Community detection in signed networks based on discrete-time model*

Jianrui Chen(陈建芮)†, Li Zhang(张莉), Weiwei Liu(刘维维), and Zaizai Yan(闫在在)

College of Science, Inner Mongolia University of Technology, Hohhot 010051, *China*

(Received 19 May 2016; revised manuscript received 26 June 2016; published online 30 November 2016)

Community detection in signed networks has been studied widely in recent years. In this paper, a discrete difference equation is proposed to imitate the consistently changing phases of the nodes. During the interaction, each node will update its phase based on the difference equation. Each node has many different nodes connected with it, and these neighbors have different influences on it. The similarity between two nodes is applied to describe the influences between them. Nodes with high positive similarities will get together and nodes with negative similarities will be far away from each other. Communities are detected ultimately when the phases of the nodes are stable. Experiments on real world and synthetic signed networks show the efficiency of detection performance. Moreover, the presented method gains better detection performance than two existing good algorithms.

Keywords: community detection, signed networks, discrete-time model, similarity

PACS: 89.75.Fb, 89.75.Hc, 84.30.Bv **DOI:** 10.1088/1674-1056/26/1/018901

1. Introduction

Community detection of complex networks has been a hot topic in the network study.[1] Many social, recommendation, and information systems can be naturally described as networks, where vertices represent entities and links denote relationships or interactions between the vertices.[2] The topological structure of many networks often features an organization of communities in which the nodes tend to divide into clusters, with dense connections within clusters and sparser connections between them.[3] There are many relationships between individuals in the real world, especially in social networks. Commonly, the social relationships between individuals have positive relationships and negative relationships.[4] It will be regarded as a positive link when two individuals have a friendly relationship, however, it will be regarded as a negative link when two individuals have a hostile relationship. Such networks are called signed networks.[5] The relationships between individuals also have different levels, the similarity can be defined by using the essential attributes of nodes: two nodes are considered to be similar if they have many common features.[6]

There are many algorithms proposed for community detections, such as KerNighanLin algorithm,[7] spectral bisection method,[8] Girvan–Newman algorithm,[9,10] etc. An agent-based heuristic algorithm (FEC) has been adopted to mine signed social networks in which both positive within-group relationships and negative between-group relationships are dense.[11] Two overlapping community detection algorithms have been introduced, one is the signed probabilistic mixture (SPM) model for overlapping community detection in signed networks,[12] and the other is the dominant label propagation algorithm (DLPA).[13] Dynamic algorithms or cluster synchronization[14] have been proposed in many papers. A dynamic community detection algorithm based on the phase synchronization of Kuramoto oscillators has been proposed.[15] It has put in negative coupling strength coefficients in order to make the phases of two unconnected oscillators in the network evolve far away, so the network can be divided into several clusters. Besides, several improved Kuramoto models have been introduced in Refs. [16]–[19]. Also, reference [20] studied the oscillatory behaviors of excitable complex networks (ECNs) and found some interesting dynamic behaviors of ECNs in oscillatory probability. The above mentioned work bridges the gap between the structure and the dynamics of networks. What is more, several algorithms based on the statistical measures for characterizing the degree correlation,[21] Potts spin model,[22,23] self-loop rescaling strategy,[24] local modularity,[25] and joint matrix factorization[26] also played an effective role in detecting communities. There are some optimization algorithms applied in community discovery, such as single objective optimization algorithm,[27] multi objective optimization algorithm,[28] etc.

In this paper, the discrete-time model is applied to solve the community detection problem of a signed network. Furthermore, a novel similarity is put forward to imitate the relationship between two nodes. The remainder of the paper is organized as follows. In Section 2, the related work is intro-

*Project supported by the National Natural Science Foundation of China (Grant Nos. 11261034, 71561020, 61503203, and 11326239), the Higher School Science and Technology Research Project of Inner Mongolia, China (Grant No. NJZY13119), and the Natural Science Foundation of Inner Mongolia, China (Grant Nos. 2015MS0103 and 2014BS0105).

†Corresponding author. E-mail: jianrui_chen@sina.com

© 2017 Chinese Physical Society and IOP Publishing Ltd

duced and the model is described. In Section 3, the theoretical analysis is given. In Section 4, the detailed algorithm is provided. In Section 5, experimental results are demonstrated. Section 6 makes a summary of this work.

2. A discrete-time model based on similarity

In this paper, the signed network $G(V,W)$ is considered, where V is the set of nodes, and W is the set of weights (similarity) . There is some literature about community detection of a signed network. An algorithm for a signed social network (DBAS) has been presented which is detailed dynamics-based.[29] The model is

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{K_\mathrm{p}}{N}\sum_{j=1}^{N}\frac{|a_{ij}|(1+a_{ij})}{2}\sin(\theta_j-\theta_i) + \frac{K_\mathrm{n}}{N}\sum_{j=1}^{N}\frac{|a_{ij}|(1-a_{ij})}{2}\sin(\theta_j-\theta_i). \quad (1)$$

Here N is the number of nodes in the positive social network, θ_i is the phase of the i-th node, $K_\mathrm{p}>0$ is the positive coupling strength, and $K_\mathrm{n}<0$ is the negative coupling strength. If the j-th node is positively connected with the i-th node, $a_{ij}=1$, if the j-th node is negatively connected with the i-th node, $a_{ij}=-1$, otherwise $a_{ij}=0$.

In general, all the non-observed links are ranked according to their scores, the links connecting more similar nodes are supposed to be of higher existence likelihoods.[30] In this situation, many local similarity measures in networks are closely related to the concept of community structures.[31] We have found that the similarity between nodes should be considered. Dynamic mechanism with similarity would gain better detection performance than the model without similarity.[32]

In Ref. [33], a novel similarity measurement was defined according to the social balance theory[34–36] for the signed network. Due to their different similarities, inner-community positive links were found and detected. A threshold has been given to decide a node's positive neighbors who have high similarity in magnitude with this node. However the algorithm in Ref. [33] has difficulty in detecting the community structure for the following two kinds of networks. (i) The total number of communities is large. (ii) The sizes of different communities are not balanced. To further improve the detection precision, we give the following similarity definition and network model.

In many applications, there are some nodes without direct connection but they still have similarities.[37] To handle this problem, a kind of similarity between two nodes is defined as

$$s_{ij}=\sum_{k=1}^{m}\lambda_k\times N_k,\quad \lambda_1\geq\lambda_2\geq\cdots\geq\lambda_m, \quad (2)$$

where N_k is the number of the paths with k hops between nodes i and j, and m is the largest hops of the path between nodes i and j. In fact, $N_1=0$ if the i-th node and the j-th node are directly connected. N_1 is the number of common neighbors of nodes i and j, which is equal to the number of one-hop paths. Moreover, λ_k in Eq. (2) is the weight of the k-hop path to the similarity, it can be calculated by

$$\lambda_k=\frac{1}{d^k},$$

where d is the average degree of the network.

In this paper, we consider the influences from both positive neighbors and negative neighbors. The similarity evaluation formula just takes the first two terms to reduce the computational cost, that is to say, only one-hop and two-hop paths have been considered. For the one-hop path, consider nodes i, j, k, here node k is the common neighbor of nodes i and j. If nodes i and j have node k as their common positive neighbor (shown in Fig. 1(a)), the similarity between i and j should be positive. If nodes i and j have node k as their common negative neighbor (shown in Fig. 1(b)), the similarity between i and j should be positive, too. In other words, we are friends if we have a common friend or common enemy. Thus the positive one-hop path is computed as

$$N_1^+=\sum_{A(i,k)A(k,j)>0}A(i,k)A(k,j). \quad (3)$$

Here $\boldsymbol{A}$ is the adjacency matrix. $A(i,j)=1$ if there is a positive link between the i-th node and the j-th node; $A(i,j)=-1$ if there is a negative link between the i-th node and the j-th node; otherwise, $A(i,j)=0$ if there is no link between the i-th node and the j-th node.

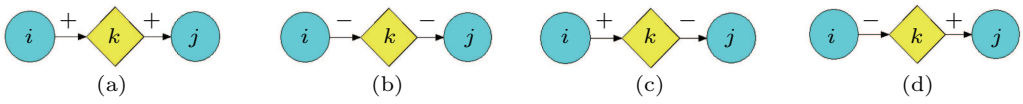


Fig. 1. (color online) Connections between one-hop path nodes. See text for the detailed description of panels (a)–(d).

However, if the k-th node is a positive neighbor of the i-th node and a negative neighbor of the j-th node (shown in Fig. 1(c)), there will be a negative influence between nodes i and j. If the k-th node is a negative neighbor of the i-th node and a positive neighbor of the j-th node (shown in Fig. 1(d)), there will also be a negative influence between nodes i and j. That is, the enemy of my friend is my enemy and the friend of my enemy is my enemy. Similarly, the negative one-hop path

is computed as

$$N_1^- = \sum_{A(i,k)A(k,j)<0} A(i,k)A(k,j). \tag{4}$$

Without loss of generality, for the two-hop path, consider nodes i, j, k, and h. There are eight kinds of situations, as shown in Fig. 2, and the different links denote different situations. Similar to the one-hop path, the positive and the negative two-hop paths can be defined as

$$N_2^+ = \sum_{A(i,k)A(k,h)A(h,j)>0} A(i,k)A(k,h)A(h,j), \tag{5}$$

$$N_2^- = \sum_{A(i,k)A(k,h)A(h,j)<0} A(i,k)A(k,h)A(h,j). \tag{6}$$

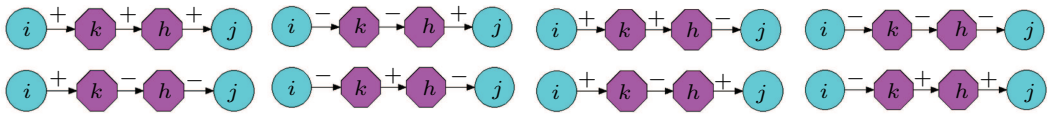


Fig. 2. (color online) Connections between two-hop path nodes.

Above all, the similarity between nodes i and j is defined as follows:

$$\mathrm{sim}_{ij} = \begin{cases} \lambda_1 \times N_1^+ + \lambda_2 \times N_2^+, & A(i,j) > 0, \\ \lambda_1 \times N_1^- + \lambda_2 \times N_2^-, & A(i,j) < 0, \\ 0, & A(i,j) = 0. \end{cases} \tag{7}$$

Furthermore by normalization, the similarity can be redefined as follows:

$$s_{ij} = \begin{cases} \dfrac{\mathrm{sim}_{ij}}{\sum_{\mathrm{sim}_{ij}>0} \mathrm{sim}_{ij}}, & A(i,j) > 0, \\ \dfrac{\mathrm{sim}_{ij}}{\sum_{\mathrm{sim}_{ij}<0} \mathrm{sim}_{ij}}, & A(i,j) < 0. \\ 0, & A(i,j) = 0. \end{cases} \tag{8}$$

After calculating the similarity of nodes, we find that the probability density has a normal distribution, and it has the characteristic that most similarities are close to the mean. Here the synthetic signed network with 500 nodes is taken as an example, as shown in Fig. 3. To better grasp the different similarity distributions in different networks, upper and lower quantile α will be adopted as a threshold. The community structure of the signed networks has the following characteristics: there are more positive links within the communities and more negative links between communities. But the community detection of the signed networks often shows the phenomenon that there are more positive links between communities and more negative links within the communities. For this reason, the sets of positive and negative neighbors are determined by the upper and lower quantile α. The definition of α is

$$P\{|s_{ij}| > U_{\alpha/2}\} = \alpha. \tag{9}$$

According to Eq. (9), the set of positive neighbors for node i is expressed as

$$N_e^+ = \{j \mid s_{ij} > U_{\alpha/2}\}.$$

The set of negative neighbors for node i is expressed as

$$N_e^- = \{j \mid s_{ij} < 0\}.$$

In this paper, we further improve the formulation from Ref. [19] as follows:

$$\begin{aligned} \theta_i(k+1) = \theta_i(k) &+ K_\mathrm{p} \sum_{j \in N_e^+} \sin\left(s_{ij}\left(\theta_j(k) - \theta_i(k)\right)\right) \\ &+ K_\mathrm{n} \sum_{j \in N_e^-} sin\left(s_{ij}\left(\theta_j(k) - \theta_i(k)\right)\right). \end{aligned} \tag{10}$$

Here K_p and K_n are predefined parameters, $K_\mathrm{p} > 0$ and $K_\mathrm{n} > 0$. The initial phase $\theta_i(0)$ is randomly and uniformly selected and distributed in $[0, \pi/2]$. The parameter α is in the range $[0.1, 0.3]$.

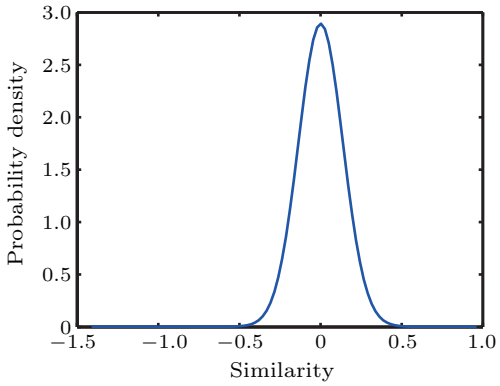


Fig. 3. (color online) Probability density estimate of similarity matrix $\boldsymbol{S}$.

3. Theoretical analysis

To simplify network model Eq. (10), the following transformation is adopted. Define

$$S^+ = \{s_{ij} \mid s_{ij} > U_{\alpha/2}\}, \quad S^- = \{s_{ij} \mid s_{ij} < 0\}.$$

All positive values higher than $U_{\alpha/2}$ in S are kept in S^+ and the other elements in S^+ are all zeros. In the same way, all negative values lower than 0 in S are kept in S^- and the other elements in S^- are all zeros. $\boldsymbol{T}^+$ is a diagonal matrix and the diagonal element is defined as

$$T_{ii}^+ = \sum_{\{j|s_{ij}>U_{\alpha/2}\}} s_{ij}, \quad i = 1, \ldots, N.$$

$\boldsymbol{T}^-$ is a diagonal matrix and the diagonal element is defined as

$$T_{ii}^{-}=\sum_{\{j|s_{ij}<0\}}s_{ij},\ \ i=1,\ldots,N.$$

Furthermore, denote $\boldsymbol{L}^{+}=\boldsymbol{S}^{+}-\frac{1}{2}\boldsymbol{T}^{+},\boldsymbol{L}^{-}=\frac{1}{2}\boldsymbol{S}^{-}-\boldsymbol{T}^{-}$.

Theorem 1 The network model is uniformly stable if the following matrix is negative semi-definite:

$$\left(2\boldsymbol{E}+\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)^{\mathrm{T}}\right)\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right),\qquad(11)$$

provided with suitable K_{p} and K_{n}.

Proof The identical transformation of Eq. (10) can be deduced as follows:

$$\begin{aligned}\theta_i(k+1)\ =\ &\theta_i(k)+K_{\mathrm{p}}\sum_{j\in N_e^{+}}\sin\left(s_{ij}\left(\theta_j(k)-\theta_i(k)\right)\right)\\&+K_{\mathrm{n}}\sum_{j\in N_e^{-}}\sin\left(s_{ij}\left(\theta_j(k)-\theta_i(k)\right)\right)\\=\ &\theta_i(k)+K_{\mathrm{p}}\sum_{j=1}^{N}\sin\left(s_{ij}^{+}\left(\theta_j(k)-\theta_i(k)\right)\right)\\&+K_{\mathrm{n}}\sum_{j=1}^{N}\sin\left(s_{ij}^{-}\left(\theta_j(k)-\theta_i(k)\right)\right).\end{aligned}$$

In this paper, $-1\le s_{ij}\le 1$ and the initial values θ_i $(i=1,\ldots,N)$ are selected randomly from the interval $[0,\pi/2]$, we can obtain $-\frac{\pi}{2}\le s_{ij}\left(\theta_j(k)-\theta_i(k)\right)\le\frac{\pi}{2}$. It can be observed that $\sin\left(s_{ij}\left(\theta_j(k)-\theta_i(k)\right)\right)\le\mu_{ij}s_{ij}\left(\theta_j(k)-\theta_i(k)\right)$, where

$$\mu_{ij}=\begin{cases}\dfrac{1}{2}, & -\dfrac{\pi}{2}\le s_{ij}\left(\theta_j(k)-\theta_i(k)\right)<0,\\ 1, & 0\le s_{ij}\left(\theta_j(k)-\theta_i(k)\right)\le\dfrac{\pi}{2}.\end{cases}$$

The above formula can be further expressed as

$$\begin{aligned}\theta_i(k+1)\ \le\ &\theta_i(k)+K_{\mathrm{p}}\sum_{j=1}^{N}\left(\mu_{ij}s_{ij}^{+}\left(\theta_j(k)-\theta_i(k)\right)\right)\\&+K_{\mathrm{n}}\sum_{j=1}^{N}\left(\mu_{i,j}s_{ij}^{-}\left(\theta_j(k)-\theta_i(k)\right)\right)\\=\ &\theta_i(k)+K_{\mathrm{p}}\sum_{j=1}^{N}\left(\mu_{ij}s_{ij}^{+}\theta_j(k)-\mu_{ij}s_{ij}^{+}\theta_i(k)\right)\\&+K_{\mathrm{n}}\sum_{j=1}^{N}\left(\mu_{ij}s_{ij}^{-}\theta_j(k)-\mu_{ij}s_{ij}^{-}\theta_i(k)\right)\\=\ &\theta_i(k)+K_{\mathrm{p}}\left(\sum_{j=1}^{N}\left(\mu_{ij}s_{ij}^{+}\theta_j(k)\right)-\sum_{j=1}^{N}\mu_{ij}s_{ij}^{+}\theta_i(k)\right)\\&+K_{\mathrm{n}}\left(\sum_{j=1}^{N}\left(\mu_{ij}s_{ij}^{-}\theta_j(k)\right)-\sum_{j=1}^{N}\mu_{ij}s_{ij}^{-}\theta_i(k)\right).\end{aligned}$$

Obviously,

$$\begin{aligned}&K_{\mathrm{p}}\sum_{j=1}^{N}\left(\mu_{ij}s_{ij}^{+}\right)\theta_j(k)\le K_{\mathrm{p}}\sum_{j=1}^{N}s_{ij}^{+}\theta_j(k),\\&-K_{\mathrm{p}}\sum_{j=1}^{N}\mu_{ij}s_{ij}^{+}\theta_i(k)\le-\frac{1}{2}K_{\mathrm{p}}\sum_{j=1}^{N}s_{ij}^{+}\theta_i(k),\\&K_{\mathrm{n}}\sum_{j=1}^{N}\left(\mu_{ij}s_{ij}^{-}\right)\theta_j(k)\le\frac{1}{2}K_{\mathrm{n}}\sum_{j=1}^{N}s_{ij}^{-}\theta_j(k),\\&-K_{\mathrm{n}}\sum_{j=1}^{N}\mu_{ij}s_{ij}^{-}\theta_i(k)\le-K_{\mathrm{n}}\sum_{j=1}^{N}s_{ij}^{-}\theta_i(k).\end{aligned}$$

Then

$$\begin{aligned}\theta_i(k+1)\ \le\ &\theta_i(k)+K_{\mathrm{p}}\sum_{j=1}^{N}s_{ij}^{+}\theta_j(k)-\frac{1}{2}K_{\mathrm{p}}\sum_{j=1}^{N}s_{ij}^{+}\theta_i(k)\\&+\frac{1}{2}K_{\mathrm{n}}\sum_{j=1}^{N}s_{ij}^{-}\theta_j(k)-K_{\mathrm{n}}\sum_{j=1}^{N}s_{ij}^{-}\theta_i(k).\end{aligned}$$

Denote $\boldsymbol{\theta}(k)=\left(\theta_1(k),\theta_2(k),\ldots,\theta_N(k)\right)^{\mathrm{T}}$. The matrix form can be written as

$$\begin{aligned}\boldsymbol{\theta}(k+1)\ \le\ &\boldsymbol{\theta}(k)+K_{\mathrm{p}}\left(\boldsymbol{S}^{+}\boldsymbol{\theta}(k)-\frac{1}{2}\boldsymbol{T}^{+}\boldsymbol{\theta}(k)\right)\\&+K_{\mathrm{n}}\left(\frac{1}{2}\boldsymbol{S}^{-}\boldsymbol{\theta}(k)-\boldsymbol{T}^{-}\boldsymbol{\theta}(k)\right)\\=\ &\boldsymbol{\theta}(k)+K_{\mathrm{p}}\left(\left(\boldsymbol{S}^{+}-\frac{1}{2}\boldsymbol{T}^{+}\right)\boldsymbol{\theta}(k)\right)\\&+\boldsymbol{K}_{\mathrm{n}}\left(\left(\frac{1}{2}\boldsymbol{S}^{-}-\boldsymbol{T}^{-}\right)\boldsymbol{\theta}(k)\right)\\=\ &\boldsymbol{\theta}(k)+K_{\mathrm{p}}\boldsymbol{L}^{+}\boldsymbol{\theta}(k)+K_{\mathrm{n}}\boldsymbol{L}^{-}\boldsymbol{\theta}(k).\end{aligned}$$

Define the Lyapunov function

$$V=\sum_{i=1}^{N}\theta_i^2(k),$$

then

$$\begin{aligned}\Delta V\ =\ &V(k+1)-V(k)\\=\ &\sum_{i=1}^{N}\theta_i^2(k+1)-\sum_{i=1}^{N}\theta_i^2(k)\\=\ &\boldsymbol{\theta}(k+1)^{\mathrm{T}}\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)^{\mathrm{T}}\boldsymbol{\theta}(k)\\\le\ &\left[\boldsymbol{\theta}(k)+K_{\mathrm{p}}\boldsymbol{L}^{+}\boldsymbol{\theta}(k)+K_{\mathrm{n}}\boldsymbol{L}^{-}\boldsymbol{\theta}(k)\right]^{\mathrm{T}}\\&\times\left[\boldsymbol{\theta}(k)+K_{\mathrm{p}}\boldsymbol{L}^{+}\boldsymbol{\theta}(k)+K_{\mathrm{n}}\boldsymbol{L}^{-}\boldsymbol{\theta}(k)\right]-\boldsymbol{\theta}(k)^{\mathrm{T}}\boldsymbol{\theta}(k)\\=\ &\boldsymbol{\theta}(k)^{\mathrm{T}}\boldsymbol{\theta}(k)+\boldsymbol{\theta}(k)^{\mathrm{T}}\left[K_{\mathrm{p}}\boldsymbol{L}^{+}\boldsymbol{\theta}(k)+K_{\mathrm{n}}\boldsymbol{L}^{-}\boldsymbol{\theta}(k)\right]\\&+\left[K_{\mathrm{p}}\boldsymbol{L}^{+}\boldsymbol{\theta}(k)+K_{\mathrm{n}}\boldsymbol{L}^{-}\boldsymbol{\theta}(k)\right]^{\mathrm{T}}\boldsymbol{\theta}(k)\\&+\left[K_{\mathrm{p}}\boldsymbol{L}^{+}\boldsymbol{\theta}(k)+K_{\mathrm{n}}\boldsymbol{L}^{-}\boldsymbol{\theta}(k)\right]^{\mathrm{T}}\\&\times\left[K_{\mathrm{p}}\boldsymbol{L}^{+}\boldsymbol{\theta}(k)+K_{\mathrm{n}}\boldsymbol{L}^{-}\boldsymbol{\theta}(k)\right]-\boldsymbol{\theta}(k)^{\mathrm{T}}\boldsymbol{\theta}(k)\\=\ &2\boldsymbol{\theta}(k)^{\mathrm{T}}\left[K_{\mathrm{p}}\boldsymbol{L}^{+}\boldsymbol{\theta}(k)+K_{\mathrm{n}}\boldsymbol{L}^{-}\boldsymbol{\theta}(k)\right]\\&+\left[\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)\boldsymbol{\theta}(k)\right]^{\mathrm{T}}\left[\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)\boldsymbol{\theta}(k)\right]\\=\ &2\boldsymbol{\theta}(k)^{\mathrm{T}}\left[\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)\boldsymbol{\theta}(k)\right]\\&+\boldsymbol{\theta}(k)^{\mathrm{T}}\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)^{\mathrm{T}}\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)\boldsymbol{\theta}(k)\\=\ &\boldsymbol{\theta}(k)^{\mathrm{T}}\left[2\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)+\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)^{\mathrm{T}}\right.\\&\left.\times\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)\right]\boldsymbol{\theta}(k)\\=\ &\boldsymbol{\theta}(k)^{\mathrm{T}}\left[\left(2E+\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)^{\mathrm{T}}\right)\right.\\&\left.\times\left(K_{\mathrm{p}}\boldsymbol{L}^{+}+K_{\mathrm{n}}\boldsymbol{L}^{-}\right)\right]\boldsymbol{\theta}(k).\end{aligned}$$

Here, E is an $N \times N$ identity matrix. If K_{p} and K_{n} are selected to satisfy that the following matrix

$$\left(2E + \left(K_{\mathrm{p}}\boldsymbol{L}^{+} + K_{\mathrm{n}}\boldsymbol{L}^{-}\right)^{\mathrm{T}}\right)\left(K_{\mathrm{p}}\boldsymbol{L}^{+} + K_{\mathrm{n}}\boldsymbol{L}^{-}\right)$$

is negative semi-definite. According to the Lyapunov stability theory,[38] the network model is uniformly stable.

The above theoretical analyses show that our proposed network model would be uniformly stable. With time evolving, the community structure would be formed according to the stable results.

Parameter analysis In fact, it can be found that the stable condition in Theorem 1 can be held with suitable K_{p} and K_{n}. We give further demonstration.

Denote $\boldsymbol{P} = \left(2\boldsymbol{E} + \left(K_{\mathrm{p}}\boldsymbol{L}^{+} + K_{\mathrm{n}}\boldsymbol{L}^{-}\right)^{\mathrm{T}}\right)\left(K_{\mathrm{p}}\boldsymbol{L}^{+} + K_{\mathrm{n}}\boldsymbol{L}^{-}\right)$. The diagonal elements of $\boldsymbol{P}$ are

$$\left(2 + K_{\mathrm{p}}\left(S_{ii}^{+} - \sum_{j\in N_e^{+}}\frac{1}{2}S_{ij}\right) + K_{\mathrm{n}}\left(\frac{1}{2}S_{ii}^{-} - \sum_{j\in N_e^{-}}S_{ij}\right)\right) \times \left(K_{\mathrm{p}}\left(S_{ii}^{+} - \sum_{j\in N_e^{+}}\frac{1}{2}S_{ij}\right) + K_{\mathrm{n}}\left(\frac{1}{2}S_{ii}^{-} - \sum_{j\in N_e^{-}}S_{ij}\right)\right).$$

If suitable K_{p} and K_{n} are provided, each diagonal element in $\boldsymbol{P}$ would satisfy the following condition:

$$\left(2 + K_{\mathrm{p}}\left(S_{ii}^{+} - \sum_{j\in N_e^{+}}\frac{1}{2}S_{ij}\right) + K_{\mathrm{n}}\left(\frac{1}{2}S_{ii}^{-} - \sum_{j\in N_e^{-}}S_{ij}\right)\right) \times \left(K_{\mathrm{p}}\left(S_{ii}^{+} - \sum_{j\in N_e^{+}}\frac{1}{2}S_{ij}\right) + K_{\mathrm{n}}\left(\frac{1}{2}S_{ii}^{-} - \sum_{j\in N_e^{-}}S_{ij}\right)\right) < 0.$$

The diagonal elements in $\boldsymbol{S}^{+}$ and $\boldsymbol{S}^{-}$ are equal to 0 due to the diagonal elements in the similarity matrix being equal to 0. In this case, $S_{ii}^{+} = 0, S_{ii}^{-} = 0$. For arbitrarily $m < 0$, $\boldsymbol{P} + m\boldsymbol{I}$ can satisfy the condition of strictly diagonally dominant matrix

$$|P_{ii}| > \sum_{j=1, j\neq i}^{N} |P_{ij}|.$$

It can be found that $\boldsymbol{P}$ is invertible. Besides, the linear equation $(\boldsymbol{P} + m\boldsymbol{I})x = 0$ merely has zero solution. This means $-m > 0$ (m is any given negative number) is not the eigenvalue of $\boldsymbol{P}$, all eigenvalues of $\boldsymbol{P}$ are not positive numbers. Consequently, $\boldsymbol{P}$ is a negative definite matrix.

The stable condition can be satisfied as K_{p} is ranged in $[0.7, 1]$ and K_{n} is ranged in $[0, 0.07]$ for different signed networks. However, in Ref. [19], for different scales of signed networks, the values of K_{p} and K_{n} are adjusted in a very wide range.

4. The algorithm

The discrete-time based model for community detection in signed network (denoted as DTCDS) is further described in detail.

For a signed network, the connections should be positive and dense inside communities, and the connections should be negative and sparse between communities.[33] During the iteration of Eq. (10), the nodes with high positive similarities will evolve together, while the nodes with negative similarities will evolve far away. The algorithm flow of DTCDS is given as follows.

Algorithm 1 DTCDS

Require adjacency matrix, $K_{\mathrm{p}} > 0$, $K_{\mathrm{n}} > 0$, $\alpha > 0$, $\varepsilon = 0.01$
Ensure communities C
1. //Calculate the similarity matrix
2. **for** each node $i \in V$ **do**
3. compute the positive one-hop path N_1^{+} and positive two-hop path N_2^{+} according to Eqs. (3) and (5);
4. compute the negative one-hop path N_1^{-} and negative two-hop path N_2^{-} according to Eqs. (4) and (6);
5. compute the similarity based on Eqs. (7) and (8)
6. **end for**
7. //Update the phase of nodes
8. **for** each node $i \in V$ **do**
9. **if** the phase difference in one step is higher than threshold ε **then**
10. update the phase of nodes according to Eq. (10)
11. **end if**
12. **end for**
13. //Find communities
14. **for** each node $i \in V$ **do**
15. **if** $|\theta_j(k) - \theta_i(k)| < 0.01$ **then**
16. nodes i and j are put into the same community
17. **end if**
18. **end for**

Step 1: Input the adjacency matrix. According to Ref. [33], it has positive links inside communities and negative links between communities. The adjacency matrix $\boldsymbol{A}$ of a synthetic network is input and figure 4(a) gives the network structure

$$A = \begin{bmatrix} 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 & -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & -1 \\ -1 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 & 0 & 1 & -1 & 0 & 1 \\ -1 & -1 & 0 & 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 0 & -1 & -1 & -1 & 0 & -1 & -1 & 0 & 1 & 1 \\ -1 & 0 & 0 & -1 & 0 & 0 & -1 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 & -1 & -1 & 1 & 1 & 0 \end{bmatrix}.$$

Step 2: Calculate the similarity matrix. The similarity matrix is calculated by Eq. (8)

$$S = \begin{bmatrix} 0 & 0.55 & 0.45 & -0.35 & 0 & 0 & -0.47 & 0 & -0.19 & 0 \\ 0.51 & 0 & 0.49 & 0 & -0.28 & 0 & -0.42 & -0.30 & 0 & 0 \\ 0.46 & 0.54 & 0 & 0 & 0 & -0.23 & 0 & -0.45 & 0 & -0.32 \\ -0.25 & 0 & 0 & 0 & 0.24 & 0.33 & 0.44 & -0.42 & -0.33 & 0 \\ 0 & -1 & 0 & 0.31 & 0 & 0.30 & 0.39 & 0 & 0 & 0 \\ 0 & 0 & -0.17 & 0.34 & 0.24 & 0 & 0.42 & -0.47 & 0 & -0.36 \\ -0.17 & -0.15 & 0 & 0.38 & 0.26 & 0.36 & 0 & -0.27 & -0.21 & -0.21 \\ 0 & -0.11 & -0.16 & -0.22 & 0 & -0.22 & -0.28 & 0 & 0.48 & 0.52 \\ -0.15 & 0 & 0 & -0.38 & 0 & 0 & -0.47 & 0.56 & 0 & 0.44 \\ 0 & 0 & -0.22 & 0 & 0 & -0.34 & -0.43 & 0.59 & 0.41 & 0 \end{bmatrix}.$$

The values only keep two decimal places after the point for simplicity.

Step 3: Generate the initial phases of the nodes. The initial phases (θ_i) are generated randomly in the range $[0, \pi/2]$. Give three values for the parameters K_p , K_n, and α. The initial phases of the 10 nodes are expressed by a vector

$$\theta_i(0) = (0.46, 0.68, 0.02, 1.55, 0.26, 0.17, 0.58, 0.31, 0.77, 0.53).$$

The values only keep two decimal places after the point for simplicity.

Step 4: Update the phases of the nodes. Update the phases of the nodes by Eq. (10), the terminal condition is

$$|\theta_j(k) - \theta_j(k-1)| < \varepsilon, \quad j = 1, \ldots, N,$$

where $\varepsilon = 0.01$ in this paper.

Step 5: Detect the community structure. Obtain the distribution curve of phases based on network model Eq. (10), as shown in Fig. 4(b). To better detect the communities, the stable phases are multiplied by 100. If the two nodes i and j satisfy the condition $|\theta_j(k) - \theta_i(k)| < \tau$ (a relative small positive number), the nodes are regarded as in the same community. It can be found that 10 nodes are divided into three communities, and they are $C_1 = \{1,2,3\}$, $C_2 = \{4,5,6,7\}$, and $C_3 = \{8,9,10\}$, respectively. This means that the community structure can be detected successfully according to the different phases based on our proposed method.

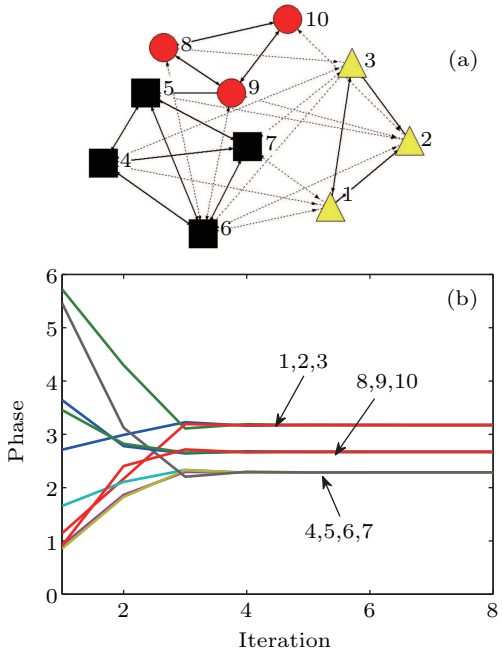


Fig. 4. (color online) Synthetic network. (a) The community structure detected by DTCDS. Triangle nodes are in the first community, square nodes are in the second community, and circle nodes are in the third community. (b) Phases of the nodes evolving with time.

5. Simulations

In this section, the proposed algorithm DTCDS is applied to both real world signed networks and synthetic signed networks. All the algorithms are coded in Matlab 2013a, the experiments have been executed on Intel(R) Xeon(R) CPU E5-2680 v2 @ 2.80 GHz, with 64 GB memory. The operating system is Windows Server 2008 R2 Enterprise.

5.1. Results on real world signed networks

The following real world signed networks are used to test the algorithm DTCDS.

U.S. supreme court (USC) network The USC network is about 9 justices' voting behavior on the U.S. supreme court for the 2006–2007 term.[39] The justices are Stevens (1), Ginsburg (2), Souter (3), Breyer (4), Kennedy (5), Alito (6), Roberts (7), Scalia (8), and Thomas (9). Figure 5(b) gives the community detection results based on algorithm DTCDS.

In Ref. [39], it was shown that the USC network was partitioned into three communities or two communities. The topology structure is shown in Fig. 5(a). The difficulty to detect the community comes from Justice Kennedy. In fact, he has positive links with all the justices of the conservative wing (Stevens, Ginsburg, Souter, Breyer), and has three positive links and one negative link with the liberal wing (Alito, Roberts, Scalia, Thomas). Therefore, there would be two kinds of results. Based on DTCDS, the USC network has two communities: $C_1 = \{1,2,3,4\}$, and $C_2 = \{5,6,7,8,9\}$, as shown in Fig. 5(b). The line of node 5 is found a little far from the communities, sometimes it can be considered as a single

community when there is no demand for the number of nodes in a community. The parameters are $K_p = 0.9$, $K_n = 0.03$, and $\alpha = 0.3$.

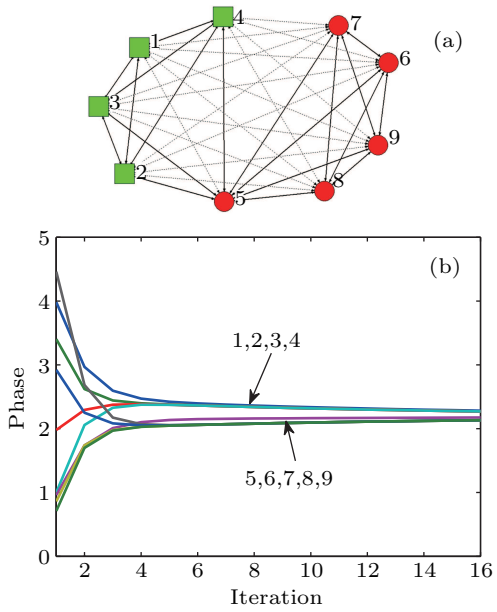


Fig. 5. (color online) USC network. (a) The community structure detected by DTCDS. Square nodes are in the first community, and circle nodes are in the second community. (b) Phases of the nodes evolving with time.

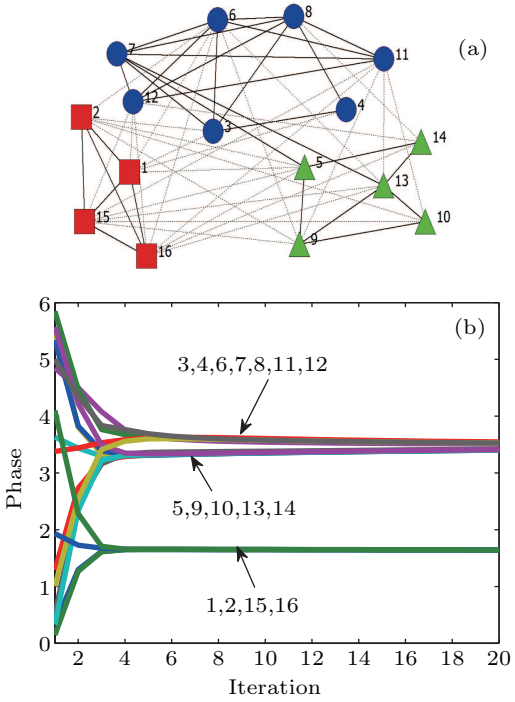


Fig. 6. (color online) GGS network. (a) The community structure detected by DTCDS. Square nodes are in the first community, triangle nodes are in the second community, and circle nodes are in the third community. (b) Phases of the nodes evolving with time.

Gahuku–Gama subtribes (GGS) network The GGS network came from a research about the cultures of New Guinea Highland, which is the political alliances and oppositions among 16 Gahuku–Gama subtribes.[40] Figure 6(a) shows that the political relationship among the subtribes can be positive or negative. Node 5 has two positive links with C_2, and one positive link with C_3. Node 13 has three positive links with C_2 and one positive link with C_3. It increases the difficulty to separate the communities C_2 and C_3. Through the DTCDS algorithm, 16 nodes are divided into three communities, and they are $C_1 = \{1,2,15,16\}$, $C_2 = \{5,9,10,13,14\}$, and $C_3 = \{3,4,6,7,8,11,12\}$. Comparing with Ref. [40], we can detect the communities successfully, which is shown in Fig. 6(b). The parameters are $K_p = 0.7$, $K_n = 0.001$, and $\alpha = 0.3$.

5.2. Results on synthetic signed networks

In this part, two kinds of synthetic networks are applied to test the efficiency of the DTCDS algorithm.

Small-benchmark synthetic network Two synthetic signed networks are widely used for tests,[29,32,33] which came from Ref. [11]. Figure 7 shows the topology structure of these two networks. They consist of 28 nodes, and have three communities. From Fig. 7(a), it can be found that the network has three communities. Figure 7(b) has 7 more negative links compared to Fig. 7(a), and it is unbalanced. The three communities in Fig. 7(a) are $C_1 = \{1,2,3,10,11,12,19,20,21,28\}$, $C_2 = \{4,5,6,7,13,14,15,16,22,23,24,25\}$, and $C_3 = \{8,9,17,18,26,27\}$. Based on the DTCDS algorithm, three communities are detected in the two networks, as shown in Fig. 8. The parameters are $K_p = 0.8$, $K_n = 0$, and $\alpha = 0.3$.

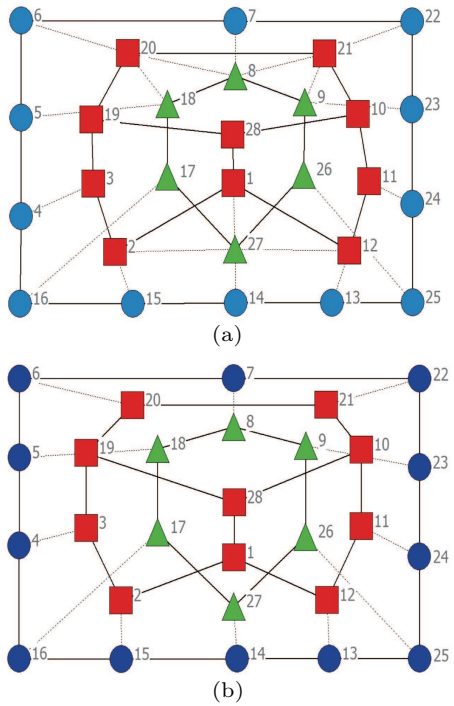


Fig. 7. (color online) The community structure detected by DTCDS. Square nodes are in the first community, circle nodes are in the second community, and triangle nodes are in the third community. (a) Synthetic network 1. (b) Synthetic network 2.

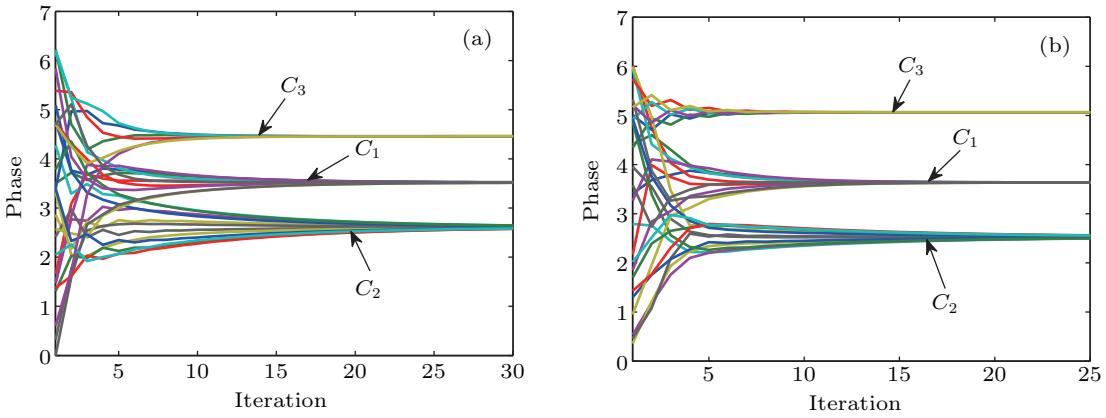


Fig. 8. (color online) Phases of the nodes evolving with time based on DTCDS. C_1 is the community with square nodes, C_2 is the community with circle nodes, and C_3 is the community with triangle nodes. (a) Synthetic network 1. (b) Synthetic network 2.

Large-scale synthetic network The method to generate synthetic signed networks is introduced by Refs. [4], [33], and [41]. The synthetic signed networks are denoted as SN(N,g,p,q), where N is the number of nodes, g is the number of communities, p is the ratio of the number of inter-community positive links to that of the intra-community positive links, and q is the probability of randomly flipping the signs of the existing links. Three different size synthetic networks are generated as follows:

$$\begin{aligned}
&\mathrm{SN}(500,10,p,q)=(5,10,15,30,40,60,70,80,90,100),\\
&\mathrm{SN}(1000,10,p,q)=(5,10,25,50,60,100,100,150,200,300),\\
&\mathrm{SN}(2000,15,p,q)=(5,10,15,40,60,80,100,100,100,130,160,200,300,350,350).
\end{aligned}$$

The results of the community detection for the signed networks SN$(500,10,0.1,0)$ (Fig. 9(a)), SN$(1000,10,0.1,0)$ (Fig. 9(b)), and SN $(2000,15,0.1,0)$ (Fig. 9(c)) are obtained. It can be found that SN$(500,10,0.1,0)$ has ten communities, SN$(1000,10,0.1,0)$ also has ten communities, and SN$(2000,15,0.1,0)$ has fifteen communities. The parameters are $K_{\mathrm{p}}=1$, $K_{\mathrm{n}}=0$, and $\alpha=0.2$.

In this paper, the normalized mutual information (NMI) is adopted to evaluate the efficiency of the proposed algorithm when the ground truth of the network is known.[42] Two existing algorithms based on the dynamic mechanism are applied to compare with the proposed algorithm. One is the dynamic evolutionary clustering algorithm for signed networks (DEC),[33] and the other is the dynamic-based algorithm for signed social networks (DBAS).[29] They are algorithms recently proposed for signed networks and gain good performances. The two algorithms are efficient and can run on large-scale signed networks. For the above three synthetic signed networks, different values of p and q are given as $p=0.1, p=0.3, p=0.5$ and $q=0, q=0.05, q=0.1, q=0.15, q=0.2$, respectively. Figure 10 shows the NMI corresponding to different p and q. The vertical axis NMI is the average of 50 runs.

From Fig. 10, it can be observed that the DTCDS algorithm outperforms DBAS and DEC when $q=0.1, 0.15$. DTCDS is relatively in balance with different q except for SN$(2000,15,0.5,q)$. The main reason is that the negative links would cause little effect on the detection performance with the same p, so parameter K_{n} should be adjusted to advance the value of NMI. The detection results of DBAS and DEC are always lower than DTCDS, and they only obtain higher detection accuracy than DTCDS at $q=0.2$, as shown in Fig. 10(i).

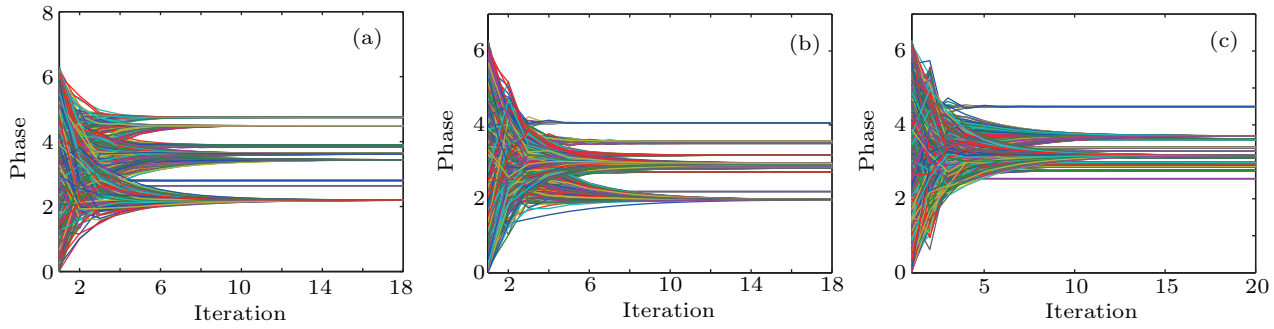


Fig. 9. (color online) Phases of the nodes evolving with time based on DTCDS: (a) SN$(500,10,0.1,0)$, (b) SN$(1000,10,0.1,0)$, (c) SN$(2000,15,0.1,0)$.

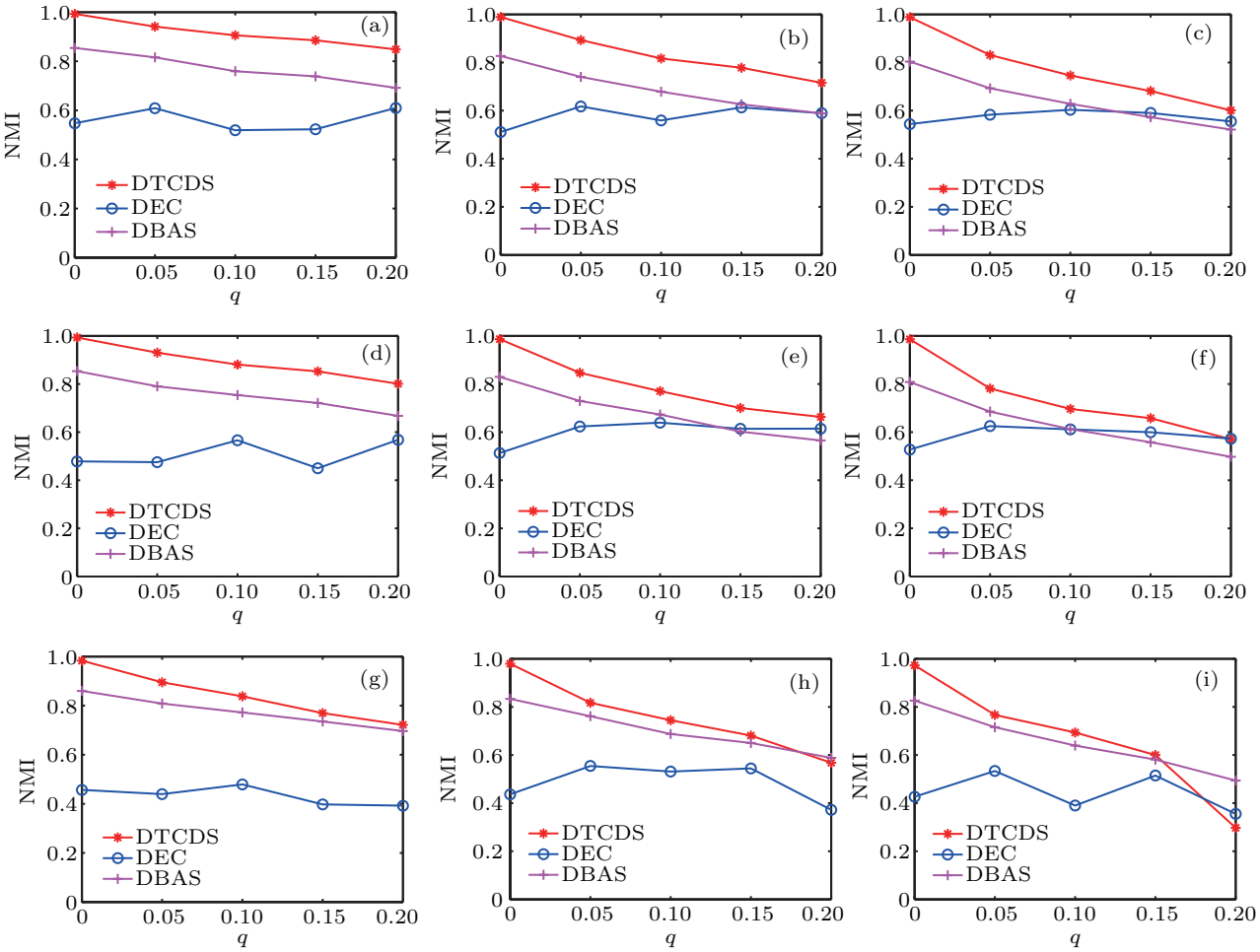


Fig. 10. (color online) Comparing between DTCDS and existing algorithms: (a) SN(500,10,0.1,q), (b) SN(500,10,0.3,q), (c) SN(500,10,0.5,q), (d) SN(1000,10,0.1,q), (e) SN(1000,10,0.3,q), (f) SN(1000,10,0.5,q), (g) SN(2000,15,0.1,q), (h) SN(2000,15,0.3,q), (i) SN(2000,15,0.5,q).

In addition, we compare the running time of these algorithms, as shown in Table 1. The running time is almost equal when the number of nodes is the same, so we only give three signed networks results, i.e., SN(500,10,0.1,0), SN(1000,10,0.1,0), SN(2000,15,0.1,0). It can be observed that the running time of our DTCDS algorithm is shorter than the DBAS algorithms. As the number of nodes increases, the running time is growing slowly, for instance, 1.3419 s for 500 nodes and 10.5984 s for 1000 nodes. But for the DBAS algorithm, the running time is growing quickly. DEC runs the fastest among the three algorithms, but it obtains the lowest NMI.

Table 1. Comparing running time (in units of s) among DTCDS, DEC, and DBAS for SN(N,C,0.1,0).

	SN(N,C,0.1,0)		
N	500	1000	2000
DEC[33]	0.6047	5.1945	47.9185
DTCDS	1.3419	10.5984	31.5768
DBAS[29]	6.0803	37.1417	466.9949

In this section, the simulations are performed for the synthetic signed networks with a large number of nodes and abundant communities, which increase the difficulty on community detection, even though the DTCDS algorithm can obtain a higher NMI. For instance, as shown in Fig. 10, the NMI can reach 0.9834 when the number of nodes is 2000 and the number of communities is 15.

6. Conclusion

In this paper, a discrete-time-model-based method is designed for signed networks community detection. The similarity between nodes is used to make aggregation speed and separating rate quicker. The positive similarity makes two nodes evolve together easier, and the negative similarity makes two nodes evolve far away. Community structure can be discovered as the phases are stable based on the dynamic mechanism of the proposed network model. This algorithm is appropriate not only for the small-scale signed networks but also for large-scale signed networks. Moreover, this algorithm can also be applied to the networks with more communities, and it has better accuracy and runs faster than the compared algorithms. The accuracy of this algorithm is closely related to the similarity of the nodes, new similarity can be found to improve the

accuracy, which needs further research.

References

[1] Chen J, Jiao L C, Wu J S and Wang X D 2010 *Nonlinear Analysis: Real World Applications* **4** 3045
[2] Jiang M S, Chen Y X and Chen L 2015 *arXiv:* 1502. 04380
[3] Newman M J 2004 *Phys. Rev. E* **69** 066133
[4] Wu L T, Ying X W, Wu X T, Lu A D and Zhou Z H 2012 *International Journal of Social Network Mining* **1** 91
[5] Doreian P and Mrvar A 1996 *Social Networks* **2** 149
[6] Lin D K 1998 *ICML* **98** 296
[7] Kernighan B W and Lin S 1970 *Bell System Technical Journal* **2** 291
[8] Pothen A, Simon H D and Liou K P 1990 *SIAM Journal on Matrix Analysis and Applications* **3** 430
[9] Newman M E J 2004 *Proc. Natl. Acad. Sci. USA* **1** 5200
[10] Newman M E J 2006 *Proc. Natl. Acad. Sci. USA* **23** 8577
[11] Yang B, Cheung W K and Liu J 2007 *Knowledge and Data Engineering, IEEE Transactions on* **10** 1333
[12] Chen Y, Wang X L, Yuan B and Tang B Z 2014 *Journal of Statistical Mechanics: Theory and Experiment* **2014** 03021
[13] Sun L H, Huang J B, Tian Y Q, Song Q B and Liu H L 2015 *Chin. Phys. B* **24** 551
[14] Wang Y and Cao J 2013 *Nonlinear Analysis: Real World Applications* **1** 842
[15] Wu J S, Jiao L C and Jin C 2012 *Phys. Rev. E* **1** 016115
[16] Wu J C, Wang X H and Jiao L C 2012 *Physica A* **3** 508
[17] Almendral J A, Leyva I, Li D, Sendiña-Nadal, Havlin S and Boccaletti S 2010 *Phys. Rev. E* **1** 016115
[18] Li D, Leyva I and Almendral J A 2008 *Phys. Rev. Lett.* **16** 168701
[19] Wu J S and Jiao Y 2014 *Chaos* **3** 033104
[20] Zhang L S, Gu W F, Hu G and Mi Y Y 2014 *Chin. Phys. B* **10** 626
[21] Xiang J, Hu K, Zhang Y, Hu T and Li J M 2015 *Europhys. Lett.* **111** 48003
[22] Xiang J, Hu T, Hu K, Tang Y N, Gao Y Y, Chai C H and Liu X J 2015 *Canadian Journal of Physics* **93** 418
[23] Esmailian P and Jalili M 2015 *Scientific Reports* **5** 14339
[24] Xiang J, Tang Y N, Gao Y Y, Zhang Y, Deng K, Xu X K and Hu K 2015 *Physica A* **432** 127
[25] Xiang J, Hu T, Zhang Y, Hu K, Li J M, Xu X K, Liu C C and Chen S 2016 *Physica A* **443** 451
[26] Chang Z C, Liu Y, Yu H T, and Huang R Y 2015 *Acta Phys. Sin.* **21** 218901 (in Chinese)
[27] Pizzuti C 2008 *Parallel Problem Solving from Nature - PPSN X, 10th International Conference*, September 13–17, 2008, Dortmund, Germany, p. 1081
[28] Liu C L, Liu J and Jiang Z Z 2014 *IEEE Transactions on Cybernetics* **12** 2274
[29] Wu J S, Zhang L, Li Y and Jiao Y 2016 *Physica A* **443** 568
[30] Lü L Y and Zhou T 2011 *Physica A* **6** 1150
[31] Xiang J, Hu K, Zhang Y, Bao M H, Tang L, Tang Y N, Gao Y Y, Li J M, Chen B Y and Hu J B 2016 *Journal of Statistical Mechanics: Theory and Experiment* **2016** 033405
[32] Chen J R, Hong Z M, Wang L N and Wu L 2014 *Chin. Phys. B* **11** 118903
[33] Chen J R, Wang H, Wang L N and Liu W W 2015 *Physica A* **447** 482
[34] Tang J, Lou T C and Kleinberg J 2012 *WSDM* 743
[35] Easley D and Kleinberg J 2010 *Networks, Crowds, and Markets: Reasoning about a Highly Connected World* (Cambridge: Cambridge University Press) p. 10
[36] Khadivi A, Rad A A and Hasler M 2010 *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on IEEE* 3777
[37] Wu J S, Hou Y T, Jiao Y, Li Y, Li X X and Jiao L C 2015 *Physica A* **433** 218
[38] Sideris T C 2013 *Ordinary Differential Equations and Dynamical Systems* (New York: Atlantis Press)
[39] Doreian P and Mrvar A 2009 *Social Networks* **31** 1
[40] Read K E 1954 *Southwestern Jouranal of Anthropopgy* **10** 1
[41] Zeng Y and Liu J 2015 *Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems* (Heidelberg: Springer) p. 259
[42] Wu F and Huberman B A 2004 *Eur. Phys. J. B* **2** 331