COMBATING CROWDSOURCED MANIPULATION OF SOCIAL MEDIA

A Thesis

by

PRITHIVI TAMILARASAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	James Caverlee
Committee Members,	Riccardo Bettati
	Laura Mandell
Head of Department,	Duncan Moore Hank Walker

August 2013

Major Subject: Computer Engineering

ABSTRACT

Crowdsourcing systems - like Ushahidi (for crisis mapping), Foldit (for protein folding) and Duolingo (for foreign language learning and translation) - have shown the effectiveness of intelligently organizing large numbers of people to solve traditionally vexing problems. Unfortunately, new crowdsourcing platforms are emerging to support the coordinated dissemination of spam, misinformation, and propaganda. These "crowdturfing" systems are a sinister counterpart to the enormous positive opportunities of crowdsourcing; they combine the organizational capabilities of crowdsourcing with the ability to widely spread artificial grass root support (so called "astroturfing"). This thesis begins a study of crowdturfing that targets social media and proposes a framework for "pulling back the curtain" on crowdturfers to reveal their underlying ecosystem. Concretely, this thesis (i) analyzes the types of campaigns hosted on multiple crowdsourcing sites; (ii) links campaigns and their workers on crowdsourcing sites to social media; (iii) analyzes the relationship structure connecting these workers, their profile, activity, and linguistic characteristics, in comparison with a random sample of regular social media users; and (iv) proposes and develops statistical user models to automatically identify crowdturfers in social media. Since many crowdturfing campaigns are hidden, it is important to understand the potential of learning models from known campaigns to detect these unknown campaigns. Our experimental results show that the statistical user models built can predict crowdturfers with very high accuracy.

DEDICATION

To my family and friends.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Caverlee for his guidance and support throughout this research and my graduate studies at Texas A&M University. I am very privileged to have been a part of his research group (Infolab).

I would also like to thank my committee members, Dr. Bettati and Dr. Mandell for agreeing to serve on my committee.

I would also like to appreciate the help provided by the members of Infolab - especially Dr. Kyumin Lee for his support during this research.

Finally, thanks to my parents, my brother, Karthik Prabhu and friends for their constant encouragement and support.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

1. INTRODUCTION

1.1 Opportunity

Crowdsourcing systems have successfully leveraged the attention and capacity of millions of “crowdsourced” workers to tackle traditionally vexing problems. There are two entities involved in a crowdsourcing system – requesters (people who have tasks) and workers (people who work on these tasks and get paid). Figure 1 shows the typical organization of a crowdsourcing system.

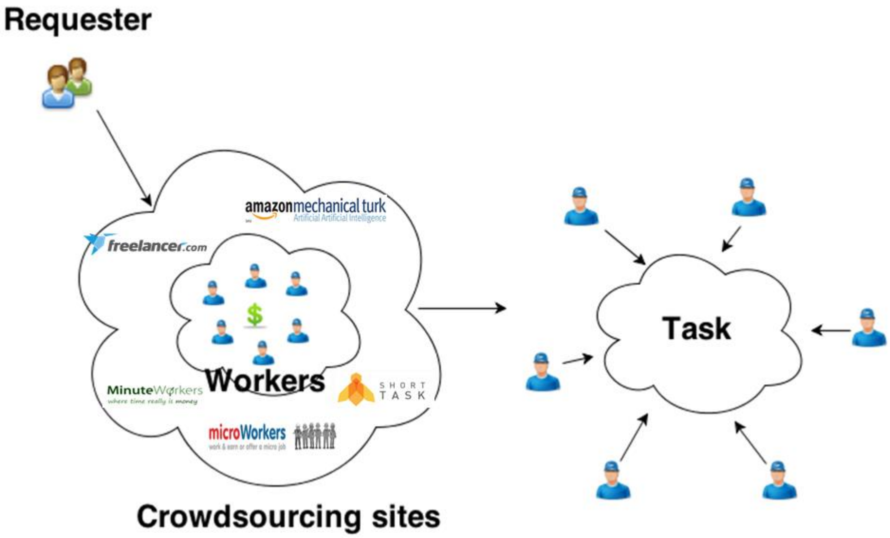


Figure 1: Crowdsourcing

Crowdsourcing systems are used to effectively organize large numbers of people. Based on the incentives in play, the crowdsourcing systems can be grouped into two categories: specialized crowdsourcing systems which are unpaid such as Ushahidi (for crisis mapping), Foldit (for protein folding) and Duolingo (for translation) versus general-purpose crowdsourcing marketplaces that provide monetary gains such as Amazon Mechanical Turk, ShortTask and Crowdflower.

In specialized crowdsourcing systems such as Duolingo – the users translate text from web while learning a language (which is the incentive they receive). In Ushahidi, the users volunteer to do crisis mapping. Such specialized crowdsourcing systems are dedicated for a single and well defined purpose without providing any monetary benefits for the users. These are different from crowdsourcing marketplaces such as Amazon Mechanical Turk and Shorttask where the users get paid for the services they provide. These are not specialized and can be used to get any type of task done by paying the users. Both specialized crowdsourcing systems and crowdsourcing marketplaces have a lot of benefits – a large number of users can be obtained in no time, the users can be used to tackle large and complex problems and a diverse group of users (from around the world) can be got together to solve a given problem.

Our research is aimed at studying the ecosystem of general-purpose crowdsourcing marketplaces, specifically the types of campaigns and the characteristics of users who participate in these campaigns. We also intend to analyze the extent of "crowdturfing" campaigns in these systems. Crowdturfing is a sinister counterpart to the positive opportunities of crowdsourcing marketplaces, 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. Figure 2 presents how crowdturfing works.

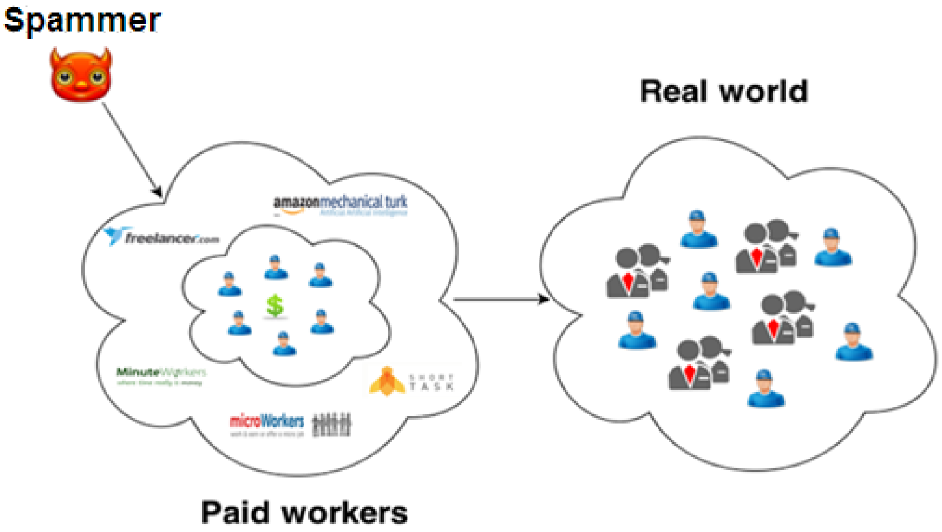
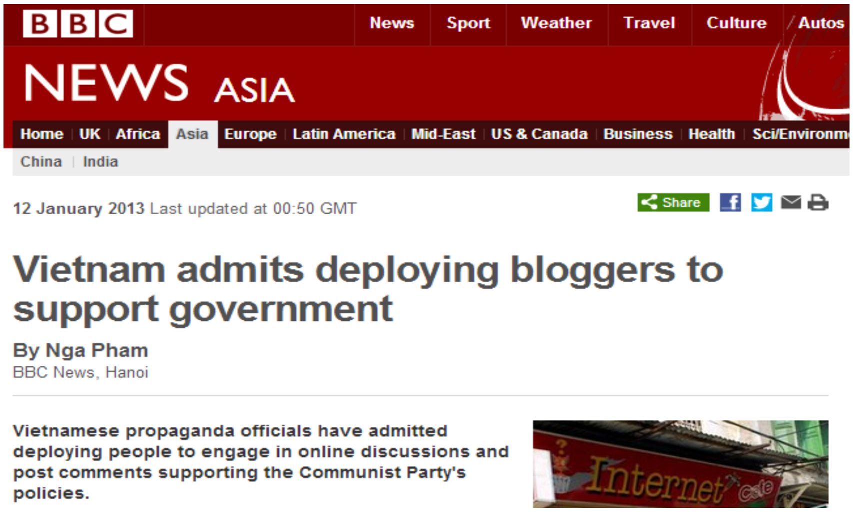


Figure 2 : Crowdturfing

For example, it has been recently reported that Vietnamese propaganda officials (Figure 3) deployed 1,000 crowdturfers to engage in online discussions and post comments supporting the Communist Party's policies [1].



12 January 2013 Last updated at 00:50 GMT

Vietnam admits deploying bloggers to support government

By Nga Pham
BBC News, Hanoi

Vietnamese propaganda officials have admitted deploying people to engage in online discussions and post comments supporting the Communist Party's policies.

Figure 3 : News article about crowdsourced propaganda by Vietnamese officials

Similarly, the Chinese "Internet Water Army" can be hired to post positive comments for the government or commercial products, as well as disparage rivals[2-4]. Mass organized crowdturfers are also targeting popular services like iTunes [5] and attracting the attention of US intelligence operations [6]. 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. [7] found that about 90% of all tasks were for crowdturfing.

1.2 Challenges

There are various challenges involved in identifying crowdturfing campaigns and linking them to their workers in social media. Some of the challenges are listed below,

- Crowdsourcing sites provide very limited information about workers. Hence it is hard to map the workers to social media.
- Many spam campaigns in crowdturfing sites go undetected as they are hidden – i.e., the details of the campaign are directly emailed to the workers who accept to do the given task.
- With the number of crowdsourcing sites on the Internet increasing by the day, it is difficult to keep track of all of them.
- Crowdturfers in social media exhibit a behavior which is entirely different from spam bots. Hence traditional spam detection techniques cannot be applied.
- It is difficult to develop generalized statistical models that are valid over time as workers' behavior tends to change.

1.3 Contributions

In this research 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? 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? Towards answering these questions, we make the following contributions in this research,

- 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 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.

1.4 Related work

The architecture of various crowd-sourcing sites has been studied by various previous researches. Hirth et al. [8] studied the characteristics of workers and employers in Microworkers.com. The user studies conducted by Kittur et al. [9] in Mechanical Turk have shown that a large number of workers can be hired for doing tasks within a short time and with very less cost. Similar studies [10] have shown the potential of crowdsourcing and researchers have begun developing new crowd-based platforms – e.g., [11, 12] – for augmenting traditional information retrieval and database systems, embedding crowds into workflows (like document authoring) [13], and so forth.

Wang et al. [7] coined the term “crowdturfing” (crowd-sourcing + astroturfing) to refer to crowd-sourcing systems where malicious campaigns are hosted by employers. They have studied crowd-sourcing sites based in China and the impact of these sites on one social networking site – Weibo. Chen et al. [3] have done a detailed study on detection of hidden paid posters in Sina.com and Sohu.com – both websites based in China. Ratkiewicz et al. [14] created a system for tracking the spread of astroturfing content in Microblogs with respect to “political astroturf”.

A key issue for open crowd-based systems is the control of the quality of workers and outputs. Venetis and Garcia-Molina [15] 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. [16] provided a real-time quality control strategy for relevance evaluation of search engine results using crowd workers – based on a 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 work is on crowdsourcing sites that deliberately encourage crowdturfing.

2. CROWDTURFING CAMPAIGNS

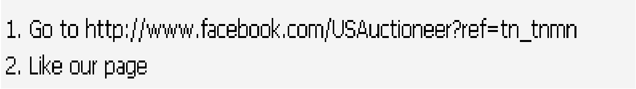
In this section, we begin our study through an examination of the different types of crowdturfing campaigns that are posted on crowdsourcing sites and study the characteristics of both requesters (who post tasks) and workers (who work on the tasks).

We collected 505 campaigns by crawling three popular crowdsourcing sites that host clear examples of crowdturfing campaigns: Microworkers.com, ShortTask.com, and Rapidworkers.com during a span of two 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 prohibit the posting of crowdturfing campaigns. Each of the 505 sampled campaigns has multiple tasks, totaling 63,042 tasks.

2.1 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 five categories. Table 1 shows the split-up of each category.

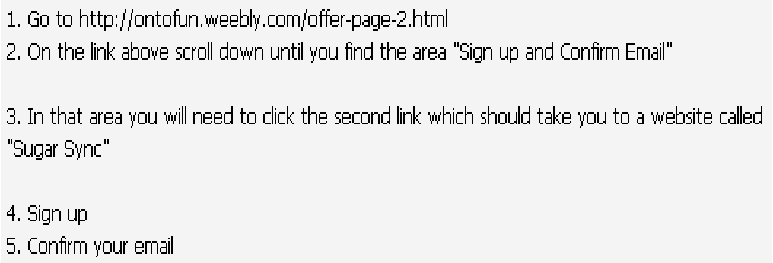
- **Social Media Manipulation** : The most popular type of campaign targets social media. Campaigns request workers to spread a meme through social media sites such as Twitter, “like” a specific Facebook profile/product page, bookmark a webpage on Stumbleupon, answer a question with a link on Yahoo! Answers, write a review for a product at Amazon.com, or write an article on a personal blog. A campaign where workers are required to “like” a Facebook page is shown in Figure 4.



1. Go to http://www.facebook.com/USAuctioneer?ref=tn_tnmn
2. Like our page

Figure 4 : A campaign which asks workers to like a Facebook page

- **Sign Up** : 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. An example of such a campaign is given in Figure 5.



1. Go to http://ontofun.weebly.com/offer-page-2.html
2. On the link above scroll down until you find the area "Sign up and Confirm Email"

3. In that area you will need to click the second link which should take you to a website called "Sugar Sync"

4. Sign up
5. Confirm your email

Figure 5 : A campaign which asks workers to sign-up on a website.

- **Search Engine Spamming** : 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), thereby increasing the rank of that link.

- **Voting** : Requesters ask workers to cast votes. In one example, a 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** : 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.

Table 1: Types of crowdturfing campaigns

Type	#Campaigns
Social media manipulation	171
Sign up	118
Search Engine Spamming	36
Voting	18
Miscellany	162
Total	505

3. CROWDTURFING AND SOCIAL MEDIA

In this section, we present the different campaigns related to two popular social media websites – Twitter and Facebook.

3.1 Facebook crowdturfing campaigns

The crowdturfing tasks targeted towards Facebook are those which ask the user to “like” a given Facebook page or “share” something in Facebook. Since Facebook does not reveal who all liked or shared a page, it was not possible for us to analyze the profiles of Facebook workers. But we could analyze the “like” statistics for these pages. From this analysis, we found that the target Facebook pages get a high degree of attention once a campaign is posted on a crowd sourcing site.

We have presented the snapshots of the “like” statistics of the target Facebook pages, for two crowdturfing campaigns which were posted on Microworkers.com and Shorttask.com in Figure 6 and Figure 7 respectively. For all these pages, the sudden increase in the number of likes corresponds to the first appearance of the crowdturfing campaign asking crowd workers to “like” the Facebook page.

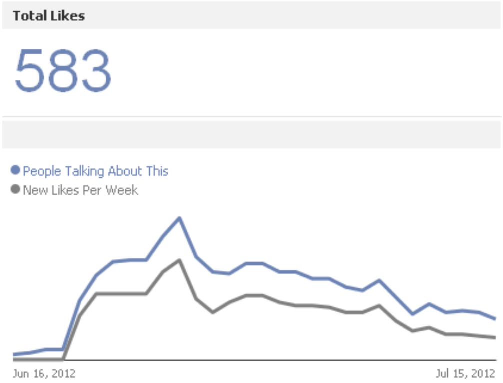


Figure 6 : Facebook statistics for http://www.facebook.com/USAuctioneer/

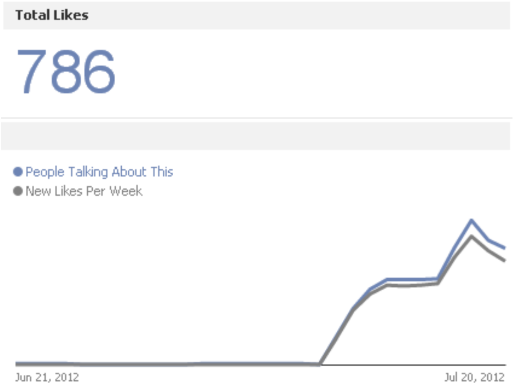
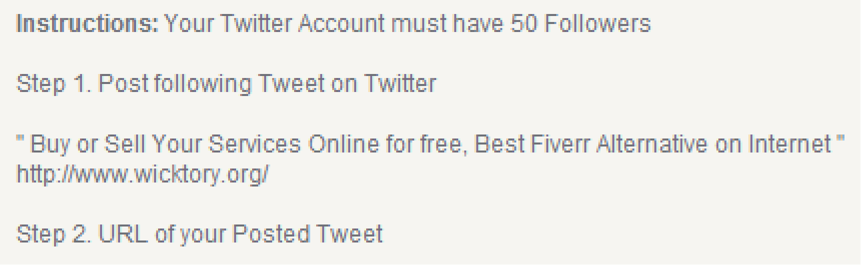


Figure 7 : Facebook statistics for https://www.facebook.com/VirtualMediaMavens

3.2 Twitter crowdturfing campaigns

Twitter crowdturfing campaigns aim to promote targeted content among Twitter users. We identified two types of Twitter crowdturfing campaigns:

- **Tweeting about a link**: These campaigns ask the Twitter workers to post a tweet including a specific URL. The objective is to spread the URL to other Twitter users, and thereby increase the number of clicks on the URL. Figure 8 shows a crowdturfing campaign requesting workers to tweet a URL. The corresponding tweets posted by the workers are shown in Figure 9.



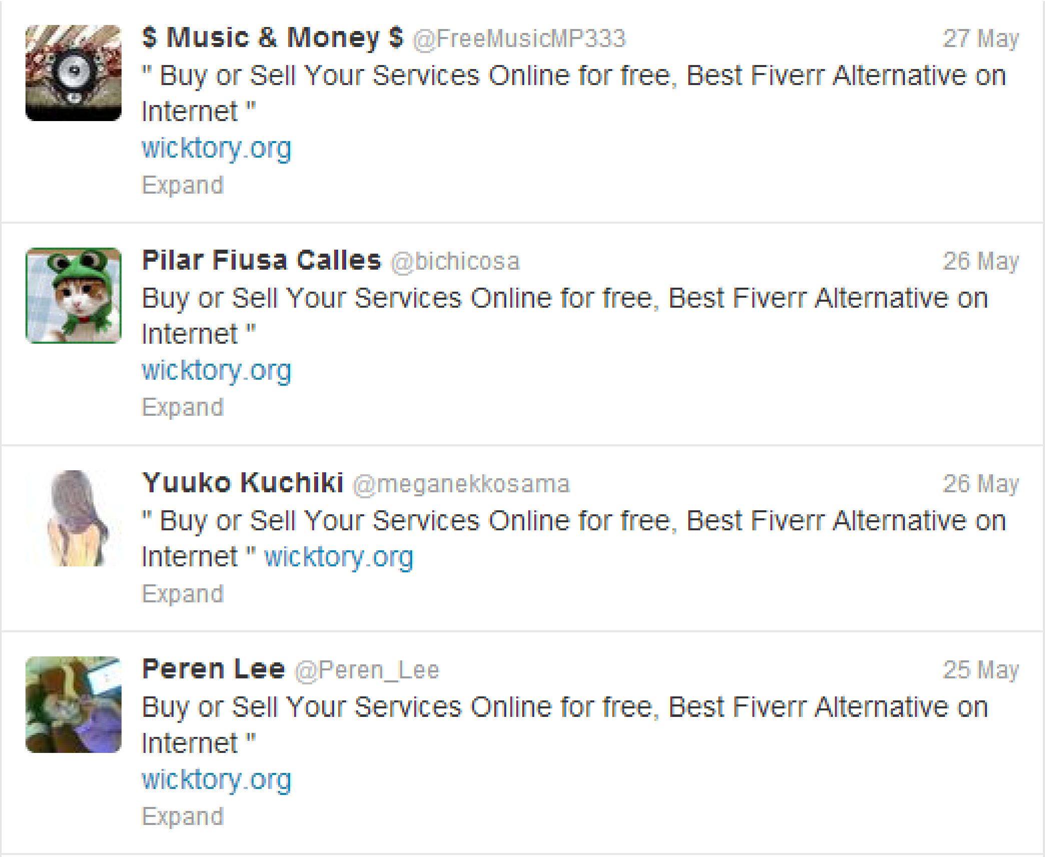
Instructions: Your Twitter Account must have 50 Followers

Step 1. Post following Tweet on Twitter

" Buy or Sell Your Services Online for free, Best Fiverr Alternative on Internet "
http://www.wicktory.org/

Step 2. URL of your Posted Tweet

Figure 8 : Campaign asking workers to tweet a URL



$ Music & Money $ @FreeMusicMP333 27 May
" Buy or Sell Your Services Online for free, Best Fiverr Alternative on Internet "
wicktory.org
Expand

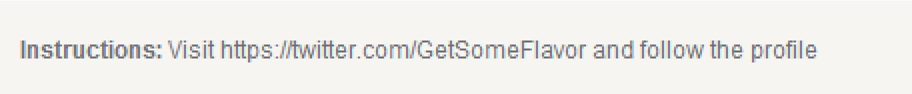
Pilar Fiusa Calles @bichicosa 26 May
Buy or Sell Your Services Online for free, Best Fiverr Alternative on Internet "
wicktory.org
Expand

Yuuko Kuchiki @meganekkosama 26 May
" Buy or Sell Your Services Online for free, Best Fiverr Alternative on Internet " wicktory.org
Expand

Peren Lee @Peren_Lee 25 May
Buy or Sell Your Services Online for free, Best Fiverr Alternative on Internet "
wicktory.org
Expand

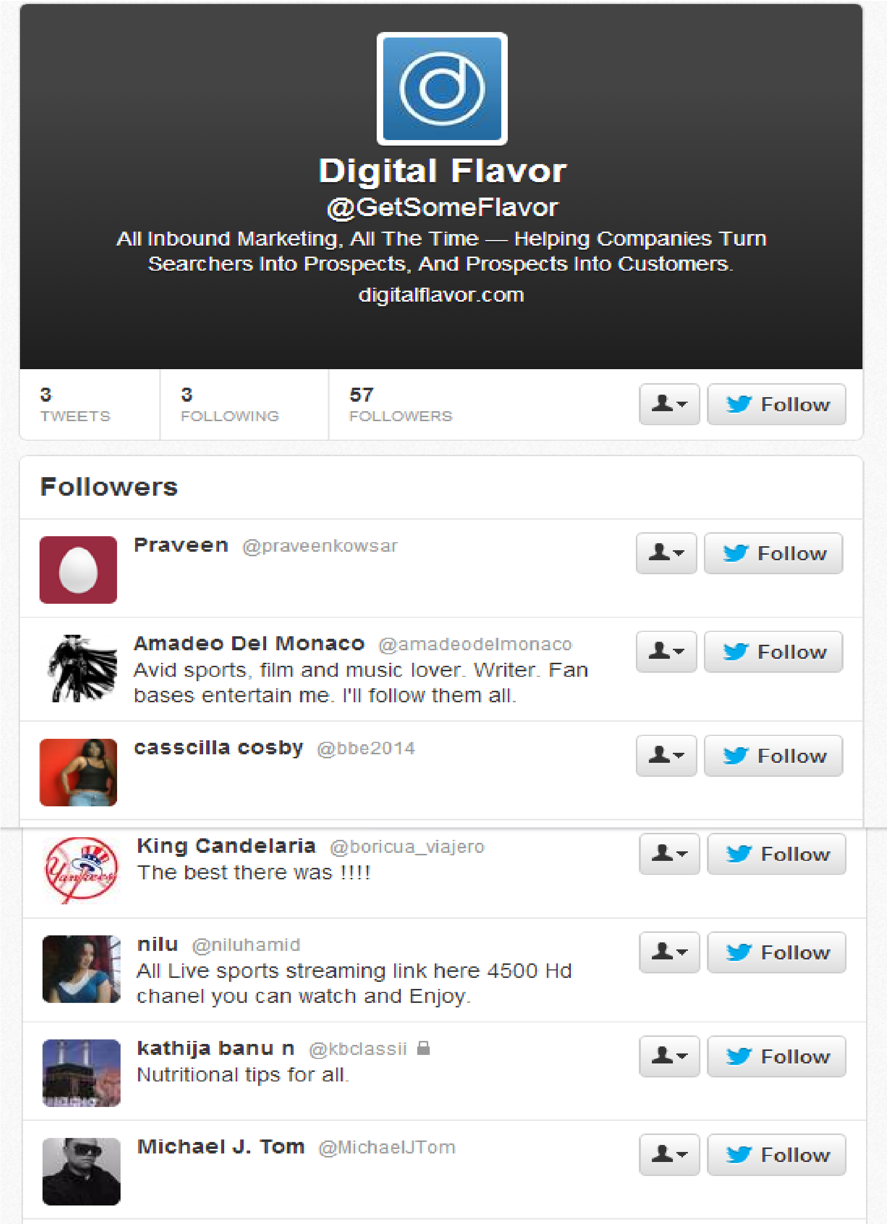
Figure 9 : Tweets posted by workers in response to the campaign of Figure 8

- **Following a twitter user**: The second type of campaign 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. Figure 10 presents a crowdturfing campaign which requests workers to follow a target Twitter profile. The corresponding Twitter profile is shown in Figure 11. As we can see, though the profile has only 3 Tweets, it has 57 followers – most of them being workers.



Instructions: Visit https://twitter.com/GetSomeFlavor and follow the profile

Figure 10 : Campaign requesting workers to follow a Twitter profile



Digital Flavor
@GetSomeFlavor
All Inbound Marketing, All The Time — Helping Companies Turn Searchers Into Prospects, And Prospects Into Customers.
digitalflavor.com

3 TWEETS | 3 FOLLOWING | 57 FOLLOWERS

Followers

Praveen @praveenkowsar

Amadeo Del Monaco @amadeodelmonaco
Avid sports, film and music lover. Writer. Fan bases entertain me. I'll follow them all.

casscilla cosby @bbe2014

King Candelaria @boricua_viajero
The best there was !!!!

nilu @niluhamid
All Live sports streaming link here 4500 Hd chanel you can watch and Enjoy.

kathija banu n @kbclassii
Nutritional tips for all.

Michael J. Tom @MichaelJTom

Figure 11 : Twitter profile followed by workers in response to the campaign in Figure 10

4. LINKING CROWDTURFING WORKERS ONTO TWITTER

We now propose a framework for beginning a more in-depth study of the ecosystem of crowdturfing by linking crowdturfing 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 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”.

4.1 Following crowd workers onto Twitter

As described in the previous section, we identified two types of Twitter campaigns – tweeting a link and following a Twitter profile. 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, their 200 most recent 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' characteristics 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 sampled Twitter 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. Based on this, we labeled the 9,878 users as non-workers. Even though there is a chance of a random worker being 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. A summary of the Twitter dataset is given in Table 2.

Table 2 : Twitter dataset

Class	#Users	#Tweets
Workers	2,864	364,581
Non-workers	9,878	1,878,434

4.2 Basic properties of Twitter workers and non-workers

In this section we present the basic profile information of workers (Table 3) and non-workers (Table 4), especially focusing on the number of followings, the number of followers they have and their total number of tweets.

We can clearly observe that the average number of followings and the average number of followers for the workers are both much larger than the corresponding numbers for non-workers, but the average number of tweets for the workers is smaller than that for non-workers. Interestingly, workers are well connected with other users, and their manipulated messages will potentially be exposed to many users.

Table 3: Properties of workers

	Followings	Followers	Tweets
Min	0	0	0
Max	300,385	51,382	189,300
Avg.	5,519	6,649	2,667
Median	429	213	194

Table 4: 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

4.3 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?

We first analyzed the Twitter workers' relationships to check whether they were connected to each other. Figure 12 depicts the worker network structure, where a node represents a worker and an edge between two nodes represents that at least one of the two 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 (defined as the ratio of number of edges existing in the graph to the total number of possible edges) of the workers as $\frac{|E|}{|V|*(|V|-1|)}$ (where E and V are the number of edges and vertices respectively), 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. [17] found the average graph density of users on Twitter to be 0.000000845, many orders of magnitude less dense.

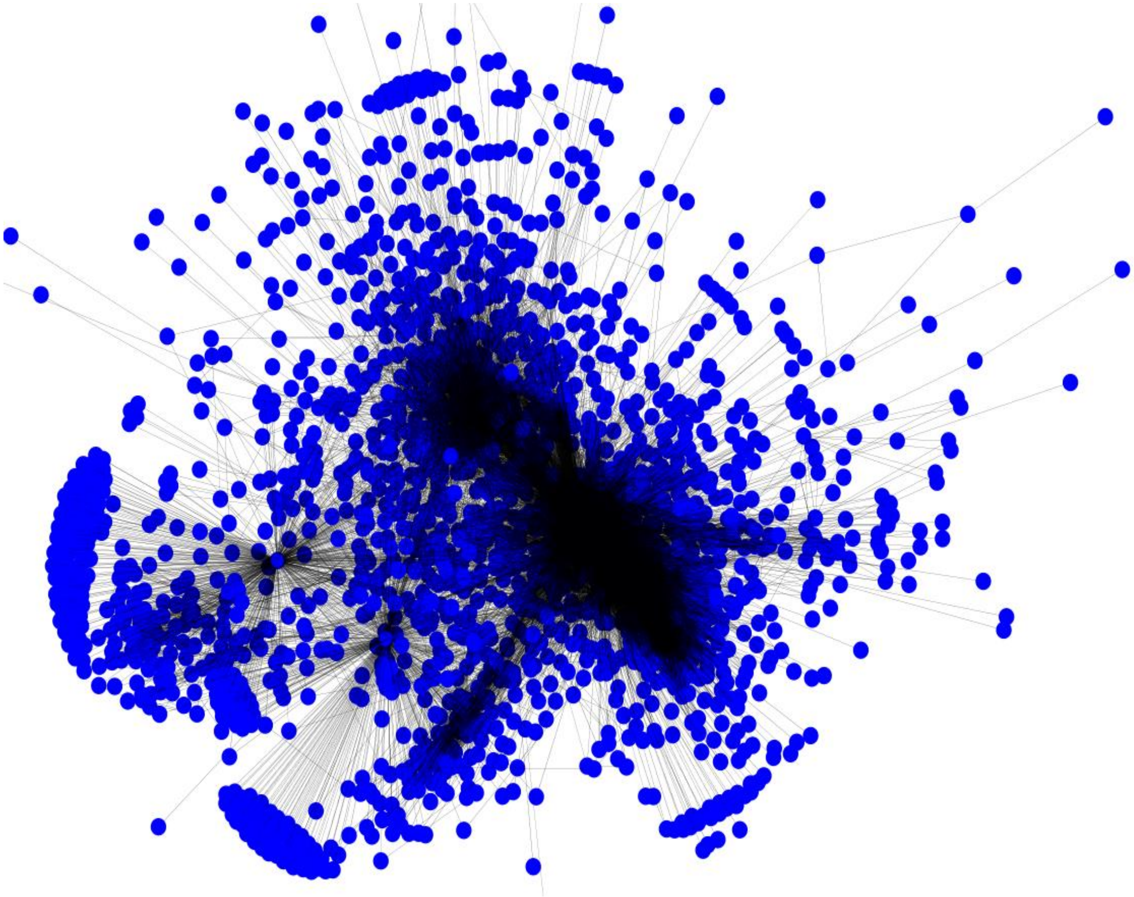


Figure 12 : Network structure of Twitter workers

5. DETECTING CROWD WORKERS

Next, we study the features which help distinguish between 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 such as email), it is important to understand the potential of learning models from known campaigns to detect these unknown campaigns.

5.1 Detection approach

To detect 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. We built the classifier using the WEKA machine learning toolkit and tested 30 classification algorithms such as naive Bayes, logistic regression, support vector machine (SVM) and tree-based algorithms using 10-fold cross-validation. For training, we relied on the dataset of 2,864 workers and 9,878 non-workers.

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).

5.2 Features

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. We created a wide variety of features belonging to one of the 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 number of followers; User Activity (UA) - features representing posting activities; and User Content (UC) - features extracted from posted tweets. From the four groups, we generated a total of 92 features as shown in Table 5.

Table 5 : List of features

Category	Feature
UD	length of the screen name
UD	length of profile description
UD	longevity of the account
UD	has description in profile
UD	has URL in profile
UFN	number of followings
UFN	number of followers
UFN	ratio of the number of followings to number of followers
UFN	percentage of bidirectional friends
UA	Total number of tweets posted
UA	number of posted tweets posted per day
UA	|links| in tweets / |tweets|

Table 5 Continued.

Category	Feature
UA	|hashtags| in tweets / |tweets|
UA	|@<username>| in tweets / |tweets|
UA	|rt| in tweets / |tweets|
UA	|tweets| per day for the 200 most recent tweets
UA	|links| in the 200 most recent tweets
UA	|hashtags| in the 200 most recent tweets
UA	|@username | in the 200 most recent tweets
UA	|rt| in tweets in the 200 most recent tweets
UA	|links| in RT tweets / |RT tweets|
UC	the average content similarity over all pairs of tweets posted
UC	the ZIP compression ratio of posted tweets
UC	LIWC features: 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

Table 5 Continued.

Category	Feature
UC	LIWC Features: 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

5.2.1 User demographics based features

These features take into account factors such as the length of the screen name, the length of profile description and the longevity of the account.

5.2.1.1 Length of profile description

Based on our observations we found that workers had a shorter profile description on Twitter, compared to normal users. Table 6 shows the average number of characters in profile description for workers and non-workers. As we can see, workers tend to have a shorter profile description.

Table 6 : Average number of characters in Twitter profile description

Class	Average length of profile description
Workers	59.5 characters
Non-workers	74.2 characters

5.2.2 User friendship network based features

These features are aimed at studying how well the workers are connected to other users. These features help us understand the information spread from the workers to other users.

5.2.2.1 Ratio of #followings to #followers

This feature is the ratio of the number of followings to the number of followers for a given user. As we see in Table 7, workers tend to follow a higher number of profiles when compared to normal users. This is because, certain tasks on crowdsourcing sites have restrictions as to the minimum number of followers required to participate in the task. In order to get a higher number of followers, the workers tend to randomly follow users hoping that they might follow back.

Table 7 : Ratio of #followings to #followers

	Avg. #Followings	Avg. #Followers	#Followings / #Followers
Workers	5,519	6,649	0.83
Non-workers	511	1,000	0.51

5.2.3 User activity based features

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

5.2.3.1 Number of URLs per tweet

Figure 13 shows the CDF plot of the URLs per tweet for both workers and non-workers. From the plot we can see that the workers tend to include a large number of URLs and one of their objectives is to spread URLs of targeted content.

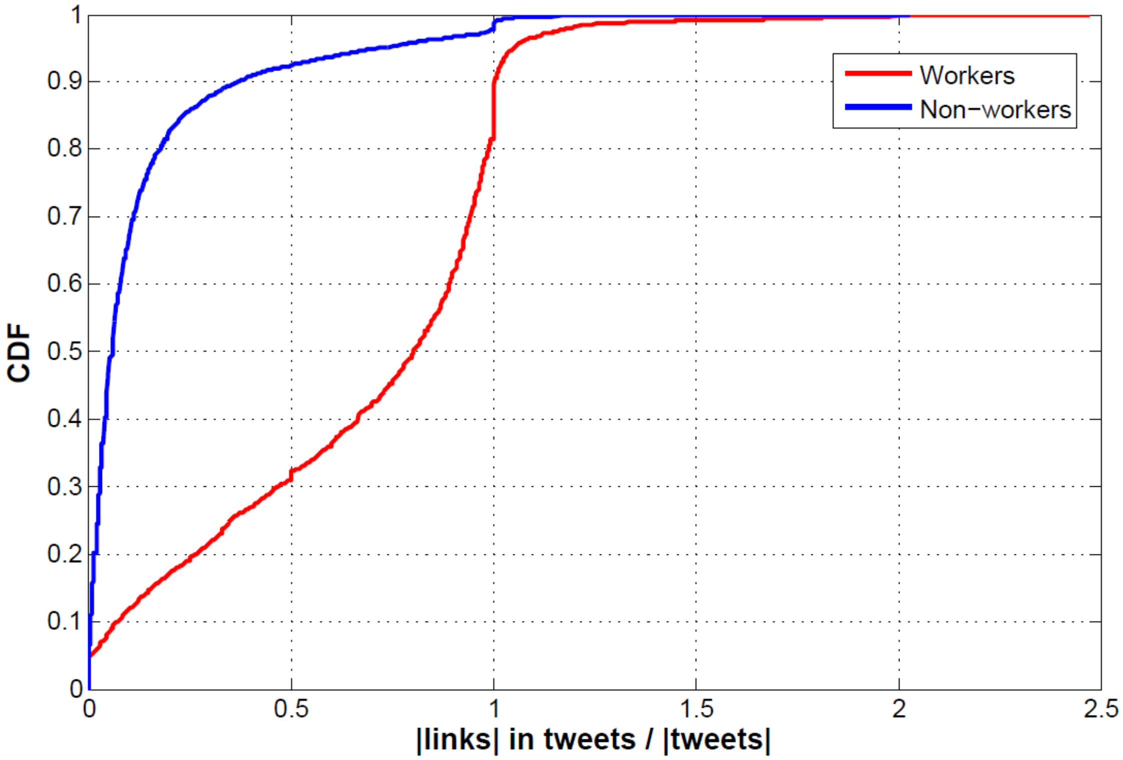


Figure 13: CDF plot of the number of URLs per tweet

5.2.3.2 Recent tweeting activity

This feature is calculated by measuring the number of tweets per day for the recent 200 tweets. The idea behind this feature is to understand the user's recent tweeting activity. Figure 14 : presents the CDF plot of this feature. Workers tend to have lesser recent tweeting activity when compared to non-workers. The reason for this behavior is that unlike normal users who post tweets frequently, workers post tweets only when they work for a Twitter campaign found in a crowdsourcing website. This also tells us that workers use their profiles almost solely for crowdturfing.

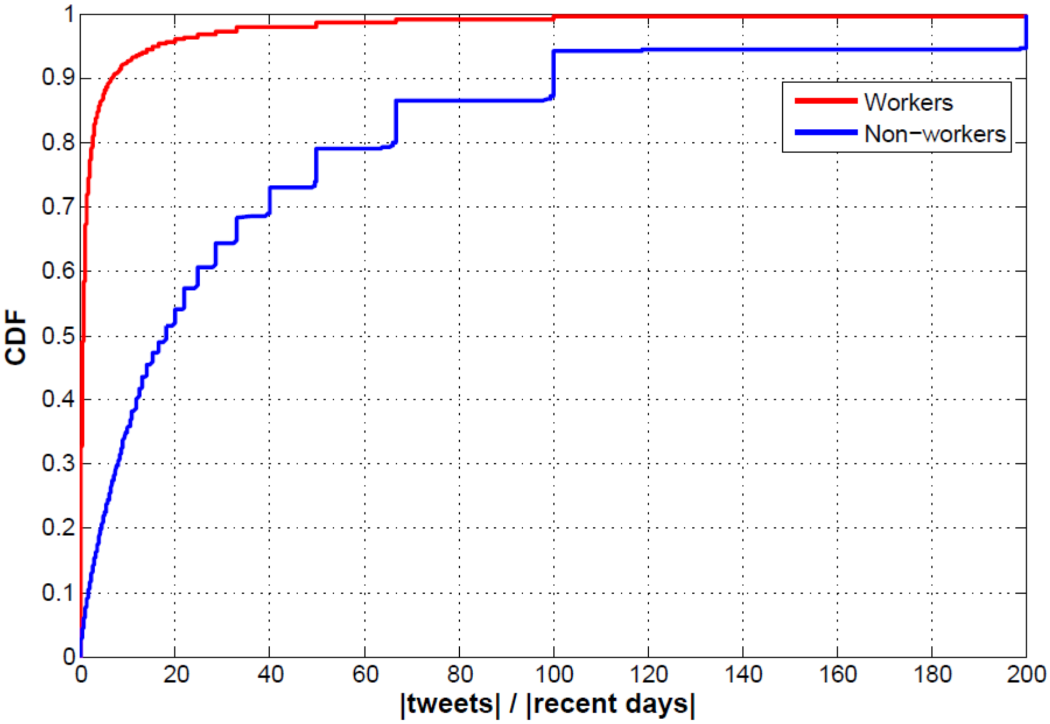


Figure 14 : CDF plot of the number of tweets per day for the 200 most recent tweets

5.2.3.3 Number of @username mentions in the 200 most recent tweets

Twitter has the unique feature of referring to other Twitter users using the "@" symbol in a tweet. This can be used to get the attention of a user to the Tweet. We count the number of @ mentions in the most recent 200 Tweets. As shown in Figure 15, we can see that workers rarely use the @ feature in their tweets, the reason being, workers post tweets from crowdsourcing sites just to get paid and their messages are not directed to a specific user.

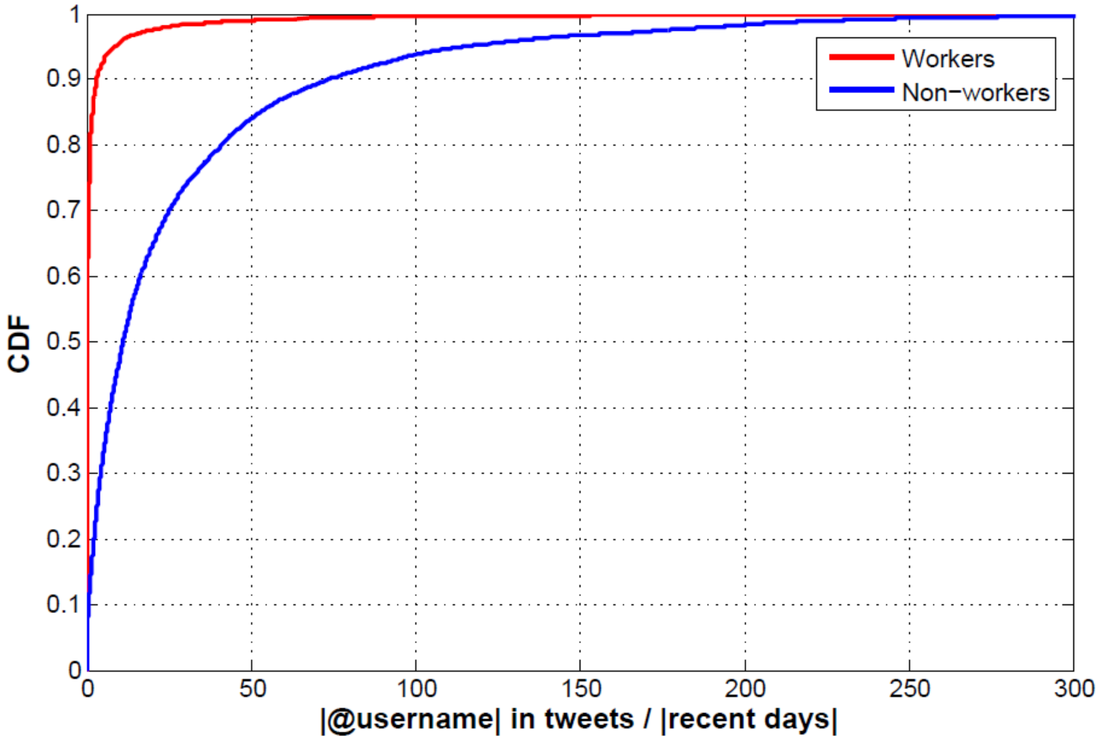


Figure 15 : CDF plot of the number of @username mentions in the 200 most recent tweets

5.2.3.4 Number of tweets posted per day (lifetime)

This feature tries to understand the overall tweeting behavior of the user. It is measured as follows,

$$Number\ of\ tweets\ posted\ \ per\ day\ (lifetime)$$
$$= \frac{Total\ number\ of\ tweets\ posted\ in\ lifetime}{Number\ of\ days\ active\ in\ Twitter}$$

As we see in Figure 16, workers tend to post a lesser number of tweets per day when compared to normal users. This is because workers post tweets only when they get a Twitter related task to work from a crowdsourcing website.

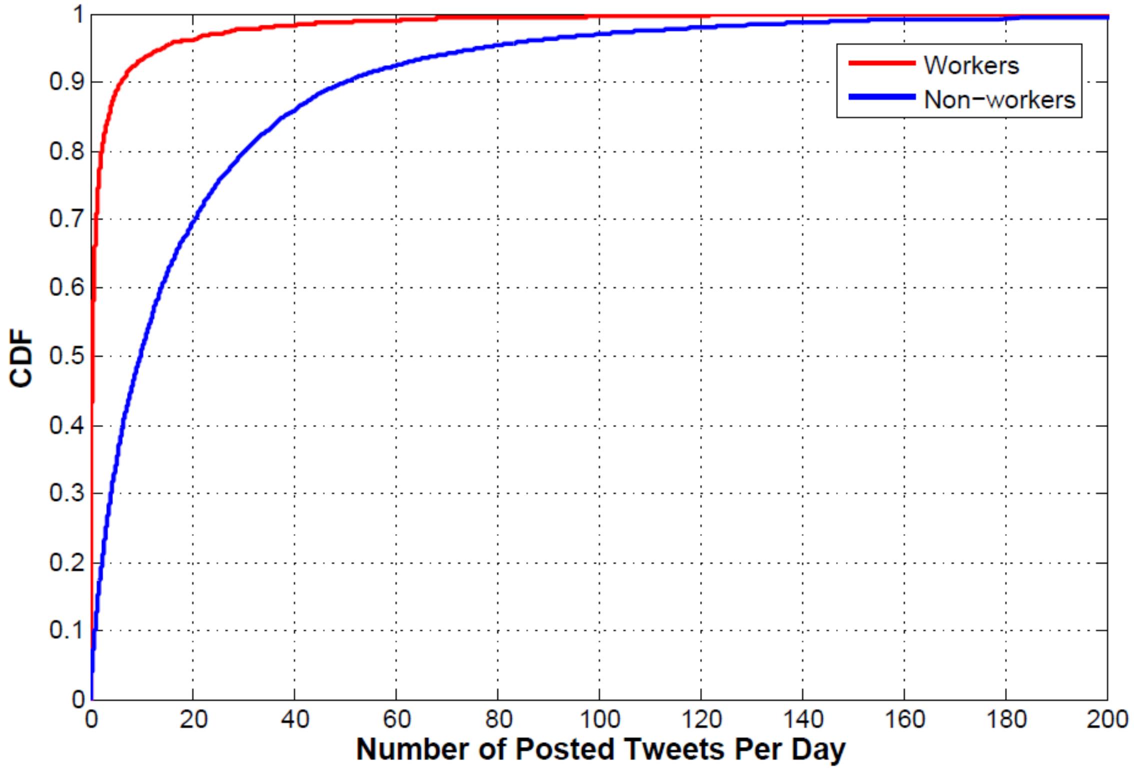


Figure 16 : CDF of the number of tweets posted per day (lifetime)

5.2.4 User content based features

Next, we study the linguistic characteristics of the tweets posted by workers and non-workers. Do workers use language differently? 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 [18]. 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 for each category based on the LIWC dictionary. The linguistic analysis shows that workers are less personal in the messages when compared to non-workers. This seems reasonable since workers intend to spread pre-defined manipulated content and URLs and thus worker tweets are less personal.

5.2.4.1 Anger in LIWC

The Anger feature (as part of LIWC) measures the fraction of words expressing anger in the tweets from a given user. There are a total of 184 words (such as hate, kill, annoyed) which are identified as expressing anger. Figure 17 shows the CDF plot of the “anger” component in the tweets for both workers and normal users. From this plot we can conclude that the tweets from workers rarely express “anger”. This is because, most of their tweets are from crowdsourcing websites promoting some meme or URL (and not their personal tweet) and hence they do not express any opinion.

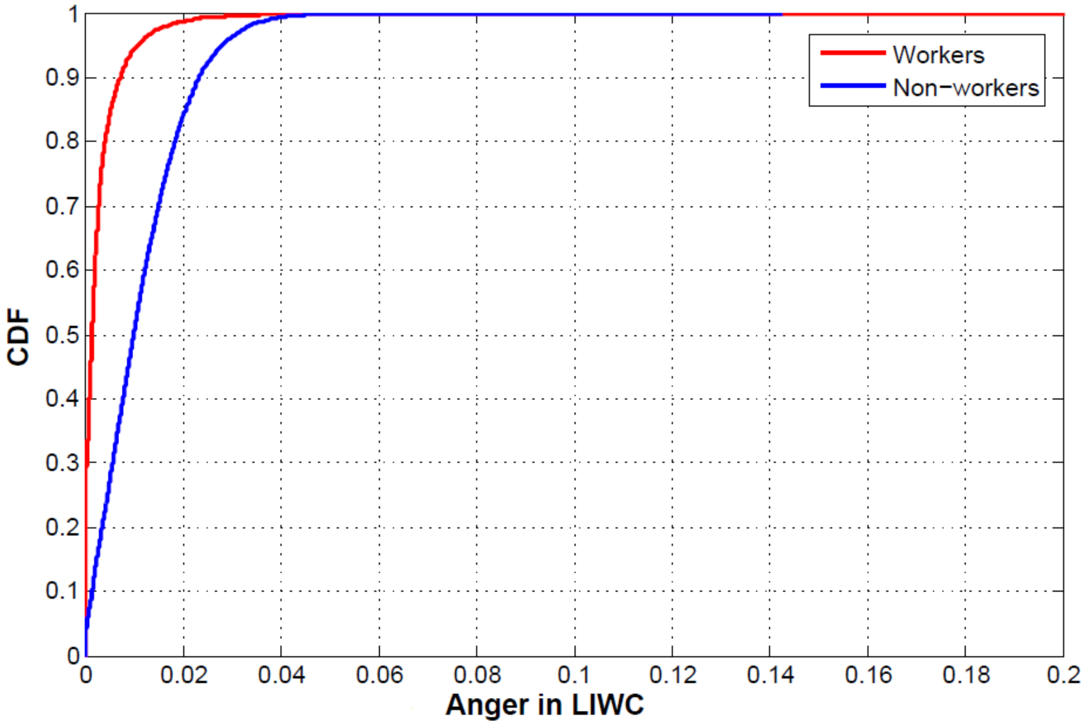


Figure 17 : CDF of the fraction of “anger” words in tweets (LIWC)

5.2.4.2 1st person singular in LIWC

The 1st person singular feature (as part of LIWC) measures the fraction of the 1st person singular pronouns (such as I, me, mine) in the tweets. The CDF plot of this feature is shown in Figure 18. Workers tend not to use personal pronouns conveying that they rarely tweet about themselves.

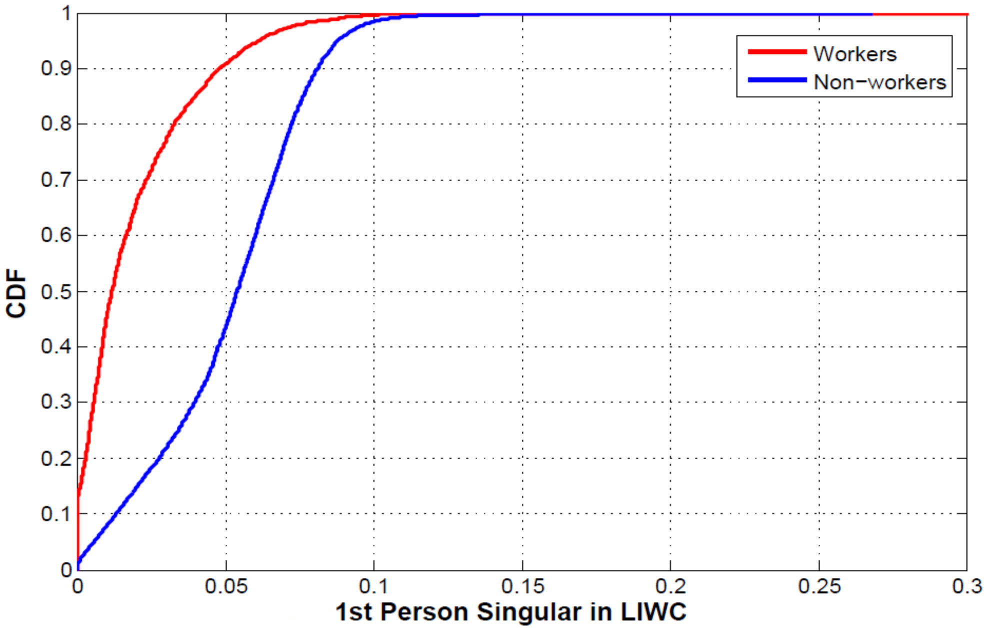


Figure 18 : CDF plot of the fraction of 1st person singular pronouns in tweets (LIWC)

5.3 Classification results

Using the four feature groups described in the previous sections, we tested 30 classification algorithms. The classification accuracies ranged from 86% to 93%. Tree-based classifiers showed the highest accuracy results. In particular, Random Forest classifier - with 25 trees each constructed while considering 50 features - produced the highest 10-fold cross validation accuracy of 93.26%. Table 8 presents the classification results.

Table 8 : Classification results

Classifier	Accuracy	F-measure	False Positive Rate	False Negative Rate
Random Forest	93.26%	0.966	0.036	0.174

In addition, we considered different training mixtures of workers and non-workers, ranging from 1% workers and 99% non-workers to 99% workers and 1% non-workers. We found that the classification quality is robust across these training mixtures. In other words, our proposed features are very strong in distinguishing between workers and non-workers.

5.4 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 section is still effective at a later point in time, we created our own 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 network information. As shown in Table 9, we collected 368 Twitter user profiles and their recent 200 messages (in total, 40,344 messages).

Table 9 : New worker dataset

Class	#Users	#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 10 presents its experimental result. It confirms that our classifier is still effective even with the passage of time with 94.3% accuracy.

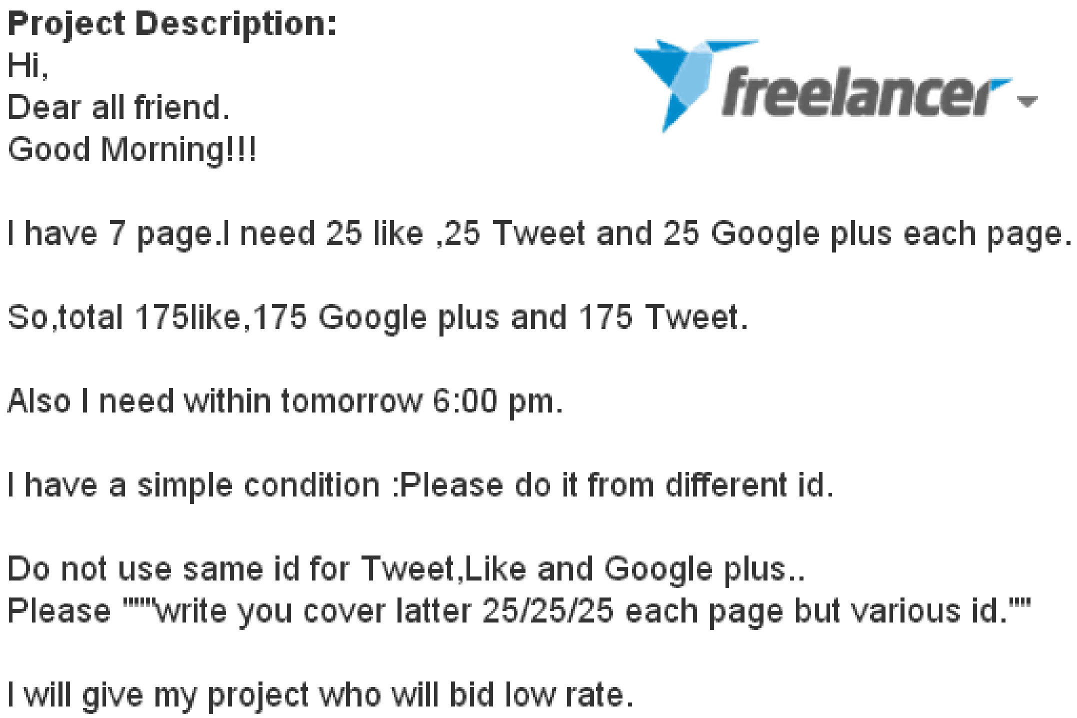
Table 10 : Classification Results on new worker dataset

Classifier	Accuracy	F-measure	False Negative Rate
Random Forest	94.3%	0.971	0.057

In summary, this positive experimental result shows that ours is a promising classification approach to identify 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 characteristics will be consistent, and will be different from regular users.

6. IDENTIFYING HIDDEN CROWDTURFING CAMPAIGNS USING LDA

In addition to the campaigns discussed in section 2, we found another interesting type of campaign – hidden campaigns – on sites such as freelancer.com and elance.com. These are special campaigns where the requesters post on the crowdsourcing site that they have some task and the workers are required to make a bid on the campaign. The worker who makes the lowest bid gets to work on the campaign. One interesting aspect of these campaigns is that the requesters post only the “type'” of the task that is needed to be done and not the exact “task'” to the done. Thus it becomes difficult to detect and curb these campaigns. Figure 19 presents such a hidden campaign.



Project Description:
Hi,
Dear all friend.
Good Morning!!!

I have 7 page.I need 25 like ,25 Tweet and 25 Google plus each page.

So,total 175like,175 Google plus and 175 Tweet.

Also I need within tomorrow 6:00 pm.

I have a simple condition :Please do it from different id.

Do not use same id for Tweet,Like and Google plus..
Please """write you cover latter 25/25/25 each page but various id.""

I will give my project who will bid low rate.

Figure 19 : An example of a hidden campaign

After we identify a group of crowdturfers in Twitter, it becomes necessary to find the hidden campaigns in which they have participated for two reasons. One, we do not know anything about hidden campaigns in which the workers have participated as the requesters do not post the exact task to be done for these campaigns and; two, we have multiple crowdsourcing sites and it is not a good idea to crawl them all to detect the hidden campaigns. We found that Latent Dirichlet Allocation (LDA) [19] with Gibbs sampling [20] can be used to identify the underlying hidden campaigns given the tweet text corpus of a set of crowdturfers.

LDA in its simplest form can be defined as a generative probabilistic model for identifying a set of hidden topics describing a text corpus. For the text corpus consisting of about 360,000 crowdturfers' tweets' text, we created multiple LDA models with Gibbs sampling having various numbers of topics such as 10, 20, 50, 100, 150 and 200 using MALLET. The hyper-parameters, alpha and beta were set to 0.5 and 0.01 respectively. We found that a model with 100 topics was able to identify most of our previously collected crowdturfing campaigns. A subset of the topics generated is presented in Table 11.

Table 11 : Sample topics for the crowdturfers' tweets from the LDA model of 100 topics

Topic#	Possible words in Topic
6	follow twitter back followers friends teamfollowback aday followback ifollowback autofollow shoutout followfriday instantfollowback ll tfb instantfollow ifollowall retweet ff

Table 11 Continued.

Topic#	Possible words in Topic
11	design services company call air service cleaning solutions area ambulance offer offers equipment professional quick years simple staff companies
14	college pro painters home young special tips pick painting students dad choose summer color process spring entrepreneurs expect father
15	blog post leave view comment photo ha write guest guys thx kind moment sharing blogspot blogging oil interesting blogger
21	kids game play games fun summer ways online playing coach passion role teach sports flash activities bingo league museum
29	baby mom cute check babies boy tip boomers girl born birth boomer sleeping sleep pregnant shower memory webdesignwijzer sweet
30	security st monitoring alarms control training sharing justin playing ready au information april bieber po opinion spend infosec pair
32	phone weight diet loss fat lose advice number call fast plan cell losing programs weightloss cash request challenge trouble
37	online education degree training public science program university master management nursing skills career leadership star academy profile wars bi

Table 11 Continued.

Topic#	Possible words in Topic
40	children nanny kids parents parenting nannies childcare moms parent teens child reasons safety tips families dads teach youth childminders
49	social media marketing infographic boy soshable socialmedia content networks networking digital automotive brand tkcarsitesinc marketers tk oc seconds engage
52	google seo search website tool increase traffic videos sharethis stand tips improve engine websites pages content powerful update results
59	vote favorite retweet fan voted anteyup win side picture pinterest poll big tcdisrupt fnboxlatamchallenge simply accessories futbol facts atlantis
74	give fiverr uk usa pr kindle link gig followers hours unique digg promo sign site logo website create messages
95	care child skin beauty natural green tea face organic products register dallas anti leaving spot essential grade skincare makeup
97	real estate miami beach south florida sale condos fl homes luxury island property cbias exciting group housing properties forbes

Of these detected topics, the topics 14, 29, 30, 32, 40, 52 can directly be matched to one campaign each from the set of Twitter campaigns we initially collected from crowdsourcing sites. This asserts that LDA is a good method for grouping related keywords from a given campaign into a single topic given a corpus of tweets from crowdturfers. Also, it is intuitive that if there are any offline campaigns in which the workers have participated, these campaigns will also appear as a separate topic when we apply LDA.

To ascertain that LDA performs well even in the presence of tweets from normal users, we created an LDA model of 200 topics for the text corpus consisting of tweets from both workers and non-workers. A subset of the topics generated is presented in Table 12. By looking through the topics we can identify that topics 13, 16, 17, 18, 35 and 43 belong to normal users and the remaining belong to crowdturfers. The topics from crowdturfers can be directly matched to the topics presented in Table 11. Thus the words in crowdturfing campaigns are grouped into a single topic irrespective of whether the text corpus contains tweets from normal users or not.

Table 12: Sample topics for the combined (both workers and non-workers) tweets corpus from the LDA model of 200 topics

Topic#	Possible words in Topic
13	play game games playing played football baseball fifa soccer xbox basketball golf team sports hunger plays ball guitar role
16	hot cold water drink beer weather shower warm bottle drinking bath winter freezing johnson cole heat stone glass vodka
17	happy birthday hope day bday enjoy eminem xx bro dear celebrate wishes wishing hint anniversary celebration present celebrating hun

Table 12 continued.

Topic#	Possible words in Topic
18	live watch tonight season show episode chat series tune premiere starts pm watching stream tim streaming emmerdale ep missed
35	feelings today hurt pisces aries leo scorpio easily aquarius capricorn taurus partner gemini emotions sagittarius emotional feeling intense jerry
43	nice rain weather sound mm wind ve speed supposed heavy storm santa raining safe km midnight standard experience sunshine
49	follow back retweet ll followers teamfollowback shoutout sbabyfollowtrain unfollow followback gain fav aday happy rts ff autofollow retweets tfb
108	food fat weight lose diet eat loss healthy health body eating fast workout fitness pounds foods exercise burn lbs
141	college pro choose company services air opinion hosting tips painting painters service medical choice home students experience student helped
175	family nanny nannies live training part group needed full families reasons free ways members parents time nyc childcare child
178	baby cute babies aw born daddy boy sweet pregnant adorable bomb gorgeous aww cutie shower blow loyalty awww poor

7. CONCLUSION

In this research, we have presented a framework for "pulling back the curtain" on crowdturfers to reveal their underlying ecosystem. We have analyzed the types of malicious campaigns hosted on multiple crowdsourcing sites. By linking campaigns and their workers on crowdsourcing sites to social media (Twitter), we have traced the activities of crowdturfers in social media and the relationship structure connecting these workers in social media. We have found 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 between regular social media users and workers. Our experimental results show that these models can effectively detect previously unknown Twitter-based workers. We also proposed a method to identify hidden campaigns using topic models.

7.1 Future work

In our current work, we concentrated on detecting Twitter workers, but still a large number of workers are involved in many other sites, such as forums, review sites and blogs. One possible extension to our current work would be to identify workers in other social media/forums and see whether their behavior is similar to that of the workers in Twitter.

Another possible extension to the current work would be to analyze the temporal variance of the characteristics of workers i.e., to study how they evolve over a period of time. Our current dataset has worker activities for only 2 months but analyzing the temporal variance of the characteristics will require collection of worker activities from social media for extended periods of time (say 1 year).

In our work we assumed that there is a one-to-one mapping between a worker and his Twitter account. But this may not be the case in reality – a single worker may maintain multiple accounts to earn multiple times from a single campaign. We can track the behaviors of multiple worker accounts simultaneously to see if they match, thereby indicating that they are actually being operated by a single worker.

Since it is not feasible to detect the workers on all sites, we plan to build a model to detect whether a given campaign from a crowdsourcing site is a spam campaign or not. We intend to achieve this by collecting campaigns from a wide range of crowdsourcing sites and extracting features unique to spam/non-spam campaigns.

REFERENCES

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

2. Sterling, B. The Chinese online ‘Water Army’. 2010; Available from: http://www.wired.com/beyond_the_beyond/2010/06/the-chinese-online-water-army.

3. Chen, C., K. Wu, V. Srinivasan, and X. Zhang, Battling the internet water army: Detection of hidden paid posters. arXiv preprint arXiv:1111.4297, 2011.

4. Sterling, B., Sui, D., Estes, A., Nolan, H., Strange, A., Internet Water Army. 2013; Available from: http://en.wikipedia.org/wiki/Internet_Water_Army.

5. Chan, C. How a Fake Erotic Fiction eBook Hit the Top 5 of iTunes. 2012; Available from: http://gizmodo.com/5933169/how-a-fakecrowdsourced-erotic-ebook-hit-the-top-5-ofitunes.

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

7. Wang, G., C. Wilson, X. Zhao, Y. Zhu, M. Mohanlal, et al. Serf and turf: Crowdturfing for fun and profit. in Proceedings of the 21st international conference on World Wide Web. 2012. ACM.

8. Hirth, M., T. Hoßfeld, and P. Tran-Gia. Anatomy of a crowdsourcing platform-using the example of microworkers.com. in Proceedings of the 2011 Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS). 2011. IEEE.

9. Kittur, A., E.H. Chi, and B. Suh. Crowdsourcing user studies with Mechanical Turk. in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008. ACM.

10. Brabham, D.C., Crowdsourcing as a model for problem solving an introduction and cases. CONVERGENCE: the international journal of research into new media technologies, 2008. 14(1): p. 75-90.

11. Alonso, O., D.E. Rose, and B. Stewart. Crowdsourcing for relevance evaluation. in ACM SIGIR Forum. 2008. ACM.

12. Franklin, M.J., D. Kossmann, T. Kraska, S. Ramesh, and R. Xin. CrowdDB: answering queries with crowdsourcing. in Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. 2011.

13. Bernstein, M.S., G. Little, R.C. Miller, B. Hartmann, M.S. Ackerman, et al. Soylent: a word processor with a crowd inside. in Proceedings of the 23nd annual ACM symposium on User interface software and technology. 2010. ACM.

14. Ratkiewicz, J., M. Conover, M. Meiss, B. Gonçalves, S. Patil, et al. Truthy: mapping the spread of astroturf in microblog streams. in Proceedings of the 20th international conference companion on World wide web. 2011. ACM.

15. Venetis, P. and H. Garcia-Molina. Quality control for comparison microtasks. in Proceedings of the First International Workshop on Crowdsourcing and Data Mining. 2012. ACM.

16. Xia, T., C. Zhang, J. Xie, and T. Li. Real-time quality control for crowdsourcing relevance evaluation. in Network Infrastructure and Digital Content (IC-NIDC), 2012 3rd IEEE International Conference on. 2012. IEEE.

17. Yang, C., R. Harkreader, J. Zhang, S. Shin, and G. Gu. Analyzing spammers' social networks for fun and profit: A case study of cyber criminal ecosystem on Twitter. in Proceedings of the 21st international conference on World Wide Web. 2012. ACM.

18. Pennebaker, J.W., M.E. Francis, and R.J. Booth, Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates, 2001.

19. Blei, D.M., A.Y. Ng, and M.I. Jordan, Latent dirichlet allocation. the Journal of machine Learning research, 2003. 3: p. 993-1022.

20. Griffiths, T.L. and M. Steyvers, Finding scientific topics. Proceedings of the National academy of Sciences of the United States of America, 2004. 101(Suppl 1): p. 5228-5235.