Optimization Application Base on the Social Complex Network Algorithm

Hua ZHANG
Hunan Information College, Changsha Hunan 410151, CHINA

Abstract: As for the present community discovery and compression algorithm ignoring the issue of network community structures, this paper proposes a community discovery GS algorithm and community compression SNC algorithm. On the basis of proposing the theorems and corollaries related to the importance of the nodes in community discovered by topological method. GS algorithm discoveries the important nodes on different levels in community, and then through social networks compresses SNC algorithm and according the importance of the node compresses the community. Experimental results show that: the proposed algorithm can maintain the relationship between the communities during the compression process. It has a good community compression, in which the ratio can up to 0.95, and at the same time it can retain the important nodes in the community or community basic structures.

Keywords: Figure Compression; Community Discovery; Routing Selection; Node

1. Introduction

Online social networking communities are some unique teams consisted by very closely linked units or individuals, which usually have the common interests and share the same theme. Community represents of the social activities of user in the network; the in-depth study of the community can learn the knowledge of online social networking information and its organizational structure; it can help network service providers effectively organize portals and product manufacturers to accurately find their interest groups.

In the real world, there are many network models, which range between complete rule and complete random called as complex networks, such as national transport networks, ecological networks, scientists cooperation network and social relations networks. Complex network has the properties as with small world and scale-free, but also it has a community structure characteristic. Links between nodes within the community is relatively close, while the links between communities is relatively sparse [1-4]. Discovery of community structures in network is helpful for understanding and applying the complex network. For example, in social networks, community structure can provide friends referral services for community members, and it can also be used for social network analysis, network culture security early warning and so on. Thus, community discovery in social network has very important practical value. Online social networking community is divided into two kinds: artificial community and potential community [5-7]. Human community is found and maintained by artificial, however, it is difficult to manage these artificial communities through artificial means; the number of potential communities is greatly larger than the artificial communities, and it is still growing [8]. Therefore, it is necessary to study a community automatic discovery technology for those potential, or is about to be the community,

With the advent of the era of social networks, it is the necessary requirements made by the times conducting the in-depth research on the social networks. The researches on visualization and knowledge discovery for social network will be involved to the social network compression [9-12]. With the increasing of social network scale, the community discovery has become an indispensable step in the process of social networking application. As an important structural feature of social networks, community retains the important nods or basic structure and maintains the relationship between them in the compression process, which has an important and meaningful value. However, from the point of view of existing figure compression method, the research taking community as compress object is still very rare.

Figure compression is widely used in semantic label network, important node discovery, network retrieval, network visualization, network analysis and other fields. In recent years, some relative typical figure compression method is appeared. The current figure compression methods can be divided into no right figure compression and right figure compression; the used compression method is generally as the combined similarity nodes. For example, when node A and node B has the same or similar common neighboring nodes, they can be merged and generated the so-called super-node; the edge merges between super nodes is called super edge. Thus, this method will produce some edges between the super nodes which does not exist in the original network, and then cause errors when decompressed. Therefore, these methods damage the compression method. Besides, there are other shortcomings of these methods, such as the need to pay a higher time cost to balance the similarity between nodes and prior knowledge of merger thresholds set required for more parameters to meet different situations.

For these problems, the thesis proposes the social network compression method based on the importance of community nodes. This approach consists of a community discovery algorithm (GS) based on greedy strategy and social network compression algorithm (SNC); social network compression, in essence, is a figure compression. However, in order to distinguish it from other networks approach, which do not consider the premise of community, the approach is called as the social network compression method. This method first uses the social network topology theory to research the community discovery of social network, and on this basis to find the importance of distinguishing community nodes, and then the important compression based on community nodes are conducted.

2. Network Community

2.1. Network Community Nodes

The topology potential theory is derived from the nuclear physics, which is used to guide the community discovery on social networks. In the nuclear physics, the non-contact interaction nuclear between the nucleons is characterized through nuclear field. Topology potential theory draws on the field theory, which considers that there is interaction of direct or indirect relationship between nodes mutual. Theorem 1 Let node u and node v locating in an attractive chain of the community representatives point v^* in social network, and u is the a_{-th} jump of v^* ; v is located in $a+1_{-th}$ jump of v^* ; $a=0,1,2...,h-1$, so the contribution ratio of the topology potential of u and v to v^* is as follows:

$$A_{v^*\leftarrow p}(\sigma_{opt},l)=\frac{1}{n}e^{-(\frac{l}{\sigma_{opt}})^2} \tag{1}$$

l is minimal hops of p leaving v^*.

From the formula (2), it is easy to know that the contribution ratio of the topology potential of u and v to v^* is:

$$A_{v^*\leftarrow u}(\sigma_{opt},a)=\frac{1}{n}e^{-(\frac{a}{\sigma_{opt}})^2} \tag{2}$$

Therefore, the contribution ratio of the two is:

$$R_{u\leftarrow v}(a,a+1)=\frac{A_{v^*\leftarrow u}(\sigma_{opt},a)}{A_{v^*\leftarrow v}(\sigma_{opt},a+1)}=e^{\frac{2a+1}{\sigma^2_{opt}}} \tag{3}$$

Corollary 1 Let node u and node v locating in an attractive chain of the community representatives point v^* in social network, and u is the a_{-th} jump of v^*; v is located in $a+1_{-th}$ jump of v^*; $a=0,1,2...,h-1$, so the contribution ratio of the topology potential of u and v to v^* is $R_u \leftarrow v(a,a+1)\succ 1$.

Proof From the theorem 1, $R_{u\leftarrow v}=e^{\frac{2a+1}{\sigma^2_{opt}}}$; $a=0,1,2...,h-1$, $\sigma opt\succ 0$, so $2a+1\succ 0$.

$\sigma^2_{opt}>0$, $\frac{2a+1}{\sigma^2_{opt}}>0$, $R_{u\leftarrow v(a,a+1)}=e^{\frac{2a+1}{\sigma^2_{opt}}}>1$

Corollary 2 Let node u, node v and node y are in an attractive chain of the social network community representatives point v^*; y is the a_{-th} jump of v^*; v is $a+1_{-th}$ jump of v^*; m is $a+2_{-th}$ jump of v^*; $a=0,1,2...,h-1$, so:

$$R_{v\leftarrow w}(a+1,a+2)>R_{u\leftarrow v}(a,a+1) \tag{4}$$

Proof Because $\sigma^2_{opt}\succ 0$ and a is the non-negative integer, for a given network σ'_{opt} is a constant value, so $\frac{2(a+1)+1}{\sigma^2_{opt}}>\frac{2a+1}{\sigma^2_{opt}}$. From the Theorem, it can be known that $R_{u\leftarrow v(a,a+1)}=e^{\frac{2a+1}{\sigma^2_{opt}}}$, $R_{u\leftarrow v(a+1,a+2)}=e^{\frac{2a+1}{\sigma^2_{opt}}}$; and ex (x>0) is a strictly monotonic increasing function, therefore $R_{u\leftarrow v(a+1,a+2)}>R_{u\leftarrow v(a,a+1)}$.

Corollary 3 Let node u, node v and node w are in an attractive chain of the social network community representatives point v^*; u is the a_{-th} jump of v^*; v is $a+1_{-th}$ jump of v^*; w is $a+2_{-th}$ jump of v^*; $a=0,1,2...,h-1$, so:

$$R_{v\leftarrow w}(a+1,a+2)=e^{\frac{2}{\sigma^2_{opt}}}R_{u\leftarrow v}(a,a+1) \tag{5}$$

Proof From Theorem 1 $R_{u\leftarrow v(a,a+1)}=e^{\frac{2a+1}{\sigma^2_{opt}}}$,

$R_{u\leftarrow v(a+1,a+2)}=e^{\frac{2a+1}{\sigma^2_{opt}}}$,so

$R_{u\leftarrow v(a+1,a+2)}/R_{u\leftarrow v(a,a+1)}=e^{\frac{2}{\sigma^2_{opt}}}R_{u\leftarrow v(a,a+1)}$, that is:

$$R_{v\leftarrow w}(a+2,a+1)=e^{\frac{2}{\sigma^2_{opt}}}R_{u\leftarrow v}(a,a+1) \tag{6}$$

Table 1 lists the contribution ratio of a number of nodes in network which is distant away 1 hop to the representative point; the above theorems and corollaries can be used to verify the correctness. Theorem 1 and its corollaries fully demonstrate that in the community discovered by topology potential method compared to distant neighbor nodes, the neighbor nodes have larger contribution to the topology potential of the representatives; with the increasing distance from the representative points, the contribution of the nodes is exponentially decreased.

Table 1. Ru ←v (A,A+1)VALUES OF NUMBER OF THE NETWORKS

The hopes of node u when leaving representative v^*	The hopes of node v when leaving representative v^*	Scene People Network	Karate Club Network	Dolphin social networks
1	2	19.789	16.873	15.852
2	3	146.432	138.421	98.063
3	4	165432	820.863	620.931

The above conclusions are exactly the same with intuitive feel of people for the community. In addition, some other aspects can also be used to illustrate the importance of community representatives' neighbor nodes is far more important than the importance of neighbors. For example, the study found that robustness and fragile is one of the basic characteristics of complex systems and complex networks, while the other important mean to trigger the vulnerability of complex networks is consciously attacking the nodes with high number.

2.2. Network Community Structures

Let the undirected network $G(V,E)$; the set of network nodes as V ; the set of network edge is as $E=\{E=(u,v)|u\in v,v\in V\}$; G is presented by the matrix A with a size as $|V|\times|A|$; if the edge $e=(i,j)\in E$, $A_{ij}=1$, otherwise, $A_{ij}=0$. Internet community structure is an m division scheme $p=(v_1,v_2,...v_m)$ of the set V of network nodes, in which V_i should meet four conditions:

$V_i\neq\varnothing(i=1,2,\cdots,m)$;

$\bigcup_{i=1}^{m}V_i=V$;

$V_i\cap V_j=\varnothing(i\neq j)$;

2.3. Evaluation Index of Network Community

1) Modularity function R

Several definitions of community structures can not be directly applied to probe the structure of complex networks in the community and can not evaluate the quality of community structures obtained by algorithm. Therefore, Girvan and Newman proposed the modularity R function, which quantitatively describes the advantage and disadvantage of the division of community structure.

2) Community level C

This thesis makes improvements in the defects of modularity function R and proposes the concept of community level. Community level index is defined as:

$$C=\frac{1}{m}\sum_{i=1}^{m}\frac{C_{in}}{n_c-(n_c-1)/2}-\frac{C_{oat}}{n_c(n-n_c)}$$

3. Algorithm Description

The proposed method is conducted by the way of compression of relative representatives from the outside to the inside; it can be compressed only leaving representative points in the network at maximum. One of the advantages of the method is embodied as follows: during the compression process, it can compress out some relatively unimportant node, which effectively reduce network size, but also keep the necessary important nodes in the community or the basic structure of the community. Different from other general figure compression method, this method need not using the experience parameters specified by the user in the compression process, but only need to assign the hops should be compressed under the guidance of the optimization range of influence determined by the automatically method. Figure 1 presents the compression sketch map of the community with the optimization range of influence as 2 hops, in which the one-way arrows represent that some hops can be compressed to another hop.

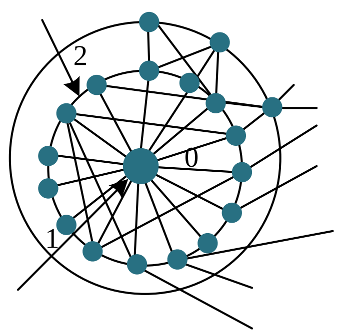


Figure 1. Compression sketch map of community

3.1. Data Structures

In order to achieve lossless compression, the proposed method in the community discovering process indicates the hops of all nodes between representative points, and these markers is marked in the list structure in the memory community. For there is possible that the relationships between communities may be loosed in the compression process, the chain table of relationships between memory communities is designed. The basic data structure used in the methods used is as follows:

```
// List structure in various communities
Typedef struct CommNode {
 int Node
 int hop
Struct CommNode*next node
} CommNode
Typedef struct {
```

int Re *pNode*

int *totalhop*

CommNode*FirstNode

} VexNode . Community Arr [maxSize]

// List structure in various communities

Typedef struct *Comm*Re*lation* {

int c_1

int c_2

Struct *Comm*Re*lation* *next

} *Comm*Re*lation*

3.2. SNC Algorithm

On the basis of the GS algorithm discovering the community, social network compression algorism (SNC) obtains the number of hops need to compressed through the interactively method, and then conducts the compression operation. The specific description of SN_c algorithm is as follows:

Input network $G=(v,e)$, $|v|=n$, $|E|=m$, *Opstigma*

Output compressed community C_i (i is the number of the community representatives, and for each community C_i it only indicates the nodes within a user-specified number of hops)

Disover Community()

$h=(\text{int})(3*Optsigma/sprt(2))$

Count<<" /*n* Optimization range of infuluence of the current network" *n* "

Count>>hop

For $i=0, i<\max Size, i++)$

P= $G.Comm[i]$

While (p) {

If $(p->hop<=hop)$ display $(p->Node)$

$p=p->next\ node$

}

}

$r=G.CommR->next$

while (r) {

display $(r->c_1, r->c_2)$

r=r->next

}

4. Experimental Simulation and Analysis

4.1. Experimental Environment and Setup

To verify the feasibility and effectiveness of this method, it is tested in two widely used sets of data: the karate club network and dolphin social networks. The node number in the two networks is the same with the number offered by Newman.

4.2. Analysis of Experimental Results

(1) Compression Test in Karate Club Network

Karate club network is drawn by Zachary according to the interaction of its members. The club eventually split into two respective groups which as the core of coaches and the core of management, shown in Figure 2.

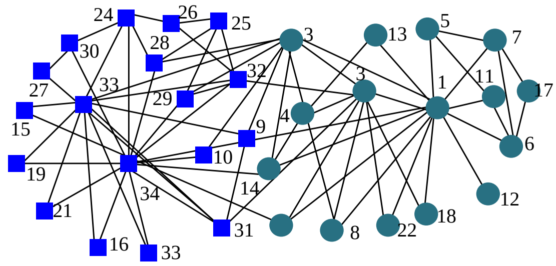


Figure 2. Karate club network

GS algorithm is applied on the karate club network to make community discovery, and the results are shown in Figure 3. In Figure 3, the circular and the square are used to mark two different communities; large icons are used to identify the community representatives; the triangle icons are used to identify communities overlap between nodes. The meaning of the icons IN Figures 4 ~ 6 are the same with Figure 3.

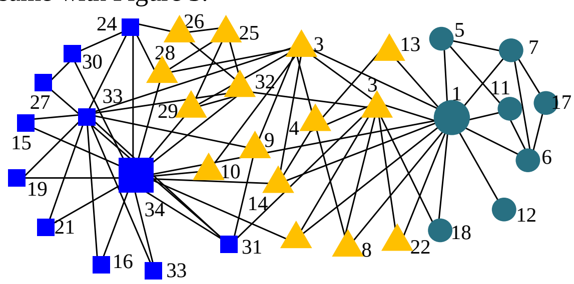
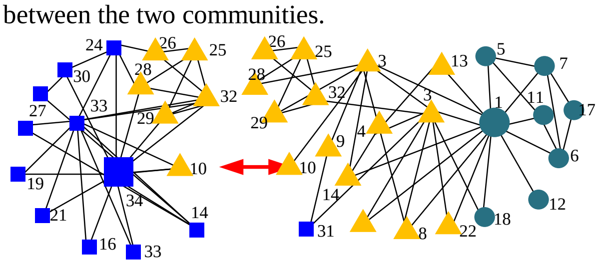


Figure 3. GS algorithm used in karate club network to discovery community

In the community found by GS algorithm, the SNC algorithm is applied to find the compression of the community with 2 hops, 1 hop skip and 0 hops, and the compression results are shown in Figure 4~6. The double-headed arrow in Figure 4~6 are used to identify the relationships between the two communities.



2 hops compression in karate club network

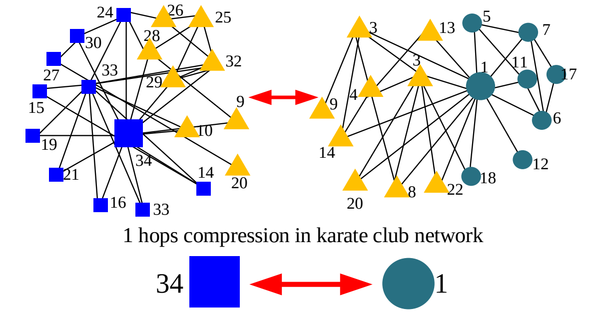


Figure 4. 0 hops compression in karate club network

(2) Compression Test on Dolphin Social Networks

In the community found by GS algorithm, the SNC algorithm is applied to find the compression of the community with 2 hops, 1 hop skip and 0 hops, and the compression results are shown in Figure 8~10. The double-headed arrow in Figure 8~10 are used to identify the relationships between the two communities.

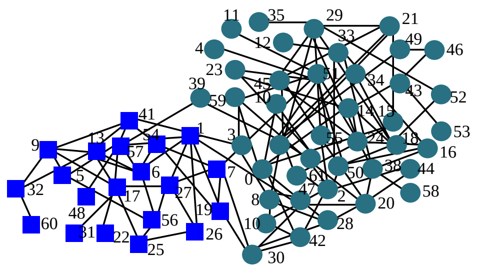


Figure 5. Dolphin social networks

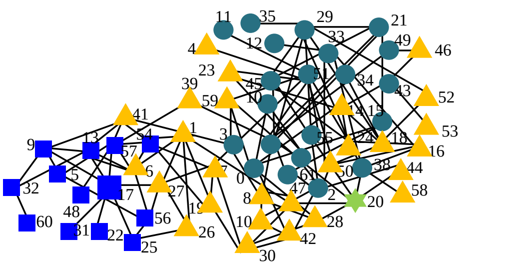


Figure 6. Communities discovered by GS algorithm in dolphin social network

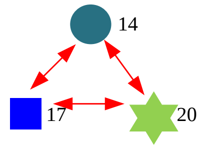


Figure 7. 2 hops compression Figure in dolphin social network

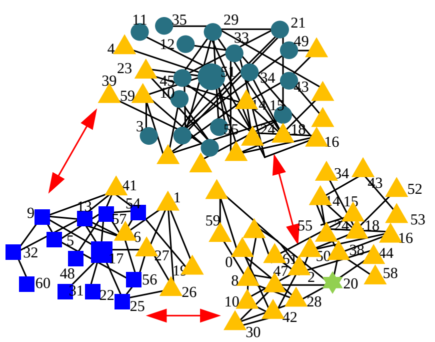


Figure 8. 0 hops compression Figure in dolphin social network

5. Conclusion

With the wide use of the figure compression methods and techniques in semantic label network, network retrieval and other fields, the related researches are growingly concerned. In order to solve the problems of figure compression methods like high time complexity, depending on prior knowledge to set the parameters, too many parameters need to be adjusted, compression beneath and ignoring network community structures, this paper presents a new compression method - the social network compression method based on the importance of community nodes. On the basis of proposing the theorems and corollaries related to the importance of community nodes discovered by the topology potential method, this method makes community discovery through the GS algorithm based on greedy strategy, excavates important nodes with different levels in community, through the social networks compresses SNC algorithm and based on the importance of nodes compresses the community. Through the experiments in classic data sets the feasibility and effectiveness of this method are verified. The experimental results show that this method can keep the relationships between communities in the compression process, has the ideal community compression ratio, in which the highest can be up to 0.95, and if necessary can reserve the important nodes in the community or community basic structures.

References

[1] Muhammad J. Mirza, Nadeem Anjum. Association of Moving Objects Across Visual Sensor Networks. Journal of Multimedia, Vol 7, No 1 (2012) pp. 2-8

[2] Aaltonen, K.A network perspective on response strategies to stakeholder pressures in global projects, In Miia Martinsuo edt Recipes for success in project-based management. Project Management Association Finland, 2008:29-59

[3] Andrews; K. M., Delahaye, B. L Influences on knowledge processes in organizational learning The psychosocial filter. Journal of Management Studies, 2000, (7):797-810.

[4] Cho H, Gay G, Davidson B, et al. Social networks, communication styles, and learning performance in a CSCL community. Computers & Education. 2007, 49(2): 309-329.

[5] ohan Bruneel, Pablo D'Este, Ammon Salter. Investigating the factors that diminish the barriers to university-industry collaboration. Research Policy, 2010(39); 858-868.

[6] Project Management Institute. Project Management Body of Knowledge, 2000,(8);876-889

[7] Stock, G.N., Tatikonda, M.V. A typology of project-level technology transfer processes. Journal of Operations Management, 2000, 18: 719-137.

[8] Toke Bjerregaard. Industry and academia in convergence: Micro-institutional dimensions of R&D collaboration. Technovation; 2010 (30):100-108.

[9] Xianyu, B., Yang, J.M. Evolutionary ultimatum game on complex networks under incomplete information. Physica A, 2010, 389: 1115-1123.

[10] Savage N. Twitter as medium and message. Communications of the ACM, March 2011, 54(3):18-20

[11] Chew C, Eysenbach G. Pandemics in the Age of Twitter-Content Analysis of Tweets during the 2009 H1N1 outbreak. PLoS ONE, November 2010, 5(11):1-13

[12] Pal A, Counts S. Identifying Topical Authorities in Microblogs. WSDM, New York, 2011:45-54