Pay Me and I'll Follow You: Detection of Crowdturfing Following Activities in Microblog Environment

Yuli Liu, Yiqun Liu*, Min Zhang, Shaoping Ma
State Key Lab of Intelligent Technology & Systems; Tsinghua National TNLIST Lab
Department of Computer Science & Technology, Tsinghua University, Beijing, 100084, China
yiqunliu@tsinghua.edu.cn

Abstract

A number of existing works have focused on the problem of malicious following activity detection in microblog services. However, most of them make the assumption that the spamming following relationships are either from fraudulent accounts or compromised legitimate users. They therefore developed detection methodologies based on the features derived from this assumption. Recently, a new type of malicious crowdturfing following relationship is provided by the follower market, called voluntary following. Followers who provide voluntary following services (or named *volowers*) are normal users who are willing to trade their following activities for profit. Since most of their behaviors follow normal patterns, it is difficult for existing methods to detect volowers and their corresponding customers. In this work, we try to solve the voluntary following problem through a newly proposed detection method named *DetectVC*. This method incorporates both structure information in user following behavior graphs and prior knowledge collected from follower markets. Experimental results on large scale practical microblog data set show that *DetectVC* is able to detect volowers and their customers simultaneously and it also significantly outperforms existing solutions.

1 Introduction

The microblogging services such as Twitter and Weibo allow users to follow accounts in which they are interested, and then receive status updates shared by these accounts. In the following relationship, the one who follows others is usually called a *follower* while the one who is followed is called a *followee*.

With the continuing popularity of the microblog, the number of followers becomes an important metric of the influence and reputation of a person's or a business entity's account [Cha et al., 2010]. Some users want to gain more social attention by attracting more followers. Companies also hope to gain more fans to promote their products or services, enlarge their business networks or increase brand awareness [Stringhini et al., 2013; Aggarwal et al., 2014].

Usually, microblog users enlarge their follower populations by creating/forwarding interesting contents or establishing close online relationships with other users. However, for some users, this kind of legitimate efforts cannot meet their needs and they turn to the follower market [Stringhini et al., 2012] to get some undeserved followers. These microblog users who have bought traded followers from follower market to follow them are called customers [Stringhini et al., 2013]. The abnormal following activities provided by the follower market impose a great threat to Microblogging services, which disrupt fair following mechanisms and help customers to gain excessive reputation or influence. Malicious entities can also spread malwares and/or perform other spamming activities if they get a large number of followers through the follower market [Wang and Konolige, 2013].

Previous works have suggested that the underground follower market mainly provides follower populations in the forms of: i) fraudulent accounts (i.e., fake accounts or Sybils) [Almaatouq et al., 2014; Motoyama et al., 2011], which are usually created and manipulated by market operators for conducting spamming activities; ii) compromised accounts owned by legitimate users whose credentials have been stolen by spammers [Stringhini et al., 2012; Egele et al., 2013] to conduct spamming activities against their will.

Besides these two types of followers, in this paper we focus on the rise of a new type of malicious following activities oriented from normal users who are willing to join the follower market and get rewards by following customers. To our best knowledge, we found no previous works on this kind of malicious following activities and therefore name it as "voluntary following" relationship. We also call the corresponding users the voluntary followers (or *volowers*) due to their personal willingness to trade the following relationship for profit. In most cases, volowers' activities are normal, and the only difference between volowers and legitimate users is their abnormal following activities which do not always happen. In voluntary following activities, customers pay to get tradable followers; market operators take charge of management; and voluntary followers answer

*Corresponding author

calls by performing following activities. We believe that this kind of abnormal following activity is a crowdturfing activity [Wang and Mohanlal, 2013] that brings great challenges to the operation of microblog platforms.

Existing malicious following detection methods [Egele et al., 2013; Cheng et al., 2013; Singh et al., 2014; Xie. 2008; Thomas. 2013] are usually designed based on the abnormal behavior or demographic characteristics of malicious accounts. They usually get exhausted in dealing with continuously evolving cheating strategies and may not be able to detect volowers and their customers effectively. This is due to the fact that volowers appear to be normal users most of the time and are difficult to be separated from legitimate users. According to our observation on a set of 3,185 volowers for two months (from 2015/01/15 to 2015/03/14), only 3.05% of these volowers were suspended by the microblog platform, which indicates that it is hard to detect the voluntary following activities by existing strategies.

In the detection of voluntary following activities, it is also important to detect followers and customers simultaneously. The follower market is likely to lose its revenue stream and be eliminated promptly only when both traded followers and customers are detected and penalized. Therefore, how to detect voluntary followers and the customers at the same time also becomes our concern.

Observations suggest that the volower's goal is to follow enough customers for profit and the customer's motivation is to gain more followers for influence or reputation to achieve self-promotion or product-promotion [Stringhini et al., 2013]. In this work, taking advantage of the inherent motivation and goal of the volowers and customers, we propose a robust and effective method *DetectVC* to detect volowers and customers simultaneously without being burdened by spammers' simulated patterns that are constantly changing. *DetectVC* incorporates both graph structure of microblogging users' following relationships and prior knowledge collected from follower market. Each user is assigned two scores by *DetectVC*, which represent the probabilities of a user being a volower and a customer.

The main contributions of the paper are outlined below:

- To our knowledge, this is the first work to investigate an emerging crowdturfing following activity in microblogging environment named voluntary following.
- We first propose an effective method *DetectVC* to conduct voluntary follower and customer detection simultaneously without using the constantly changing cheating properties. The method is based on graph structure analysis and prior knowledge application.
- With extensive experiments on real-world microblog datasets, we show the robustness and effectiveness of our framework.

The rest of this paper is organized as follows. After reviewing the related work in the next section, we describe the data set preparation process. Then we introduce the properties of voluntary follower. After that, present our approach of voluntary following detection *DetectVC* and prove the convergence of the propagation process. The following section reports the experimental results on real-world datasets. Finally, the last section concludes the paper.

2 Related Work

With the development of online social networking (OSN), detection of malicious activities has been studied for years on various OSN platforms. In previous detection approaches, spammers are usually supposed to create fake accounts to seize private information or to promote advertisements for personal gains [Zhu et al., 2012]. Researchers have tried different techniques to detect these spammers, which can be roughly grouped into two categories: social graph-based methods [Zhu et al., 2012; Yang and Liu, 2012] and content-based methods [Ribeiro et al., 2015; Teraguchi et al., 2012]. Most of these studies have shown effectiveness in their experimental settings.

Abnormal following is one common spamming activity on OSN. For example, [Thomas et al., 2013] investigated the follower market for fraudulent Twitter accounts and developed a classifier to detect them. In [Cheng et al., 2013; Wang and Mohanlal, 2012], fraudulent accounts are detected based on the attribute analysis. Compromised followers have also been well researched. [Stringhini et al., 2013; Almaatouq et la., 2014; Egele et al., 2013] gave a composition of statistical modeling compromised accounts. Besides these efforts, [Stringhini et al., 2013] presented a study of Twitter follower markets, reported in details on both the static and the dynamic properties of the customers in these markets, and tried to detect these activities based on the properties.

We can see that most existing detection methods are committed to detecting fake or compromised accounts which are manipulated by spammers relying on detailed cheating behaviors. These detection systems may not be suitable for detecting volowers. To fight against the newly emerging voluntary following problem, we propose to focus on the following relationship by propagating spamming intents discovered from existing follower market activities.

In recent years, popular Internet services have shown that remarkable things can be achieved by harnessing the power of the masses using crowd-sourcing systems [Wang et al., 2012]. However, malicious tasks can also performed by real humans en masse through malicious crowd-sourcing systems, which will pose challenges to existing security mechanisms. [Wang et al., 2012] refer to malicious crowdsourcing systems as *crowdturfing* which is a portmanteau of "crowd-sourcing" and "astroturfing".

Researchers began studying crowdturng problems and market. [Motoyama and McCoy, 2011] analyzed abusive tasks on Freelancer. [Wang et al., 2012] analyzed two largest crowdturfing sites in China reveals that $4 million dollars have already been spent on these two sites alone. [Lee et al., 2013] develop a framework for "pulling back the curtain" on crowdturfers to reveal their underlying ecosystem. [Lee et al., 2014] present a comprehensive analysis of crowdsourcing sites (i.e., Fiverr) and build crowdturfing task detection classifiers to filter these tasks.

Compared with the previous crowdturng works, we mainly study the emerging crowdturfing following activities on microblogging services and develop a unified and flexible framework to detect the two types of involved users (i.e., customers and volowers).

3 Data Preparation

In this section, we present our efforts for sampling a large set of microblog accounts U, which are grouped into labeled and unlabeled categories. The labeled user set contains the legitimate user set U_n and volower set U_v. The unlabeled user set U_u includes all the users who are the neighbors of the labeled accounts U_n and U_v. The dataset is collected from Weibo in January, 2015 through a crowd crawling platform named Pameng[1]. The datasets used in this paper are publicly available[2].

3.1 Volower

To obtain a set of volowers, we purchase a few followers from a popular follower market (named Weibo Fans) in Taobao.com which is the most popular E-commerce site in China. The market provides several different kinds of follower accounts for sale, including fake followers who are mostly zombie users, comprised followers who may unfollow if they realize the following activity is against their will and volowers. To focus on the voluntary following problem, we ask for more information about the volowers. The market owner claims that the volowers are high-quality "DaRen" users, which means that the accounts have been registered for a long time and interact with other users quite frequently, and that they will not unfollow, which confirms that the volowers are willing to follow customers.

We create 3 new Weibo accounts as customers and then spend 300 RMB yuan to purchase 3,000 volowers, with each created customer getting 1,000 followers. After our purchase, market operators provide us with a total of 3,185 followers, including 185 complimentary ones. These accounts are identified as ground-truth volowers. The traded following activities lasted about one day.

Considering the fact that volowers follow customers for profit only, volowers' behavior patterns have no relation to the purpose of customers. Therefore, behaviors of the labeled volowers can be generalized to other voluntary followers in the following market, which means that our proposed volower sampling method doesn't hurt generality.

3.2 Legitimate User

To collect a sample set of legitimate users, we identify around 200 accounts as U_1 from the authors' close friends on Weibo platform who are highly unlikely to be volowers/customers. We also select about 500 verified users[3] as U_2 because they are usually reputable accounts on the Weibo platform and are not likely to be volowers/customers. Since the users in U_1 and U_2 are unlikely to follow spammers, we obtain the followees of them as U_3 and use the union of the three sets as the normal user set U_n, which includes about thirty thousand accounts.

Through the sampling process, we collect both ordinary users and verified users for the legitimate user set. It has a good representative of the whole set of normal users in microblog environment. Although it is inevitable that the sampling is not fully randomized, we can ensure these users are normal. The detection performance will not be evaluated on this set and the analyzed methodology itself can also be transferred to other data sets if we could get a more uniform sampling of legitimate users.

3.3 Unlabeled User

All the labeled users' neighbors (followees and followers) are collected as the unlabeled user set U_u to build connections between labeled users and get a more complete social graph, which contains about millions of accounts.

Customers are the users who have bought volowers as followers, so they would be in the followee lists of the volowers and are necessarily contained in U_u. The labeling process of customers from the data set will be introduced in Section 6.1.

4 Voluntary Follower Properties

To investigate the difference between voluntary followers and legitimate users, we compare some of their properties in Table 1. We can see that accounts in U_v have registered for nearly 900 days on average, which is almost as long as normal users. The numbers of messages and original messages (which exclude forwarding messages) are also close to normal users. Volowers have slightly fewer followers (251.1 v.s. 288.6) than normal users probably because the verified users in U_n have a large number of followers due to their high reputation. Originating from the goal of volowers, they follow much more users than legitimate users. The difference in the number of followees between legitimate users and suspicious followers is also observed in [Egele et al., 2013; Aggarwal. 2014]. However, the other properties of volowers are much more similar to legitimate users. We can even find that the volowers' per message obtains much more interactions (e.g. forward and comment) compared with normal users. This shows that volowers are even more active and popular in interacting with their neighbors.

These properties indicate that voluntary follower accounts are owned by real users and can provide voluntary following services which are difficult to be detected, as claimed by market operators.

Fraudulent accounts are relatively easy to be detected, due to their obvious abnormal behaviors. Compromised followers are relatively harder to be detected because they are real accounts that have a long associated history and network relationships. Voluntary followers are also real accounts, however, different from compromised followers, volowers are willing to be inducted into the follower markets. They neither experience a sudden change in behaviors nor unfollow customers, which means that we cannot use the previously proposed detection methods for compro-

[1] The "crawl league" in Chinese, http://cnpameng.com/.
[2] http://www.thuir.cn/group/~yqliu/publications/ijcai2016.zip
[3] The users whose identities have been verified by Sina Weibo.

mised followers [Stringhini et al. 2012; Egele et al. 2013] to detect volowers.

Further analysis shows that the graph density of volowers is 1.5×10^{-5}, which is smaller than the average graph density of 4.2×10^{-3} for normal users. The graph's density was measured by $|E|/(|V|\times|V-1|)$, where V and E represent the number of nodes and edges. This result shows that the volowers are more sparsely connected to each other than normal users, so using spammers' close connections [Hu et al. 2013] is hard to detect volowers.

From above analysis, we can see that the characteristics of volowers make it rather difficult for existing methods to detect them and we should turn to other means to accomplish the detection task.

Table 1: Comparison of volowers and normal users.

	U_v	U_n
#Days since registration	882.4	934.4
#Message	519.1	588.2
#Original message	363.3	353.0
#Follower	251.1	288.6
#Followee	908.6	317.1
#Interaction per message	2.33	1.43

5 DetectVC Algorithm

We apply *DetectVC* algorithm to detect the participators of the crowdturfing following activity, i.e., the voluntary followers and customers. Before presenting our algorithm, we introduce some definitions that will be adopted in this work.

5.1 Definitions

Existing works show that, in microblogging services, a social network can be represented by a directed graph [Wang et al., 2010; Hu et al., 2013; Jiang et al., 2014]. We use graph $G = (U, E)$ to denote the social network, where nodes $u_i \in U$ represent microblog users, and each directed edge between two nodes $[u_i, u_j] \in E$ represents a following relation from u_i to u_j. Microblog users can unilaterally follow others without their prior permission, which is different from other social networks such as Facebook. In this graph, we do not have self-links, i.e., $u_i \neq u_j$.

The social graph G can be represented by its corresponding adjacency matrix $W \in \mathcal{A}^{n\times n}$, where n is the number of users. If u_i follows u_j, $W(u_i, u_j) = 1$, otherwise, $W(u_i, u_j) = 0$. The probability vectors P_v and P_c are used to denote all the users' probabilities of being volowers and customers respectively. The seed set $U_s \subset U_v$ is a small set of users that are randomly selected from labeled volowers U_v. These users are used as the prior knowledge of *DetectVC*. In practical applications, the seed set of volowers can be obtained from the follower market with a small amount of money, as the process described in Section 3.1.

In our framework, after the *DetectVC* algorithm eliminates, each user u will receive two scores $P_v(u)$ and $P_c(u)$ that indicate its possibilities of being a volower and a customer. Given $G = (U, E)$ and a set of seed users $U_s \subset U_v$, our goal is to estimate the spam probabilities of $P_v(u)$ and $P_v(u)$ for each user $u \in U$.

5.2 Assumptions

Considering the inherent motivation of the volowers and customers, we propose two assumptions:

Assumption 1 *A customer will be followed by many volowers to gain influence or reputation.*

Assumption 2 *A volower will follow many customers to gain enough profit.*

Voluntary following has become a business. Volowers get paid to follow customers. Such users cannot just follow a single customer as they would not make much money that way. Besides, customers would like to buy a large number of followers because they want to gain much influence or reputation in an easy way. So there are strong connections between volowers and customers. Based on these connections and the analysis of the graph structure of users' following relationships, we design *DetectVC*.

Using collected voluntary followers in follower market as seeds, we can propagate the anomalous following intents on the graph constructed through users' following relationships. Through recursive propagation procedure, the strong connections between volowers and customers will help mutually reinforce the spam probabilities (i.e., $P_v(u)$ and $P_c(u)$) of volowers and customers.

5.3 Voluntary Following Detection Algorithm

In this work, we propose a *DetectVC* algorithm to solve the problem of voluntary following activities. For every user u, we could calculate the probability $P_v(u)$ that u is a volower by incorporating all of the spam information of its neighbors. Similarly, we could calculate each user's customer probability $P_c(u)$. Formally procedure description is shown as follows.

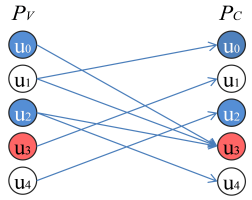


Figure 1: An example of *DetectVC* calculation.

To drive the propagation procedure of *DetectVC* and exploit the prior knowledge (i.e., seed volowers), we assign all of the users in the seed set U_s with initial spam probabilities (i.e., $P_v^{(0)}(u)=1$, $P_c^{(0)}(u)=0$). Every user $u \notin U_s$ will be assigned $P_v^{(0)}(u)=0$ and $P_c^{(0)}(u)=0$. Then we have

$$P_c^{(k)}(u_i) = \sum_{u_j:(u_j,u_i)\in E} P_v^{(k-1)}(u_j) \qquad (1)$$

The customer probability $P_c(u_i)$ of user u_i is calculated

through the volower probabilities of all the users that follow u_i. Similarly, for each user, the volower probability $P_v(u_i)$ is

$$P_v^{(k)}(u_i) = \sum_{u_j:(u_i,u_j)\in E} P_c^{(k)}(u_j) \quad (2)$$

The volower probability of user u_i is calculated by the customer probabilities of all the users that are followed by u_i.

Using the form of matrix/vector, in the k[th] iteration, equation (1) and (2) can be written as follows:

$$P_c^{(k)} = W^T P_v^{(k-1)} \text{ and } P_v^{(k)} = W P_c^{(k)} \quad (3)$$

Figure 1 indicates an example of *DetectVC* algorithm calculation through a toy graph. We suppose the blue nodes represent labeled volowers and the red nodes indicate customers. Then we get

$$W = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$P_v^{(0)} = (1,0,1,0,0)^{\mathrm{T}};\ P_c^{(1)} = (0,0,0,2,1)^{\mathrm{T}};\ P_v^{(1)} = (2,2,3,0,0)^{\mathrm{T}}$$

The *DetectVC* algorithm for volower and customer account detection is described in Algorithm 1. At each iteration, we reset the $P_v(u)$ as 1.0 for each u in U_s. The probability vectors P_v and P_c derived from each iteration are normalized by dividing their largest scores.

We now prove that our iterative *DetectVC* algorithm converges as k increases arbitrarily.

Proof. If we plug initial vector $P_v^{(0)}$ into Equation 3, then we have

$$\begin{aligned} P_c^{(k)} &= W^T \cdot P_v^{(k-1)} = W^T \cdot W \cdot P_c^{(k-1)} = W^T \cdot W \cdot W^T \cdot P_v^{(k-2)} \\ &= (W^T \cdot W)^2 \cdot P_c^{(k-2)} = \cdots = (W^T \cdot W)^{k-1} \cdot P_c^{(1)} \\ &= (W^T \cdot W)^{k-1} W^T \cdot P_c^{(0)} \end{aligned}$$

Similarly, we can get $P_v^{(k)} = (W \cdot W^T)^k \cdot P_c^{(0)}$. Based on the proofs of power method in linear algebra[4], we can know that the user spam probabilities matrix P_v and P_c converge and the convergence values are the principal eigenvectors of matrixes $(W \cdot W^T)$ and $(W^T \cdot W)$ respectively. ■

Algorithm 1: Volower and Customer Detection

Input: W: Adjacency matrix corresponds to social graph G;
U_S : Seed volowers;

1: **repeat**
2: **for** $u \in U_S$ set $P_v(u) = 1$;
3: $P_c = W^T P_v$;
4: $P_v = W P_c$;
5: **until** Convergence
6: **Output**: $P_v(u)$ and $P_c(u)$ for all the users.

6 Experimental Results and Discussions

To make comprehensive evaluation on how effective the proposed framework is in detecting voluntary followers and customers, we conduct a series of experiments. An important factor in our framework is the choice of volowers as seeds, because the spam probability of each user is inherently propagated from the selected seed users. To investigate the rationality of our seed selection and the robustness of our framework, we try our best to guarantee the randomness and diversity of seed selection.

We set m volowers in U_v as seed users, and reserve the remaining for evaluation, where m ranges from 100 to 1000 with a step size of 100. For each specific set number m, we randomly select the seed set with size m for 100 times. Each time after selecting volowers, we perform an experiment using them as the seed set U_s that drives the iteration and propagation process. So in general, one thousand (10×100) experiments are performed and corresponding performances are evaluated in this section.

6.1 Performance of Customer Detection

Each user will be assigned a score $P_c(u)$ to denote its customer probability when the iteration terminates. As described in section 3.3, customers are necessarily contained in our dataset. To evaluate the performance of our framework in customer detection, we rank all the users in descending order of $P_v(u)$ after an experiment being performed when m = 100. We randomly sample 2,000 users from top 40,000 users and manually label them as customers or non-customers. We also tried randomly sample 2,000 users from all the millions of users, but the number of customers is too small to analyze.

Inspired by existing researches, we extracted a list of signals to help judges label customers, e.g., (i) whether the account contains promotion channels [Li et al., 2015] (e.g. URL, phone number, and social media account) which lead to commercial intent Web sites in user profile. (ii) whether the account frequently posts messages that contain promotion intents [Aggarwal et al., 2014]. (iii) whether the account has much more followers than followees [Stringhini et al., 2013]. All these accounts' profiles, messages, followee and follower lists are provided as references for judges to label. We also encourage judges to use their own signals.

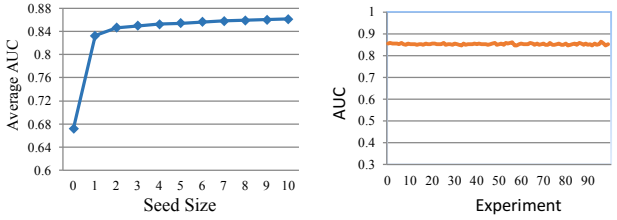
We ask three judges to annotate the 2,000 accounts. The judges have good knowledge of microblog environment and they have been informed of the phenomenon of voluntary following activity. If two or all of the judges deem an account as a customer, we label it as a customer. After labeling, 507 customers are identified from the sampled users. The average kappa coefficient of the three judges is 0.709, which means that the annotation is reliable and the customers are relatively easy to be identified manually.

With the labeled customer set, we evaluate the proposed method's performance of customer detection and the results are illustrated in Figure 2. As Figure 2(a) shows, when the size of seed set is 0 (i.e., without any prior knowledge and all the users with initial score 1) the AUC value is only about 0.68. It means that the prior knowledge incorporated in *DetectVC* is important. As the seed size (i.e., m) reaches 100, the performance of average AUC of 100 experiments

[4] http://distance-ed.math.tamu.edu/Math640/chapter6/node4.html

has largely improved to 0.84. Besides, while the m increases, the detection performance tends to rise in the beginning, but it then tends to keep stable. This result indicates that *DetectVC* is effective in customer detection and it does not require so many seed volowers (a few hundreds will be enough) to gain promising performance.

Figure 2(b) exhibits the changes of all the AUC values from 100 experiments when m = 300. As we can see, *DetectVC* method is robust to the change of seed set because the performance remains relatively stable in customer detection. It can achieve good performance without relying on the particular volower seeds.

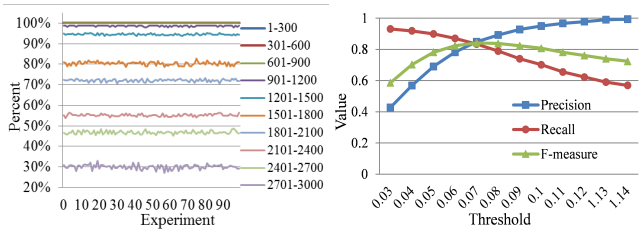


(a) Different seed sizes (b) Different experiments

Figure 2 Performance of customer detection.

6.2 Performance of Volower Detection

In this section, we use all the ground truth volowers in U_v except for the seed volowers to evaluate our method in volower detection. The assigned score $P_v(u)$ denotes the possibility of a user being a volower. All the users in our user set U are ranked in descending order of $P_c(u)$, and we use the top 3,000 users to show the effectiveness of the proposed method. The selected users are divided into ten buckets, each contains 300 users.



(a) Distribution of volowers. (b) Different thresholds

Figure 3 Performance of volower detection.

Figure 3(a) shows the proportion of volowers in each buckets of 100 experiments when m = 300 (m = 300 is a relatively stable parameter setting. We also try a number of other parameters from 100 to 1,000 and the results are similar). A line is composed of the volower proportions of all the 100 experiments of a bucket. As we can see, for each of the 100 experiments, almost 100% users are volowers in the top four buckets. With the fall of rankings, the corresponding percentage goes down. By the statistic, 85% of users are volowers among the top 3,000 users. That is, the volower will get a higher $P_v(u)$ value and the algorithm is effective. Besides, although each experiment is performed with different users as seeds, the difference among experiment results are very small, which means that *DetectVC* algorithm is also robust in volower detection.

We recognize the users with $P_v(u) > \theta_r$ as volowers. In volower detection task, we regard labeled volowers as positive samples and all the other collected users as negative samples. We adopt evaluation measures of Precision, Recall and F-measure to evaluate the effectiveness. The experimental results with different threshold settings are shown in Figure 3(b). It is shown that when θ_r = 0.07 our framework gets the best performance in volower detection, the F-measure is about 0.84.

6.3 Comparison with Baseline Methods

To make the statements above more convicing, we compare the performance of our approach with 5 popular existing malicious following activity detection methods. The first baseline [Yang et al., 2012] uses a measurement-based method to detect fraudulent accouts. The second one [Egele et al., 2013] is developed to identify compromised accounts through a composition of statistical modeling and anomaly detection. The third one [Lee et al., 2014] trains a classifier to detect the followers that might be legitimate accounts yet were compromised to perform crowdturing tasks. The fourth one [Stringhini et al., 2013] detects customers based on detailed properties. The fifth one [Aggarwal et al., 2015] builds a supervised learning model to predict suspicious following behaviours with the help of differentiating features. All these methods detect spammers based on various characteristics which represent state-of-the-art techniques.

We compare the performance of all baseline methods except the one proposed by [Stringhini et al., 2013] with the volower detection method of *DetectVC*. Meanwhile, the baselines from [Stringhini et al., 2013] and [Aggarwal et al., 2015] are compared with our customer detection method. Please be noted that only the method proposed by [Aggarwal et al., 2015] can be used to detect malicious followers and customers simultaneously as *DetectVC*. Our method can be further adopted as a discriminative feature to boost existing detection methods, so we integrate the spam probability $P_v(u)$ or $P_c(u)$ as a feature with corresponding baseline. The evaluation of customer detection is based on all the manually labeled customers and non-customers. Similar with volower detection evaluation, we set threhold θ_c with different scores to find the reasonable one.

From the results shown in Table 2 and Table 3, we can see that our proposed methods can achieve better performance in volower and customer detection compared with baselines. This shows that voluntary followers are difficult to be detected with existing strategies which are mainly designed to identify suspicious accounts based on discovered features, because volowers have less abnormal properties. Besides, our proposed detection method can be adopted to boost the performance of existing spam dete-ction approaches. This indicates that the spam probabilities generated by *DetectVC* are discriminate and promising

features to detect voluntary followers and customers.

Table 2: Comparison of F-measure scores between our volower detection method and baselines.

	Original	With $P_v(u)$
DetectVC	0.844	–
[Yang et al., 2012]	0.715	0.850 (+13.5%)
[Egele et al., 2013]	0.807	0.863 (+5.6%)
[Lee et al., 2014]	0.832	0.895 (+6.3%)
[Aggarwal et al., 2015]	0.825	0.868 (+4.3%)

Table 3: Comparison of F-measure scores between our customer detection method and baselines.

	Original	With $P_c(u)$
DetectVC	0.860	–
[Stringhini et al., 2013]	0.805	0.864 (+5.9%)
[Aggarwal et al., 2015]	0.837	0.907 (+7.0%)

6.4 Result Discussions

With the above experimental results, we can see that the proposed algorithm could effectively detect voluntary following relationships. It therefore helps us better understand this newly arisen spamming activity. For example, we can investigate a basic research question with the detection results: how many percentage of volowers' following relationships are related with spamming activities? We get all the users' $P_c(u)$ scores from an experiment when $m = 300$. We regard the user u as a customer if its $P_c(u)$ is higher than the threshold θ_c. Figure 4 shows the proportions of detected customers (not manually labeled) in labeled volowers' and legitimate users' followees with different thresholds.

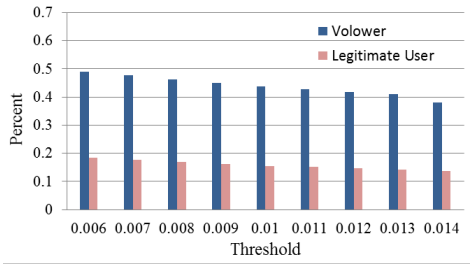


Figure 4 Percentages of customers in different user set.

From the experimental results in Figure 4, we can see that there are much more customers in the followees of vowlers than those of legitimate users, which is reasonable because volowers are more willing to follow customers than legitimate users to make profit. Another finding is that with a reasonable threshold $\theta_c = 0.009$ (which is corresponding to the best performance in customer detection), the percentage of customers in volowers' followee set is about 44%. It validates our proposed definition of volowers that they follow a majority of ordinary users (for fun/interest) and also some customers (for profit) at the same time.

7 Conclusions

Detection of voluntary following activities will help microblogs to automatically detect malicious accounts that make profit in the follower markets and potentially harm the normal operation of microblogs. In this work, we investigated the new problem of voluntary following activity detection with an effective algorithm *DetectVC* that incorporates prior knowledge and graph structure. This method does not involve any static or dynamic property features and relieves researchers and practitioners from the exhausting fight against continuously emerging spam activities. Our proposed algorithm references the HITS algorithm [Kleinberg. 1999] which is a widely used authority source distilling algorithm based on link analysis. The spam probabilities of P_v and P_v calculated from our algorithm are similar to the basic definitions of hub and authority. In the future, we would like to extend this work to other social network platforms (such as CQA) and further improve the detection performance of crowdturfing activities.

Acknowledgments

This work was supported by Tsinghua University Initiative Scientific Research Program (2014Z21032), National Key Basic Research Program (2015CB358700), Natural Science Foundation (61532011, 61472206) of China and Tsinghua-Samsung Joint Laboratory for Intelligent Media Computing.

References

[Almaatouq et al., 2014] Abdullah Almaatouq, Ahmad Alabdulkareem, Mariam Nouh, Erez Shmueli, Mansour Alsaleh, Vivek K. Singh. Twitter: who gets caught? observed trends in social micro-blogging spam. In *Proceedings of the 2014 ACM conference on Web science*, pages 33-41, New York, NY, USA, 2014.

[Aggarwal et al., 2015] Anupama Aggarwal, and Ponnurangam Kumaraguru. What they do in shadows: Twitter underground follower market. *PST*, pages 93-100, 2015.

[Cha et al., 2010] Meeyoung Cha, Hamed Haddadi, Fabr´ıcio Benevenuto, and Krishna P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *Proceedings of International Aaai Conference on Weblogs & Social*, 10(10-17), 2010.

[Cheng et al., 2013] Cheng Binlin, Fu Jianming, and Huang Jingwei. Detecting Zombie Followers in Sina Microblog based on the Number of Common Friends. *The International Journal of Advancements in Computing Technology* 5(1): 612-620, 2013.

[Egele et al., 2013] Manuel Egele, Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. COMPA: Detecting Compromised Accounts on Social Networks. *Symposium on Network and Distributed System Security*, 2013.

[Hu et al., 2013] Xia Hu, Jiliang Tang, Yanchao Zhang, and Huan Liu. Social spammer detection in microblogging. *IJCAI*, pages 2633-2639, 2013.

[Jiang et al., 2014] Meng Jiang, Peng Cui, Alex Beutel, Christos Faloutsos, and Shiqiang Yang. CatchSync: catching synchronized behavior in large directed graphs. *SIGKDD*, pages 941-950, 2014.

[Kleinberg. 1999] Kleinberg, Jon M. Hubs, authorities, and communities. *Acm Computing Surveys*, 31(2): 685-695, 1999.

[Lee et al., 2013] Kyumin Lee, Prithivi Tamilarasan, James Caverlee. Crowdturfers, Campaigns, and Social Media: Tracking and Revealing Crowdsourced Manipulation of Social Media, *ICWSM*, 2013.

[Lee et al., 2014] Kyumin Lee, Steve Webb, and Hancheng Ge. Characterizing and automatically detecting crowd-turfing in Fiverr and Twitter. *Social Network Analysis and Mining*, 5.1:1-16, 2014.

[Li et al., 2015] Xin Li, Yiqun Liu, Min Zhang, Shaoping Ma, Xuan Zhu, Jiashen Sun. Detecting Promotion Campaigns in Community Question Answering. *AAAI*, pages 2348-2354, 2015.

[Motoyama and McCoy, 2011] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M. Voelker. Dirty jobs: The role of freelance labor in web service abuse. *Proceedings of the 20th USENIX conference on Security*, 2011.

[Motoyama et al., 2011] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M. Voelker. An analysis of underground forums. *SIGCOMM*, pages 71-80, 2011.

[Ribeiro et al., 2015] Ribeiro, Bruno, and Christos Faloutsos. Modeling WebSite Popularity Competition in the Attention-Activity Marketplace. *WSDM*, 2015.

[Singh et al., 2014] Singh Monika et al. Detecting Malicious Users in Twitter using Classifiers. In *Proceedings of the 7th International Conference on Security of Information and Networks,* pages 247-254, 2014.

[Stringhini et al., 2012] Gianluca Stringhini, Manuel Egele, Christopher Kruegel, and Giovanni Vigna. Poultry markets: on the underground economy of twitter followers. In *Proceedings of the 2012 ACM workshop on Workshop on online social networks*, pages 1-6, 2012.

[Stringhini et al., 2013] Gianluca Stringhini, Gang Wang Manuel Egele et al. Follow the Green: Growth and Dynamics in Twitter Follower Markets. In *proceedings of the 14th Internet Measurement Conference*, pages 163-16, 2013.

[Teraguchi et al., 2012] Toshio Teraguchi et al. Detection Method of Blog Spam Based on Categorization and Time Series Information. In *Proceedings of the 26th International Conference on Advanced Information Networking and Applications Workshops*, pages 801-808, 2012.

[Thomas and McCoy, 2013] Kurt Thomas, McCoy Damon, Grier Chris et al. Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse. *USENIX Security*, pages 195-210, 2013.

[Wang et al., 2010] Alex Hai Wang. Don't follow me: Spam Detection Twitter. In *proceedings of the 5th Security and Cryptography,* pages 142-151, 2010.

[Wang et al., 2012] Gang Wang, Christo Wilson, Xiaohan Zhao, Yibo Zhu, Manish Mohanlal, et al. Serf and turf: crowdturfing for fun and profit. *International Conference on World Wide Web*, pages 679-688, 2012.

[Wang and Mohanlal, 2013] Gang Wang, Manish Mohanlal, Christo Wilson, Xiao Wang. Social Turing Tests: Crowdsourcing Sybil Detection. *Eprint Arxiv*, 2013.

[Wang and Konolige, 2013] Gang Wang, Tristan Konolige et al. You are how you click: Clickstream analysis for sybil detection. In *Proceedings of the 22nd USENIX Security Symposium,* pages 241-256, 2013.

[Xie et al., 2008] Yinglian Xie, Fang Yu, Kannan Achan, Rina Panigrahy, Geoff Hulten, and Ivan Osipkov. Spamming botnet: Signatures and characteristics. *SIGCOMM*, pages 171-182, 2008.

[Yang and Liu, 2012] Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. Automatic detection of rumor on Sina Weibo. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics,* 2012.

[Yang et al., 2012] Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y. Zhao, and Yafei Dai. Uncovering Social Network Sybils in the Wild. *Journal of ACM Transactions on Knowledge Discovery from Data*, 2012.

[Zhu et al., 2012] Yin Zhu, Xiao Wang, Erheng Zhong et al. Discovering Spammers in Social Networks. *In Proceedings of 26th of Association for the Advancement of Artificial Intelligence*, 2012.