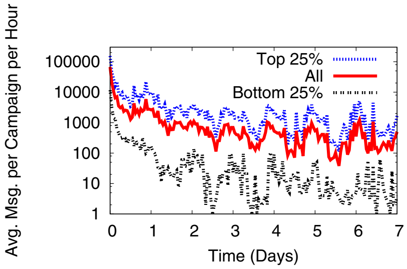

Figure 18: Message creation over time.

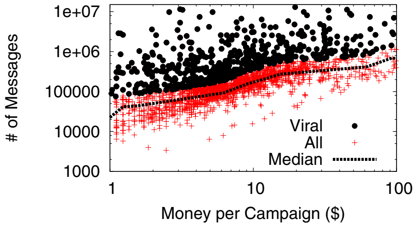

Figure 19: Cost of Weibo campaigns.

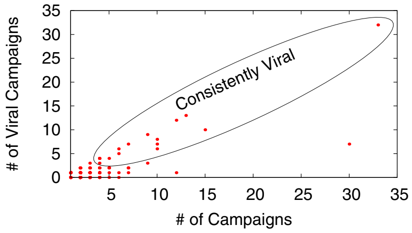

Figure 20: Viral campaigns per customer.

ily blend in. This may represent a conscious effort on the part of workers to make their Weibo accounts appear "normal" so that they will evade automatic Sybil detectors. Customers tend to have follow rates >1. This makes sense, since customers tend to be commercial entities, and are thus net information disseminators rather than information consumers.

4.3 Information Dissemination on Weibo

Much work has studied how to optimize information dissemination on social networks. We analyze our data to evaluate the level of success in crowdturfing cascades, and whether there are factors that can predict the success of social crowdturfing campaigns.

Campaign Analysis. We start by examining the number of *messages* generated by crowdturfing campaigns on Weibo. We define a message as a single entry in a Weibo timeline. A tweet from a single user generates f messages, where f is their number of followers. The number of messages in a campaign is equal to the number of messages generated by the customer, workers, and any normal users who retweet the content. Total messages per campaign represents an upper bound on the *audience size* of that campaign. Since we have an incomplete view of the Weibo social graph, we cannot quantify the number of duplicate messages per user.

Figure 16 shows the CDF of messages generated by Weibo campaigns. 50% of campaigns generate $\leq$146K messages, and 8% manage to breach the 1M-message milestone. As expected, workers are responsible for the vast majority of messages, *i.e.* there are very few retweets. Considering the low cost of these campaigns, however, these raw numbers are nonetheless impressive.

Next, we want to examine the depth of crowdturfing cascades. Figure 17 plots the depth of cascades measured as the height of each information cascade tree. Pay-per-tweet campaigns are very shallow, *i.e.* worker's tweets are rarely retweeted by normal users. In contrast, pay-per-retweet campaigns are more successful at engaging normal users: 50% reach depths >2, *i.e.* they include at least one retweet from a normal user. One possible explanation for the success of pay per retweet is that normal users may place greater trust in information that is retweeted from a popular customer, rather than content authored by random worker accounts.

Next, we examine the temporal dynamics of crowdturfing campaigns. Figure 18 shows the number of messages generated per hour after each campaign is initiated. The "all" line is averaged across all campaigns, while the top- and bottom-25% lines focus on the largest and smallest campaigns (in terms of total messages). Most messages are generated during a campaigns' first hour (10K on average), which is bolstered by the high-degree of customers (who tend to be super-nodes), and the quick responses of career crowdturfers (see Figure 7). However, by the end of the first day, the message rate drops to ≈1K per hour. There is a two order of magnitude difference between the effectiveness of the top- and bottom-25% campaigns, although they both follow the same falloff trend after day 1.

Factors Impacting Campaign Success. We now take a look at factors that may affect the performance of crowdturfing cascades. The high-level question we wish to answer is: are there specific ways to improve the probability that a campaign goes *viral*?

The first factor we examine is the cost of the campaign. Figure 19 illustrates the number of messages generated by Weibo campaigns versus their cost. The median line, around which the bulk of campaigns are clustered, reveals a linear relationship between money and messages. This result is intuitive: more money buys more workers, who in turn generate more messages. However, Figure 19 also reveals the presence of *viral* campaigns, which we define as campaigns that generate at least two times more messages than their cost would predict. There are 723 viral campaigns scattered randomly throughout the upper portion of Figure 19. This shows that viral popularity is independent of campaign budget.

We look at whether specific workers are better at generating viral campaigns. We found that individual workers are not responsible for the success of viral campaigns. The only workers consistently involved in viral campaigns are career crowdturfers, who tend to be involved in *all* campaigns, viral or not.

Surprisingly, a small number of customers exhibit a consistent ability to start viral campaigns. Figure 20 plots the total number of campaigns started by each customer vs. the number that went viral, for all customers who started at least 1 viral campaign. The vast majority of customers initiate $\leq$3 campaigns, which makes it difficult to claim correlation when one or more go viral. However, the 20 customers (1.5%) in the highlighted region do initiate a significant number of campaigns, and they go viral $\geq$50% of the time. Since many of these customers do not actively participate in their own campaigns, this suggests that campaigns go viral because their content is of interest to Weibo users, perhaps because they are related to customers such as well-known actors or performers.

5. ACTIVE EXPERIMENTS

Our next step to understanding crowdsurfing systems involves a look from the perspective of a paying customer on ZBJ. We initiate a number of benign advertising campaigns on different platforms and subjects. By redirecting the click traffic through a *measurement server* under our control, we are able to analyze the clicks of workers and of users receiving crowdturf content in real-time. We begin by describing our experimental setup before moving on to our findings, and conclude with a discussion of practical lessons we learned during this process.

5.1 Experimental Setup

Methodology. Figure 21 depicts the procedure we use to collect real-time data on crowdturfing clicks. The process begins when we post a new campaign to ZBJ that contains a brief description of the tasks, along with a URL ("Task Info" in Figure 21) that workers can click on to find details and to perform the tasks. The task details page is hosted on our measurement server, and thus any worker

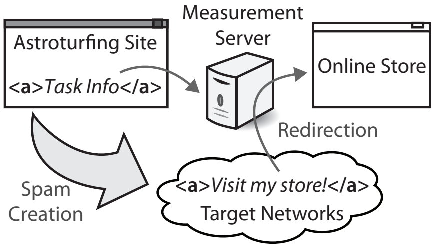

Figure 21: Crowdturfing data collection.

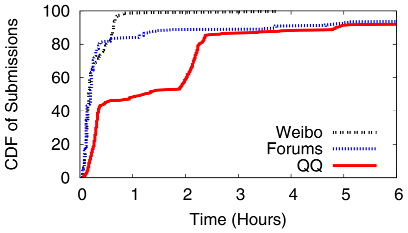

Figure 22: Response time of ZBJ workers.

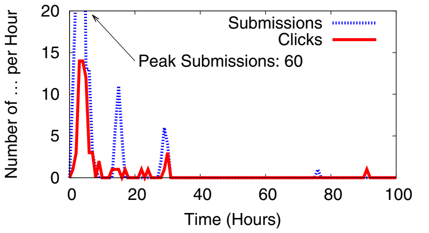

Figure 23: Long campaign characteristics.

who wants to accept our tasks must first visit our server, where we collect their information (*i.e.* IP, timestamp, etc). Referring workers to task details on external sites is a common practice on ZBJ, and does not raise suspicion among workers.

Workers that accept our tasks are directed to post spam messages that advertise real online stores to one of three target networks: Weibo, QQ instant message groups, and discussion forums. The posted messages urge normal users to click embedded links ("Visit my store!" in Figure 21) that take them to our measurement server. The measurement server records some user data before transparently redirecting them to the real online store.

We took care to preserve the integrity of our experimental setup. Because some Chinese Internet users have limited access to websites hosted outside of mainland China, we placed our measurement server in China, and only advertised legitimate Chinese e-commerce sites. In addition, we also identified many search engines and bots generating clicks on our links, and filtered them out before analyzing our logs.

Campaign Details. In order to experiment with a variety of topics and venues, we posted nine total campaigns to ZBJ in October 2011. As shown in Table 3, we created three different advertising campaigns (*iPhone4S*, *Maldives*, and *Raffle*), and targeted each at three distinct networks. We discuss a fourth campaign, *OceanPark*, later in the section.

The first campaign promotes an unofficial iPhone dealer who imports iPhones from North America and sells them in China. We launched this campaign on October 4, 2011, immediately after Apple officially unveiled the iPhone 4S. In the task requirements, we required workers to post messages advertising a discount price from the dealer on the iPhone 4S ($970).

The second campaign tried to sell a tour package to the Maldives (a popular tourist destination in China). The spam advertises a 30% group-purchase discount offered by the seller that saves $600 on the total trip price ($1542 after discount). The third campaign tells users about an online raffle hosted by a car company. Anyone could participate in the raffle for free, and the prizes were 200 pre-paid calling cards worth $4.63 each.

All campaigns shared the same set of baseline requirements. Each campaign had a budget of $15 on each target network, and workers had a time limit of 7 days to perform tasks. The desired number of tasks was set to either 50 or 100, depending on the campaign type. Submissions were not accepted if the content generated by the worker was deleted by spam detection systems within 24 hours of creation. These baseline requirements closely match the expected norms for campaigns on ZBJ (see Figure 4 and Table 2).

We applied additional requirements for campaigns on specific networks. For campaigns on the QQ instant messaging network, workers were required to generate content in groups with a minimum of 300 members. For campaigns on user discussion forums, workers were only allowed to post content on a predefined list of forums that receive at least 1,000 hits per day.

Each campaign type had additional, variable requirements. For Maldives and Raffle campaigns, the price per task was set to $0.154, meaning 100 submissions would be accepted. However, the price for iPhone4S tasks was doubled to $0.308 with an expectation of 50 submissions. iPhone 4S tasks were more challenging for two reasons. On Weibo, workers were required to tweet using accounts with at least 3,000 followers. On QQ, workers needed to spam two groups instead of one. Finally, on forums, the list of acceptable sites was reduced to only include the most popular forums.

5.2 Results and Analysis

Table 3 lists the high level results of from our crowdturfing campaigns, including 9 short campaigns and the "OceanPark" campaign. Seven of the short campaigns received sufficient submissions, and six were completed within a few hours (Time column). Interestingly, workers continued submitting to campaigns even after they were "full," in the hopes that earlier submissions would be rejected, and they would claim the reward. In total, the short campaigns garnered 894 submissions from 224 distinct workers.

Figure 22 shows the response times of workers for campaigns targeting different networks. We aggregate the data across campaign types rather than networks because workers' ability to complete tasks is based on the number of accounts they control on each network. More than 80% of submissions are generated within an hour for Weibo and forum campaigns, and within six hours for QQ.

The "Msgs" column lists the number of *messages* generated by each campaign. For Weibo campaigns, we calculate messages using the same methodology as in Section 4. For QQ campaigns, messages are calculated as the number of users in all QQ groups that received spam from our workers. We cannot estimate the number of messages for forums because we do not know how many users browse these sites.

We can understand the effectiveness of different crowdturfing strategies by comparing the number of messages generated to the number of clicks (responses by normal users, "Clicks" column in Table 3). We see that QQ campaigns are the most effective, and generate more clicks than Weibo campaigns despite generating only 1/5 as many messages as Weibo. One possible reason is that QQ messages pop-up directly on users' desktops, leading to more views and clicks. Tweets on Weibo, on the other hand, are not as invasive, and may get lost in the flood of tweets in each user's timeline. Forums perform the worst of the three, most likely because admins on popular forums are diligent about deleting spammy posts.

Finally, we try to detect the presence of Sybil accounts (multiple accounts controlled by one user) on crowdturfing sites. Column "W/IP" in Table 3 compares the number of distinct workers (W) to the number of distinct IPs (IP) that click on the "Task Info" link (see Figure 21) in each campaign. If $W > IP$, then not all ZBJ workers clicked the link to read the instructions. This suggests that multiple ZBJ worker accounts are controlled by a single user, who viewed the instructions once before completing tasks from multiple

Campaign	Network	Subm.	Time	Msgs.	Clicks	W/IP
iPhone4S	Weibo	47	45min	197K	204	24/54
	QQ	41	6hr	35K	244	34/36
	Forums	71	3day	N/A	43	40/22
Maldives	Weibo	108	3h	220K	28	35/30
	QQ	118	4h	46K	187	24/29
	Forums	123	4h	N/A	3	18/11
Raffle	Weibo	131	2h	311K	47	67/38
	QQ	131	6day	60K	78	29/33
	Forums	124	1day	N/A	0	28/9
OceanPark	Weibo	204	4day	369K	63	204/99

Table 3: Results from our crowdturfing campaigns.

accounts. Our results show that $W > IP$ for 66% of our campaigns. Thus, not only do crowdturfers utilize multiple accounts on target websites to complete tasks (Figure 14), but they also have multiple accounts on crowdturfing sites themselves.

Long Campaigns. The campaigns we have analyzed thus far all required $\leq$100 tasks, and many were completed within about an hour by workers (see Figure 22). These short campaigns favor career crowdturfers, who control many accounts on target websites and move rapidly to generate submissions.

To observe the actions of less prolific workers, we experimented with a longer campaign that required 300 tasks. This campaign included an additional restriction to limit career crowdturfers: each ZBJ worker account could only submit once. The goal of the campaign was to advertise discount tickets to an ocean-themed amusement park in Hong Kong on Weibo. This campaign is listed as *OceanPark* in Table 3.

Figure 23 plots the number of worker submissions and clicks from Weibo users over time for the OceanPark campaign. Just as in previous experiments, the first 100 submissions were generated within the first few hours. Clicks from users on the advertised links closely track worker submission patterns. Overall, 191 submissions were received on day one, 11 more on day two, and 2 final submissions on day four, for a total of 204 submissions. This indicates that there are $\approx$200 active Weibo workers on ZBJ: if there were more, they would have submitted to claim one of the 97 incomplete tasks in our campaign.

Discussion. Our real-world experiments demonstrate the feasibility of crowd-sourced spamming. The iPhone4S and Maldives campaigns were able to generate 491 and 218 click-backs (respectively) while only costing $45 each. Considering that the iPhone 4S sells for $970 in China, and the Maldives tour package costs $1,542, just a single sale of either item would be more than enough to recoup the entire crowdturfing fee. The *cost per click* (CPC) of these campaigns are $0.21 and $0.09, respectively, which is more expensive than observed CPC rates ($0.01) for traditional display advertising on the web [30]. However, with improved targeting (*i.e.* omitting underperformers like forum spam) the costs could be reduced, bringing CPC more in line with display advertising.

Our Maldives campaign is a good indicator of the effectiveness of crowdturfing. The tour website listed 4 Maldives trips sold to 2 people in the month before our campaign. However, the day our Maldives campaign went live, 11 trips were sold to 2 people. In the month after our campaign, no additional trips were sold. While we cannot be sure, it is likely that the 218 clicks from our campaign were responsible for these sales.

6. CROWDTURFING GOES GLOBAL

In previous sections, we focused on the crowdturfing market in China. We now take a global view and survey the market for crowdturfing systems in the U.S. and India. Additional crawls conducted by us, as well as prior work from other researchers, demonstrates that crowdturfing systems in the U.S. are very active, and are supported by an international workforce.

Website	Campaigns	% Crowdturfing	Tasks	$ per Subm.
Amazon Turk (US)	41K	12%	2.9M	$0.092
ShortTask* (US)	30K	95%	527K	$0.096
MinuteWorkers (US)	710	70%	10K	$0.241
MyEasyTask (US)	166	83%	4K	$0.149
Microworkers (US)	267	89%	84K	$0.175
Paisalive (India)	107	N/A	N/A	$0.01

Table 4: Details of U.S. and Indian crowd-sourcing sites. Data encompasses one month of campaigns, except ShortTask which is one year.

Mechanical Turk. Although prior work has found that 41% of tasks on Mechanical Turk were spam related in 2010 [15], our measurements indicate that this is no longer the case. We performed hourly crawls of Mechanical Turk for one month in October 2011, and used keyword analysis to classify tasks. As shown in Table 4, crowdturfing now only accounts for only 12% of campaigns.

Other U.S. Based Sites. However, the drop in crowdturfing on Mechanical Turk does not mean this problem has gone away. Instead, crowdturfing has just shifted to alternative websites. For example, recent work has shown that 31% of the jobs on Freelancer over the last seven years were related to search engine optimization (SEO), Sybil account creation, and spam [24]. Many SEO products are also available on eBay: trivial keyword searches turn up many sellers offering bulk Facebook likes/fans and Twitter followers.

To confirm this finding, we crawled four U.S. based crowd-sourcing sites that have been active since 2009. Since they do not provide information on past tasks, we crawled MinuteWorkers, MyEasyTask, and Microworkers once a day during the month of October 2011. ShortTask does provide historical data for tasks going back one year, hence we only crawled them once. As shown in Table 4, keyword classification reveals that between 70-95% of campaigns on these sites are crowdturfing. We manually verified that the remaining campaigns were not malicious. The types of campaigns on these sites closely matches the types found on Freelancer, *i.e.* the most prevalent campaign type is SEO [24].

Sites like ShortTask, Microworkers, and MyEasyTask fill two needs in the underground market. First, they do not enforce any restrictions against crowdturfing. This contrasts with Mechanical Turk, which actively enforces policies against spammy jobs [6]. Second, these sites enable a truly international workforce by supporting a wide range of payment methods. Amazon requires workers to have U.S. bank accounts, or to accept cheques in Indian Rupees, and hence most “turkers” are located in the US (46.8%) and India (34%) [14]. However, alternative crowd-sourcing sites support payments through systems like Paypal and E-Gold, which makes them accessible to non-U.S. and non-Indian workers. For example, Microworkers come from Indonesia (18%), Bangladesh (17%), Philippines (5%), and Romania (5%) [12]. Freelancers are also located in the United Kingdom and Pakistan [24].

Paisalive. We located one crowdturfing site in India called Paisalive that takes globalization even further. As shown in Table 4, Paisalive is very small and the wages are very low compared to other services. However, the interesting feature of Paisalive is that it is e-mail based: workers sign up on the website, and afterwards all task requests and submissions are handled through e-mail. This design is geared towards enabling workers in rural populations constrained by low-bandwidth, intermittent Internet connectivity.

7. RELATED WORK

Crowd-sourcing Research. Since coming online in 2005, Amazon's Mechanical Turk has been scrutinized by the research community. This includes studies of worker demographics [14, 28], task pricing [7, 13], and even meta-studies on how to use Mechanical Turk to conduct user studies [18]. The characteristics of Micro Workers have also been thoroughly studied [12].

OSN Spam and Detection. Researchers have identified copious amounts of fake accounts and spam campaigns on large OSNs like Facebook [8], Twitter [11, 32], and Renren [34]. The growing threat posed by this malicious activity has spurred work that aims to detect and stop OSN spam using machine learning techniques [3, 33, 31]. This body of research has focused on analyzing and defending against the outward manifestations of OSN spam. In contrast, our work identifies some of the underlying systems used by attackers to generate spam and evade security measures.

Opinion Spam. Spam that attempts to influence the opinions and actions of normal people has become more prevalent in recent years [16]. Researchers have been working on detecting and characterizing fake product reviews [22, 17], fake comments on news sites [4], and astroturf political campaigns on Twitter [27]. The authors of [26] created a model to help classify deceptive reviews generated by Mechanical Turk workers. These works reaffirm our results, that crowdturfing is a growing, global threat on the web.

8. CONCLUSION

In this paper, we contribute to the growing pool of knowledge about malicious crowd-sourcing systems. Our analysis of the two largest crowdturfing sites in China reveals that $4 million dollars have already been spent on these two sites alone. The number of campaigns and dollars spent on ZBJ and SDH are growing exponentially, meaning that the problems associated with crowdturfing will continue to get worse in the future.

We measure the real-world ramifications of crowdturfing by looking at spam dissemination on Weibo, and by becoming active customers of ZBJ. Our results reveal the presence of career crowdturfers that control thousands of accounts on OSNs, and manage them carefully by hand. We find that these workers are capable of generating large information cascades, while avoiding the security systems that are designed to catch automated spam. We also observe that this spam is highly effective, driving hundreds of clicks from normal users.

Finally, our survey of crowdturfing sites in the U.S. and elsewhere demonstrates the global nature of this problem. Unscrupulous crowd-sourcing sites, coupled with international payment systems, have enabled a burgeoning crowdturfing market that targets U.S. websites, fueled by a global workforce. As part of ongoing work, we are exploring the design and quantifying the effectiveness of both passive and active defenses against these systems.

Acknowledgments

The authors wish to thank Zengbin Zhang for his help on ZBJ experiments. This project is supported by NSF under IIS-0916307, and also supported by WCU (World Class University) program under the National Research Foundation of Korea and funded by the Ministry of Eduation, Science and Technology of Korea (Project No: R31-30007).

9. REFERENCES

[1] China's internet users targeted in online rumour probes. BBC, 2011.

[2] Websites shut over illegal PR deals. China Daily, August 2011.

[3] BENEVENUTO, F., MAGNO, G., RODRIGUES, T., AND ALMEIDA, V. Detecting spammers on twitter. In *Proc. of CEAS* (2010).

[4] CHEN, C., ET AL. Battling the internet water army: Detection of hidden paid posters. *CoRR abs/1111.4297* (2011).

[5] DUAN, Y. The invisible hands behind web postings. China Daily, June 2010.

[6] EATON, K. Mechanical turk's unsavory side effect: Massive spam generation. Fast Company, December 2010.

[7] FARIDANI, S., HARTMANN, B., AND IPEIROTIS, P. G. What's the right price? pricing tasks for finishing on time. In *Proc. of AAAI Workshop on Human Computation* (2011).

[8] GAO, H., HU, J., WILSON, C., LI, Z., CHEN, Y., AND ZHAO, B. Y. Detecting and characterizing social spam campaigns. In *Proc. of IMC* (2010).

[9] GE, L., AND LIU, T. 'water army' whistleblower threatened. Global Times, January 2011.

[10] GILES, J. Inside facebook's massive cyber-security system. New Scientist, October 2011.

[11] GRIER, C., THOMAS, K., PAXSON, V., AND ZHANG, M. @spam: the underground on 140 characters or less. In *Proc. of CCS* (2010).

[12] HIRTH, M., HOSSFELD, T., AND TRAN-GIA, P. Anatomy of a crowdsourcing platform - using the example of microworkers.com. In *Proc. of IMIS* (2011).

[13] IPEIROTIS, P. G. Analyzing the amazon mechanical turk marketplace. *XRDS 17* (December 2010), 16–21.

[14] IPEIROTIS, P. G. Demographics of mechanical turk. NYU Working Paper, 2010.

[15] IPEIROTIS, P. G. Mechanical turk: Now with 40.92% spam. Behind Enemy Lines blog, December 2010.

[16] JINDAL, N., AND LI, B. Opinion spam and analysis. In *Proc. of WSDM* (2008).

[17] JINDAL, N., LIU, B., AND LIM, E.-P. Finding unusual review patterns using unexpected rules. In *Proc. of CIKM* (2010).

[18] KITTUR, A., CHI, H., AND SUH, B. Crowdsourcing user studies with mechanical turk. In *Proc. of CHI* (2008).

[19] KWAK, H., LEE, C., PARK, H., AND MOON, S. What is twitter, a social network or a news media? In *Proc. of WWW* (2010).

[20] LAU, A. Reputation management: PR vs. search vs. china's water army. Search Engine Watch, May 2011.

[21] LE, Z. Average salaries up 13-14% last year as income disparity increases. Global Times, May 2011.

[22] LIM, E.-P., ET AL. Detecting product review spammers using rating behaviors. In *Proc. of CIKM* (2010).

[23] MASNICK, M. Bot-on-bot ebay scamming. Techdirt, July 2006.

[24] MOTOYAMA, M., MCCOY, D., LEVCHENKO, K., SAVAGE, S., AND VOELKER, G. M. Dirty jobs: The role of freelance labor in web service abuse. In *Proc. of Usenix Security* (2011).

[25] NI, V. China's internet users hit 485 million, weibo users and group buyers surge. China Briefing, July 2011.

[26] OTT, M., CHOI, Y., CARDIE, C., AND HANCOCK, J. T. Finding deceptive opinion spam by any stretch of the imagination. In *Proc. of ACL* (2011).

[27] RATKIEWICZ, J., ET AL. Detecting and tracking political abuse in social media. In *Proc. of ICWSM* (2011).

[28] ROSS, J., ET AL. Who are the crowdworkers?: Shifting demographics in amazon mechanical turk. In *Proc. of CHI* (2010).

[29] SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. Photo tourism: Exploring photo collections in 3d. *ACM ToG 25*, 3 (2006).

[30] STONE-GROSS, B., ET AL. Understanding fraudulent activities in online ad exchanges. In *Proc. of IMC* (2011).

[31] STRINGHINI, G., KRUEGEL, C., AND VIGNA, G. Detecting spammers on social networks. In *Proc. of ACSAC* (2010).

[32] THOMAS, K., ET AL. Suspended accounts in retrospect: An analysis of twitter spam. In *Proc. of IMC* (2011).

[33] WANG, A. H. Don't follow me: Spam detection on twitter. In *Proc. of SECRYPT* (2010).

[34] YANG, Z., WILSON, C., WANG, X., GAO, T., ZHAO, B. Y., AND DAI, Y. Uncovering social network sybils in the wild. In *Proc. of IMC* (2011).