

Community Detection Based on Regularized Semi-Nonnegative Matrix Tri-Factorization in Signed Networks

Zhen Li¹ · Jian Chen^{1,2} · Ying Fu¹ · Guyu Hu¹ · Zhisong Pan¹  · Liangliang Zhang¹

Published online: 30 May 2017
© Springer Science+Business Media New York 2017

Abstract Community detection is a fundamental task in the social network analysis field, which is beneficial for many real-world applications such as recommendation systems and telephone fraud detection. Community detection in unsigned networks has been extensively studied, however, few works focus on community detection in signed networks. Under this background, we propose a framework based on regularized semi-nonnegative matrix tri-factorization which maps the signed network from high-dimensional space to low-dimensional space, such that the communities of the signed network can be derived. In addition, to improve the detection accuracy, we introduce a graph regularization to distribute the pair of nodes which are connected with negative links into different communities. The experimental results on both synthetic datasets and real-world datasets verify the effectiveness of the proposed method.

Keywords Community detection · Signed networks · Negative links · Regularized semi-nonnegative matrix tri-factorization

1 Introduction

Community structure is a prominent feature of social networks. Community detection is a fundamental task in the Social Network Analysis (SNA) field that helps us understand the organizational structure and functional components of a social network [1]. It also promotes other SNA tasks such recommendation systems [1]. Recommendation system is usually based on the users' past preferences, while community detection provides complementary information for it by deriving user communities. For example, if a community whose users have similar requirements and interests is detected, the telecommunications operator can recommend relevant services to a user according to the subscribed services of other users in the same community, so that better recommendations are made. Moreover, community detection has been applied to some real-world missions [2–4]. In the against telephone fraud missions, community detection helps to identify fraudster groups. According to the communication records and other online activities of a suspect, community detection helps the police to have a rough idea of the suspected fraudster group. Besides, in the against terrorism missions, community detection together with other SNA technics plays an important role in terrorist organization identification. Thus, community detection has attracted many attentions.

Various methods have been developed to capture the network structural characteristics during the past years, however, most of them can only deal with networks with only positive links, i.e., unsigned networks [5–7]. While many complex systems in the real world are represented by networks with both positive and negative links,

✉ Zhisong Pan
hotpzs@hotmail.com

¹ College of Command Information Systems, PLA University of Science and Technology, 88 Houbiaoying Road, Nanjing, 210007, China

² School of Information Science and Technology, Southeast University, Nanjing, 210007, China

i.e., signed networks. The relationships between users in some real-world systems are various, including not only positive relations like friendships, trusts and likes, but also negative relations like antagonisms, distrusts and dislike. For example, we have both contact list and black list in the telephone address book. These systems are modeled as signed networks, whose positive links represent positive relations and negative links represent negative relations [8]. Kunegis et al. [9] shows that negative links have a significant influence on the structure of the social network. However, the existing algorithms developed for unsigned networks cannot handle negative links [10], thus, mining signed networks is still a young field which needs further and separate efforts.

In the unsigned network, a community can be defined as a group of nodes that contact with each other more frequently than with those outside the group [11]. While in the signed network, not only the density of links but also the signs of links are considered. Two principles are developed: (1) within communities the links should be positive and dense and (2) between communities the links should be negative [12]. Community detection in the signed network is to find out antagonistic communities so that entities within the same community have a positive relationship with each other and entities in different communities have a negative relationship with each other. The existence of negative links exerts great challenges on mining communities in signed networks. The negative links may have a stronger influence on the community structure. In other words, a negative link indicates that the connected users have a high probability to belong to different communities while a positive link may not indicate a high probability of the same belongingness. Besides, the real-world systems are often complex, it is natural to have some positive links between communities and some negative links within communities. Thus, how to properly retain or disregard some positive and negative links to find out the communities accurately is a problem. Furthermore, most existing community detection algorithms designed for unsigned networks can only deal with positive links, methods for signed networks are needed. Some strategies based on graph partition have been proposed which solve two optimizations: maximizing the positive links in communities and maximizing the negative links across communities. However, the community detection accuracy is not high.

Keeping the intrinsic properties of communities in the signed network in mind, in this paper we propose a community detection method based on Regularized Semi-Nonnegative Matrix tri-Factorization (ReS-NMF) which considers both link density and link signs. Furthermore, in order to improve the detection accuracy, we introduce a graph regularization to keep all the negative links between

communities. The main contributions are summarized as follows:

1. We take into account of the mutual influences of positive and negative links in the signed network, build a latent space model to map the signed network from the high-dimensional space to the low-dimensional space, then derive the underlying communities in the new space.
2. We investigate the effects of the negative links on the structure of the network. In order to detect the communities more accurately, we introduce a graph regularization to distribute the pair of nodes which are connected by negative links into different communities. Furthermore, we enforce the sparsity constraints on the model to avoid overlapping distributions.
3. Experiments are carried out on both synthetic and real-world datasets, the results show the effectiveness of the proposed method.

The rest of this paper is organized below: Section 2 includes a brief review of the related work. The proposed framework is described in details in Section 3. Experimental results on both synthetic and real-world datasets are presented in Section 4. Finally, conclusions are provided in Section 5.

2 Related work

Some algorithms designed for the unsigned network have been extended to derive communities in the signed network. Wu et al. [13] investigated the impacts of the negative links on the network structure and examined the patterns in the spectral space of the adjacency matrix. Kunegis et al. [14] extended the spectral clustering algorithm to deal with negative links by developing the signed graph Laplacian matrix. But it had been proved to have a theoretical weakness when it was extended from 2-way to k -way ($k > 2$) clustering problem [15]. Anchuri et al. [16] extended modularity [17] to signed modularity to evaluate the quality of community partitions in the signed network. There are also multiobjective optimization based methods developed. In paper [16], Anchuri et al. proposed frustration to access community quality which was defined as the sum of the negative links inside the communities and positive links across communities. The communities were detected by minimizing the frustration. Doreian et al. [8] developed an objective function which is similar to the frustration except that it worked out the weighted sum. In the detection process, the nodes were first randomly distributed to k communities, then the above objective function was optimized by reallocating the nodes to more proper communities.

Liu et al. [18] designed two objective functions based on the signed similarity and the natural contradiction between positive and negative links, and developed a multiobjective evolutionary algorithm. The algorithm could handle large-scale signed networks and detect overlapping communities directly. Amelio et al. [19] proposed a multiobjective framework based on genetic algorithm, where the first objective was modularity maximization, and the second was frustration minimization. The multiobjective genetic algorithm evolved a population of candidate solutions by trying to obtain the best trade-off between high modularity and low frustration. Chiang et al. [15] proposed positive ratio cut and negative ratio association to detect communities in the signed network. The positive ratio cut objective was to minimize the positive links between communities relative to the community size, and the negative ratio association objective was to minimize the negative links within each community relative to its size. Some model-based methods also have been proposed. Chen [20] proposed a signed probabilistic mixture model to derive overlapping communities in the signed network. The model generated positive and negative links with different probabilities, and provided soft memberships which indicated the strength of a node belonging to a community. Yang et al. [21] proposed an FEC framework based on random walk model which could be applied to detect communities in both unsigned and signed networks. FEC adopted an agent-based heuristic method, it included two phases - (1) the FC phase computed the transition probability vector and sorted them for each row via random walks, and (2) the EC phase utilized a cutoff criterion to divide the transformed adjacency matrix into two block matrices, which corresponded to two subgraphs. One of the subgraphs was the identified community, and the other was to be processed recursively by the above two phases.

There are some shortcomings in the above methods: (1) The community detection accuracy needs to be improved, more effective and robust methods are needed. (2) Most of them deal with positive and negative links separately by solving the above mentioned objectives, neglecting the mutual influence of positive and negative links. In this paper, we develop a variant of Non-negative Matrix Factorization (NMF) to approximate the adjacency matrix of the signed network with low-dimensional matrices, so that communities can be detected in the latent space. NMF [22] is a popular matrix factorization method where all the elements are restricted to be nonnegative, it performs well in the SNA field [11, 23]. As NMF is limited to deal with matrix with nonnegative matrices, Ding et al. [24] proposed a semi-NMF algorithm which did not require the matrix to be nonnegative, extended the application of NMF ideas. Semi-NMF has been applied to address issues such as nonlinear

unmixing of hyperspectral data analysis [25] and motion segmentation [26], achieving good performances. Based on the idea of semi-NMF, we propose semi-Nonnegative Matrix Tri-Factorization (semi-NMTF) to factorize the adjacency matrix of the signed network, map the original network to a low-dimensional space where the underlying structure information is more explicit and clear. Moreover, we introduce a graph regularization to enhance the detection accuracy. And we develop the iterative algorithm to discover communities in the signed network.

3 ReS-NMF framework for community detection in signed networks

In this section, we introduce the definition and main symbols used in this paper, and then describe the proposed method in details.

3.1 Notations

The symbols used in this paper are shown in Table 1. Given a signed network with n nodes which are grouped into k communities. The topology is represented by the adjacency matrix $A \in R^{n \times n}$ where $A_{ij} = 1$, $A_{ij} = -1$, and $A_{ij} = 0$ denote a positive link, a negative link and no link between node i and node j respectively. The matrix $A^p \in R^{n \times n}$ represents a positive network where $A_{ij}^p = 1$ denotes a positive link between node i and node j and $A_{ij}^p = 0$ otherwise. Similarly, $A^n \in R^{n \times n}$ is used to represent a negative network where $A_{ij}^n = 1$ denotes a negative link and $A_{ij}^n = 0$ otherwise. It is easy to convert one representation into the other representation with the following rules: $A = A^p - A^n$, $A^p = \frac{|A|+A}{2}$, $A^n = \frac{|A|-A}{2}$. The matrix $H \in R_+^{n \times k}$ is the community indicator whose element H_{ij} denotes the probability that node i belongs to j -th community. The matrix $S \in R^{k \times k}$ is the association matrix indicating the relationships between communities.

Table 1 Symbol definitions

Symbols	Explanations	Size
n	number of nodes	–
k	number of communities	–
A	signed network’s adjacency matrix	$n \times n$
A^p	positive network’s adjacency matrix	$n \times n$
A^n	negative network’s adjacency matrix	$n \times n$
S	association matrix	$k \times k$
H	community indicator matrix	$n \times k$
γ_1, γ_2	the community indicator matrix	–

3.2 Problem statement

With the notations given above, the community detection problem in the signed network is formally defined as follows.

Problem 1 Given a signed network $G = (V, E)$ with adjacency matrix A , positive adjacency matrix A^p and negative adjacency matrix A^n , find k communities $\{C_1, \dots, C_k\}$ with maximal positive links within communities and maximal negative links across communities, i.e., $\max \sum_{ij} A_{ij}^p \delta(c_i, c_j)$ and $\max \sum_{ij} A_{ij}^n (1 - \delta(c_i, c_j))$, where $\delta(c_i, c_j)$ is the Kronecker delta function which is 1 if node i and node j belong to the same community, 0 otherwise.

The problem is NP-hard, we attempt to solve it by building a latent space model. In the model, a variant of semi-NMF to approximate the adjacency matrix is developed, which maps the network into a low-dimensional space to make the underlying structure information explicit. Besides, we introduce a graph regularization to minimize the number of negative links inside each community, so that negative links between communities are maximized.

3.3 Proposed method

In this subsection, the proposed framework Res-NMF is presented in details.

In paper [24], the 2-factor semi-NMF, i.e., $A \approx FH^T$, was proposed where only H is restricted to be nonnegative, it has been proved useful to capture node similarity and mutual influences between links. However, 2-factor factorization only captures limited features, while 3-factor factorization can capture relations between communities and analyze the network more precisely. Then we propose semi-NMTF to approximate the adjacency matrix of the signed network, i.e., $A \approx HSH^T$, where $H \in R_+^{n \times k}$ is the community indicator and $S \in R^{k \times k}$ is the community relationship indicator. The element H_{ij} is nonnegative which denotes the likelihood that node i belongs to the j -th community. If node i belongs to the j -th community, $H_{ij} = 1$, otherwise, $H_{ij} = 0$. The positive element S_{ij} of S denotes the positive relationship between i -th community and j -th community, similarly, the negative element denotes the negative relationship. The proposed semi-NMTF is written as follows:

$$\min_{H \in R^{n \times k}, S \in R^{k \times k}} \|A - HSH^T\|_F^2 \quad s.t. H_{ij} \in \{0, 1\}, \quad (1)$$

where A is the adjacency matrix of the signed network, $\|\cdot\|_F$ denotes the matrix Frobenius norm which is defined as $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2}$.

Besides, the negative links have substantial influence on the community structure. In order to keep negative links between communities, we introduce a graph regularization to minimize the number of negative links inside each community relative to its size. In particular, the regularization is defined as follows:

$$\min \sum_{j=1}^k \frac{h_j^T A^n h_j}{h_j^T h_j}, \quad (2)$$

where A^n is the adjacency matrix of a negative network, h_j is the j -th vector of matrix H and $H = [h_1; \dots; h_k]$. The regularization enforces the nodes connected with negative links to be distributed into different communities, then the goal of keeping negative links between communities dense can be achieved.

Because the above optimization problems are NP-hard [1], in this paper, we solve this problem with some relaxations. Suppose the indicator H has the following form:

$$H_{ij} = \begin{cases} \sqrt{1/\pi_j}, & \text{node } i \in \pi_j \\ 0, & \text{otherwise} \end{cases}, \quad (3)$$

where $|\pi_j|$ is the number of nodes in the j -th community. Then the objective function (1) can be written as follows:

$$\min_{H \in R^{n \times k}, S \in R^{k \times k}} \|A - HSH^T\|_F^2 \quad s.t. H_{ij} \geq 0, \forall i, j. \quad (4)$$

Moreover, we observe that $tr(H^T A^n H) = \sum_{j=1}^k \frac{h_j^T A^n h_j}{h_j^T h_j}$. Then the solution to the optimization problem (2) can be obtained by addressing the following optimization problem:

$$\min_{H \in R^{n \times k}} tr(H^T A^n H) \quad s.t. H_{ij} \geq 0, \forall i, j. \quad (5)$$

Furthermore, in order to make the node try its best to belong to only one community, we introduce l_1 norm $\sum_{j=1}^n \|H(j, :)\|_1^2$ to control the sparsity of the model.

Combining the regularizations into semi-NMTF, the final optimization objective we develop is defined as Formula (6):

$$\min_{H \in R^{n \times k}, S \in R^{k \times k}} \|A - HSH^T\|_F^2 + \gamma_1 tr(H^T A^n H) + \gamma_2 \sum_{j=1}^n \|H(j, :)\|_1^2 \quad s.t. H_{ij} \geq 0, \forall i, j, \quad (6)$$

where γ_1, γ_2 are parameters to control the weights of the regularization terms.

3.4 The updating rules

In this section, we discuss how to solve the optimization problem (6). The optimal solution can be achieved using the iterative update algorithm we develop. We define a matrix $N_{k \times k}$ whose elements are 1, let α be the Lagrange matrix

for constraint $H \geq 0$, then the Lagrangian function can be written as:

$$\begin{aligned}
 L &= \|A - HSH^T\|_F^2 + \gamma_1 \text{tr}(H^T A^n H) \\
 &\quad + \gamma_2 \sum_{j=1}^n \|H(j, :)\|_1^2 - \alpha H \\
 &= \text{tr}((A - HSH^T)(A - HSH^T)^T) \\
 &\quad + \gamma_1 \text{tr}(H^T A^n H) + \gamma_2 \text{tr}(H^T H H^T) - \alpha H \\
 &= -2\text{tr}(AHS^T H^T) + \text{tr}(HSH^T HS^T H^T) - \alpha \\
 &\quad + \text{tr}(AA^T) + \gamma_1 \text{tr}(H^T A^n H) + \gamma_2 \text{tr}(H^T H H^T)
 \end{aligned} \tag{7}$$

Then the derivatives of L with respect to H and S are:

$$\frac{\partial L}{\partial S} = -2H^T AH + 2H^T HSH^T H, \tag{8}$$

$$\frac{\partial L}{\partial H} = -4\Phi_1 + 4H\Phi_2 + 2\gamma_1 A^n H + 2\gamma_2 HN - \alpha, \tag{9}$$

where $\Phi_1 = AHS^T$, $\Phi_2 = SH^T HS$. Using the KKT conditions that $\alpha H = 0$, and according to methods in [16], the updating rules are shown as follows:

$$S \leftarrow (H^T H)^{-1} H^T AH (H^T H)^{-1}, \tag{10}$$

$$H_{ij} \leftarrow H_{ij} \sqrt{\frac{(2\Phi_1^+ + 2H\Phi_2^-)_{ij}}{(2\Phi_1^- + 2H\Phi_2^+ + \gamma_1 A^n H + \gamma_2 HN)_{ij}}}, \tag{11}$$

where $X_{ij}^+ = (|X_{ij}| + X_{ij})/2$, $X_{ij}^- = (|X_{ij}| - X_{ij})/2$.

Algorithm 1 Res-NMF community detection in signed social network

Input: number of communities k , adjacency matrices: A and A^n , parameters: γ_1 and γ_2

Output: communities C_1, \dots, C_k

1: Initialize $H \geq 0, S$

2: **while** not convergent **do**

3: update S according to Formula (10)

4: update H according to Formula (11)

5: **end while**

6: obtain C_j from H by finding the max value and its index in each row of H .

With the above updating rules, the optimization algorithm is presented in Algorithm 1. Note that we update one matrix while fixing the other matrices in each step and the iterative process is stopped if these cluster matrices converge or the number of iterative reaches a given threshold.

4 Experiments

In this section, we evaluate the effectiveness of the proposed algorithm in detecting communities from signed social networks. Experiments are carried out on both synthetic and real-world datasets.

4.1 Evaluation measures and baseline methods

In this paper, we apply Normalized Mutual Information (NMI) [27], Purity [28] and Signed Modularity [29] to evaluate the performances of different algorithms. Both NMI and Purity adopt the ground truth as a baseline, the values of the two measures range from 0 to 1 and higher value means better performance. Signed Modularity measures the quality of the derived community structure from the perspective of how far it deviates from a random network, the larger value it is, the stronger the strength of community structure is.

NMI estimates the similarity between the ground truth communities and the detected. Let C and G denote our partition the ground truth respectively, F be the confusion matrix whose element F_{ij} is the number of nodes in community i of the partition C that are also in the community j of the partition G , then NMI is defined as follows:

$$NMI = \frac{-2 \sum_{i=1}^{f_C} \sum_{j=1}^{f_G} F_{ij} \log(F_{ij}n / F_i \cdot F_j)}{\sum_{i=1}^{f_C} F_i \cdot \log(F_i/n) + \sum_{j=1}^{f_G} F_j \log(F_j/n)}, \tag{12}$$

where f_C (f_G) is the number of communities in the partition C (G), F_i (F_j) is the sum of the elements of F in row i (column j), and n is the number of nodes. If $C = G$, $NMI = 1$. If C and G are completely different, $NMI = 0$.

Purity is measured by computing the number of nodes assigned with the same labels in all communities. Let $G = G_1, \dots, G_k$ be the set of communities in the ground truth and $C = C_1, \dots, C_t$ be the t communities extracted by different approaches. Purity is defined as:

$$Purity = \frac{1}{n} \sum_{i=1}^t \max_j |C_i \cap G_j|, \tag{13}$$

where n is the number of nodes.

The original Modularity Q is designed for unsigned networks [17], it is defined as the probability of having links falling within communities in the networks minus the expected probability in an equivalent network with the same number of nodes, and links placed at random preserving the nodes degree. However, Q cannot deal with negative links. Gmez et al. [29] developed Signed Modularity as follows:

$$Q_s = \frac{1}{2w^+ + 2w^-} \sum_i \sum_j \left[w_{ij} - \left(\frac{w_i^+ w_j^+}{2w^+} - \frac{w_i^- w_j^-}{2w^-} \right) \right], \delta(c_i, c_j) \tag{14}$$

where w_i^+ (w_j^+) denotes the sum of all positive weights of node i (j), and w_i^- (w_j^-) represents the sum of all negative weights of node i (j). Besides w^+ (w^-) denotes total positive (negative) strength. If nodes i and j are in the same community, $\delta(c_i, c_j) = 1$, otherwise, $\delta(c_i, c_j) = 0$. Normally

the larger the value of Q_s is, the better the separation of the community structure is.

In order to measure the effectiveness of ReS-NMF framework, three existing community detection methods for signed networks, i.e., the signed spectral clustering algorithm [14] and FEC algorithm [21], and MEAs-SN algorithm [18] are compared in our experiments.

In paper [14], a signed variant of the graph Laplacian is proposed, then the signed spectral clustering algorithm is developed based on it. In the method, the top k eigenvectors of signed Laplacian matrix associated with the signed network are calculated, then k -means is applied to the eigenvectors to obtain disjoint partitions of the network.

FEC is proposed in [21], it regards the sign and the density of relations as the clustering attributes. The method adopts an agent-based approach for formulating the communities. It contains two main phases: (1) the FC phase transforms the adjacency matrix to compute their transition probability vector and sorts them for each row and (2) the EC phase applies a cutoff criterion to the transformed adjacency matrix and divides it into two block matrices, which correspond to two subgraphs. One of these subgraphs is the identified community, and the other is the matrix to be processed recursively following the above mentioned two phases.

MEAs-SN is a multiobjective evolutionary algorithm proposed in [18], it models community detection problem as a multiobjective problem considering both link density and signs. In the method, signed similarity is extended from the original similarity based on the social balance theory so that it can deal with the negative links. Two objectives are optimized together, one is target to put all positive links in communities while the other is target to keep all negative links between communities.

4.2 Experimental results on synthetic datasets

We first validate ReS-NMF on synthetic datasets. We apply the LFR benchmark to generate synthetic networks for community detection. The LFR benchmark is declared to support some main statistical properties of the real networks. The details of the synthetic networks are shown in Table 2, where DS denotes the datasets, N denotes the number of nodes, PL and NL denote the number of positive links and negative links respectively, and C denotes the number of communities.

The scores of NMI, Purity and Signed Modularity for different methods on five synthetic datasets are presented in Table 3. Note that all experiment results are averaged over 100 runs. From the results shown in the table, some conclusions can be drawn.

It is evident that the proposed ReS-NMF algorithm performs best among all the methods on all the synthetic datasets. For example, in terms of *Network 1*, our method can gain more than 78%, 44% and 2% improvements on NMI respectively, over 4%, 5% and 8% improvements on Purity respectively, and nearly 17%, 9% and 9% improvements on Signed Modularity, compared with signed spectral clustering algorithm, FEC algorithm and MEAs-SN algorithm. Furthermore, let us pay attention to *Network 2* and *Network 3* whose generator parameters are the same apart from μ . With higher μ , the network structure is more fuzzy and the communities are more difficult to detect. It can be observed that our method has a better performance than the others and the performance does not deteriorate as μ increases. *Network 4* and *Network 5* have more nodes and links than the other networks, and the structures are more complex. It can be seen that on the three evaluation metrics, ReS-NMF also performs better than the other methods. The reasons may be as follows: (1) The ReS-NMF framework maps the signed networks into a low-dimensional space, makes the implicit structures clear in the latent space. (2) The graph regularization contributes to distribute the nodes connected with negative links into correct communities, reducing the wrong partitions.

4.3 Experimental results on real-world datasets

In this subsection, we evaluate ReS-NMF method on real-world datasets. Two widely used real-world signed networks, i.e., the Slovene Parliamentary Party network and the Gahuku-Gama Subtribes network, are employed to validate the effectiveness of our method. Moreover, although ReS-NMF is designed for signed networks, it can also be used to unsigned networks. Therefore, two popular unsigned benchmark networks, the Zachary karate club network and the Dolphin network, are tested.

The Slovene Parliamentary Party network describes the relations between ten parties of the Slovene Parliamentary in 1994 [30]. Positive links mean that two parties' activities are similar, while negative links mean their activities are

Table 2 Details of the synthetic networks

DS	N	PL	NL	C
Network 1	500	573	140	14
Network 2	1000	319	3146	36
Network 3	1000	620	2587	34
Network 4	2500	10452	2340	81
Network 5	3000	5684	565	26

Table 3 NMI, Purity and Signed Modularity for four methods on five datasets, bold emphasis represents the best results

Datasets	Methods	NMI	Purity	Q_s
Network 1	Signed spectral clustering	0.3294±0.059774	0.7206±0.045486	0.3036±0.046860
	FEC	0.5704±0.010482	0.7120±0.025043	0.3205±0.047126
	MEAs-SN	0.5815±0.033578	0.6940±0.025486	0.3252±0.059202
	ReS-NMF	0.5956±0.017464	0.7520±0.048744	0.3522±0.001772
Network 2	Signed spectral clustering	0.8241±0.045210	0.8078±0.045585	0.3892±0.042913
	FEC	0.8190±0.009725	0.7520±0.040748	0.2339±0.023427
	MEAs-SN	0.8080±0.014999	0.7300±0.024237	0.2182±0.021301
	ReS-NMF	0.8333±0.015231	0.8129±0.040693	0.4458±0.019375
Network 3	Signed spectral clustering	0.8278±0.032434	0.8348±0.048010	0.4542±0.037864
	FEC	0.8347±0.028379	0.8110±0.032763	0.4566±0.049632
	MEAs-SN	0.8537±0.014439	0.8150±0.042761	0.4919±0.025808
	ReS-NMF	0.8662±0.057148	0.8678±0.057879	0.5155±0.041690
Network 4	Signed spectral clustering	0.8719±0.035888	0.8031±0.042813	0.4474±0.034027
	FEC	0.8397±0.024441	0.7312±0.035280	0.3561±0.028616
	MEAs-SN	0.8071±0.039360	0.7624±0.028661	0.3995±0.016607
	ReS-NMF	0.8768±0.007280	0.8565±0.016792	0.4765±0.015263
Network 5	Signed spectral clustering	0.3259±0.021	0.7155±0.025942	0.2221±0.029524
	FEC	0.6234±0.026230	0.5683±0.026552	0.2572±0.050847
	MEAs-SN	0.6319±0.030079	0.5643±0.059696	0.2482±0.015226
	ReS-NMF	0.6540±0.023388	0.7754±0.043600	0.2976±0.010135

dissimilar. Figure 1a shows the original topological structure, and the community structure derived by ReS-NMF is shown in Fig. 1b. As can be seen, all parties are separated into two opponent communities. Note that this result is the same as that given by Kropivnik and Mrvar [30].

The Gahuku-Gama Subtribes dataset contains 16 tribal groups of the Eastern Central Highlands in New Guinea, the relations between tribal groups can be friendly or antagonistic [31] which are denoted by positive or negative links respectively. The 16 tribal groups are distributed into 3 communities in reality. Figure 2a presents the original

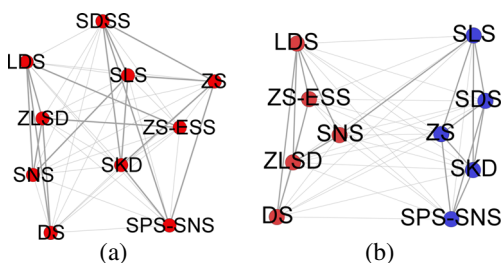


Fig. 1 a Topological structure of the Slovene Parliamentary Party network. b Community structure obtained by ReS-NMF. In the figures, the thick edges between nodes represent positive links while the thin edges represent negative links. Different node colors in b correspond to different communities

topological structure, and the communities obtained by ReS-NMF are shown in Fig. 2b. The community detection result is identical to that in reality.

Then we further evaluate the effectiveness of ReS-NMF on unsigned networks. Figure 3 shows the communities found by ReS-NMF for the Zachary karate club network and the Dolphin network. As can be seen, the Zachary karate club network and the dolphin network are divided into two and four communities respectively. The community structures found are meaningful and identical to those in reality.

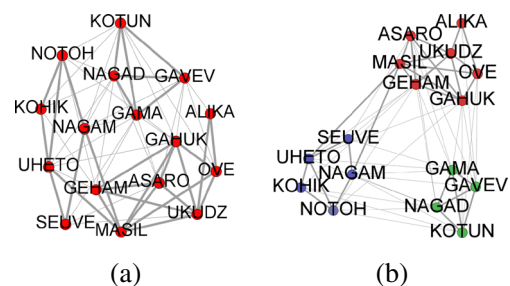


Fig. 2 a Topological structure of the Gahuku-Gama Subtribes network. b Community structure obtained by ReS-NMF. In the figures, the thick edges between nodes represent friendship while the thin edges represent antagonism. Different node colors in b correspond to different communities

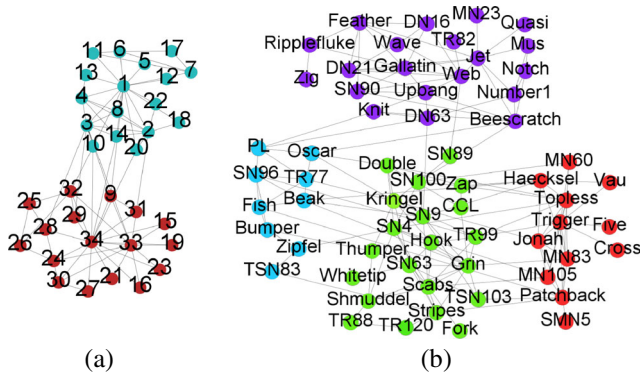


Fig. 3 **a** Community structure found by ReS-NMF for Zachary karate club network. **b** Community structure obtained by ReS-NMF for Dolphin network. Different node colors correspond to different communities

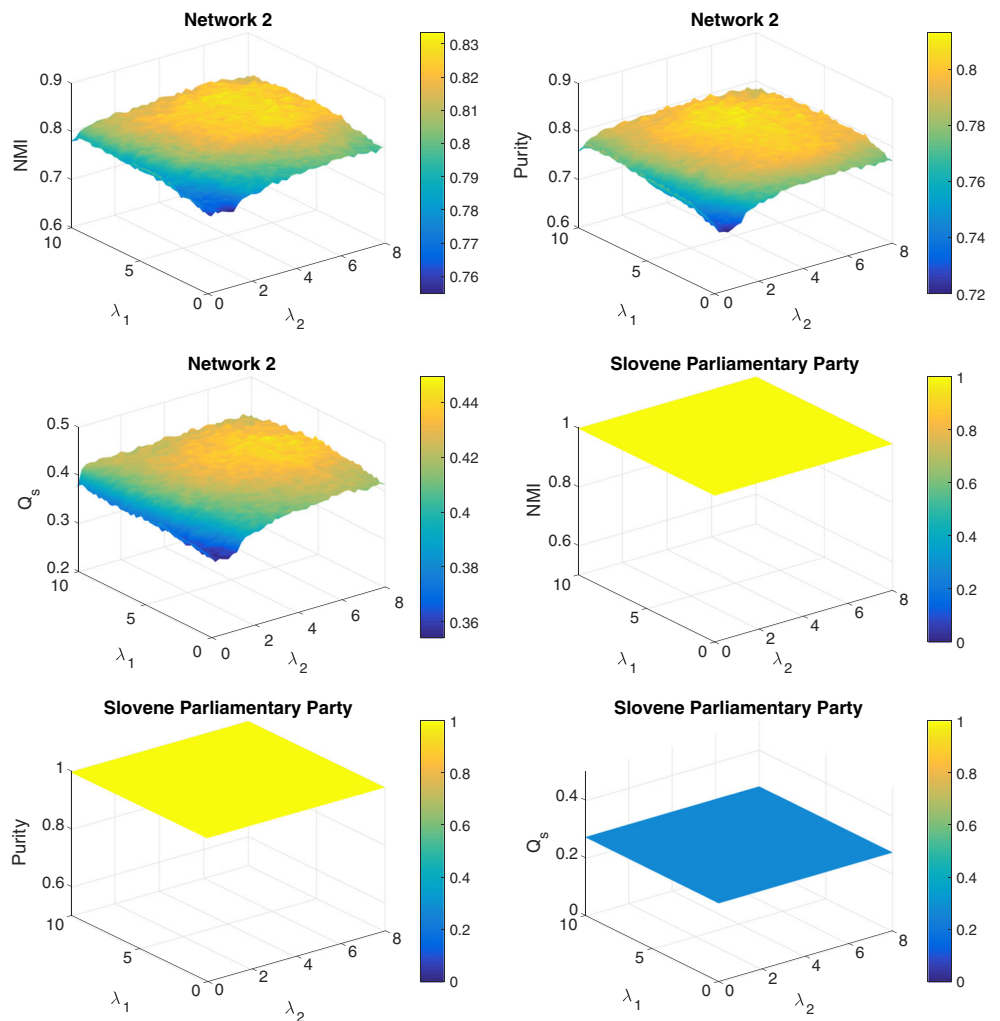
4.4 Parameter study

In this section, we evaluate ReS-NMF in terms of its sensitivity to the regularization parameters γ_1 and γ_2 . The

datasets we use include *Network 2* and the Slovene Parliamentary Party network. We systematically test the effect of γ_1 and γ_2 , where γ_1 increases from 0.1 to 10 and γ_2 increases from 0.1 to 8. For each combination of the two parameters, 50 independent runs of ReS-NMF are conducted, and the averaged NMI, Purity and Signed Modularity are shown in Fig. 4.

We note that γ_2 is fixed, with the increase of γ_1 , the performance of ReS-NMF on *Network 2* first increases. When $\gamma_1 > 2$, ReS-NMF keeps a relatively stable performance, and when $\gamma_1 > 8$, the values of NMI, Purity and Signed Modularity drop a little. With the fixed γ_1 , we observe that with the increase of γ_2 , the performance on *Network 2* first increases correspondingly, when $\gamma_2 > 4$, the performance is relatively stable. It can be observed that our method is relatively stable on γ_1 and γ_2 across a wide range. Then we test ReS-NMF on Slovene Parliamentary Party network, across all the values of γ_1 and γ_2 in the experiment, the values of NMI and Purity keep to be 1, and Signed Modularity also has stable score. This is because without the help of the regularizations, semi-NMFT can extract the

Fig. 4 Parameter sensitivity evaluation on *Network 2* dataset and Slovene Parliamentary Party network



communities of this network accurately. From Fig. 4 we can see that ReS-NMF achieves relatively stable and outstanding performances, and is not sensitive to the regularization parameters.

5 Conclusions

In this paper, we develop a ReS-NMF community detection algorithm for signed networks. The ReS-NMF algorithm captures communities by mapping the signed network into a latent space, moreover, improves the detection accuracy by studying the effects of negative links and enforcing the sparsity constraints on the indicator matrix. Experiments on both synthetic data and real-world data all demonstrate that ReS-NMF has better performance and regularization parameter sensitivity compared with other algorithms. For the future work, we plan to design faster algorithm for regularized matrix factorization.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No.61473149, No.61301159), Natural Science Foundation of Jiangsu Province, China (No. BK20140073) and China Postdoctoral Science Foundation (No.2015M571635, No. 2016T90404).

References

1. Tang L, Liu H Community detection and mining in social media. www.morganclaypool.com
2. Sageman M (2004) Understanding terror networks. University of Pennsylvania Press, Philadelphia
3. Xu J, Chen H (2005) CrimelNet explorer: a framework for criminal network knowledge discovery. *ACM Trans Inf Syst* 23(2):201–226
4. Chen H, Chung W, Xu JJ et al (2004) Crime data mining: a general framework and some examples. *Computer* 37(4):50–56
5. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(3):75–174
6. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci* 99(12):7821–7826
7. Greene D, Cunningham P (2013) Producing a unified graph representation from multiple social network views. In: Proceedings of the 5th annual ACM web science conference, pp 118–121
8. Doreian P, Mrvar A (1996) A partitioning approach to structural balance. *Soc Netw* 18(2):149–168
9. Kunegis J, Preusse J, Schwagereit F (2013) What is the added value of negative links in online social networks? In: International conference on world wide web, pp 727–736
10. Tang J, Aggarwal C, Liu H (2012) A survey of mining signed networks in social media. In: IEEE transactions on knowledge and data engineering
11. Pei Y, Chakraborty N, Sycara K (2015) Nonnegative matrix tri-factorization with graph regularization for community detection in social networks. In: International conference on artificial intelligence. AAAI Press
12. Chiang KY, Hsieh CJ, Natarajan N et al (2014) Prediction and clustering in signed networks: a local to global perspective. *J Mach Learn Res* 15(1):1177–1213
13. Wu L, Ying XW, Wu XT et al (2011) Spectral analysis of kbalance signed graphs. In: Advances in knowledge discovery and data mining, pp 1–12
14. Kunegis J, Schmidt S, Lommatzsch A et al (2010) Spectral analysis of signed graphs for clustering, prediction and visualization. In: Siam international conference on data mining, SDM 2010, pp 975–978
15. Chiang K-Y, Whang JJ, Dhillon IS (2012) Scalable clustering of signed networks using balance normalized cut. In: Proceedings of the 21st ACM international conference on information and knowledge management. ACM
16. Anchuri P, Magdon-Ismael M (2012) Communities and balance in signed networks: a spectral approach. In: 2012 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 235–242
17. Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69(2):026113
18. Liu C, Liu J, Jiang Z (2014) A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks. *IEEE Trans Cybern* 44(12):2274–2287
19. Amelio A, Pizzuti C (2013) Community mining in signed networks: a multiobjective approach. In: IEEE/ACM international conference on advances in social networks analysis and mining, pp 95–99
20. Chen Y, Wang X-l, Yuan B (2013) Overlapping community detection in signed networks. arXiv:1310-4023
21. Yang B, Cheung WK, Liu J (2007) Community mining from signed social networks. *IEEE Trans Knowl Data Eng* 19(10):1333–1348
22. Lee DD, Sebastian Seung H (2001) Algorithms for nonnegative matrix factorization. In: Advances in neural information processing systems, pp 556–562
23. Wang F, Li T, Wang X et al (2011) Community discovery using nonnegative matrix factorization. *Data Min Knowl Disc* 22(3):493–521
24. Chris D, Tao LI, Jordan MI (2010) Convex and seminonnegative matrix factorizations. In: IEEE transactions on pattern analysis and machine intelligence
25. Yokoya N, Chanussot J, Iwasaki A (2014) Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization. *IEEE Trans Geosci Remote Sens* 52(2):1430–1437
26. Mo Q, Draper BA (2012) Semi-nonnegative matrix factorization for motion segmentation with missing data. *Lect Notes Comput Sci* 7578(1):402–415
27. Strehl A, Ghosh J (2003) Cluster ensemble—a knowledge reuse framework for combining multiple partitions. *J Mach Learn* 3:583–617
28. Lin W, Kong X, Yu PS et al (1948) Community detection in incomplete information networks. *N Engl J Med* 319(5):313–324
29. Gmez S, Jensen P, Arenas A (2009) Analysis of community structure in networks of correlated data. *Phys Rev E* 80(1):016114
30. Kropivnik S, Mrvar A (1996) An analysis of the slovene parliamentary parties network. In: Proceedings of develop data analysis, pp 209–216
31. Read KE (1954) Cultures of the central highlands, New Guinea. *Southwest J Anthropol* 1:1–43