more complex metric (used by Google) is the Pagerank of the user computed over the social network graph [25]. Table 1 shows the median, average and 90^{th} percentile statistics for spammers' follower count. As a baseline for comparison, we also show similar statistics for a randomly selected sample of 300,000 Twitter users. We find that spammers have an order of magnitude higher number of followers compared to random Twitter users.

Next, we compute the PageRank for our 41,352 spammers in the Twitter social network, and show them in Figure 3. We observe that some of the spammers are able to acquire very high Pageranks – 7 spammers rank within the top 10,000 (0.018% of all users) while 304 and 2,131 spammers rank within the top 100,000 (0.18% of all users) and 1 million (1.8% of all users) users according to Pagerank, respectively. Thus, our data shows that some spammers do succeed in acquiring high influence ranks through link farming.

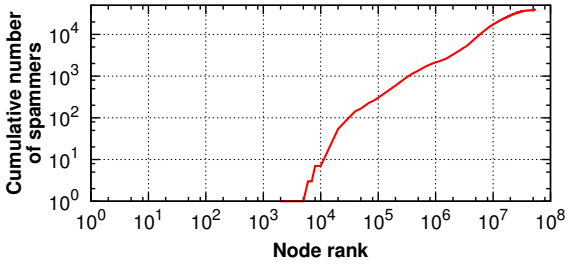


Figure 3: Number of spammers (among the 41,352 identified ones) who rank within the top K according to Pagerank

3.5 Most farmed links come from few users

We now probe further into the susceptibility of Twitter users to link farming. Specifically, we investigate whether some users are more likely than others to reciprocate links to spammers, and whether spammers exploit this susceptibility to farm most of their links from such users.

We show that a fraction of the targeted spam-followers are likely to reciprocate most links from spammers, i.e., they regularly follow spammers back. For this, we rank the targeted spam-followers based on number of links they create to the 41,352 spammers. Figure 4 shows the fraction of links from spammers to a spam-follower that are reciprocated by the spam-follower (or fraction of reciprocated in-links of the spam-follower), as a function of the spam-follower rank. We observe that the top spam-followers (based on the ranking stated above) exhibit a very high reciprocation to links from spammers. In fact, the top 100,000 spam-followers exhibit a reciprocation of 0.8 on average. As a result, a large majority of the links acquired by spammers come from these users – we found that the top 100,000 spam-followers account for 60% of all links acquired by spammers. Thus, spammers acquire a majority of their links from a small number (about 100,000) of Twitter users, who tend to reciprocate to any spammer who links to them.

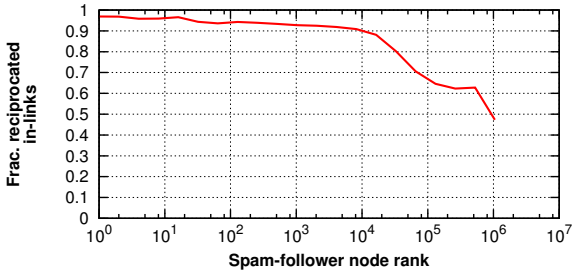


Figure 4: Fraction of reciprocated in-links from spammers vs spam-follower node rank (based on number of links to spammers). Node ranks on the x-axis is based on log-binning. Top spam-followers tend to reciprocate all links established to them by spammers.

3.6 Verifying link farming activity

To further confirm our findings about link farming activity, we conducted a small-scale real-world experiment on the Twitter network. We created a Twitter account with a common user name without any profile details, and posted two tweets expressing joy over discovering Twitter. We then used this account to follow a set of 500 users randomly sampled from among the top 100,000 spam-followers identified above as potential collaborators for link farming. We established links to all 500 users on the same day. Within a span of 3 days, 65 (i.e., 13% out of 500) of the users responded by following our account back. As a result of the followers acquired, our 3-day old account ranked among the top-9% of all Twitter users according to the PageRank influence metric. Our experiment demonstrates the ease with which links and thereby, influence can be acquired in the current Twitter social network.

3.7 Summary

In this section, we studied link farming in the Twitter network by analyzing how spammers acquire followers. We found that spammers attempt to acquire followers by establishing links to a large fraction of Twitter users. We also found that spammers succeed not only in acquiring considerably more followers than random Twitter users, but that some of the spammers also rank amongst the most influential Twitter users. Interestingly, we found that a majority of farmed links come from a small fraction of Twitter users, who tend to reciprocate to anyone who connects to them. We verified our findings through a simple real-world experiment. In the next section, we focus on characterizing these users, who are most susceptible to link farming.

4. ANALYSIS OF LINK FARMERS

Our goal in this section is to get a better insight into what drives link farming in Twitter. For this, we analyze the characteristics (network connectivity and tweeting activity) of the users who are willing to reciprocate links from arbitrary users, and their potential reasons for engaging in link farming.

4.1 Popular users more likely to farm links

Based on conventional wisdom, one might expect that lay users with few followers would be more likely to reciprocate links from spammers than popular users with lots of followers. After all, lay users would be eager to gain more followers, while popular users with lots of followers might be concerned about the damage that following spammers may cause to their reputation. Figure 5 shows how the probability of a user reciprocating to a link from spammers varies with the user's indegree (number of followers). The plot shows something very unexpected – users with low indegree, which constitute the bulk of Twitter's user population, rarely respond back to spammers. Rather counter-intuitively, responsiveness generally increases with indegree, with the exception of very high indegree values, where there are too few nodes and the probability of response starts dropping. Thus, barring a few extremely popular celebrities and media sites, users' tendency to farm links increases with their popularity in the network.

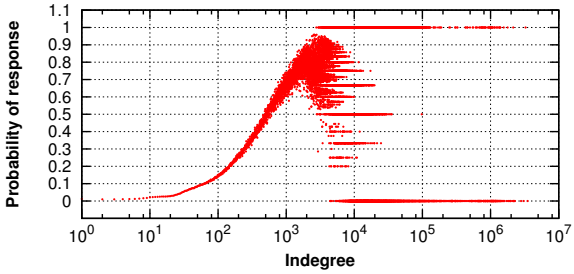
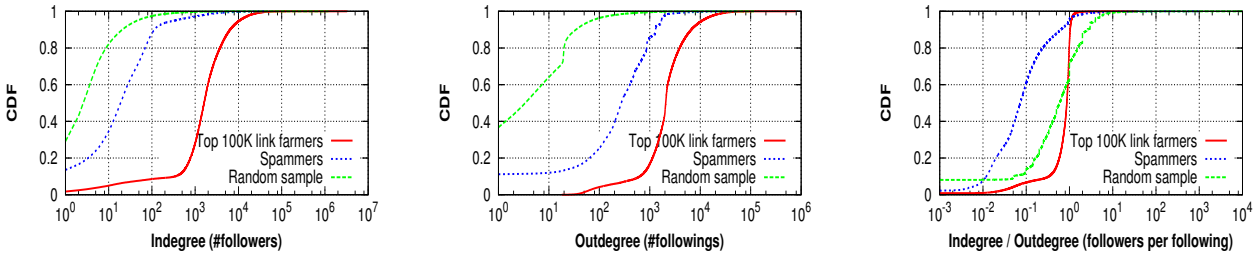


Figure 5: Probability of response vs indegree for all users targeted by spammers. Users with low indegree (few followers) do not reciprocate to links from spammers. Responsiveness increases with number of followers.

To further confirm our surprising finding, we examined the follower-counts of the top 100,000 spam-followers (the set of users who account for a majority of the links farmed by the 41,352 spammers). We refer to these users as the `top link farmers`. From Fig. 5, we would expect most of them to be popular users with lots of followers. Figure 6(a) shows the cumulative distributions of indegrees of top link farmers as well as a randomly selected sample of 300,000 Twitter users. We find that top link farmers have one order to two orders of magnitude higher follower counts than random Twitter users. In fact, 71% of the top link farmers have more than 1000 followers, while less than 0.3% of the random sample have more than 1000 followers. Thus the top link farmers



(a) Indegree (b) Outdegree (c) Indegree/outdegree ratio

Figure 6: Node degree distributions of top 100K link farmers, spammers and a random sample of Twitter users. Top link farmers have very high indegree and outdegree compared to both spammers and a random population. Also, most of the top link farmers have indegree/outdegree ratios near 1.

are popular nodes with lots of followers. In the rest of this section, we investigate their characteristics and motivations in greater detail.

4.2 Top link farmers are not spammers

We begin by checking if the top link farmers are spammers or accounts controlled by the spammers themselves. We analyzed the current status of the top 100,000 link farmers (as in July 2011). Out of the 100,000, 18,826 have been suspended (hence, possible spammers) and 4,768 were reported as 'Not Found' (i.e., the users deleted their accounts). However, more than 76% of the accounts still exist, and have not been suspended by Twitter in the last two years, suggesting that a large majority of these accounts are most likely to be legitimate users and *not* spammers. Furthermore, 235 of the top link farmers have actually been 'verified' by Twitter as real, well-known users.

To further confirm that top link farmers are indeed legitimate users, we manually verified 100 randomly selected accounts from among the top 100,000 link farmers who still exist in Twitter. To minimize the impact of human error, three volunteers independently analyzed each account. Volunteers were instructed to determine whether the accounts are from real users or they look suspicious. In total, 86 accounts were considered to be real by all three volunteers. Analyzing the tweeting behavior of the 86 accounts, all three volunteers agreed that a majority of them belong to one of two categories: (i) users tweeting on topics like Internet marketing, entrepreneurship, money, and social media (ii) business firms whose tweets attempt to promote their websites.

Finally, we also compared the network connectivity of top link farmers with those of spammers. Figs. 6(a), (b), and (c) respectively show the cumulative distributions of indegree, outdegree, and ratio of indegree to outdegree for the top link farmers and the 41,352 spammers. Not only do top link farmers have one order to two orders of magnitude higher indegree and outdegrees than spammers, but also their indegree-to-outdegree ratios are considerably higher than those of spammers (and close to 1). The fact that top link farmers exhibit very different network connectivity than spammers further suggests that a majority of top link farmers are not spammers.

4.3 Top link farmers are active contributors

To gain a better understanding of who the top link farmers really are, we analyzed information on their profile pages and their tweeting activity. To this end, we crawled the profile pages of the top 100,000 link farmers in July 2011, collecting more detailed information including the users' bios (a short description of a user posted by the user herself), profile pictures, and so on.

(a) Top link farmers (b) Random users

Figure 7: Word-cloud of words in the Twitter account bio of top 100,000 link farmers and a random sample

We compared a number of profile characteristics of top link farmers with those of random Twitter users in an attempt to define their distinguishing features. Table 2 summarizes the characteristics we investigated. In comparison with a random sample of 300,000 users, top link farmers put in considerable effort to improve their accounts and define their Twitter identities, by changing profile themes, providing pictures and user location, editing their public biography, and also exhibit a link to an external website in their profile page. For example, 87% and 79% of top link farmers provide bio and URLs (links) to their external webpages, while only 25% and 14% of the Twitter population provide this information. Similarly, 23% of the top link farmers have created at least one *List*, a recently introduced Twitter feature that allow users to organize the people they follow. In comparison, only 4% of the random sample used this feature. Our analysis suggests that, as compared to random Twitter users, the top link farmers are active users that make more heavy use of their profile information and explore more of the features provided by Twitter.

In order to gain more insight into the topical expertise of top link farmers, we generated a *word-cloud* of the most frequent words in the bio of these accounts. Figure 7 compares these word-clouds for the top link farmers and a set of 300,000 random users. Link farmers usually describe themselves using words like marketing, business, Internet, entrepreneur, and media, suggesting that these users are promoting their own businesses or content, or are talking about trends in a particular domain. Also, a manual analysis of

	Has Bio	Has URL	Profile Pic	Changed profile theme	Has Location	Has Lists
Top link farmers	87%	79%	96%	84%	84%	23%
Random sample	25%	14%	50%	40%	36%	4%

Table 2: Characteristics from profile and activity of the Top 100,000 link farmers

Top 5 link farmers according to	
#links to spammers	**Pagerank**
Larry Wentz: Internet, Affiliate Marketing	Barack Obama: Obama 2012 campaign staff
Judy Rey Wasserman: Artist, founder	Britney Spears: It's Britney
Chris Latko: Interested in tech. Will follow back	NPR Politics: Political coverage and conversation
Paul Merriwether: helping others, let's talk soon	UK Prime Minister: PM's office
Aaron Lee: Social Media Manager	JetBlue Airways: Follow us and let us help

Table 3: Names and extracts from Twitter account bios of 10 link farmers – the ones having most links to spammers and the highest ranked according to Pagerank.

100 randomly selected top link farmers (as described in Section 4.2) showed that a majority of their tweets contains links to legitimate external web pages. This is in contrast to the general Twitter population (the random sample), who describe themselves using words such as love, life, live, music, student, and friend, and most of whom never tweet links to external web pages.

Table 3 shows the names and bios of some selected top link farmers. They include celebrities like Britney Spears, politicians like Barack Obama, news media like NPR Politics, businesses like JetBlue airways as well as individual artists, technology enthusiasts and marketers. Thus, top link farmers also happen to be popular and highly active contributors to information on the Twitter network. Even though they are trying to promote their tweets, they are not spammers themselves, and the content they are promoting is legitimate.

4.4 Top link farmers are `social capitalists`

Having studied the characteristics of top link farmers, we now explore potential reasons for why they participate in link farming. Specifically, we ask the following question: *what motivates legitimate, popular, and actively contributing Twitter users to indiscriminately follow back anyone who connects to them?*

One simple and intuitive explanation is that these users have similar incentives as spammers. Like spammers, they seek to amass social capital and influence in the network, and leverage it to promote their legitimate tweets. So it is only natural that they would interconnect with others having a similar desire to amass social capital, including with the spammers. In fact, the bio of a number of these accounts contain phrases such as "will follow back" or "follow me, I'll follow you" (see Table 3). Furthermore, connecting back to one's followers might be polite social etiquette and increase the chance of retaining the followers in the long term [18]. Since desire for social capital drives their link farming behavior, we call such users `social capitalists`.

We studied the set of social capitalists constituted by the top 100,000 link farmers. We found that social capitalists connect to a vast majority (over 80%) of their network neighbors via reciprocated links. We also found evidence that social capitalists heavily interconnect with one another to enhance their influence. The Twitter sub-graph formed by the 100,000 social capitalists is densely connected with 180,741,277 links, which implies a high network density of 0.018 (in comparison, the entire Twitter network has a density of 6.5×10^{-7}). Thus, we find that social capitalists heavily interconnect with each other to increase their mutual influence.

Finally, we analyzed the influence of the social capitalists in the network. We calculated capitalists' influence according to several well known metrics based on the network structure and the user's activity. Specifically, we computed the following three widely used metrics: (a) *Follower-rank*: this metric simply ranks users based on their number of followers and it is used by Twitter itself to rank users [29], (b) *Page-rank*: several search-engines, including Google, estimate the importance of tweets based on the Pagerank of the user posting the tweet, in order to return tweets as search results [25], (c) *Retweeted-rank*: this metric measures the number of times the tweets posted by a user are retweeted. It indicates the ability of a user to generate content with pass-along value [9]. We found that *a significant majority of social capitalists appear within the top 5% of most influential Twitter users, independent of the ranking scheme used, confirming that social capitalists yield considerable influence in the network*. It is ironic that the capitalists who wield the most influence in the Twitter network are most susceptible to link farming. It is also worrisome because by acquiring influential social capitalists as followers, spammers stand to gain the most influence (as seen in Section 3.4).

4.5 Summary

In this section, we analyzed the characteristics of the link farmers in order to get a better insight into the factors driving link farming in Twitter. Rather surprisingly, we find that legitimate, popular, and highly active users in Twitter, such as bloggers and domain experts, are the users most likely to engage in link farming. We conjectured that the motivating factor for such users might be the desire to acquire social capital and thereby, influence. We showed evidence that these social capitalists connect with others with a similar desire to amass social capital, including each other and spammers.

5. COMBATING LINK FARMING

In its early days, Twitter provided individual users a platform to publish their messages and enabled other interested users to sign up for the updates. In those days, a follow-link from user A to user B implicitly suggested that user A likes to read user B's tweets. As Twitter became more valuable and important as a platform for sharing real-time information, Twitter saw the rise of spammers and social capitalists. As seen in the earlier Sections, the social capitalists are legitimate users who try to acquire social capital (i.e., links) so that they can promote their content. Thus, social cap-

Algorithm 1 Collusionrank

Input: network, G; set of known spammers, S; decay factor for biased Pagerank, α
Output: Collusionrank scores, c

initialize score vector d for all nodes n in G

$$d(n) \leftarrow \begin{cases} \frac{-1}{|S|} & \text{if } n \in S \\ 0 & \text{otherwise} \end{cases}$$

/* compute Collusionrank scores */
$c \leftarrow d$
while c not converged **do**
  for all nodes n in G **do**

$$tmp \leftarrow \sum_{nbr \in followings(n)} \frac{c(nbr)}{|followers(nbr)|}$$

$$c(n) \leftarrow \alpha \times tmp + (1-\alpha) \times d(n)$$

  end for
end while
return c

italists represent an inversion in incentives for establishing links. One unfortunate side-effect of these incentives is that spammers, who also aim at acquiring social links, can exploit them to acquire links from social capitalists. We next propose an approach to combat link farming that discourages social capitalists from colluding with spammers.

5.1 Collusionrank

Our approach borrows ideas from spam-defense strategies proposed for the Web graph (discussed in Section 2). The key idea is to penalize web-pages that link to spam pages, assuming that a page that links to bad pages must itself be a bad one [24,30,31]. Such a strategy can deter link farming in Twitter, as it would penalize those users who follow a large number of spammers. By lowering the influence scores of users connecting to spammers, our approach incentivizes users to be more careful about who they connect with.

We assume a system model where Twitter operators periodically identify and suspend spammers, just as they do today. We use the set of identified spammers to penalize users who connect to them. As the same Twitter users tend to follow all spammers, both identified and unidentified, this has the effect of recursively lowering the scores of other yet unidentified spammers (as we show in Section 5.2).

We propose *Collusionrank*, a Pagerank-like approach, to combat link farming in Twitter. In the original Pagerank algorithm, each page starts with the same initial score, which is then recursively modified based on the scores of the pages which link to this page. Variants of the Pagerank algorithm, such as topic-sensitive Pagerank [16] and Trustrank [15], have demonstrated the benefits of biasing these initial scores towards a certain subset of nodes deemed as more relevant or trustworthy. Here, we use a similar strategy to identify 'bad' nodes in the social network, but with the following two modifications. First, we *negatively* bias the initial scores towards a set of bad nodes, i.e., nodes identified as spammers. Second, since a user should be penalized for following spammers and *not* for being followed by spammers, the Collusionrank score of a node is computed based on the score of its followings (instead of its followers, as it is done in Pagerank or Trustrank). Thus users who follow a larger number of spammers, or who follow those who in turn follow spammers, get a negative score of higher magnitude and are pushed down in the ranking.

Algorithm 1 explains our approach. The static score vector d is initialized by setting the entries that correspond to a set of known spammers to a negative score, and the rest to 0, such that all entries of d sum to -1. Collusionrank scores are then computed using a method similar to a biased Pagerank computation (with $\alpha = 0.85$, the most commonly used value), but the score of a given node n is computed based on the scores of the nodes which are followed by n (indicated as the set $followings(n)$). In effect, if a user u follows another user v who has a low Collusionrank score (i.e., a negative score of high magnitude), the score of u gets reduced, by an amount that depends on the score of v (i.e., how 'bad' a node is u following) and the number of followers of v.

5.1.1 Collusionrank + Pagerank

It is to be noted that Collusionrank alone *cannot* be used to find popular or trustworthy users. Rather, acquiring a low Collusionrank score (a negative score of higher magnitude) indicates that a user is colluding with spammers (or with those who are colluding with spammers), for which this user should be penalized. However, Collusionrank can be combined with any ranking strategy used to identify reputed users, in order to filter out users who gain high ranks by means of link farming. This is similar to strategies proposed to combat Web spam by combining trust and distrust scores of pages in order to filter out untrustworthy pages from rankings [31].

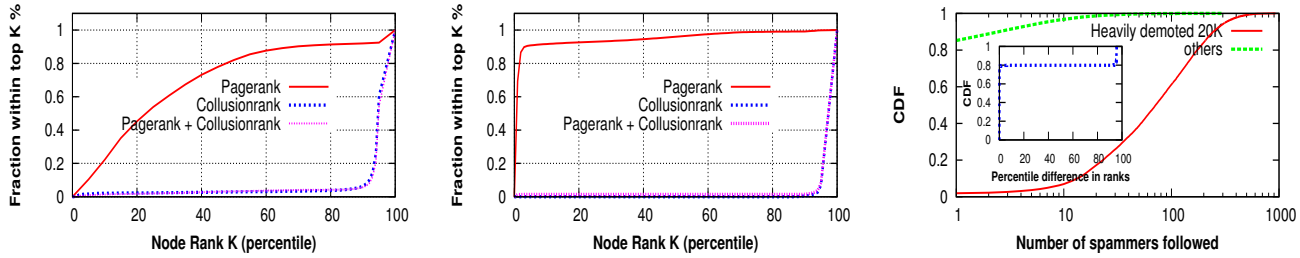
Collusionrank can be used to filter out spammers and their followers from any influence ranking strategy, such as retweetrank [9], klout [2], or any topic-sensitive Pagerank-like algorithm [29]. For simplicity, here we consider the basic Pagerank algorithm computed on the entire Twitter graph as an approach to rank users. To combine these two ranks, we take the sum of normalized Pagerank scores and normalized Collusionrank scores. As normalized values of Pagerank scores vary in the range [0,1] and normalized values of Collusionrank scores vary in the range of [-1,0], the (Pagerank + Collusionrank) score varies in the range [-1, 1].

5.2 Evaluating Collusionrank

Our goals when evaluating our approach are two-fold: (i) verify that our approach effectively lowers the reputation rankings of spammers (including those that have not yet been identified as spammers) and spam-followers, and (ii) ensure that our approach does not penalize normal users who are not spammers or spam-followers. We evaluated whether these objectives are achieved by measuring how different types of users are ranked according to Pagerank, Collusionrank, and the combination of Pagerank and Collusionrank.

We computed the Collusionrank scores of all users in the Twitter social network, considering as the set of identified spammers S, a randomly selected subset of **600** out of the 41,352 spammers. Experiments using three different randomly selected subsets of 600 spammers yielded almost identical results, which are shown in Figure 8 and described below.

Effect on rankings of spammers: Figure 8(a) shows that whereas more than 40% of the 41,352 spammers appear within the top 20% positions in Pagerank, 94% of them are demoted to the *last* 10% positions in Collusionrank. More importantly, when we look at the combined Pagerank



(a) Rankings of all 41,352 spammers (b) Rankings of social capitalists (c) Selective penalization of spammers and social capitalists

Figure 8: Using Collusionrank considering only 600 known spammers (randomly selected out of the 41,352 spammers) along with Pagerank filters out most (a) unidentified spammers and (b) social capitalists from among the top influence ranks; (c) users following a large number of spammers are selectively demoted without largely affecting the majority of common users.

+ Collusionrank, 94% of the spammers also appear in the last 10% positions. Specifically, out of the 304 spammers who ranked within the top 100,000 Pageranks (see Section 3.4), 284 (i.e., 93%) have been pushed down to very low ranks beyond 40 million in the combined (Pagerank + Collusionrank). Manual verification of the other 20 (who rank within the top 0.6 million users according to Pagerank + Collusionrank) reveals that these were either *fake* celebrity accounts, or genuine accounts which got compromised (and hence were suspended by Twitter), after which the owners abandoned these accounts and created new ones. Hence, these accounts have a relatively large number of non-spam followers, due to which they acquire relatively high ranks even in the combined Pagerank + Collusionrank. Thus the proposed approach (which uses only a small set of identified spammers) can effectively filter out most unidentified spammers from the top rankings.

Effect on rankings of social capitalists following spammers: As our approach aims to deter the social capitalists from linking with arbitrary users, it is equally important to assign low ranks to those capitalists who follow a large number of spammers. Figure 8(b) shows the rankings of the social capitalists according to different schemes. In comparison with spammers, while social capitalists are ranked much higher according to Pagerank, they rank even lower than spammers according to Collusionrank – just as social capitalists acquire high Pagerank scores due to collusion, in Collusionrank they accumulate negative scores of higher magnitude for colluding with spammers and other social capitalists (who in turn are colluding with spammers). As a result, in the combined (Pagerank + Collusionrank), 98% of the social capitalists appear in the last 10% positions.

In more detail, 18,869 out of 100K social capitalists rank within the top 100,000 according to Pagerank. Among these, 17,493 were demoted heavily by (Pagerank + Collusionrank), while the rankings of the remaining 1,376 were not affected much. This is not surprising, since we have already seen that there are several genuinely popular users among the social capitalists (e.g., Barack Obama and Britney Spears) who have large numbers of non-spammer followers. Hence they acquire high ranks even in the combined Pagerank + Collusionrank. We also observe that the median number of spammers followed by the 17,493 heavily demoted capitalists is 88, while that for the rest is 27. Thus the combined (Pagerank + Collusionrank) scheme is heavily demoting those capitalists who follow a large number of spammers. Even though we focused on social capitalists following spammers, we verified that our approach is also quite effective in lowering the reputations of all spam-followers, including those of non-targeted spam-followers i.e., the fake accounts explicitly created by spammers to farm links.

Effect on normal users who are neither spammers nor spam-followers: The results given above show that Collusionrank is able to assign very low ranks to spammers and to those who follow a large number of spammers, thus filtering them out of the higher ranks in Pagerank. However, we still need to determine whether the scheme selectively penalizes only spammers and frequent spam-followers, or whether the entire ordering of Pagerank gets changed when we combine the two rankings. For this, we consider the top 100,000 users according to Pagerank, and measure their percentile difference in ranks when ranked according to Pagerank and (Pagerank + Collusionrank). Let the rank of a given user according to Pagerank and (Pagerank + Collusionrank) be P and PC respectively. The percentile difference in the rankings of this user is computed as $\frac{|PC-P|}{N} \times 100$, where N is the total number of users in the network.

Figure 8(c) (inset) shows the CDF for the percentile difference in rank positions. Out of the top 100K users according to Pagerank, about 20K are pushed down to very low ranks in the combined (Pagerank + Collusionrank), while the rankings of the rest of the users are not affected much. Figure 8(c) (the main figure) shows the distribution of the number of spammers followed by the two sets of users – it is evident that the heavily demoted set of users follow many more spammers than the rest who are not demoted heavily.

Taken together, these observations indicate the following: (i) even when only a small set of 600 known spammers is used, our approach selectively filtered out from the top positions of Pagerank, most of the unidentified spammers and social capitalists who follow a large number of spammers, and (ii) this strategy selectively filters out spammers and frequent spam-followers without interfering with the rank-

ings of other normal users. As a result, using this approach would potentially encourage users (especially the social capitalists) to show more discretion when establishing follow links, and thus prevent spammers from easily farming links in Twitter.

6. CONCLUSION

As Twitter emerges as a popular platform for sharing real-time information on the Web, it has become a target for spammers, who try to infiltrate its social network, gain influence, and promote their tweets by acquiring (farming) follower links. In this paper, we first investigated link farming activity in Twitter and then proposed approaches to deter the activity. Our analysis of link farming resulted in a surprising finding: a small number of legitimate, popular, and highly active Twitter users account for a majority of the link farming activity. These elite users unwittingly resort to link farming as they seek to amass social capital by indiscriminately following back any user who follows them. Spammers exploit their behavior to gain followers and reputation in the network. To discourage social capitalists from connecting to unknown users, we proposed a ranking scheme, where users are penalized for following spammers. Our evaluation shows that our ranking scheme effectively lowers the influence of spammers and their followers in the network.

Acknowledgment We thank the anonymous reviewers whose suggestions helped to improve the paper. This research was supported in part by a grant from the Indo-German Max Planck Centre for Computer Science (IMPECS).

7. REFERENCES

[1] bitly blog - Spam and Malware Protection. `http://tinyurl.com/nv2oer`.

[2] Klout | The Standard for Influence. `http://klout.com/home`.

[3] There Are Now 155m Tweets Posted Per Day, Triple the Number a Year Ago. `http://rww.to/gv4VqA`, April 2011.

[4] Twitter help center: The Twitter rules. `http://tinyurl.com/22obg56`, 2011.

[5] L. Becchetti, C. Castillo, D. Donato, R. Baeza-Yates, and S. Leonardi. Link analysis for web spam detection. *ACM Transactions on the Web*, 2:1–42, March 2008.

[6] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting spammers on Twitter. In *Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS)*, 2010.

[7] K. Bharat and M. R. Henzinger. Improved algorithms for topic distillation in a hyperlinked environment. In *ACM Int'l Conference on Research and Development in Information Retrieval (SIGIR)*, 1998.

[8] C. Castillo, D. Donato, A. Gionis, V. Murdock, and F. Silvestri. Know your neighbors: web spam detection using the web topology. In *ACM Int'l Conference on Research and Development in Information Retrieval (SIGIR)*, 2007.

[9] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring user influence in Twitter: the million follower fallacy. In *AAAI Int'l Conference on Weblogs and Social Media (ICWSM)*, 2010.

[10] S. Chakrabarti. Integrating the document object model with hyperlinks for enhanced topic distillation and information extraction. In *ACM Int'l Conference on World Wide Web (WWW)*, 2001.

[11] B. D. Davison. Recognizing nepotistic links on the web. In *AAAI Workshop on Artificial Intelligence for Web Search*, 2000.

[12] D. Gayo-Avello and D. J. Brenes. Overcoming Spammers in Twitter - a tale of five algorithms. In *Spanish Conference on Information Retrieval (CERI)*, 2010.

[13] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: the underground on 140 characters or less. In *ACM Int'l Conference on Computer and Communications Security (CCS)*, 2010.

[14] Z. Gyöngyi and H. Garcia-Molina. Link spam alliances. In *Int'l Conference on Very Large Data Bases (VLDB)*, 2005.

[15] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen. Combating web spam with trustrank. In *Int'l Conference on Very Large Data Bases (VLDB)*, 2004.

[16] T. H. Haveliwala. Topic-sensitive pagerank. In *ACM Int'l Conference on World Wide Web (WWW)*, 2002.

[17] US confirms it asked Twitter to stay open to help Iran protesters. `http://tinyurl.com/klv36p`.

[18] H. Kwak, H. Chun, and S. Moon. Fragile online relationship: a first look at unfollow dynamics in Twitter. In *Annual Conference on Human Factors in Computing Systems (CHI)*, 2011.

[19] K. Lee, J. Caverlee, and S. Webb. Uncovering social spammers: social honeypots + machine learning. In *ACM Int'l Conference on Research and Development in Information Retrieval (SIGIR)*, 2010.

[20] K. Lee, B. D. Eoff, and J. Caverlee. Seven months with the devils: a long-term study of content polluters on Twitter. In *AAAI Int'l Conference on Weblogs and Social Media (ICWSM)*, 2011.

[21] R. Lempel and S. Moran. The stochastic approach for link-structure analysis (SALSA) and the TKC effect. *Computer Networks*, 33:387–401, Jun 2000.

[22] A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. *SIGCOMM Computer Communication Review*, 36:291–302, Aug 2006.

[23] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *ACM Int'l Conference on World Wide Web (WWW)*, 2010.

[24] M. Sobek. Google PageRank - PR 0. `http://pr.efactory.de/e-pr0.shtml`.

[25] D. Talbot. How Google Ranks Tweets. `http://www.technologyreview.in/web/24353/`.

[26] J. Teevan, D. Ramage, and M. R. Morris. #TwitterSearch: a comparison of microblog search and web search. In *ACM Int'l Conference on Web Search and Data Mining (WSDM)*, 2011.

[27] K. Thomas, C. Grier, V. Paxson, and D. Song. Suspended accounts in retrospect: an analysis of Twitter spam. In *ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2011.

[28] L. Rao, Twitter Seeing 90 Million Tweets Per Day, 25 Percent Contain Links, *TechCrunch*, 2010. `http://tinyurl.com/27x5cay`.

[29] J. Weng, E.-P. Lim, J. Jiang, and Q. He. TwitterRank: finding topic-sensitive influential Twitterers. In *ACM Int'l Conference on Web Search and Data Mining (WSDM)*, 2010.

[30] B. Wu and B. D. Davison. Identifying link farm spam pages. In *ACM Int'l Conference on World Wide Web (WWW)*, 2005.

[31] B. Wu, V. Goel, and B. D. Davison. Propagating trust and distrust to demote web spam. In *Workshop on Models of Trust for the Web*, 2006.

[32] S. Yardi, D. Romero, G. Schoenebeck, and D. M. Boyd. Detecting spam in a twitter network. *First Monday*, 15(1):1–13, Jan 2010.

[33] C. M. Zhang and V. Paxson. Detecting and analyzing automated activity on Twitter. In *Int'l Conference on Passive and Active Measurement (PAM)*, 2011.