You are Where You Have Been: Sybil Detection via Geo-location Analysis in OSNs

Xiaokuan Zhang, Haizhong Zheng, Xiaolong Li, Suguo Du, Haojin Zhu
Shanghai Jiao Tong University, Shanghai, China
{number_x, bingo_ms, ymingchen_123, sgdu, zhu-hj}@sjtu.edu.cn

***Abstract*—Online Social Networks (OSNs) are facing an increasing threat of sybil attacks. Sybil detection is regarded as one of major challenges for OSN security. The existing sybil detection proposals that leverage graph theory or exploit the unique clickstream patterns are either based on unrealistic assumptions or limited to the service providers. In this study, we introduce a novel sybil detection approach by exploiting the fundamental mobility patterns that separate real users from sybil ones. The proposed approach is motivated as follows. On the one hand, OSNs including Yelp and Dianping allow us to obtain the users' mobility trajectories based on their online reviews and the locations of their visited shops/restaurants. On the other side, a real user's mobility is generally predictable and confined to a limited neighborhood while the sybils' mobility is forged based on the paid review missions. To exploit the mobility differences between the real and sybil users, we introduce an entropy based definition to capture users' mobility patterns. Then we design a new sybil detection model by incorporating the newly defined location entropy based metrics into other traditional feature sets. The proposed sybil detection model can significantly improve the performance of sybil detections, which is well demonstrated by extensive evaluations based on the data set from Dianping.**

Keywords – **Sybil Detection, Location-Based Feature, Minimum Covering Circle, Entropy**

I. Introduction

Online Social Networks (OSNs) have become some of the fastest growing Web services with a massive user base, and are attracting millions of Internet users. For example, Facebook alone boasts over 500 million users, and has recently surpassed Google as the most visited site on the Internet. Sina Weibo has 503 million registered users, which posted about 100 million messages per day.

Despite their super popularity, OSNs are facing the increasing threat of sybil accounts. Sybil accounts represent fake identities that are often controlled by a small number of real users, and are increasingly used in coordinated campaigns to spread spams and malwares. Recently, Facebook reveals that up to 83 million of its users may be fake [1], while a recent study shows that more than 50% Sina Weibo users may be fake accounts. According to MarketWatch, it is reported that 20% of Yelp reviews are fraudulent. Posting fake reviews online is regarded as "the 21st century's version of false advertising". Recent evidence shows that popular OSNs are increasingly becoming the target of phishing attacks. By using compromised or fake accounts (or sybils), attackers can turn the trusted OSN environments against their users by masquerading spam messages as communications from friends and family members [2].

There are quite a few researches which have explored a number of potential solutions on sybil detection. Most proposals focus on detecting sybils in social networks by leveraging the assumption that sybils are more likely to connect to each other and form strongly connected subgraphs. Based on this assumption, sybils can be distinguished from the others by using graph theoretic approaches. However, this assumption has been challenged by the recent researches from Chinese Renren network that the large majority of sybils have actively and successfully integrated themselves into real user communities [3]. Other sybil detection approaches exploit the unique behavior patterns, clickstream models to detect fake identities in OSNs [4]. However, the clickstream patterns are not available to the public, which makes clickstream based sybil detection approach strictly limited to the service providers and cannot be performed by any third party without clickstream data.

In this study, we describe a novel sybil detection approach by exploiting the fundamental mobility patterns that separate real users from sybil ones. The proposed approach is based on the following observations. On the one hand, OSNs including Yelp and Dianping [5] provide a new platform for the users to share their experiences and comments on they received services (e.g., online review for the visited restaurants). Especially for Yelp and Dianping, the comments are normally made on the basis of the shops or restaurants and a real user needs to visit a shop/restaurant before he posts an online review, which allows us to obtain the users' mobility trajectories based on their online reviews and the locations of their visited shops/restaurants. On the other side, According to the existing studies, the human mobility is generally predictable and 94% indiviuals' daily activity is confined to a limited neighborhood of less than 100 km [6]. This makes the mobility pattern of a real user quite different from that of a paid poster. The mobility of the latter is driven by the paid review tasks instead of the normal human's mobility pattern. Intuitively, a real Yelp or Dianping user is more likely to choose the restaurants within his activity region (e.g., near his home or office) while the sybils will write fake reviews for the restaurants that are obviously beyond of his active region.

To exploit the mobility differences between the real and sybil users, we introduce an entropy based definition to capture the degree of predictability of the user's whereabouts. We

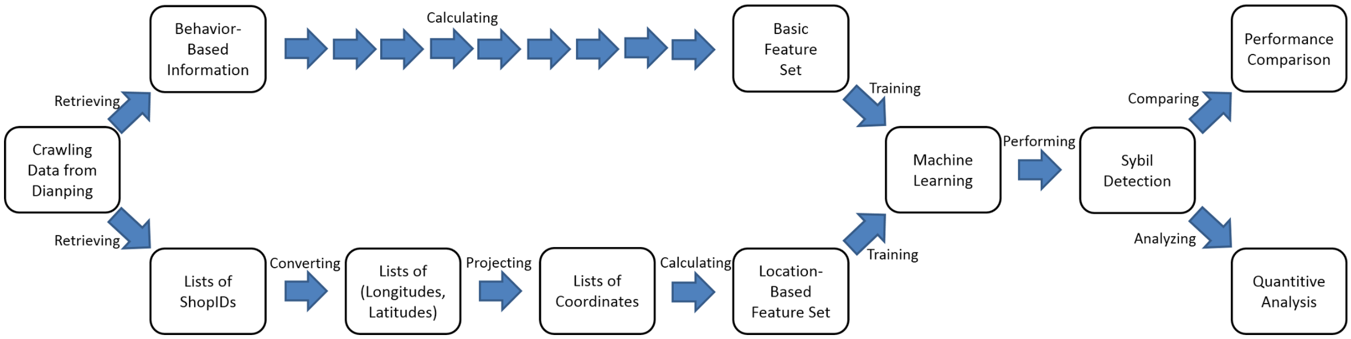

Fig. 1. Overall Procedures of Our Work

design a new sybil detection model by incorporating the newly defined location entropy based metrics into other traditional feature sets (e.g., the number of reviews per day, interval between every two consecutive comments, the average score and etc), and then perform SVM based sybil detection. We validate our models using the crawled data from Dianping, which is the largest online consumer review website in China. It is interesting to point out that, our investigation shows that the average entropy for real users' mobility pattern is about 1.68, which is in line with the previous research results that the human's mobility entropy is about 1.74 [7]. Our extensive experiment results also show that this new location information based sybil detection approach can increase the sybil detection accuracy significantly.

The rest of paper is organized as follows. Firstly, we briefly introduce the researches related to our work in Section II. Then, Section III explains the process of building our entire system. Next, some basic detection features are proposed for sybil classification in Section IV, and Section V gives another two location-based characteristics that enhance our detection. Section VI analyses the performance of our classifications and compares several features between sybil users and normal ones. Finally, we draw our conclusion in Section VII.

II. RELATED WORK

To fight against sybil attacks, there are a number of potential solutions proposed in the past several years [4] [8]. Most proposals focus on detecting sybils in social networks by leveraging the assumption that sybils will find it difficult to befriend real users. This forces sybils to connect to each other and form strongly connected subgraphs that can be detected using graph theoretic approaches [8].

The latest research shows that sybil attacks are now becoming a new kind of crowd-sourcing system, or referred as crowdturfing systems, which is dedicated to organizing workers to perform malicious tasks [9]. A crowdturfing system could be defined as systems where customers initiate "campaigns", and a significant number of users obtain financial compensation in exchange for performing simple "tasks" that go against accepted user policies. The campaigns on these crowdturfing systems are highly effective at reaching users, and their continuing growth poses a concrete threat to online communities both in the US and elsewhere, especially for those profit function OSNs such as Amazon, Ebay and Dianping. In Wang et al.'s work, it demonstrates the existence of campaigns that provide unsolicited advertisements for legitimate businesses [9]. All of these demonstrate that sybil attacks as well as crowdturfing systems pose a real threat to OSNs.

III. SYSTEM DESIGN

Established in 2003, Dianping is the first third-part review site on local consumption service in China and right now it is the biggest and the most used search portal on local and personal consumption services [5]. According to the statistics from Dianping, as of the fourth quarter of 2013, Dianping had more than 90 million monthly active users, over 30 million reviews, and more than 8 million local businesses covering approximately 2,300 cities across China [10]. The reasons for choosing Dianping are twofold. On one hand, the good reviews and high ratings on Dianping serve as good advertisements, which encourages more users having similar interests to enjoy the services provided by this restaurant/shop/hotel. On the other hand, it also attracts a lot of sybil users to launch the campaigns that provide unsolicited advertisements to attract real users to some particular restaurants.

In this section, the system that we build to do our research will be briefly described. First of all, this system will crawl raw data from Dianping, which will be stored in our database. Table I shows the four tables in the database and their contents. After that, some basic computations will be done to construct our first feature set. This set is concerned about a user's behavior, e.g., the number of reviews per day, interval between every two consecutive comments, the average score, etc. Next, location-based information is taken into consideration. Lists of shopIDs will be retrieved from the database, which will be converted into longitudes and latitudes first, then projected into Cartesian coordinate system to calculate two novel features. Then we will compare the performances of sybil detections using these different feature sets, which reveals several characteristics that can be used to distinguish sybils. Therefore, some quantitive analysis will be done to show the huge difference between sybil users and normal ones in these features. Fig. 1 shows the overall procedures of our work.

Table	Contents of the Table
shopInfo	shopID, shop's name and address, and other information about this shop, including average score, total amount of comments, coupons, etc.
shopCmt	all comments in the shop's page, each comment includes a user's userID, the date he gives the comment, the score he gives, and the verbal content
userInfo	basic information of this user, like name, gender, number of comments, number of photos, etc.
userCmt	all comments in the user's page, each comment includes a shopID, the time stamp he gives the comment with an accuracy of one minute, the score he gives, and the verbal content

TABLE I
TABLES IN THE DATABASE

IV. SYBIL DETECTION VIA BASIC FEATURES

In this section, we discuss the traditional features that can be exploited to distinguish the sybil users from the real users. The traditional features have been exploited in the previous works for behavior-based analysis including average interval time of post [9] [11], active day [11], clickstream pattern [4] and others. By jointly considering the techniques adopted by previous works and the practical issues of Dianping (e.g. system setting), we mainly consider four features, including Maximum Daily Number of Posting Reviews, $maxDailyCmt$, Maximum Daily Number of Posting Reviews within A Certain Interval, $max\ Daily\ Suspicious\ Interval\ Count$ ($maxDailySIC$), the average score, $aveScore$, and variance of the score, $scoreVariance$ in our basic detection set, which are introduced in detail as follow.

1) maxDailyCmt: We calculate the number of reviews of a user each day, and this term represents the maximum number among them. A normal user is less likely to post a huge amount of comments within a day.

2) maxDailySIC: We compute the interval between every two consecutive comments from the same user. If it is less than a threshold, it would be considered suspicious and the SIC of the day that he post the latter comment increases by one. This term represents the maximum count amongst these days. According to our experiments, the threshold is set to 3 minutes. A sybil user is more likely to post consecutive comments within a very short interval time, because they tend to maximize their financial rewards by posting as many reviews as possible in a fixed time.

3) aveScore: In each comment a user gives, it contains a score (from 1 to 5) to evaluate the services provided by the shop. We calculate the average of these scores. A sybil user tends to give an extremely high score to the shop which hires them for paid reviews while gives a low score to its rivals.

4) scoreVariance: Similar to $aveScore$, we calculate the variance of the collected scores. If a user is a sybil one, he would give almost every comment a 5 point score, therefore the variance would be lower than normal users.

It is important to point out that, it is possible to exploit other new features to launch the sybil detection. In the next section, we will show how to exploit the mobility information of the users for sybil detection.

V. EXPLOITING MOBILITY PATTERN FOR SYBIL DETECTION

In addition to the above mentioned traditional sybil detection features, we would like to introduce a new features, mobility information of the users, into sybil detection. The geography information based sybil detection is motivated by the observation that, in Yelp or Dianping, the online reviews on a restaurant show that this user should recently visit this restaurant, which can help us to figure out the user's recent location from this restaurant's location, and even the user's mobility traces in a certain period. According to the existing researches, a real user should be highly predictable and should be confined to a certain coverage. On the contrary, the sybils write paid reviews based on their missions. Thus, their mobility traces are not predictable, and normally beyond the coverage that a real user can have. This insight can help us classify the real users from the sybils.

To achieve mobility pattern based sybil detection, we perform the following steps. Firstly, for each targeted user, we crawl the published reviews as well as the publishing time. Since the reviews are given based on a shop or restaurant, we could obtain users' location traces based on the location of this shop/restaurant using Dianping's API. Such a location and publishing time pair can constitute a spatial, temporal footprint of this user. We suppose that sybil users have no ability to fool the API. Secondly, by collecting a user's footprint for a certain duration, we aim to have his activity region size, which is measured by radius of geographical minimum covering circle. Lastly, we adopt an entropy based definition for users' mobility predication. Then, we input them to SVM classifier for sybil detection.

A. Radius of Geographical Minimum Covering Circles

To capture a user's mobility pattern, we define the concept of smallest radius of a user's daily minimum covering circles, $minRadius$, which is used to measure the radius of a user's activity region. Suppose a user gives comments to n different shops a day. If $n \geqslant 2$, the activity region of the user is defined as the minimum covering circle which covers all n nodes. The radius of this circle is called $minRadius$. Fig. 2 shows several minimum covering circles of different users, in which each color represents a user. According to the existing researches, a real user's daily activity region should be within a certain geographical range. In other words, if a user's average $minRadius$ is too large, he is more likely to be a sybil user, like the green one in Fig. 2. Therefore, average $minRadius$ is used as a feature in the location-based feature set.

So here is our method to calculate $minRadius$. Suppose that there're n shopIDs, we firstly should acquire their longitudes and latitudes via API, and feed them into our database. Then, they will be transformed into a Cartesian coordinate system for the convenience of calculating.

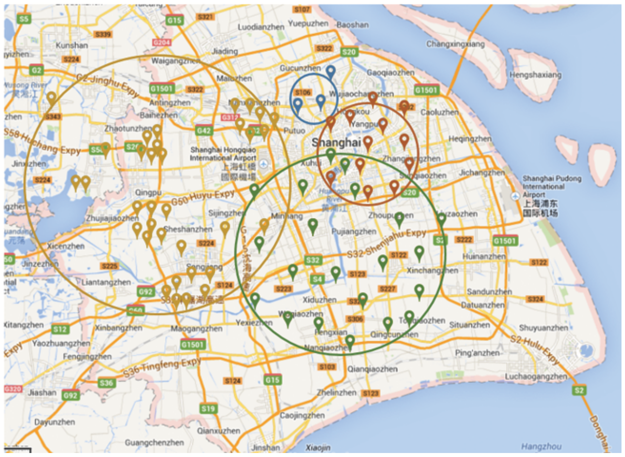

Fig. 2. Minimum Covering Circles

Algorithm 1 can calculate the radius of the minimum covering circle of n shops that a user has given comments to within a day [12]. Because the location information is stored in our database as several pairs of longitude and latitude, they should be firstly converted into x, y form as in Cartesian coordinate system. For that purpose, we use the Miller cylindrical projection to achieve this goal [13]. Then random increment is used to get the minimum circle covering n points.

The main idea of this algorithm is to determine two points that are on the border of the minimum covering circle. At the beginning, two points are chosen as endpoints of the diameter to construct a circle O. Then we add other points to this circle one by one. If any point is not in this circle, this point must be on the border of a new circle which covers the point itself. Provided that p_1 is found like that, this algorithm forms a new circle O_1 whose endpoints of the diameter are this one and another random point. Then the points that are already in O would be added to O_1 one by one. Again, if there is a point p_2 that is not in O_1, it must on the border of a new circle O_2 which covers the point itself. Finally, two points, p_1 and p_2, are found, and they are on the border of O_2. Therefore, this algorithm enumerates the third point that is already in O_2 to construct a bunch of new circles, knowing that three non-collinear points can determine a circle. We name these circles O_3, O_4, ..., O_m. After that, the circle O would be updated with the circle that has the largest radius among O_1, ..., O_m. So, after all iterations, the circle O is the smallest circle, and the radius of it is the $minRadius$ of the user in a certain day. The time complexity is $O(n^3)$.

Algorithm 1: MinRadius Algorithm

```
elicit $n$ shops' (longitude, latitude) that a user has
  given comments to within a day from the database;
set $s_i = (longitude_i, latitude_i)$ and
  $\mathscr{S} = (s_0, s_1, \cdots, s_n)$;
put $\mathscr{S}$ to Miller's cylindrical projection and attain
  the set $p_i = (x_i, y_i)$ and $\mathscr{P}$ = $(p_0, p_1, \cdots, p_n)$;
if $\mathscr{P}$ just has one element then
  return 0;
else
  take $p_0$,$p_1$ from $\mathscr{P}$;
  compute a initial circle $\mathscr{O}$ whose endpoints of
    diameter are $p_0$,$p_1$;
  while $\mathscr{P}$ is not empty do
    take $p_i$ from $\mathscr{P}$
    if $p_i$ is in $\mathscr{O}$ then
      continue;
    else
      $\mathscr{O} \leftarrow$ compute a circle whose endpoints of
        diameter are $p_0, p_i$;
      for $j = 1$; $j < i$; $j++$ do
        if $p_j$ is in $\mathscr{O}$ then
          continue;
        else
          for $k = 1$,$k < j$,$k++$ do
            $\mathscr{O} \leftarrow max(\mathscr{O}$,circumcircle($p_i$,$p_j$,$p_k$));
          end for
        end if
      end for
    end if
  end while
  return radius of $\mathscr{O}$;
end if
```

B. Entropy of Geographical Locations

In this part, *entropy* is proposed to measure the predictability of users' mobility. In information theory, *entropy*, or *Shannon entropy*, is a measure of the uncertainty in a random variable [14]. Song et al. find that a human being is more likely to go to some fixed locations, rather than explore new places [15]. So *entropy* can be used to model a user's mobility and determine whether he is sybil or not. In Fig. 2, the yellow circle's $minRadius$ is very large, but the locations are distributed densely, not sparsely, unlike sybil users. So, only taking $minRadius$ into consideration is not enough. Therefore, *entropy* is brought into our location-based feature set to enhance our detection.

The follows show how to compute *entropy*. Suppose a user has given comments to n different shops, and they belong to m regions. There're n_i shops locate in region i. So if we randomly pick one shop, the probability of it belonging to region i would be:

$$P_i = \frac{n_i}{n} \tag{1}$$

We use P_i as our probability mass function. Therefore, a user's *entropy* would be:

$$E(user) = -\sum_{i=1}^{m} P_i log_2 P_i \tag{2}$$

From Equation (1) and (2), we can learn that the *entropy* is determined by both of the number of regions m and the distribution of n shops. So, $E(user)$ has the maximum value when every region contains the same number of shops. In other words, the $entropy$ reaches its peak when P_is are uniform. On the other hand, $E(user)$ would be minimum if a user have n shops all in one region. In that case, $E(user)$ would be zero.

The existing works demonstrate that a user's active region should be confined to a certain range. If a user's comments mainly focus on one or two regions, the $entropy$ would be small, which means he is a real user. On the contrary, a sybil user's comments would be distributed sparsely on the map, thus leading to a larger $entropy$. Surprisingly, our real-world data experiments results show that the real users' mobility entropy is 1.68. This result confirms the recent work from Song et al., which shows that the entropy of a human being is 1.74 [7].

VI. Evaluation

After constructing these features, they are used to detect sybil users. First of all, we choose a hotpot restaurant on Dianping to crawl. Our team manually checks 2000 users from this shop to see if they're sybil users, based on the interval of their comments, verbal contents, personal information and so on. After that, there are 463 sybil users found among the 2000 accounts. The sybil rate is over 23%. Then these 2000 users are split into two parts and we use the first half as our training set, the second half as the test set. These 2000 users have a total amount of 67721 comments, which have also been crawled for future usage.

A. Classification Performances: Basic vs. Advanced

For classification purpose, Support Vector Machine (SVM) is chosen as our tool to measure the detection performance when using different feature sets [16]. The objective of our classification system is to classify each user as a sybil or normal user using the features chosen before. This detection is modeled as a binary classification problem in this paper. We feed the training set to SVM, and then use the test set to measure the performance. After classification, the performance of SVM is evaluated using the six metrics: $TP\ rate$, $FP\ rate$, $precision$, $recall$, $F-measure$ and $accuracy$. Note that these metrics are well known and broadly used in the evaluation of a classification system.

The performance of three kinds of feature sets is shown in Fig. 3. When using the basic feature set only, SVM only achieves the accuracy of 79.8%. However, by jointly considering the tradition and newly defined location-based features, the performance can be significantly improved to nearly 85%. From Fig. 3, we can clearly learn that SVM's performance is improved when using the combined feature sets. Not only does the FP rate go down, but the other metrics all rise.

B. Characteristic Analysis: Normal vs. Sybil

After our classification, we find that there're several distinctive lines separating sybil users and the normal ones. Only some qualitative analysis have been done before, so next, our paper are going to find out the attributes that distinguish sybil users in quantity. In this part, we will clearly state these features and analyse the quantitative differences in these characteristics between malicious accounts and ordinary ones.

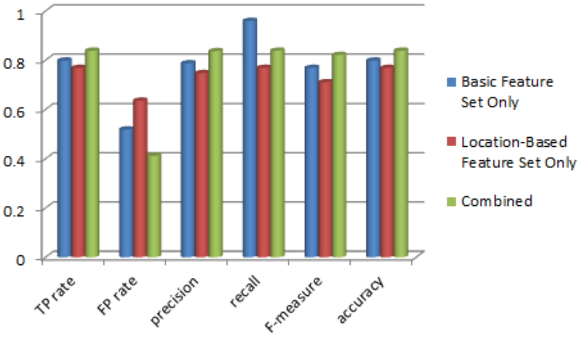

Fig. 3. Performance of SVM

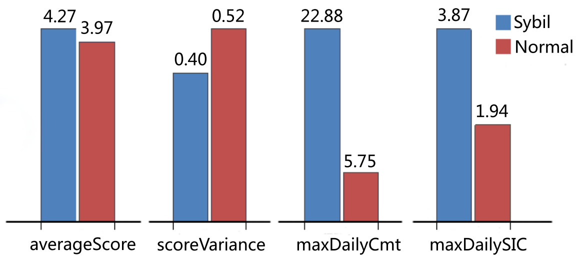

Fig. 4. Average of Basic Features

1) maxDailyCmt: It's shown to us in Fig. 4 that a sybil user's $maxDailyCmt$ is extremely high. On the contrary, a normal user would only post 5 or 6 reviews, maximum. This characteristic is the most important one to distinguish them. Sybil users log in to earn money, thus it would be a large amount of comments for them to post every time. When it comes to normal users, they comment for their own sake. So there is no need to give a bunch of reviews.

2) maxDailySIC: Fig. 4 also confirms our another hypothesis in IV-2. We can learn from Fig. 4 that sybil users' average $maxDailySIC$ is almost twice the number of a common user. Just as mentioned before, if a sybil user wants to make some profits from posting comments, he will do it in a very short time in order to make more profits.

3) aveScore: This paper assumes that a sybil user would always give a high score in IV-3. From Fig. 4, we can safely conclude that our assumption is correct. sybil user's $aveScore$ is approximately 4.3, while that of normal user is slightly below 4. It's not a small difference, given that the scale of the score is from 1 to 5, and only integers are allowed in the system.

4) scoreVariance: Surprisingly, the $scoreVariance$ difference between sybil and ordinary account shown in Fig. 4 is no so large as we thought before. Still, the variance of normal user is almost 1.5 times versus that of sybil users. Maybe it's because the sybil ones have learnt to disguise themselves like giving a low score occasionally as to increase their credibility.

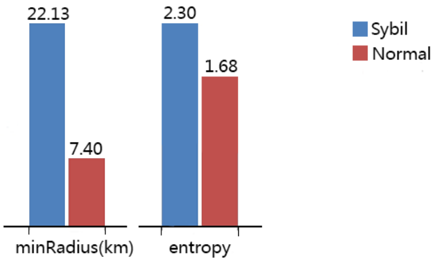

Fig. 5. Average of MinRadius and Entropy

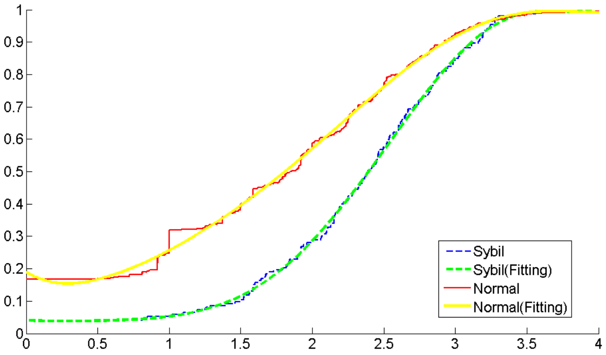

Fig. 6. CDF of Entropy

5) minRadius: To see the difference of $minRadius$ among sybil and normal users, we calculate the mean of $minRadius$ of every user. There is also a huge difference between them. From Fig. 5, the average $minRadius$ of sybil users are 22 kilometers, which is three times the number of ordinary accounts. It's also reasonable because malicious users' comments are fake, thus they are not constrained by geographical locations.

6) entropy: We also compare the mean of $entropy$ between users and plot them in Fig. 5. Clearly, sybil accounts have a greater one, about 1.5 times larger than normal users. Besides, the average $entropy$ of normal users is roughly 1.7, which is almost the same as in Song et al.'s experiment [7]. This result supports our hypothesis about the $entropy$ of these two kinds of users. Malicious accounts really have a higher $entropy$.

Besides, Fig. 6 shows the Cumulative Distribution Function(CDF) of $entropy$. The red line and blue line are raw data of $entropy$, while the yellow one and green one are lines generated by polynomial fitting. We can see sybil users' $entropy$ mainly concentrate in area where it's larger than 2, while most of normal users' are lower than 2.5. Besides, there're almost 20% of normal users whose $entropy$ are zero, meaning the shops they have given comments to are all in one region. On the contrary, this number of sybil users is less than 5%. Again, this figure clearly confirms our assumptions.

VII. CONCLUSION

Sybil attacks as well as crowdturfing are posing a serious threat towards online social networks. The existing studies demonstrate the existence of the crowdturfing system, which could be exploited by the sybil attackers to post spam reviews. Though there are extensive studies on sybil detection using behavior-based features or semantic analysis, few attention have been paid to location-based features that can be used to distinguish sybil users and normal ones. Our studies start from crawling online customers' comments from Dianping, then build two different feature sets, one of which contains location-based features and one does not, and apply the sybil detection on the users using these feature sets. Our results and further analysis show that location-based features are important factors to detect sybil users. Besides, there are distinctive differences in several characteristics between malicious accounts and normal ones, especially in location-based features. In a word, locations do talk.

ACKNOWLEDGEMENT

This research is supported by National Natural Science Foundation of China (Grant No. 61272444, 61161140320).

REFERENCES

[1] K. Lee, P. Tamilarasan, and J. Caverlee, "Crowdturfers, campaigns, and social media: tracking and revealing crowdsourced manipulation of social media," in *7th international AAAI conference on weblogs and social media (ICWSM), Cambridge*, 2013.

[2] G. Wang, M. Mohanlal, C. Wilson, X. Wang, M. Metzger, H. Zheng, and B. Y. Zhao, "Social turing tests: Crowdsourcing sybil detection," *arXiv preprint arXiv:1205.3856*, 2012.

[3] Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai, "Uncovering social network sybils in the wild," in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, 2011, pp. 259–268.

[4] G. Wang, T. Konolige, C. Wilson, X. Wang, H. Zheng, and B. Y. Zhao, "You are how you click: Clickstream analysis for sybil detection," *Sybil*, vol. 9, pp. 1–008, 2013.

[5] "dianping.com," http://www.dianping.com/.

[6] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, 2013.

[7] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[8] J. Xue, Z. Yang, X. Yang, X. Wang, L. Chen, and Y. Dai, "Votetrust: Leveraging friend invitation graph to defend against social network sybils," in *INFOCOM, 2013 Proceedings IEEE*, 2013, pp. 2400–2408.

[9] G. Wang, C. Wilson, X. Zhao, Y. Zhu, M. Mohanlal, H. Zheng, and B. Y. Zhao, "Serf and turf: Crowdturfing for fun and profit," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 679–688.

[10] "Dianping about us," http://www.dianping.com/aboutus.

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

[12] J. Elzinga and D. W. Hearn, "Geometrical solutions for some minimax location problems," *Transportation Science*, vol. 6, no. 4, pp. 379–394, 1972.

[13] O. M. Miller, "Notes on cylindrical world map projections," *Geographical Review*, pp. 424–430, 1942.

[14] C. E. Shannon, "The mathematical theory of communication," *Bell System Technical Journal*, vol. 30, pp. 50–64, 1948.

[15] C. Song, T. Koren, P. Wang, and A.-L. Barabási, "Modelling the scaling properties of human mobility," *Nature Physics*, vol. 6, no. 10, pp. 818–823, 2010.

[16] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.