International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.3 (2016), pp.209-222
http://dx.doi.org/10.14257/ijmue.2016.11.3.20

Weak Connection Edges Independent Discriminant of Rapid Spanning Tree Recommendation of Social Network Community

Cao Ke-xing[1], Li Zhao-xing[2], Li Xin[3*] and Lv Zhi-han[4]

1. Information Engineering school of Yulin University in Yuyang, Yulin, Shaanxi Province, China
2. Information Engineering school of Yulin University in Yuyang, Yulin, Shaanxi Province, China
3. School of Urban Design, Wuhan University, Wuhan, China
4. SIAT, Chinese Academy of Science, Shenzhen, China
** Corresponding Author*
E-mail

Abstract

Aiming at the problem of low accuracy and high computational complexity of the traditional social network community recommendation algorithm, a rapid spanning tree detection algorithm is proposed to independently discriminate social network community with the weak connected edge, in order to improve the accuracy of community recommendation and reduce the complexity of algorithm. Firstly, according to the characteristics of social network community recommendation, the maximum spanning tree algorithm is proposed, which is based on the edge weight distribution node similarity, to realize the effective detection of social network community. Secondly, for the proposed algorithm having the problems of repeated adding and deleting of weakly connected edges and the waste of computing resources, a rapid spanning tree detection algorithm based on the independent discrimination of weakly connected edge is proposed so as to further improve the calculation efficiency of the algorithm. Lastly, the effectiveness of the proposed algorithm is verified by comparing the experimental results in the standard test database.

Keywords: *Weakly connected edge; Independent discriminant; Social network; Rapid spanning tree*

1. Introduction

Social network, with its fast and convenient communication, is becoming more and more popular. At the same time, some software and applications based on the internet gradually take social networks as the main framework, users as nodes, the information link between nodes as a marginal connection in social networks, and according to the relationship between nodes, social networks can be expressed as a directed or undirected graph, in which the dense community recommendation is a very valuable and challenging key issue in social network research [1-2].

In recent years, a new social network form typically represented by Twitter and Sina Weibo rises gradually, and its main feature is the existence of a large number of weak connection edges. Because of the one-way property of weak connection edges, social networks represent obvious heterogeneity. For example, the nodes of natural persons based on the interests pay one-way attention to the nodes of media and other topics, while the nodes between natural persons based on real relationship will build two-way attention relationship. In real networks, the bidirectional strong connection is far less than the one-

ISSN: 1975-0080 IJMUE

way weak connection subscription relationship. Therefore, efficient processing of weak connections is the key factor to improve the detection and recognition of social networks.

At present, there are many research on social network community recommendation algorithms, and research direction involves collaborative filtering, content recommendation, clustering and association rules, *etc.* Among them, collaborative filtering is based on the similarity of user interest points as the evaluation index for recommendation [4], which needs to interact with the users, reflected through the evaluation scoring form, but because social network does not have evaluation scoring concept, and data is extremely sparse, resulting in collaborative filtering unable to be applied in real social networks. And recommendation based on contents is to have community recommendation according to the previous interest points of users, which is suitable for the function of friend recommendation, as Literature [5] with natural language to express user node tweet and to find and recommend congenial friends. The design of Literature [6] is the algorithm based on Twitter users' interest of fuzzy concept set for identification community recommendation, *etc*. However, this recommendation method based on the contents is poor in universality; except for the good results of friend recommendation, the detection performance of other communities is general. In addition, a community recommendation algorithm based on association rules is proposed in Literature [7], and the clustering social network community recommendation algorithm is proposed by Literature [8].

As can be seen from the above algorithms, social network community recommendation is similar to a sort or clustering; from the inter node relationship, in fact, it is more suitable to use the spanning tree algorithm for recommendation. This paper firstly studied on the implementation of spanning tree algorithm of the social network community recommendation. In the research of spanning tree, Literature [9] proposed a modified minimum spanning tree algorithm. Literature [10] proposed a parallel algorithm of GPU minimum spanning tree, *etc.*. Different from the above literatures, this paper first learns from the standard maximum spanning tree algorithm for social network community recommendation. Then, different from the proposed design concept of social network community, this paper introduces the pre-sentence processing of weak connection edges, which can effectively prevent the large number of weak connections that leads to the problem of redundant operation of the spanning tree algorithm, thus greatly reduces the computational complexity of the algorithm.

Therefore, the main work of this paper is in two aspects as follows: first, the introduction of the maximum spanning tree algorithm to the social network community recommendation for research; second, the introduction of the pre judgment processing of weak connection edges, which effectively reduces the computational complexity of the algorithm and improves the practical application value of the algorithm.

2. Problem Description

Literature [11] proposed a modularity density (D value) theory of network community, and put forward a simplified mixed integer nonlinear programming model (MINLP) based on it, but this model requires that the number of communities is known, or else it cannot start the algorithm. Firstly, the D value network community model is introduced as follows:

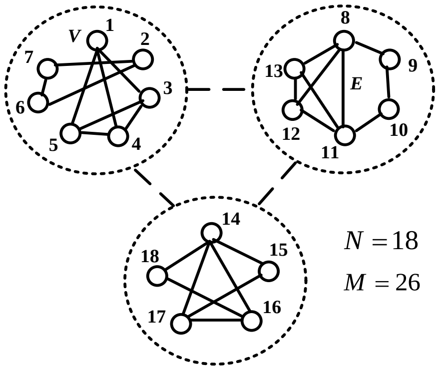


Figure 1. Simplified Model of Network Community

For the given network: $G = (V, E)$, there are N nodes and M edges, V is the collection of N nodes, E is the collection of M edges (node connection), shown as in Figure 1. Then the C community model of D value can be described as:

$$D = \sum_{i=1}^{C} d(G_i) = \sum_{i=1}^{C} (1-\lambda) L(V_i, V_i) - \lambda L(V_i, -V_i) / |V_i| \tag{1}$$

$$D = \sum_{i=1}^{C} d(G_i) = \sum_{i=1}^{C} (1-\lambda) L(V_i, V_i) - \lambda L(V_i, -V_i) / |V_i| \tag{1}$$

In the formula, λ is adjustment parameter, V_i is the node collection of community i, $-V_i$ means node collection that does not belong to community i in the network, $|V_i|$ is the cardinal number of community i, then the definition of function L is as follows:

$$L(V_a, V_b) = \sum_{i \in V_a} \sum_{i \in V_b} a_{i,j} \tag{2}$$

$$L(V_a, V_b) = \sum_{i \in V_a} \sum_{i \in V_b} a_{i,j} \tag{2}$$

In the formula, $a_{i,j}$ refers to the j column i row of adjacent matrix A, for the establishment of mathematical model of complex network community.

The network community model can be defined as:

$$\begin{cases} f = \underset{\forall 1 \le k \le n}{\arg\max} CN_{G_k}(G) = \{CM_{G_i}(G)\} | \forall i, j \\ 1 \le i, j \le n, CM_{G_i}(G) \ge CM_{G_j}(G) \end{cases} \tag{3}$$

$$\begin{cases} f = \underset{\forall 1 \le k \le n}{\arg\max} CN_{G_k}(G) = \{CM_{G_i}(G)\} | \forall i, j \\ 1 \le i, j \le n, CM_{G_i}(G) \ge CM_{G_j}(G) \end{cases} \tag{3}$$

In the formula, for any i, j, to satisfy $1 \le i, j \le n$ and $G_{si}, G_{sj} \in G_s$, $V_{si} \cap V_{sj} = V_{sij}$, $E_{si} \cap E_{sj} = E_{sij}$. If for all combination of i, j, V_{sij} and E_{sij} are both null set $\varnothing$, then G_s means a set of disjoint independent communities, or else G_s is overlapping communities, among which, V_{sij} and E_{sij} are the node set edge collection of overlapping part G_{si}, G_{sj}.

According to the D value theory model of formula (4) in Literature [7], it can be expressed as:

$$\begin{cases} f = \max \sum_{l=1}^{C} \left[(z_1 - z_2) \Big/ \sum_{j=1}^{N} x_{il} \right] \\ z_1 = (1-\lambda) \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_{il} x_{jl} \\ z_2 = \lambda \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_{il} (1 - x_{jl}) \end{cases} \quad (4)$$

$$\begin{cases} f = \max \sum_{l=1}^{C} \left[(z_1 - z_2) \Big/ \sum_{j=1}^{N} x_{il} \right] \\ z_1 = (1-\lambda) \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_{il} x_{jl} \\ z_2 = \lambda \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_{il} (1 - x_{jl}) \end{cases} \quad (4)$$

$$CM(G_s) = \sum_{G_{si} \in G_s} \left[\frac{|E(G_{si})|}{m} - \left(\frac{\sum_{v \in G_{si}} \deg(v)}{2m} \right)^2 \right] \quad (5)$$

In this paper, the above model is improved and the quality of network coverage is detected and evaluated by using the modular method. The evaluation index can be expressed as:

From the formula (5) modular network community recommendation model, if it is made for the maximum value, it should strive to increase the $|E(G_{si})| / m$ item and reduce $\left(\sum_{v \in G_{si}} \deg(v) / 2m \right)^2$ item. The former equals to the maximum of the inner-cluster edge, and the latter is equivalent to the minimum of inter-cluster edge.

Similarly, according to the D value theory, the model of formula (5) can be expressed as follows:

$$f = \max \sum_{l=1}^{C} \left[\begin{pmatrix} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_{il} x_{jl} \\ -\lambda \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_{il} \end{pmatrix} \Big/ \left(\sum_{j=1}^{N} x_{il} + b_1 \right) \right] \quad (6)$$

$$f = \max \sum_{l=1}^{C} \left[\begin{pmatrix} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_{il} x_{jl} \\ -\lambda \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_{il} \end{pmatrix} \Big/ \left(\sum_{j=1}^{N} x_{il} + b_1 \right) \right] \quad (6)$$

In the formula, λ is adjustment coefficient, existed in negative form, indicating that the model can be the maximum number of edges for each community scoring. In the detection of large communities, λ should be given a greater value to punish the endless node; correspondingly, in the detection of small communities, make $\lambda \to 0$. Although the model can seek for optimal λ to solve the problem, for the sake of simplicity, we consider λ be equal to 0.5.

3. Maximum Spanning Tree Community Recommendation

3.1. Algorithm Steps

In this paper, the solution of the community recommendation problem is depends on the triangular structure of the social network, which is considered as the smallest unit of a community. First, use the minimum spanning tree algorithm to design social network community recommendation algorithm framework. Then, the divisive hierarchical clustering algorithm is used to keep removing the layered weak edge of the split node to node. The algorithm steps are as follows:

Step 1: (the edge weight assignment). Make $G(V,E)$ as non-weighted network topology, $G'(V,E,W)$ as weighted network topology, among them, W is all the edge weights set, so $|W| = M$. The above derived edge weights represent the number of common neighbor nodes between adjacent nodes.

If the network side node is a part of the triangle structure, it will be deleted at the beginning stage of the algorithm. In order to obtain the network topology with weight $G'(V,E,W)$, it needs to take $G(V,E)$ as the input, and then to obtain the similarity matrix $\omega(v_i,v_j)$ of each edge $e(v_i,v_j) \in E$, the form is as follows:

$$\omega(v_i,v_j) = \frac{|Nghb(v_i) \cap Nghb(v_j)|}{|Nghb(v_i) \cup Nghb(v_j)|} \tag{7}$$

$$\omega(v_i,v_j) = \frac{|Nghb(v_i) \cap Nghb(v_j)|}{|Nghb(v_i) \cup Nghb(v_j)|} \tag{7}$$

In the formula, $Nghb(v)$ is the neighboring node collection of node v , then the set of edge weights can be defined as:

$$W = \mathrm{n}_{v_i,v_j \in V} \omega(v_i,v_j) \tag{8}$$

$$W = \mathrm{n}_{v_i,v_j \in V} \omega(v_i,v_j) \tag{8}$$

The above similarity matrix is similar to the Jaccard similarity measure. If the similarity between v_i, v_j is high, then $\omega(v_i,v_j)$ has larger value; conversely, $\omega(v_i,v_j)$ has smaller value. In this paper, the value range of $\omega(v_i,v_j)$ is $\omega(v_i,v_j) \in [0,1]$.

Step 2: (Maximum Cost Spanning Tree). In the process of constructing a maximum cost spanning tree T in weight network $G'(V,E,W)$, the networks are divided into non-connected subnets by deleting the smaller value edge $\omega(v_i,v_j)$.

In this construction mode, the node v will be connected to the node with the largest proportion of the joint node. The construction of the maximum spanning tree is based on the following lemma given by Literature [12]:

Lemma: For the weight network G, the removal of the minimum weight edge of the maximum cost spanning tree T is more able to retain more of the triangle network community structure.

Step 3: (Isolated Node Post-processing). After the completion of the above processes, some nodes may be out of the community because of some reasons, which are considered as noise points, or they are assigned to the community with the most common neighbor nodes. Then, the specific process of the maximum spanning tree social network community recommendation algorithm is shown as in the pseudo code 1.

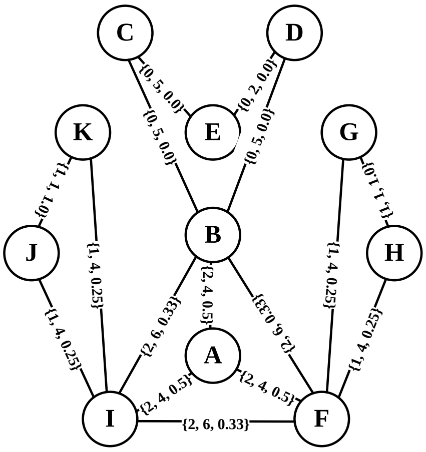
Pseudo Code 1 Maximum Spanning Tree Community Recommendation

Input: non-weight network $G(V,E)$;
Output: k network communities, $G_s = \{G_{s1}, \mathrm{n}, G_{sk}\}$;
1. begin
2. for $v \in V$ do
3. if $\deg(v) = 1$ then $v \leftarrow V \setminus \{v\}$; end
4. end
5. for $e(v_i,v_j) \in E$ do

$\omega(v_i,v_j) \leftarrow \left|Nghb(v_i) \cap Nghb(v_j)\right| / \left|Nghb(v_i) \cup Nghb(v_j)\right|$;
  if $\omega(v_i,v_j)=0$ do $E \leftarrow E \setminus \{e(v_i,v_j)\}$; end
end
$T(V,E_T) \leftarrow MST-PRIM(G,W^{-1},root)$;
$Q \leftarrow W_T$; $s \leftarrow 0$; $CM(G_0) \leftarrow 0$;
while $s=0 \,\|\, CM(G_s) > CM(G_{s-1})$ do
  if $Q=\varnothing$ do $return(G_s)$;
    $\omega(v_p,v_q) \leftarrow extract\_min(Q)$;
  end
  If $remove(e(v_p,v_q))$ creating number of new community nodes is less than 4, do
  continue;
  else do
    $remove(e(v_p,v_q))$; $s \leftarrow s+1$;
  end
  Export and save the sub community G_s from the spanning tree T
end
$return(G_{s-1})$
end

3.2. Instance Analysis

Suppose network is $M=(V,E)$, the vertex set of network M is $V=\{A \sim K\}$, the network edge set contains 16 edges, shown as Figure 2.



$c=\{|I(v_i,v_j)|,|U(v_i,v_j)|,\omega(v_i,v_j)\}$

Figure 2. Undirected Connection Graph

Step 1: As mentioned before, the suspension nodes are first removed, and the nodes either are noise points or are distributed to the adjacent nodes. Select the edge (A,I), and find all the common adjacent nodes (not including end node of the edge). The intersections are $2\left(I\left(A,I\right)=\{B,F\}\right)$ and $4\left(U\left(A,I\right)=\{J,K,B,F\}\right)$. For the other edge $\left(B,F\right)$, the intersection and union base are respectively 2 and 6. Its common neighboring nodes of the intersection are $\{A,I\}$, neighbor nodes of union are $\{C,D,G,H,A,I\}$. Weight calculation values are shown in Figure 2.

Step 2: Construct the maximum spanning tree of the undirected connected graph, and take the weight as an additional input: $M=\left(V,E,W\right)$. M is transformed as spanning tree $M'=\left(V,E',W\right)$, in which $E'\in E$. Give network graph $M=\left(V,E,W\right)$ as the input of algorithm, the created maximum spanning tree is shown as Figure 3.

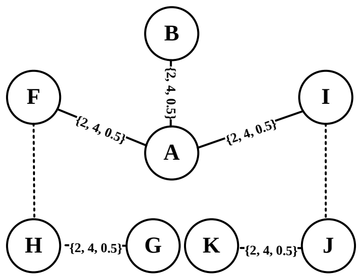


Figure 3. Maximum Spanning Tree

If the edges are two or more disjoint sets of the connecting bridge, that is, the edge weight value is 0, it should be deleted, so that the maximum spanning tree is an acyclic graph, which will be described in the following steps.

Step 3: In Figure 2, the thickness of maximum spanning tree edge is in direct proportion to the W value of their respective weights. The vertices of the spanning tree are the endpoints of the minimum weighted edges, for example, the edges $\left(F,H\right)$ and $\left(I,J\right)$ have larger weights than their connected edges $\left(H,G\right)$, $\left(J,K\right)$, $\left(A,F\right)$ and $\left(A,I\right)$. This means, nodes F and I may be attached to more groups, such as F may be attached to $\{F,G,H\}$ and $\{A,B,F,I\}$, and I may be attached to $\{I,J,K\}$ and $\{A,B,F,I\}$. Deleting weak edges $\{F,H\}$ and $\{I,J\}$ will produce disjoint sets $\{G,H\}$ 、 $\{A,B,F,I\}$ and $\{K,J\}$.

Step 4: for the suspension nodes that are assigned to their adjacent node groups and the terminal nodes that are not included in the former steps, such as nodes C, D and E, now join them in. in the original network diagram, the two edges $\left(B,C\right)$ and $\left(C,E\right)$ of the node C. Because the two edges of C are both 0, they cannot be included in the final spanning tree, which also means that node B is the cut point. Thus, when there is no weak edges that can be deleted, the $\{A,B,F,I\}$, $\{F,G,H\}$, $\{I,K,J\}$ and $\{B,C,D,E\}$ can be available as the final results of the algorithm.

4. WERSTA Community Recommendation

In the above spanning tree algorithm, the spanning tree model is firstly established, and then the weak connection edges are eliminated one by one, and finally the module is maximized. One drawback of the above algorithms is that the same edges are chosen in the process of constructing the maximum spanning tree, while they are deleted in the process of hierarchical clustering. The process of offset does not contribute to community recommendation, but increases the computational complexity of the algorithm. In this regard, the weak link edge independent discriminant of the rapid spanning tree detection algorithm of social network community is proposed as follows:

Step 1: In a dynamic hierarchical clustering method, all nodes are selected as one population or called as community at the start stage. It is assumed that E_T is the finally selected the set of edges, E_R is the collection of the edges not involved in the clustering, so the initial stage is $E_T = \varnothing$, $E_R = E$, Such as the 2-10 lines of pseudo code 2.

Step 2: The sorting according to the connection strength of the edges is similar to the above method, using Jaccard matrix (formula (7)) as the similarity measure in order to determine the intensity of the edge. In the process of detecting the mentioned sort of edges from strong to weak, detect whether adding edges will improve the module coverage, if so, remove them from E_R and add into E_T, such as the 11-14 line of pseudo code 2.

Step 3: For all the remaining edges in E_R, detect whether adding edges will improve the module coverage (independent of step 2), if so, remove it from E_R and add into E_T. The increase of module coverage and the edge sort in E_R is independent of each other, such as the 20-30 lines of pseudo code 2.

Step 4: All the nodes that the Jaccard coefficient value of all suspension nodes and all edges are 0 are added to the final spanning tree graph. The process is the post-processing process, not reflected in the pseudo code.

The computational complexity of algorithm is $O\left(|V|+|E|\right)\log|V|$, which is consistent with the algorithm in Chapter 2. But after using this method, it can avoid the redundant operation of the weak connection edge repeatedly adding and deleting.

Pseudo code 2 WERSTA spanning tree

Input: non-weight network $G(V,E)$
Output: k network communities $G_s = \{G_{s1}, \mathrm{n}, G_{sk}\}$
1. begin
2. for $i = 1:n$ do
3. $C_i = Makeset(v_i)$;
4. end
5. for $e(v_i, v_j) \in E$ do
6. $\omega(v_i, v_j) \leftarrow \left|Nghb(v_i) \cap Nghb(v_j)\right| / \left|Nghb(v_i) \cup Nghb(v_j)\right|$;
7. if $\omega(v_i, v_j) = 0$ do $E \leftarrow E \setminus \{e(v_i, v_j)\}$; end
8. end
9. $Q = W^{-1}$;
10. $j \leftarrow i \leftarrow k \leftarrow n$; $E' \leftarrow \varnothing$; $CM(G_n) \leftarrow CM(G)$;
11. repeat
12. if $Q = \varnothing$ do $return(G_s)$;
13. $\omega(v_p, v_q) \leftarrow extract\_min(Q)$;
14. end
15. if $find-set(v_p) \neq find-set(v_q)$ find-set(v_p) do
16. $Union(v_p, v_q); E' \leftarrow E' \cup e(v_p, v_q)$;
17. $E \leftarrow E \setminus e(v_p, v_q); i \leftarrow j; k \leftarrow k-1; j \leftarrow k$;
18. end

until $CM(G_i) > CM(G_j)$
  for $e(v_i, v_j) \in E$ do
    if $find-set(v_p) \neq find-set(v_q)$ find-set(v_p) do
      $Union(v_p, v_q)$;
      if $CM(G') > CM(G_j)$ do
        $i \leftarrow i-1; G_i \leftarrow G'$;
      end
    end
  Sub-graph is derived from G_i and stored as G_s
  end
  return G_s ;
end

5. Experiment and Analysis

Hardware equipment: Intel(R) Core(TM)2 Duo CPU 2.5GHz, RAM ddr3 1333 4G. Simulation software: Matlab 2013a. Operating system: Win 7 Ultimate.

5.1. Standard Test Database Verification

Benchmarking network selects three data sets with different sequence and size: Karate Club(KC), Dolphin Network(DN), American College Football(ACF), Wikivote, Enron, Amazon, DBLP seven baseline networks. The contrast algorithm selects the standard spanning tree algorithm (STA) and the Hierarchical clustering detection measure (CNM) proposed by Literature [13]. The precision and recall of evaluation index selection are shown as in Table 1. The comparison results of precision rate, recall rate, network coverage and running time of algorithm are shown in Figure 4. Among them, the precision rate and the recall rate indexes are defined as follows:

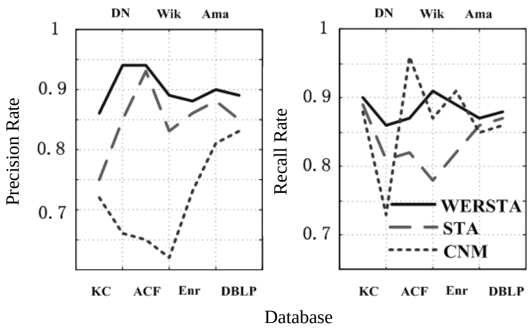
$$\begin{cases} \text{Precision rate} = \dfrac{\text{number of extracted correct information}}{\text{total number of extracted information}} \\ \text{Recall rate} = \dfrac{\text{number of extracted correct information}}{\text{total number of sample information}} \end{cases} \tag{9}$$

The precision rate and recall rate reflect the precision of community recommendation information and comprehensive rate. From simulation results of Table 1, in the precision index, WERSTA algorithm is superior to the contrast STA algorithm and CNM algorithm in the tested network. The average precision index of WERSTA algorithm was 90%, the average precision index of STA algorithm was 85%, the average precision index of CNM algorithm was 72%; the STA algorithm is superior to the CNM algorithm. In the recall rate index, WERSTA algorithm compared with STA algorithm and CNM algorithm, the performance are close, with, the average recall rate index of WERSTA algorithm being 88%, average recall rate index of STA algorithm being 84%, average recall rate index of CNM algorithm being 86%; the STA algorithm is slightly worse than the CNM algorithm.

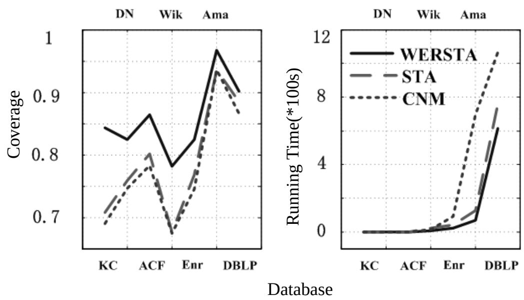
Table 1. Comparison of Precision Rate and Recall Rate

network	WERSTA		STA		CNM	
	P	R	P	R	P	R
KC	0.86	0.90	0.75	0.89	0.72	0.88

DN	0.94	0.86	0.85	0.81	0.66	0.73
ACF	0.94	0.87	0.93	0.82	0.65	0.96
Wikivote	0.89	0.91	0.83	0.78	0.62	0.87
Enron	0.88	0.89	0.86	0.82	0.73	0.91
Amazon	0.90	0.87	0.88	0.86	0.81	0.85
DBLP	0.89	0.88	0.85	0.87	0.83	0.86



Database
(a) Precision Rate and Recall Rate



(b) Coverage and Running Time

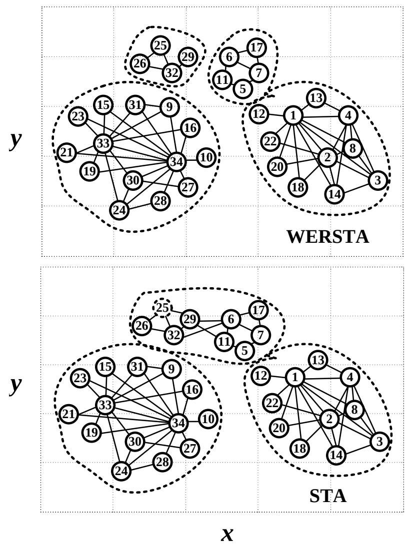
Figure 4. Experimental Contrast Curves

Figure 4 (a) is a graphical form of Table 1 data, which can be seen the WERSTA algorithm is higher than the contrast algorithm in the precision, and the results of the algorithm is more stable. In the recall rate, the WERSTA algorithm is relatively stable, the results of the CNM algorithm are more volatile, and the recall rate of STA algorithm is relatively poor. Figure 4 (b) is the contrast curve of three kinds of algorithm coverage and running time, which can be seen that the coverage index of the WERSTA algorithm is better than the STA algorithm, both of them are better than the CNM algorithm. In the running time index, the WERSTA algorithm and STA algorithm are similar, both better than the CNM algorithm.

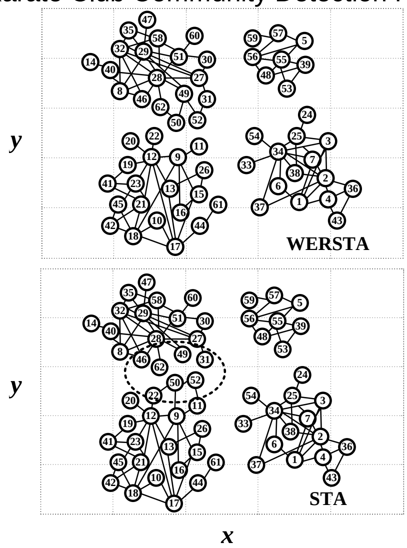
Taking benchmark test networks of Karate Club and Dolphin Network as an example, the sub community classification results of WERSTA algorithm and STA algorithm are given, as shown in Figure 5.

Figure 5 shows the community identification results (WERSTA algorithm and STA algorithm) of the two benchmark test networks, and marks the differences between the two algorithms. Because the rest of the testing data models are more complex, it is not expressed here.

The above simulation experiment is based on the standard test network, the connection between the nodes, through the common neighboring node numbers in each other to define, thus reflects the weak connection, which is a kind of suppositional weak connection. The next chapter will be based on the perspective of actual data for the simulation verification of the algorithm in this paper.



(a) Karate Club Community Detection Result



(a) Dolphin Network Community Detection Result

Figure 5. Community Detection Results Sample

5.2. Real Data Experiment

The data of Sina Weibo platform was obtained by API, and then, by the way of setting the seed users, the weak relationship community network was obtained from the method of producing interest diagram step by step in the network, so as to construct heterogeneous characteristics samples of social network. The production process is: first, five adjacent nodes are taken as the seed user nodes; second, grab the adjacent node based on the depth-first strategy or the breadth-first strategy to focus on topic node; third is real-time adjustment of grab ratio based on the network situation,.

Finally the grabbed data include more than 600 user nodes, 65 topic nodes and more than 30000 pieces of microblogging information. The program is achieved based on

Python-Java hybrid programming. STA and CNM algorithms are still selected as comparison algorithms, evaluation indexes choose precision rate and recall rate, and add harmonic mean of the two algorithms F_{meas} as a comprehensive index, which is defined as follows:

$$F_{meas} = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{10}$$

$$F_{meas} = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{11}$$

The sparse degree of network data has a weak influence on the edge number of the network. Therefore, this experiment mainly detects the influence on the above mentioned indexes with the change of sparsity degree of network. The experimental results are shown in Figure 6.

Figure 6 gives the community identification results of three comparative algorithms in the actual data. It can be seen that in the case of low sparsity, the performance of CNM algorithm is the highest, and the STA algorithm is the worst. With the sparse degree gradually increased, the performance of the algorithm is rapidly improved due to its consideration of weak connection, and gradually becomes the best algorithm. While the performance of the STA and CNM algorithms are not greatly improved for not considering weak connection.

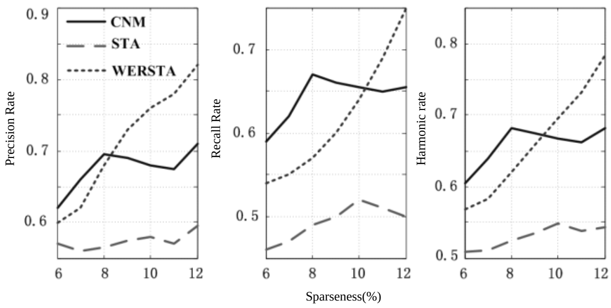


Figure 6. Experimental Comparisons of Real Data

6. Conclusion

This paper proposes a weak connection edge independent discriminant of rapid spanning tree detection algorithm of social network community, to improve the robustness of detection algorithm to the changes of weak connection edge numbers. Through the experimental comparisons of data in benchmarking network and real network, the effectiveness of proposed algorithm is demonstrated. In the next steps of research, the introduction of feedback mechanism is considered in that the evaluations of recommendation results by users are comprehensively considered for real time adjustment in the process of spanning tree detection. At the same time, in the background of big data era, consider the parallel implementation of the algorithm to improve the practical application value of the algorithm.

Acknowledgement

This work was supported by the Science and Technology Research and Development Project of Shaanxi Province (2015NY047), Industry University Research Project of Yulin City(2014cxy-05), National Natural Science Foundation of China (No. 51408442).

References

[1] Y. Wang, Y. Su and G. Agrawal, "A Novel Approach for Approximate Aggregations Over Arrays", In Proceedings of the 27th international conference on scientific and statistical database management, ACM, **(2015)**.

[2] J. Hu and Z. Gao, "Modules identification in gene positive networks of hepatocellular carcinoma using Pearson agglomerative method and Pearson cohesion coupling modularity", Journal of Applied Mathematics, vol. 2012, **(2012)**.

[3] X. Li, Z. Lv, J. Hu, L. Yin, B. Zhang and S. Feng, "Virtual Reality GIS Based Traffic Analysis and Visualization System, "Advances in Engineering Software, **(2015)**.

[4] Z. Lv, C. Esteve, J. Chirivella and P. Gagliardo, "Clinical Feedback and Technology Selection of Game Based Dysphonic Rehabilitation Tool", 2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth2015), IEEE, **(2015)**.

[5] W. Ke, "Next generation job management systems for extreme-scale ensemble computing", Proceedings of the 23rd international symposium on High-performance parallel and distributed computing. ACM, **(2014)**.

[6] L. Tonglin, "Distributed Key-Value Store on HPC and Cloud Systems", 2nd Greater Chicago Area System Research Workshop (GCASR), **(2013)**.

[7] T. Su, W. Wang, Z. Lv, W. Wu and X. Li, "Rapid Delaunay Triangulation for Random Distributed Point Cloud Data Using Adaptive Hilbert Curve", Computers & Graphics, **(2015)**.

[8] J. Hu, Z. Gao and W. Pan, "Multi-angle Social Network Recommendation Algorithms and Similarity Network Evaluation", Journal of Applied Mathematics, vol. 2013, **(2013)**.

[9] Z. Lv, T. Yin, Y. Han, Y. Chen and G. Chen, "WebVR——web virtual reality engine based on P2P network", Journal of Networks, vol. 6, no. 7, **(2011)**, pp. 990-998.

[10] J. Yang, B. Chen, J. Zhou and Z. Lv, "A portable biomedical device for respiratory monitoring with a stable power source", Sensors, **(2015)**.

[11] Z. Dongfang, "FusionFS: Toward supporting data-intensive scientific applications on extreme-scale high-performance computing systems", Big Data (Big Data), 2014 IEEE International Conference on. IEEE, **(2014)**.

[12] S. Dang, J. Ju, D. Matthews, X. Feng and C. Zuo, "Efficient solar power heating system based on lenticular condensation", Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference on, **(2014)**.

[13] X. Zhang, Y. Han, D. Hao and Z. Lv. "ARPPS: Augmented Reality Pipeline Prospect System", 22th International Conference on Neural Information Processing (ICONIP), Istanbul, Turkey. In press, **(2015)**.

[14] J. Hu and Z. Gao, "Distinction immune genes of hepatitis-induced heptatocellular carcinoma", Bioinformatics, vol. 28, no. 24, **(2012)**, pp. 3191-3194.

[15] W. Ke, "Overcoming Hadoop Scaling Limitations through Distributed Task Execution".

[16] Z. Su, X. Zhang and X. Ou, "After we knew it: empirical study and modeling of cost-effectiveness of exploiting prevalent known vulnerabilities across IAAS cloud", Proceedings of the 9th ACM symposium on Information, computer and communications security. ACM, **(2014)**.

[17] G. Bao, L. Mi, Y. Geng and K. Pahlavan, "A computer vision based speed estimation technique for localizing the wireless capsule endoscope inside small intestine", 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), **(2014)**.

[18] W. Gu, Z. Lv and M. Hao, "Change detection method for remote sensing images based on an improved Markov random field", Multimedia Tools and Applications, **(2016)**.

[19] Z. Lu, C. Esteve, J. Chirivella and P. Gagliardo, "A Game Based Assistive Tool for Rehabilitation of Dysphonic Patients", 3rd International Workshop on Virtual and Augmented Assistive Technology (VAAT) at IEEE Virtual Reality 2015 (VR2015), Arles, France, IEEE, **(2015)**.

[20] Z. Chen, W. Huang and Z. Lv, "Uncorrelated Discriminant Sparse Preserving Projection Based Face Recognition Method", Multimedia Tools and Applications, **(2016)**.

[21] Z. Lv, A. Halawani, S. Feng, H. Li and S. U. Rehman, "Multimodal Hand and Foot Gesture Interaction for Handheld Devices", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM). 11, 1s, Article 10, **(2014)**, pp. 19.

[22] K. Leng, W. Shi, J. Chen and Z. Lv, "Designing of a I-shaped less-than-truckload cross-dock: A simulation experiments study", International Journal of Bifurcation and Chaos, **(2015)**.

[23] Y. Lin, J. Yang, Z. Lv, W. Wei and H. Song, "A Self-Assessment Stereo Capture Model Applicable to the Internet of Things", Sensors, **(2015)**.

[24] J. He, Y. Geng and K. Pahlavan, "Toward Accurate Human Tracking: Modeling Time-of-Arrival for Wireless Wearable Sensors in Multipath Environment", IEEE Sensor Journal, vol. 14, no. 11, **(2014)**, pp. 3996-4006.

[25] W. Ou, Z. Lv and Z. Xie, "Spatially Regularized Latent topic Model for Simultaneous object discovery and segmentation", The 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC), **(2015)**.

[26] W. Wang, Z. Lu, X. Li, W. Xu, B. Zhang and X. Zhang, "Virtual Reality Based GIS Analysis Platform", 22th International Conference on Neural Information Processing (ICONIP), Istanbul, Turkey, **(2015)**.

[27] W. Ke, "Using Simulation to Explore Distributed Key-Value Stores for Exascale System Services", 2nd Greater Chicago Area System Research Workshop (GCASR), **(2013)**.
[28] J. He, Y. Geng, F. Liu and C. Xu, "CC-KF: Enhanced TOA Performance in Multipath and NLOS Indoor Extreme Environment", IEEE Sensor Journal, vol. 14, no. 11, **(2014)**, pp. 3766-3774.
[29] S. Zhou, L. Mi, H. Chen and Y. Geng, "Building detection in Digital surface model", 2013 IEEE International Conference on Imaging Systems and Techniques (IST), **(2012)**.
[30] N. Lu, C. Lu, Z. Yang and Y. Geng, "Modeling Framework for Mining Lifecycle Management", Journal of Networks, vol. 9, no. 3, **(2014)**, pp. 719-725.
[31] Y. Geng and K. Pahlavan, "On the accuracy of rf and image processing based hybrid localization for wireless capsule endoscopy", IEEE Wireless Communications and Networking Conference (WCNC), **(2015)**.
[32] G. Liu, Y. Geng and K. Pahlavan, "Effects of calibration RFID tags on performance of inertial navigation in indoor environment", 2015 International Conference on Computing, Networking and Communications (ICNC), **(2015)**.
[33] J. He, Y. Geng, Y. Wan, S. Li and K. Pahlavan, "A cyber physical test-bed for virtualization of RF access environment for body sensor network", IEEE Sensor Journal, vol. 13, no. 10, **(2013)**, pp. 3826-3836.
[34] W. Huang and Y. Geng, "Identification Method of Attack Path Based on Immune Intrusion Detection", Journal of Networks, vol. 9, no. 4, **(2014)**, pp. 964-971.
[35] Y. Su, "In-situ bitmaps generation and efficient data analysis based on bitmaps", In Proceedings of the 24th International Symposium on High-Performance Parallel and Distributed Computing, ACM, **(2015)**, pp. 61-72.
[36] G. Yan, Y. Lv, Q. Wang and Y. Geng, "Routing algorithm based on delay rate in wireless cognitive radio network", Journal of Networks, vol. 9, no. 4, **(2014)**, pp. 948-955.
[37] D. Jiang, X. Ying, Y. Han and Z. Lv, "Collaborative Multi-hop Routing in Cognitive Wireless Networks", Wireless Personal Communications, **(2015)**.
[38] Z. Xiaobing, "Exploring Distributed Resource Allocation Techniques in the SLURM Job Management System", Illinois Institute of Technology, Department of Computer Science, Technical Report, **(2013)**.
[39] G. Bao, L. Mi, Y. Geng, M. Zhou and K. Pahlavan, "A video-based speed estimation technique for localizing the wireless capsule endoscope inside gastrointestinal tract", 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), **(2014)**.
[40] D. Zeng and Y. Geng, "Content distribution mechanism in mobile P2P network", Journal of Networks, vol. 9, no. 5, **(2014)**, pp. 1229-1236.
[41] L. Tonglin, "ZHT: A light-weight reliable persistent dynamic scalable zero-hop distributed hash table", Parallel & Distributed Processing (IPDPS), 2013 IEEE 27th International Symposium on. IEEE, **(2013)**.
[42] W. Ke, "Optimizing load balancing and data-locality with data-aware scheduling", Big Data (Big Data), 2014 IEEE International Conference on. IEEE, **(2014)**.
[43] M. Zhou, G. Bao, Y. Geng, B. Alkandari and X. Li, "Polyp detection and radius measurement in small intestine using video capsule endoscopy", 2014 7th International Conference on Biomedical Engineering and Informatics (BMEI), **(2014)**.
[44] X. Song and Y. Geng, "Distributed community detection optimization algorithm for complex networks", Journal of Networks, vol. 9, no. 10, **(2014)**, pp. 2758-2765.
[45] J. Yang, S. He, Y. Lin and Z. Lv, "Multimedia cloud transmission and storage system based on internet of things", Multimedia Tools and Applications, **(2016)**.

Author

Cao Kexing, He received his M.S. degree in Management Science from Southwest University in Chongqing, China. He is currently a lecturer in the School of Information Engineering at Yulin University. His research interest is mainly in the area of wireless sensor network, software engineering. He has published several research papers in scholarly journals in the above research areas.

