ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RESEARCH ON LOSSLESS NETWORK COMPRESSION OF RAILWAY DATA BASED ON TOPOLOGY POTENTIAL COMMUNITY

[1]WANG XUEMING, [2]LI XIAOCUN, [1]JIN JING

Department of Locomotive Vehicle, Jilin Railway Vocational and Technology College, Jilin China

School of Urban Rail Transportation, Soochow University, Suzhou China

ABSTRACT

Research of lossless network compression based on topology potential community is carried out. To meet the different needs, two approaches of lossless network compression are proposed in this research. One approach, judging importance of the nodes according to their roles playing in the community composition, quantifies the importance of every node in communities, and achieves lossless network compression through layers; another approach, judging importance of the nodes according to the distances from the community representative nodes to them, differentiates the nodes with different distances, and achieves lossless network compression through compression ratio. Comparative experiments show that the two approaches not only can achieve perfect compression ratio, and retain the relationship between the communities, but also can reserve the important nodes or basic community structures during the compression process according to the needs.

Keywords: *Network; Compression, Topology, Potential Community, Railway*

1. INTRODUCTION

Map Compression (Graph Compression), also known as graph simplifying (Graph Simplication) (Graph Summarization) or abstract, and can be widely applied to the semantic label network, important node discovery, network retrieval, visual network, network analysis and other fields. In recent years, some typical map compression methods have come out, such as [1] method. The current map compression algorithm can be divided into right map compression and the right map compression two categories, but also by the compression method is generally combined with similarity of nodes, such as A node and the node B with the same or similar common neighbor node merging, generating so-called super node (Super node), and the super node the boundary with super edge (Super edge). Thus, this method will produce some original super nodes in network does not exist in the side, resulting in decompression (Decompression) error. Therefore, these methods are lossy compression methods. In addition, these methods also have other deficiencies; such as a measure of similarity between nodes often need to pay higher cost of time, require a priori knowledge of set merge threshold, in order to meet the different situation need to set more parameters.

Along with the network social era, on the social network research is the inevitable requirement of times. The social network visualization, knowledge discovery and research should be related to social network compression. With the increase of the scale of social network, community discovery has become a social network application process, one of the indispensable important step [2]. The community, as a social network's important structural features in the compression process, is retained in its critical nodes or basic structure and maintains the relationship between them has important significance and value. However, from the existing picture compression method, to the community for the compression object research is still very rare.

In view of the above questions, based on the complex network of community discovery method based on network, this paper study on compression method, put forward two kinds of lossless network compression method. The first method called social network compression algorithm SNC (Social Network Compression), essence is still a picture compression method. But in order to distinguish it

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

from other network community based precondition method, this method is called social network compression method. The method will be the first to use topology potential theory social network community found and distinguished community the importance of nodes, and then based on the community of node importance level compression. Second methods of nondestructive social network compression method will still based on topological potential theory and the community discovery node importance were quantified, then according to the compression rate of the network compression.

Although more than two kinds of methods are presented for the social network, but due to social network also belong to the complex network (although the two do exist some differences, such as the social network node is generally referred to as actors (Actor), more emphasis on the joint initiative), and complex networks also exist in the community structure of [3], so the two methods are fully applicable to complex network compression.

2. SOCIAL NETWORK SNC LOSSLESS COMPRESSION METHOD

As the SNC method will be in topology potential community found on the basis of the importance of nodes, according to the community network compression, so this section of the first topology potential community discovery methods found in the community of node importance analysis.

2.1 Community node importance analysis.

From the community composition level based on the topology of potential theory method for community detection that community in the importance of nodes there are different. To illustrate the importance of nodes and community differences show that the judgment is given below, the following theorem and inference.

Theorem 1 let u, v is a network of nodes in the v^* community represents a attract chain, and u located in the v^* a jump, v located in the v^* $a+1$ jump, $a=0,1,2,...,h-1$, u、v, on the topology potential contribution ratio of $R_{u\leftarrow v}(a,a+1) = e^{\frac{2a+1}{\sigma_{opt}^2}}$ 。.

Proof of any one at the node p to attract chain community representative point v^* topology potential contribution to

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

The l p left v^* minimum hops.

Based on the formula, u、v, is on v^* topology potential contribution respectively

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

and

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

Therefore the two contribution ratio of quantity

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

Corollary 1 a node u、v in a network community representative point v^* a attract chain, and u located in the v^* a jump, v located in the v^* $a+1$ jump, $a=0,1,2,...,h-1$ u、v, v^* topology potential contribution ratio $R_{u\leftarrow v}(a,a+1)>1$.

Proved by Theorem 1 and known knowledge $R_{u\leftarrow v}=e^{\frac{2a+1}{\sigma_{opt}^2}}$, $a=0,1,2,...,h-1$, $\sigma_{opt}>0$, then $2a+1>0$, $\sigma_{opt}^2>0$, $\frac{2a+1}{\sigma_{opt}^2}>0$ and then there is

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

Corollary 2 a node u、v、w, a network of community representatives, at point v^* a attract chain, and u located in the v^* a jump, v located in the v^* $a+1$ jump, w located in the v^* $a+2$ jump, $a=0,1,2,...,h-2$, then $R_{v\leftarrow w}(a+1,a+2) > R_{u\leftarrow v}(a,a+1)$.

Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Prove that for $\sigma_{opt}^2 > 0$, a non-negative integer, and for a given network σ_{opt} to a certain value, so $\frac{2a+1}{\sigma_{opt}^2} > \frac{2a+1}{\sigma_{opt}^2}$. From theorem 1, $R_{u\leftarrow v}(a,a+1) = e^{\frac{2a+1}{\sigma_{opt}^2}}$, $R_{v\leftarrow w}(a+1,a+2) = e^{\frac{2(a+1)+1}{\sigma_{opt}^2}}$, and e^x （x>0）is a strictly monotone increasing function, so have $R_{v\leftarrow w}(a+1,a+2) > R_{u\leftarrow v}(a,a+1)$

Corollary 3 a node u、v、w, a network of community representatives, at point v^* a attract chain, and u located in the v^* a jump, v located in the v^* $a+1$ jump, w located in the v^* $a+2$ jump, $a = 0,1,2,...,h-1$, then $R_{v\leftarrow w}(a+1,a+2) = e^{\frac{2}{\sigma_{opt}^2}} R_{u\leftarrow v}(a,a+1)$.

Proved by Theorem 1 know $R_{u\leftarrow v}(a,a+1) = e^{\frac{2a+1}{\sigma_{opt}^2}}$, $R_{v\leftarrow w}(a+1,a+2) = e^{\frac{2(a+1)+1}{\sigma_{opt}^2}}$, So $R_{v\leftarrow w}(a+1,a+2) / R_{u\leftarrow v}(a,a+1) = e^{\frac{2(a+1)+1}{\sigma_{opt}^2}} / e^{\frac{2a+1}{\sigma_{opt}^2}} = e^{\frac{2}{\sigma_{opt}^2}}$, That is $R_{v\leftarrow w}(a+2,a+1) = e^{\frac{2}{\sigma_{opt}^2}} R_{u\leftarrow v}(a,a+1)$。

Corollary 4 a node u、v、x、y in a network of community representatives v^* a attract chain, and u located in the v^* a jump, v located in the v^* $a+1$ jump, x located in the v^* b jump, y located in the v^* $b+1$ Jump, $a,b = 0,1,2,...,h-1$ And $b > a$, There are $R_{x\leftarrow y}(b,b+1) = e^{\frac{2(b-a)}{\sigma_{opt}^2}} R_{u\leftarrow v}(a,a+1)$。

Proved by Theorem 1 know $R_{x\leftarrow y}(b,b+1) = e^{\frac{2b+1}{\sigma_{opt}^2}}$, $R_{u\leftarrow v}(a,a+1) = e^{\frac{2a+1}{\sigma_{opt}^2}}$, So $R_{x\leftarrow y}(b,b+1) / R_{u\leftarrow v}(a,a+1) = e^{\frac{2b+1}{\sigma_{opt}^2}} / e^{\frac{2a+1}{\sigma_{opt}^2}} = e^{\frac{2(b-a)}{\sigma_{opt}^2}}$, That is $R_{x\leftarrow y}(b,b+1) = e^{\frac{2(b-a)}{\sigma_{opt}^2}} R_{u\leftarrow v}(a,a+1)$。

Table 1 lists a number of network distance jump node to represent points contribution ratio, can be used to verify the correctness of the theorem and corollary. Table 1, theorem 1 and corollary of 1~4 shows, in HCD methods found in the community, neighboring nodes than the neighbor node on behalf of some topological potential contribution to greater; along with the representative point distance increase, node contribution exponential decline. Therefore, with the local extremum of the community representative point of its neighboring nodes were more in number, the connection between them is also more closely and formed a community core structure; neighbor nodes to represent the topological potential contribution is relatively small, number is also relatively less, they are the link between more sparse. To sum up, from the community composition level, representing the point nearest neighbor node than the neighbor node is more importance.

2.2 SNC method the basic idea.

From the preceding analysis can know, in the HCD method to find the community representative points of the neighbor node importance is the hop-by-hop reduced. Accordingly, the SNC method will use the first based on the topological potential theory of community discovery, then network compression step in the realization of network scale effectively reduce.

SNC method using relative representative points from the inside outward compression manner, most can be compressed into the network only representative points. The advantages of this method: one is the embodiment of compression in the process can not only compress some relatively important nodes to reduce the scale of the network, but also in the necessary to retain an important node in the community or community of the basic structure.

Different from the general picture compression method, the SNC method in the process of

Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

compression without user specified parameters, and only in accordance with the method for automatically determining optimal effect range of h under the guidance of specified to be compressed to jump number can be. Figure 1 shows a design range of 2 jump community compression diagram, wherein the one-way arrows represent a jump from a compressible to another jump.

Table 1 $R_{u \leftarrow v}(a, a+1)$ *Values Of Several Networks*

Node uaway from the representative point v*hops a	Nodev away from the representative point v*hops a+1	Karate club (σopt=1.0204)	Dolph in Society (σopt=1.1782)	Word adjacencies (σopt=1.0043)	Les miserables (σopt=1.0435)	Books about US politics (σopt=0.9803)
1	2	17.8365	8.6810	19.5772	15.7225	22.6869
2	3	121.7630	36.6679	142.2059	98.6741	181.8129
3	4	831.2309	154.8820	1.0330e+003	619.2763	1.4571e+003

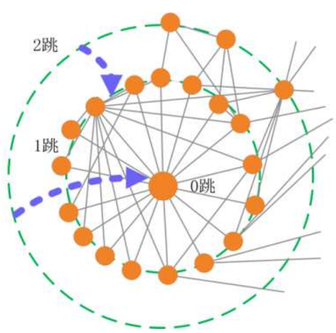

Fig. 1 Community Compression Schematic Diagram

The basic theory of SNC is the first topology potential community discovery method for community detection, and then compression. According to this idea, the proposed SNC method consists of two parts. The first part is intended to be used with the third chapter of the greedy strategy based on the overlapping community finding algorithm similar to GS algorithm for community detection. The algorithm considers not only overcome the community edge nodes and other community links in the node is cut apart artificially problems and overcoming method HCD overlapping node number is too small, and considerations for the subsequent network compressed to provide support for the. To avoid ambiguity, hereinafter referred to as the algorithm is NGS (New Greedy Strategy). The second part puts forward a kind of according to the importance of network node level compression algorithm, referred to as BL (Based on Layers).

2.3 SNC method of basic data structure.

In order to realize the lossless compression, the first SNC is adopted to design method of data structure. In addition to some of the basic data structure, data structure design must consider the following factors: the SNC method in a community discovery process is marked all the nodes representing the distance the hop count, while in the process of compression needed to save the community relations. Based on the above considerations, design the SNC method of data structure. The SNC method data structure is defined as follows:

```
/ / storage each community structure
    typedef struct CommNode{
        int    Node;
        int    hop;
        Struct CommNode    *nextnode;
    }CommNode;
    typedef struct {
        int RepNode;
        int totalhop;
        CommNode * FirstNode;
    }VexNode,CommunityArr [maxSize];
  / / define map types
    typedef struct{
      int adjMatrix[maxSize][maxSize];
        double potential[maxSize];
        CommNode *RepSet;
        int TagArr[maxSize];
        CommunityArr Comm;
        MulAttrNodeArr BoundNode;
```

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

```
        CommRelation *CommR;
    }Graph;
   / / storage community relation list type
    typedef struct CommRelation{
        int c1;
        int c2;
        struct CommRelation * next;
    }CommRelation;
```

/ / storage overlapping nodes linked list structure

```
    typedef struct RepNode{
        int RepNodeNum;
        double probability;
        struct RepNode *next;
    }RepNode;
    typedef struct MulAttrNode{
        int NodeNum;
        RepNode* FirAttrNode;
    }MulAttrNode,MulAttrNodeArr[maxSize];
```

3. SNC METHOD DESCRIPTION.

3.1Greedy Strategy Based On The Network Of Overlapping Community Finding Algorithm NGS.

NGS algorithm will be identified community representatives to greedy strategy along the attracting chain traversal is representative points to attract all nodes, is described as follows:

Method name: greedy strategy based on the network of overlapping community finding algorithm NGS

Algorithm input: NetworkG=(V,E) （|V|=n，|E|=m）

The algorithm output: CommunityCi（iAs a community representative point number）

The algorithm steps：

（1） InitComm(); InitTag(TagArr,-1); InitBoundNode();

（2）for(each node v in RepSet){

（3） i = GetRepNode(v); j = 0; InsertComm(i, i, j);

（4） InitQueue(Q1); InitQueue(Q2); EnQueue(&Q1, i); TagArr[i] = i;

（5） while(Q1 is not empty){

（6） j ++;

（7） while(Q1 is not empty){

（8） DeQueue(&Q1, &u);

（9） for(each u's neighbour w which potential less than or equal to u's potential)

（10） if(TagArr[w] != i){

（11） InsertComm(i, w, j);

（12） if(TagArr[w] == -1){TagArr[w] = i; EnQueue(&Q2, w);}

（13） else{ if(relation between community TagArr[w] and community i is

not in link list CommR) InsertRelation(CommR, TagArr[w], i);

（14） InsertRelation(CommR, TagArr[w], i);

（15） if(BoundNode[w] is empty) InsertBoundNode(w,TagArr[w]);

（16） InsertBoundNode(w, i);

（17） TagArr[w] = i; EnQueue(&Q2, w);

（18） }

（19） }

（20） }

（21） while(Q2.front != Q2.rear) {DeQueue(&Q2, &u); EnQueue(&Q1, u);}

（22） }

（23） G.Comm[i].totalhop = j;

（24）}

3.2 According To The Importance Of Node In A Hierarchical Network Compression Algorithm BL.

BL algorithm through the interactive way get to be compressed to jump number, then the compression operation. BL algorithm described in detail as follows:

Algorithm of node importance: according to the hierarchical network compression algorithm BL

Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Algorithm input: NetworkG=(V,E) （|V|=n，|E|=m ），Optimization of influencing factorsOptSigma

Output: compression algorithmCi (i community represented point number, each Cidisplays only user specified hop nodes inside)

The algorithm steps:

```
（1）DiscoverCommunity();
（2）h = (int)(3*OptSigma/sqrt(2));
（3）cout<<"\nThe current network optimization effect range："<<h;
（4）cout<<"\nPlease input to display the hop"<<"(<="<<h<<")：";
（5）cin>>hop;
（6）for(i = 0; i < maxSize; i ++){
（7）    p = G.Comm[i].FirstNode;
（8）    while(p) {
（9）        if(p->hop <= hop) display(p->Node);
（10）       p = p->nextnode;
（11）   }
（12）}
（13）r = G.CommR->next;
（14）while(r){
（15）    display(r->c1, r->c2);
（16）    r = r->next;
（17）}
```

SNC time complexity analysis.The SNC method mainly relates to a community found in the NGS algorithm and BL algorithm network compression. By comparison, the NGS algorithm time complexity is higher. Because the NGS algorithm in the worst case scenario is not more than O(n2) (n for the network node number), so the SNC method of time complexity is not more than O(n2).

4．SNC EXPERIMENT AND ANALYSIS

In order to verify the feasibility of the proposed method and effectiveness, through the experiment in the karate club network [3] and dolphin social network [4] two widely used data sets on the methods for testing.

4.1 Karate Club Network Compression Experiment.

Application of NGS algorithm in the karate club network of community discovery, results as shown in figure 2. In Figure 2 circular and square icon to indicate two different communities, large icons are used to indicate the community representative point, triangle icon to indicate the community overlaps between nodes. Figure5 and Figure 2: the same meaning of icon.

In the NGS algorithm to discover the community, BL algorithm is applied to find community were 2, 1 and 0 jump compression, compression results as shown in figure5. Figure5. Double Arrow to Indicate the Two Community Relations.

4.2 Dolphin Social Network Compression Experiment.

Application of NGS algorithm in dolphin social network community found, results as shown in figure 6. In Figure 5-6 circular, square and star icon to indicate three different communities, large icons are used to indicate the community representative point, triangle icon to indicate the community overlap between nodes. Figure 7~9 and Figure 6: the same meaning of icon.

In the NGS algorithm to discover the community, BL algorithm is applied to find community were 2, 1 and 0 jump compression, compression results as shown in figure 7~9. Figure 7~9 double arrow also used to mark the two community relations.

4.3 Experiments Analysis.

The experiments show that the classic data sets, using greedy strategy based on the overlapping community finding algorithm NGS and node importance hierarchy network compression algorithm BL, communities in the node number to get effective compression. Table 2 The Karate Club network and dolphin social network communities in 2, 1 and 0 jump compressed data rate (second columns in a community numbering and community representatives numbered aligned). The compression ratio is defined as follows

Definition 1 if the network is a $G^{'}=(V^{'},E^{'})$ network $G=(V,E)$ network compression, then $R=\dfrac{|G|-|G^{'}|}{|G|}$ is called network G compression rate

Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

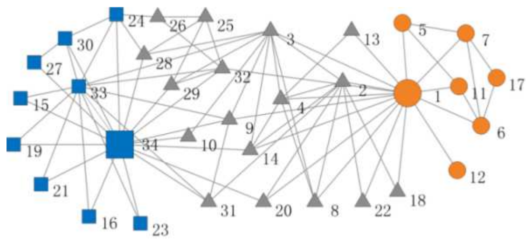

Fig. 2 Communities Discovered By GS Algorithm On Karate Club Network

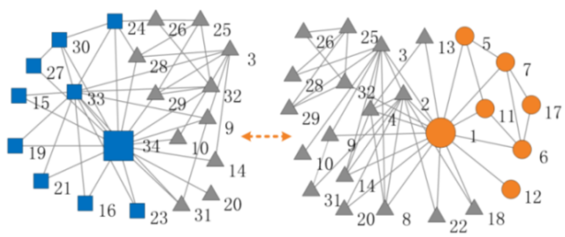

Fig. 3 Two Hops Compression On Karate Club Network

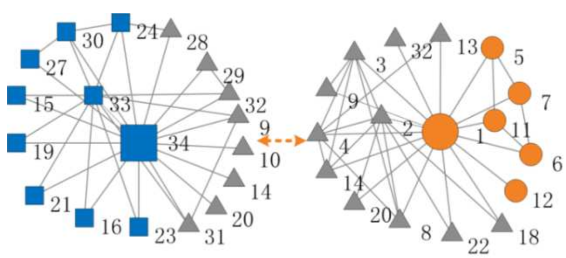

Fig. 4 One Hop Compression On Karate Club network

Fig. 5 Zero Hops Compression On Karate Club Network

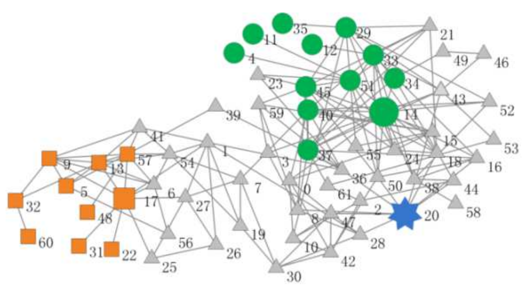

Fig. 6 Communities Discovered By GS Algorithm On Dolphin Social Network

Due to the two network optimization influence range for h = 2 so this may lead to certain communities in the node is unable to attract other communities in the node, and the compression rate of 0, such as the karate club network community C1 in compression to 2 jumps when the compression ratio is the case. In general, in the optimization of community of some nodes will still have to attract other community node capability, so the karate club network of community C34 and dolphin social networks in the community of C14、C17and C20 in compression to 2 jumps when the compression ratio is 0, the highest compression ratio up to 0.4314. In reference [5] and method identify communities were compared, in compression to 1 jump after compressing the community or still maintained its basic structure or retains the important node, while the maximum compression rates 0.75, the minimum is 0.2917. In the compression to 0 jumps, each community compression rate reached the highest, in more than 0.95.

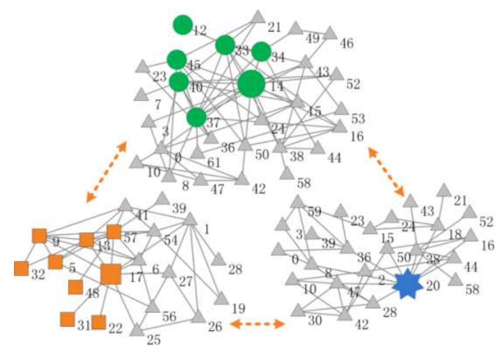

Fig. 7 Two Hops Compression On Dolphin Social Network

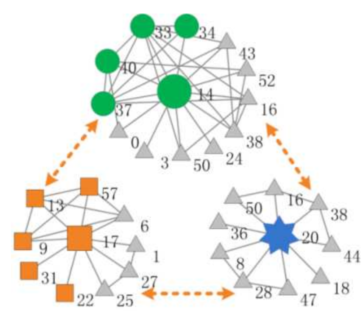

Fig. 8 One Hop Compression On Dolphin Social Network

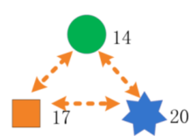

Fig. 9 Zero Hops Compression On Dolphin Social Network

5. SNC PROBLEMS

The current map compression algorithm has high time complexity, rely on a priori knowledge of parameters setting, need to adjust the parameters too much, and compression, and ignore the network community structure. Fousing on these problems, we should base on the community of node importance in social network compression SNC. In the network community for the compression object, according to the importance of nodes in the hierarchy are compressed, the compression process may need to choose whether to retain the community the important node or basic structure,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and can keep the community relationships. However, the SNC method also exists beyond the control of the compression ratio of the problem. Aiming at this problem, the following will present a new lossless compression method of social network NSNC.

Table 2 Community Compression Rate List

Network name	Community name	Community number of nodes	Compression to hop		
			2	1	0
Karate club network	C1	24	0	0.2917	0.9583
	C34	27	0.2222	0.3333	0.9630
Dolphin social network	C14	51	0.4314	0.7451	0.9811
	C17	23	0.1304	0.5652	0.9565
	C20	40	0.3750	0.7500	0.9750

5.1social NETWORK Nsnc LOSSLESS COMPRESSION METHOD

The NSNC method is with two algorithms. The method proposed is applied first to the third chapter based on attribution uncertainty community node important degree sorting algorithm for node IS, the importance of the community quantitative characterization, followed by the algorithm according to the compression rate on compression. For convenience, hereinafter is referred to as the NSNC methods in the compression algorithm for BIV (Based on Importance values). As the NSNC method will directly using IS algorithm, so this section of IS algorithm no longer.

5.2 Snc And Nsnc Node Importance In Difference.

Although the NSNC and SNC methods using topology potential theory to determine the importance of nodes in the community, but the two method of node importance meaning in the difference with bigger presence. The SNC method to nodes with representatives of the community of point relative distance as the basis for the node importance of judgment,, to distinguish between the different distances of node importance. The SNC method of the importance of nodes is a distance from the level of distinction, has a level of wholeness. The NSNC method to nodes in the community composition in the role as the basis to judge the importance of nodes for each node is to realize the importance of quantifying.

6. CONCLUSIONS

As the map compression methods and techniques in the semantic tag network, network retrieval and many other fields more and more widely, related research is concerned. According to the map compression algorithm in the presence of high time complexity, rely on a priori knowledge of parameters setting, need to adjust the parameters too much, compression and ignore the network community structure and other issues, launched a lossless network coding based on. Firstly, it should be based on the topology potential community method to determine community social network node importance lossless compression method SNC. In the proposed method and the topological potential found community node importance in relevant theorems and inferences on the basis, the method begins by greedy strategy based on the NGS algorithm for community discovery and mining communities in the different levels of importance of the nodes, followed by social network compression algorithm SNC based on the importance of nodes on the community compression. The feasibility and effectiveness of the method through the classic data sets of experiments were performed to verify the. The experimental results show that, this method not only in the process of compression can keep the community relations, but also has the ideal community compression rate, up to 0.95 or more, and can retain the community in need of an important node in the basic structure or community. Secondly, in view of the network compression method in the SNC can flexibly control the compression ratio; it provides another lossless compression method for NSNC network. In the NSNC method, it uses the topology potential community method to determine community node important degree. The experimental results show that, compared with the SNC method, NSNC method can not only achieve equivalent compression effect, but also can specify any compression ratio.

ACKNOWLEDGEMENTS

The work is supported by 2010 Plan project of national science of education under No.FFB108075.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES

[1] Tian Y, Hankins R A, Patel J M. Efficient aggregation for graph summarization. Proceedings of the ACM SIGMOD International Conference on Management of Data. New York: Association for Computing Machinery, 2008: 567-579P

[2] Hauguel S, Zhai Chengxiang, Han Jiawei. Parallel PathFinder algorithms for mining structures from graphs. 2009 Ninth IEEE International Conference on Data Mining. Miami: Institute of Electrical and Electronics Engineers Inc., 2009: 812-817P

[3] Navlakha S, Rastogi R, Shrivastava N. Graph summarization with bounded error. 2008 ACM SIGMOD International Conference on Management of Data. New York: Association for Computing Machinery, 2008: 419-432P

[4] Zhang Ning, Tian Yuanyuan, Patel J M. Discovery-driven graph summarization. 26th IEEE International Conference on Data Engineering. Long Beach: IEEE Computer Society, 2010: 880-891P

[5] Toivonen H, Zhou Fang, Hartikainen A, et al. Compression of weighted graphs. The 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. San Diego: Association for Computing Machinery, 2011: 965-973P

[6] Toivonen H, Mahler S, Zhou F. A frame work for path-oriented network simplification. Advances in Intelligent Data Analysis IX, 2010, 6065(2010): 220-231P