Noname manuscript No.
(will be inserted by the editor)

Characterizing and Automatically Detecting Crowdturfing in Fiverr and Twitter

Kyumin Lee · Steve Webb · Hancheng Ge

Received: date / Accepted: date

Abstract As human computation on crowdsourcing systems has become popular and powerful for performing tasks, malicious users have started misusing these systems by posting malicious tasks, propagating manipulated contents, and targeting popular web services such as online social networks and search engines. Recently, these malicious users moved to Fiverr, a fast growing micro-task marketplace, where workers can post crowdturfing tasks (i.e., astroturfing campaigns run by crowd workers) and malicious customers can purchase those tasks for only $5. In this manuscript, we present a comprehensive analysis of crowdturfing in Fiverr and Twitter, and we develop predictive models to detect and prevent crowdturfing tasks in Fiverr and malicious crowd workers in Twitter. First, we identify the most popular types of crowdturfing tasks found in Fiverr and conduct case studies for these crowdturfing tasks. Second, we build crowdturfing task detection classifiers to filter these tasks and prevent them from becoming active in the marketplace. Our experimental results

An early version of this manuscript appeared in the 2014 International AAAI Conference on Weblogs and Social Media (ICWSM) [17].

Kyumin Lee
Department of Computer Science
Utah State University
Logan, UT 84322
E-mail: kyumin.lee@usu.edu

Steve Webb
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332
E-mail: steve.webb@gmail.com

Hancheng Ge
Department of Computer Science and Engineering
Texas A&M University
College Station, TX 77843
E-mail: hge@cse.tamu.edu

show that the proposed classification approach effectively detects crowdturfing tasks, achieving 97.35% accuracy. Third, we analyze the real world impact of crowdturfing tasks by purchasing active Fiverr tasks and quantifying their impact on a target site (Twitter). As part of this analysis, we show that current security systems inadequately detect crowdsourced manipulation, which confirms the necessity of our proposed crowdturfing task detection approach. Finally, we analyze characteristics of paid Twitter workers, find distinguishing features between these workers and legitimate Twitter accounts, and use these features to build classifiers that detect Twitter workers. Our experimental results show that our classifiers are able to detect Twitter workers effectively, achieving 99.29% accuracy.

1 Introduction

Crowdsourcing systems are becoming more and more popular because they can quickly accomplish tasks that are difficult for computers but easy for humans. For example, a word document can be summarized and proofread by crowd workers while the document is still being written by its author [5], and missing data in database systems can be populated by crowd workers [8]. As the popularity of crowdsourcing has increased, various systems have emerged – from general-purpose crowdsourcing platforms such as Amazon Mechanical Turk, Crowdflower and Fiverr, to specialized systems such as Ushahidi (for crisis information) and Foldit (for protein folding).

These systems offer numerous positive benefits because they efficiently distribute jobs to a workforce of willing individuals. However, malicious customers and unethical workers have started misusing these systems, spreading malicious URLs in social media, posting fake reviews and ratings, forming artificial grassroots campaigns, and manipulating search engines (e.g., creating numerous backlinks to targeted pages and artificially increasing user traffic). Recently, news media reported that 1,000 crowdturfers – workers performing crowdturfing tasks on behalf of buyers – were hired by Vietnamese propaganda officials to post comments that supported the government [21], and the "Internet water army" in China created an artificial campaign to advertise an online computer game [6, 23]. These types of crowdsourced manipulations reduce the quality of online social media, degrade trust in search engines, manipulate political opinion, and eventually threaten the security and trustworthiness of online web services. Recent studies found that ∼90% of all tasks in crowdsourcing sites were for "crowdturfing" – astroturfing campaigns run by crowd workers on behalf of customers – [29], and most malicious tasks in crowdsourcing systems target either online social networks (56%) or search engines (33%) [16].

Unfortunately, very little is known about the properties of crowdturfing tasks, their impact on the web ecosystem, or how to detect and prevent them. Hence, in this manuscript we are interested in analyzing Fiverr – a fast growing micro-task marketplace and the 125th most popular site in the world [1]

– to be the first to answer the following questions: what are the most important characteristics of buyers (a.k.a. customers) and sellers (a.k.a. workers)? What types of tasks, including crowdturfing tasks, are available? What sites do crowdturfers target? How much do they earn? Based on this analysis and the corresponding observations, can we automatically detect these crowdturfing tasks? Can we measure the impact of these crowdturfing tasks? Can the current security systems for targeted sites adequately detect crowdsourced manipulation? Do paid Twitter workers have different characteristics than legitimate Twitter accounts? Can we develop classifiers only based on Twitter-based information to detect paid Twitter workers?

To answer these questions, we make the following contributions in this manuscript:

- First, we collect a large number of active tasks (these are called gigs in Fiverr) from all categories in Fiverr. Then, we analyze the properties of buyers and sellers as well as the types of crowdturfing tasks found in this marketplace. To our knowledge, this is the first study to focus primarily on Fiverr.
- Second, we conduct a statistical analysis of the properties of crowdturfing and legitimate tasks, and we build a machine learning based crowdturfing task classifier to actively filter out existing and new malicious tasks, preventing the propagation of crowdsourced manipulation to other web sites. To our knowledge, this is the first study to detect crowdturfing tasks automatically.
- Third, we feature case studies of three specific types of crowdturfing tasks: social media targeting gigs, search engine targeting gigs and user traffic targeting gigs.
- Fourth, we purchase active crowdturfing tasks targeting a popular social media site, Twitter, and measure the impact of these tasks on the targeted site. We then test how many crowdsourced manipulations Twitter's security systems can detect, and we confirm the necessity of our proposed crowdturfing detection approach.
- Finally, we analyze characteristics of paid Twitter workers and find distinguishing features between the paid Twitter workers and legitimate accounts. Based on these features, we develop classifiers that detect the paid workers with 99.29% accuracy.

2 Related Work

In this section, we introduce some crowdsourcing research work which focused on understanding workers' demographic information, filtering low quality answers and spammers, and analyzing crowdturfing tasks and market.

Ross et al. [22] analyzed user demographics on Amazon Mechanical Turk, and found that the number of non-US workers has been increased, especially led by Indian workers who were mostly young, well-educated males. Heymann and Garcia-Molina [12] proposed a novel analytics tool for crowdsourcing systems to gather logging events such as workers' location and used browser type.

Crowd workers have been used to identify sybils (fake accounts) in Facebook and Renren [28]. Baba et al. [3] hired crowd workers (non-experts) to identify improper tasks in a Japanese crowdsourcing site, Lancers[1] and found that these workers were able to correctly identify these tasks.

Other researchers studied how to control quality of crowdsourced work, aiming at getting high quality results and filtering spammers who produce low quality answers. Venetis and Garcia-Molina [27] compared various low quality answer filtering approaches such as gold standard, plurality and work time, found that the more number of workers participated in a task, the better result was produced. Halpin and Blanco [11] used a machine learning technique to detect spammers at Amazon Mechanical Truck. Allahbakhsh et al. [2] classify existing task quality control approaches into two categories such as design-time and run-time.

Researchers began studying crowdturfing problems and market. Motoyama et al. [18] analyzed abusive tasks on Freelancer. Wang et al. [29] analyzed two Chinese crowdsourcing sites and estimated that 90% of all tasks were crowdturfing tasks. Lee et al. [16] analyzed three Western crowdsourcing sites (e.g., Microworkers.com, ShortTask.com and Rapidworkers.com) and found that mainly targeted systems were online social networks (56%) and search engines (33%). Stringhini et al. [24] and Thomas et al. [25] studied Twitter follower market and Twitter account market, respectively.

Compared with the previous research work, we collect a large number of active tasks in Fiverr and analyze crowdturfing tasks among them. We then develop crowdturfing task detection classifiers for the first time and effectively detect crowdturfing tasks. We measure the impact of these crowdturfing tasks in Twitter. Finally, we develop classifiers to detect malicious crowd workers in a target site, Twitter, and our approach effectively detect these workers. Our research will complement the existing research work.

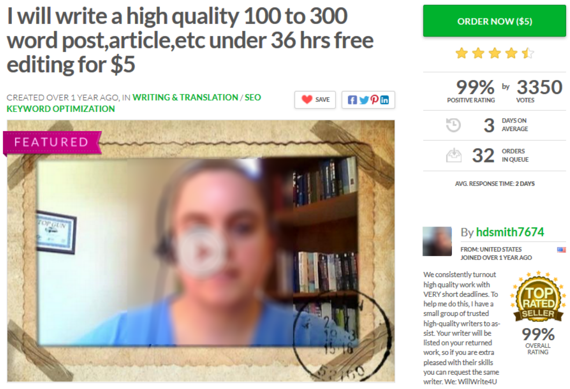
3 Background

Fiverr is a micro-task marketplace where users can buy and sell services, which are called *gigs*. The site has over 1.7 million registered users, and it has listed more than 2 million gigs[2]. As of November 2013, it is the 125th most visited site in the world according to Alexa [1].

Fiverr gigs do not exist in other e-commerce sites, and some of them are humorous (e.g., "I will paint a logo on my back" and "I will storyboard your script"). In the marketplace, a *buyer* purchases a gig from a *seller* (the default purchase price is $5). A user can be a buyer and/or a seller. A buyer can post a review about the gig and the corresponding seller. Each seller can be promoted

[1] http://www.lancers.jp

[2] http://blog.Fiverr/2013/08/12/fiverr-community-milestone-two-million-reasons-to-celebrate-iamfiverr/



I will write a high quality 100 to 300 word post,article,etc under 36 hrs free editing for $5

CREATED OVER 1 YEAR AGO, IN WRITING & TRANSLATION / SEO KEYWORD OPTIMIZATION

99% POSITIVE RATING by 3350 VOTES

3 DAYS ON AVERAGE

32 ORDERS IN QUEUE

AVG. RESPONSE TIME: 2 DAYS

By hdsmith7674

FROM: UNITED STATES
JOINED OVER 1 YEAR AGO

We consistently turnout high quality work with VERY short deadlines. To help me do this, I have a small group of trusted high-quality writers to assist. Your writer will be listed on your returned work, so if you are extra pleased with their skills you can request the same writer. We: WillWrite4U

Fig. 1 An example of a Fiverr gig listing.

to a 1st level seller, a 2nd level seller, or a top level seller by selling more gigs. Higher level sellers can sell additional features (called "gig extras") for a higher price (i.e., more than $5). For example, one seller offers the following regular gig: "I will write a high quality 100 to 300 word post,article,etc under 36 hrs free editing for $5". For an additional $10, she will "make the gig between 600 to 700 words in length", and for an additional $20, she will "make the gig between 800 to 1000 words in length". By selling these extra gigs, the promoted seller can earn more money. Each user also has a profile page that displays the user's bio, location, reviews, seller level, gig titles (i.e., the titles of registered services), and number of sold gigs.

Figure 1 shows an example of a gig listing on Fiverr. The listing's human-readable URL is http://Fiverr.com/hdsmith7674/write-a-high-quality-100-to-300-word-postarticleetc-under-36-hrs-free-editing, which was automatically created by Fiverr based on the title of the gig. The user name is "hdsmith7674", and the user is a top rated seller.

Ultimately, there are two types of Fiverr sellers: (1) legitimate sellers and (2) unethical (malicious) sellers, as shown in Figure 2. Legitimate sellers post legitimate gigs that do not harm other users or other web sites. Examples of legitimate gigs are "I will color your logo" and "I will sing a punkrock happy birthday". On the other hand, unethical sellers post crowdturfing gigs on Fiverr that target sites such as online social networks and search engines. Examples of crowdturfing gigs are "I will provide 2000+ perfect looking twitter followers" and "I will create 2,000 Wiki Backlinks". These gigs are clearly used to manipulate their targeted sites and provide an unfair advantage for their buyers.

4 Fiverr Characterization

In this section, we present our data collection methodology. Then, we measure the number of the active Fiverr gig listings and estimate the number of listings

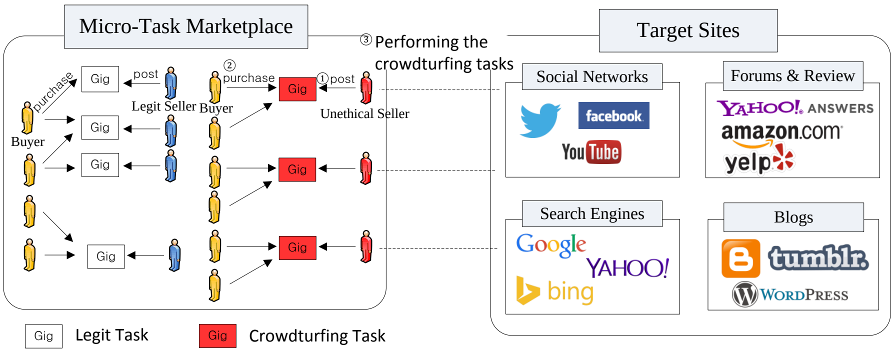


Fig. 2 The interactions between buyers and legitimate sellers on Fiverr, contrasted with the interactions between buyers and unethical sellers.

that have ever been created. Finally, we analyze the characteristics of Fiverr buyers and sellers.

4.1 Dataset

To collect gig listings, we built a custom Fiverr crawler. This crawler initially visited the Fiverr homepage and extracted its embedded URLs for gig listings. Then, the crawler visited each of those URLs and extracted new URLs for gig listings using a depth-first search. By doing this process, the crawler accessed and downloaded each gig listing from all of the gig categories between July and August 2013. From each listing, we also extracted the URL of the associated seller and downloaded the corresponding profile. Overall, we collected 89,667 gig listings and 31,021 corresponding user profiles.

4.2 Gig Analysis

First, we will analyze the gig listings in our dataset and answer relevant questions.

How much data was covered? We attempted to collect every active gig listing from every gig category in Fiverr. To check how many active listings we collected, we used a sampling approach. When a listing is created, Fiverr internally assigns a sequentially increasing numerical id to the gig. For example, the first created listing received 1 as the id, and the second listing received 2. Using this number scheme, we can access a listing using the following URL format: http://Fiverr.com/[GIG_NUMERICAL_ID], which will be redirected to the human-readable URL that is automatically assigned based on the gig's title.

As part of our sampling approach, we sampled 1,000 gigs whose assigned id numbers are between 1,980,000 and 1,980,999 (e.g., http://Fiverr.com/1980000).

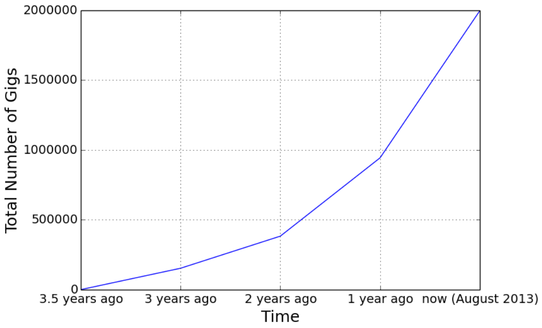


Fig. 3 Total number of created gigs over time.

Then, we checked how many of those gigs are still active because gigs are often paused or deleted. 615 of the 1,000 gigs were still active. Next, we crossreferenced these active listings with our dataset to see how many listings overlapped. Our dataset contained 517 of the 615 active listings, and based on this analysis, we can approximate that our dataset covered 84% of the active gigs on Fiverr. This analysis also shows that gig listings can become stale quickly due to frequent pauses and deletions.

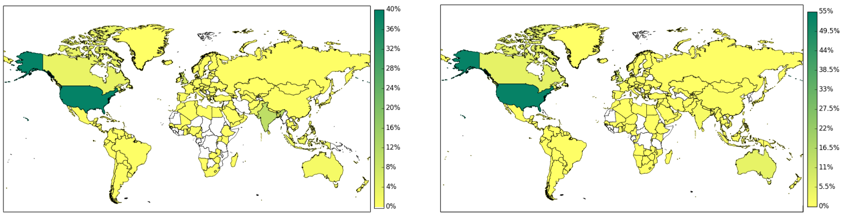
Initially, we attempted to collect listings using gig id numbers (e.g., http://Fiverr.com/1980000), but Fiverr's Safety Team blocked our computers' IP addresses because accessing the id-based URLs is not officially supported by the site. To abide by the site's policies, we used the human-readable URLs, and as our sampling approach shows, we still collected a significant number of active Fiverr gig listings.

How many gigs have been created over time? A gig listing contains the gig's numerical id and its creation time, which is displayed as days, months, or years. Based on this information, we can measure how many gigs have been created over time. In Figure 3, we plotted the approximate total number of gigs that have been created each year. The graph follows the exponential distribution in macro-scale (again, yearly) even though the micro-scaled plot may show us a clearer growth rate. This plot shows that Fiverr has been getting more popular, and in August 2013, the site reached 2 million listed gigs.

4.3 User Analysis

Next, we will analyze the characteristics of Fiverr buyers and sellers in the dataset.

Where are sellers from? Are the sellers distributed all over the world? In previous research, sellers (i.e., workers) in other crowdsourcing sites were usually from developing countries [16]. To determine if Fiverr has the same demographics, Figure 4(a) shows the distribution of sellers on the world map.



(a) Distribution of sellers in the world map. (b) Distribution of buyers in the world map.

Fig. 4 Distribution of all sellers and buyers in the world map.

Sellers are from 168 countries, and surprisingly, the largest group of sellers are from the United States (39.4% of the all sellers), which is very different from other sites. The next largest group of sellers is from India (10.3%), followed by the United Kingdom (6.2%), Canada (3.4%), Pakistan (2.8%), Bangladesh (2.6%), Indonesia (2.4%), Sri Lanka (2.2%), Philippines (2%), and Australia (1.6%). Overall, the majority of sellers (50.6%) were from the western countries.

Table 1 Top 10 sellers.

Username	|Sold Gigs|	Eared	|Gigs|	Gig Category	Crowdturfing
crorkservice	601,210	3,006,050	29	Online Marketing	yes
dino_stark	283,420	1,417,100	3	Online Marketing	yes
volarex	173,030	865,150	15	Online Marketing	yes
alanletsgo	171,240	856,200	29	Online Marketing	yes
portron	167,945	839,725	3	Online Marketing	yes
mikemeth	149,090	745,450	19	Online Marketing	yes
actualreviewnet	125,530	627,650	6	Graphics	no
bestoftwitter	123,725	618,625	8	Online Marketing	yes
amazesolutions	99,890	499,450	1	Online Marketing	yes
sarit11	99,320	496,600	2	Online Marketing	yes

What is Fiverr's market size? We analyzed the distribution of purchased gigs in our dataset and found that a total of 4,335,253 gigs were purchased from the 89,667 unique listings. In other words, the 31,021 users in our dataset sold more than 4.3 million gigs and earned at least $21.6 million, assuming each gig's price was $5. Since some gigs cost more than $5 (due to gig extras), the total gig-related revenue is probably even higher. Obviously, Fiverr is a huge marketplace, but where are the buyers coming from? Figure 4(b) shows the distribution of sold gigs on the world map. Gigs were bought from all over the world (208 total countries), and the largest number of gigs (53.6% of the 4,335,253 sold gigs) were purchased by buyers in the United States. The next most frequent buyers are the United Kingdom (10.3%), followed by Canada (5.5%), Australia (5.2%), and India (1.7%). Based on this analysis, the majority of the gigs were purchased by the western countries.

Who are the top sellers? The top 10 sellers are listed in Table 1. Amazingly, one seller (crorkservice) has sold 601,210 gigs and earned at least $3 million over the past 2 years. In other words, one user from Moldova has earned at least $1.5 million/year, which is orders of magnitude larger than $2,070, the GNI (Gross National Income) per capita of Moldova [4]. Even the 10th highest seller has earned almost $500,000. Another interesting observation is that 9 of the top 10 sellers have had multiple gigs that were categorized as online marketing, advertising, or business. The most popular category of these gigs was online marketing.

We carefully investigated the top sellers' gig descriptions to identify which gigs they offered and sold to buyers. Gigs provided by the top sellers (except actualreviewnet) are all crowdturfing tasks, which require sellers to manipulate a web page's PageRank score, artificially propagate a message through a social network, or artificially add friends to a social networking account. This observation indicates that despite the positive aspects of Fiverr, some sellers and buyers have abused the micro-task marketplace, and these crowdturfing tasks have become the most popular gigs. These crowdturfing tasks threaten the entire web ecosystem because they degrade the trustworthiness of information. Other researchers have raised similar concerns about crowdturfing problems and concluded that these artificial manipulations should be detected and prevented [29,16]. However, previous work has not studied how to detect these tasks. For the remainder of this manuscript, we will analyze and detect these crowdturfing tasks in Fiverr.

5 Analyzing and Detecting Crowdturfing Gigs

In the previous section, we observed that top sellers have earned millions of dollars by selling crowdturfing gigs. Based on this observation, we now turn our attention to studying these crowdturfing gigs in detail and automatically detect them.

5.1 Data Labeling and 3 Types of Crowdturfing Gigs

To understand what percentage of gigs in our dataset are associated with crowdturfing, we randomly selected 1,550 out of the 89,667 gigs and labeled them as a legitimate or crowdturfing task. Table 2 presents the labeled distribution of gigs across 12 top level gig categories predefined by Fiverr. 121 of the 1,550 gigs (6%) were crowdturfing tasks, which is a significant percentage of the micro-task marketplace. Among these crowdturfing tasks, most of them were categorized as online marketing. In fact, 55.3% of all online marketing gigs in the sample data were crowdturfing tasks.

Next, we manually categorized the 121 crowdturfing gigs into three groups: (1) social media targeting gigs, (2) search engine targeting gigs, and (3) user traffic targeting gigs.

Table 2 Labeled data of randomly selected 1,550 gigs.

Category	|Gigs|	|Crowdturfing|	Crowdtufing%
Advertising	99	4	4%
Business	51	1	2%
Fun&Bizarre	81	0	0%
Gifts	67	0	0%
Graphics&Design	347	1	0.3%
Lifestyle	114	0	0%
Music&Audio	123	0	0%
Online Marketing	206	114	55.3%
Other	20	0	0
Programming...	84	0	0
Video&Animation	201	0	0
Writing&Trans...	157	1	0.6%
Total	1,550	121	6%

Social media targeting gigs. 65 of the 121 crowdturfing gigs targeted social media sites such as Facebook, Twitter and Youtube. The gig sellers know that buyers want to have more friends or followers on these sites, promote their messages or URLs, and increase the number of views associated with their videos. The buyers expect these manipulation to result in more effective information propagation, higher conversion rates, and positive social signals for their web pages and products.

Search engine targeting gigs. 47 of the 121 crowdturfing gigs targeted search engines by artificially creating backlinks for a targeted site. This is a traditional attack against search engines. However, instead of creating backlinks on their own, the buyers take advantage of sellers to create a large number of backlinks so that the targeted page will receive a higher PageRank score (and have a better chance of ranking at the top of search results). The top seller in Table 1 (crorkservice) has sold search engine targeting gigs and earned $3 million with 100% positive ratings and more than 47,000 positive comments from buyers who purchased the gigs. This fact indicates that the search engine targeting gigs are popular and profitable.

User traffic targeting gigs. 9 of the 121 crowdturfing gigs claimed to pass user traffic to a targeted site. Sellers in this group know that buyers want to generate user traffic (visitors) for a pre-selected web site or web page. With higher traffic, the buyers hope to abuse Google AdSense, which provides advertisements on each buyer's web page, when the visitors click the advertisements. Another goal of purchasing these traffic gigs is for the visitors to purchase products from the pre-selected page.

To this point, we have analyzed the labeled crowdturfing gigs and identified monetization as the primary motivation for purchasing these gigs. By abusing the web ecosystem with crowd-based manipulation, buyers attempt to maximize their profits [18, 29]. In the next section, we will develop an approach to detect these crowdturfing gigs automatically.

Table 3 Confusion matrix

		Predicted	
		Crowdturfing	Legitimate
Actual	Crowdturfing Gig	*a*	*b*
	Legit Gig	*c*	*d*

5.2 Detecting Crowdturfing Gigs

Automatically detecting crowdturfing gigs is an important task because it allows us to remove the gigs before buyers can purchase them, and eventually, it will allow us to prohibit sellers from posting these gigs. To detect crowdturfing gigs, we built machine-learned models using the manually labeled 1,550 gig dataset.

The performance of a classifier depends on the quality of features, which have distinguishing power between crowdturfing gigs and legitimate gigs in this context. Our feature set consists of the title of a gig, the gig's description, a top level category, a second level category (each gig is categorized to a top level and then a second level – e.g., "online marketing" as the top level and "social marketing" as the second level), ratings associated with a gig, the number of votes for a gig, a gig's longevity, a seller's response time for a gig request, a seller's country, seller longevity, seller level (e.g., top level seller or 2nd level seller), a world domination rate (the number of countries where buyers of the gig were from, divided by the total number of countries), and distribution of buyers by country (e.g., entropy and standard deviation). For the title and job description of a gig, we converted these texts into bag-of-word models in which each distinct word becomes a feature. We also used *tf-idf* to measure values for these text features.

To understand which feature has distinguishing power between crowdturfing gigs and legitimate gigs, we measured the chi-square of the features. The most interesting features among the top features, based on chi-square, are category features (top level and second level), a world domination rate, and bag-of-words features such as "link", "backlink", "follow", "twitter", "rank", "traffic", and "bookmark".

Since we don't know which machine learning algorithm (or classifier) would perform best in this domain, we tried over 30 machine learning algorithms such as Naive Bayes, Support Vector Machine (SVM), and tree-based algorithms by using the Weka machine learning toolkit with default values for all parameters [30]. We used 10-fold cross-validation with 1,550 gigs for each machine learning algorithm.

We compute precision, recall, F-measure, accuracy, false positive rate (FPR) and false negative rate (FNR) as metrics to evaluate our classifiers. Overall, SVM outperformed the other classification algorithms. Its classification result is shown in Table 4. It achieved 97.35% accuracy, 0.974 F_1, 0.008 FPR, and 0.248 FNR. This positive result shows that our classification approach works well and that it is possible to automatically detect crowdturfing gigs.

Table 4 SVM-based classification result

Accuracy	F_1	FPR	FNR
97.35%	0.974	0.008	0.248

6 Detecting Crowdturfing Gigs in the Wild and Case Studies

In this section, we apply our classification approach to a large dataset to find new crowdturfing gigs and conduct case studies of the crowdturfing gigs in detail.

6.1 Newly Detected Crowdturfing Gigs

In this study, we detect crowdturfing gigs in the wild, analyze newly detected crowdturfing gigs, and categorize each crowdturfing gig to one of the three crowdturfing types (social media targeting gig, search engine targeting gig, or user traffic targeting gig) revealed in the previous section.

First, we trained our SVM-based classifier with the 1,550 labeled gigs, using the same features as the previous experiment in the previous section. However, unlike the previous experiment, we used all 1,550 gigs as the training set. Since we used the 1,550 gigs for training purposes, we removed those gigs (and 299 other gigs associated with the users that posted the 1,550 gigs) from the large dataset containing 89,667 gigs. After this filtering, the remaining 87,818 gigs were used as the testing set.

We built the SVM-based classifier with the training set and predicted class labels of the gigs in the testing set. 19,904 of the 87,818 gigs were predicted as crowdturfing gigs. Since this classification approach was evaluated in the previous section and achieved high accuracy with a small number of misclassifications for legitimate gigs, almost all of these 19,904 gigs should be real crowdturfing gigs. To make verify this conclusion, we manually scanned the titles of all of these gigs and confirmed that our approach worked well. Here are some examples of these gig titles: “I will 100+ Canada real facebook likes just within 1 day for $5”, “I will send 5,000 USA only traffic to your website/blog for $5”, and “I will create 1000 BACKLINKS guaranteed + bonus for $5”.

To understand and visualize what terms crowdturfing gigs often contain, we generated a word cloud of titles for these 19,904 crowdturfing gigs. First, we extracted the titles of the gigs and tokenized them to generate unigrams. Then, we removed stop words. Figure 5 shows the word cloud of crowdturfing gigs. The most popular terms are online social network names (e.g., Facebook, Twitter, and YouTube), targeted goals for the online social networks (e.g., likes and followers), and search engine related terms (e.g., backlinks, website, and Google). This word cloud also helps confirm that our classifier accurately identified crowdturfing gigs.

Next, we are interested in analyzing the top 10 countries of buyers and sellers in the crowdturfing gigs. Can we identify different country distributions

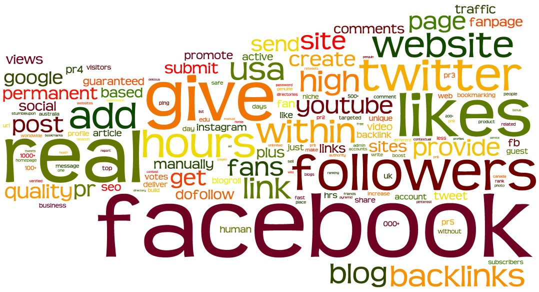
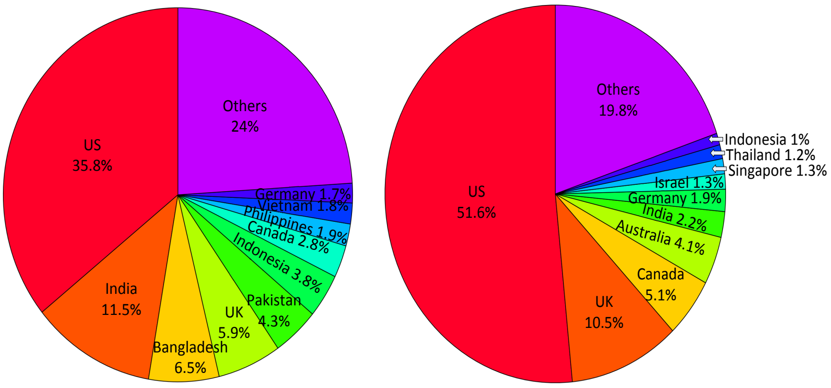


Fig. 5 Word cloud of crowdturfing gigs.

compared with the distributions of the overall Fiverr sellers and buyers shown in Figure 4? Are country distributions of sellers and buyers in the crowdsourcing gigs in Fiverr different from distribution of users in other crowdsourcing sites? Interestingly, the most frequent sellers of the crowdturfing gigs in Figure 6(a) were from the United States (35.8%), following a similar distribution as the overall Fiverr sellers. This distribution is very different from another research result [16], in which the most frequent sellers (called "workers" in that research) in another crowdsourcing site, Microworkers.com, were from Bangladesh. This observation might imply that Fiverr is more attractive than Microworkers.com for U.S. residents since selling a gig on Fiverr gives them higher profits (each gig costs at least \$5 but only 50 cents at Microworkers.com). The country distribution for buyers of the crowdturfing gigs in Figure 6(b) is similar with the previous research result [16], in which the majority of buyers (called "requesters" in that research) were from English-speaking countries. This is also consistent with the distribution of the overall Fiverr buyers. Based on this analysis, we conclude that the majority of buyers and sellers of the crowdturfing gigs were from the U.S. and other western countries, and these gigs targeted major web sites such as social media sites and search engines.

6.2 Case Studies of 3 Types of Crowdturfing Gigs

From the previous section, the classifier detected 19,904 crowdturfing gigs. In this section, we classify these 19,904 gigs into the three crowdturfing gig groups in order to feature case studies for the three groups in detail. To further classify the 19,904 gigs into three crowdturfing groups, we built another classifier that was trained using the 121 crowdturfing gigs (used in the previous section), consisting of 65 social media targeting gigs, 47 search engine targeting gigs, and 9 user traffic targeting gigs. The classifier classified the 19,904 gigs as 14,065 social media targeting gigs (70.7%), 5,438 search engine targeting gigs (27.3%), and 401 user traffic targeting gigs (2%). We manually verified that these classifications were correct by scanning the titles of the gigs. Next, we will present our case studies for each of the three types of crowdturfing gigs.



(a) Sellers. (b) Buyers.

Fig. 6 Top 10 countries of sellers and buyers in crowdturfing gigs.

Social media targeting gigs. In Figure 7, we identify the social media sites (including social networking sites) that were targeted the most by the crowdturfing sellers. Overall, most well known social media sites were targeted by the sellers. Among the 14,065 social media targeting gigs, 7,032 (50%) and 3,744 (26.6%) gigs targeted Facebook and Twitter, respectively. Other popular social media sites such as Youtube, Google+, and Instagram were also targeted. Some sellers targeted multiple social media sites in a single crowdturfing gig. Example titles for these social media targeting gigs are “I will deliver 100+ real fb likes from france to you facebook fanpage for $5” and “I will provide 2000+ perfect looking twitter followers without password in 24 hours for $5”.

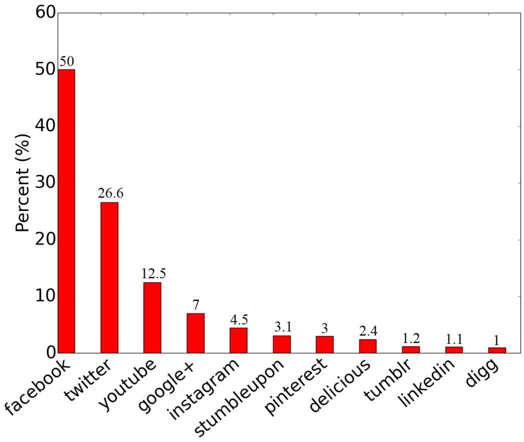


Fig. 7 Social media sites targeted by the crowdturfing sellers.

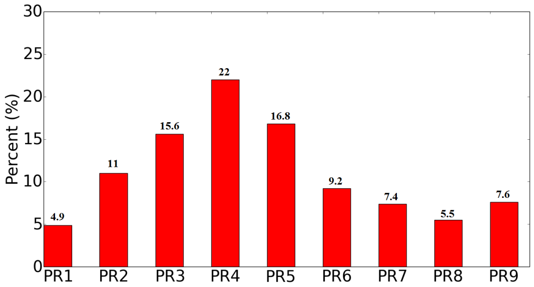


Fig. 8 PageRank scores of web pages managed by the crowdturfing gig sellers and used to link to a buyer's web page.

Search engine targeting gigs. People operating a company always have a desire for their web site to be highly ranked in search results generated by search engines such as Google and Bing. The web site's rank order affects the site's profit since web surfers (users) usually only click the top results, and click-through rates of the top pages decline exponentially from #1 to #10 positions in a search result [19]. One popular way to boost the ranking of a web site is to get links from other web sites because search engines measure a web site's importance based on its link structure. If a web site is cited or linked by a well known web site such as cnn.com, the web site will be ranked in a higher position than before. Google's famous ranking algorithm, PageRank, is computed based on the link structure and quality of links. To artificially boost the ranking of web sites, search engine targeting gigs provide a web site linking service. Example titles for these gigs are "I will build a Linkwheel manually from 12 PR9 Web20 + 500 Wiki Backlinks+Premium Index for $5" and "I will give you a PR5 EDUCATION Nice permanent link on the homepage for $5".

As shown in the examples, the sellers of these gigs shared a PageRank score for the web pages that would be used to link to buyers' web sites. PageRank score ranges between 1 and 9, and a higher score means the page's link is more likely to boost the target page's ranking. To understand what types of web pages the sellers provided, we analyzed the titles of the search engine targeting gigs. Specifically, titles of 3,164 (58%) of the 5,438 search engine targeting gigs explicitly contained a PageRank score of their web pages so we extracted PageRank scores from the titles and grouped the gigs by a PageRank score, as shown in Figure 8. The percentage of web pages between PR1 and PR4 increased from 4.9% to 22%. Then, the percentage of web pages between PR5 and PR8 decreased because owning or managing higher PageRank pages is more difficult. Surprisingly, the percentage of PR9 web pages increased. We conjecture that the buyers owning PR9 pages invested time and resources carefully to maintain highly ranked pages because they knew the corresponding gigs would be more popular than others (and much more profitable).

User traffic targeting gigs. Web site owners want to increase the number of visitors to their sites, called "user traffic", to maximize the value of the web

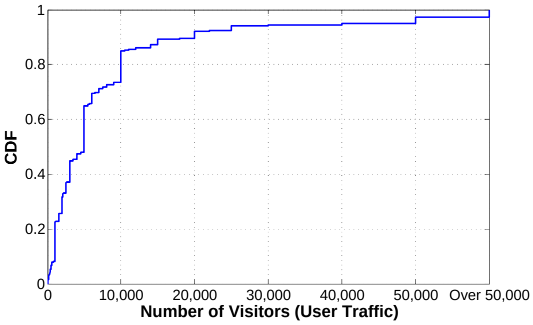


Fig. 9 The number of visitors (User Traffic) provided by the sellers.

site and its revenue. Ultimately, they want these visitors buy products on the site or click advertisements. For example, owners can earn money based on the number of clicks on advertisements supplied from Google AdSense [10]. 401 crowdturfing gigs fulfilled these owners' needs by passing user traffic to buyers' web sites.

An interesting research question is, "How many visitors does a seller pass to the destination site of a buyer?" To answer this question, we analyzed titles of the 401 gigs and extracted the number of visitors by using regular expressions (with manual verification). 307 of the 401 crowdturfing gigs contained a number of expected visitors explicitly in their titles. To visualize these numbers, we plotted the cumulative distribution function (CDF) of the number of promised visitors in Figure 9. While 73% of sellers guaranteed that they will pass less than 10,000 visitors, the rest of the sellers guaranteed that they will pass 10,000 or more visitors. Even 2.3% of sellers advertised that they will pass more than 50,000 visitors. Examples of titles for these user traffic targeting gigs are "I will send 7000+ Adsense Safe Visitors To Your Website/Blog for $5" and "I will send 15000 real human visitors to your website for $5". By only paying $5, the buyers can get a large number of visitors who might buy products or click advertisements on the destination site.

In summary, we identified 19,904 (22.2%) of the 89,667 gigs as crowdturfing tasks. Among those gigs, 70.7% targeted social media sites, 27.3% targeted search engines, and 2% passed user traffic. The case studies reveal that crowdturfing gigs can be a serious problem to the entire web ecosystem because malicious users can target any popular web service.

7 Impact of Crowdturfing Gigs

Thus far, we have studied how to detect crowdturfing gigs and presented case studies for three types of crowdturfing gigs. We have also hypothesized that crowdturfing gigs pose a serious threat, but an obvious question is whether they actually affect to the web ecosystem. To answer this question, we measured the real world impact of crowdturfing gigs. Specifically, we purchased a

Table 5 The five gigs' sellers, the number of followers sent by these sellers and the period time took to send all of these followers.

Seller Name	|Sent Followers|	The Period of Time
spyguyz	5,502	within 5 hours
tweet_retweet	33,284	within 47 hours
fiver_expert	5,503	within 1 hour
sukmoglea4863	756	within 6 hours
myeasycache	1,315	within 1 hour

few crowdturfing gigs targeting Twitter, primarily because Twitter is one of the most targeted social media sites. A common goal of these crowdturfing gigs is to send Twitter followers to a buyer's Twitter account (i.e., artificially following the buyer's Twitter account) to increase the account's influence on Twitter.

To measure the impact of these crowdturfing gigs, we first created five Twitter accounts as the target accounts. Each of the Twitter accounts has a profile photo to pretend to be a human's account, and only one tweet was posted to each account. These accounts did not have any followers, and they did not follow any other accounts to ensure that they are not influential and do not have any friends. The impact of these crowdturfing gigs was measured as a Klout score, which is a numerical value between 0 and 100 that is used to measure a user's influence by Klout[3]. The higher the Klout score is, the more influential the user's Twitter account is. In this setting, the initial Klout scores of our Twitter accounts were all 0.

Then, we selected five gigs that claimed to send followers to a buyer's Twitter account, and we purchased them, using the screen names of our five Twitter accounts. Each of the five gig sellers would pass followers to a specific one of our Twitter accounts (i.e., there was a one seller to one buyer mapping). The five sellers' Fiverr account names and the titles of their five gigs are as follows:

spyguyz I will send you stable 5,000 Twitter FOLLOWERS in 2 days for $5
tweet_retweet I will instantly add 32000 twitter followers to your twitter account safely $5
fiver_expert I will add 1000+ Facebook likes Or 5000+ Twitter follower for $5
sukmoglea4863 I will add 600 Twitter Followers for you, no admin is required for $5
myeasycache I will add 1000 real twitter followers permanent for $5

These sellers advertised sending 5,000, 32,000, 5,000, 600 and 1,000 Twitter followers, respectively. First, we measured how many followers they actually sent us (i.e., do they actually send the promised number of followers?), and then, we identify how quickly they sent the followers. Table 5 presents the experimental result. Surprisingly, all of the sellers sent a larger number of followers than they originally promised. Even tweet_retweet sent almost 33,000

[3] http://klout.com

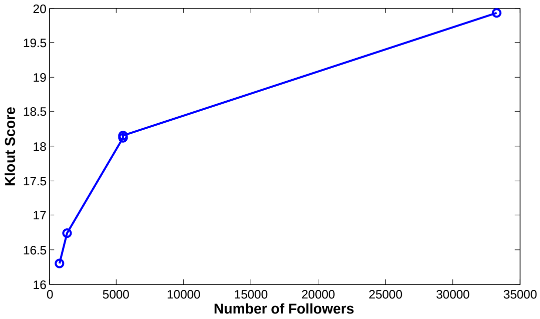


Fig. 10 Klout scores of our five Twitter accounts were correlated with number of followers of them.

followers for just $5. While tweet_retweet sent the followers within 47 hours (within 2 days, as the seller promised), the other four sellers sent followers within 6 hours (two of them sent followers within 1 hour). In summary, we were able to get a large number of followers (more than 45,000 followers in total) by paying only $25, and these followers were sent to us very quickly.

Next, we measured the impact of these artificial Twitter followers by checking the Klout scores for our five Twitter accounts (again, our Twitter accounts' initial Klout scores were 0). Specifically, after our Twitter accounts received the above followers from the Fiverr sellers, we checked their Klout scores to see whether artificially getting followers improved the influence of our accounts. In Klout, the higher a user's Klout score is, the more influential the user is [14]. Surprisingly, the Klout scores of our accounts were increased to 18.12, 19.93, 18.15, 16.3 and 16.74, which corresponded to 5,502, 33,284, 5,503, 756 and 1,316 followers. From this experimental result, we learned that an account's Klout score is correlated with its number of followers, as shown in Figure 10. Apparently, getting followers (even artificially) increased the Klout scores of our accounts and made them more influential. In summary, our crowdsourced manipulations had a real world impact on a real system.

The Followers Suspended By Twitter. Another interesting research question is, "Can current security systems detect crowdturfers?". Specifically, can Twitter's security system detect the artificial followers that were used for crowdsourced manipulation? To answer this question, we checked how many of our new followers were suspended by the Twitter Safety team two months after we collected them through Fiverr. We accessed each follower's Twitter profile page by using Twitter API. If the follower had been suspended by Twitter security system, the API returned the following error message: "The account was suspended because of abnormal and suspicious behaviors". Surprisingly, only 11,358 (24.6%) of the 46,176 followers were suspended after two months. This indicates that Twitter's security system is not effectively detecting these manipulative followers (a.k.a. crowdturfers). This fact confirms that the web ecosystem and services need our crowdturfing task detection system to detect crowdturfing tasks and reduce the impact of these tasks on other web sites.

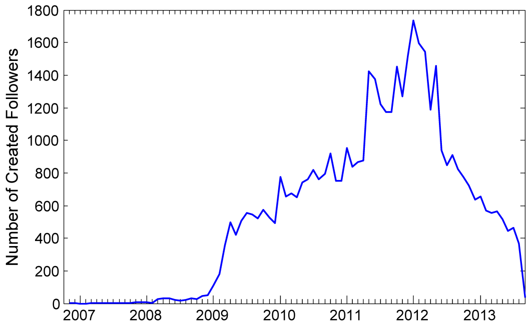


Fig. 11 Number of Twitter worker accounts created in each month

8 Analysis and Detection of Twitter Workers

So far, we verified the impact of crowdturfing gigs targeting Twitter by measuring Klout scores, and we found that current Twitter security systems are not effective for detecting the paid followers. In this section, we analyze the behaviors and characteristics of these Twitter workers (i.e., the paid followers)[4], and we build classifiers to automatically detect these workers.

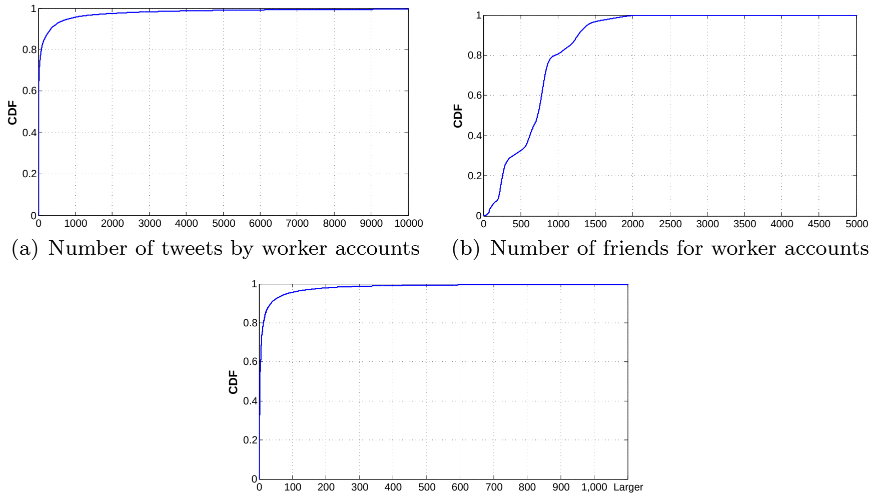
8.1 Characteristics of Twitter Workers

First, we analyze the characteristics of 46,176 Twitter workers that followed our target accounts. Specifically, we answer a number of research questions. When were Twitter worker accounts created? How many Tweets have they posted? How many followers and friends do they have? Did the accounts follow each other?

When were Twitter worker accounts created? To answer this research question, we analyzed creation dates for these worker acccounts. Figure 11 depicts the number of worker accounts created in each month. Interestingly, half of the worker accounts were created before August 2011, and the most popular month was January 2012 (i.e., the largest number of Twitter workers were created during this month). 354 accounts were created before 2009, which suggests they were carefully managed by sellers to avoid Twitter's spam detection systems. Alternatively, some of these long-lived worker accounts might be legitimate accounts that were compromised to perform crowdturfing tasks.

Posting activity, friends and followers. Next, we analyze the cumulative distributions of the number of posted tweets, friends, and followers for worker accounts (as shown in Figure 12). The minimum and maximum number of posted tweets among the workers were 0 and 93,745, respectively. 50% of the workers posted less than 5 tweets, and about 82% of the workers posted less than 100 tweets. This result indicates that most workers do not actively post

[4] We refer to paid followers as workers for the remainder of this manuscript.



(a) Number of tweets by worker accounts (b) Number of friends for worker accounts

(c) Number of followers for worker accounts

Fig. 12 CDFs for the number of tweets, friends and followers for worker accounts

tweets (i.e., they are typically only active when performing crowdturfing tasks). In Figure 12(b) and 12(c), we can observe that the number of friends for the workers were larger than the number of followers (by an order of magnitude, in many cases). We conjecture that these worker accounts have been used to follow other users, in an attempt to expand their network reach. Specifically, 50% of these workers followed more than 730 users, and the top 10% workers followed more than 1,260 users. In contrast, 50% of these workers had less than 4 followers, and 92% of these workers had less than 50 followers (7,232 workers did not have any followers). We also measured a ratio for the number of friends and followers for each worker. More than 90% of the workers had a ratio higher than 10, which means they had at least 10 times more friends than followers.

Changing number of friends over time. The unbalanced number of friends and followers for each worker motivated us to investigate how the number of friends for these workers has been evolving over time. Our analysis revealed that many workers frequently change their friend counts over time. Figure 13 depicts the friend count evolution for two specific workers. As the figure shows, these workers quickly followed numerous accounts before suddenly unfollowing a large percentage of those accounts. Workers engage in this following/unfollowing behavior because it helps bolster their own follower counts (some users follow back the workers as a courtesy). We conjecture that workers need more followers to expand their network reach and to be able to follow even more accounts. If a worker follows too many accounts in a short time period, the Twitter Safety team will easily identify the worker as abnormal and suspend the account [26]. However, if the number of friends and followers remains roughly equivalent, the worker account will not be suspended,

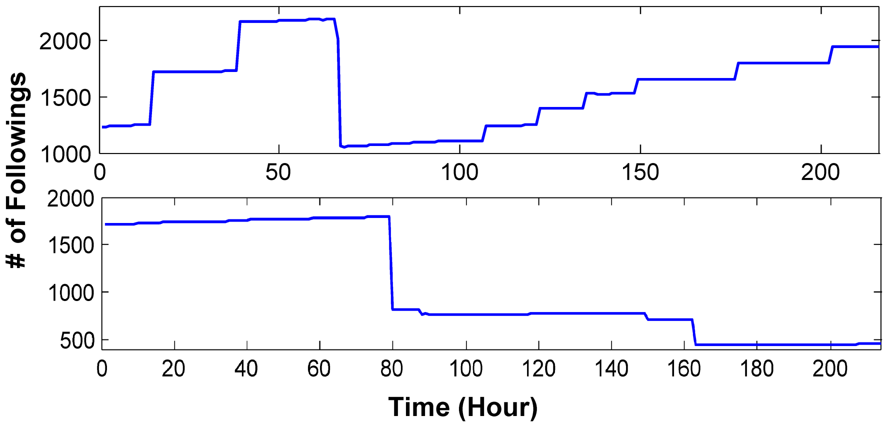


Fig. 13 Evolving number of friends for two workers over a period of time

despite the volatile friend count. It's important to note that this behavior was also observed for social spammers [15].

Graph density. Next, we measured the graph density of the worker accounts. By performing this measurement, we can determine if the workers follow each other, and we can observe specific connections between the accounts? In our graph, each worker is a node (vertex), and we add an edge between two workers if one of them follows the other. The graph's density was measured by $\frac{|E|}{|V| \times |V-1|}$, where V and E represent the number of nodes and edges, respectively, in the graph. Given 46,176 workers, there were 193 edges. This graph's density is 0.0000000905, which is smaller than the average graph density of 0.000000845 for normal accounts on Twitter [31]. In other words, these workers are more sparsely connected than normal users. This observation makes sense because most of the workers exclusively followed customer accounts (i.e., buyers) instead of following each other.

Digging a little deeper, what happens if we measure the graph density for each subgraph containing a set of worker accounts belonging to each Fiverr seller? To answer this question, we measured the graph density for workers associated with each of the five sellers we originally used to purchase gigs. Each seller's worker graph density was 0.00000009911, 0.0000001462, 0.000000132, 0.0000035, and 0. Interestingly, the graph density for workers from each seller except the last one (whose graph density is 0) was slightly larger than the graph density of all of the 46,176 workers. This means at least some of the workers from each seller (except the last one) followed each other. However, each seller's worker graph density is still smaller than the average graph density associated with normal Twitter accounts. This result also makes sense because each seller earns money by following customers, and there is little to no incentive for a seller to make his worker accounts follow each other.

Table 6 Twitter dataset.

Class	|User Profiles|	|Tweets|
Workers	46,176	1,760,580
Legitimate Users	82,775	15,815,141

8.2 Detection of the Twitter Workers

In the previous section, we analyzed the behaviors of Twitter workers and observed unique patterns that might distinguish them from legitimate (normal) user accounts. In this section, we leverage those observations to build worker detection classifiers that automatically detect these workers using only their Twitter information.

Twitter Dataset. In order to build worker detection classifiers, we need Twitter information for workers and legitimate user accounts. Using Twitter's streaming API, we randomly selected 86,126 accounts, and then, we used Twitter's Rest APIs to collect each account's profile, recent tweets, and friendship list with temporal friendship information (to measure the evolution of friend and follower counts over time). To make sure the 86,126 accounts are not spammers or malicious participants, we checked their account status once per hour. Based on these checks, we removed 3,351 out of 86,126 user accounts because they were either suspended by Twitter or deleted by users. Our final Twitter dataset contained 82,775 legitimate user accounts and 46,176 workers, as shown in Table 6.

Training and Testing Sets. To build and test a classifier, we randomly split the Twitter dataset into training and testing sets. The training set contains 2/3 data, and the testing set contains the remaining 1/3 data. The two sets were stratified and contained the same ratio of workers and legitimate users. Specifically, the training set consists of 30,784 workers and 55,183 legitimate users, and the testing set consists of 15,392 workers and 27,592 legitimate users.

Features. To train a classifier, we need to convert each account's raw data into feature values. The performance of the classifier is dependent on high quality features that have distinguishing power for workers and legitimate users. Based on our previous analysis and observations, we created 103 features and grouped them into the following 7 categories:

- **Profile Features (PF):** extracted from an account's profile (e.g., the longevity of the account and whether it has a description in its profile).
- **Activity Features (AF):** measure a user's activity patterns. Examples of these features include how often a user posts a tweet with a URL, how many tweets a user posts daily, and the tweeting steadiness, which was computed as the standard deviation of the elapsed time between consecutive tweets from the most recent 200 tweets.

- **Social Network Features (SNF):** extracted from an account's social network to understand the user's social relationship (e.g., number of friends and followers).
- **Content Features (CF):** extracted from a user's contents (tweets). One of these features is 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.
- **Klout Feature (KF):** extracted by Klout's service [13], which measures a user's influence. Given a user's Twitter id, the Klout API returns the user's Klout score.
- **Personality Features (PNF):** extracted features using Linguistic Inquiry and Word Count (LIWC), which is a standard approach for mapping text to psychologically-meaningful categories such as "Positive Emotions" and "Family" [20]. We may understand a user's personality based on word usage, similar to previous research on essays and blogs [7,9]. In particular, the LIWC 2001 dictionary defines 68 different categories, each of which contains dozens to hundreds of words. Given each user's tweets, we computed the user's score for each category based on the LIWC dictionary: (i) we counted the total number of words in the tweets (N); (ii) we counted the number of words in the tweets that overlapped with the words in each category i in the LIWC dictionary (C_i); and (iii) we computed the score for a category i as C_i/N. Finally, each category and the score for each category become a feature and its feature value, respectively. The personality features contained 68 features in total.
- **Temporal Features (TF):** extracted from snapshots of user profiles, which were saved once per hour. We measure the standard deviation (SD) for the number of friends over time, the number of followers over time, and their ratio. Additionally, since we are interested in how much a user's friend and follower counts changed between two consecutive snapshots, we measured the change rate (CR) as follows:

$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(t_{i+1}-t_i)}$$

  where n is the total number of recorded temporal information extracted from the snapshots, and t_i means temporal information of the user (e.g., number of friends) in the ith snapshot.

The detailed information regarding the 103 features is presented in Table 7.

Feature Selection. Next, we computed the χ^2 value [32] for each of the features to see whether all features are positively contributing to build a good classifier. The larger the χ^2 value is, the higher discriminative power the corresponding feature has. The results showed that all features had positive discrimination power in spite of different relative strengths. Table 8 shows the top 10 features with average feature values for workers and legitimate users, which illustrate how behaviors of workers and legitimate users are quite distinct. For

Table 7 Features.

Group	Feature
PF	the length of the screen name
PF	the length of the description
PF	the longevity of the account
PF	has description in profile
PF	has URL in profile
AF	the number of posted tweets
AF	the number of posted tweets per day
AF	|links| in tweets / |tweets|
AF	|hashtags| in tweets / |tweets|
AF	|@username| in tweets / |tweets|
AF	|rt| in tweets / |tweets|
AF	|tweets| / |recent days|
AF	|links| in tweets / |recent days|
AF	|hashtags| in tweets / |recent days|
AF	|@username | in tweets / |recent days|
AF	|rt| in tweets in tweets / |recent days|
AF	|links| in RT tweets / |RT tweets|
AF	tweeting steadiness
SNF	the number of friends
SNF	the number of followers
SNF	the ratio of the number of friends and followers
SNF	the percentage of bidirectional friends: $\frac{\lvert friends \cap followers \rvert}{\lvert friends \rvert}$ and $\frac{\lvert friends \cap followers \rvert}{\lvert followers \rvert}$
SNF	standard deviation (SD) for followee IDs
SNF	standard deviation (SD) for follower IDs
CF	the average content similarity over all pairs of tweets posted: $\frac{\sum similarity(a,b)}{\lvert \text{set of pairs in tweets} \rvert}$, where $a, b \in$ set of pairs in tweets
CF	the ZIP compression ratio of posted tweets: $\frac{uncompressed\ size\ of\ tweets}{compressed\ size\ of\ tweets}$
PNF	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
KF	Klout score
TF	standard deviation (SD) for number of friends over time
TF	standard deviation (SD) for number of followers over time
TF	ratio of standard deviation (SD) for number of friends over time and standard deviation (SD) for number of followers over time
TF	the change rate (CR) for number of friends over time
TF	the change rate (CR) for number of followers over time
TF	ratio of the change rate (CR) for number of friends over time and the change rate (CR) for number of followers over time

Table 8 Top 10 features.

Feature	Workers	Legitimate
ratio of SD for |friends| over time and SD for |followers| over time	57.52	1.17
ratio of |friends| and |followers|	311	1.7
Klout score	6.7	34.5
the CR for number of friends over time	3.8	0.9
the number of posted tweets per day	0.3	20.8
SD for number of friends over time	30	4.8
$\frac{\|friends \cap followers\|}{\|friends\|}$	0.022	0.599
the CR for number of followers over time	0.085	0.901
the number of posted tweets	258	14,166
|tweets| / |recent days|	2.1	19.5

example, workers have a much larger ratio of SD for |friends| over time and SD for |followers| over time than legitimate users (57.52 vs. 1.17). This indicates that workers regularly increase and decrease their friend counts (as shown in Figure 13). Workers also have lower Klout scores than legitimate users (6.7 vs. 34.5), and workers post fewer tweets per day than legitimate users (0.3 vs. 20.8). Overall, this feature selection study clearly shows that workers exhibit different characteristics than legitimate users.

Predictive Models and Evaluation Metrics. We computed feature values for each user in the training and testing sets, according to the previously described features. Then, we selected five popular classification algorithms: J48, Random Forest, SMO (SVM), Naive Bayes and Logistic Regression. Using the Weka machine learning toolkit's implementation of these algorithms, we developed five classifiers to predict whether a user is a worker or a legitimate user. For evaluation, we used the same metrics described in Section 5.2 (e.g., accuracy, F_1 measure, FPR, and FNR).

Experimental Results. Each of the five trained classifiers classified each of the users in the testing set consisting of 15,392 workers and 27,592 legitimate users as either a worker or a legitimate user. Experimental results are shown in Table 9. All of the classifiers achieved over 96% accuracy, which is much higher than the 64.19% accuracy of the baseline approach measured by assigning all of the users in the testing set to the majority class (legitimate users). Among the five classifiers, Random Forest outperformed the others, achieving 99.28% accuracy, 0.993 F_1 measure, 0.004 FPR and 0.013 FNR. This result proves that automatically detecting workers is possible, and our classification approach successfully detected workers.

9 Conclusion

In this manuscript, we presented a comprehensive analysis of gigs and users in Fiverr, and we identified three types of crowdturfing gigs: social media targeting gigs, search engine targeting gigs and user traffic targeting gigs. Based

Table 9 The performance result of Classifiers

Classifier	Accuracy	F_1	FPR	FNR
J48	99.1%	0.991	0.007	0.012
Random Forest	**99.29%**	0.993	0.005	0.011
SMO (SVM)	98.26%	0.983	0.008	0.034
Naive Bayes	96.26%	0.963	0.04	0.033
Logistic Regression	98.18%	0.982	0.008	0.036

on this analysis, we proposed and developed statistical classification models to automatically differentiate between legitimate gigs and crowdturfing gigs, and we provided the first study to detect crowdturfing tasks automatically. Our experimental results show that these models can effectively detect crowdturfing gigs with an accuracy rate of 97.35%. Using these classification models, we identified 19,904 crowdturfing gigs in Fiverr, and we found that 70.7% were social media targeting gigs, 27.3% were search engine targeting gigs, and 2% were user traffic targeting gigs. Then, we presented detailed case studies that identified important characteristics for each of these three types of crowdturfing gigs.

We also measured the real world impact of crowdturfing by purchasing active Fiverr crowdturfing gigs that targeted Twitter. The purchased gigs generated tens of thousands of artificial followers for our Twitter accounts. Our experimental results show that these crowdturfing gigs have a tangible impact on a real system. Specifically, our Twitter accounts were able to obtain increased (and undeserved) influence on Twitter. We also tested Twitter's existing security systems to measure their ability to detect and remove the artificial followers we obtained through crowdturfing. Surprisingly, after two months, the Twitter Safety team was only able to successfully detect 25% of the artificial followers.

Finally, to complement existing Twitter security systems, we analyzed characteristics of 46,176 paid Twitter workers, found distinguishing patterns between the paid Twitter workers and 82,775 legitimate Twitter users, and built classifiers to automatically detect Twitter workers. Our experimental results show that the classifiers successfully detect Twitter workers, achieving 99.29% accuracy. In the near future, we plan to widely deploy our detection system and greatly reduce the impact of these crowdturfing tasks on other sites.

10 Acknowledgements

This work was supported in part by Google Faculty Research Award, Research Catalyst grant and faculty startup funds from Utah State University, and 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

1. Alexa: Fiverr.com site info - alexa. http://www.alexa.com/siteinfo/fiverr.com (2013)
2. Allahbakhsh, M., Benatallah, B., Ignjatovic, A., Nezhad, H.R.M., Bertino, E., Dustdar, S.: Quality control in crowdsourcing systems: Issues and directions. IEEE Internet Computing **17**(2), 76–81 (2013). URL http://dblp.uni-trier.de/db/journals/internet/internet17.html#AllahbakhshBIMBD13
3. Baba, Y., Kashima, H., Kinoshita, K., Yamaguchi, G., Akiyoshi, Y.: Leveraging non-expert crowdsourcing workers for improper task detection in crowdsourcing marketplaces. Expert Syst. Appl. **41**(6), 2678–2687 (2014). URL http://dblp.uni-trier.de/db/journals/eswa/eswa41.html#BabaKKYA14
4. Bank, T.W.: Doing business in moldova - world bank group. http://www.doingbusiness.org/data/exploreeconomies/moldova/ (2013)
5. Bernstein, M.S., Little, G., Miller, R.C., Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D., Panovich, K.: Soylent: A word processor with a crowd inside. In: UIST (2010)
6. Chen, C., Wu, K., Srinivasan, V., Zhang, X.: Battling the internet water army: Detection of hidden paid posters. CoRR **abs/1111.4297** (2011)
7. Fast, L.A., Funder, D.C.: Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. Journal of personality and social psychology **94**(2), 334 (2008)
8. Franklin, M.J., Kossmann, D., Kraska, T., Ramesh, S., Xin, R.: Crowddb: Answering queries with crowdsourcing. In: SIGMOD (2011)
9. Gill, A.J., Nowson, S., Oberlander, J.: What are they blogging about? personality, topic and motivation in blogs. In: ICWSM (2009)
10. Google: Google adsense ? maximize revenue from your online content. www.google.com/adsense (2013)
11. Halpin, H., Blanco, R.: Machine-learning for spammer detection in crowd-sourcing. In: Human Computation workshop in conjunction with AAAI (2012)
12. Heymann, P., Garcia-Molina, H.: Turkalytics: Analytics for human computation. In: WWW (2011)
13. Klout: Klout — the standard for influence. http://klout.com/ (2013)
14. Klout: See how it works. - klout. http://klout.com/corp/how-it-works (2013)
15. Lee, K., Eoff, B.D., Caverlee, J.: Seven months with the devils: A long-term study of content polluters on twitter. In: ICWSM (2011)
16. Lee, K., Tamilarasan, P., Caverlee, J.: Crowdturfers, campaigns, and social media: Tracking and revealing crowdsourced manipulation of social media. In: ICWSM (2013)
17. Lee, K., Webb, S., Ge, H.: The dark side of micro-task marketplaces: Characterizing fiverr and automatically detecting crowdturfing. In: ICWSM (2014)
18. Motoyama, M., McCoy, D., Levchenko, K., Savage, S., Voelker, G.M.: Dirty jobs: The role of freelance labor in web service abuse. In: USENIX Security (2011)
19. Moz: How people use search engines - the beginners guide to seo. http://moz.com/beginners-guide-to-seo/how-people-interact-with-search-engines (2013)
20. Pennebaker, J., Francis, M., Booth, R.: Linguistic Inquiry and Word Count. Erlbaum Publishers (2001)
21. Pham, N.: Vietnam admits deploying bloggers to support government. http://www.bbc.co.uk/news/world-asia-20982985 (2013)
22. Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., Tomlinson, B.: Who are the crowdworkers?: Shifting demographics in mechanical turk. In: CHI (2010)
23. Sterling, B.: The chinese online 'water army'. http://www.wired.com/beyond_the_beyond/2010/06/the-chinese-online-water-army/ (2010)
24. Stringhini, G., Wang, G., Egele, M., Kruegel, C., Vigna, G., Zheng, H., Zhao, B.Y.: Follow the green: growth and dynamics in twitter follower markets. In: IMC (2013)
25. Thomas, K., McCoy, D., Grier, C., Kolcz, A., Paxson, V.: Trafficking fraudulent accounts: the role of the underground market in twitter spam and abuse. In: USENIX Security (2013)
26. Twitter: The twitter rules. https://support.twitter.com/articles/18311-the-twitter-rules (2013)

27. Venetis, P., Garcia-Molina, H.: Quality control for comparison microtasks. In: Crowd-KDD workshop in conjunction with KDD (2012)
28. Wang, G., Mohanlal, M., Wilson, C., Wang, X., Metzger, M.J., Zheng, H., Zhao, B.Y.: Social turing tests: Crowdsourcing sybil detection. In: NDSS (2013)
29. Wang, G., Wilson, C., Zhao, X., Zhu, Y., Mohanlal, M., Zheng, H., Zhao, B.Y.: Serf and turf: crowdturfing for fun and profit. In: WWW (2012)
30. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). Morgan Kaufmann (2005). URL `http://www.cs.waikato.ac.nz/~ml/weka/book.html`
31. Yang, C., Harkreader, R., Zhang, J., Shin, S., Gu, G.: Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In: WWW (2012)
32. Yang, Y., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: ICML (1997)