Establishing Ties with Bounded Capacity and Limited Network Access

Qian Chen, Hongyi Su, Jiamou Liu, Hong Zheng and Bo Yan

December 19, 2018

Establishing Ties with Bounded Capacity and Limited Network Access

Qian Chen per 1st Affiliation *(Author)*
Computer Science&Technology
School of Computer Science and Technology,BIT
Beijing, China
18801156596@163.com

Hongyi Su per 2nd Affiliation *(Author)*
Computer Science&Technology
School of Computer Science and Technology,BIT
Beijing, China
henrysu@bit.edu.cn

Jiamou Liu per 3rd Affiliation *(Author)*
Computer Science
University of Auckland
Auckland, New Zealand
jiamou.liu@auckland.ac.nz

Hong Zheng per 4th Affiliation *(Author)*
Computer Science&Technology
School of Computer Science and Technology,BIT
Beijing, China
hongzheng@bit.edu.cn

Bo Yan per 5th Affiliation *(Author)*
Computer Science
University of Auckland
Auckland, New Zealand
Yanbo@bit.edu.cn

***Abstract*— This paper explores how individuals build relationships in a social network from the perspective of computer science, mathematics and sociology. Interpersonal relationships in social networks can boost up information dissemination and enhance individuals' dominant position in the network. The following question is put forward in this paper: how do individuals build relationships with members of social networks in order to improve their self-importance? Therefore, three effective algorithms are raised to solve this problem from the perspective of network operators. Network operators can selectively provide users with different rights, which enables users to see the local information, global information and partial community information in the network, and develops corresponding social strategies from the perspective of the users' The paper also proposes a connection-based and time-based restriction model and compares the performance of the three strategies based on it. It then examines the different performance of these three strategies on network structure attributes, such as embeddedness and clustering coefficient. What is more, for the community strategy, the impact of different scales of community on the community strategy to improve users' centricity and embeddedness is also discussed. Experiments were conducted on simulated random networks and real dynamic network datasets. Finally, this paper compares the three strategies and makes detailed analysis.**

Keywords-component; social network, community, network operators, authority

I. INTRODUCTION

By taking a more central place in a social network, individuals in it can gain greater control over the flow of information in the network, earning more social status by obtaining access to more knowledge and skills. Based on the rules above, individuals can improve their social status by building relationships with other members of the social network[7]. According to Six Degrees of Separation[5] Theory, an individual can establish a relationship with a stranger through six people, which means users can achieve his or her purpose of making friends in the limited time and cost, while the Dunbar Number[2] reveals that people's ability to make friends in the social circle is limited. The number of people who communicate with each other is accurately 20 people, is considered to be the basis of many human resource management as well as sns. Due to the limited ability of users to make friends, the scale of the social circle which the user is in also affects the efficiency of the improvement of users' self-importance.

Analyzed from the perspective of the users', they can adopt local strategy, greedy strategy and community strategy. In the local strategy, the user can only observe part of the network conditions. However, in the global strategy, the user can observe the entire network. Community strategy is also put forward. Users can know "friends in their own communities", and to a certain extent, it expands the users' perspective and protects the users' privacy.

The problems of local networks and local community networks are proposed in this paper. Chapter Two describes a local network formed by the combination of the target node V and other neighbor nodes. In Chapter Three, the local strategy, global strategy and community strategy are defined for the global network, local network and local community network, and each strategy in different network structure model is tested. Chapter Four illustrates the performance of different strategies under various evolutionary models of social networks and real networks by comparison. Community strategies can help to increase the self-importance of users in a faster way with limited time cost and connection cost, and enable users to build strong, trustworthy relationships in social networks[11]. What is more, in order to find the certain scale of community that can faster increase the self-importance of users in the network, the influences of community strategy on embeddedness and centrality under different community scales are analyzed. Chapter Five proposes the analysis and outlook of the paper.

This research has the following meanings:

The new social model places more emphasis on protecting users' privacy, while traditional social models allow users to gain more information and develop a new social strategy that both protects the privacy of users to a certain extent and allows users to improve their self-importance in the network in less time.

The existing social network friend recommendation strategies a pan-social model that discloses user information, and provide theoretical support for the recommendation of the friend network for the social network.

II. RELATED WORK

The issues of privacy risks in social networks were analyzed in Risks of Friendships on Social Networks[1]. In terms of the privacy issues existing in social networks, Primates: A Privacy Management System for Social Networks[7]proposes a social network privacy management system which can allow users to specify access control rules for their resources and implement access control for all shared resources. From the perspective of the user, access rights are set for other users, and the access of other users to the user is limited and controlled, which protects the privacy of the user to some extent. From the perspective of the network operators, it may be considered to provide users with different permission options: the user may choose to set different permissions including all public and partial public to release information to other users. Different permissions have different scope of vision for other users. All public information means that the user can observe all the nodes in the network, while partial public information indicates that the user can only know the local information in the network.

For the problem of community structure, Composing Activity Groups in Social Networks [4] proposes to establish high-quality activity groups based on users' information in social networks to achieve good communication among members of the group. The concept of "group" is "a small community" to a certain degree. It also proposes that the group formation process in current social networking services is tedious and may include inappropriate team members or missing relevant team members. The quality of the community will influence the communication and exchange among the members of the community to a certain extent[10].

III. REALATED THEORIES AND DEFINITIONS

Following standard convention, we view a social network as a graph $G(V,E)$, where V: a set of nodes, E: a set of undirected edges on V of the form uv where $u \neq v \in V$, U: a set of all the communities, $C(v)$: a set of nodes in the community in which V is located and $S(v)$: a set of nodes in the community of v which is connected to other communities. A path (with the length of k) is a sequence of nodes $u_0, u_1, \dots\dots u_i$ where $u_i,\ u_{i+1} \in E$ for any $0 \le i \le k$. The (geodesic) distance between u and v, denoted by $dist_G(uv)$, is the length of a shortest path between u and v.

Definition 1: In the local strategy, v will establish its own local network G_l, and G_l, is composed of v and node sequence $D(v)$ and $D(D(v))$.

Definition 2: In the community strategy, v will establish its own local community network G_c, and G_c is composed of v and node sequence $S(v)$ and $C(v)$, which means that the set of points that are connected by v and all the other communities in the community.

In order to ensure that the target node can get the maximum benefit and rapidly increase its importance in the network, it should choose to be able to establish or release the relationship between the node that maximizes its own revenue when facing the situation of establishing or disassociating the relationship of the selected node:

Definition 3 : v is a node in G, and the establishment or dissolution of relationship of the revenue is defined as follows:

When selecting a node to establish a relationship, the target node traverses all the nodes in the network to which no relationship is established, and calculates the raising value of closeness centrality after establishing the relationship with each node, and selects the node with the highest centrality. It can be calculated through $max(Changecls(v))$, and establish a relationship with it.

$$C_{g_{cls}}(v) = \max(C_{cls_{before}} - C_{cls_{after}}) \qquad (1)$$

When the relationship of the selected node is disassociated, the target node traverses all the neighboring nodes. The change amount of closeness centrality should be calculated. The change amount includes the raising value and the decreasing value of closeness centrality. For the convenience of calculation, if the approaching centrality value is increased, the rising value is taken as the opposite and the node with the lowest absolute value which can make itself close to the change of the centering degree is selected, that is, the node whose height is increased or decreased by a minimum is approached. It can be calculated through $min(abs(Changecls(v)))$and disassociate the relationship.

The two restrictions in this paper are the restriction of the number of edges and restriction of the number of time steps. Dunbar Number reveals that people's ability to make friends in the social circle is limited, so the number of edges of v set by the restriction is limited and a person's ability to make friends is limited. According to Six Degrees of Separation Theory, any two individuals can establish a relationship with others through six people. Therefore, anyone can make friends with a certain number of people within the limited number of time steps. Therefore, the number of time steps to limit the number of new edges to be set is limited. Detailed definition is as follows:

Definition 4: v is a node in G, the restriction of the number of edges and steps are defined as follows:

1. e is the restriction of the number of edges..

$e = ad(G)$ the average degree value of the network

2. s is the restriction of the number of steps.

$s = ap(G)$ the average shortest path of the network

A user can belong to one community or multiple communities, so there will be community overlap[9], and users in overlapping community segments also belong to two or more communities at the same time.

Louvain algorithm[3] is a community detection algorithm based on modularity., which performs well both in efficiency and effectiveness, and can discover hierarchical community structures. The community division algorithm is defined as follows:

Definition 5: For any connected graph $G(V,E)$, the nodes in the network are continuously traversed, trying to add a single node to the community that can maximize the modularity until all nodes are no longer changed. Small communities are merged into a super node to reconstruct the network, then the weight of the edge is the sum of the edge weights of all the original nodes in the two nodes. It will continue iterating until the algorithm is stable. The definition of Modularity is:

$$Q = \left(k_{i,in} - \frac{\sum_{tot} k_i}{m}\right) \tag{2}$$

"m" represents the number of edges in the network; "$k_{i,in}$" represents the sum of the weights of the edges within the Community c; and "$\sum_{tot} k_i$" represents the sum of the weights connected to the nodes in the community c

For any connected graph $G(V,E)$ and node $v \in V$, the closeness centrality of V is defined as follows:

$$C_{cls}(v) = \frac{|V|-1}{\sum_{u\in V\setminus v} dist(u,v)} \tag{3}$$

A higher value of C_{cls} means that v is closer to the average distance of other nodes, which occupies a more central position in the network.

Now, a complete definition of social circle strategy can be given: in a given connected graph G and a target node v, we can use different strategies of adding and deleting edges to let v obtain higher cls value within limited edges and steps.

IV. GLOBAL STRATEGY LOCAL STRATEDY AND COMMUNITY STRATEGY

A. Local strategy(lg)

In local strategy, the user's network is local network G_l and the user can only notice the neighbor node and the neighbor node of the neighbor node. So v is always connected with his friend and two nodes without the edge of "friend of friends" node. Each time the target node runs with a strategy, it will establish a relationship with one node and disassociate the relationship with one of the neighboring nodes according to the limitation of the number of sides in the restriction model. Because the target node's neighbor relationship has changed, G_l will also be updated. Local strategy

B. Global strategy(gg)

Unlike the local strategy, the user can notice the layout of the entire network in the global strategy. For example, Weibo will recommend people follow the person with higher "follows". In the global strategy, users will traverse all nodes in the network, and select the node with the highest revenue as the output node and establish or disassociate the relationship with it according to the definition of 3. Global greedy strategy

C. Community strategy(lc)

In social networks, a group can be seen as a small community and the people in the small community will also have contact with users in other communities. This part of the connection with the outside is in a pivotal position, which can connect several small communities, so it is necessary to connect this part of the node. Each time the target node runs with a strategy, it will establish a relationship with one node and disassociate the relationship with one of the neighboring nodes according to the limitation of the number of sides in the restriction model. Because the target node's neighbor relationship has changed, G_c will also be updated. Table 3 describes an algorithm for community to output a node under a time step.

V. EXPERIMENTAL RESULTS

A. Network Model

Community structure network model can be divided into strong community structure network model and weak community structure network model. Four well-known community structure network models (WS model, NW model, Random Planted Partition model and LFR model) are discussed in this paper. Each model's network is generated by its generation mechanism.

B. Experiment purposes and node selection

100 nodes were randomly selected for experiments. In order to avoid the influence of nodes with maximum and minimum centrality on the experimental results, the centrality is ranked when the nodes are randomly selected for experiments. The average closeness centrality is calculated, and 50 nodes are selected randomly in the nodes larger than the average closeness centrality value, and the rest nodes are selected in the nodes which are smaller than the average closeness centrality value.

C. Experiment Results

In order to know the average values of the change of centrality, time consumption and embeddedness, three strategies were tested 100 times in different scales of networks (100, 200, 500, 1000) in three models respectively of the first type of experiments, and the average values of the change of centrality of the target nodes were obtained at the connection cost and time cost.

The higher the centrality is, the greater the importance of the user in the network is. Fig. 1 is the experimental results of the changes of centrality.

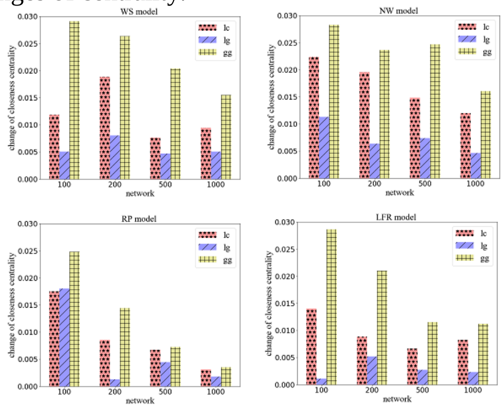

Figure 1. the average value of change of closeness centrality in four models with different size of networks (WS model NW model RP model LFR model)

As can be seen from the results of the experiments, The global strategy does good to the improvement of the closeness centrality because the target node can know the distribution of the nodes in the whole network and according to the revenue model. The target node chooses the highest center of their own highest node to establish or disassociate relations to get the maximum benefit. The community strategy also has a good performance, the target node can establish relations of different nodes and users will be more conducive to make more friends in the community. Correspondingly, the local strategy can only know "friends of friends", while the community strategy can know "friends in the community" and get more information, so the community strategy is better than the local strategy.

A better strategy can not only affect the target nodes in the change of centrality and other indicators, but also be more rapid enough to improve the importance of the target node. Fig. 2 is the experimental results of different time consumption.

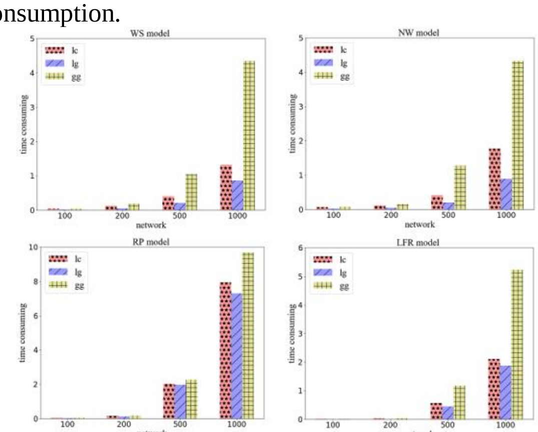

Figure 2. The average value of time consuming in four models with different size of networks(WS model NW model RP model LFRmodel)

The following results can be obtained from the results shown in Figure 2: The average time of running a global strategy is the longest. This is due to the view that the target node is the entire network, so it will be longer to traverse all nodes. The community strategy takes more time than the local strategy. This is because in the community strategy, the target node is "friends in the community", while in the local strategy, the target node is "friends of friends", which narrows down the part of the scope. And in the community strategy, it is necessary to run the community algorithm which increases the time consumption.

In establishing social relations, trust, the strength of relationships and personal character are other important dimensions of analysis other than centrality. They are all mainly affected by embeddedness. In social networks[12], embeddedness refers to the extent to which an individual is affected by social relations, which is usually used to measure the degree of trust of an individual. The embeddedness of an edge between "x" and "y" under a node is defined as the number of common neighbor nodes $D(x) \cap D(y)$. This paper redefines the embeddedness used in this experiment as follow

$$embed(v)\ \frac{\sum_{uv\in E}|\{w\in V \mid wu,wv\in E\}|}{(|V|-1)(|V|-2)} \tag{5}$$

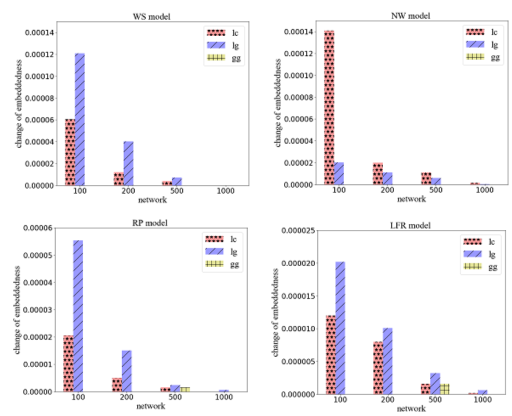

Figure 3. the average value of change of embeddedness in four models with different size of networks(WS model NW model RPmodel LFRmodel)

Fig. 3 is the experimental results of the changes of embeddedness

As can be seen from the results of the experiments, For the small world model, the local strategy and community strategy are the most significant for the improvement of embeddedness, among which, the local strategy is much better. There is a high degree of trust between nodes, because the target node is "friends of friends" and "friends in the community" in the local strategy and community strategy. As with real-world social networks, users are generally more trustworthy about relationships between acquaintances and easier to build relationships.

For the community model, the local strategy is significant for the improvement of embeddedness. Because the local greedy strategy only considers the neighbor nodes and the neighbor nodes of these nodes, which are all closely related nodes. The community strategy is also superior to the global strategy in improving the embeddedness.

D. Real network experiment

The experiment is also conducted from the real data sets of BlogCatalog (4924 nodes, 120776edges), Douban (5996 nodes, 18796 edges), Youtube (4999 nodes, 45958 edges), Facebook (3927 nodes, 84210 edges) and email-Eu-core (1005 nodes, 25571 edges).The email-Eu data set is a data set with real community distribution. Different from the other four data sets, the community dividing algorithm needs to be used for community division. The email-Eu data set uses the community division of its own data set: n order to know the details, three strategies are tested 100 times in the real data sets. Fig. 6 is the experimental results.

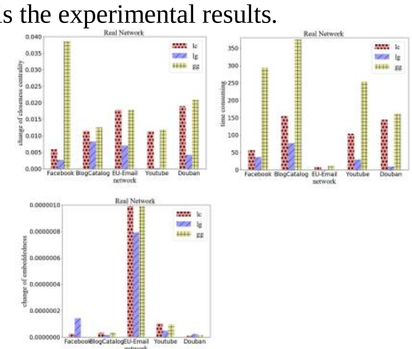

Figure 4. the average value of change of closeness centrality/time consumin/change of embeddedness in five real datasets(Facebook Youtube Douban BlogCatalog Email-eu)

As can be seen from the results of the experiment, In terms of the improvement of closeness centrality, the global strategy is the best, and the community strategy and the local strategy share nearly the same results.

In terms of time consumption, obviously, global strategy takes longer than local strategy and community strategy. And the local greedy strategy takes less time than the local community strategy, because the community strategy needs to consider the issue of community division.

In the aspect of improving embeddedness, community strategy performs the best, which can effectively improve the embeddedness of nodes in the network. However, the global strategy can be ignored.

VI. CONCLUSION AND FUTURE WORK

This paper analyzes the strategies in the social circle based on the community structure network and proposes local strategy, global strategy and community strategy, from the perspective of network operators who provide users with certain rights so that users can choose to observe different scopes of network structures. The paper also discusses the performance of different strategies under the limited connection cost and time cost. In the experiment, according to the community's strong and weak structure, four different network models of community structure are selected to test the simulated data set, and three strategies are tested under the social network of real data sets. Other future research directions include:

1. In the real world network, most of the network is not static and immutable. So the users also need to consider the evolution of the network in the process of change when developing a social circle strategy. The dynamic evolution of networks can be applied in the future work.

2. In the real world network, other users will change with the change of the network. In the future work, we can consider selecting nodes in the network, taking the same strategy as the target node changing, and observing the impact on the target node after adding other change of the nodes.

3. Nowadays, social model is mostly pan social model. The user's privacy cannot be well protected, and community mode of social protection can be a certain extent to protect the privacy of users, and will not affect the user to expand social circles and so on. Therefore, social strategies can be applied to the recommendation of friends on social networking sites.

REFERENCES

[1] C. G. Akcora, B. Carminati, and E. Ferrari. 2013. "Risks of Friendships on Social Networks. In IEEE International Conference on Data Mining". 810–815.

[2] P. Mac Carron, K. Kaski, and R. Dunbar. 2016. "Calling Dunbar's numbers". Social Networks 47 (2016), 151–155.

[3] Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. 2011. "Generalized Louvain Method for Community Detection in Large Networks". 79, 6 (2011), 88–93.

[4] Imen Ben Dhia, Talel Abdessalem, and Mauro Sozio. 2012. "Primates: a privacy management system for social networks". In ACM International Conference on Information and Knowledge Management. 2746–2748.

[5] Hakan Kardes, Abdullah Sevincer, Mehmet Hadi Gunes, and Murat Yuksel. 2013. "Six Degrees of Separation among US Researchers". In Ieee/acm International Conference on Advances in Social Networks Analysis and Mining. 654–659.

[6] Baozhen Lee, Weiguo Fan, Anna C. Squicciarini, Shilun Ge, and Yun Huang. 2014. "The relativity of privacy preservation based on social tagging". Information Sciences An International Journal 288, C (2014), 87–107.

[7] Loet Leydesdorff, Thomas Schank, Andrea Scharnhorst, and Wouter De Nooy. 2008. "Animating the development of Social Networks over time using a dynamic extension of multidimensional scaling. Physics 17, 6 (2008), 611–626."

[8] B Uzzi and S Dunlap. 2005. "How to build your network". Harvard Business Review 83, 12 (2005), 53–60.

[9] Bingying Xu,Lei Deng,Yan Jia,Bin Zhou,and Yi Han.2013. "Overlapping Community Detectionon Dynamic Social Network".In Sixth International Symposium on Computational Intelligence and Design.321–326

[10] Jianhua Ruan and Weixiong Zhang. 2006."Identification and evaluation of weak community structures in networks". (2006), 470–475.

[11] Song, Y., Xu, R.: "Affective ties that bind: Investigating the affordances of social networking sites for commemoration of traumatic events". Social Science Computer Review p. 0894439318770960 (2018)

[12] Burt, R.S.:"Structural holes versus network closure as social capital". In: Social capital, pp. 31–56. Routledge (2017)