Social Bots Detection on Mobile Social Networks

Cheng Binlin[1], Fu Jianming[2]
(Corresponding author: Cheng Binlin)

School of Computer Science, Wuhan University[1]
Wuhan 430079, Hubei, P.R. China
School of Computer Science, Wuhan University[2]
Wuhan 430079, Hubei, P.R. China
(Email: binlincheng@163.com)

(Received Dec. 30, 2015; revised and accepted Mar. 11 & Apr. 9, 2016)

Abstract

Mobile social networks have become very popular in recent years. The popularity of mobile social networks has attracted a large number of companies to do marketing on it. However, the social marketing suffer from social bots, a kind of bot accounts. In this paper, SBDSF(social bots detection based on the number of shared friends), a social graph based approach is proposed to detect social bots . SBDSF use the feature of social graph to detect social bots. We measure the effectiveness of SBDSF, the result of evaluation shows that SBDSF can achieve 96.1% accuracy and 95.1% precision using the Neural Network classifier.

Keywords: Mobile social networks, security, social bots

1 Introduction

Mobile social networks have become a trend in this modern time. It spreads rapidly attracting a lot of new users [4, 5].

As mobile social networks becomes increasingly popular in the world, The popularity of mobile social networks has attracted a large number of companies to do marketing on it. However, the social marketing suffer from social bots, a kind of bot accounts [3, 6, 8].

As the bot accounts registered by automatic program, social bots have many profile features different with real human accounts. For Example, social bots have less followers and post less tweets. There are some applications on Mobile social networks attempt to detect social bots by these profile features. To evade detection, the social bots have evolved from low-lever social bots to high-lever social bots . High-lever social bots have a certain amount of followers, publish tweets every day, the profile feature based approach can't detect the high-lever social bots.

As the profile features are easy to alter, we aim to answer the question: can we design a social bots detection approach which not relies on the profile feature? Some users announced that they no more used the Mobile social networks as they don't like the social bots , if the Mobile social networks provider can detect social bots in its system effectively, it can improve the experience of its users and attract more companies to do the Mobile social networks marketing in the Mobile social networks.

We propose SBDSF (social bots detection approach based on the number of Shared Friends), a social graph based social bots approach. SBDSF does not rely on the profile features of Mobile social networks account; instead, our approach is focused on the social graph structure.

In the paper, we plan to give the formal description of Shared Friends feature, and propose a social graph based social bots approach, then we validate the efficacy of the classification system based on the feature of the number of Shared Friends.

The rest of the paper is organized as follows. Section 2 introduces the goals of our system. Section 3 gives the approach our system. Section 4 presents our experimental classification results of social bots on Wechat. Section 5 concludes.

2 Goals

SBDSF aims to identify social bots accounts using the social graph based feature. Our design has the following main goals:

1) Effectiveness.
The approach should identify mostly social bots (low false positives), while limiting the number of normal followers considered social bots (low false negatives).

2) Efficiency.
The system should have a low performance overhead. It should be feasible to deploy to handle large mobile social networks by Microblog providers in practice.

3) Robustness.
The approach should be robust under various at-

tack strategies, and be immune to sophisticated evasion techniques. We assume that the attacker knows about the SBDSF techniques and will try to avoid detection.

3 Approach

In this section, we describe the basic idea of social bots detection, formalize the problem of social bots detection.

Definition 1. ***Follower*** *node v_j is a follower of node v_i if the edge $a = (j, i)$ is contained in the set of edges, followers are the incoming links of a node.*

Definition 2. ***Friend*** *node v_j is a follower of node v_i if the edge $a = (i, j)$ is contained in the set of edges, followers are the incoming links of a node.*

Definition 3. ***Shared Friend.*** *In the Weibo social graph, we use set $F(i)$ denoting the vertex which is the friend of $vertex_i$: $F(i) = \{ j \mid \{vertex_j\ is\ the\ friend\ of\ vertex_i\ \}$; then the common friends of $vertex_i$ and $vertex_j$ can be defined as:*

$$C(i, j) = F(i) \cap F(j). \tag{1}$$

Definition 4. ***Shared Friends Graph.*** *For a Weibo social graph $G(V, E)$, we define Shared Friends Graph as $G_C = (V_C, E_C)$. The V_C is the vertices set which is identical to the V of graph $G(V, E)$. An weighted edge $E_C = (i, j)$ is linking V_{C_i} and V_{C_j}, which stands for there are Shared Friends between V_i and V_j, and the weights of E_C is equal to the number of Shared Friends of V_i and V_j. Given a $E_C = (i, j)$, there is a equation shown as follow:*

$$Weights\ (E_C) = C(i, j). \tag{2}$$

In a summary, social graph $G(V, E)$ is unweighed directed graph, while Shared Friends Graph $G_C(V_C, E_C)$ is weighted undirected graph.

Example 1. *Consider a social graph $G(V, E)$ in Figure 1, the corresponding Shared Friends Graph is shown as G_C (V_C, E_C) in Figure 2. User b and d has a common friend which is User e; User c and d has two common friend User a and e.*

In order to see the difference of Shared Friends Graph on zombie friends and normal ones intuitively, we generate the Shared Friends Graph on the sample of normal Friends and zombie friends respectively. Figure 3 is the example of Shared Friends Graph on normal friends Sample, while Figure 4 is the example of Shared Friends Graph on zombie friends Sample. We can find that the number of Shared Friends among zombie followers is greater than that of normal ones significantly.

Definition 5. ***Community Aggregation Degree.*** *The Community Aggregation Degree denotes the degree of the community aggregation:*

$$D_{community} = E_{community}/E_{total} \tag{3}$$

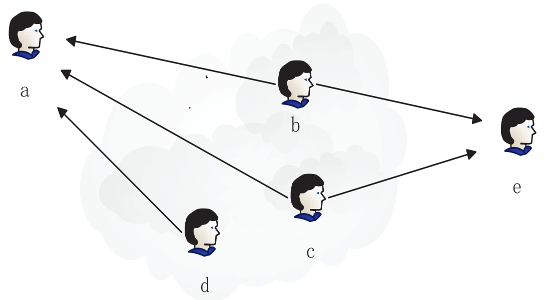

Figure 1: Social graph

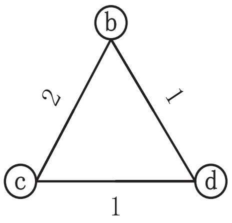

Figure 2: Shared friends graph

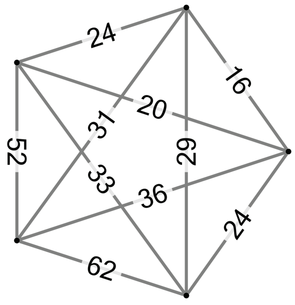

Figure 3: Sample of normal account

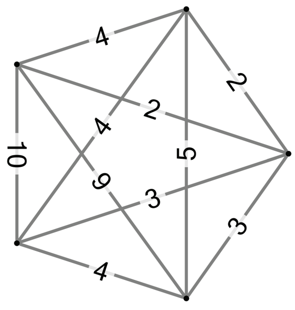

Figure 4: Sample of social bots

The $D_{community}$ denotes Community Aggregation Degree of the community, $E_{community}$ denotes the number of the edges in the Community, E_{total} denotes the number of the edges of the vertices in the Community.

We use the Community Aggregation Degree of the Shared Friends Graph the detect the social bots in this paper.

4 Evaluation

This section discusses the evaluation of our system.

4.1 Datasets

Wechat is the top mobile social networks in China, it is known to all that Twitter and Facebook are two popular ones in American and Erope while Wechat is the most widely used application in China [1]. Wechat offers API [7, 2], and we use it to crawl and collect data.

The sample set contains two subsets (Table 1): benign set and bots set. Benign set consist of 5,000 known, benign crawled ID from 10 seed ID which is randomly select, the bots set consist of 5,000 social bots which we purchased from 5 different sellers on Taobao site (Chinese version of Ebay).

Table 1: Evaluation dataset

Set	Source
5,000 benign accounts	crawled from Wechat
5,000 social bots	purchased on Taobao site

4.2 Evaluation Result

To increase the difficulty of detection, we mix randomly with the Benign set and bots set to build the Follower set. In our evaluation, bots set consist of 5,000 samples; the ratio of social bots account for all the followers is 40%, which is quite a low ratio in practice but nevertheless presents no problems for social bots detection.

We use the classifiers based on the Community Aggregation Degree of Community Aggregation Degree to detect the social bots. All the classifiers reported in the evaluation are computed using 10-fold cross validation. Table 2shows the evaluation result of different classification algorithms.

Table 2: Evaluation dataset

Classifier	Accuracy	Precision
Decision Tree	92.1%	93.1%
Neural Network	96.1%	95.1%
Support Vector Machines	95.0%	94.1%
Native Bayesian	90.3%	88.6%

It can be seen that Neural Network classifier has the best overall performance compared with other algorithms.

5 Conclusions

The popularity of mobile social networks makes it being a great platform to do marketing. Social bots become the major threat of social marketing. The social bots is in evolution, previous work can detect low-lever social bots but not high-lever social bots. To this end, we have proposed SBDSF, a social graph based approach to detect social bots, which use the feature that the number of shared friends among the social bots from a purchaser is usually greater than that of the normal users.

The popularity of mobile social networks makes it being a great platform to do marketing. Social bots become the major threat of social marketing. The social bots is in evolution, previous work can detect low-lever social bots but not high-lever social bots. To this end, we have proposed SBDSF, a social graph based approach to detect social bots, which use the feature that the number of shared friends among the social bots from a purchaser is usually greater than that of the normal users.

References

[1] F. Gao and Y. Zhang, “Analysis of wechat on iphone,” in *2nd International Symposium on Computer, Communication, Control and Automation*, vol. 69, pp. 278–281, 2013.

[2] C. H. Lien and Y. Cao, “Examining wechat users’ motivations, trust, attitudes, and positive word-of-mouth: Evidence from china,” *Computers in Human Behavior*, vol. 41, pp. 104–111, 2014.

[3] A. Muthana, A. A. A. Ghani, and R. Mahmod, “It cannot get away: An approach to enhance security of user account in online social networks,” *International Journal of Computer Science and Network Security*, vol. 15, no. 4, pp. 1, 2015.

[4] V. Nincic, M. Chang, and F. Lin, “Dmlrid-an xml-based proof-of-concept mobiledrm framework for sharing learning contents amongmobile networks,” *International Journal of Information and Electronics Engineering*, vol. 3, no. 2, pp. 180, 2013.

[5] M. Romoozi and H. Babaei, “A P2P fuzzy reputation system based on security policies for mobile ad-hoc networks,” *International Journal of Information and Electronics Engineering*, vol. 3, no. 5, pp. 481, 2013.

[6] S. Sarpong, C. Xu, and X. Zhang, “An authenticated privacy-preserving attribute matchmaking protocol for mobile social networks,” *International Journal of Network Security*, vol. 17, no. 3, pp. 357–364, 2015.

[7] L. Wu and X. Yang, “A research on the services of university mobile library based on the wechat public platform [j],” *Research on Library Science*, vol. 18, pp. 013, 2013.

[8] P. Wuttidittachotti and T. Daengsi, “Quality evaluation of mobile networks using voip applications: A case study with skype and line based-on stationary tests in bangkok,” *International Journal of Computer*

Network and Information Security, vol. 7, no. 12, pp. 28, 2015.

Cheng Binlin received his M.S degree from Wuhan University in 2009, China. Currently he is a doctor candidate in Wuhan University. Student member of China Computer Federation. His main research interests include software security and network security.

Fu Jianming received his Ph.D. degree from Wuhan University in 2000, China. Now he is a professor in Wuhan University. Senior member of China Computer Federation. His main research interests include software security and network security.