

A Community Influence Prediction Calculation Method Based on Link Similarity in Social Network

Song Qianqian

College of Information, Mechanical and Electrical
Engineering
Shanghai Normal University
Shanghai 200234, China
e-mail: 296078985@qq.com

Zhang Bo

The Third Research Institute of Ministry of Public
Security
Shanghai, China
e-mail: zhangbo@shnu.edu.cn

Yang Tao*

The Third Research Institute of Ministry of Public
Security
Shanghai 200234, China
Corresponding author e-mail: yangtao@stars.org.cn

Hu Sihui

College of Information, Mechanical and Electrical
Engineering
Shanghai Normal University
Shanghai 200234, China
e-mail: 756148873@xyz.com

Abstract—In this paper, we propose a community influence prediction calculation method based link similarity in social network. First, we present a model of community influence prediction, and then describe its related models formally. Secondly, we give a set of factors, including label similarity between users, similarity between uses based common neighbors, and their calculation methods, which can evaluate the similarity between users and then estimate whether a new user would join the community or not in near future. Furthermore, we address a method of renewing the community structure and influence respectively. Finally, experimental results verify the effectiveness and feasibility of our proposed method.

Keywords—Link similarity; Community Influence; Prediction; Social network

I. INTRODUCTION

With the rapid development of the social network, online community has become one of the most influential organizations for information acquiring, sharing, and propagating [1]. In recent years, there are many large-scale and continuously expanding communities which exist in large social network sites, such as Facebook, Twitter, and Friendster [2]. Many existing studies have proven that community influence analysis can help us to find interesting information, and also use the community resources reasonably [3]. However, how to evaluate the likelihood of a user joining the community and then estimate the future influence of community is still an important problem of related research field.

Traditionally, influence researches in social network are focused on the individual influence calculation of users, and fixed-scale community. There is little attention which has been paid to scale-expanded community influence analysis [4]. Wu Liang and his colleges proposed an influence evaluation method based on topic and the correlations between community and users to get more detailed and accurate results of user influence[5-6]. Cha M

and his colleges analyzed user influence based on the degree of user through some factors, such as the times of user be mentioned in the content and the times of user's content be forwarded or reference [7]; Page L and his colleges analyzed user influence by PageRank sort based on relationship between users [8]; Zhou Wenan proposed UserRank model based on the number of user's friends and quality to analysis user influence [9]. Skehill R proposed InfluenceRank model based on the concept of spread factor of microblog to analysis user influence [10]. Zhao Fazhen revealed that the more the site be linked, the higher utilization of the site, the greater influence of the site in his work [1]. Guo Jingshi et al defined influence activation probability of each other according to the edge attribute between nodes, and gave a measurement method for user influence, namely the scope of influence dissemination and the delay of influence dissemination. And then, they dig out the core nodes of the community itself by comparing the node influence to evaluate regional influence in the community [11].

In this paper, we propose a community influence prediction calculation method based link similarity in social network. First, we define the community influence prediction model and relevant properties; secondly, we give a set of factors, including label similarity between users, similarity between uses based on common neighbors, and their calculation methods, which can evaluate the similarity between users and then estimate whether a new user would join the community or not in near future. Finally, we address a method of renewing the community structure and influence respectively. Experimental results show the feasibility and effectiveness of our proposed method.

II. COMMUNITY INFLUENCE PREDICTION CALCULATION METHOD

2.1 Community influence prediction model

Community is composed of many users and their relationships. Due to the distinctiveness of individual user and the complex of relationships between users, there is a certain complexity of community structure and community influence. In this paper, we give the community influence prediction model and describe the community influence prediction model formally based on the description method of social network graph model.

Definition 1 Community influence prediction model (CIPM). Community influence aims to describe community influence in new social network structure, which is formed by user's joining or exiting, and connecting or disconnecting between users in the future.

In this paper, the structure in social network is described as $G = (V, E)$, $V = V_{in} \cup V_{out}$, $V_{in} = \{v_1, v_2, \dots, v_n\}$ denotes n users in the community, namely n nodes, $V_{out} = \{u_1, u_2, \dots, u_m\}$ denotes m nodes which are outside of the community, and $E = \{(v_1, v_2), (v_1, v_3), \dots\}$ denotes the relationships between nodes in social network. The node pairs are ordered and relationships between users are two-way, such as: if $(v_1, v_2) = 1$ and $(v_2, v_1) = 0$, namely user v_1 has paid attention to user v_2 , but user v_2 has not paid attention to user v_1 .

The current structure of community is described as $G_s = (V_s, E_s)$, $V_s = V_{in}$, E_s denotes relationship between users in community. The future community structure is described as $G_s' = (V_s', E_s')$, V_s' denotes the nodes in the community in the future, and E_s' denotes the relationships between users in the community in the future.

Community influence prediction model is described as $CIPM(G_s') = f(G, G_s, G_s')$, $CIPM(G_s')$ denotes the influence of community G_s' , function $f(G, G_s, G_s')$ denotes influence prediction of community, and G_s' is based on G and G_s .

2.2 Factors of community influence prediction

The key of community influence prediction model is to estimate the link probability of user to other users, which will determine whether user join the community or not, and it is related to the label similarity between users and the common neighbors of them.

Definition 2 The prediction credibility for linking between users (CP). For user u and user v , the prediction credibility for linking between users describes the prediction trust value of linking u to v in the future.

Users have their own labels and the label of user v is described as $lab_v = \{attribute1, attribute2, \dots\}$, where

attribute1 denotes one of his interest, such as reading; the label of user u is described as $lab_u = \{attribute1, attribute2, \dots\}$. Label similarity between user v and user u is calculated as:

$$Lab_{uv} = \frac{|lab_v \cap lab_u|}{|lab_v \cup lab_u| + 1} \quad (1)$$

On the above formula, the degree of label similarity between users is determined by their own interests together. In order to avoid the situation of non-common hobbies, we add 1 in the denominator to make the calculation method feasible in any circumstances.

The common neighbors between user v and user u are the users who link to user v and user u simultaneously. Similarity between user v and user u based on common neighbors is calculated as follows:

$$Sim_{uv} = \begin{cases} \sum_{w \in CN(u,v)} P(link/w) \times Sim_{vw} \times Sim_{uw}, & \text{if } u \text{ has not} \\ & \text{paid attention to } v \\ S_i, & \text{otherwise} \end{cases} \quad (2)$$

$$P(link/w) = \frac{|O(v) \cap O(u)|}{\sum_{i \in V, i \neq w} D_{iw} \times \sum_{j \in V, j \neq w} D_{wj}} \quad (3)$$

From the above formula, we propose that if user v has relationship to user u , the calculation method as follows:

$$S_i = Sim_{vw} = \frac{1}{\sum_{j \in V, j \neq v} D_{vj}} \times \frac{1}{\sum_{j \in V, j \neq w} D_{wj}} + \sum_{h \in CN(v,w)} P(link/h) \times Sim_{vh} \times Sim_{wh} \quad (4)$$

In the formula (2), Sim_{uv} is the similarity between user u and user v ; $CN(u, v)$ is the collection of common neighbors of them; $P(link/w)$ denotes probability of linking between nodes, and in the situation that they regard w as their common neighbor; Sim_{vw} and Sim_{uw} represent similarity between user v and user w and similarity between user u and user w respectively; S_i is calculation method for the situation that user u has paid attention to user v . In formula (2), if user u has not connected to user v , we calculate the similarity between them; otherwise we can use formula (4) to calculate their similarity.

In the above formula (3), D denotes degree matrix, $O(v)$ and $O(u)$ represent the set of out-degree neighbors of user v and set of out-degree neighbors of user u respectively; $\sum_{i \in V, i \neq w} D_{iw}$ and $\sum_{j \in V, j \neq w} D_{wj}$ represent in-degree and out-degree of user w respectively.

The prediction credibility which describes the likelihood of user u connecting to user v in the future is calculated as follows:

$$CP_{uv} = \exp\{-(\alpha Lab_{uv} + (1-\alpha)Sim_{uv})\} \quad (5)$$

Where α is the influence parameter between $[0,1]$. By selecting random probability p , u will link to v if the condition $p < CP_{uv}$ is satisfied, otherwise, u would not link to v .

This paper proposes a method for renewing the community structure after determining whether the nodes are linked or not, as follows:

Assume that u is outside the community. The influence of node which is in the community to node u is calculated as,

$$Influ_{in} = \sum_{i \in I(u)} \frac{\sum_{j \in V, j \neq i} D_{ji}}{\sum_{l \in V, l \neq u} D_{ul}},$$

where $I(u)$ is the set of nodes in the community which node u links to.

The influence of nodes in the community to node u outside the community is determined by the number of nodes' in-degree and out-degree of node u . If $Influ_{in} \geq Influ_{out}$, node u would join the community; otherwise, the node would not join the community.

2.3 Calculation of community influence prediction

Due to the new users joining and old users exiting with the time passing, the community influence must be dynamically renewed. The renewing calculation of community influence prediction is as follows:

$$CIPM(t + \Delta t) = (1 - \rho) \times CIPM(t) + \Delta influ \quad (6)$$

In the above formula, $CIPM(t + \Delta t)$ denotes the community influence after time Δt ; ρ is the community influence erosion factor after time Δt ; $CIPM(t)$ denotes the current community influence of community; $\Delta influ$ is the change of community influence after time Δt .

In addition, community influence is related to the nodes which join the community. Then, the change of community influence is:

$$\Delta influ = \sum_{u \in NewAdd(G_S)} \left(\sum_{m \in In(u)} \frac{\delta_{mu}}{\sum_{l \in V, l \neq u} D_{ml}} \right) \quad (7)$$

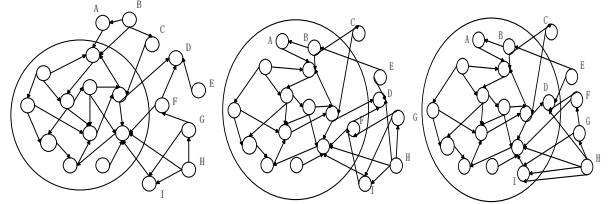
In above formula, $\Delta influ$ is the change situation of community influence for new node joining community; $NewAdd(G_S)$ is the set of nodes new joining the community G_S ; $In(u)$ is the set of in-degree node of node u ; δ_{mu} is used to marked whether in-degree node of node u is in the community or not. if m is outside the community, δ_{mu} is 1, otherwise δ_{mu} is 0.

III. EXPERIMENTAL ANALYSIS

In this paper, we testify our proposed method based on the data set which was collected manually by us through the latest data of microblog. After analyzing and preprocessing the data, there are about 10,000 records for our examination. For the convenience of experiments, select 4000 records randomly.

First, we verify the accuracy of prediction credibility. This experiment selected 20 users to analysis the label similarity between users and the similarity between nodes based common neighbors, and then, we calculate the prediction credibility for links between users. In this experiment, we selected different p for analyzing the predict situation of user joining the community and the actual situation of the community, so that we can get the relative optimal value of P , and verify the effectiveness and feasibility of our proposed method.

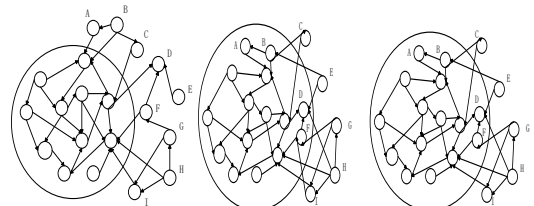
(1) When $p=0.95$, the community current structure, the prediction structure of community in the future, and the actual structure of community in the future are shown in figure 1.



(1) Community current structure (2) The prediction structure of community in the future (3) The actual structure of community in the future

Figure 1 Community structure graph

(2) When $p=0.90$, the community current structure, the prediction structure of community in the future, and the actual structure of community in the future are shown in figure 2.

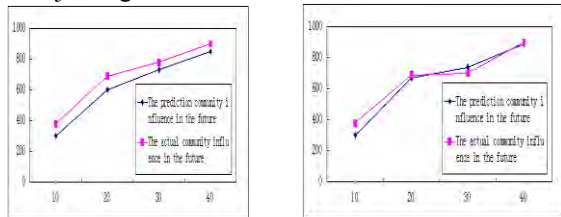


(1) Community current structure (2) The prediction structure of community in the future (3) The actual structure of community in the future

Figure 2 Community structure graph

In the above, we know that when p is 0.95 the prediction structure of community in the future and p is 0.90 the actual structure of community in the future. By contrast, we can know that when p is bigger not mean the prediction is more accurate, when $p=0.95$, the prediction is less accurate than $p=0.90$. In Fig. 1(2), node D has not joined the community, but actual node D joined the community. In graph 2(2), node D joined the community, from graph 2(3) we can see that the prediction structure is almost same with the actual structure of the community in the future, so when $p=0.90$, the prediction is more accurate than $p=0.95$.

Secondly, we verify the accuracy of community influence prediction model. In this examination, we selected 300 records at random, and then calculated the community influence which is consisted by the 300 users. Based on the calculation method, we calculated the prediction values of community influences based on prediction values of joining some new nodes.



(1) $p=0.95$, contrast of community influence
(2) $p=0.90$, contrast of community influence

Figure 3 Contrast of community influence

As shown in Fig.3 we can see that the results of community influence prediction were impacted by parameter p . A proper value of p would make the result of community influence prediction almost same with the actual community influence, namely the prediction will more accurate. So it is necessary for use to get a proper p so that he can make decision whether he should join the community or not in the future.

IV. CONCLUSION

Community is an important platform for users' communication. In this paper, we propose a community influence prediction calculation method based on link similarity in social network. We establish the social network model, the current structure of community model and community influence prediction model formally

according to the label similarity between users and common neighbor similarity between nodes. Then, we propose the method of calculating the prediction credibility for linking between users, which determines whether users join the community or not, and then we present the renewing method of community structure. Finally, we give the method of calculating the community influence in the future to realize the target of community influence prediction. Next, we will focus on the research of multiple communities influence.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (61103069), Key Lab of Information Network Security, Ministry of Public Security (C14602), Innovation Program of Shanghai Municipal Education Commission (13YZ052), and the Program of Shanghai Normal University (DCL201302).

REFERENCES

- [1]Zhao Fazhen.Influence of Online Communities Based on Link Analysis.J.Journal of Modern Information,2013,33(6):91-95.
- [2]Wei Chen, Yajun Wang, Siyu Yang.Efficient Influence Maximization in Social Networks.J.ACM,KDD'09,June 28-July 1,2009.
- [3]Gibbs, Susannah E, Le Linh Cu, et al. Peer and community influences on the acceptance of premarital sex among Vietnamese adolescents. J. Journal of Paediatrics & Child Health, 2014, 50(6): 438-443.
- [4]Wu Xindong, Li Yi, Li Lei. Influence Analysis of Online Social Networks, J. Chinese Journal of Computers, 2014, 37(4): 735-752.
- [5]Wu Liang, Huang Weijing,etc.ALT-LDA: A Probabilistic Model of Topic, Community and User Influence.J.Journal of Frontiers of Computer Science and Technology,2013,7(8):719-728.
- [6]Yang Jing, Xin Yu, et al.Semantics Social Network Community Detection Algorithm Based on Topic Comprehensive Factor Analysis.J.Journal of Computer Research and Development,2014,51(1):559-569.
- [7]Cha M, Haddadi H, Benevenuto F, etc.Measuring user influence in Twitter:the million follower fallacy.C.Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM 10), 2010: 10-17.
- [8]Page L, Brin S, Motwani R, et al. The PageRank citation ranking: bringing order to the Web.R. Stanford InfoLab, 1999.
- [9]Zhou Wenan, Fen Runjun, et al. Heterogeneous/fusion network QoS management and control technology.M. Beijing:Publishing House of Electronics Industry, 2009 : 2-3.
- [10]SKEHILL R, BARRY M, KENT W, et al. The common RRM approach to admission control for converged heterogeneous wireless networks.J.IEEE Wireless Communications, 2007, 14(2) : 48 — 56.
- [11]Guo Jingshi, Tang Hongbo, et al.Influence optimization model based on community structure.J.Journal of Computer Applications,2013,33(9):2436-2439.