Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 14, no. 7, July 2018, pp. 1590-1599
DOI: 10.23940/ijpe.18.07.p23.15901599

Dynamic Community Mining based on Behavior Prediction

Xiao Chen[a,*], Xinzhuan Hu[b], Xiao Pan[c], and Jingfeng Guo[d,e]

[a]*Network Technology Center, Hebei Normal University of Science and Technology, Qinhuangdao, 066004, China*
[b]*College of Economics and Management, YanShan University, Qinhuangdao, 066004, China*
[c]*College of Economic and Management, Shijiazhuang Tiedao University, Shijiazhuang, 050043, China*
[d]*College of Information Science and Engineering, YanShan University, Qinhuangdao, 066004, China*
[e]*The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao, 066004, China*

Abstract

Dynamic network research has been a new trend in recent years. Based on the influence of vertex behavior on community structure, this paper studies signed network dynamic community mining. Firstly, the set pair connection degree is introduced to describe the relation between vertices, and the edge prediction model of signed network is proposed by taking into account the variability of the relation between vertices. Secondly, based on the prediction model, a set pair signed networks dynamic model is proposed by adding time axis *T* to the signed network. Then, based on the dynamic model, the evolution of signed networks and community discovering are studied. Finally, network evolution law and community stability are analyzed by using the connection trend and connection entropy in set pair theory, and the accuracy and validity of the dynamic community mining algorithm are verified by experiments.

Keywords: signed networks; network evolution; dynamic community mining; set pair theory; behavior prediction

(Submitted on March 29, 2018; Revised on May 5, 2018; Accepted on June 8, 2018)

1. Introduction

Currently, signed networks have been widely used in many fields and have achieved some research results. The research is mainly divided into three categories [5]: (1) Theory research based on structural balance theory [9] and social status theory [4]. (2) Static topological analysis, which includes measurement standards [13], community structure [2,3,23], network imbalance [7,10,19]. (3) Evolutionary dynamics research [1,8,14,15,16,17,18,20,21]. The existing research on signed networks is mainly based on the static topology. Researchers integrate the network data to mine the static community structure. However, the number of vertices and the relation between vertices changes with time, which leads to network structure changes. The existing static network analysis methods conceal the dynamic changes and cannot detect the changes. Therefore, the research of the dynamic community discovering and evolving has become one of the research hotspots in signed network analysis.

The essential characteristic of signed network dynamics is the evolutionary behavior of vertices. As shown in Figure 1, there is no connection between *v*2 and *v*3 at the current moment, whereas *v*2 and *v*3 have 4 common neighbors, which are *v*1, *v*4, *v*5, and *v*7, including common friends, common enemies, etc. Then, at the next time, does *v*2 and *v*3 generate link? If there is a link, is it positive or negative? If we grasp the behavior characteristics of vertices, the relation between vertices at the next moment can be predicted. Therefore, it is possible to analyze the dynamic evolutionary behavior of a network or community by predicting the vertex behavior tendency.

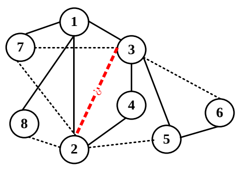


Figure 1. Signed network

* Corresponding author.
E-mail address: chenxiao0604@163.com

This paper focuses on the influence on the change of the network community structure because of the possible changes of vertex behavior, when the number of vertices remains constant. The main contributions are as follows: (1) According to the characteristics of topological structure and signed attribute, based on the set pair theory, the connection degree between vertices is depicted by defining the identical-discrepancy-contrary relation between vertices. Then, the prediction method of the relation between vertices is given. (2) Based on this prediction method, dynamic network model with time axis T is established, that is $SN_T=\{SN_{(1)}, SN_{(2)},\ldots, SN_{(T)}\}$, and a new dynamic community mining method based on the signed network is proposed. (3) The communities are mining at each point of time. The change of individual and community structure is analyzed by a connection trend. The network is aggregated or decomposed. The changing trends of dynamic communities are explored and the stability and evolution state of the network community is analyzed by connection entropy. (4) The correctness of the community mining method and the rationality of set pair analysis are proven by experiments.

2. Related Work

According to the dynamic components in the network, the research of the dynamic community can be divided into two kinds. (1) The dynamic model based on vertex and edge, such as the classic BA model [11,12]. (2) The dynamic model based on edge, such as the ER model [6] or WS model [22]. Currently, there is some research on the dynamic evolution of signed networks. These mainly focus on the extraction and superposition of multiple discrete network snapshots based on time slices, so as to realize the dynamic community evolution of signed networks [20]. For example, Wang [21] proposed a discrete time evolution model in signed networks based on Monte Carlo simulation. In the initial time, the model assigns a positive or negative sign to the edge randomly, and at every moment, the triangles are checked in turn to determine whether they are in a state of structural balance. For the unbalanced triangle, one edge is randomly selected, and its sign is changed. In addition to the above model, there are some other signed network dynamics modeling ideas, such as statistical physics modeling methods [18]. Radicchi [17] explores the similarity between unbalanced networks and an unstable spin glass model, and proves the feasibility of using energy spectrum, spin glass and instability [16]. Szell [19] proposed the STC (Signed Triadic Closure) model, which can simulate the dynamic evolution of different types of triangles in large-scale game networks. Malekzadeh [14] proposed a game theory model, which achieved more balanced triangles by setting the vertex adjustment strategy and found the signed network would eventually achieve balance regardless of any initial state. The evolution of dynamic communities in signed networks is still in the early stage of research. The main purposes are to explore the possible evolution model of the real system with signed attributes, to analyze the function of the corresponding mechanism, and to deepen the understanding of this system. Although the existing research has many limitations, they lay a foundation for further exploring the dynamic community and evolution mechanism.

3. Preliminary

3.1. Signed Network Model

Signed network, as shown in Figure 1, is defined as a three-tuple $SN=(V,E,Sign)$. $Sign{:}A\rightarrow\{+1,-1,0\}(A\subseteq V\times V)$ is a signed function. If $\forall v_k,v_s\in V$, if $Sign(v_k,v_s)=+1$, then the link between v_k and v_s is positive. If $Sign(v_k,v_s)=-1$, then the link is negative. In signed networks, $N(v_k)_L=\begin{cases}\{v_i|(v_i,v_k)\in E\} & L=1\\ \{v_i|(v_i,v_j)\in E \text{ where } v_j\in N(v_k)_{L-1} \text{ and } v_i\notin\bigcup_{l=1}^{L-1}N(v_k)_l\} & L\geq 2\end{cases}$ is the L-level neighbor set of v_k.

$CN(v_k,v_s)_{L\cap M}=N(v_k)_L\cap N(v_s)_M=\{v_i|v_i\in N(v_k)_L\cap v_i\in N(v_s)_M\}$ is the insertion set between v_k' L-level neighbor set $N(v_k)_L$ and v_s' M-level neighbor set $N(v_s)_M$, short for the $L\cap M$-common neighbor set. If $L=M$, then $CN(v_k,v_s)_L$ is the L-common neighbor set.

$CN(v_k,v_s)_1$ is a special case for $CN(v_k,v_s)_L$. For any $v_i\in CN(v_k,v_s)_1$, v_i links with v_k and v_s directly. According to the link signs (positive or negative), we can further divide the items in $CN(v_k,v_s)_1$ into two categories. If the signs of (v_i,v_k) and (v_i,v_s) are the same, v_i is an identical neighbor to v_k and v_s. The identical neighbors constitute a set $SCN(v_k,v_s)_1=\{v_i|v_i\in N(v_k)_L\cap v_i\in N(v_s)_M, Sign(v_i,v_k)=Sign(v_i,v_s)\}$. If the signs of (v_i,v_k) and (v_i,v_s) are contrary, v_i is a contrary neighbor to v_k and v_s. The contrary neighbors constitute a set $PCN(v_k,v_s)_1=\{v_i|v_i\in N(v_k)_L\cap v_i\in N(v_s)_M, Sign(v_i,v_k)=-Sign(v_i,v_s)\}$.

3.2. Model of Connection Degree Between Vertices in Signed Networks

Set pair theory points out that any system both contain certain information (identical or contrary) and uncertain information (discrepancy). To some extent, the uncertain information can transform into certain information. Given two objects, they have $N=|S+F+P|$ attributes. Where S is the number of the identical attributes, P is the number of the contrary attributes, and

F is the number of the discrepancy attributes. Thus, in view of the characteristics of signed networks, based on the idea of identical-discrepancy-contrary, we propose the identical-discrepancy-contrary relation between vertices.

Given $SN=(V,E,Sign)$, for any two vertices v_k and v_s, if they share a common neighbor v_i with the same sign, v_i is an identical attribute, that is $S=|SCN(v_k,v_s)_1|$. If the two vertices share a common neighbor v_i with a different sign, v_i is a contrary attribute, that is $P=|PCN(v_k,v_s)_1|$. If the two vertices don't share any common neighbors, the vertices are the discrepancy attributes. For simplicity, we only consider the $1\cap2$, $2\cap1$ and 2-level common neighbors between v_k and v_s, that is $F=|CN(v_k,v_s)_{1\cap 2}|+|CN(v_k,v_s)_{2\cap 1}|+|CN(v_k,v_s)_2|$. Then, the connection degree between vertices is the sum of the ratio of the identical, discrepancy and contrary relation with weight to the total relation, that is $\mu(v_k,v_s)=\frac{(1)_{1\times S}\times(w(v_i))_{S\times 1}}{N}+\frac{(1)_{1\times F}}{N}\times(w(v_i)\times Sign\times i(v_i))_{F\times 1}+\frac{(1)_{1\times P}\times(w(v_i))_{P\times 1}}{N}j$. Where $(1)_{1\times S}$, $(1)_{1\times F}$ and $(1)_{1\times P}$ respectively represent the row vector of the identical, discrepancy and contrary attributes. Each vector value is 1. v_i represents some attribute relation between v_k and v_s. $w(v_i)$ is the weight value and is quantified by vertex degree $d(v_i)$. $Sign\times i(v_i)$ is the different value for v_i, where $Sign$ is the link's positive or negative, and is quantified by the weak balance theory. $i(v_i)$ is the probability of the discrepancy attribute v_i turn to the identical or contrary and is quantified by the clustering coefficient. j=-1 is only an indicator.

3.3. Prediction and Addition of Edges

Based on the idea of connection degree, the discrepancy attributes can be converted into the identical or contrary attributes under certain conditions. $Sign\times i(v_i)$ depicts how the vertex as a discrepancy attribute transforms into the identical or contrary attribute and its possibility. That is, the possible relation between vertices at the next moment is predicted. If the v_i is converted to the identical or contrary attribute of v_k and v_s, then v_i has the link with v_k or v_s, that is, the new edge is added to the network. New methods for predicting and adding new edges are as follows: (1) If $v_i\in CN(v_k,v_s)_{1\cap 2}$ or $v_i\in CN(v_k,v_s)_{2\cap 1}$, and $i(v_i)=\sum_{p\in CN(v_i,v_s)_1} CC(v_q)\big/|CN(v_i,v_s)_1|>0$, then the new edge (v_i,v_s) is added. If $Pos(v_k,v_s)>Neg(v_k,v_s)$, then $Sign(v_k,v_s)$ =+1; else $Sign(v_k,v_s)$=-1. (2) If $v_i\in CN(v_k,v_s)_2$, and $i(v_i)=\sum_{p\in CN(v_i,v_s)_1} CC(v_q)\big/|CN(v_i,v_s)_1|\times\sum_{p\in CN(v_i,v_k)_1} CC(v_p)\big/|CN(v_i,v_k)_1|>0$, then the edge (v_i,v_k) and (v_i,v_s) are added. Similarly, if $Pos(v_k,v_s)>Neg(v_k,v_s)$, then $Sign(v_i,v_s)$=+1; else $Sign(v_i,v_s)$=-1.

4. Dynamic Community Mining of Signed Network

4.1. Relation Definition

Based on the prediction and addition of the edges, the relation between vertices at the next moment has also changed. In this prediction mechanism, when the network size is constant, the network structure will change as time goes on. Therefore, under the static network analysis, when the timeline T is added, the whole network is modelled as a snapshot with T discrete time points, that is $SN_T=\{SN_{(1)}, SN_{(2)},\ldots, SN_{(T)}\}$. $SN_{(1)}$ is the initial network. $SN_{(t)}$ ($t=2,\ldots,T$-1) is the network that adds predictive edges to $SN_{(t-1)}$. $SN_{(T)}$ is the termination network, which is the stable network without predicted edges. Based on the above idea, the connection degree between vertices in time t is given, as shown in Equation (1).

$$\mu(v_k,v_s)_t=\frac{(1)_{1\times|S|}\times(w(v_i)_t)_{|S|\times 1}}{N}+\frac{(1)_{1\times|F|}}{N}(w(v_i)_t\times(Sign\times i(v_i))_t)_{|F|\times 1}+\frac{(1)_{1\times|P|}\times(w(v_i)_t)_{|P|\times 1}}{N}j \tag{1}$$

Definition 1. Given $SN=(V,E,Sign)$, $\forall v_k\in V$, at the t moment, the connection degree of v_k is the mean of connection degree between v_k and the other vertices, that is $\mu(v_k)_t$, as shown in Equation (2).

$$\mu(v_k)_t=\sum_{s=1}^{N}\mu(v_k,v_s)_t\big/N=a(v_k)_t+b(v_k)_t i+c(v_k)_t j \tag{2}$$

Definition 2. Given $SN=(V,E,Sign)$, at the t moment, the connection degree of the whole network is the mean of all vertices' connection degree, that is $\mu(SN)_t$, as shown in Equation (3).

$$\mu(SN)_t=a(SN)_t+b(SN)_t i+c(SN)_t j=\sum_{k=1}^{N}\mu(v_k)_t\big/N=\sum_{k=1}^{N}\sum_{s=1}^{N}\mu(v_k,v_s)_t\big/N\times N \tag{3}$$

Definition 3. Given two communities $C_K=(V_K,E_K,Sign_K)$ and $C_S=(V_S,E_S,Sign_S)$, at the t moment, connection degree between C_K and C_S is denoted as $\mu(C_K,C_S)_t$, as shown in Equation (4), where $v_k\in V_K$ and $v_s\in V_S$.

$$\mu\left(C_K,C_S\right)_t=\begin{cases}\sum_{s=1}^{|V_S|}\sum_{k=1}^{|V_K|}\mu\left(v_k,v_s\right)_t \Big/ |C_K|\times|C_S| & \exists Sign\left(v_k,v_s\right)=1\\ -1 & \forall Sign\left(v_k,v_s\right)=-1 \text{ or } 0\end{cases} \tag{4}$$

4.2. Connection Trend and Entropy of Signed Network and Their Properties

In connection degree $\mu=a+bi+cj$, based on the analysis of a, b and c, as well as the value of i and j, the situation change and the steady state of the whole network can be obtained by connection trend and connection entropy. Given connection degree $\mu=a+bi+cj$, when $c\neq 0$, the ratio a/c is the connection trend in the given problem background, in which a is the identical degree and c is the contrary degree, that is $Shi(H)=a/c$. The connection trend reflects the degree of the identical-discrepancy-contrary relation between two subjects. However, we need to ensure that the denominator is not zero. In order to extend its versatility, we extend a, b and c to the relative identical degree e^a, the relative discrepancy degree e^b and the relative contrary degree e^c, respectively. The related definitions and properties are as follows.

Definition 4. Given the connection degree of signed network $\mu(SN)_t=a(SN)_t+b(SN)_ti+c(SN)_tj$, the connection trend is the ratio of $e^{a(SN)t}$ to $e^{c(SN)t}$, that is $Trend(SN)_t=e^{a(SN)t}/e^{c(SN)t}$. The connection close trend is the ratio of sum of $e^{a(SN)t}$ and $e^{b(SN)t}$ to $e^{c(SN)t}$, that is $CTrend(SN)_t=e^{a(SN)t+b(SN)t}/e^{c(SN)t}$. The connection loose trend is the ratio of $e^{a(SN)t}$ to sum of $e^{b(SN)t}$ and $e^{c(SN)t}$, that is $LTrend(SN)_t=e^{a(SN)t}/e^{b(SN)+c(SN)t}$.

Definition 5. Given connection degree $\mu=a+bi+cj$, when a, b, $c\neq 0$, connection entropy is denoted as S, as shown in Equation (5), where S_S is the identical entropy, S_F is the discrepancy entropy, and S_P is the contrary entropy.

$$S=S_S+S_F+S_P=\sum_{k=1}^{N}a_k ln a_k+i\sum_{k=1}^{N}b_k ln b_k+j\sum_{k=1}^{N}c_k ln c_k \tag{5}$$

The stability of aggregation or fragmentation of signed networks can be measured by the connection entropy. In definition 6, it is necessary to ensure that the logarithm of the index is not zero. In order to extend its versatility, we extend a, b and c to the relative identical degree $(1+a)/2$, the relative discrepancy degree $(1+b)/2$ and the relative contrary degree $(1+c)/2$ respectively.

Definition 6. Given the connection degree of all vertices $\mu(v_k)_t=a(v_k)_t+b(v_k)_ti+c(v_k)_tj$, $(k=2,\dots,N)$, the connection entropy of the signed network is denoted as $S(SN)_t$, as shown in Equation (6).

$$S(SN)_t=\sum_{k=1}^{N}\frac{1+a\left(v_k\right)_t}{-2}ln\frac{1+a\left(v_k\right)_t}{2}+i\sum_{k=1}^{N}\frac{1+b\left(v_k\right)_t}{-2}ln\frac{1+b\left(v_k\right)_t}{2}+j\sum_{k=1}^{N}\frac{1+c\left(v_k\right)_t}{-2}ln\frac{1+c\left(v_k\right)_t}{2} \tag{6}$$

4.3. Dynamic Community Mining Algorithm

For the opposite and unity relation in the signed network and the identical-discrepancy-contrary system of a set pair theory, a new dynamic community mining algorithm DCD for signed networks is proposed. In the process of community clustering, the level of communities is determined by the modularity of signed networks. As the standard to measure the quality of network partition, modularity function Q has been widely used. Algorithm DCD is shown in Algorithm 1, which is mainly divided into two steps. (1) Initialize variables, as shown in line (1). (2) Dynamic community mining, as shown in lines (2)-(8). Firstly, the connection degree between vertices is calculated, and the network connection degree matrix is obtained. Secondly, compute $Trend(SN)_t$, $CTrend(SN)_t$, $LTrend(SN)_t$ and $S(SN)_t$ of the current network. Thirdly, the community of the current network is mined by $Community\ Discovering(A_t,t)$ as shown in Algorithm 2. Then, add a prediction edge to the network. Loop until no new edge joins the network, and then the network evolution is over.

Algorithm 1 DCD

Input: $SN=(V,E,Sign)$, A

Output: $\mu(SN)_t$, $Trend(SN)_t$, $CTrend(SN)_t$, $LTrend(SN)_t$, $S(SN)_t$, $C_t=\{C_{ti}\}$

(1) $t=1$, $A_t=A$, $Q_t=0$, $Flag=False$

(2) *Do*

(3) Calculate $\mu(v_k,v_s)_t$, obtain R_t

(4) Calculate $Trend(SN)_t$, $CTrend(SN)_t$, $LTrend(SN)_t$ and $S(SN)_t$
(5) *Community Discovering*(A_t,t)
(6) Insert to the prediction edges, obtain A_{t+1}
(7) If $A_{t+1}=A_t$ Then $Flag=True$, $t=t+1$
(8) While ($Flag$)

Algorithm 2 is mainly divided into four steps. (1) Initialize the variables and initialize each vertex as a community, as shown in line (1). (2) Merge the independent communities and merge the vertices with the greatest connection degree into a community first, as shown in lines (2)-(8). (3) The rest of the independent vertices in the second step are merged into the community that has the most connections with it, and then the ultimate initial community forms, as shown in lines (9)-(11). (4) The initial communities merge based on the connection degree between communities; small communities merge one by one until all vertices are merged into one community, as shown in lines (12)-(18).

Algorithm 2 *Community Discovering*(A_t,t)

Input: A_t,t
Output: $C_t=\{C_{ti}\}$
(1) $V=\{v_k\}(k=1,\ldots,|V|)$, $VN=|V|$, $C_t=\{C_{ti}\}=\varnothing$, $CN=0$, $UV=\varnothing$, $UVN=0$
(2) **While** ($VN\neq 0$)
(3) Select $Max\{\mu(v_k,v_s)_t\}$
(4) If $\forall v_k\in V$, $\mu(v_k,v_s)_t\geq\mu(v_k,v_i)_t$ and $\mu(v_k,v_s)_t\geq\mu(v_s,v_i)_t$ Then
(5) $C_{new}= v_k\cup v_s$, $C_t=C_t\cup C_{new}$, $VN=VN-1$
(6) Else If $\forall v_i\in C_{ti}$, $\mu(v_k,v_i)_t\geq\mu(v_k,v_s)_t$ Then
(7) $C_{new}=C_{ti}\cup v_k$, $C_t=(C_t-C_{ti})\cup C_{new}$, $VN=VN-1$
(8) Else $UV=UV\cup(v_k,v_s)$, $VN=VN-2$, $UVN=UVN+2$
(9) **For** each v_k in UV **Do**
(10) Select $Max\{\mu(v_k,v_s)_t\}$
(11) If $v_k\in C_{ti}$ Then $C_{ti}=C_{ti}\cup v_k$
(12) $QFlag=True$
(13) **While** ($QFlag$)
(14) Calculate $\mu(C_{ti},C_{tj})_t$
(15) Select $Max\{\mu(C_{ti},C_{tj})_t\}$ Then $C_{new}=C_{ti}\cup C_{tj}$
(16) Calculate Q_{t+1}
(17) If $Q_{t+1}\leq Q_t$ Then $Flag=False$
(18) Else $C_t=(C_t-C_{ti}-C_{tj})\cup C_{new}$, update $\mu(C_{ti},C_{new})_t$

5. Experimental Results and Analysis

The experimental platform is an Intel Core Duo E7500 CPU@1.8GHz with 2GB memory. The algorithm is implemented in Java JDK 1.6, Eclipse 4.3 and Matlab R2008a. The experimental data consist of two real networks and two artificial networks. The real signed networks include the SPP (Slovene Parliamentary Parties) network [7], as shown in Figure 2(a), and GGS (Gahuku-Gama Subtribes) network [7], as shown in Figure 3(a). The artificial signed networks include the illustrative network A [7], as shown in Figure 5(a), and illustrative network B [7], as shown in Figure 6(a).

The main purpose of the experiments is to verify the rationality and correctness of the prediction method based on the set pair in signed network dynamic community mining. In the experiments, adding new edges to the network is divided into 2 cases, $i(v_i)>0$, $i(v_i)\geq Threshold$. Therefore, this section introduces the evolution and dynamic community mining of signed networks in two cases. The correctness of community mining is verified by NMI, and the situation and stability of the network are analysed by set pair theory. The NMI of information theory is an index to evaluate the quality of community partition.

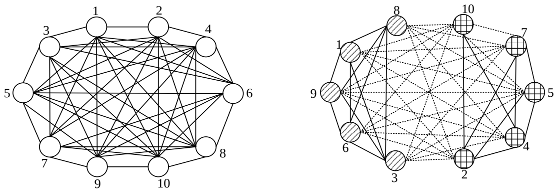
5.1. Dynamic Evolution Analysis when $i(v_i)>0$

- SPP network

In the SPP network, there are 120 triangles, of which 105 are balanced triangles and 116 are weakly balanced triangles. The dynamic community mining information is shown in Table 1. After only 1 round of evolution, the network has reached a stable state, and two communities have been obtained, namely (1, 3, 6, 8, 9) and (2, 4, 5, 7, 10), as shown in Figure 2.

Table 1. Evolution information of SPP network

Network	N/P Edge	All/Balanced/Weakly Triangles	Connection Degree	Connection Trend	Connection Entropy	NMI
$SPP_{(1)}$	18/27	120/105/116	0.20978+0.0i+0.15022j	1.06136	0.30405+0.34657i+0.31811j	1



(a) (b)

Figure 2. (a) $SPP_{(1)}$ (b) Community mining results of $SPP_{(1)}$

We can see from Table 1, at the t_1 moment, $\mu(SPP)_1$=0.20978+0.0i+0.15022j. $a(SPP)_1>c(SPP)_1>b(SPP)_1=0$ shows that the network tends to be in a strong identical trend and will evolve toward a tight convergence. $a(SPP)_1+b(SPP)_1>c(SPP)_1$ and $c(SPP)_1=0.15022\neq 0$ shows that the network belongs to a close non-identical trend, and there exists the opposite relation. $a(SPP)_1>c(SPP)_1+b(SPP)_1$ shows that the network belongs to a loose non-identical trend, and there is also opposite relation. $b(SPP)_1=0$ shows that there is no uncertainty relation in the network, and no new edges are added, which is consistent with the SPP network itself as a complete graph. The SPP network density is 1, the ratio of balance triangle is 87.50%, and the ratio of weak balance triangle is 96.67%. $S_S(SPP)_1=0.30405\approx S_P(SPP)_1=0.31811$ indicates that the network is in the stable state. Because of the existence of the contrary trend and $S_S(SPP)_1:S_P(SPP)_1=7:5$, the network is in a stable state with two completely opposite communities.

- GGS network

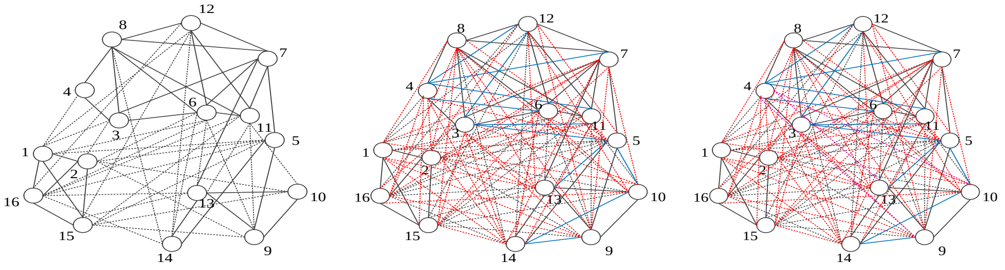
In the GGS network, there are 68 triangles, of which 59 are balance triangles and 66 are weak balance triangles. The dynamic community mining information is shown in Table 2. The network evolution process is shown in Figure 3. The dynamic community mining process is shown in Figure 4. After 3 rounds evolution, the GGS network reaches a stable state and gets three communities, namely (1, 2, 15, 16), (3, 4, 6, 7, 8, 11, 12) and (5, 9, 10, 13, 14).

We can see from Table 2, at the t_1 moment, $\mu(GGS)_1$=0.10319+0.07326i+0.04061j. $a(GGS)_1>b(GGS)_1>c(GGS)_1$ shows that the network tends to be in a weak identical trend and will develop towards a tight convergence. But, the development is slowed. $a(GGS)_1+b(GGS)_1>c(GGS)_1$ and $c(GGS)_1=0.04061$ shows that the network is in a close non-identical trend, and there exists an opposite relation. $a(GGS)_1>b(GGS)_1+c(GGS)_1$ shows that the network belongs to a loose non-identical trend. $b(GGS)_1=0.07326$ shows that there is strong uncertainty and new edges will be added to the network. The GGS network density is 0.48, the ratio of balance triangle is 10.53%, and the ratio of weak balance triangle is 11.79%. It shows that the network is in a very unbalanced state will continue to evolve, which is consistent with the calculation result of the connection degree. In the process of evolution, although the close identical trend is the main trend, the growth of $a(GGS)_t$ is slowed and $c(GGS)_t$ grows rapidly, which shows that the network has strong opposition and tends to split in a certain state. $b(GGS)_t$ continues to decrease until zero, which indicates that all uncertain relations in the network are transformed into certain relations. In the GGS network, the discrepancy mostly turns into the contrary, and it shows that there are strong opposing forces, which leads to the existence of the opposing community in the network.

The GGS network eventually evolves into a complete graph. The ratio of balanced triangle is 72.50%, the ratio of weak balanced triangle is 95.36%, and $S_S(SPP)_t\approx S_P(SPP)_t$, which indicates that the network is in a stable state. Because of the existence of contrary trend and $S_S(SPP)_t:S_P(SPP)_t=13:10$, there exists the antagonistic communities in the network. Meanwhile, the community mining results are the same at different times of the evolution, indicating that the network is stable.

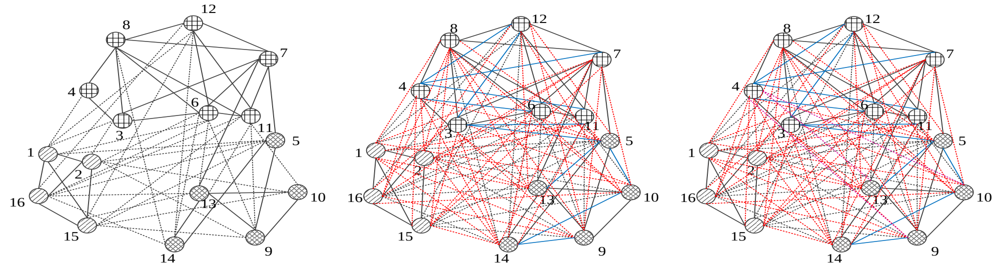
Table 2. Evolution information of GGS network

Network	N/P Edge	All/Balanced/Weakly Triangles	Connection Degree	Connection Trend	Connection Entropy	NMI
$GGS_{(1)}$	29/29	68/59/66	0.10319+0.07326i+0.04061j	1.06459/1.14551/0.98938	0.32813+0.33402i+0.33987j	1
$GGS_{(2)}$	40/77	521/380/496	0.13042+0.00282i+0.09957j	1.03131/1.03423/1.02840	0.32245+0.34614i+0.32884j	1
$GGS_{(3)}$	40/80	560/406/534	0.13145+0.0i+0.10293j	1.02893/1.02893/1.02893	0.32222+0.34657i+0.32818j	1



(a) (b) (c)

Figure 3. (a) $GGS_{(1)}$ (b) $GGS_{(2)}$ (c) $GGS_{(3)}$



(a) (b) (c)

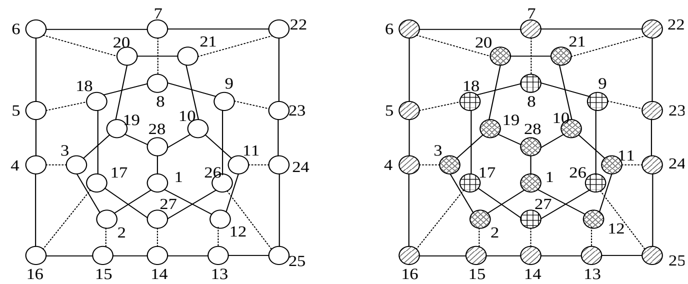
Figure 4. (a) community mining results of $GGS_{(1)}$ (b) community mining results of $GGS_{(2)}$ (c) community mining results of $GGS_{(3)}$

- A network

In the A network, there are 0 triangles, of which there are 0 balance triangles and 0 weak balance triangles. The dynamic community mining information is shown in Table 3. After only 1 round evolution, the A network has reached a stable state, and three communities have been obtained, namely (1, 2, 3, 10, 11, 12, 19, 21, 20, 28), (8, 9, 17, 18, 26, 27) and (4, 5, 6, 7, 13, 14, 15, 16, 22, 23, 24, 25), as shown in Figure 5.

Table 3. Evolution information of A network

Network	N/P Edge	All/Balanced/Weakly Triangles	Connection Degree	Connection Trend	Connection Entropy	NMI
$A_{(1)}$	29/12	0/0/0	0.05002+0.03609i+0.00777j	1.04315/1.08148/1.00618	0.33828+ 0.34072i+ 0.34536j	1



(a) (b)

Figure 5. (a) $A_{(1)}$ (b) community mining results of $A_{(1)}$

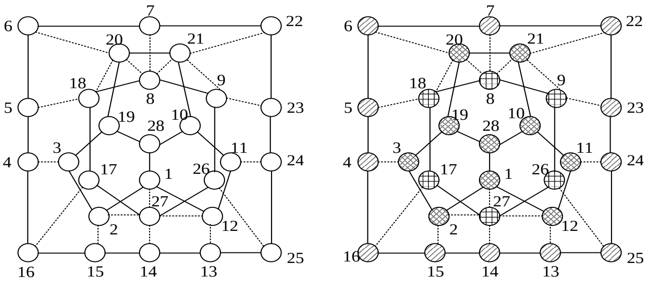
We can see from Table 3, at the t_1 moment, $\mu(A)_1$=0.05002+0.03609i+0.00777j. $a(A)_1>b(A)_1>c(A)_1$ shows that the network tends to be in a weak identical trend and will evolve toward tight convergence. $a(A)_1+b(A)_1>c(A)_1$ and $c(A)_1=0.00777\neq 0$ shows that the network is in a close non-identical trend, and there is a weak opposition. $b(A)_1$=0.03609 shows that there is uncertainty in the network. The A network density is 0.085, the ratio of balance triangle is 0%, and the ratio of weak balance triangle is 0%, which shows that the network is a sparse unbalance network. Since no triangles exist in the network, the clustering coefficient of each vertex is 0. The uncertainty of the network does not change, so the edges are not added and changed and it is relatively stable. Because of the existence of contrary trend and $S_S(A)_1{:}S_P(A)_1$=50:7, the network is in the stable state of the three opposing communities.

- B network

In the B network, there are 4 triangles, of which 4 are balance triangles and the 4 are weak balance triangles. The dynamic community mining information is shown in Table 4. After 4 rounds evolution, the B network reaches a stable state and gets three communities, namely (1, 2, 3, 10, 11, 12, 19, 21, 20, 28), (8, 9, 17, 18, 26, 27) and (4, 5, 6, 7, 13, 14, 15, 16, 22, 23, 24, 25), as shown in Figure 6.

Table 4. Evolution information of B network

Network	N/P Edge	All/Balanced/Weakly Triangles	Connection Degree	Connection Trend	Connection Entropy	NMI
$B_{(1)}$	29/19	5/5/5	0.04837+0.04092i+0.00947j	1.03967/1.08309/0.99798	0.33857+0.33988i+0.34509j	1
$B_{(2)}$	42/65	150/111/150	0.04916+0.05074i+0.01796j	1.03169/1.08539/0.98065	0.33843+0.33814i+0.34372j	1
$B_{(3)}$	76/198	1417/922/1380	0.06352+0.02275i+0.04581j	1.01787/1.04129/0.99497	0.33583+0.34295i+0.33899j	1
$B_{(4)}$	95/283	3276/1930/3161	0.07343+0.0i+0.06432j	1.00915/1.00915/1.00915	0.33399+0.34657i+0.33569j	1



(a) (b)

Figure 6. (a) $B_{(1)}$ (b) dynamic community mining results of $B_{(4)}$

We can see from Table 4, at the t_1 moment, $\mu(A)_1$=0.04837+0.04092i+0.00947j. $a(B)_1>b(B)_1>c(B)_1$ shows that the network tends to be in a weak identical trend and will develop to a tight convergence but the development is slow. $a(B)_1+b(B)_1>c(B)_1$ and $c(B)_1=0.00947\neq0$ shows that the network is in a close non-identical trend, and there exists an opposite relation. $a(B)_1>b(B)_1+c(B)_1$ shows that the network belongs to a loose non-identical trend. $b(B)_1$=0.04092 shows that there is a strong uncertainty and new edges will be added to the network. The B network density is 0.13, the ratio of balance triangle is 0.15%, and the ratio of weak balance triangle is 0.15%, which indicates that the network is in a very unbalanced state and the network will continue to evolve, which is consistent with the calculation result of the connection degree. In the process of evolution, although the close identical trend is the main trend, the growth of $a(B)_t$ is slow and the $c(B)_t$ is rapid, which shows that the network has a strong opposition and tends to split in a certain state. $b(B)_t$ continues to decrease until zero, indicating that all uncertain relations in the network are transformed into certain relations. In the B network, the discrepancy mostly turns into the contrary, and it also shows that there are strong opposing forces, which leads to the existence of the opposing community in the network.

The B network eventually evolves into a complete graph, the ratio of balance triangle is 58.91%, the ratio of weak balance triangle is 96.49%, and $S_S(B)_t \approx S_P(B)_t$, indicating that the network is in a stable state. Because of the existence of contrary trend and $S_S(B)_t$:$S_P(B)_t$=7:6, there exists the antagonistic communities in the network. Meanwhile, the community mining results are the same at different times in the evolution, indicating that the network is stable in three communities.

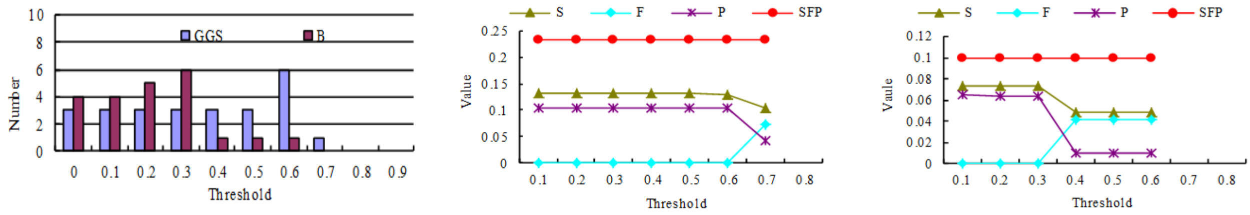
Through the study of dynamic communities in four signed networks, the conclusions are as follows. (1) The network structure is obvious (positive edges within communities is density, and negative edges between communities is density), which makes the community structure relatively stable during the evolution of the network. (2) The density of the network is large, and the network evolution speed is fast. (3) The uncertainty of the network is small and the network evolution speed is fast. (4) There is no triangle relation in the network, which hinders the evolution of the network to a certain extent.

5.2. Dynamic Evolution Analysis when $i(v_i)\geq Threshold$

The number of network evolution is between 1 to 4 rounds, and soon no new edges are added and the network reaches the stable state. The analysis shows that the speed of the network evolution is determined by the speed of the addition of new edges, while the speed of new edges increase is determined by $i(v_i)$. In the evolution process, when $i(v_i)$, new edges are added. When the sign is not considered, $i(v_i)\in[0,1]$. What impact does $i(v_i)$ have on the evolution of the network? Therefore, the threshold is set. If $i(v_i)\geq Threshold$, then the new edges are added, otherwise they are not added. Then, the influence of $i(v_i)$ on network evolution is analysed from two aspects. (1) The influence on the evolution number and (2) The influence on the evolution result.

With the increase of *Threshold*, network evolution number shows normal distribution. That is, evolution number increases first, and then decreases, as shown in Figure 7(a). When $Threshold\approx0$, there is little impact on the evolution. When the *Threshold* value is too large, the network evolution number decreases or does not evolve. For example, when *Threshold*=0.6, the GGS and B networks will evolve. When *Threshold*=0.7, only the GGS network will evolve. When *Threshold*≥0.8, the networks will not evolve. Thus, the $i(v_i)$ is between 0.2 and 0.6, and the evolution number of the network is appropriate.

The connection degree of GGS and B network in evolution termination time is shown in Figure 7(b)-(c). When the network can evolve into a complete graph, the discrepancy is 0. Then, the change of *Threshold* will not affect the network connection degree. That is, the ultimate evolution state of the network remains the same. When the network cannot evolve into a complete graph, with an increase of *Threshold*, the discrepancy increases, the identical and contrary decreases, and the network connection degree decrease. For a network with obvious community structure that cannot evolve into a complete graph, although the evolution state changes, the final network connection degree tends to be consistent.



(a) (b) (c)

Figure 7. (a) Number of network evolution under different thresholds (b) GGS Situation changes of identical-discrepancy-contrary value under different threshold (c) B Situation changes of identical-discrepancy-contrary value under different threshold

The analysis shows that network size, density and the balance triangle number have a direct influence on the evolution of the network. Meanwhile, as the network evolves, $i(v_i)$ changes, so it is not reasonable to set a fixed threshold. We set the mean of all vertices $i(v_i)$ as the threshold. Since the SPP network is a complete graph, and the $i(v_i)$ is 0 in the A network, the two networks have no evolution process. Next, the main analysis is the dynamic network evolution process of GGS and B when $i(v_i) \geq \overline{i(v_i)}$ .To sum up, when the *Threshold* is $\overline{i(v_i)}$, the evolution speed slows down. Because the *Threshold* is greater than zero, some networks cannot evolve into a complete graph, but eventually, the network can reach a stable state.

6. Conclusions

Based on the set pair theory, this paper studies the dynamic evolution and community mining in signed networks. Firstly, the signed network prediction model is presented based on the possibility that uncertainty between vertices is transformed into a certainty. Secondly, the time axis is added to the network and the dynamic model of the set pair signed network is given. Finally, the performance of network evolution and dynamic community mining is investigated by setting different thresholds. The experimental results show that the evolution law of network connection degree is consistent with the actual network evolution, and the stability of the community mining result is proven by evolution. The next research goal is to study the dynamic evolution and dynamic community mining method, mine more potential relation in the network, and better predict the developmental trend of the network.

Acknowledgements

This work is supported by the National Youth Science Foundation of Hebei (No. F2017209070), the National Science Foundation of China, (No. 61472340, No. 61303017), the National Youth Science Foundation of China (No. 61602401), and the Natural Science Foundation of Hebei Province (No. F2014210068).

References

1. T. Antal, P. L. Krapivsky and S. Redner, "Social Balance on Networks: the Dynamics of Friendship and Enmity," *Physical D: Nonlinear Phenomena*, vol. 224, no. 2, pp. 130-136, February 2006
2. M. Brusco, P. Doreian, A. Mrvar and D. Steinley, "Two Algorithms for Relaxed Structural Balance Partitioning: Linking Theory, Models and Data to Understand Social Network Phenomena," *Sociological Methods and Research*, vol. 40, no. 1, pp. 57-87, January 2011
3. S. Banerjee, K. Sarkar, S. Gokalp, A. Sen and H. Davulcu, "Partitioning Signed Bipartite Graphs for Classification of Individuals and Organizations," *Social Computing, Behavioral-Cultural Modeling and Prediction*, pp. 196-204, Berlin: 2012
4. D. Cartwright and F. Harary, "Structural Balance: A Generalization of Heider's Theory," *Psychological Review*, vol. 63, no. 5, pp. 277-293, May 1956

5. S. Q. Cheng, H. W. Shen, G. Q. Zhang and X. Q. Cheng, "Survey of Signed Network Research," *Journal of Software*, vol. 25, no. 1, pp. 1-15, January 2014
6. P. Erdos and A. Renyi, "On Random Graphs," *Publications Mathematical*, no. 6, pp. 290-297, June 1959
7. G. Facchetti, G. Iacono, and C. Altafini, "Computing Global Structural Balance in Large-Scale Signed Social Networks," in *Proceedings of the 1th Conference on National Academy of Science*, vol. 108, no. 52, pp. 20953-20958, University of Rome, Italy, October 2011
8. P. Gawroński, P. Gronek and K. Kulakowski, "The Heider Balance and Social Distance". Actual Physical Polemical B, vol. 36, no. 8, pp. 2549-2558, August 2005
9. F. Heider, "Attitudes and Cognitive Organization," *Journal of Psychology*, vol. 21, no. 1, pp. 107-112, January 1946
10. J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. deLuca, and S. Albayrak, "Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization," in *Proceedings of the 1th SIAM Conference on Data Mining*, pp. 559-570, Philadelphia: SIAM, April 2010
11. K. Klemm and V. M. Egufluz, "Growing Scale-Free Networks with Small-World Behavior," *Physical Review E*, vol. 65, no. 5, pp. 057102, May 2002
12. K. Klemm and V. M. Egufluz, "Highly Clustered Scale-Free Networks," *Physical Review*, vol. 65, no. 3, pp. 036123, March 2002
13. J. Leskovec, D. Huttenlocher and J. Kleinberg, "Signed Networks in Social Media," in *Proceedings of the 10th SIGCHI Conference on Human Factors in Computing Systems*, *pp.* 1361-1370, New York, USA, October 2010
14. M. Malekzadeh, M. A. Fazli, P. J. Khalidabadi, M. Safariy and H. R. Rabieey, "Social Balance and Signed Network Formation Games," in *Proceedings of the 5th Workshop on Social Network Mining and Analysis*, New York, USA, 2011
15. S. A. Marvel, J. Kleinberg, R. D. Kleinberg and S. H. Strogatz, "Continuous-Time Model of Structural Balance," in *Proceedings of the 1th Conference on National Academy of Science*, pp. 1771-1776 United States of America, October 2011
16. F. Radicchi, D. Vilone and H. Meyer-Ortmanns, "Universality Class of Triad Dynamics on a Triangular Lattice," *Physical Review E*, vol. 75, no. 2, pp. 021118, February 2007
17. F. Radicchi, D. Vilone, S. Yoon and H. Meyer-Ortmanns, "Social Balance as a Satiability Problem of Computer Science," *Physical Review E*, vol. 75, no. 2, pp. 026106, February 2007
18. S. Redner, "Social Balance on Networks: the Dynamics of Friendship and Hatred," *Physical D Nonlinear Phenomena*, vol. 224, no. 1-2, pp. 130-136, January 2006
19. M. Szell, R. Lambiotte and S. Thurner, "Multirelational Organization of Large-Scale Social Networks in an Online World," in *Proceedings of the 1th Conference on National Academy of Sciences*, vol. 107, no. 3, pp. 13636-13641, USA, July 2010
20. L. Wang and X. Q. Cheng, "Dynamic Community in Online Social Networks," *Chinese Journal Of Computers*, vol. 38, no. 2, pp. 219-237, February 2015
21. Z. G. Wang and W. Thorngate, "Sentiment and Social Mitosis: Implications of Heider's Balance Theory," *Journal of Artificial Societies and Social Simulation*, vol. 6, no. 3, pp. 26-45, March 2003
22. D. J. Watts and S. H. Strogatz, "Collective Dynamics of Small World Networks," *Nature*, no.393, pp. 440-442, June 1998
23. B. Yang, W. K. Cheung and J. M. Liu, "Community Mining from Signed Social Networks," *IEEE Trans. on Knowledge and Data Engineering*, vol. 19, no. 10, pp. 1333-1348, October 2007

Xiao Chen graduated from the College of Information Science and Engineering, YanShan University, China, with her of Master's and Ph. D. She is a lecturer at Hebei Normal University of Science and Technology. Her research interests include graph mining, social network analysis, etc.

XinZhuan Hu graduated from the College of Information Science and Engineering, YanShan University, China, with her Bachelor's, Master's and Ph. D. She is an associate professor at YanShan University. Her research interests include social network analysis, etc.

Xiao Pan is an associate professor at Shijiazhuang Tiedao University, China. She was a visiting scholar in the Department of Computer Science, University of Illinois at Chicago, USA. She received her Ph.D. in Computer Science from Renmin University of China in 2010. Her research interests include data management on moving objects, location based social networks, and privacy-aware computing.

Jingfeng Guo graduated from the College of Information Science and Engineering, YanShan University, China, with his Bachelor's, Master's and Ph. D. He is a professor and Ph.D. supervisor at YanShan University. His research interests include database theory and application, data mining, social network analysis, etc.