Crowdturfers, Campaigns, and Social Media: Tracking and Revealing Crowdsourced Manipulation of Social Media

Kyumin Lee*, **Prithivi Tamilarasan***, **James Caverlee**
Texas A&M University
College Station, TX 77843
{kyumin, prithivi, caverlee}@cse.tamu.edu

Abstract

Crowdturfing has recently been identified as a sinister counterpart to the enormous positive opportunities of crowdsourcing. Crowdturfers leverage human-powered crowdsourcing platforms to spread malicious URLs in social media, form "astroturf" campaigns, and manipulate search engines, ultimately degrading the quality of online information and threatening the usefulness of these systems. In this paper we present a framework for "pulling back the curtain" on crowdturfers to reveal their underlying ecosystem. Concretely, we analyze the types of malicious tasks and the properties of requesters and workers in crowdsourcing sites such as Microworkers.com, ShortTask.com and Rapidworkers.com, and link these tasks (and their associated workers) on crowdsourcing sites to social media, by monitoring the activities of social media participants. Based on this linkage, we identify the relationship structure connecting these workers in social media, which can reveal the implicit power structure of crowdturfers identified on crowdsourcing sites. We identify three classes of crowdturfers – professional workers, casual workers, and middlemen – and we develop statistical user models to automatically differentiate these workers and regular social media users.

Introduction

Crowdsourcing systems have successfully leveraged the attention of millions of "crowdsourced" workers to tackle traditionally vexing problems. From specialized systems like Ushahidi (for crisis mapping), Foldit (for protein folding) and Duolingo (for translation) to general-purpose crowdsourcing platforms like Amazon Mechanical Turk and Crowdflower – these systems have shown the effectiveness of intelligently organizing large numbers of people.

However, these positive opportunities have a sinister counterpart: large-scale "crowdturfing," wherein masses of cheaply paid shills can be organized to spread malicious URLs in social media, form artificial grassroots campaigns ("astroturf"), and manipulate search engines. For example, it has been recently reported that Vietnamese propaganda officials deployed 1,000 crowdturfers to engage in online discussions and post comments supporting the Communist Party's policies (Pham 2013). Similarly, the Chinese "Internet Water Army" can be hired to post positive comments for the government or commercial products, as well as disparage rivals (Sterling 2010; Wikipedia 2013). Mass organized crowdturfers are also targeting popular services like iTunes (Chan 2012) and attracting the attention of US intelligence operations (Fielding and Cobain 2011). And increasingly, these campaigns are being launched from commercial crowdsourcing sites, potentially leading to the commoditization of large-scale turfing campaigns. In a recent study of the two largest Chinese crowdsourcing sites Zhubajie and Sandaha, Wang et al. (Wang et al. 2012) found that ∼90% of all tasks were for crowdturfing.

Hence, in this paper we are interested to explore the ecosystem of crowdturfers. Who are these participants? What are their roles? And what types of campaigns are they engaged in? Unfortunately, crowdsourcing sites typically only reveal very limited information about each worker – like a username and a date joined – meaning that detailed analysis is inherently challenging. As a result, we propose to link workers to their activity in social media. By using this linkage, can we find crowd workers in social media? Can we uncover the implicit power structure of crowdturfers? Can we automatically distinguish between the behaviors of crowdturfers and regular social media users? Toward answering these questions, we make the following contributions in this paper:

- We first analyze the types of malicious tasks and the properties of requesters and workers on Western crowdsourcing sites such as Microworkers.com, ShortTask.com and Rapidworkers.com. Previous researchers have investigated Chinese-based crowdsourcing sites; to our knowledge this is the first study to focus primarily on Western crowdsourcing sites.
- Second, we propose a framework for linking tasks (and their workers) on crowdsourcing sites to social media, by monitoring the activities of social media participants on Twitter. In this way, we can track the activities of crowdturfers in social media where their behavior, social network topology, and other cues may leak information about the underlying crowdturfing ecosystem.
- Based on this framework, we identify the hidden information propagation structure connecting these workers in social media, which can reveal the implicit power structure of crowdturfers identified on crowdsourcing

*Authors with equal contributions.

sites. Specifically, we identify three classes of crowdturfers – professional workers, casual workers, and middlemen – and we demonstrate how their roles and behaviors are different in social media.

- Finally, we propose and develop statistical user models to automatically differentiate among regular social media users and workers. Our experimental results show that these models can effectively detect previously unknown Twitter-based workers.

Related Work

With the rise in popularity of commercial crowdsourcing services, there have been many efforts to analyze the nature of jobs available and their characteristics. For example Kittur et al. (Kittur, Chi, and Suh 2008) studied Amazon Mechanical Turk and found that a large number of workers can be hired within a short time and for low cost. Similar studies – e.g., (Brabham 2008) – have shown the potential of crowdsourcing. And researchers have begun developing new crowd-based platforms – e.g., (Alonso, Rose, and Stewart 2008) (Franklin et al. 2011) – for augmenting traditional information retrieval and database systems, embedding crowds into workflows (like document authoring) (Bernstein et al. 2010), and so forth.

A key question for these crowd-based systems is how to control the quality of workers and outputs due to the openness of these sites. For example, Venetis and Garcia-Molina (Venetis and Garcia-Molina 2012) described two quality control mechanisms. The first mechanism repeats each task multiple times and combines the results from multiple users. The second mechanism defines a score for each worker and eliminates the work from users with low scores. Xia et al. (Xia et al. 2012) provided a real-time quality control strategy for workers who evaluate the relevance of search engine results based on the combination of a qualification test of the workers and the time spent on the actual task. The results are promising and these strategies facilitate reducing the number of bad workers. Note, however, that our interest in this paper is on crowdsourcing sites that deliberately encourage crowdturfing.

Recently, Wang et al. (Wang et al. 2012) coined the term "crowdturfing" (crowdsourcing + astroturfing) to refer to crowdsourcing systems where malicious campaigns are hosted by employers. They have studied crowdsourcing sites based in China and the impact of these sites on one social networking site – Weibo.

Analysis of Crowdturfing Tasks and Participants

In this section, we begin our study through an examination of the different types of crowdturfing campaigns that are posted in Western crowdsourcing sites and study the characteristics of both *requesters* (who post jobs) and *workers* (who actually perform the jobs).

We collected 505 campaigns by crawling three popular Western crowdsourcing sites that host clear examples of crowdturfing campaigns: Microworkers.com, ShortTask.com, and Rapidworkers.com during a span of two

Twitter Post: Getmine

1. Go to http://getminecraftforfree.org
2. Click on the tweet button on the left side
3. Tweet something like "how to play minecraft for free" or "check this site out"
4. Include link to the site

Figure 1: An example social media manipulation campaign.

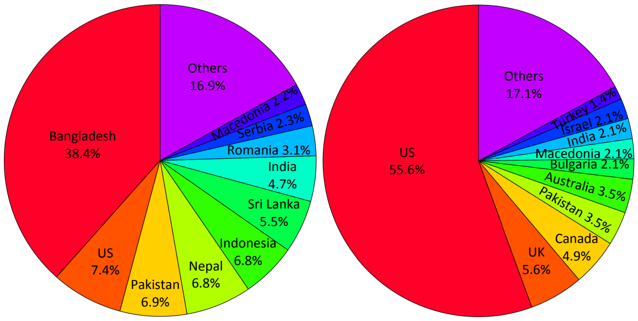
months in 2012. Almost all campaigns in these sites are crowdturfing campaigns, and these sites are active in terms of number of new campaigns. Note that even though Amazon Mechanical Turk is one of the most popular crowdsourcing sites, we excluded it in our study because it has only a small number of crowdturfing campaigns and its terms of service officially prohibits the posting of crowdturfing campaigns.[1] For the 505 sampled campaigns, each has multiple tasks, totaling 63,042 tasks.

Types of Crowdturfing Campaigns

Analyzing the types of crowdturfing campaigns available in crowdsourcing sites is essential to understand the tactics of the requesters. Hence, we first manually grouped the 505 campaigns into the following five categories:

- **Social Media Manipulation [56%]:** The most popular type of campaign targets social media. Example campaigns request workers to spread a meme through social media sites such as Twitter, click the "like" button of a specific Facebook profile/product page, bookmark a webpage on Stumbleupon, answer a question with a link on Yahoo! Answers, write a review for a product at Amazon.com, or write an article on a personal blog. An example campaign is shown in Figure 1, where workers are requested to post a tweet including a specific URL.
- **Sign Up [26%]:** Requesters ask workers to sign up on a website for several reasons, for example to increase the user pool, to harvest user information like name and email, and to promote advertisements.
- **Search Engine Spamming [7%]:** For this type of campaign, workers are asked to search for a certain keyword on a search engine, and then click the specified link (which is affiliated with the campaign's requester), toward increasing the rank of the page.
- **Vote Stuffing [4%]:** Requesters ask workers to cast votes. In one example, the requester asked workers to vote for "Tommy Marsh and Bad Dog" to get the best blue band award in the Ventura County Music Awards (which the band ended up winning!).
- **Miscellany [7%]:** Finally, a number of campaigns engaged in some other activity: for example, some requested workers to download, install, and rate a particular software package; others requested workers to participate in a survey or join an online game.

[1] Perhaps surprisingly, Microworkers.com is ranked by Alexa.com at the 4,699th most popular website while Amazon Mechanical Turk is ranked 7,173.



(a) Workers (b) Requesters

Figure 2: Top 10 countries of workers and requesters.

Table 1: Characteristics of Crowdturfing Workers.

	# of Tasks	Total Earned ($)	Longevity (day)
Max	24,016	3,699	1,215
Avg	738	117	368
Median	166	23	320
Min	10	1	5

We see that most crowdturfing campaigns target social media; for this reason, we will return in the following section with a framework for harvesting user activity in social media for further exploring the ecosystem of crowdturfing.

Requesters and Workers

We now turn to an analysis of the requesters who have posted these jobs and the workers who have actually completed them. Since this type of information is potentially quite revealing, both ShortTask.com and Rapidworkers.com do not reveal any information about their requesters and workers (aside from username). Luckily, Microworkers.com does provide payment information, country of origin, and other detailed characteristics of both requesters and workers. Hence, we collected 144 requesters' profiles and 4,012 workers' profiles from Microworkers.com – where all campaigns in our sample data are crowdturfing campaigns and other researchers have found that 89% of campaigns hosted at Microworkers.com are indeed crowdturfing (Wang et al. 2012).

Worker Characteristics. First, we analyzed the workers' profile information consisting of the country, account longevity, number of tasks done and profit (how much they have earned).

We found that the workers are from 75 countries. Figure 2(a) shows the top-10 countries which have the highest portion of workers. 83% of the workers are from these countries. An interesting observation is that a major portion of the workers in Microworkers.com are from Bangladesh – where 38% workers (1,539 workers) come from – whereas in Amazon Mechanical Turk over 90% workers are from the United States and India (Ross et al. 2010).

The 4,012 workers have completed 2,962,897 tasks and earned $467,453 so far, which suggests the entirety of the crowdturfing market is substantial. Interestingly, the average price per task is higher on a crowdturfing site (for Microworkers.com, the average is $0.51) than on the legitimate Amazon Mechanical Turk where 90 percent of all tasks pay less than $0.10 (Ipeirotis 2010).

Table 2: Characteristics of Crowdturfing Requesters.

	# of campaigns	# of paid tasks	Longevity (day)
Max	4,137	455,994	1,091
Avg	68	7,030	329
Median	7	306	259
Min	1	0	3

Table 1 presents the maximum, average, median and minimum number of tasks done, how much they have earned, and the account longevity for the sampled workers. We observe that there are professional workers who have earned reasonable money from the site to survive. For example, a user who earned $3,699 for slightly more than 3 years (1,215 days) lives in Bangladesh where the GNI (Gross National Income) per capita is $770 in 2011 as estimated by the World Bank (TradingEconomics 2011). Surprisingly, she has earned even more money per year ($1,120) than the average income per year ($770) of a person in Bangladesh.

Requester Characteristics. Next, we examine the characteristics of those who post the crowdturfing jobs.

We found that requesters are from 31 countries. Figure 2(b) shows the top-10 countries which have the highest portion of requesters. Interestingly, 55% of all requesters are from the United States, and 70% of all requesters are from the English-speaking countries: United States, UK, Canada, and Australia. We can see an imbalance between the country of origin of requesters and of the workers, but that the ultimate goal is to propagate artificial content through the English-speaking web.

The requesters' profile information reveals their account longevity, number of paid tasks and expense/cost for campaigns. As shown in table 2, many workers have created multiple campaigns with lots of tasks (on an average – 68 campaigns and 7,030 paid tasks). The most active requester in our dataset initiated 4,137 campaigns associated with 455,994 paid tasks. In other words, he has spent a quarter million dollar ($232,557) – again a task costs $0.51 on an average. In total, 144 requesters have created 9720 campaigns with 1,012,333 tasks and have paid a half million dollars ($516,289). This sample analysis shows us how the dark market is big enough to tempt users from the developing countries to become workers.

Down the Rabbit Hole: Linking Crowdsourcing Workers to Social Media

So far, we have seen that most crowdturfers target social media and that the crowdturfing economy is significant: with hundreds of thousands of tasks and dollars supporting it, based on just a fairly small sample. We now propose a framework for beginning a more in-depth study of the ecosystem of crowdturfing by linking crowdsourcing workers to social media. Specifically, we focus on Twitter-related campaigns and their workers. Of the social media targets of interest by crowdturfers, Twitter has the advantage of being open for sampling (in contrast to Facebook and others). Our

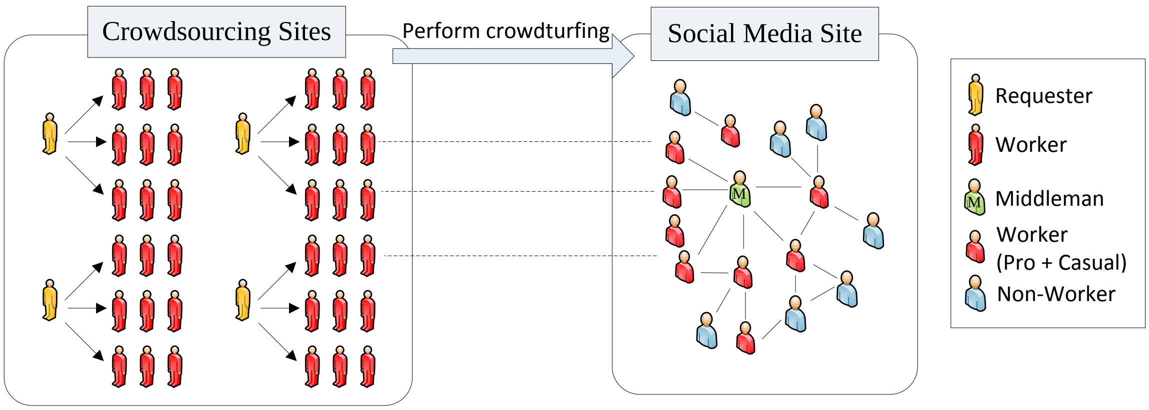


Figure 3: Linking crowdsourcing workers to social media.

Table 3: Twitter dataset.

Class	|User Profiles|	|Tweets|
Workers	2,864	364,581
Non-Workers	9,878	1,878,434

goal is to better understand the behavior of Twitter workers, how they are organized, and to find identifying characteristics so that we may potentially find workers "in the wild".

Following Crowd Workers onto Twitter

Based on our sample of 505 campaigns, we found that 65 specifically targeted Twitter. Of these, there were two types:

- **Tweeting about a link:** These tasks ask the Twitter workers to post a tweet including a specific URL (as in the example in Figure 1). The objective is to spread a URL to other Twitter users, and thereby increase the number of clicks on the URL.
- **Following a twitter user:** The second task type requires a Twitter worker to follow a requester's Twitter account. These campaigns can increase the visibility of the requester's account (for targeting larger future audiences) as well as impacting link analysis algorithms (like PageRank and HITS) used in Twitter search or in general Web search engines that incorporate linkage relationships in social media.

Next we identified the Twitter accounts associated with these workers (see the overall framework in Figure 3). For campaigns of the first type, we used the Twitter search API to find all Twitter users who had posted the URL. For campaigns of the second type, we identified all users who had followed the requester's Twitter account. In total, we identified 2,864 Twitter workers. For these workers, we additionally collected their Twitter profile information, most recent 200 tweets, and social relationships (followings and followers). The majority of the identified Twitter workers participated in multiple campaigns; we assume that the probability that they tweeted a requester's URL or followed a requester's account by chance is very low.

In order to compare how these workers' properties are different from non-workers, we randomly sampled 10,000 Twitter users. Since we have no guarantees that these sampled users are indeed non-workers, we monitored the accounts for one month to see if they were still active and not suspended by Twitter. After one month, we found that 9,878 users were still active. In addition, we randomly selected 200 users out of the 9,878 users and manually checked their profiles, and found that only 6 out of 200 users seemed suspicious. Based on these verifications, we labeled the 9,878 users as non-workers. Even though there is a chance of a false positive in the non-worker set, the results of any analysis should give us at worst a lower bound since the introduction of possible noise would only degrade our results.

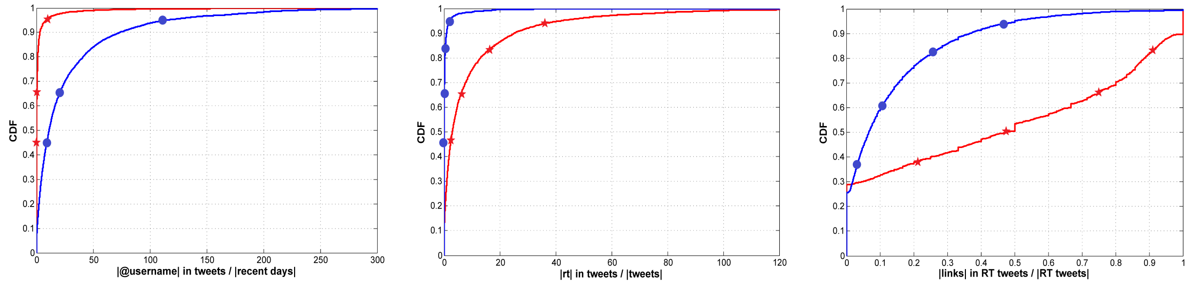
The basic property information of the workers and non-workers are shown in Table 3. In total, we collected 2,864 twitter workers' profile information, their 364,581 messages and their social relationships, and 9,878 twitter non-workers' profile information, their 1,878,434 messages and their social relationships.

Analysis of Twitter Workers: By Profile, Activity, and Linguistic Characteristics

In this section we conduct a deeper analysis regarding the Twitter workers and non-workers based on their profile information, activity within Twitter, and linguistic information revealed in their tweets. Are workers on Twitter fundamentally different from regular users? And if so, in what ways? Note that our analysis that follows considers the entirety of the characteristics of these workers and not just the messages associated with crowdturfing campaigns.

First, we compare profile information of workers and non-workers, especially focusing on the number of following, the number of followers, and the number of tweets. In Table 4 and 5, we can clearly observe that the average number of followings and followers of the workers are much larger than non-workers, but the average number of tweets of the workers is smaller than non-workers. Interestingly, workers are well connected with other users, and potentially their manipulated messages will be exposed to many users.

Next, we study how workers' activity-based characteristics differ from non-workers. We analyzed many activity-based features, including the average number of links per tweet, the average number of hashtags per tweet, and the average number of @username per tweet. In Figure 4, we



(a) |@username| in tweets / |recent days| (b) |rt| in tweets / |tweets| (c) |links| in RT tweets / |RT tweets|

Figure 4: Three activity-based characteristics of workers (red line with stars) and non-workers (blue line with circles). Workers tend to mention few other users, but retweet more often, and include links more often than non-workers.

Table 4: Properties of workers.

	|Followings|	|Followers|	|Tweets|
Min.	0	0	0
Max.	300,385	751,382	189,300
Avg.	5,519	6,649	2,667
Median	429	213	194

Table 5: Properties of non-workers.

	|Followings|	|Followers|	|Tweets|
Min.	0	0	0
Max.	50,496	1,097,911	655,556
Avg.	511	1,000	10,128
Median	244	231	4,018

report the cumulative distribution function for three clearly distinct activity-based characteristics: the average number of @usernames per day during the recent days (in this case, the past month), the average number of retweet message per tweet, and the average number of links per retweet message.

We can clearly observe that workers rarely communicate with other users via @username while non-workers are often communicating with other users. This distinctive behavior makes sense because workers mostly post a message containing a meme or a URL instead of personally talking to another user. However, workers often retweet messages so that these messages may reach their followers and include links more often than non-workers.

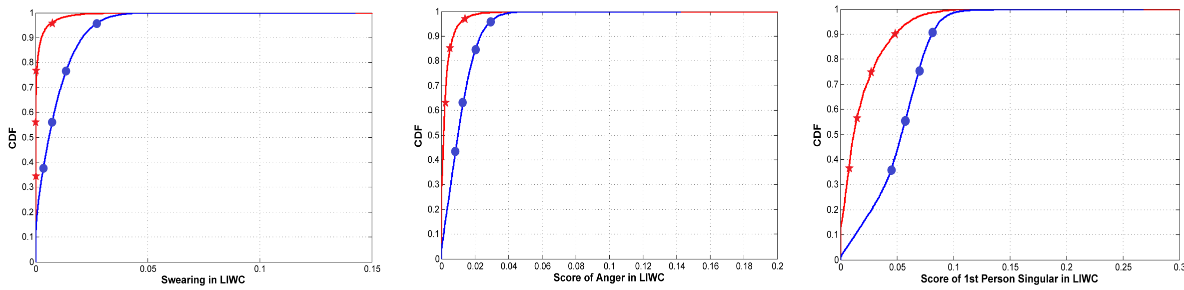
Next, we study the linguistic characteristics of the tweets posted by workers and non-workers. Do workers engage in different language use? To answer this question, we used the Linguistic Inquiry and Word Count (LIWC) dictionary, which is a standard approach for mapping text to psychologically-meaningful categories (Pennebaker, Francis, and Booth 2001). LIWC-2001 defines 68 different categories, each of which contains several dozens to hundreds of words. Given each user's tweets, we measured his linguistic characteristics in the 68 categories by computing his score of each category based on LIWC dictionary. First we counted the total number of words in his tweets (N). Next we counted the number of words in his tweets overlapped with the words in each category i on LIWC dictionary (C_i). Then, we computed his score of a category i as C_i/N. In Figure 5, we show the cumulative distribution functions for three of the most distinguishing linguistic characteristics: Swearing, Anger, and Use of 1st Person Singular. Interestingly, we see that workers tend to swear less, use anger less (e.g., they don't use words like "hate" or "pissed"), and use the 1st-person singular less than non-workers. That is, this linguistic analysis shows that workers are less personal in the messages they post than are non-workers. On one hand, this seems reasonable since workers intend to spread pre-defined manipulated content and URLs (and hence worker tweets should not focus on themselves). However, recall that our data collection includes for each worker all of their recent tweets and not just their crowdturfing-related tweets; so this result may be surprising that the entirety of a worker's tweets show such a clear linguistic division from non-workers.

Network Structure of Twitter Workers

We next explore the network structure of workers by considering the social network topology of their Twitter accounts. What does this network look like? Are workers connected? More generally, can we uncover the implicit power structure of crowdturfers?

A Close-Knit Network? We first analyzed the Twitter workers' following-follower relationship to check whether they were connected to each other. Figure 6 depicts the induced network structure, where a node represents a worker and an edge between two nodes represents that at least one of workers is following the other (in some cases both of them follow each other). Surprisingly, we observed that some workers are densely connected to each other, forming a closely knit network. We measured the graph density of the workers as $\frac{|E|}{|V| \times |V-1|}$ to compare whether these workers form a denser network than the average graph density of users in Twitter. Confirming what visual observation of the network indicates, we found that the workers' graph density was 0.0039 while Yang et al. (Yang et al. 2012) found the average graph density of users on Twitter to be 0.000000845, many orders of magnitude less dense.

Hubs and Authorities. We next examine who in the network is significant. Concretely, we adopted the well-known HITS (Kleinberg 1999) algorithm to identify the hubs (workers who follow many other workers) and authorities (workers who are followed by many other workers) of the network:



(a) Swearing in LIWC (b) Anger in LIWC (c) 1st Person Singular

Figure 5: Three linguistic characteristics of workers (red line with stars) and non-workers (blue line with circles). Workers tend to swear less, use anger less, and use the 1st-person singular less than non-workers.

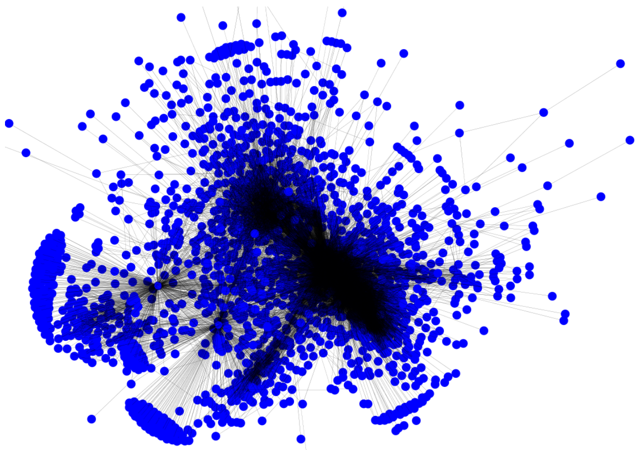


Figure 6: Network structure of all workers.

$$\vec{a} \leftarrow A^T \vec{h}$$
$$\vec{h} \leftarrow A \vec{a}$$

where $\vec{h}$ and $\vec{a}$ denote the vectors of all hub and all authority scores, respectively. A is a square matrix with one row and one column for each worker (user) in the worker graph. If there is an edge between worker i and worker j, the entry A_{ij} is 1 and otherwise 0. We iterate the computation of $\vec{h}$ and $\vec{a}$ until both $\vec{h}$ and $\vec{a}$ are converged. We initialized each worker's hub and authority scores as *1/n* – where *n* is the number of workers in the graph – and then computed HITS until the scores converged.

Table 6 and 7 present the top-10 hubs and top-10 authorities in the workers' following-follower graph. Interestingly, most top-10 hubs are top-10 authorities. It means that these workers are very well connected with other workers, following them and followed by them. The top hub and top authority is *NannyDotNet*, a user who has been both a requester of crowdturfing jobs and a worker on the jobs of others. The other nine workers have a large number of followings and followers. This behavior is similar with "social capitalists", who are eager to follow other users and increase a number of followers as noted in (Ghosh et al. 2012). Even *_Woman_health*'s profile description shows "Always follow back within 24 hours", indicating her intention increasing a number of followers. Interestingly, their Twitter profiles are fully filled, sharing what they are working for or why they are using Twitter, location information and a profile photo.

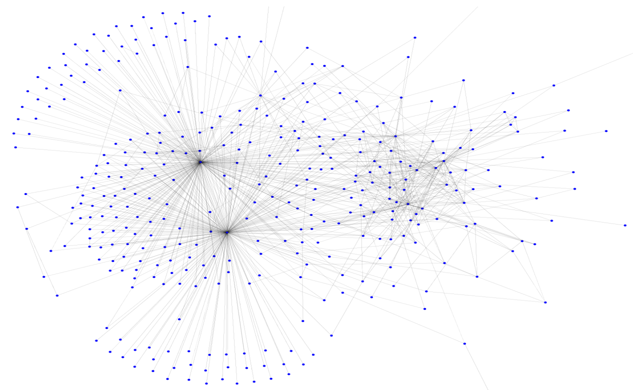


Figure 7: Network structure of professional workers.

Professional Workers. In our examination of workers, we noticed that some workers engaged in many jobs, while others only participated in one or two. We call these workers who occasionally participate "casual workers", while we refer to workers in three or more campaigns as "professional workers". Since these professional workers often worked for multiple campaigns, understanding their behaviors is important to discern the characteristics of the quasi-permanent crowdturfing workforce.

Of the 2,864 workers in total, there were 187 professional workers who participated in at least 3 Twitter campaigns in our collection. Figure 7 depicts their network structure. We can clearly observe that these professional workers are also densely connected. Surprisingly, their graph density is 0.028 which is even higher than all workers' graph density (0.0039).

So far, we only looked at the relationship between these professional workers in their following-follower relationship (i.e., the restricted graph). Next we extend the following-follower relationship to all users (i.e., the open graph including all followings and followers of these professional workers). Table 8 and 9 present top-10 followings and followers of these professional workers, respectively.

One observation from Table 8 is that these professional

Screen Name	|Followings|	|Followers|	|Tweets|
NannyDotNet	1,311	753	332
_Woman_health	210,465	207,589	33,976
Jet739	290,624	290,001	22,079
CollChris	300,385	300,656	8,867
familyfocusblog	40,254	39,810	22,094
tinastullracing	171,813	184,039	73,004
drhenslin	98,388	100,547	10,528
moneyartist	257,773	264,724	1,689
pragmaticmom	30,832	41,418	21,843
Dede_Watson	37,397	36,833	47,105

Table 6: Top-10 hubs of the workers.

Screen Name	|Followings|	|Followers|	|Tweets|
NannyDotNet	1,311	753	332
_Woman_health	210,465	207,589	33,976
CollChris	300,385	300,656	8,867
familyfocusblog	40,254	39,810	22,094
tinastullracing	171,813	184,039	73,004
pragmaticmom	30,832	41,418	21,843
Jet739	290,624	290,001	22,079
moneyartist	257,773	264,724	1,689
drhenslin	98,388	100,547	10,528
ceebee308	283,301	296,857	169,061

Table 7: Top-10 authorities of the workers.

Screen Name	Fre.	|Followings|	|Followers|	|Tweets|
Alexambroz	52	53	51,463	307
Talenthouse	51	14,369	99,880	13,356
Thestoryvine	47	230	127	97
Nettime	42	307	401	1,305
0boy	41	847	108,929	10,827
TheRealAliLee	38	323	9,060	1,509
consumeraware	37	10	845	60
WebsiteBuilderr	36	509	235	81
Ijsfondue	35	100	900	98
ProveeduriaT	33	87	171	312

Table 8: Top-10 followings of the professional workers.

Screen Name	Fre.	|Followings|	|Followers|	|Tweets|
TrueTobacco	29	1,893	866	150
Honest_Solution	28	9,759	14,620	440
Choroibati	27	1,676	567	77
Mostafizurrr	26	34,229	36,809	1,612
YourSjop	24	3,610	3,236	6
SunnieBrees	23	89	56	7
TeamHustleBunny	21	88,331	99,038	9,129
Tarek0593	21	1,055	546	2,302
TinyGems	21	112,417	102,181	8,704
Checkdent	20	2,923	4,002	334

Table 9: Top-10 followers of the professional workers.

workers commonly retweeted messages generated by the two users named *Alexambroz* and *0boy*. We conjecture that these users are *middlemen* who create messages to promote a website or a meme. Since the professional workers follow the middlemen, they will receive the messages automatically (if a user A follows a user B, A will receive B's postings/messages automatically in Twitter) and retweet them to their followers so that the messages are exposed to these workers' followers. The middlemen and professional workers strategically use Twitter as a tool to effectively propagate targeted messages. In another perspective, since these professional workers follow the middlemen and retweet the middlemen's messages, the middlemen get higher rank in a link analysis method such as PageRank (Brin and Page 1998) and HITS (Kleinberg 1999). As the result, the middlemen's profiles and messages will be ranked in the top position in a search result returned by a search engine like Google. These middlemen and professional workers are difficult to detect as evidences by the long lifespan of their accounts, compared with traditional Twitter spammers. For example, while middlemen's average lifespan and professional workers' average lifespan in our dataset are 1,067 days and 614 days, respectively (which are similar or even longer than regular users), twitter spammers' average lifespan is 279 days (Lee, Eoff, and Caverlee 2011).

Table 9 shows the top-10 followers of the professional workers. *Honest_Solution*, *Choroibati*, *Mostafizurrr* and *Tarek0593* followed many professional workers and they are also professional workers in our dataset, demonstrating that these professional workers are connected to each other. Why? We conjecture that they can increase their number of followers and pretend to be legitimate users. Another reason is that some crowdturfing tasks require a minimum number of followers to become eligible workers for certain tasks (e.g., at least larger than 100 followers) because these requesters want their URLs to reach more users. In addition, these professional workers are followed by some business accounts and random user accounts who will be potential victims.

Digging Deeper: Middlemen

We have seen that workers (and especially, professional workers) often retweet middlemen's messages so that more people including the followers of the workers were exposed to the messages. This observation naturally led us to study how to reveal middlemen. First, we investigated the messages of 187 professional workers and extracted retweeted messages containing a URL because the intention of middlemen and professional workers for spreading messages is not only to share the message, but also to tempt the message recipient (e.g., a follower of a professional worker) to visit a web page of the URL. Second, we counted how many professional workers retweeted each one of the extracted retweeted messages. Third, we sorted the extracted retweeted messages by descending order of the number of frequencies. Then, extracted an origin, who is a user (a potential middleman) creating and posting the original message of a retweeted message, from each retweeted message. Our hunch is that the more professional workers retweeted an origin's message, the higher probability to become a middleman the origin has because these professional workers make profit by posting or retweeting an astroturfing (artificial) meme or message, so if many professional workers retweeted an origin's message, he will be a middleman.

By using our approach, we found 575 potential middlemen, and one or two professional workers retweeted messages of 486 out of 575 potential middlemen. Because sometimes a few professional workers retweet the same user's message by chance, we considered the potential middlemen, whose messages are retweeted by at least 10 professional workers, the middlemen. Then there were 41 middlemen. Table 10 shows the top-10 middlemen whose messages are retweeted the most by the professional workers.

Table 10: Top-10 middlemen with the number of professional workers.

Middleman	|Pro-Workers|	|Followings|	|Followers|
0boy	139	847	108,929
louiebaur	95	285	68,772
hasai	63	6,360	41,587
soshable	57	956	22,676
virtualmember	56	5,618	5,625
scarlettmadi	55	5,344	26,439
SocialPros	54	10,775	22,985
cqlivingston	54	6,377	28,556
huntergreene	49	27,390	25,207
TKCarsitesInc	48	1,015	18,661

Interestingly, the top-10 middlemen have a large number of followers (5,625 $\sim$ 108,929), and most of the middlemen disclosed they are interested in social media strategy, social marketing and SEO on their profiles. *hasai* and *SocialPros* have the same homepage on their profiles which is http://hasai.com/ advertising social media marketing. Several middlemen opened their location as Orange County, CA. Some of these middlemen also often retweeted other middlemen's messages. These observations led us to conclude that some of these middlemen accounts are connected or controlled by the same people/organization.

Next, we measured which messages are most retweeted by professional workers, and 10 most retweeted messages are shown in Table 11. Nine messages except the first one have promoting and advertising flavors. We conjecture that sometimes middlemen post regular messages like the first message in the table and professional workers retweet them so that they can pretend a regular user, avoid spam detection from Twitter Safety team and be alive longer on Twitter.

In summary, by finding professional workers we can potentially find middlemen, and by finding the most retweeted messages from professional workers we can potentially find hidden workers who retweeted the messages many times.

Table 11: 10 most retweeted messages by professional workers.

Freq.	Message
29	RT @alexambroz: RT @Twitter for our beautiful planet #earth with #peace and #happiness to all people. One retweet could change Our world.
23	RT @viawomen: Check out the great pregnancy info on http://t.co/5NiVbh6v. Love the celeb parenting blog posts! #pregnancy #pregnancysy ...
22	RT @BidsanityDeals: Bid now! Auctions are ending. Get DVDs, gift cards, jewelry, IPad accessories, books, handmade goods, and much more! ...
20	RT @ik8sqi: Family Tracker free iPhone app lets you track friends and family, even shows past GPS locations as breadcrumbs on map http:/ ...
17	RT @0boy: Here's an interesting marketing idea with lots of #flavor http://t.co/EPl24WZ2 #BucaVIP
17	RT @JeremyReis: 7 Reasons to Sponsor a Child - http://t.co/weg0Tq0y #childsponsorship @food4thehungry
17	RT @louiebaur: StumbleUpon Paid Discovery Is Getting Massive http://t.co/OvYJv2ne via @0boy
16	RT @evaporizing: #ECigarette Save EXTRA @v2cigs today http://t.co/BNbJl1cX use V2 Cigs Coupon EVAPE15 or EVAPE10 plus 20% Memorial Day S ...
16	RT @evaporizing: The Best #ECIGARETTE Deal Of The YEAR Today Only @V2Cigs 4th July Sale + Coupon EVAPE15 40% OFF http://t.co/yrrhTYDy
15	RT @DoYouNeedaJob: Internet Millionaire Looking For Students! Give me 30 days & I will mold you into my next success story!...Visit http ...

Detecting Crowd Workers

So far we have identified three classes of crowdturfers – professional workers, casual workers, and middlemen – and their relationship structure connecting these workers in social media. Next, we turn to study features which have distinguishing power between workers (including professional workers and casual workers) and non-workers. Our goal is to validate that it is possible to detect crowd workers from Twitter "in the wild", with no knowledge of the original crowdturfing task posted on a crowdsourcing site. Since many crowdturfing campaigns are hidden from us (as in the case of campaigns organized through off-network communication channels), it is important to understand the potential of learning models from known campaigns to detect these unknown campaigns.

Detection Approach and Metrics

To profile workers on Twitter, we follow a classification framework where the goal is to predict whether a candidate twitter user u is a worker or a non-worker. To build a classifier c

$$c : u \rightarrow \{worker, non-worker\}$$

we used the Weka machine learning toolkit (Witten and Frank 2005) to test 30 classification algorithms, such as naive bayes, logistic regression, support vector machine (SVM) and tree-based algorithms, all with default values for all parameters using 10-fold cross-validation. 10-fold cross-validation involves dividing the original sample (data) into 10 equally-sized sub-samples, and performing 10 training and validation steps. In each step, 9 sub-samples are used as the training set and the remaining sub-sample is used as the validation set. Each sub-sample is used as the validation set once. For training, we relied on the dataset of 2,864 workers and 9,878 non-workers in Table 3.

To measure the effectiveness of a classifier, we compute precision, recall, F-measure, accuracy, area under the ROC curve (AUC), false positive rate (FPR) and false negative rate (FNR) as metrics to evaluate our classifier.

Table 12: Features.

Group	Feature
UD	the length of the screen name
UD	the length of description
UD	the longevity of the account
UD	has description in profile
UD	has URL in profile
UFN	the number of followings
UFN	the number of followers
UFN	the ratio of the number of followings and followers
UFN	the percentage of bidirectional friends: $\frac{\|followings \cap followers\|}{\|followings\|}$ and $\frac{\|followings \cap followers\|}{\|followers\|}$
UA	the number of posted tweets
UA	the number of posted tweets per day
UA	|links| in tweets / |tweets|
UA	|hashtags| in tweets / |tweets|
UA	|@username| in tweets / |tweets|
UA	|rt| in tweets / |tweets|
UA	|tweets| / |recent days|
UA	|links| in tweets / |recent days|
UA	|hashtags| in tweets / |recent days|
UA	|@username | in tweets / |recent days|
UA	|rt| in tweets in tweets / |recent days|
UA	|links| in RT tweets / |RT tweets|
UC	the average content similarity over all pairs of tweets posted: $\frac{\sum similarity(a,b)}{\|\text{set of pairs in tweets}\|}$, where $a, b \in$ set of pairs in tweets
UC	the ZIP compression ratio of posted tweets: $\frac{uncompressed\ size\ of\ tweets}{compressed\ size\ of\ tweets}$
UC	68 LIWC features which are Total Pronouns, 1st Person Singular, 1st Person Plural, 1st Person, 2nd Person, 3rd Person, Negation, Assent, Articles, Prepositions, Numbers, Affect, Positive Emotions, Positive Feelings, Optimism, Negative Emotions, Anxiety, Anger, Sadness, Cognitive Processes, Causation, Insight, Discrepancy, Inhibition, Tentative, Certainty, Sensory Processes, Seeing, Hearing, Touch, Social Processes, Communication, Other References to People, Friends, Family, Humans, Time, Past Tense Verb, Present Tense Verb, Future, Space, Up, Down, Inclusive, Exclusive, Motion, Occupation, School, Job/Work, Achievement, Leisure, Home, Sports, TV/Movies, Music, Money, Metaphysical States, Religion, Death, Physical States, Body States, Sexual, Eating, Sleeping, Grooming, Swearing, Nonfluencies, and Fillers

Features

The quality of a classifier is dependent on the discriminative power of the features. Based on our observations, we created a wide variety of features belonging to one of four groups: User Demographics (**UD**): features extracted from descriptive information about a user and his account; User Friendship Networks (**UFN**): features extracted from friendship information such as the number of followings and followers; User Activity (**UA**): features representing posting activities; and User Content (**UC**): features extracted from posted tweets. From the four groups, we generated total 92 features as shown in Table 12.

To determine our proposed features' discriminative power, we computed the χ^2 value (Yang and Pedersen 1997) of each of the features. The larger the χ^2 value is, the higher discriminative power the corresponding feature has. The results showed all features had positive discrimination power, though with different relative strengths. Table 13 shows the top-10 features with the average feature values of workers and non-workers. We see that workers and non-workers have different characteristics across many dimensions.

Table 13: Top-10 features.

Feature	Workers	Non-workers
|links| in tweets / |tweets|	0.696	0.142
|tweets| / |recent days|	4	37
|@username| in tweets / |recent days|	2	28
the number of posted tweets per day	3	21
|rt| in tweets / |tweets|	0.7	9.7
Swearing in LIWC	0.001	0.009
|links| in RT tweets / |RT tweets|	0.589	0.142
Anger in LIWC	0.003	0.012
Total Pronouns in LIWC	0.054	0.107
1st Person Singular in LIWC	0.019	0.051

Table 14: Worker detection: Results.

Classifier	Accuracy	F_1	AUC	FPR	FNR
Random Forest	93.26%	0.966	0.955	0.036	0.174

Detection Results

Using the classification setup described above and these feature groups, we tested 30 classification algorithms using the Weka machine learning toolkit (Witten and Frank 2005). To test which classification algorithm returns the highest accuracy, we ran over 30 classification algorithms such as Naive Bayes, Logistic Regression and SMO (SVM) with the default setting. Their accuracies ranges from 86% to 91%. Tree-based classifiers showed the highest accuracy results. In particular, Random Forest produced the highest accuracy which was 91.85%. By changing input parameter values of Random Forest, we achieved 93.26% accuracy and 0.932 F_1 as shown in Table 14.

We additionally considered different training mixtures of workers and non-workers, ranging from 1% worker and 99% non-worker to 99% worker and 1% non-worker. We find that the classification quality is robust across these training mixtures. In other words, our proposed features have good discriminative power between workers and non-workers.

Consistency of Worker Detection over Time

As time passes, a pre-built classifier can lose its classification accuracy because crowdturfing workers may change their behavioral patterns to hide their true identities from the classifier. In order to test whether the classifier built in the previous sub-section is still effective at a later point in time, we created Twitter campaigns a month later in three crowdsourcing sites – Microworkers.com, ShortTask.com and Rapidworkers.com – to collect new workers' Twitter account information consisting of their profile information, tweets and following-follower information. As shown in Table 15, we collected 368 Twitter user profiles and their recent 200 messages (in total, 40,344 messages).

Table 15: A New Worker Dataset.

Class	User Profiles	Tweets
Workers	368	40,344

Next, we evaluated our previously built classifier, with this dataset as the testing set, by measuring how many workers in the set are correctly predicted. Table 16 presents its experimental result. It confirms that our classifier is still effective even with the passage of time with 94.3% accuracy, 0.971 F_1 measure and 0.057 false negative rates.

Table 16: Worker detection over the new dataset using the original detection model.

Classifier	Accuracy	F_1	FNR
Random Forest	94.3%	0.971	0.057

In summary, this positive experimental result shows that our classification approach is promising to find new workers in the future. Our proposed framework linking crowdsourcing workers to social media works effectively. Even though workers may change memes or URLs which they want to spread as the time passes, their behaviors and observable features such as activity patterns and linguistic characters will be similar and are different from regular users.

Conclusion

In this paper, we have presented a framework for "pulling back the curtain" on crowdturfers to reveal their underlying ecosystem. We have analyzed the types of malicious tasks and the properties of requesters and workers on Western crowdsourcing sites. By linking tasks and their workers on crowdsourcing site to social media, we have traced the activities of crowdturfers in social media and have identified three classes of crowdturfers – professional workers, casual workers, and middlemen – and their relationship structure connecting these workers in social media. We have revealed that these workers' profile, activity and linguistic characters are different from regular social media users. Based on these observations, we have proposed and developed statistical user models to automatically differentiate among regular social media users and workers. Our experimental results show that these models can effectively detect previously unknown Twitter-based workers.

Acknowledgements

This work was supported in part by AFOSR Grant FA9550-12-1-0363. Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsor.

References

Alonso, O.; Rose, D. E.; and Stewart, B. 2008. Crowdsourcing for relevance evaluation. *SIGIR Forum* 42(2):9–15.

Bernstein, M. S.; Little, G.; Miller, R. C.; Hartmann, B.; Ackerman, M. S.; Karger, D. R.; Crowell, D.; and Panovich, K. 2010. Soylent: a word processor with a crowd inside. In *UIST*.

Brabham, D. C. 2008. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence: The International Journal of Research into New Media Technologies* 14(1):75–90.

Brin, S., and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. In *WWW*.

Chan, C. 2012. How a fake erotic fiction ebook hit the top 5 of itunes. http://gizmodo.com/5933169/how-a-fake-crowdsourced-erotic-ebook-hit-the-top-5-of-itunes.

Fielding, N., and Cobain, I. 2011. Revealed: US spy operation that manipulates social media. http://www.guardian.co.uk/technology/2011/mar/17/us-spy-operation-social-networks.

Franklin, M. J.; Kossmann, D.; Kraska, T.; Ramesh, S.; and Xin, R. 2011. Crowddb: answering queries with crowdsourcing. In *SIGMOD*.

Ghosh, S.; Viswanath, B.; Kooti, F.; Sharma, N. K.; Korlam, G.; Benevenuto, F.; Ganguly, N.; and Gummadi, K. P. 2012. Understanding and combating link farming in the twitter social network. In *WWW*.

Ipeirotis, P. G. 2010. Analyzing the amazon mechanical turk marketplace. *XRDS* 17(2):16–21.

Kittur, A.; Chi, E. H.; and Suh, B. 2008. Crowdsourcing user studies with mechanical turk. In *CHI*.

Kleinberg, J. M. 1999. Authoritative sources in a hyperlinked environment. *J. ACM* 46(5):604–632.

Lee, K.; Eoff, B. D.; and Caverlee, J. 2011. Seven months with the devils: A long-term study of content polluters on twitter. In *ICWSM*.

Pennebaker, J.; Francis, M.; and Booth, R. 2001. *Linguistic Inquiry and Word Count*. Erlbaum Publishers.

Pham, N. 2013. Vietnam admits deploying bloggers to support government. http://www.bbc.co.uk/news/world-asia-20982985.

Ross, J.; Irani, L.; Silberman, M. S.; Zaldivar, A.; and Tomlinson, B. 2010. Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI Extended Abstracts on Human Factors in Computing Systems*.

Sterling, B. 2010. The chinese online 'water army'. http://www.wired.com/beyond_the_beyond/2010/06/the-chinese-online-water-army/.

TradingEconomics. 2011. Gni per capita; atlas method (us dollar) in bangladesh. http://www.tradingeconomics.com/bangladesh/gni-per-capita-atlas-method-us-dollar-wb-data.html.

Venetis, P., and Garcia-Molina, H. 2012. Quality control for comparison microtasks. In *CrowdKDD 2012*.

Wang, G.; Wilson, C.; Zhao, X.; Zhu, Y.; Mohanlal, M.; Zheng, H.; and Zhao, B. Y. 2012. Serf and turf: crowdturfing for fun and profit. In *WWW*.

Wikipedia. 2013. Internet water army. http://en.wikipedia.org/wiki/Internet_Water_Army.

Witten, I. H., and Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques (Second Edition)*. Morgan Kaufmann.

Xia, T.; Zhang, C.; Xie, J.; and Li, T. 2012. Real-time quality control for crowdsourcing relevance evaluation. In *Network Infrastructure and Digital Content (IC-NIDC)*.

Yang, Y., and Pedersen, J. O. 1997. A comparative study on feature selection in text categorization. In *ICML*.

Yang, C.; Harkreader, R.; Zhang, J.; Shin, S.; and Gu, G. 2012. Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In *WWW*.