Research and application on set pair entity similarity model of social network

Yanjun Zhao*, Chunying Zhang

College of Science, Hebei Untied University, Tangshan, 063009, China

Received 1 March 2014, www.tsi.lv

Abstract

In allusion to the certain and uncertain value which exist in the node and relationship attributes of social network, From the attributes and relation angle of the entity to analyse the similarity degree of the affirmative, negative and uncertain between them, then build the set pair entity similarity model based on the set pair analytical method and apply it to the network association detection. First of all, applying the generalized pair close potential and the generalized set loose potential in social network based on set pair analysis method, and see it as the basis of association detection; secondly, giving the set similarity calculation method based on the entity attribute and relation, from the point of view of node attribute and relations attribute to calculate respectively, by setting the weight to consolidated calculate the set pair similarity of the entity; thirdly, utilizing entity set similarity to divide network association into clustering problem, then give the association partitioning algorithm; finally, integrating with the network instance to verify the effectiveness of the new network association partitioning algorithm.

Keywords: set pair, social network, entity similarity, attribute and relation, association partitioning

1 Introduction

The social network is one relation network [1], social network analysis is one analysis method for relation to study [1, 2]. The method has a very wide range application in the social network, owing to the entity is of feature attribute and relation attribute in social network. And the social network analysis aims at studying the relation attribute [3-6], the purpose of network association partitioning algorithm is to excavate the community structure in the network, It is a clustering problem in essence, by defining the similarity degree between the network node, to adopt the clustering algorithm for the node in network to conduct clustering, then to achieve the purpose of dividing network association.

In the similarity constructor method of network node, one constructor method based on the local node information [7], Considering neighbour information of the node, if the two nodes have the same or similar neighbour nodes then regard as them are similar. The Constructor method is based on the network topology information [8] regarding the node in network as the node of receiving and transmitting signals. Guo Jingfeng [9] has proposed one namesake entity partitioning algorithm based on attribute relation figure. By calculating the similarity degree of feature attribute of the entity and the similarity of the link and through the clustering to distinguish the individuals of the same name in the social network, making the accuracy of the analysis of social networks has improved. Zhou Li [10] has proposed one partitioning calculation method of the node similarity degree in graph. By the node block structure feature in the social network to conduct the similarity degree calculation, and the experiments has proved the effectiveness of the algorithm. Xiang et al has proposed partitioning algorithm based on subgraph similarity degree [11]; this algorithm is superior to the algorithm of optimization module of Clause prominent [12]. Pan Ying has proposed association partitioning algorithm based on the node similarity degree in no weight graph [13]. But Leicht E A et al [14] made amendment for the no weight graph algorithm, and then proposed the structure partitioning algorithm of the weighted network community based on node similarity degree. But these proposed similarity calculation method and entity recognition algorithm can only be confined to one separate issue, the entity in the social network has various attributes, and the uncertain phenomenon exists in these property, How to resolve such problems, according to the authors has propose the entity in social network is constituted of the attribute [15]. Therefore, in this paper proposes the same entity recognition model based on set pair analysis theory, utilizing the attribute of entity set and relation degree of to conduct the similarity calculation. Finally, by the clustering method to set the threshold to identify the entity set, then utilize an example to demonstrate the effectiveness of the algorithm.

2 Set pair social network model

Assuming in the social relation network, any two objects v_k and v_s have the number of attribute is N (that is the

* ***Corresponding author*** e-mail: teacher_jsj@126.com

number of attribute set of two objects). Assuming the two research object have the number of the same attribute is *S*, the number of the different attribute is *P*, the number of the uncertain attribute is *F*, commanding:

$$\frac{S}{N}=a,\frac{F}{N}=b,\frac{P}{N}=c\ . \tag{1}$$

In allusion to the complexity of the nodes in social network and the uncertainty in the relation among the node, we give the following definition:

Definition 2.1 [10]: Assuming the domain set of research object $U=\{v_1,v_2,...,v_n\}$, thereinto the attribute set of the object v_k and v_s is $A(v_k)=\{x_{k1},x_{k2},..x_{km}\}$ and $A(v_s)=\{x_{s1},x_{s2},..x_{sm}\}$, so at one moment the contact degree of v_k and v_s is

$$\begin{aligned}&\rho(v_k,v_s)(t)=a_{k,s}(t)+b_{k,s}(t)i+\\&c_{k,s}(t)j\rho(e_k,e_s)(t)=a(t)+b(t)i+c(t)j\end{aligned}, \tag{2}$$

where $v_k\in U$ and $v_s\in U$, *i* is the differential label and takes the different in [-1,1] depending on the circumstance. *j* only plays a marked role and takes the value is -1.

If the weight of attribute is different, assuming:

$$\omega_k(k=1,2,\cdots,N,\sum_{k=1}^{N}\omega_k=1).$$

Assuming attributes according to the order of *S*, *F*, *P* to align and consecutive number. The contact degree is:

$$\begin{aligned}&\rho(v_k,v_s)(t)=a_{k,s}(t)+b_{k,s}(t)i+c_{k,s}(t)j=\\&\sum_{k=1}^{S}\omega_k(t)+\sum_{K=S+1}^{S+F}\omega_k(t)i+\sum_{k=S+F+1}^{N}\omega_k(t)j\end{aligned}, \tag{3}$$

where $i\in[-1,1];j=-1;a_{k,s}(t)+b_{k,s}(t)+c_{k,s}(t)=1$.

Definition 2.2: Assuming the matrix $R=(\rho(\mathrm{v}_k,v_s))_{k\times s}$, R is called the set pair contact relation matrix, herein $\rho(v_k,v_s)$ is the element of the set pair contact relation matrix, and behalf of the set pair contact degree between v_k and v_s. $R(t)=(\rho(v_k,v_s)(t))_{n\times n}$ is behalf of the relation degree about every research object in set pair social network, the matrix can be expressed:

$$R(t)=\begin{bmatrix}\rho(v_1,v_1)(t) & \rho(v_1,v_2)(t) & ... & \rho(v_1,v_n)(t)\\ \rho(v_2,v_1)(t) & \rho(v_2,v_2(t) & ... & \rho(v_2,v_n)(t)\\ .... & ... & ... & ...\\ \rho(v_n,v_1)(t) & \rho(v_n,v_2)(t) & ... & \rho(v_n,v_n)(t)\end{bmatrix} \tag{4}$$

Do not take the time *t* into consideration, the *R* matrix is a static relation matrix, otherwise it is the dynamic relation matrix, with the time goes by, the nodes in social network constantly change, thereby to obtain a relation matrix is constantly updated. By analyzing the relation matrix at the different moments, then to find the trend of the social network.

Definition 2.3: Assuming the vertex set of social network figure at the moment of *t* is $V(t)=\{v_1,v_2,...,v_n\}$, the attribute of vertex is $VX(t)=\{vx_1,vx_2,...,vx_m\}$, the set pair relation matrix at the moment of *t* is $R(t)=(\rho(v_k(t),v_s(t)))_{n\times n}$, so the dynamic social network analysis model is $GR(t)=(V(t)(VX),R(t))$. Do not take the time *t* into consideration, the model is the static social network analysis model.

Definition 2.4: The two individuals' relation contact degree in social network is $\rho(v_k,v_s)=a+bi+cj$, the ratio of relative same degree e^a and the relative opposed degree e^c is called the individual generalized set pair potential in the social network, that is:

$$Tread(v_k,v_s)_G=\frac{e^a}{e^c}\ . \tag{5}$$

Even if $c=0$, we can still judge the relation trend of the two nodes in network.

Definition 2.5: The two individuals' relation contact degree in social network is $\rho(v_k,v_s)=a+bi+cj$, the ratio of relative same degree e^a and product of the relative opposed degree e^c the relative different degree e^b is called the individual generalized set pair loose potential in social network, that is:

$$Tread(v_k,v_s)_G=\frac{e^a}{e^{c+b}}\ . \tag{6}$$

The uncertainty term (different degree) is converted to opposition term (opposed degree) to study the trend that the node in social network may withdraw the network.

Definition 2.6: The two individuals' relation contact degree in social network is $\rho(v_k,v_s)=a+bi+cj$, the ratio of relative same degree e^a and product of the relative different degree e^b the relative opposed degree e^c is called the individual generalized set pair tight potential in social network, that is:

$$Tread(v_k,v_s)_G=\frac{e^{a+b}}{e^c} \tag{7}$$

The uncertainties term (different degree) are converted to the same term (same degree) to study the trend that the node and other nodes in social network.

For the generalized set pair loose potential and generalized set pair tight potential to analyse respectively from the angle of node relation may be looser or tighter. They can also be divided again.

The generalized set pair loose potential is the lower limit of the generalized set pair potential, the generalized set pair tight potential is the upper limit of the generalized set pair potential, within this range, the loose and tight relation between the two nodes is mutual restraint and mutual influence, and mutual conversion under certain conditions.

3 The entity set pair similarity degree calculation method based on attribute relation

The social network is a complex network, the relation of the nodes and among them all have the corresponding attribute feature, this paper will build attribute figure based on the attribute feature, in the meanwhile to describe the relation of entity attributes and among them. To calculate the similarity of the entity in the network, in the meanwhile to concern attribute information of the node and relation information among them. At present, the clustering method of taking the object attribute and relation information among the objects into account and collectively called the clustering based on attribute-relation. Thereinto, the hypertext document clustering, scientist cooperative relation clustering, telecom customer division, and so on. Because of the rich data resource, attribute of the research object and relation information is easier to obtain, and has taken the lead to develop.

An important class algorithm among the clustering method based on attribute-relation is the method based on similarity degree. The representative is the HyPursuit algorithm of Weiss et al [16], the M-S algorithm of Modha et al [17], Wu Lingyu of University of Science and Technology Beijing has proposed the calculation method of the integrated similarity degree among the object and the similarity degree among the class. And design the appropriate strategy to achieve clustering from the bottom up. The algorithm calculated based on the attributes of the object distance and relation distance, assuming these attribute values and relation are certain and completed. In fact, the attribute value or relation is vague, uncertain and incomplete, so only utilize the attribute difference value to calculate the similarity degree among the object is not enough. This section proposes the set pair similarity calculation method of attribute-relation based on set pair analysis theory. In the meanwhile, consider the same, non-same and unsure of the attribute and relation whether is the same three-dimensional similarity, so as to solve the unicity problem in the work [16-18].

3.1 THE ENTITY SET PAIR ATTRIBUTE CONTACT DEGREE BASED ON ATTRIBUTE

In allusion to the complexity of the node in social network, and the uncertainty of node relation, we give the following definition:

Definition 3.1: (entity set pair attribute contact degree) assuming all the entities in the social network as domain set $U=\{v_1,v_2,\ldots,v_n\}$, v_k and v_s are entities, they are of the feature attribute set are $\alpha(v_k)=\{\alpha_{k1},\alpha_{k2},\ldots,\alpha_{km}\}$ and $a(v_s)=\{a_{s1},a_{s2},\ldots,a_{sm}\}$, the number of the same attribute value of v_k and v_s is p, the number of the different attribute value is q, the number of the uncertain attribute value is r, and $p+q+r=m$, the similarity contact degree of the entity v_k and v_s in network is:

$$\tau(v_k,v_s)=\frac{p}{m}+\frac{r}{m}i+\frac{q}{m}j=a_{ks}+b_{ks}i+c_{ks}j,\quad 0\le\tau(v_k,v_s)\le 1. \tag{8}$$

Thereinto $v_k\in U$ and $v_s\in U$, i is the differential label and takes the different in [-1,1] depending on the circumstance. j only plays a marked role and takes the value is -1.

Definition 3.2: (Set pair attribute similarity contact vector) assuming in network, the node v_k has the number of neighbour node is m, the similarity degree of its neighbour node and among them to build the vector $L(v_k)=(\tau(v_k,v_s))_{m\times m}$, $L(v_k)$ is called the set pair attribute similarity contact vector, it is expressed as:

$$R(v_k)=\{\tau(v_k,v_{k1}),\tau(v_k,v_{k2}),\ldots,\tau(v_k,v_{km})\}. \tag{9}$$

In allusion to the node in social network, then calculate the set pair attribute contact degree of two entity object of connecting the edge, and signs in network to build the weighted network based set pair attribute contact degree, as showed in Figure 1.

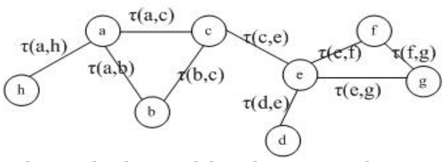

FIGURE 1 The weighted network based set pair attribute contact degree

The set pair attribute similarity contact vector of every node is showed in Table 1.

TABLE 1 The set pair attribute similarity contact vector of Figure 1

Node	Similar vector
a	$\{\tau(a,h), \tau(a,b), \tau(a,c)\}$
b	$\{\tau(a,b), \tau(a,c)\}$
c	$\{\tau(a,c), \tau(b,c), \tau(c,e)\}$
d	$\{\tau(d,e)\}$
e	$\{\tau(c,e), \tau(d,e), \tau(e,f), \tau(e,g)\}$
f	$\{\tau(e,f), \tau(f,g)\}$
g	$\{\tau(f,g), \tau(e,g)\}$
h	$\{\tau(a,h)\}$

When to calculate the set pair attribute similarity contact based on the node, we should only consider the node connected with it and ignore the node without association. The vector makes preparation for calculating the comprehensive set pair similarity degree later.

3.2 THE ENTITY SET PAIR ATTRIBUTE SIMILARITY CONTACT DEGREE BASED ON RELATION

Definition 3.3: (the one level neighbour entity) the entity which connected to certain entity v_q in network is called one level neighbour entity of this entity, all the one level neighbour entity of the entity v_q to build the set and denoted as $L(v_q)$.

As showed in Figure 1, the one level neighbour entity of the entity *e* is *c*, *d*, *f*, *g* and denoted as $L(e)=\{c,d,f,g\}$.

Definition 3.4: (the common one level neighbour entity) if certain entity v_q and entity v_r, v_s connect directly in network, we call the entity v_q is the common one level neighbour entity for entity v_r, v_s .The common one level neighbour entity of entity v_r, v_s to build the set and denoted as $LQ(v_r,v_s)$. Obviously:

$$LQ(v_r,v_s)=L(v_r)\bigcap L(v_s). \tag{10}$$

Definition 3.5: (the two level neighbour entity): the one level neighbour entity of the one level neighbour entity of certain entity v_q in network, we call all the two level neighbour entity of the entity v_q to build the set and denoted as $LE(v_q)$.

As shown in Figure 1, the one level neighbour entity of the entity *e* is *a*, *b* and denoted as $L(e)=\{c,d,f,g\}$.

Definition 3.6: (the common two level neighbour entity) if certain entity v_q is the two level neighbour entity of entity v_r and v_s in network, we call entity v_q is the common two level neighbour entity of entity v_r and v_s, the two level neighbour entity of entity v_r and v_s build the set and denoted as $LEQ(v_r,v_s)$. Obviously:

$$LEQ(v_r,v_s)=LE(v_r)\bigcap LE(v_s). \tag{11}$$

Here, we think that the two neighbour entity may have relation with the entity v_q and may not. So, it is uncertain, yet, the node in addition to one and two neighbour entities have relation with entity is slim chance, so the contact with the other entities is to the contrary.

Definition 3.7: (the entity set pair similarity contact degree based on relation) assuming the number of entity is n in network, the entity v_r and v_s have the common one level entity set is $LQ(v_r,v_s)$ and the common two level entity set is $LEQ(v_r,v_s)$, and denoted as:

$$\mu(v_r,v_s)=\frac{|LQ(v_r,v_s)|}{n}+\frac{|LEQ(v_r,v_s)|}{n}i+\frac{|n-LQ(v_r,v_s)-LEQ(v_r,v_s)|}{n}j. \tag{12}$$

The entity set pair similarity contact degree based on relation, it is abbreviated as:

$$\mu(v_r,v_s)=u_a+u_b i+u_c j \tag{13}$$

3.3 THE ENTITY SET PAIR ATTRIBUTE SIMILARITY CONTACT DEGREE BASED ON ATTRIBUTE-RELATION

Definition 3.8: (The entity set pair attribute synthesized similarity contact degree based on attribute-relation) assuming v_k, v_s are the two entities based on attribute contact degree in weighted network, so the entity set pair synthesized similarity contact degree based on attribute-relation of v_k, v_s is:

$$Sim(v_k,v_s)=\alpha\times\tau(v_k,v_s)+(1-\alpha)\times u(v_k,v_s), \tag{14}$$

abbreviated as:

$$Sim(v_k,v_s)=Sim_a+Sim_b i+Sim_c j, \tag{15}$$

where $\tau(v_k,v_s)$ is set pair contact degree for the entity v_k, v_s, $u(v_k,v_s)$ is the entity set pair similarity contact degree for the entity v_k, v_s. α is parameter and $\alpha\in[0,1]$.

The entity set pair synthesized similarity contact degree based on attribute-relation will make further fusion for attribute and relation of the entity, than the single entity set pair similarity contact degree based on attribute or the entity set pair similarity contact degree based on relation can better reflect the true similarity degree among the entities, and have better entity recognition ability.

3.4 THE SUB GRAPH SET PAIR SYNTHESIZED SIMILARITY CONTACT DEGREE BASED ON ATTRIBUTE-RELATION

Definition 3.9: (The block graph set pair attribute synthesized similarity contact degree based on attribute-relation) assuming $G_p=\{V_1^p,V_2^p,...,V_{|G_p|}^p\}$ and $G_q=\{V_1^q,V_2^q,...,V_{|G_q|}^q\}$ is the two sub graph based on in weighted network set pair attribute contact degree, so the sub graph set pair synthesized similarity contact degree based on attribute-relation for the sub graph G_p and G_q is:

$$Sim(G_p, G_q) = \frac{\sum_{j=1}^{|G_q|}\sum_{i=1}^{|G_p|} Sim(V_i^P, V_j^q)}{|G_p| \times |G_q|}. \tag{16}$$

4 The network associations' detection algorithm

After giving the definition of entity set pair synthesized similarity contact degree based on attribute-relation, the network association partitioning has transformed into a clustering problem.

4.1 THE ALGORITHM IDEA

First, taking each entity as an initial sub graph, iterative selected two sub graph of the highest value of the entity set pair synthesized similarity contact degree based on attribute-relation to merge, until all the entities are divided into one sub graph. The clustering results is final to output one tree, the root node contains all entities, the leaf nodes as a single entity.

When to solve the practical problem can terminate the iterative process, with the fusion of sub graph, the biggest similarity contact value has showed a decreasing trend, So we can set the set pair synthesized similarity degree threshold χ, When the value of Sim_a of set pair synthesized similarity degree current maximum sub graph is less than the threshold value, considering it has finished the clustering of similar entities, the algorithm is over and output sub graph and isolated node (sub graph contains only one entity).

4.2 THE ALGORITHM STEPS

Input:
The social network consists of m entities, the value set of every entity has n attribute feature, the calculation parameter of set pair synthesized similarity degree is α and $\alpha \in [0,1]$, the set pair similarity degree threshold is χ.
Output:
The associations and isolated node.
Steps:
1) Assuming the social network attribute figure of describing the n nodes is $GA = (V(VA, LV), E(EA, LE))$, the vertex attribute threshold as showed in Table 2.

TABLE 2 The vertex threshold table

v	va_1	va_2	…	$va_{\|V\|}$
v_1	Lva_{11}	Lva_{12}	…	$Lva_{1\|V\|}$
v_2	Lva_{21}	Lva_{22}	…	$Lva_{2\|V\|}$
…	…	…	…	…
$v_{\|V\|}$	$Lva_{\|V\|1}$	$Lva_{\|V\|2}$	…	$Lva_{\|V\|\|V\|}$

2) From the threshold table to calculate the attribute contact degree $\tau(v_k, v_s)$, $(k, s = 1, 2, 3, ..., n)$ between the two nodes, then to get the set pair attribute contact degree matrix, The Table 3 shows the set pair attribute contact degree matrix of possessing the 5 nodes.

Table 3 set pair attribute contact degree matrix

	v_1	v_2	v_3	v_4	v_5
v_1	$\tau(v_1,v_1)$	$\tau(v_1,v_2)$	$\tau(v_1,v_3)$	$\tau(v_1,v_4)$	$\tau(v_1,v_5)$
v_2	$\tau(v_2,v_1)$	$\tau(v_2,v_2)$	$\tau(v_2,v_3)$	$\tau(v_2,v_4)$	$\tau(v_2,v_5)$
v_3	$\tau(v_3,v_1)$	$\tau(v_3,v_2)$	$\tau(v_3,v_3)$	$\tau(v_3,v_4)$	$\tau(v_3,v_5)$
v_4	$\tau(v_4,v_1)$	$\tau(v_4,v_2)$	$\tau(v_4,v_3)$	$\tau(v_4,v_4)$	$\tau(v_4,v_5)$
v_5	$\tau(v_5,v_1)$	$\tau(v_5,v_2)$	$\tau(v_5,v_3)$	$\tau(v_5,v_4)$	$\tau(v_5,v_5)$

3) According to the attribute contact degree and the edge adjacency matrix to form weighted network set pair similarity contact vector which based on set pair attribute contact degree.
4) In weighted network, identifying the common one level neighbour entity and the common two level neighbour entity of any two nodes v_k, v_s, calculating the entity set pair similarity contact degree $\mu(v_k, v_s) = u_a + u_b i + u_c j$, where $(k, s = 1, 2, \ldots, n)$ based on relation.
5) According to the value of α to calculate the entity set pair synthesized similarity contact degree $Sim(v_k, v_s) = Sim_a + Sim_b i + Sim_c j$ of any two nodes v_k, v_s based on attribute-relation.
6) Each node is initialized to a sub graph.
7) Identifying the two looser sub graphs of the current state of the entity set pair synthesized similarity contact degree under the generalized set pair tight potential, calculating the contact degree among the sub graph block.
8) Judging the set pair synthesized similarity contact degree $Sim_a > \chi$ among $Sim(G_p, G_q)$ if it is set up then merge the entity node G_p, G_q and denote $G_{\max} = G_p \cup G_q$, otherwise end the algorithm and output sub graph and isolated node.
9) Applying the Equation (16) to update the set pair synthesized similarity contact degree of $G_{\max}$ and other sub graphs. Returning to 8).

By the analysis, the clustering process is actually the process of network association partitioning its time complexity is $O(tn)$. Herein, the number of entity node in network is n, the number of iterations is t, the efficiency of the algorithm is higher. In addition, when to calculate the similarity contact degree, at the same time to take into the certain-uncertain of node attribute and the distance of node attribute set pair distance and the object, the result of division of the algorithm has more ideal.

5 The calculation example

Next, we set one simple network figure as example to illustrate the process of partitioning, as shown in Figure 2.

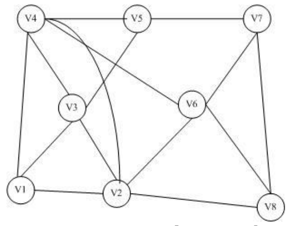

FIGURE 2 One simple network

Assuming the attribute set of eight nodes is shown in Table 4, thereinto, the blank indicates unknown, which is the uncertainty.

TABLE 4 The node attribute set in Figure 3

Attribute / node	A1	A2	A3	A4	A5
v_1	F	JS	Jsj	Shuxue	
v_2	M	FJS	Jsj		fayu
v_3	F		Dianzi	Shuxue	yingyu
v_4	M	JiangS		Shuxue	Yingyu
v_5	M	Zhuj	Zdh		fayu
v_6		JS	Jsj	Shuxue	fayu
v_7	M	Zhuj		wuli	yingyu
v_8	M		Zdh	wuli	fayu

1) Calculating the attribute set pair contact degree matrix between the two nodes.

2) Building the weighted network based on attribute set, as shown in Figure 3:

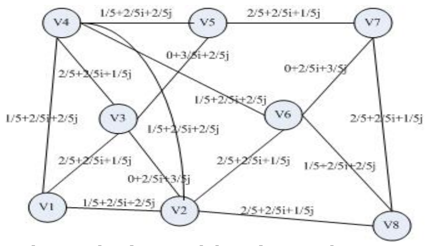

FIGURE 3 the weighted network based on attribute set pair contact degree

3) According the weighted network to build the set pair similarity vector, as shown in Table 5.

4) According to the weighted network and adjacency matrix to identify the common one level neighbour entity and the common two level neighbour entity, then calculate the set pair contact degree based on relation.

5) Setting $\alpha = 0.5$, then calculate the synthesized set pair similarity degree of every node, as shown in Table 6.

TABLE 5 The set pair similarity vector of the nodes

node	similarity vector
v_1	{1/5+2/5*i*+2/5*j*, 2/5+2/5*i*+1/5*j*, 1/5+2/5*i*+2/5*j*}
v_2	{1/5+2/5*i*+2/5*j*, 2/5*i*+3/5*j*, 1/5+2/5*i*+2/5*j*, 2/5+2/5*i*+1/5*j*, 2/5+2/5*i*+1/5*j*}
v_3	{2/5+2/5*i*+1/5*j*, 2/5*i*+3/5*j*, 2/5+2/5*i*+1/5*j*, 3/5*i*+2/5*j*}
v_4	{1/5+2/5*i*+2/5*j*, 2/5+2/5*i*+1/5*j*, 1/5+2/5*i*+2/5*j*, 1/5+2/5*i*+2/5*j*, 1/5+2/5*i*+2/5*j*}
v_5	{3/5*i*+2/5*j*, 1/5+2/5*j*+2/5*j*, 2/5+2/5*i*+1/5*j*}
v_6	{2/5+2/5*i*+1/5*j*, 1/5+2/5*i*+2/5*j*, 1/5+2/5*i*+2/5*j*, 1/5+2/5*i*+3/5*j*, 1/5+2/5*i*+2/5*j*}
v_7	{2/5+2/5*i*+1/5*j*, 2/5*i*+3/5*j*, 2/5+2/5*i*+1/5*j*}
v_8	{2/5+2/5*i*+1/5*j*, 1/5+2/5*j*+2/5*j*, 2/5+2/5*i*+1/5*j*}

TABLE 6 The result of the synthesized set pair contact degree of all the node in FIGURE 3

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
v_1	1+0*i*+0*j*	0.2250+	0.3250+0.	0.2250+	0.1250+0.3250*i*+0.5500*j*	0.3625+0.16	0.0000+0.2000*i*+0.	0.0625+0.2625*i*+0.6750*j*
v_2		1+0*i*+0*j*	0.1250+0.	0.2875+	0.3250+0.1000*i*+0.5750*j*	0.4250+0.16	0.2250+0.2625*i*+0.	0.2650+0.3250*i*+0.4125*j*
v_3			1+0*i*+0*j*	0.3875+	0.0625+0.4250*i*+0.5125*j*	0.1625+0.20	0.1625+0.2000*i*+0.	0.2625+0.2000*i*+0.5375*j*
v_4				1+0*i*+0*j*	0.1625+0.3250*i*+0.5125*j*	0.2250+0.20	0.3250+0.1625*i*+0.	0.2250+0.2000*i*+0.5750
v_5					1+0*i*+0*j*	0.2250+0.26	0.2000+0.2625*i*+0.	0.0625+0.3625*i*+0.5750*j*
v_6						1+0*i*+0*j*	0.0625+0.2625*i*+0.	0.2875+0.3875*i*+0.3250*j*
v_7							1+0*i*+0*j*	0.3250+0.3250*i*+0.3500*j*
v_8								1+0*i*+0*j*

6) Calculating the generalized set pair tight potential among every node and other nodes, if the generalized set pair tight potential greater than or equal to 1, then divided into a network, calculating the synthesized set pair similarity degree of the network sub graph until all the nodes are divided into the corresponding network or become isolated nodes.

The result of the final partition in this instance, as shown in Figure 4.

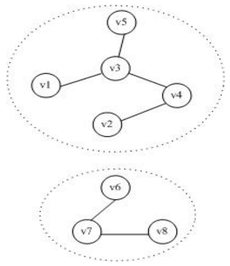

FIGURE 4 the network partition result of Figure 3

6 Conclusions

The better your paper looks, the better the Journal looks. Thanks for your cooperation and contribution. This article aims at the node and its relation in network to conduct analysis based on set pair analysis ideology, Consider the node has attribute and attribute values may exist the uncertainty, and the node relation contains the one neighbour and two neighbour, and the two neighbour can transform into the one neighbour, The other level neighbour is regarded as the relation is no close enough or unrelated, From two angles of the node attribute and the relation to calculate respectively its set pair similarity degree, then to integrate to get the synthesized set pair similarity degree. So, it is more consistent with the actual situation then only considering the calculation of one level relation similarity degree, and the result is more reasonable. Applying the method of calculating node similarity degree into the network association detection, then get the effect of good partition.

Acknowledgments

This work is supported by Hebei Province Natural Science Fund Project (No.F2012209019).

References

[1] Scott J 2000 Social Network Analysis: a Handbook *Sage Publications* London
[2] Lin J 2009 The social network analysis, method and its application *Beijing Normal University Press* Beijing
[3] Girvan M, Newman M E J 2002 Community Structure in Social and Biological Networks *PNAS* 1999 **12** 7821-6
[4] Budak C, Agrawal D, Abbadi A E 2011 Structural trend analysis for online social networks *Proceedings of the VLDB Endowment* **4**(10) 101-11
[5] Deng C, Zheng S, He X, Yan X, Han J 2005 Mining Hidden Community in Heterogeneous Social Networks *LinkKDD'05* Chicago, IL USA
[6] Yang D, Xiao Y, Wang W 2010 Statistical Model-Based Analysis and Predication of Collective Attentions in social networks *Journal of Computer Research and Development* **47**(Suppl) 378-84
[7] Liu Z, Li P, Zheng Y, Sun M 2008 Community detection by affinity propagation, *Technical Reports No. 001, 200, Dep of Computer Science and Technology* Tsinghua University Beijing China
[8] Hu Y, Li M, Zhang P, Fan Y, Di Z 2008 Community detection by signaling on complex networks *Phys Rev E* **78** 016115
[9] Guo J, Hao D 2010 One Same Name Entity Distinguish Algorithm Based on the Attribute Relationship Diagram *ICIC Express Letters (in Chines)*
[10] Zhang C, Liang R, Liu L 2011 The model of set pair social network analysis and its application *Journal of Hebei Polytechnic University* **33**(3) 99-103
[11] Xiang B, Chen B H, Zhou T 2009 Finding community structure based on subgraph similarity *Studies in Computational Intelligence* **207**(5) 73-81
[12] Claset A, Newman M E J, Moore C Finding 2004 *community structure in very large networks* Physical Review E **70**(6) 066111
[13] Pan Y, Li D-H, Liu G-J, Liang J-Z 2010 Detecting community structure in complex networks via node similarity *Physica A: statistical Mechanics and its Applications* **389**(14) 2849-57
[14] Leicht E A,Holme P, Newman M E J 2006 *Vertex similarity in networks* Physical Review E 73(2) 026120
[15] Guo J, Zhang C, Chen X 2011 Attribute Graph and Its Structure *ICIC Express Letters* **5**(8A) 2611-6
[16] Weiss R, Velez B, Sheldon M 1996 HyPursuit: a hierarchical network search engine that exploits content-link hypertext clustering *Proceedings of the 7tth ACM conference on Hypertext, ACM Press* New York 180-93
[17] Modha D, Spangler W 2004 Clustering hypertext with applications to Web searching USA 2004/0049503AI 3-11
[18] Wu L, Gao X, Wu S 2011 Clustering algorithm based on attribute-relationship integrated similarity *Application Research of Computer* **28**(1) 44-7

Authors

Yanjun Zhao, born on July 5, 1977, Tangshan, China

Current position, grades: master of engineering, lecturer of Hebei Untied University.
University studies: master's degree at Hebei Technology University.
Scientific interest: data mining, social network, computer network.
Publications: 10 papers.
Experience: Teacher of computer fundamentals for more than ten years in Hebei Untied University.

Zhang Chunying, born on October 3, 1969, Tangshan, China

Current position, grades: Ph.D. candidate, master supervisor, professor of Hebei Untied University.
University studies: Ph.D degree in computer science and its application at Yanshan University in 2012.
Scientific interest: data mining, social network.
Publications: 30 papers.