Modeling for User Interaction by Influence Transfer Effect in Online Social Networks

Qindong Sun, Nan Wang
Shaanxi Key Laboratory of Network Computing and Security, Xi'an University of Technology
Xi'an, China
sqd@xaut.edu.cn
wn_xaut@163.com

Yadong Zhou
MOE KLINNS Lab, Xi'an Jiaotong University
Xi'an, China
ydzhou@ xjtu.edu.cn

Hanqin Wang, Liansheng Sui
Shaanxi Key Laboratory of Network Computing and Security Xi'an University of Technology Xi'an, China
171413696@qq.com
slsitrd@163.com

***Abstract*—User interaction is one of the most important features of online social networks, and is the basis of research of user behavior analysis, information spreading model, etc. However, existing approaches focus on the interactions between adjacent nodes, which do not fully take the interactions and relationship between local region users into consideration as well as the details of interaction process. In this paper, we find that there exists influence transfer effect in the process of user interactions, and present a regional user interaction model to analyze and understand interactions between users in a local region by influence transfer effect. Based on real data from Sina Weibo, we validate the effectiveness of our model by the experiments of user type classification, influential user identification and zombie user identification in online social networks. The experimental results show that our model present better performance than the PageRank based method and machine learning method.**

***Index Terms*— Online social networks, microblogging, user interaction model, influence transfer effect.**

I. Introduction

With the great impact of social networking services, more and more users communicate through online social networks for the communication of emotion and information. Compared with the traditional Internet communication platform (e-mail, forums, newsgroups), the huge number of users and efficient dissemination of information have become the characteristics of social networking sites [1]. Many issues of online social networks such as evolution models and network dynamic process have been hot topics of academic research. User interaction is one of the most fundamental and important features in online social networks. There are already many research results about user interaction such as interaction pattern and the regular range of the user interaction behavior. These researches are mostly focus on the interactions and relationship between the two adjacent nodes. Even the models based on the nodes in a local zone are also focus on the connected nodes. However, in the actual network, the interactions and influence between different nodes with two or three hops are away of importance to the evolution of network and some dynamic processes in the network [2]

As information sharing is one of the most important functions of online social networks, the interaction can even exist in the disconnected users. Therefore, the interaction behaviors not only exist in adjacent users but also exist between users who are not directly connected. The existing researches could not fully analyze the interaction between users in random location within local zone. During the information diffusion, users can influence others with the interaction between different users. This situation is similar to the energy transfer in field model in the physical theory. The field model could be used to describe the interaction and relationship between different objects based on various kinds of energy transfer. The influence transfer effect is formed by the interaction behaviors in online social networks.

According to our analysis of user interaction in online social networks, in this paper we proposed a regional user interaction model based on influence transfer effect for analyzing and understanding user interactions in online social networks. Based on the data captured from Sina Weibo*, it is validated that this regional user interaction model is capable of identifying node influence and detecting zombie users, which has better performance than the traditional methods according to the experiments. In this paper, our model is presented for the case of directed social network. Considering that undirected network and bipartite network can be regarded as special cases of directed network, the model for directed network can be improved to be used in these two types of networks mentioned above.

II. Related works

For the user interaction, there are already some researches on the behavior spreading or local user interaction. Centola analyzed how the users got influenced by the behaviors of their neighbors in health-interest online social network [3]. Cascading model is a frequently-used method for behavior spreading. Holthoefer and Fowlera are both utilized the cascade model to study the interaction between different users in online social networks [4][5]. These exiting researches can express the transitivity of user behavior, but the relation or influence between users in a random distance within a local zone have not been well analyzed.

* http://www.weibo.com/

User influence is an essential feature in online social networks. Cha used the follower, retweet and mention features to measure user influence directly [6]. For a more complex method based on user features Cappelletti used the tendency of a user being retweeted or mentioned to evaluate the user influence [7]. Bakshy used the cascade size of information diffusion to evaluate the user influence [8]. PageRank algorithm is a common method for user influence. Tang has used the PageRank to rank user influence in a weighted social network [9]. We evaluate user influence in a different way based on the user interaction model proposed in this paper.

III. Methodological Analysis

Inspired by the energy field model, the regional user interaction model is proposed in this paper. In this model, the influence used to analyze the user interaction in a certain region. We analyze the impact of influence transfer on user interactions and carry out three experiments of user type classification, influential user identification and zombie user identification to verify the effectiveness of our model. The schema of the regional user interaction model is shown in Fig. 1.

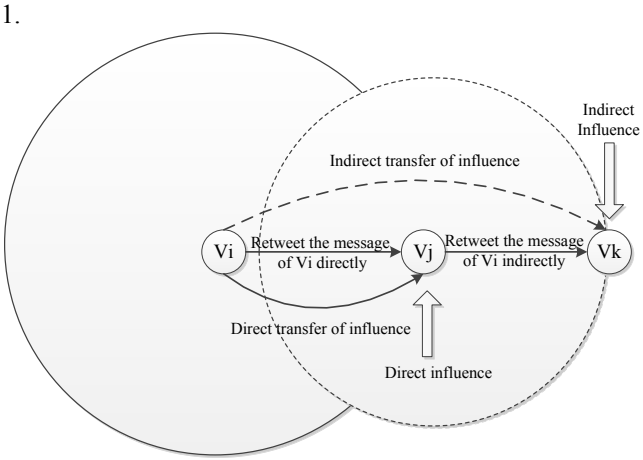

Fig.1. Schema of regional user interaction model

Suppose that there are three nodes V_i, V_j, V_k and the relationships of these node are V_j is the follower of V_i and V_k is the follower of V_j. The specific process of user interaction effecting on influence could be described as the following process: If V_j retweet information tweeted by V_i, it is thought that V_i has effect on the behavior of V_j, V_i transfer part of influence to V_j as the tweet content is attractive or V_i has a very high reputation. The influence of V_i will increase and here the transferred influence we regard as the some as the influence increased. If V_k retweet the information retweeted by V_j from V_i, it is thought that V_i has effect on V_k, and V_k get part influence of V_i. The influence works on other nodes will be weakened with the increase of distance between different nodes. According to the distance the influence of a user has two types: the direct influence ($E_{r=1}$) and the indirect influence ($E_{r>1}$), and the influence can be expressed as:

$$\begin{cases} E_{r=1} = f_1(U_P, C_{rt}) \\ E_{r>1} = f_2(E(V_{fa}), r) \end{cases} \tag{1}$$

In Eq. 1, when user V_n retweets the tweets of V_i and the distance between the two nodes equals 1, the influence ($E_{r=1}$) increased by V_i is related to its user characteristics U_p and C_{rt} the tweet count retweeted by V_n from V_i. If the distance is greater than 1, the influence ($E_{r>1}$) increased by V_i is related to the influence transferred from V_i to V_f(the father node of V_n), and also to the distance between V_i and V_n. As shown in Fig. 1, the influence of V_k's father node is the influence transferred from V_i to V_j. U_p is the user characteristics of V_i, including the follower count, following count and tweet count, etc.

As shown in Fig. 1, V_j retweets tweet from V_i, when the distance between V_i and the node retweeting V_i's tweet equals 1, the influence acquired by V_i is defined as follows:

$$E_{r=1} = \frac{d\left[B(V_i,t)\cdot\dfrac{C_{rt}(V_j,t)}{C_p(V_j,t)}\right]}{dt} \tag{2}$$

where $C_{rt}(V_j,t)$ is the retweet count of V_j from V_i at time t, $C_p(V_j,t)$ is the total tweet count of V_j at time t, $B(V_i,t)$ is the follower count of V_i at time t.

When the distance is greater than 1, there being some intermediate nodes in the retweet process, the influence transferred from V_i to the remote nodes is related to their father nodes' influence acquired from V_i. The measurement of the ration of transferred tweet based on statistic method can protect the precision of experimental results while effectively reducing the time complexity. Therefore, the influence acquired by V_i is defined as follows:

$$E_{r>1} = \begin{cases} E(V_{fa})\cdot k^r, r \in Z^+ \\ 0, \qquad r \to \infty \end{cases} \tag{3}$$

where k is the retweet probability, r is the path length between the nodes, $E(V_{fa})$ is the influence value of V_k's father nodes acquired from V_i.

According the above two kind of influence, the total influence of V_i is:

$$E = \sum_n E_{r=1} + \sum_m E_{r>1} \tag{4}$$

where n denotes the number of users have interaction relationship on the target node V_i with the distance r equaling 1, and m denotes the number of users causing the influence increase with the distance r greater than 1.

IV. Empirical Results

In this section we will show the analysis results by applying our model into the user types identification.

A. DataSet

The regional user interaction model can be easily performing research on a directed network, undirected network and bipartite network. To validate the effectiveness of this model, it will be applied to Sina Weibo, the most influential Chinese microblogging website.

By Sina *API* and the program of crawlers, we have collected the sample data of 500 Sina Weibo users and another one hundred thousand unfiltered user profiles in February 2013.

B. Classification of Typical Types of Users

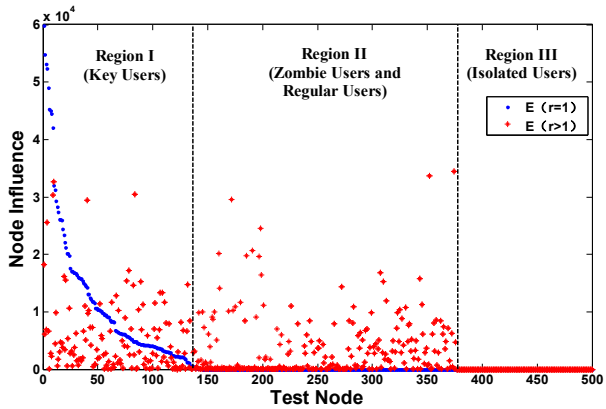

Fig.2. The distribution of transferred influence of each node

According to the influence, we could estimate whether a user can have impact on the interaction behavior of others. Different influence makes different user types, therefore, we divided users into key users, regular users, zombie users and isolated users according to the influence transfer value. The key users are the groups who showed great influence in the experimental period, regular users and zombie users are the inactive groups but with followers of a certain scale, and isolated users are those with zero follower and followee number in microblogging networks. This kind of user could barely have impact on behavior of other users during the information diffusion [11].

Fig. 2 is the schematic diagram of the influence of nodes in the experimental period. We used the blue dots and red dots represent $E(r=1)$ and $E(r>1)$ of nodes respectively ($E(r=1)$ is the direct influence caused by retweet of the adjacent users and $E(r>1)$ is the indirect influence caused by retweet of the remote users), and ranged the influence of the node greater than zero in descending order on the basis of their $E(r=1)$. Region I is the area of the active users within the experimental period in which the key users are located. Region II is the gathering area of zombie users and the regular users that are not active in this experimental period but with a certain scale of followers. Region III is the area of isolated microblogging users who do not have a decisive role in conducting the public opinions in the entire network.

C. Influential User Identification

We evaluated influential users based on the value of transferred influence of users. In order to verify the effectiveness of regional user interaction model, we have carried out some comparative experiments with PageRank-based methods [12]. The PageRank-based methods use the quantifiable features (e.g. retweet count and follower count) and the PageRank process to evaluate user influence. On the basis of the follower count and the tweet count, the comparison of top 50 influential users between PageRank based method and the regional user interaction model is shown in Fig. 3.

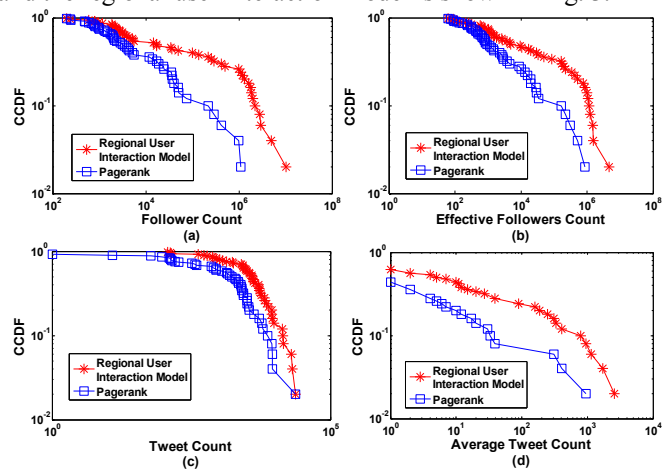

Fig.3. The attributes distribution of top 50 users from regional user interaction model and PageRank: a) The contrastive distribution of the amount of follower; b) The contrastive distribution of the effective followers; c) The contrastive distribution of the amount of tweet; d) The contrastive distribution of the average amount of retweet

According to the analysis of the experimental data, we have obtained that the retweet between disconnected users are often existed in the users with large scale of followers. In the present study stage the coefficient $B(V_i,t)$ in *Eq*.2 denotes the follower count of V_i at time t. Therefore, the PageRank method is based on the follower count. We will improve our model in the feature research and put forward more reasonable parameters.

As shown in Fig. 3, the key users identified by regional user interaction model have a larger scale of the four selected features which could have great impact on user interaction. The key users identified by our model are superior to the users indentified by the PageRank-based method.

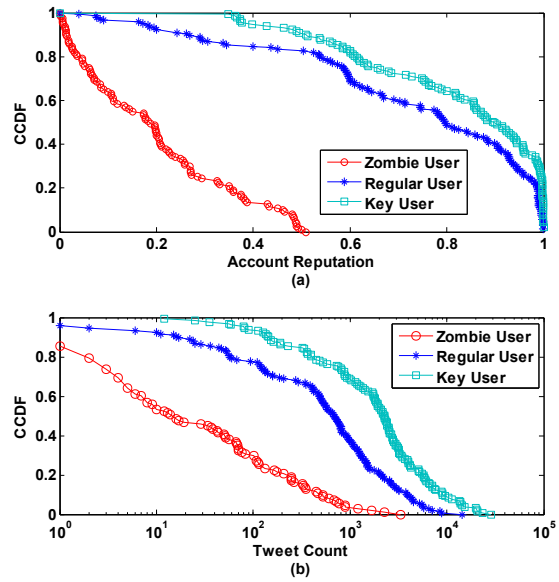

Fig.4. The multi-index contrastive schematic of key users, regular users and zombie users: a) The contrastive distribution of account reputation; b) The contrastive distribution of tweet count

D. Zombie User Identification

In Region II shown in Fig. 2 those zombie users can be identified with $E(r>1)$. The regular users and the zombie users are both with zero $E(r=1)$, but there are less microblogging interactions with the follower for the zombie users so that their $E(r>1)$ is weaker. Therefore, in accordance with the $E(r>1)$, the zombie users can be identified effectively. In Chu's studies of Twitter [10], the methods are proposed for identifying zombie users through the use of information entropy, account reputation and some other features.

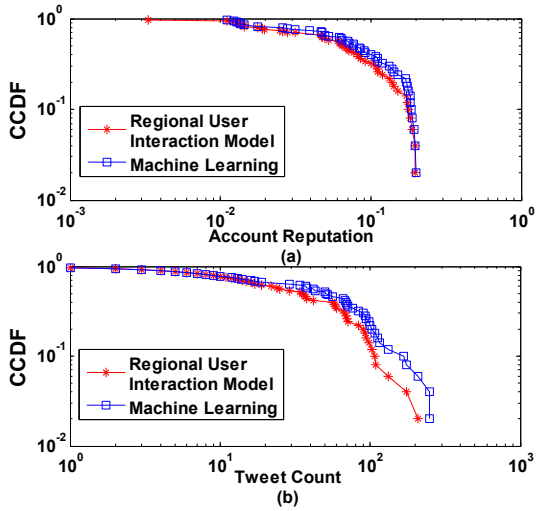

Fig.5. The contrast of zombie users' indentified by the regional user interaction model and machine learning (random forests): a) The contrast of Account Reputation; b) The contrast of tweet count

We choose the Account Reputation as well as the number of tweet as the assessment criteria to investigate the efficiency of our model, and draw out the CCDF (Complementary Cumulative Distribution Function) for the key users, regular users and zombie users in the regional user interaction model in Fig. 4. As shown in Fig. 4(a) and Fig. 4(b), the characteristics of these three types of users can be well shown from each other through the division of the regional user interaction model.

The machine learning is an effective method to distinguish user types [10], but if there is insufficient features for the machine learning method, the result may be different. Then we make a comparative experiment for analyzing the information of identified zombie users from our regional user interaction model and machine learning method (random forests algorithm).As Fig. 5 (a) and (b) manifest that the zombie users who identified by the local/regional user interaction model have a lower number of Account Reputation and tweet. The final result shows that the zombie users identified by the regional user interaction model have a higher accuracy. For this reason, we conclude that regional user interaction model can identify deeper hidden zombie users.

V. Conclusion and Discussion

In this paper, we analyzed the user interactions based on the influence transfer effect in online social networks, and found that traditional models are not suitable for describing dynamic process in online social networks. On this basis, we present a regional user interaction model to study the interaction process of different users in online social networks. This model can evaluate user influence and can be easily applied to user identification, user behavior analysis, etc. The experimental results show that this model can effectively identify important users and zombie ones and that the ranking of user influence can reasonably reflects the distribution of users. The model presented in this paper is of flexibility which could be applied to some other research fields such as recommender systems in and commercial marketing in the future works.

Acknowledgment

The research presented in this paper is supported in part by the National Natural Science Foundation (61172124, 61202392), the International Cooperation Project of Shaanxi Province (No.: 2013KW01-01), and The Key Laboratories Development Program of Shaanxi Province Education Department (No.: 13JS085).

References

[1] A R M Teutle. "Twitter: Network Properties Analysis." Electronics, Communications and Computer, 20th International Conference on, IEEE, 2010, pp.180-186.

[2] Y D Zhou, X H Guan, Q H Zheng, et al. "Group dynamics in discussing incidental topics over online social networks". Network, IEEE, vol. 24, issue. 6, 2010, pp. 42-47.

[3] D Centola. "The spread of behavior in an online social network experiment", Science, 329(5596), 2010, pp. 1194-1197.

[4] J B Holthoefer, R A Baños. "Cascading behaviour in complex socio-technical networks", Journal of Complex Networks, Vol. 1, 2013, pp. 3-24.

[5] JH Fowler, NA Christakis. "Cooperative behavior cascades in human social networks", Proceedings of the National Academy of Sciences, Vol. 107, issue. 12, 2010, pp. 5334-5338.

[6] M Cha, H Haddadi, F Benevenuto, "Measuring User Influence in Twitter: The Million Follower Fallacy", 4th AAAI Conference on Weblogs and Social Media. 2010, pp. 10-18.

[7] R Cappelletti, N Sastry. "IARank: Ranking Users on Twitter in Near Real-time, Based on their Information Amplification Potential", Social Informatics, IEEE, 2012, pp. 70-77.

[8] E Bakshy, J M Hofman, W A. Mason. "Everyone's an influencer: quantifying influence on twitter", the fourth ACM international conference on Web search and data mining, 2011, pp.65-74.

[9] X N Tang, C C Yang. "Ranking user influence in healthcare social media", ACM Transactions on Intelligent Systems and Technology, Vol. 3, issue. 4, 2012, No.73.

[10] Z Chu, S Gianvecchio, H N Wang, et al. "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" Dependable and Secure Computing, vol. 9, issue. 6, 2012, pp. 811-824.

[11] Zhang Z, Zhao B, Qian W, et al. "Generating profiles for a lurking user by its followees' social context in microblogs", Web Information Systems and Applications Conference, 2012, pp, 135-140.

[12] Y Chang, X Wang, Q Mei, "Towards Twitter context summarization with user influence models," Proceedings of the sixth ACM international conference on Web search and data mining, 2013, pp. 527-536.