Fake Reviews Tell No Tales? Dissecting Click Farming in Content-Generated Social Networks

Neng Li[1], Suguo Du[1], Haizhong Zheng[1], Minhui Xue[2], Haojin Zhu[1]

[1] Shanghai Jiao Tong University, Shanghai 200240, China
[2] New York University Shanghai, Shanghai 200122, China

Abstract: Recently, there has been a radial shift from traditional online social networks to content-generated social networks (CGSNs). Contemporary CGSNs, such as Dianping and TripAdvisor, are often the targets of click farming in which fake reviews are posted in order to boost or diminish the ratings of listed products and services simply through clicking. Click farming often emanates from a collection of multiple fake or compromised accounts, which we call click farmers. In this paper, we conduct a three-phase methodology to detect click farming. We begin by clustering communities based on newly-defined collusion networks. We then apply the Louvain community detection method to detecting communities. We finally perform a binary classification on detected-communities. Our results of over a year-long study show that (1) the prevalence of click farming is different across CGSNs; (2) most click farmers are lowly-rated; (3) click-farming communities have relatively tight relations between users; (4) more highly-ranked stores have a greater portion of fake reviews.

Key words: Click Farming; community detection; Content-Generated Social Networks

I. INTRODUCTION

With the rapid development of social networking, there is a plethora of research [1-3] focusing on the security and privacy problems of online social networks. Social bots have been tremendously popular in online social networks over the past decade. They are automated agents that produce content and interact with humans on social media, attempting to influence the working of systems. Social bots are created for many purposes, such as email spamming, instant messaging to collaborative content rating, recommendation, political infiltration, and malicious content delivery. Online social networks have witnessed huge spurt for social bot intrusion. Twitter reported in 2014 that 5% million accounts are either fake or spam [4]. Facebook revealed that up to 83 million of its users are fake [5]. Traditional defenses against social bots rely on using properties of the social network's structure [6, 7]. However, in a social network, there exist a limited number of attack edges connecting between benign and bot users, thus, rendering strong trusts lacking in real social networks, such as RenRen [8] and Facebook [9]. In parallel, the arms race has also driven the corresponding countermeasures [10-13].

Recently, there has been a radial shift from traditional online social networks to content-generated social networks (CGSNs) in terms of the nature of content generation. User-generated content such as product reviews has become increasingly ``social,'' in the sense that consumers take suggestions not only from

the general community but also from their own online social connections. Many CGSNs, such as Dianping and Yelp, have made great effort to build connected-review communities, and some others, such as TripAdvisor, have partnered with Facebook to allow users to share reviews using Facebook accounts. As the reviewing process is taken anonymously and users have limited information on individuals who post reviews, a major concern is that the credibility of reviews can be undermined by a new type of bot attacks, which we call *click farming*. Click farming is typically launched by multiple fake or compromised accounts which are used to generate fake reviews that masquerade as testimonials from ordinary people simply through clicking. The goal of click farming is to deceive ordinary users into making decisions favorable to the products. Recent evidence suggests that many CGSNs, such as Yelp and TripAdvisor, are often the targets of click farming. In specific, Yelp profile pages featured ``consumer alerts'' on several sneaky businesses which got caught red-handed trying to buy reviews, crafted by Yelp ``elite'' users, for these businesses. TripAdvisor has also put up similar warning notices.

In this paper, we are particularly interested in exploring the role of click farming in CGSNs, and how behavioral characteristics of click farmers differ from real users. To achieve this, we conduct a three-phase methodology to detect click farming. We cluster communities based on newly-defined collusion networks. We then apply the Louvain community detection method to detecting communities. We finally perform a binary classification on detected-communities, echoing that a large number of fake reviews are usually posted by the malicious community in which all click farmers reside. This paper presents the results of over a year-long study of click farming in two CGSNs -- TripAdvisor and Dianping. Dianping is by far the most popular CGSN in China. By analyzing 10,541,931 reviews, 32,940 stores, and 3,555,154 users from Dianping and 363,196 reviews, and 3,845 stores, and 67,172 users from TripAdvisor, respectively, our re-

search shows that the methodology developed in this paper is effective in detecting click farming and can be well-generalized across CGSNs.

The main contributions of this paper are as follows:

1) We design a novel methodology to detect click-farming communities by building new social colluding relations between users.

2) We evaluate our detection system on two CGSN datasets in the wild, Dianping and TripAdvisor. For Dianping dataset, our detection system achieves a precision of 96.74%. For TripAdvisor dataset, our detection system achieves a precision of 94.74%. Furthermore, our detection system detects in total 566 click-farming communities on Dianping and 92 click-farming communities on TripAdvisor.

3) We analyze the characteristics with respect to click-farming communities across two CGSNs. We show that most click farmers are lowly-rated and click-farming communities have relatively tight relations between users.

4) We show the relations between the portion of fake reviews and store ranks on TripAdvisor and find that more highly-ranked stores have a greater portion of fake reviews.

The rest of paper is organized as follows. To begin with, we briefly introduce two content-generated social networks, Dianping and TripAdvisor, which this paper focus on in Section II.

Then, Section III describes the details of our proposed detection system.

Next, in Section IV, we evaluate our detection system on two real world CGSN datasets. Besides evaluation, we analyze some characteristics of click farming communities including distribution of user levels, behavior patterns of click farming communities and structure of click farming communities. Furthermore, we also analyze the relations between the portion of fake reviews and store ranks.

Finally, we draw our conclusion in Section

In this paper, we conduct a three-phase methodology to detect click farming.

V.

II. Background

In this section, we take a look at the different ways that content-generated social networks (CGSNs) manage the network of trust and the process of submitting a review. This is important because the nature of the network of trust provides click farmers with different options for promoting malicious advertisements or messages.

2.1 The dianping content-generated social network

Dianping is by far the most popular CGSN in China, where users can review local businesses such as restaurants, hotels, and stores. The revenue of Dianping comes from three sources: (1) selling display and keyword search advertising; (2) offering online coupons in return for an advertising fee; and (3) offering discount card and group-buying to members and getting a share from participating restaurants. These promotional activities will spur the willingness-to-pay of restaurants and somehow breed the click farming in parallel.

When a user uses Dianping, she or he can search for a restaurant based on geo-location information, the pricing, the cuisine-type, the quality-type, etc. Dianping will return to the user with a list of restaurant choices in order of overall quality-rating. The quality-rating of a restaurant review is typically scaled from 1 (worst) to 5 (best), mainly depending on the restaurant service. Users can also submit pictures of restaurants and dishes. Users can vote ''helpful'' if the review is informative and useful. Users are also assigned star-ratings. These star-ratings vary from 0 stars (rookie) to 6 stars (expert), depending on the longevity of the user account, the number of reviews posted, and the number of ''helpful votes'' received. A higher star-rating indicates that the user is more experienced and more likely to be perceived as an expert reviewer. Similar to ''Elite User'' on Yelp, a senior level user (*e.g.*, 4-star, 5-star, or 6-star user) is supposed to be a small group of in-the-know users who have a large impact on their local community. Dianping has established its user reputation system that classifies user reviews into ''normal reviews'' and ''filtered reviews,'' but the details of the algorithm remain unknown to the public.

Data Collection. We develop a web crawler to analyze HTML structure of store pages and user pages on Dianping. All reviews are crawled by web crawler from January 1, 2014 to June 15, 2015. Starting with a seed store list with 4 stores, we crawl all reviews belonging to those stores on the store list. Next, we use users who write these reviews to extend the user list and crawl all reviews from the page of these users. The web crawler repeats these two steps until reaching 32,940 stores on the store list. At last, the Dianping dataset has in total 10,541,931 reviews, 32,940 stores, and 3,555,154 users. We rely on manually labeled data for detecting click-farming communities. We browse the homepages of users in each randomly selected community and judge whether the community is a click-farming community based on users' behaviors. We randomly pick up 170 communities and further label them into 117 click-farming communities and 53 benign communities.

2.2 The tripadvisor content-generated social network

TripAdvisor is a content-generated social network which enables travelers to plan and book their trip based on other traveler's reviews. TripAdvisor is one of the largest travel communities, operating in 45 countries worldwide, and it currently reaches more than 100 million travel reviews on accommodations, restaurants, and attractions.

TripAdvisor's primary function is to disseminate user-generated content, such as reviews, ratings, photos, and videos on a specific domain. Users can consult quantitative and qualitative comments on any accommodation, restaurant, and attraction, all posted by other travelers. When submitting a review, users are required to rate each experience on a five-

star scale from 1 (worst) to 5 (best) in terms of check-in quality or comfort of the room. Users also have the opportunity to upload photos and videos to support their reviews. The quantitative rating provided by users is considered to generate a summary score and rank the properties within a destination in terms of overall popularity. Details of the algorithm used by TripAdvisor to calculate this ranking are not public knowledge, but definitely take into consideration the quantity, quality, and age of the reviews submitted.

Data Collection. We use a Python-based crawler to crawl the data from store pages on TripAdvisor from April 1, 2014 to March 31, 2017. First, we get the stores' URLs directly from the store list on TripAdvisor and limit the scope to New York City. Second, we visit stores' web pages and analyze the structure of HTML file. Then, we collect all users' data in stores' pages including users' basic information (user IDs and usernames), user levels, posted reviews, and helpful votes. We have crawled in total 700,922 reviews, 3,845 stores, and 304,546 users. Since inactive accounts do not have enough data for click farming detection, we only focus on those users who publish at least three reviews. Finally, we obtain a dataset of totally 363,196 reviews, 3,845 stores, and 67,172 users. Similar to Dianping, we rely on manually labeled data for detecting click-farming communities. We browse the homepages of users in each randomly selected community and judge whether the community is a click-farming community based on users' behaviors. We randomly pick up 103 communities and further label them into 19 click-farming communities and 84 benign communities.

III. OUR DETECTION METHODOLOGY

In this section, we will describe the methodology that we build to detect click farming. The main insight of our methodology is based on the fact that click farmers belong to the same community tend to post reviews in similar stores. The methodology mainly takes three steps. First, we set up social relations between users, since, unlike users in general online social networks, users in content-generated social networks (CGSNs) tend to have a much sparser relation and click farmers who reside in the same community are less likely to follow each other. Instead, we try to define a novel relation based on the similarity between pairs of reviews posted by different users, yielding users to be better characterized in CGSNs. We derive a social graph by using the defined similarity metric. We then apply the Louvain community detection method [22] to the derived social graph. The Louvain method can detect out the communities of which the users tend to post reviews in similar stores. Finally, because not all communities are click-farming communities, we apply supervised machine learning techniques to distinguishing click-farming communities from communities composed of real users (real-user communities).

3.1 Building social relations between users

In order to cluster users in communities, the first step is to build social relations between users. Reviews posted, by different users, in the same store for the same purpose (boosting or depreciating the store) within the same time period will be considered as colluding reviews. The more colluding reviews two arbitrarily users share, the more similar two users are. Previous work [11, 12] generally adopted Jaccard similarity metric, which is extensively used to measure similarity between sets. However, we emphasize that those review sets by simply defining Jaccard similarity do not satisfy mathematical equivalence, which means that simply applying Jaccard similarity cannot even work in our problem. We show the newly-built similarity metric between users in Algorithm 1.

Algorithm 1 takes as input two review sets from different users. First, Algorithm 1 sets all reviews to be unflagged. Next, Algorithm 1 compares each review between two review sets. If two reviews, both 1-star of 5-stars, are

posted in the same store within the same time period ΔT, Algorithm 1 will flag these two reviews. Finally, Algorithm 1 takes all flagged reviews as the intersection of two review sets, and calculates the similarity in the way that is different from Jaccard similarity in principle. After setting up the similarity between each user, we construct a social link between users whose similarity outnumbers a certain threshold.

3.2 Detecting click-farming communities

With a derived social graph in hand, we apply the Louvain community detection method to detecting communities.

The Louvain community detection method is a greedy optimization method that tries to optimize the modularity of a partition of the network and is composed of two steps. At the first step, the Louvain method optimizes modularity locally to look for small local communities. The second step aggregates nodes in the same community to a singular node to build a new network. Finally, the Louvain method repeats these two steps until the network attains a maximum modularity.

By using the Louvain method, we successfully obtain communities of which the

users present a strong colluding relation between each other. However, we find that simply applying the Louvain method is not adequate to separate out click-farming communities, because users in these communities probably reside in the proximity, and their reviews are mistakenly considered in similar stores. To distinguish these communities from click-farming communities, we use supervised machine learning techniques to classify communities into click-farming communities and real-user communities.

3.3 Classifying detected communities

At the final step of our methodology, we apply machine learning classifiers to discriminating click-farming communities from real-user communities. To make the classification more effective, we choose two types of features containing totally 8 features which are tabulated in table 1. In order to provide a comprehensive portrait of data, we use both community-based features and user-based features for the classifiers.

1) Community-based features provide statistics of network topology of the dataset. Score deviation and average number of reviews are two basic features of communities. Entropy of the number of reviews in each stores can be used to distinguish click-farming communities that only post reviews in a few stores. We also use entropy of districts of stores, a location-based feature which is widely used in prior research [23-25], because the mobility pattern of real users are different from that of click farmers. Average similarity shows the similarity between users in the same community. Click-farming communities tend to have a higher average similarity. Global clustering coefficient characterizes the degree that nodes are to be clustered together. Because click farmers tend to work collaboratively, it is more likely for click-farming communities to have a higher global clustering coefficient.

2) User-based features provide more detailed

Algorithm 1. Calculate similarity between users

Input: S_U, S_V:Two review sets derived from two different users U and V
Output: *Similarity*: A metric describes the similarity of reviews posted by U and
        V.

1: Set all reviews in S_U, S_V to the unflagged.
2: **for** $R_U \in S_U$ **do**
3:     **for** $R_V \in S_V$ **do**
4:        **i f** $R_V^{store} = R_U^{store}$ a n d $\left| R_V^{time} - R_U^{time} \right| < \Delta T$ a n d
       $(R_V^{rating} = R_U^{rating} = 1 star \, or \, R_V^{rating} = R_U^{rating} = 5 star)$ **then**
5:         Set S_U, S_V to be flagged
6:     **end if**
7:     **end for**
8: **end for**
9: L: The number of flagged reviews in S_U, S_V
10:*Similarity=* $\dfrac{L}{\left| S_U \right| + \left| S_V \right| - L}$
11:**return** *Similarity*

---

behavioral characteristics of users. Click farmers will frequent some stores and repeatedly post reviews in those stores. Unique review ratio and maximum number of duplication are two features that reflect the user-level behaviors.

IV. EVALUATION AND MEASUREMENT

In this section, we evaluate the performance of our methodology on two CGSNs and then dissect several characteristics of click farmers and real users. We begin to evaluate results by precision, recall, F1 score, and AUC. Next, we compare the distribution of user levels between click farmers and real users on Dianping and TripAdvisor. Then, we compare the entropy of the number of reviews appearing in different stores across two CGSNs. In addition, we analyze the structure of click-farming communities and find that users in click-farming communities generally have relatively tight relations. Finally, we analyze relations between the portion of fake reviews and store ranks on TripAdvisor.

4.1 Performance of classification

To evaluate the performance of our methodology, we apply the methodology to two large-scale CGSNs in the wild, which are Dianping and TripAdvisor. We evaluate the performance of classification by using standard metrics, such as accuracy, precision, and recall for each dataset. For Dianping dataset, our methodology detects out in total 710 communities. To apply supervised machine learning, we randomly sampled 170 communities and manually labeled these communities into 117 click-farming communities and 53 real-user communities, as our training set. For TripAdvisor dataset, our methodology detects out 495 communities. To apply supervised machine learning, we randomly sampled 103 communities and manually labeled these communities into 19 click-farming communities and 84 real-user communities, as our training set.

With the 8 features proposed in Subsection III-3.3, we compare with several standard machine learning classifiers implemented by *scikit-learn* library [26]. We evaluate each classifier by weighted precision, weighted recall, weighted F1 score, using 5-fold cross-validation. For Dianping dataset, table 2 shows that all classifiers have an excellent performance in classification. Particularly, SVM (support vector machine) performs best overall with 96.75% precision, 96.47% recall, 96.50% F1 score, and 99.42% AUC. For TripAdvisor dataset, table 3 shows that SVM (support vector machine) also achieves the best overall performance with 94.74% precision, 90.00% recall, 92.31% F1 score, and 92.73% AUC. We see that our methodology largely performs well on two datasets, which indicates it can be well generalized across CGSNs. The classification better performs on Dianping than on TripAdvisor is perhaps because TripAdvisor dataset has less users and cannot contain all click-farming communities in nature. Due to different social network topologies, we acknowledge that the TripAdvisor dataset may be sensitive to features used for click-farming communities.

4.2 Distribution of user levels

For prediction, we identify 566 click-farm-

Table I. *Types of features.*

Types of Features	Features
Community-based Features	Score deviation, Average number of reviews, Entropy of the number of reviews in each store, Entropy of districts of stores, Average similarity, Global clustering coefficient
User-based Features	Unique review ratio, Maximum number of duplication

Table II. *Classification performance for Dianping dataset*

Classifier	Precision	Recall	F1	AUC
SVM	96.74%	96.47%	96.45%	99.42%
KNN	96.75%	96.47%	96.50%	97.45%
Random forest	93.16%	94.01%	92.99%	97.42%

Table III. *Classification performance for TripAdvisor dataset*

Classifier	Precision	Recall	F1	AUC
SVM	94.74%	90.00%	92.31%	92.73%
KNN	89.47%	85.00%	87.18%	86.85%
Random forest	89.47%	94.44%	91.89%	91.32%

ing communities with 22,324 users, and 144 real-user communities with 5,222 users for Dianping dataset. For TripAdvisor dataset, we identify 92 click-farming communities with 524 users, and 403 real-user communities with 7,345 users. Surprisingly, we find that the portion of click-farming communities are somehow contrary on two datasets. We reason

this observation as follows: (1) The monetary reward per click in China is relatively lower than that in US, which directly entices more stores to mount click farming on Dianping. (2) The size of TripAdvisor dataset may naturally omit click-farming communities and inherit more moderate-sized real-user communities.

To boost the reputation of a store, a click-farming community requires a great number of click farmers to post fake reviews. Registering new accounts is a major approach to gaining accounts at a low cost, but new accounts generally have lower levels, reducing user levels of the most clicker farmers. In this subsection, we compare the user levels between real users and click farmers on Dianping and TripAdvisor, respectively. As shown in figure 1, for Dianping dataset, most click farmers have user levels below 1-stars. Meanwhile, the distribution of the user levels of real users almost complies with normal distribution, centered between 3- and 4-stars. TripAdvisor has an analogous interpretation: *The user levels of click farmers are lower than those of real users*.

Through this comparison, we can find that click-farming communities prefer using accounts with lower levels, since these accounts are much easier to obtain. From figure 1, it is obvious that user levels of users on TripAdvisor are higher than those on Dianping. We think this is mainly due to the disparate standards taken by two CGSNs, as it is much easier for TripAdvisor users to level up.

Fig. 1 *Comparison of user levels between different groups.*

4.3 Behavioral patterns of click-farming communities

Although click-farming communities on both Dianping and TripAdvisor have the same goal of boosting reputation of stores, click-farming communities on Dianping and TripAdvisor are characterized by different behavioral patterns due to the different topology of two CGSNs. In this subsection, we try to mine out the different behavioral patterns of communities from Dianping and TripAdvisor. figure 2 shows the CDF of the entropy of the number of reviews in each store boosted in a click-farming com-
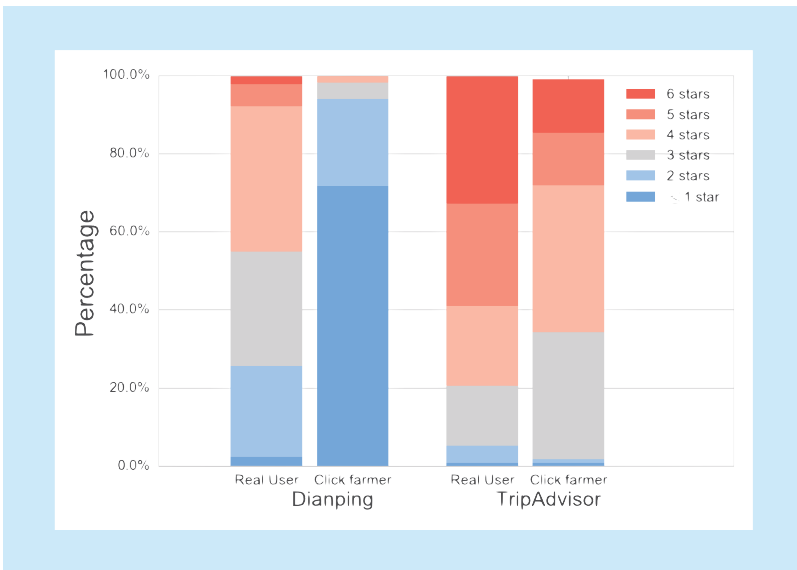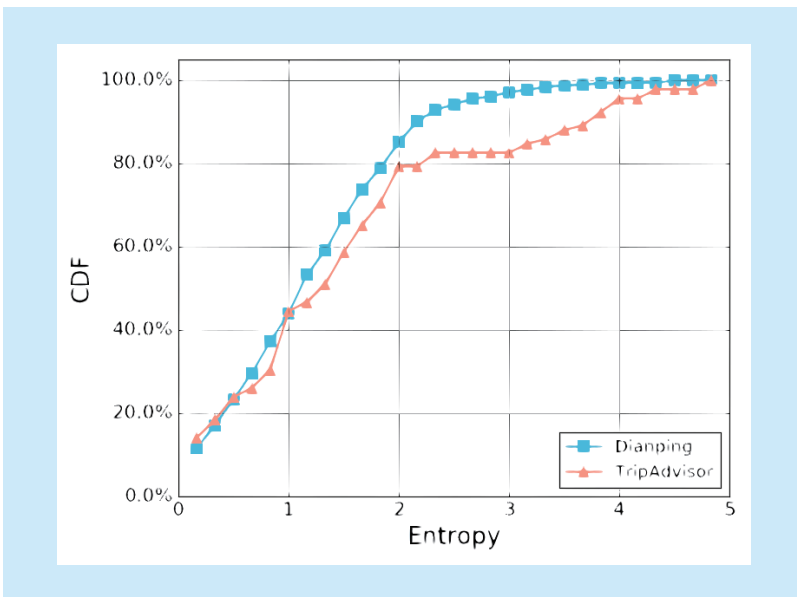

Fig. 2. *The CDF of entropy of the number of reviews in each store boosted in a click-farming community. Click-farming communities may boost reputation for several stores simultaneously. The entropy of the number of reviews in each store boosted in a click-farming community can present the concentration degree of reviews posted by a click-farming community.*

munity. From figure 2, we can find that, for both datasets, approximately 80% communities have an entropy less than 2, which means that most click-farming communities only post fake reviews in limited stores. It is interesting that there are very few click-farming communities of which the entropy is between 2 and 3 for TripAdvisor dataset. Compared with the Dianping dataset where click-farming communities largely have an entropy less than 3, the TripAdvisor dataset has approximately 20% click-farming communities of which the entropy is larger than 3. This suggests that a small number of click-farming communities are mounted by a large number of stores.

4. 4 Structure of Communities

In this subsection, we analyze the structure of click-farming communities on Dianping and TripAdvisor. figure 3 shows that the CDF of global clustering coefficient of click-farming communities in two CGSNs largely follows the same curve, with few communities having a global clustering coefficient less than 0.4. This indicates that almost all communities have relatively tight relations between users. We can also find that approximately 20% communities on Dianping have a global clustering coefficient close to 1, which indicates these communities form complete graphs. Click farmers in these communities generally post reviews in the same stores. We speculate that these click farmers are probably manipulated by a single person or organization.

4.5 Comparison of users' activities range

Our click-farming detection works well on two CSGNs, but there still exist differences between these two CSGNs, especially for users' range of activities. We compare thedistributions of users' shopping positions in these two CGSNs. Users on TripAdvisor have a larger range of activities compared with users on Dianping. TripAdvisor users are usually located in several cities even in several countries while the main activities positions of Dianping are distributed in a small area such as a city or a province.
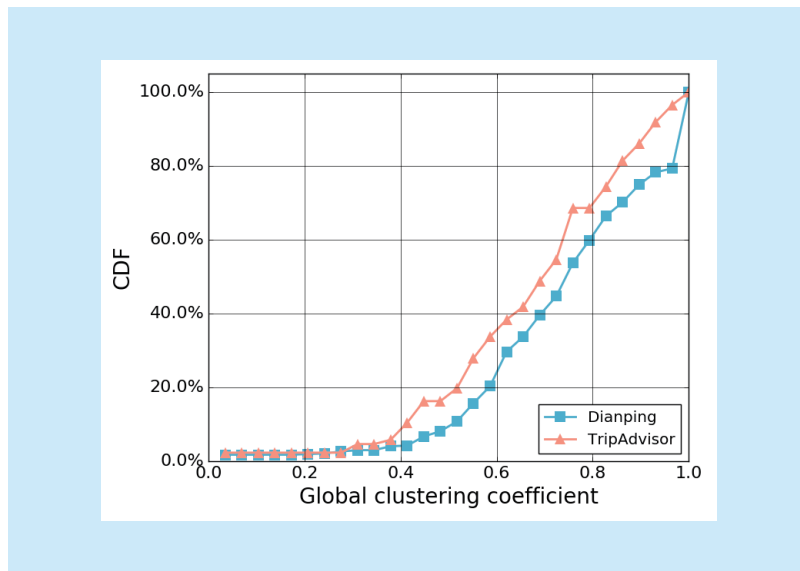

Fig. 3. *The CDF of global clustering coefficient. Global cluster coefficient is a metric to measure the degree of a cluster. The higher the global clustering coefficient a community has, the tighter the relation between users in a community is.*

The reason of above phenomenon is perhaps due to the different market positioning of these two CGSNs. TripAdvisor is a CGSN focusing on travel services, which leads to the fact that users of TripAdvisor are mainly tourists. Dianping is quite different with respect to their target users, which assists business to attract users through discounts. Tourists prefer using TripAdvisor to book hotels or find local restaurants when they are traveling and might leave their reviews in many cities or states. However, they rarely use TripAdvisor if there is no need for a trip. On the other hand, users of Dianping usually use it in their daily life for the coupon or discount which causes that the most user reviews on Dianping concentrate on their living cities. Such different marketing strategies lead to the difference of two CGSNs.

4.6 Relations between the portion of fake reviews and store ranks

The goal of click farming is to boost the reputation of stores in CGSNs. The owners of stores who mount click-farming generally wish click farming could gain huge influence on their stores. In this subsection, we analyze relations between the portion of fake reviews

and store ranks that indicate the influence of stores. figure 4 shows the trend of the portion of fake reviews with the increasing ranks of top 1,500 stores. Each red dot encodes a store and the blue curve is applied to optimally fit to all red dots. From figure 4, we find that more highly-ranked stores have a greater portion of fake reviews, which indicates that fake reviews do facilitate the ranking of stores. However, the portion of fake reviews of top 200 stores are lower than that of stores of which the ranks are between 200 and 400. We reason this observation by proposing two possible insights: (1) A significantly great number of reviews appearing in top stores generally dilute the portion of fake reviews in these stores. (2) The reputation of top stores naturally inherits a great number of highly-rated reviews, reducing demand for click farming.

V. Related Work

Over the past few years, the success of CGSNs has attracted the attention of security researchers. Review-spam detection can be considered as a binary classification or ranking problem. Previous research provides several approaches of detection.
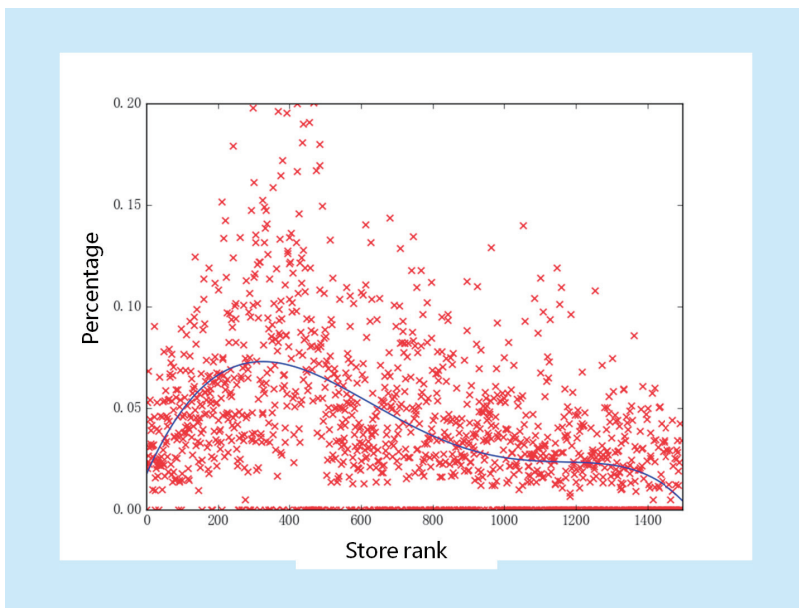

Fig. 4. *Relations between the portion of fake reviews and store ranks. X-axis represents store ranks and Y-axis represents the portion of fake reviews. Each red dot encodes a store.*

The first type of review-spam detection is based on the feature of users. Ott *et al.* [14] used unigrams and bigrams while Mukherjee *et al.* [15] incorporated many behavioral features into detection. Compared with other approaches, users' behavior features are easier to get and encode. Usually, there are clear differences between behaviors of real users and those of click farmers. Therefore, such detection often considers the information of user behavior such as profiles, activities, etc. Then they transfer these features into feature matrices and classify the users based on feature matrices. However, review-spam detection based on user features ignores the relationship between users. Only considering the features of single users will lose the information of the cooperators of click farmers. Meanwhile, click farmers can imitate the behavior of real users to avoid the detection. Some click farmers will post reviews which are similar to real users' reviews recently. It increases the difficulty of detection if we only concern the features of user. Therefore, review-spam detection based on features works well on adversarial attacks, but it needs further improvement to detect the sophisticated situation.

In recent years, researchers leveraged network relations into opinion-spam detection. Different from the detection based on users' features, most of them constructed a heterogeneous network of reviewers/reviews and products, such as using HITS-like ranking algorithms [16] and Loopy Belief Propagation [17, 18]. The idea of these detection is to construct a graph of network and detect the click farmers based on relations in network. The approach based on network focus on the relationship between reviews and reviewers. It exploits the network effect among reviewers and products to detect the click farmers. Another advantage of this approaches is that the run time grows linearly with network size which is extremely important in a large-scale CSGN.

Other work [11, 12, 19] focused on detecting clusters of users. Specifically, CopyCatch [19] and SynchroTrap [11], implementing mixed approaches, scored comparatively low

false positive rates with respect to single feature-based approaches.

Most recently, we observe that several studies discuss how to identify the malicious users in CGSNs by exploiting crowdsourcing-based approaches [5, 20], or model-based detection [21] that limits their broad applicability. These approaches can achieve a high accuracy of click farmers detection. However, the cost of such detection is extremely high especially in a large-scale network. Detection based on manual check will also lead to the privacy problem, since workers can access all the profiles of users.

Therefore, in this work, we try to build a social network using collusion relations and further incorporate community detection and supervised machine learning to our detection methodology, which is shown to be more effective in capturing click farming phenomenon.

VI. DISCUSSION AND LIMITATION

6.1 Comparison with prior click-farming detection

Comparing with the previous research, the most important advantage of our click-farming detection is that we consider the internal working mechanism of the click-farming activities and try to understand the way they work.

Instead of using personal features of users for detection, we build the user communities based on their reviews to group the users who often post reviews in the same stores. Because click-farming communities will post reviews in the same stores within the same time period, building user communities is beneficial to click farmers detection. Simultaneously, we can avoid the loss of information in feature-based detection.

Different from the detection based on relationship in network, we do not only use the features of communities. We select the community-based features to classify the community. Meanwhile, we also consider the user-based features which provide more detailed behavioral characteristics of users. Therefore, we can achieve a better performance of click-farming detection. The result shows that our click-farming detection is effective in detect click farmers and can achieve a high accuracy in both two large-scale CSGNs.

6.2 Limitations

Similar to other real-world systems, our work also has its limitations. Although our click-farming detection achieves a high accuracy and a very low false negative rate in both CGSNs, it can still cause large quantities false negative cases, especially when the CGSN has millions of users. These cases should be manual checked, although it may cause lots of labor cost and bad user experience. This limitation can be relieved by creating buffer for uncertain click farmer users instead of a binary classification.

Meanwhile, the click farmers can post numbers of unrelated reviews to evade the detection. Such behavior can reduce the similarity with click-farming users.

Finally, if click farmers hire real user to post the fake reviews, it is hard to detect since such users have their own communities and quite different from click-farming communities.

VII. CONCLUSION

In this paper, we have taken the first steps toward developing a deeper understanding of how click farming works on two popular content-generated social networks (CGSNs) domestically and internationally. We found that, despite their considerable differences across CGSNs, click-farming communities form relatively tight relations between users. We also took a large-scale measurement analysis of detected click-farming communities. Evaluation on both real-world datasets showed that our proposed methodology is fundamentally reliable to stop the spread of click farming.

As digital credibility becomes more important, it is apparent that the potential for online misconduct will increase, thereby necessitating

anti-click-farming detection. Understanding the internal structure of click farming associated with alternative detection approaches will pave the way for a full-fledged deployment.

References

[1] M. Xue, C. Ballard, K. Liu, C. Nemelka, Y. Wu, K. Ross, and H. Qian, You can yak but you can't hide: Localizing anonymous social network sers," in *Proc. Proceedings of the 2016 ACM on Internet Measurement conference.* ACM, 2016, pp. 25–31.

[2] H. Li, Q. Chen, H. Zhu, and D. Ma, "Hybrid de-anonymization across real-world heterogeneous social networks," in *Proc. Proceedings of the ACM Turing 50th Celebration Conference-China.* ACM, 2017, p. 33.

[3] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: Attacks and defense," *IEEE Transactions on Dependable and Secure Computing*, 2016.

[4] Z. Gilani, R. Farahbakhsh, G. Tyson, L. Wang, and J. Crowcroft, "An in-depth characterisation of Bots and Humans on Twitter," *arXiv preprint* arXiv:1704.01508, 2017.

[5] K. Lee, P. Tamilarasan, and J. Caverlee, "Crowdturfers, campaigns, and social media: Tracking and revealing crowdsourced manipulation of social media," in *ICWSM*, 2013.

[6] Y. Boshmaf, K. Beznosov, and M. Ripeanu, "Graph-based Sybil detec-tion in social and information systems," in *Proc. Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference* on. IEEE, 2013, pp. 466–473.

[7] N. Z. Gong, M. Frank, and P. Mittal, "Sybilbelief: A semi-supervised learning approach for structure-based Sybil detection," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 6, pp. 976–987, 2014.

[8] Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai, "Uncovering social network Sybils in the wild," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 8, no. 1, p. 2, 2014.

[9] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "The socialbot network: when bots socialize for fame and money," in *Proc. Proceedings of the 27th Annual Computer Security Applications Conference.* ACM, 2011, pp. 93–102.

[10] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "COMPA: Detecting compromised accounts on social networks," in *Proc. NDSS*, 2013.

[11] Q. Cao, X. Yang, J. Yu, and C. Palow, "Uncovering large groups of active malicious accounts in online social networks," in *Proc. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security.* ACM, 2014, pp. 477–488.

[12] K. Thomas, F. Li, C. Grier, and V. Paxson, "Consequences of connectivity: Characterizing account hijacking on Twitter," in *Proc. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security.* ACM, 2014, pp. 489–500.

[13] Y. Boshmaf, D. Logothetis, G. Siganos, J. Lerí a, J. Lorenzo, M. Ripeanu, and K. Beznosov, "Integro: Leveraging victim prediction for robust fake account detection in OSNs," in *Proc. NDSS*, vol. 15, 2015, pp. 8–11.

[14] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in *Proc. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*: Human Language Technologies, vol. 1. Association for Computational Linguistics, 2011, pp. 309–319.

[15] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What Yelp fake review filter might be doing?" in *Proc. Seventh International AAAI Conference on Weblogs and Social Media*, 2013.

[16] G. Wang, S. Xie, B. Liu, and S. Y. Philip, "Review graph based online store review spammer detection," in *Proc. Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE, 2011, pp. 1242–1247.

[17] L. Akoglu, R. Chandy, and C. Faloutsos, "Opinion fraud detection in online reviews by network effects," *ICWSM*, vol. 13, pp. 2–11, 2013.

[18] S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," in *Proc. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 2015, pp. 985–994.

[19] A. Beutel, W. Xu, V. Guruswami, C. Palow, and C. Faloutsos, "Copy-catch: stopping group attacks by spotting lockstep behavior in social networks," in *Proc. Proceedings of the 22nd International Conference on World Wide Web.* ACM, 2013, pp. 119–130.

[20] J. Song, S. Lee, and J. Kim, "Crowdtarget: Target-based detection of crowdturfing in online social networks," in *Proc. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security.* ACM, 2015, pp. 793–804.

[21] H. Li, G. Fei, S. Wang, B. Liu, W. Shao, A. Mukherjee, and J. Shao, "Bimodal distribution and co-bursting in review spam detection," in

Proc. Proceedings of the 26th International Conference on World Wide Web. ACM, 2017.

[22] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

[23] M. Xue, L. Yang, K. W. Ross, and H. Qian, "Characterizing user behaviors in location-based find-and-flirt services: Anonymity and demographics," *Peer-to-Peer Networking and Applications*, vol. 10, no. 2, pp. 357–367, 2017.

[24] M. Xue, Y. Liu, K. W. Ross, and H. Qian, "Thwarting location privacy protection in location-based social discovery services," *Security and Communication Networks*, vol. 9, no. 11, pp. 1496–1508, 2016.

[25] X. Zhang, H. Zheng, X. Li, S. Du, and H. Zhu, "You are where you have been: Sybil detection via geo-location analysis in osns," in *Proc. Global Communications Conference (GLOBECOM)*, 2014 IEEE. IEEE, 2014, pp. 698–703.
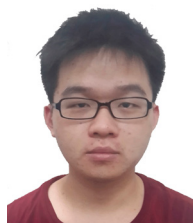
[26] Scikit-learn. [Online]. Available: http://scikit-learn.org/

Biographies

Neng Li, is currently a master candidate in computer science department, Shanghai Jiao Tong University and he completed his bachelor's degree in computer science from SJTU in 2017. He is now interested in security of advertisement. Email: lnfjpt@gmail.com
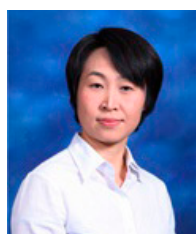
Suguo Du, is currently an Associate Professor of Management Science Department in Antai College of Economics & Management, Shanghai Jiao Tong University, China. She received her PhD in Mathematical and Information Sciences from Coventry University, U.K., in 2002, the MSc degree in Mathematics from Nanyang Technological University, Singapore, in 1998, and the BSc degree in Applied Mathematics from Ocean University of Qingdao, China, in 1993. Her current research interests include Risk and Reliability Assessment, Vehicular Networks Security and Privacy Protection and Social Networks Security Management.

Haizhong Zheng, is currently a master candidate in computer science department, Shanghai Jiao Tong University and he completed his bachelor's degree in computer science from SJTU in 2015. He is interested in online social network, adversarial machine learning and privacy of deep learning.
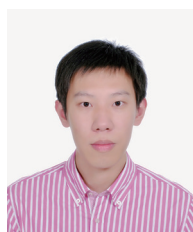
Minhui Xue, is pursuing his Ph.D. degree at the School of Computer Science and Software Engineering of East China Normal University, recipient of the Outstanding Graduate Scholarship in 2016. He is now a researcher at New York University Shanghai and a visiting Ph.D. student at Shanghai Jiao Tong University, China. Previously, he was a visiting Ph.D. student at the Courant Institute of Mathematical Sciences atNewYorkUniversitythroughout2016.Hiscurrent research interests are in data-driven analysis of online social networks. He is the recipient of the IEEE best paper award and his work has been featured in the mainstream press, including The New York Times, Science Daily, PR Newswire, and Yahoo

Haojin Zhu, Is currently a Full Professor with Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. He received his B.Sc. degree (2002) from Wuhan University (China), his M.Sc. (2005) degree from Shanghai Jiao Tong University (China), both in computer science and the Ph.D. in Electrical and Computer Engineering from the University of Waterloo (Canada), in 2009. His current research interests include network security and privacy protection. He received the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award (2014), Top 100 Most Cited Chinese Papers Published in International Journals of 2014, Supervisor of Shanghai Excellent Master Thesis, and the best paper award for IEEE ICC 2007, Chinacom 2008 and best paper award runner up for Globecom 2014 and WASA 2017. He is a senior member of IEEE and member of ACM. He serves as the Associate/ Guest Editor of IEEE Internet of Things Journal, IEEE Wireless Communications, IEEE Network, and Peer-to-Peer Networking and Applications.