International Journal of Innovative Computing, Information and Control
Volume **10**, Number **6**, December **2014**

ICIC International ©2014 ISSN 1349-4198
pp. 2221–2235

RESEARCH OF COMMUNITY MINING IN SIGNED SOCIAL NETWORK BASED ON GAME THEORY

Xiao Chen[1,2], Xinrun Du[1], Junli Yu[1], Jingfeng Guo[2], Zhaoxia Cui[1] and Liqun Fu[1]

[1]Qian'an College
Hebei United University
No. 5096, Yanshan Road, Qian'an 064000, P. R. China
chenxiao0604@163.com

[2]College of Information Science and Engineering
Yanshan University
No. 438, Hebei Avenue, Qinhuangdao 066004, P. R. China
jfguo@ysu.edu.cn

Received January 2014; revised May 2014

Abstract. *According to the characteristics of signed social networks with the attributes of links and nodes' signed, this paper proposed a new effective method of signed network community partition (SNP) based on game theory. Firstly, the transition probability matrix P is established based on the impact of signed attribute (positive/negative link) on the community structure, and the tightness between the node matrix R^L within L-step is calculated through the P, based on the random walk. Secondly, the community center nodes are selected, taking into account the signs and the density properties. Then, the community members with the central nodes are determined based on the maximum average value of the community. Finally, the correctness and effectiveness of the algorithm are checked through the experiments.*
Keywords: Signed social network, Random walk, Game theory, Community mining

1. **Introduction.** The new social network such as Blog provides a good platform for people with common interests and people with same hobbies. On the platform, users can publish any content, and other users can also comment on it. Those comments basically include two types: supportive and opposing. When modeling the community structures of this category, if we model positive links for having same idea and negative links for having different idea, then the networks with both positive and negative links are called signed social networks (short for SSN) [1], as shown in Figure 1. In the figures, solid lines denote positive links, and dashed lines denote negative links. These new social networks connect the real society with the virtual network space, bringing together a large number of human society digital footprints, such as relation, behavior and some other perception which can be calculated. Thus, the research on the social network relationship has become a popular topic.

In traditional social network, most of mining communities' techniques focus merely on topological structure, which is the community of nodes within which the links are dense but between which the links are sparse. However, the signed social network communities not only have the structure relationship between individuals (topological structure), but also have the individual behavior attributes (evaluation attitude, etc.), so applying the existing techniques to signed networks is not feasible.

However, in SSN, the community structure can be illustrated by graph theory method, and behavior attributes can be solved by establishing game theory model. The game theory model should have the following three basic factors. 1) No less than two game players. 2) Each player has a set of strategies. 3) Each player receives a payoff for each choice of strategies, or social optimality. In SSN, the node can be considered to be the player of the game; the links between nodes can be regarded as the cooperation or competition relationship between the players; individual payoff is the community contributions after each node joins a community. Finally, the overall income maximizes as a measure of partitioned communities. Therefore, we can study the community game segmentation and formation theory to replace the community mining in SSN.

According to the characteristics of SSN, this paper proposes a new effective method of signed network community partition based on game theory, which considers the attributes of links and nodes' signed. Firstly, transition probability matrix P is established through adjacency matrix A of signed networks. Secondly, based on the random walk method, L-step within the tightness between the node matrix R^L is calculated through P. Thirdly, the initial center node of community partition is determined by the center influence of node, arranging the node tightness in descending order. Fourthly, based on game theory, community is partitioned by calculating the maximum value of community. Finally, check the correctness and effectiveness of the algorithm through experiments.

2. Problem Statement and Relevance.

2.1. Problem dentition and motivation.

At present, the SSN is mainly divided into two classes. The first class: an SSN can be partitioned into two or more communities, which only include positive links within the same communities and negative links between the different communities. The second class: an SSN can be partitioned into two or more communities, where both the positive links within same communities and negative links between the different communities are dense, and where both the negative links within same communities and positive links between the different communities are sparse. As can be shown in Figure 1, the network can be divided into two communities including (6, 7, 22, 23, 24, 25, 13, 14, 15, 16, 4, 5) and (8, 9, 26, 27, 17, 18, 20, 21, 10, 11, 12, 1, 2, 3, 19, 28). Figure 1(a) belongs to the first class SSN, Figure 1(b) belongs to the second one.

Intuitively, identifying communities in the first class SSN can easily be achieved by cutting all the negative links, which is an ideal situation. In fact, most of the existing social networks are belonging to the second class; most natural partition cannot be arrived at only by cutting negative links, but many signed networks often result in several large communities accompanied with a few isolated nodes after all negative links are cut out,

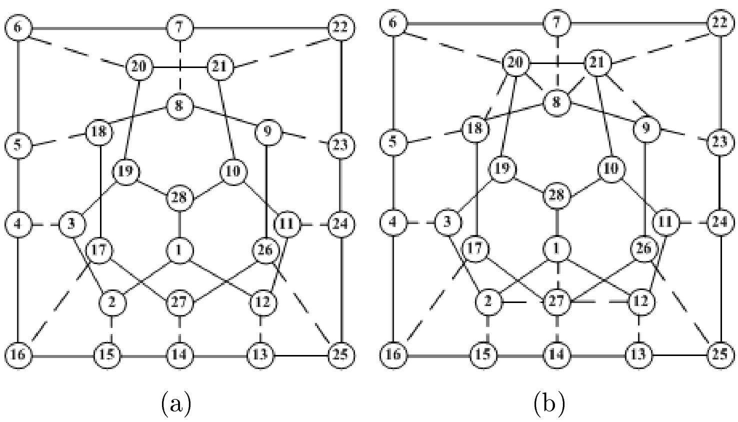


FIGURE 1. Signed social network

which have positive and negative links in the communities. Therefore, the main research work is mining communities in the second class SSN.

SSN is a two-tuple $SSN = (G, \sigma)$, where $G = (V, A)$ is a graph of undirected and unloop; V is a set of nodes; adjacency matrix $A \subseteq V * V$ is a set of edge; $\sigma : A \to \{P, N\}$ is a signed function, where P is positive links and N is negative link, that is $\sigma : A \to \{+1, -1\}$. For any v_i, $v_j \in V$, adjacency matrix A is shown in Formula (1), where $|V|$ is the number of nodes.

$$A = (a_{ij})_{|V|*|V|} = \begin{cases} 1 & (v_i, v_j) \in P \\ -1 & (v_i, v_j) \in N \\ 0 & (v_i, v_j) \notin P \text{ and } (v_i, v_j) \notin N \end{cases} \tag{1}$$

An SSN is partitioned into K disjoint communities $G_i = (V_i, A_i)$ $(i = 1, \ldots, K)$, for any i, $j \in K$ $(i \neq j)$, $V = \cup_{i=1}^{K} V_i$, and $V_i \cap V_j = \phi$. Figure 1 corresponds to the community matrix shown in Figure 2. In Figure 2, in the first row and the first listed as

V	1	12	2	28	11	3	19	10	21	20	27	26	18	17	9	8	7	6	5	4	16	15	14	13	25	24	23	22
1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0
2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0
28	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0
3	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0
19	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	-1	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	-1	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	-1	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	-1	0
8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	-1	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	1	0	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
4	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	1	0	1	0	0	0	0	0	0
15	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
13	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
24	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	1
22	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0

(a)

V	28	19	10	1	3	11	2	12	20	21	17	26	18	9	8	27	7	14	23	5	25	16	22	6	13	15	24	4
28	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1
11	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0
2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0
12	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	-1	0	0	0
20	0	1	0	0	0	0	0	0	0	1	0	0	-1	0	-1	0	0	0	0	0	0	0	0	-1	0	0	0	0
21	0	0	1	0	0	0	0	0	1	0	0	0	0	-1	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	-1	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	-1	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	-1	0	1	0	0	0	1	0	0	0	0	-1	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	-1	0	1	0	0	1	0	0	0	-1	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	-1	-1	0	0	1	1	0	0	-1	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	-1	0	0	-1	-1	0	0	1	1	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1	1	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	1	1	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	1	0	0	0	1	0
5	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
25	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
16	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
22	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
15	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
24	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
4	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0

(b)

FIGURE 2. Matrix communities graph

node identifier, the cross location of rows and lists is the edge value of nodes' pairs, the shadow of disjoint section shows two different communities, and shading said the links relation of nodes within community. There only are positive links within the communities and negative links between the different communities in Figure 2(a); there are both the densely positive links and sparsely negative links within community in Figure 2(b).

An SSN should achieve a good balance between the following two properties. 1) Nodes within one community are close to each other in terms of structure, while nodes between communities are distant from each other. 2) Nodes within one community have dense positive links and sparse negative links, while nodes between communities could have sparse positive links and dense negative links. Therefore, this paper mainly deals with the following there works in communities' partitions in SSN. 1) Calculation of tightness between nodes. 2) Choice of initial center node in community. 3) Partitioned communities based on game theory.

The main contributions of this paper are as follows.

1) Establish transition probability matrix based on the structure and attributes of signs. Because the random walk model can accurately describe the connectivity between nodes in the networks, you need to construct the P between the nodes in the networks in order to apply the random walk model into the community partition. To construct the P in the traditional social network, one only needs to consider connection between nodes; however, in the signed social networks, there are signed attributes between the nodes. Therefore, this paper will combine the connectivity between nodes with the signed attributes in the network to propose a new method to establish the transition probability matrix.

2) Function of the center node influence of community. When partitioning a community, selecting different nodes can get different community distributions, especially when the selected nodes are bridge nodes or edge nodes, the effects of the community partition will be greatly affected. In the traditional social network, the probability of the higher degree nodes becoming the center of the community will be relative higher. Similarly, this paper puts forward a new function of the center node influence as a selection basis for the community center nodes, taking into account both the signs and the density properties.

3) Community partition based on game theory. Based on game theory, the node can be taken as the game participant, the community selection as a strategic choice, then calculate earnings based on the tightness between nodes to make sure that the entire community average income can be max. So we propose a new method of partitioning the community.

2.2. **Practical applications.** There are opposite relationships in many real networks, especially in social, biological and information fields, such as Facebook, Foodspotting and Fitocracy. Using the sign properties of edges to analyze and predict can not only analyze the properties of the users which to be classified, but also can provide personalized recommendations for the users, and help the companies to advertise the products, so as to get better social profits. Research on signed social network based on game theory, also has a huge research space and a wide range of applications in the field of participants behavior, the statistical behavior, game action and collective knowledge mining, etc. For example, in the electronic commerce activity, analysis of both marketing and resisting the strategy behavior, sale, paid search advertising market bit allocation and pricing, as well as the advertising market equilibrium, reflects the value of using the network to customer. It can also solve the optimal auction and optimal equilibrium in the sale agent market; solve the allocation problem of network bandwidth in peer-to-peer systems; solve the game model in limited budget under the conditions of social network formation, in which specific to the problem that each participant wants to own the center position maximization, etc.

3. Related Work.

3.1. **Community partition.** In non-signed social networks, many algorithms have been developed to detect positive network communities. They can generally be divided into three main categories: 1) graph theoretic methods [1], 2) hierarchical methods [2], 3) methods for detecting hyperlink-based Web communities [3]. Next, the details of the algorithm in this article will be introduced in the following. In 2007, Yang et al. put forward a kind of algorithm through the random walk to find the label in the network community, which is called FEC [4]. The algorithm first starts the random walk from the node whose community not being signed; examining migration steps can reach the node set and corresponding probability, ignore all the negative edges, in the process of migration and then based on a truncation function determine which initial nodes and the beginning node of the random walk belong to the same community. FEC in time and precision of recognition shows the good performance, however, some uncertain factors of this algorithm may lead to its nonstable performance, such as the choice of initial seed node, set of random walk step. Aiming at these deficiencies, Sharma et al., put forward the CRA [5] algorithm in 2009. The algorithm can be divided into two steps: 1) ignore the negative connection, based on BFS for clustering is connected; 2) considering the negative connection of vertex to the community by clustering. Using the greedy algorithm, this method can be used in the negative connection between communities, among which there are dense and unlabeled networks connected together. The present signed network partition method either directly improves traditional community division metric function, divides the two-step processing; all not directly consider the effect of positive and negative edge weights.

3.2. **Random walk.** Since the 1980s, the random walk model has been widely used in the field of data mining, and the Internet. A large number of researchers, including top graph expert Lovász, etc., all studied the connection relation between random walk and graph. They theoretically, deeply studied the average first arrival time of the random walk, the average round-trip time, average coverage time, eigenvalue and such kind of problems. Lovász gave a reasonable theoretical research review about figure on the random walk [6]. The random walk based on graph model has been widely developed in various fields. For example, in terms of complex networks, random walk has also been used in analysis of protein network [7] and social network [8], etc.

3.3. **Game theory.** Game theory mainly studies the behavior of decision-making body direct interaction that occurs in the decision-making and the balance of this decision-making problem. John Nash and L. Shapley made a fundamental contribution to game theory. At present, the game theory is mainly divided into non-cooperative game and cooperative game two aspects. Among them, the non-cooperative game mainly participates to study how to make decisions in the game, and cooperative game can research participants in a binding agreement under the desired profit distribution. Game theory as an important analysis tool, is not only widely used in economics and political science, but also used in the transportation network and communication network. In 2010, Moretti et al.'s measurement was used in biological networks [9]. In 2013, Zhou et al. and other people applied league game to proceed community classification [10].

4. Methodology.

4.1. Calculation of tightness between nodes.

In unsigned social network, the tightness degree between the nodes mainly depends on the distance between nodes and the communication of their structures. If node v_i and node v_j are directly connected or connected through multiple paths, then v_i can get to v_j easily, in other words, v_i and v_j are in close proximity. On the other hand, if there is little or no path connected between the nodes, then v_i is far away from v_j, and the probability of them being in the same community is very small. Based on this feature, there are many ways to calculate the distance between the nodes; this paper adopts the random walk model, to calculate the tightness between nodes, combining the connectivity between nodes with the signed nodes in the SSN. In order to accurately describe the tightness between nodes, we will introduce the relevant definitions below.

Definition 4.1. *In $SSN = (G, \sigma)$, the degree $D(v_i)$ for any $v_i \in V$ is the number of edges linked to v_i, that is the sum of the ith rows value in adjacency matrix A, as shown in Formula (2), where $DP(v_i)$ is the number of positive links and $DN(v_i)$ is the number of negative links.*

$$D(v_i) = DP(v_i) + DN(v_i) = \sum_{j=1}^{|V|} |a_{ij}| \tag{2}$$

Definition 4.2. *Set random processes $\{X_{(t)}, t \in T\}$, where time $T = \{0, 1, \ldots, n, \ldots\}$, state-space $I = \{0, 1, 2 \ldots\}$, for any time n and any state $i_0, i_1, \ldots, i_{n-1}, i, j$, if existing Formula (3), then $\{X_{(t)}, t \in T\}$ is Markov Chain, that is $\{X_n, n \geq 0\}$.*

$$P\{X_{(n+1)} = j | X_{(n)} = i, X_{(n-1)} = i_{n-1}, \ldots, X_{(1)} = i_1, X_{(0)} = i_0\} = P\{X_{(n+1)} = j | X_{(n)} = i\} \tag{3}$$

Definition 4.3. *One-step transition probability is conditional probability $p_{ij}(n)$ of transitioning from one state i to another state j in a step in time index n, as shown in Formula (4).*

$$p_{ij}(n) = P\{X_{(n+1)} = j | X_{(n)} = i\} \tag{4}$$

Definition 4.4. *The t-step transition probability is the probability $p_{ij}^{(t)}$ of transitioning from state i to state j in t steps, as shown in Formula (5), where $s \geq 0$, $t \geq 1$.*

$$p_{ij}^{(t)} = P\{X_{s+t} = j | X_s = i\} \tag{5}$$

Definition 4.5. *A one-dimensional random walk can also be looked at as a markov chain. In SSN, if existing at least one path L between nodes v_i and v_j, then start from the node v_i, randomly select a neighbor node with probability P_{it}, and repeat the process until reaching the end v_j. Model created in this way is called Markov random walk model coming from v_i to v_j in L-steps. The value R_{ij}^L of markov random walk model is the sum of probability coming from v_i to v_j within L-steps, as shown in Formula (5). t is the reached node v_t in random walk $L - 1$ steps, R_{it}^L is the probability of v_i reaching the node through $L - 1$, and P_{tj} is the probability of v_t reaching v_j by one step.*

$$R_{ij}^L = \sum R_{it}^{L-1} P_{tj} \tag{6}$$

To apply the random walk model, you need to calculate the P according to the A of the SSN. The P has to ensure that to any node v_i in the network, of all the neighbor nodes of v_i, the further a node is to the v_i, the lower the possibility of walking to this node. Therefore, for an undirected, unweighted social network G, for any v_i, $v_j \in V$, to transfer one step from v_i to v_j, the probability can be simply denoted as $p_{ij} = {}^1/_{D(v_i)}$, that is to say, the probability of departing from v_i to any neighbor nodes is same. Similarly, in the SSN, a pair of the nodes having a negative connectivity may belong to the same community, thus, to walk to the negative signed edge from v_i may have the same possibility $p_{ij} = {}^1/_{D(v_i)}$.

However, in a desired community division there is only the positive connection in the same community. In order to reduce the influence of the negative signed attributes on the community construction, one can reduce the transition probability of pairs of the negative signed attribute nodes, such as reducing the reciprocal of the number of the negative side ($\frac{-1}{DN(v_i)}$). There are only two simplest calculating methods: one is $\frac{1}{D(v_i)} + \frac{-1}{DN(v_i)}$, and the other is $\frac{1}{D(v_i)} * \frac{-1}{DN(v_i)}$, as shown in Table 1. To the p_{ij} of different negative side numbers, the p_{ij} comes out of the first calculating method sometimes increases and sometimes decreases; however, the second method can not only ensure the $p_{ij} \in ({}^{-1}/_{D(v_i)}, 0)$, and when the degree of the nodes is the same, the p_{ij} can decrease as the number of negative sides increase. Therefore, this paper adopts the $\frac{1}{D(v_i)} * \frac{-1}{DN(v_i)}$ as the method to calculate the P.

Definition 4.6. *In SSN, the A corresponds to the P, as shown in Formula (7).*

$$P = (p_{ij})_{|V|*|V|} = \begin{cases} \frac{1}{D(v_i)} & a_{ij} = +1 \\ \frac{1}{D(v_i)} * \frac{-1}{DN(v_i)} & a_{ij} = -1 \\ 0 & a_{ij} = 0 \end{cases} \tag{7}$$

In SSN, considering the impact of the negative connectivity on the community construction, in order to reduce the transfer between the pairs of negative nodes, allowing the probability as a negative one, so the P has the following properties.

Property 4.1. *If $(v_i, v_j) \in P$, then $p_{ij} > 0$; if $(v_i, v_j) \in N$, then ${}^{-1}/_{DN(v_i)} \le p_{ij} < 0$; else $p_{ij} = 0$.*

Property 4.2. *For any $v_i \in V$, if $(v_i, v_j) \in P$, then $\sum_{j=1}^{|V|} p_{ij} = 1$; if $(v_i, v_j) \in N$, then $\sum_{j=1}^{|V|} p_{ij} = -1$; else $\sum_{j=1}^{|V|} p_{ij} = \frac{1}{D(v_i)}(DP(v_i) - DN(v_i))$.*

Definition 4.7. *In SSN, from any node v_i random walk to v_j in L-steps, tightness between node v_i and v_j as show in Formula (8). Where l is path length, R^l_{ij} is the transition*

Table 1. Comparison of the two calculating effects of transition probability

v_i	Positive Links	Negative Links	Degree	Transition Probability
$\frac{1}{D(v_i)} + \frac{-1}{DN(v_i)}$	2	1	3	$p_{ij} = \frac{1}{3} + \frac{-1}{1} = -\frac{2}{3} < -\frac{1}{3}$
	2	2	4	$p_{ij} = \frac{1}{4} + \frac{-1}{2} = -\frac{1}{4} = -\frac{1}{4}$
	2	3	5	$p_{ij} = \frac{1}{5} + \frac{-1}{3} = -\frac{2}{15} > -\frac{1}{5}$
$\frac{1}{D(v_i)} * \frac{-1}{DN(v_i)}$	2	1	3	$p_{ij} = \frac{1}{3} * \frac{-1}{1} = -\frac{1}{3} = -\frac{1}{3}$
	2	2	4	$p_{ij} = \frac{1}{4} * \frac{-1}{2} = -\frac{1}{8} > -\frac{1}{4}$
	2	3	5	$p_{ij} = \frac{1}{5} * \frac{-1}{3} = -\frac{1}{15} > -\frac{1}{5}$
	3	2	5	$p_{ij} = \frac{1}{5} * \frac{-1}{2} = -\frac{1}{10} > -\frac{1}{5}$
	4	1	5	$p_{ij} = \frac{1}{5} * \frac{-1}{1} = -\frac{1}{5} = -\frac{1}{5}$

probability. For any node, distance from themselves to 1.

$$d(v_i, v_j) = \begin{cases} \sum_{l=1}^{L} R_{ij}^l & i \neq j \\ 1 & i = j \end{cases} \tag{8}$$

The node tightness matrix: $R = (r_{ij})_{|V|*|V|} = (d(v_i, v_j))_{|V|*|V|}$.

The node tightness matrix in step L:

$$R^L = P * R^{L-1} - diag(diag(P * R^{L-1})) + I \tag{9}$$

The nodes tightness has the following related property.

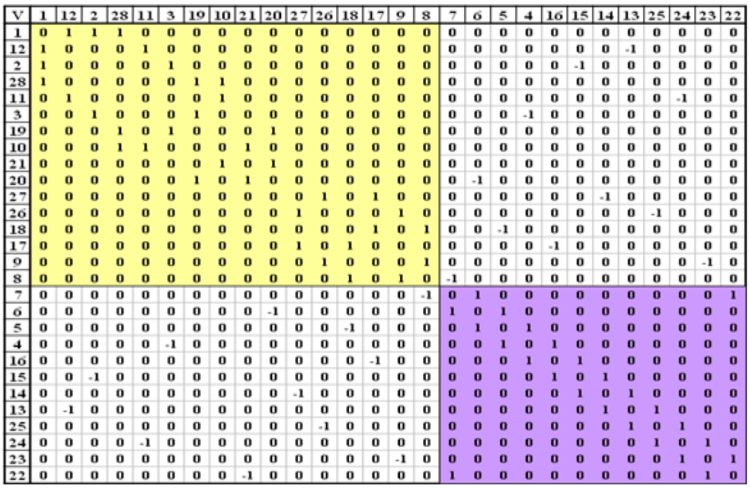
Property 4.3. *Given the two nodes v_i, $v_j \in V$, if $P_{ij}^l \approx 0$ or $P_{ij}^l < 0$, then to the node $v_k \in V$, if v_k and v_i are in the same community, not in the same community with v_j, then $P_{ik}^l > P_{jk}^l$.*

As can be seen from Formula (9), L is the only parameter to calculate the node tightness. If the value of L is too small, then the steps of random walking are too few, thus it is difficult to form a fairly large-scale community. If the value of L is too large or tends to infinity, then there will be no way to identify the community, and the whole network will form a community. Therefore, choosing a reasonable L will be helpful in dividing the community. Because most of the social networks belong to small-world networks, according to the "six degrees of separation" theory, any two nodes in the community can be reached by six steps, which is the average distance within the community being 6. Therefore, we select $L = 6$ as the length of a random walk step.

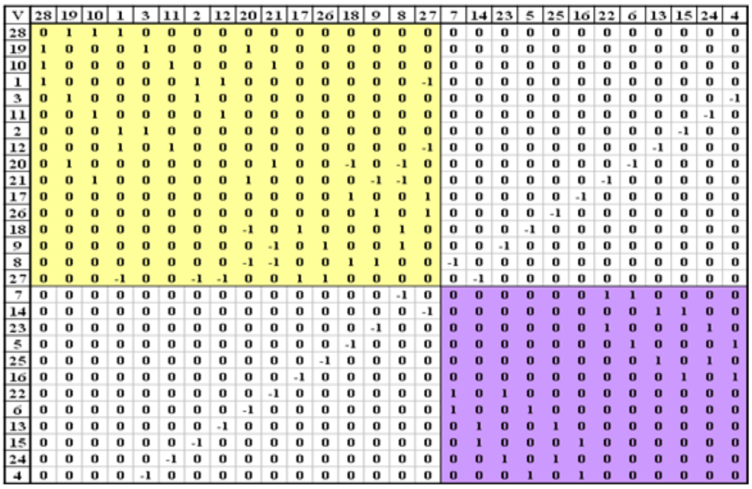
4.2. **Choice of initial center node in community.** When partitioning a community, a node is usually chosen to divide the community according to the tightness between the nodes. However, the choice of different nodes will lead to a different community distribution, especially when the selected nodes are bridge nodes or edge nodes, the effect of the community division will be seriously affected.

For the community corresponds to Figure 2(a), choosing the bridge node 5 to divide the community, as shown in Figure 3, the nodes of the same color in Figure 3 are the nodes of the same community, Figure 3(a) is the community distribution with the node 5 as the community center, as can be seen from the figure, it is hard to get an ideal community distribution. However, if node 1 is chosen as the central node of the community, then an ideal community distribution can be got, as shown in Figure 3(b) below. Therefore, it is necessary to divide the nodes in a community. Choosing a good initial central node plays an important role in successfully dividing a community.

In traditional social networks, there are many ways to calculate the importance of nodes in the network, for example, clustering coefficient, betweenness, degree. These methods just calculate starting from the structural properties of the network, however, the two attribute of network structure and signs are involved in the social network, so the conventional methods cannot be used to directly measure the community center node, thus the conventional methods need to be improved. Among them, the degree of a node as a simple measurement is widely used in the traditional network; usually the greater the degree is, the more likely it can become the central node of the community. Therefore, this paper based on the degree of nodes; taking into account the central node of a community not only needs to be of higher degree, but also needs to have more positive edges and fewer negative edges. This paper puts forward a simple and effective method to calculate the node center influence in the signed social network, as following.



(a)



(b)

FIGURE 3. Graph of community partition

Definition 4.8. *In SSN, for any node $v_i \in V$, the center influence $F(v_i)$ is shown in Formula (10).*

$$F(v_i) = e^{DP(v_i)} \big/ e^{DN(v_i)} \tag{10}$$

The node center influence has the following properties:

Property 4.4. *When $D(v_i) = D(v_j)$, if $DP(v_i) \geq DP(v_j)$, then $F(v_i) \geq F(v_j)$.*

Property 4.5. *When $DP(v_i)/DN(v_i) = DP(v_j)/DN(v_j)$ and $DN(v_i) \neq 0$, $DN(v_j) \neq 0$, if $DP(v_i) \geq DP(v_j)$, then $F(v_i) \geq F(v_j)$.*

This method not only reflects the relative ratio of the number of the positive sides and the negative edges in the node v_i, it also reflects the absolute ratio of the number of the positive edges and the negative edges. This method avoids getting a fairly large value when the degree is small, avoids being unable to represent the influence difference of the positive and negative edges, and also avoids incorrect calculation gotten when $DN(v_i) = 0$ and the dividend number is 0. Furthermore, it ensures that $F(v_i) = e^{DP(v_i)} \big/ e^{DN(v_i)}$ and $DP(v_i)/DN(v_i)$ has the same function monotony. Therefore, in SSN, this paper puts forward the center node influence $F(v_i)$ as a selection basis for the community center nodes. The descending orders by $F(v_i)$, the node which has the max value $F(v_i)$ is the center in community. When there are multiple identical maximum, you can choose one.

4.3. **Partitioned communities in SSN based on game theory.** In the SSN, the community structure widely exists. The nodes in the network can be regarded as the participants, and the positive connection between the nodes can be regarded as the cooperative relationship between the participants, the negative connection between the nodes can be regarded as the competitive relationship between the participants. The nodes choosing different communities corresponds to different strategies, the individual gains represent that every node should make a marginal contribution to the community after joining a community, and finally the maximum of the community average value can be taken as the standard to measure the community partition. Thus, the game theory can be applied to research the formation and the segmentation theory in the signed social network.

The main purpose of this paper is to divide the signed social network into m numbers disjoint sub-communities and to realize $G = g_1 \cup g_2 \cup \ldots \cup g_m$, and $g_i \cap g_j = \phi$, $(i \neq j)$. Some concepts of community and community partition based on game theory are as follows.

Definition 4.9. *Community is a two-tuple* (SSN, v)*, for any* $SSN = (G, \sigma)$*, where* $G = (V, A)$*, the node set* $V = \{1, 2, \ldots, n\}$ *denotes a finite set of players;* $A \subseteq V * V$ *is the adjacency matrix;* $v : \{g, g \subset g^{|V|}\} \to A$ *is the characteristic function; for any* $g \subseteq G$*,* g *is called a community;* $v(g)$ *is called the value of this community.*

Definition 4.10. *Community partitioned in SSN is defined as* $SNP = (G, v)$*, the community set is* $\{(g_1, v_{i1}(g_1)), \ldots, (g_m, v_{im}(g_m))\}$*, for any* $g_k \subseteq G$*, the community value* $v_i(g_k)$ *that* v_i *as the center node is shown in Formula (11), where* $d(v_i, v_j)$ *is tightness between nodes.*

$$v_i(g_k) = \begin{cases} 0 & g_k \subseteq G, |g_k| = 1 \text{ or } g_k = \phi \\ \sum\limits_{\substack{j \in g_k \\ i \neq j}} d(v_i, v_j)/|g_k| & g_k \subseteq G, |g_k| \geq 2, i \text{ is initial center vertex} \end{cases} \tag{11}$$

4.4. **Community partition algorithm description.** The detailed steps of signed social network partition (SNP) based on game theory are as shown in Figure 4.

5. **Experimental.** In this section, we performed extensive experiments to evaluate the performance of algorithm SNP on both illustrative signed network and real graph datasets. Illustrative signed network is shown in Figure 1. The real data set is Gahuku-Gama Subtribes Network. This was created based on reader's study on the cultures of highland New Guinea (obtained from http://mrvar.fdv.uni-lj.si/sola/info4/andrej/prpart5.htm).

5.1. **An illustrative signed network.** To verify the accuracy and validity of the algorithm, we have applied the SNP algorithm to the illustrative signed network, as given in Figure 1. Figure 5 presents the SNP outputs for the network. In the figure, the labels above are the node indices, and then are adjacency matrix and the initial community ID 0, the followed by their corresponding community IDs. For example, in the first row of Figure 5(a), "1" denotes "node 1" belonging to "community 1". As SNP is a recursive algorithm, the labels bellow indicate the sequence of extracting communities one by one. In addition, the matrix in Figure 5 contains diagonal blocks corresponding to different communities. In particular, three communities have been detected by SNP, which are community 1 (20, 19, 21, 3, 28, 10, 2, 1, 11, 12), community 2 (4, 5, 16, 6, 15, 7, 14, 22, 13, 23, 25, 24), and community 3 (8, 26, 17, 9, 18, 27).

Input: Adjacency matrix A

Output: Community set $\{(g_1, v_{i1}(g_1)), \ldots, (g_m, v_{im}(g_m))\}$

Begin

1. According to the signed social network $SSN = (G, \sigma)$, give the initial matrix A.

2. According to the equation (7), calculate the transition probability matrix P, and then calculate the random walk matrix R^L within L-step.

3. According to equation (10), calculate the central influence F of the community node.

4. According to the central influence F in descending order, select highest central node v_i, to arrange the tightness of v_i and other nodes in descending order matrix R^L.

5. Determine the community members where the initial node belong to according to whether r_{ij}>0, and divide the social network into two parts $g_k(v_j \in g_k)$ and $G - g_k$.

6. According to the formula (10), select the minimum node in tightness v_j to the initial central node v_i, to arrange the tightness of v_j and other nodes in descending order, to check the need for further examination of the community g_k.

1) If any r_{in} >0 and $v_i(g_k)$ is the max, then there's no need to further divide the community.

2) If there exits r_{in} <0, then divide g_k into $g_{k1}(r_{in} > 0)$ and $g_{k2}(r_{in} \leq 0)$ according to the positive and negative of g_k.

a) If $v_i(g_{k1}) < v_i(g_k)$ and $v_i(g_{k2}) < v_i(g_k)$, there is no need to further divide the community.

b) Otherwise continue dividing the community in step 4, 5, 6 until all the community average $v_i(g_k)$ is the greatest.

7. Among the nodes in the remaining community $G - g_k$, select the node with the highest central influence, and repeat steps 4, 5, 6.

End

FIGURE 4. Community partition algorithm description

V	1	28	12	2	19	10	3	11	20	21	27	26	17	9	18	8	7	23	5	22	6	16	25	4	24	14	15	13	Initial	No.1	No.2	No.3	No.4	No.5
1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	11	1		
28	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	11	1		
12	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	1	11	1		
2	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	1	11	1		
19	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	11	1		
10	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	11	1		
3	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	1	11	1		
11	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	1	11	1		
20	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	1	11	1		
21	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	11	1		
27	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	1	12		3	
26	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	1	12		3	
17	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1	12		3	
9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	-1	0	0	0	0	0	0	0	0	0	0	0	1	12		3	
18	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	12		3	
8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	1	12		3	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	1	1	0	0	0	0	0	0	0	0	2				2
23	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	2				2
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	2				2
22	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2				2
6	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2				2
16	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	2				2
25	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	2				2
4	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	2				2
24	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	2				2
14	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	2				2
15	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	2				2
13	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	2				2

(a)

V	28	19	10	1	3	11	2	12	20	21	4	16	5	15	6	14	7	13	22	25	23	24	9	26	8	27	18	17	Initial	No.1	No.2	No.3	No.4	No.5
28	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1			
19	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1			
10	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1			
1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	1	1			
3	0	1	0	0	0	0	1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1			
11	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1	1			
2	0	0	0	1	1	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1			
12	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	1	1			
20	0	1	0	0	0	0	0	0	0	1	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	-1	0	-1	0	0	1	1			
21	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	-1	0	0	0	-1	0	-1	0	0	0	0	1	1			
4	0	0	0	0	-1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2		21	2	
16	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	2		21	2	
5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	2		21	2	
15	0	0	0	0	0	0	-1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2		21	2	
6	0	0	0	0	0	0	0	0	-1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2		21	2	
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	-1	0	0	0	2		21	2	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	-1	0	0	0	0	2		21	2	
13	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	2		21	2	
22	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	2		21	2	
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	-1	0	0	0	0	0	2		21	2	
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	-1	0	0	0	0	0	0	2		21	2	
24	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	2		21	2	
9	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	-1	0	0	1	1	0	0	0	0	2		22		3
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	1	0	0	1	0	0	0	2		22		3
8	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	0	0	-1	0	0	0	0	0	1	0	0	0	1	0	0	2		22		3
27	0	0	0	-1	0	0	-1	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1	0	0	0	1	0	2		22		3
18	0	0	0	0	0	0	0	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	2		22		3
17	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2		22		3

(b)

FIGURE 5. The output for the illustrative signed network, as shown in Figure 1

V	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	+	−	D(v)	F(v)
1 GAVEV	0	1	−1	−1	−1	−1	0	0	0	0	0	−1	0	0	1	1	3	5	8	0.135335
2 KOTUN	1	0	−1	0	−1	−1	0	0	−1	−1	0	0	0	0	1	1	3	5	8	0.135335
3 OVE	−1	−1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	4	2	6	7.389056
4 ALIKA	−1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	2	1	3	2.718282
5 NAGAM	−1	−1	0	0	0	0	1	0	1	0	0	0	0	1	−1	−1	3	4	7	0.367879
6 GAHUK	−1	−1	1	0	0	0	1	1	−1	0	1	1	−1	0	0	−1	5	5	10	1
7 MASIL	0	0	1	0	1	1	0	1	0	0	1	1	1	0	0	0	7	0	7	1096.633
8 UKUDZ	0	0	1	1	0	1	1	0	0	0	1	1	0	−1	0	0	6	1	7	148.4132
9 NOTOH	0	−1	0	0	1	−1	0	0	0	1	−1	0	1	0	−1	0	3	4	7	0.367879
10 KOHIK	0	−1	0	0	0	0	0	0	1	0	−1	0	1	0	−1	0	2	3	5	0.367879
11 GEHAM	0	0	0	0	0	1	1	1	−1	−1	0	1	−1	0	−1	−1	4	5	9	0.367879
12 ASARO	−1	0	0	0	0	1	1	1	0	0	1	0	0	−1	−1	−1	4	4	8	1
13 UHETO	0	0	0	0	0	−1	1	0	1	1	−1	0	0	1	−1	−1	4	4	8	1
14 SEUVE	0	0	0	0	1	0	0	−1	0	0	0	−1	1	0	0	−1	2	3	5	0.367879
15 NAGAD	1	1	0	0	−1	0	0	0	−1	−1	−1	−1	−1	0	0	1	3	6	9	0.049787
16 GAMA	1	1	0	0	−1	−1	0	0	0	0	−1	−1	−1	−1	1	0	3	6	9	0.049787

(a) Adjacency matrix and node center influence

V	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	0	1/7	0	1/7	1/7	0	1/7	0	0	1/7	1/7	1/7	0	0	0
2	0	0	1/7	1/7	0	1/7	1/7	0	0	0	1/7	1/7	0	-1/7	0	0
3	-1/12	-1/12	0	1/6	0	1/6	1/6	1/6	0	0	0	0	0	0	0	0
4	-1/3	0	1/3	0	0	0	0	1/3	0	0	0	0	0	0	0	0
5	-1/50	-1/50	1/10	0	0	0	1/10	1/10	-1/50	0	1/10	1/10	-1/50	0	0	-1/50
6	-1/32	0	0	0	0	1/8	1/8	1/8	0	0	1/8	0	0	-1/32	-1/32	-1/32
7	0	0	0	0	0	-1/32	1/8	0	1/8	1/8	-1/32	0	0	1/8	-1/32	-1/32
8	-1/28	-1/28	0	0	0	0	1/7	0	1/7	0	0	0	0	1/7	-1/28	-1/28
9	0	-1/28	0	0	1/7	-1/28	0	0	0	1/7	-1/28	0	1/7	0	-1/28	0
10	0	-1/15	0	0	0	0	0	0	1/5	0	-1/15	0	1/5	0	-1/15	0
11	0	0	0	0	0	1/9	1/9	1/9	-1/45	-1/45	0	1/9	-1/45	0	-1/45	-1/45
12	0	0	0	0	1/5	0	0	-1/15	0	0	0	-1/15	1/5	0	0	-1/15
13	0	1/8	-1/40	-1/40	-1/40	-1/40	0	0	0	0	0	-1/40	0	0	1/8	1/8
14	1/8	0	-1/40	0	-1/40	-1/40	0	0	-1/40	-1/40	0	0	0	0	1/8	1/8
15	1/9	1/9	0	0	-1/54	0	0	0	-1/54	-1/54	-1/54	-1/54	-1/54	0	0	1/9
16	1/9	1/9	0	0	-1/54	-1/54	0	0	0	0	-1/54	-1/54	-1/54	-1/54	1/9	0

(b) Transition probability matrix

V	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.000	0.173	−0.053	−0.037	−0.043	−0.058	−0.034	−0.037	−0.016	−0.011	−0.026	−0.049	−0.014	−0.004	0.173	0.174
2	0.173	1.000	−0.041	−0.014	−0.047	−0.046	−0.027	−0.022	−0.043	−0.037	−0.016	−0.023	−0.023	−0.009	0.174	0.172
3	−0.216	−0.142	1.000	0.228	0.046	0.311	0.281	0.325	0.006	0.006	0.143	0.149	0.028	−0.046	−0.061	−0.063
4	−0.443	−0.120	0.446	1.000	0.032	0.224	0.196	0.452	0.003	0.003	0.143	0.154	0.011	−0.071	−0.088	−0.087
5	−0.070	−0.073	0.037	0.012	1.000	0.037	0.155	0.031	0.162	0.037	0.028	0.029	0.085	0.151	−0.075	−0.077
6	−0.078	−0.054	0.172	0.061	0.024	1.000	0.193	0.200	−0.022	−0.006	0.190	0.188	−0.005	−0.035	−0.030	−0.049
7	−0.086	−0.057	0.241	0.085	0.170	0.295	1.000	0.279	0.039	0.022	0.258	0.261	0.156	−0.008	−0.050	−0.052
8	−0.128	−0.054	0.291	0.209	0.013	0.323	0.288	1.000	−0.011	−0.007	0.278	0.284	−0.005	−0.177	−0.039	−0.035
9	−0.024	−0.068	0.001	−0.001	0.162	−0.040	0.032	−0.008	1.000	0.174	−0.052	−0.007	0.192	0.050	−0.074	−0.033
10	−0.028	−0.096	0.002	0.000	0.052	−0.020	0.022	−0.010	0.242	1.000	−0.080	−0.009	0.243	0.041	−0.103	−0.037
11	−0.044	−0.023	0.088	0.044	0.020	0.206	0.183	0.188	−0.030	−0.028	1.000	0.192	−0.013	−0.035	−0.039	−0.043
12	−0.079	−0.036	0.100	0.050	0.024	0.230	0.208	0.212	−0.003	−0.002	0.217	1.000	0.012	−0.061	−0.059	−0.060
13	−0.024	−0.037	0.022	0.006	0.076	−0.011	0.125	0.012	0.172	0.157	−0.024	0.011	1.000	0.139	−0.066	−0.059
14	−0.015	−0.026	−0.013	−0.013	0.215	−0.029	0.025	−0.072	0.069	0.040	−0.030	−0.075	0.218	1.000	−0.032	−0.088
15	0.152	0.153	−0.017	−0.008	−0.039	−0.023	−0.022	−0.016	−0.037	−0.031	−0.028	−0.034	−0.036	−0.010	1.000	0.155
16	0.152	0.152	−0.019	−0.009	−0.039	−0.040	−0.025	−0.018	−0.017	−0.012	−0.033	−0.036	−0.032	−0.024	0.155	1.000

(c) Matrix of tightness between nodes

V	Initial	V	7	No.1	V	14	No.2	V	10	No.3	V	11	No.4	V	1	No.5
1 GAVEV	0	7 MASIL	1.000	1	14 SEUVE	1.000	11	10 KOHIK	1.000	1						
2 KOTUN	0	8 UKUDZ	0.288	1	5 NAGAM	0.151	11	9 NOTOH	0.174	1						
3 OVE	0	3 OVE	0.281	1	13 UHETO	0.139	11	13 UHETO	0.157	1						
4 ALIKA	0	12 ASARO	0.208	1	9 NOTOH	0.050	11	14 SEUVE	0.040	1						
5 NAGAM	0	4 ALIKA	0.196	1	10 KOHIK	0.041	11	5 NAGAM	0.037	1						
6 GAHUK	0	6 GAHUK	0.193	1	7 MASIL	-0.008	12				11 GEHAM	1.000	3			
7 MASIL	0	11 GEHAM	0.183	1	6 GAHUK	-0.035	12				8 UKUDZ	0.278	3			
8 UKUDZ	0	5 NAGAM	0.155	1	11 GEHAM	-0.035	12				7 MASIL	0.258	3			
9 NOTOH	0	13 UHETO	0.125	1	3 OVE	-0.046	12				12 ASARO	0.217	3			
10 KOHIK	0	9 NOTOH	0.032	1	12 ASARO	-0.061	12				6 GAHUK	0.190	3			
11 GEHAM	0	14 SEUVE	0.025	1	4 ALIKA	-0.071	12				3 OVE	0.143	3			
12 ASARO	0	10 KOHIK	0.022	1	8 UKUDZ	-0.177	12				4 ALIKA	0.143	3			
13 UHETO	0	15 NAGAD	-0.022	2										1 GAVEV	1.000	2
14 SEUVE	0	16 GAMA	-0.025	2										2 KOTUN	0.173	2
15 NAGAD	0	2 KOTUN	-0.027	2										16 GAMA	0.152	2
16 GAMA	0	1 GAVEV	-0.034	2										15 NAGAD	0.152	2

(d) Community partition steps

V	14	5	13	9	10	7	6	11	12	3	4	8	15	16	2	1
14 SEUVE	0	1	1	0	0	0	0	0	-1	0	0	-1	0	-1	0	0
5 NAGAM	1	0	0	1	0	1	0	0	0	0	0	0	-1	-1	-1	-1
13 UHETO	1	0	0	1	1	1	-1	-1	0	0	0	0	-1	-1	0	0
9 NOTOH	0	1	1	0	1	0	-1	-1	0	0	0	0	-1	0	-1	0
10 KOHIK	0	0	1	1	0	0	0	-1	0	0	0	0	-1	0	-1	0
7 MASIL	0	1	1	0	0	0	1	1	1	1	0	1	0	0	0	0
6 GAHUK	0	0	-1	-1	0	1	0	1	1	1	0	1	0	-1	-1	-1
11 GEHAM	0	0	-1	-1	-1	1	1	0	1	0	0	1	-1	-1	0	0
3 OVE	0	0	0	0	0	1	1	0	0	0	1	1	0	0	-1	-1
12 ASARO	-1	0	0	0	0	1	1	1	0	0	0	1	-1	-1	0	-1
4 ALIKA	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	-1
8 UKUDZ	-1	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0
15 NAGAD	0	-1	-1	-1	-1	0	0	-1	-1	0	0	0	0	1	1	1
16 GAMA	-1	-1	-1	0	0	0	-1	-1	-1	0	0	0	1	0	1	1
2 KOTUN	0	-1	0	-1	-1	0	-1	0	0	-1	0	0	1	1	0	1
1 GAVEV	0	-1	0	0	0	0	-1	0	-1	-1	-1	0	1	1	1	0

(e) Output matrix of SNP algorithm

FIGURE 6. Community mining from Gahuku-Gama subtribes network using SNP

5.2. **Gahuku-Gama subtribes network.** It describes the political alliances and oppositions among 16 Gahuku-Gama subtribes, which were distributed in a particular area and were engaged in warfare with one another in 1954, as shown in Figure 6(a). The positive and negative links of the network correspond to political arrangements with positive and negative ties, respectively. Using the SNP algorithm, we analyzed this signed network and detected its community structure, as presented in Figure 6.

Firstly, according to the A, calculate the P, as shown in Figure 6(b), then calculate the random walk matrix R^L within L-step, as the tightness between nodes, as shown in Figure 6(c). Secondly, according to Equation (10), calculate the F of the community node, as shown in Figure 6(a); according to F in descending order, select highest central node "7" as the central node, to arrange the tightness of "7" and other nodes in descending order R^L. Finally, determine the community members of the community where the node "7" belongs to according to whether r_{ij} is greater than 0, and divide the social network into two parts, and then check the need for further partitioned the community according to whether $v_7(g_1)$ is the max, iterate until completing community partitions, as shown in Figure 6(d).

Three communities were detected as a result: (ASARO, GEHAM, OVE, UKUDZ, GAHUK, MASIL, ALIKA), (NAGAM,UHETO, KOHIK, NOTOH, SEUVE), and (NAGAD, GAMA, GAVEV,KOTUN), as shown in Figure 6(e), which was identical to the three-way partition of the same network reported by using FEC and CRA algorithms.

5.3. **Evaluation.** We have tested our proposed community mining algorithm SNP with a number of benchmark networks and randomly created networks so as to evaluate its effectiveness and efficiency. To measure the partitioning quality, we define the error rate of a partition C of a signed network as shown in Formula (12), where $P(C)$ is $\sum_k \sum_{i,j\in C_k} \max(0, -p_{ij}) + \sum_{r\neq s} \sum_{i\in C_r, i\in C_s} \max(0, p_{ij})$, and A is the adjacency matrix of the signed network.

$$error(C) = \frac{P(C)}{\sum_i \sum_j |a_{ij}|} * 100\% \tag{12}$$

If all the within-community links are positive, and the between-community links are negative, then $P(C) = \sum_i \sum_j |a_{ij}|$, and the error will be zero. If all the within-community links are negative, and all the between-community links are positive, then $P(C) = 0$, and the error will be 100 percent.

The error rate is shown in Table 2, about community partition SNP algorithm and FEC algorithm. Obviously, for illustrative signed network, due to only having positive links within-community and negative links between-community, the error of the partition is 0. However, for Gahuku-Gama subtribes signed network, the error ratio of the partition is 3.45 percent. The error rate of the best partition will be more than 0 due to some negative within-community links or positive between-community links. Therefore, the smaller the value of error(C) is the better the partitioning quality becomes.

TABLE 2. Error rate of signed network partition

Signed Social Network	**Error Rate**		
	SNP	**CRA**	**FEC**
Illustrative Signed Network(A)	0	0	0
Illustrative Signed Network(B)	0	0	0
Gahuku-Gama Subtribes Network	3.45	3.45	3.45

5.4. **Actual-time performances.** In paper, all the experiments are done on a 2.26GHZ Intel Core i3 PC with 2G main memory, Windows XP Professional SP3. All algorithms were implemented in Matlab 2012. We repeated SNP 100 times for each network, and the averaged actual computational time taken is shown in Table 3.

In addition, we have applied three different algorithms to signed social network with different size to examine how the run time would change with respect to the network size, and the relationship between run time and nodes number are shown in Figure 7. Comparisons of experimental results, we can see through the data curves. 1) The running time is low when the number is high. 2) The running time is approximately linear with respect to the network size. 3) With the increasing of the node number, the efficiency of the SPN algorithm is superior to the algorithms of CRA and FEC.

TABLE 3. Average actual computational time for different algorithms

Signed Social Network	**SNP(s)**	**CRA(s)**	**FEC(s)**
Illustrative Signed Network(A)	0.036	0.04	0.043
Illustrative Signed Network(B)	0.036	0.04	0.043
Gahuku-Gama Subtribes Network	0.028	0.035	0.039

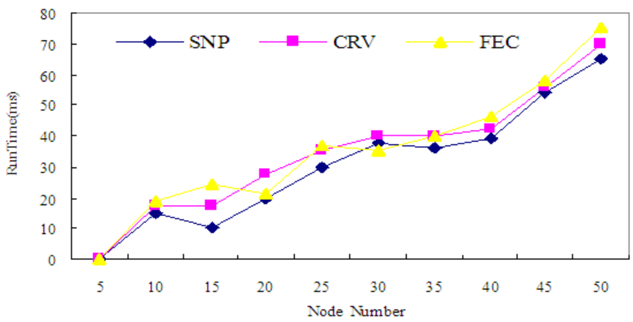


FIGURE 7. Relationship between nodes and time for different algorithms

6. **Conclusions and Future Work.** A new method of community partition (SNP) in signed social network is proposed in the paper, which considers the attributes of links and signed of the nodes based on game theory. Firstly, the transition probability matrix P is established based on the impact of signed attribute to the community structure. Secondly, based on the random walk method, L-step within the tightness between the node matrix R^L are calculated through the transition probability matrix P. Thirdly, the center influence of node is a selection basis for the community initial center nodes, taking into account the signs and the density properties. Fourthly, based on game theory, partition communities through the maximum average value of community. Finally, check the correctness and effectiveness of the algorithm through experiments. At present, this paper only considers the simple effect of signed attribute to the community partition. How to use the game theory to handle complex signed social networks is the focus of future work.

Acknowledgment. This work is supported by the National Science Foundation of China, (No. 60673136), National Science Foundation of Hebei (No. F2012209019) and Major Basic Research of Hebei Applied Basic Research Projects (No. 10963527D). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] P. Pons and M. Latapy, Computing communities in large networks using random walks, *Proc. of the 20th Int'l Symp. Computer and Information Sciences*, pp.284-293, 2005.

[2] M. Girvan and M. E. J. Newman, Community structure in social and biological networks, *Proc. of Nat'l Academy of Science*, vol.99, no.12, pp.7821-7826, 2002.

[3] G. W. Flake, S. Lawrence, C. L. Giles and F. M. Coetzee, Self-organization and identification of web communities, *Computer*, vol.35, no.3, pp.66-71, 2002.

[4] B. Yang, W. K. Cheung and J. Liu, Community mining from signed social networks, *IEEE/KDE*, vol.19, no.10, pp.1333-1348, 2007.

[5] T. Sharma, A. Charls and P. K. Singh, *Community Mining in Signed Social Networks – An Automated Approach*, IACSIT Press, Singapore, 2011.

[6] L. Lovász, Random walks on graphs: A survey, *Combinatorics, Paul Erdos is Eighty*, vol.2, pp.353-397, 1993.

[7] J. Weston, A. Elisseeff, D. Zhou, C. S. Leslie and W. S. Noble, Protein ranking: From local to global structure in the protein similarity network, *Proc. of Nat. Acad. Sci.*, vol.101, no.17, pp.6569-6563, 2004.

[8] W. Zheng, C. Wang, Z. Liu and J. Wang, A multi-label classification algorithm based on random walk model, *Chinese Journal of Computer*, vol.33, no.8, pp.1418-1426, 2010.

[9] S. Moretti et al., Using coalitional games on biological networks to measure centrality and power of genes, *Bioinformatics*, vol.26, no.21, pp.2721-2730, 2010.

[10] L. Zhou, C. Cheng, K. Lü and H. Chen, Using coalitional games to detect communities in social networks, *Proc. of WAIM 2013*, vol.7923, pp.326-331, 2013.