



# Multi-constraint non-negative matrix factorization for community detection: orthogonal regular sparse constraint non-negative matrix factorization

Zigang Chen<sup>1,2</sup> · Qi Xiao<sup>1</sup> · Tao Leng<sup>2</sup> · Zhenjiang Zhang<sup>1</sup> · Ding Pan<sup>1</sup> · Yuhong Liu<sup>3</sup> · Xiaoyong Li<sup>4</sup>

Received: 17 February 2023 / Accepted: 22 February 2024  
© The Author(s) 2024

## Abstract

Community detection is an important method to analyze the characteristics and structure of community networks, which can excavate the potential links between nodes and further discover subgroups from complex networks. However, most of the existing methods only unilaterally consider the direct link topology without comprehensively considering the internal and external characteristics of the community as well as the result itself, which fails to maximize the access to the network information, thus affecting the effectiveness of community detection. To compensate for this deficiency, we propose a new community detection method based on multi-constraint non-negative matrix factorization, named orthogonal regular sparse constraint non-negative matrix factorization (ORSNMF). Based on the network topology, the ORSNMF algorithm models the differences of the outside of the community, the similarities of the nodes inside the community, and the sparseness of the community membership matrices at the same time, which together guides the iterative learning process to better reflect the underlying information and inherent attributes of the community structure in order to improve the correct rate of dividing subgroups. An algorithm with convergence guarantee is also proposed to solve the model, and finally a large number of comparative experiments are conducted, and the results show that the algorithm has good results.

**Keywords** Community detection · Non-negative matrix factorization (NMF) · Community dissimilarity · Node similarity

## Introduction

In real life, community networks are ubiquitous, and they consist of highly interconnected entities from the natural world and society [1]. These networks typically share a common characteristic: closely related or similar nodes within a community network often belong to the same category, while

nodes with weaker connections or opposing characteristics tend to belong to different categories. Based on these characteristics, we can extract valuable feature information from community networks to achieve clustering effects, ultimately applying them in relevant fields. For instance, in the field of bioinformatics, we can partition biological molecules to discover those with similar structures and functions or identify protein complexes within protein–protein interaction (PPI) networks. In the realm of social media, we can perform opinion analysis, recommend products to users, and find potential friends [2]. Therefore, analyzing these community networks holds significant importance. However, community detection is an effective means to analyze the characteristics and structure of community networks, and important features such as explicit and implicit of community networks are mined by optimizing the community detection method to achieve the effect of effective division of community structure, so the study of community detection is of great significance to understand the deeper characteristics and functions of the network.

✉ Zigang Chen  
chenzg@cqupt.edu.cn

✉ Tao Leng  
lengtao@iie.ac.cn

<sup>1</sup> School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

<sup>2</sup> Intelligent Policing Key Laboratory of Sichuan Province, Sichuan Police College, Sichuan 646000, China

<sup>3</sup> Department of Computer Science and Engineering, Santa Clara University, Santa Clara, CA 95053, USA

<sup>4</sup> School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China

In recent years, the field of community detection has attracted the attention of many researchers and many community detection methods have been proposed [1, 3–5], such as modularization and minimal cut [6]. In addition, non-negative matrix factorization algorithm (NMF) is also an important method for community detection, which aims to decompose a high-dimensional matrix into two or several low-dimensional non-negative matrices, whose product can be approximated equal to the original matrix. Compared with other methods, NMF has the following main advantages in the application of community detection [7]: (1) high interpretability: by representing the community network using an adjacency matrix and utilizing this matrix as the feature matrix in NMF, each value in the community result matrix obtained through factorization can be understood as the probability that a node belongs to the corresponding community. For example, in the community result matrix,  $Z_{ij}$  represents the probability or strength of node  $V_i$  belonging to community  $C_j$ . This enhances the interpretability and explainability of the results; (2) high adaptability: real-world networks come in various forms, including overlapping and non-overlapping networks, directed and undirected networks, attribute networks, dynamic networks, and more. and the NMF and its related variants can be applied to any of the above networks. For example, in the case of overlapping networks, it only requires setting a probability threshold to detect nodes' membership in multiple communities; (3) high integration: existing information within the community network can be incorporated into the NMF learning process to improve the accuracy of community detection. For instance, existing attributes or node labels in the network can be integrated into the objective function to iteratively learn more refined results. Building upon these advantages, researchers have conducted in-depth studies on the application of NMF in community detection. For topological networks [8, 9], which exclusively contain structural information, such as directed or undirected networks, NMF can be directly applied to detect communities. Many researchers have further improved this by modeling communities [3] or enhancing performance by incorporating additional information [10]. For signed networks [11, 12], i.e., the relationship between nodes can be expressed as positive or negative correlation, where positive correlation means that the nodes are friends and negative correlation means that the nodes are enemies, and thus the adjacency matrix is a matrix with sign. Compared to traditional networks, signed networks not only consider the closeness between nodes but also require positively correlated nodes to be in the same community and negatively correlated nodes to be in different communities. For attribute networks [13, 14], where nodes possess labels or attribute information in addition to link structure, these attribute details often better represent unique node characteristics and complement topological information for achieving high-quality community

detection. It is evident that NMF and its variants can address community detection problems in various types of networks, playing a crucial role in community mining.

Although current research has achieved certain effectiveness, mining potential information in community networks remains insufficient, mainly due to the following shortcomings: (1) lack of consideration of homology among all nodes within the community: some nodes may not have a direct relationship, but through common neighbor nodes, potential relationships between them can be discovered. Typically, nodes with a large number of common neighbors are more closely related than those with few or no common neighbors, making them more likely to be assigned to the same community. For example, if both Paper 1 and Paper 2 cite Paper 3, while Paper 4 has no citation or referencing relationship with the aforementioned three papers, Papers 1 and 2 are more likely to be classified into the same topic category with a higher probability, indicating a potential common theme. However, it is uncertain whether Paper 4 shares a common theme with the aforementioned three papers. This illustrates the potential relationships that can be uncovered through common neighbor nodes. Therefore, considering common neighbors in calculating similarity is particularly important; (2) Lack of consideration for heterogeneity between communities (i.e., between communities): community detection aims not only for nodes within communities to have more similar features but also for differences between communities to be more distinct. The greater the differences between communities, with more focused features within each community, the clearer the community detection partition and the better the results; (3) Lack of optimization for the community membership matrix: for the probability matrix obtained from community detection, we often assign the community with the highest probability as the belonging community for the corresponding node. However, due to constraints from initialization methods and the number of iterations, the values in the probability matrix often lack clear distinctiveness. In summary, considering the three elements mentioned above simultaneously during the model learning process is crucial for community detection. Optimization from the internal, external, and inherent characteristics of community networks can enhance the effectiveness of community detection. However, addressing these aspects simultaneously is often challenging and represents a research problem that urgently needs solutions.

Motivated by existing community detection methods based on NMF, this study aims to address the following issues: (1) addressing the homogeneity issue, we will consider the similarity between nodes and measure the similarity in various ways to uncover hidden information in the network; (2) In order to address the disparities between communities, we will introduce orthogonal constraints to ensure diversity between communities. We will guide the learning process of the objective function with a mutual constraint

relationship between the diversity between communities and the homogeneity among nodes, aiming for improved detection results; (3) For optimizing the result matrix, we will add constraints to the community membership matrix. This will lead to learning better results during each iteration, producing more very small and some larger values, making the probability of nodes belonging to communities more distinct. In summary, we propose a new multi-constraint non-negative matrix factorization community detection method, named orthogonal regular L1-norm sparsity constrained non-negative matrix factorization (ORSNMF). Building on the foundation of traditional network topology modeling, we simultaneously model the diversity between communities, the similarity among nodes, and the sparsity of the community membership matrix. This comprehensive modeling aims to better characterize the features of community structures. Finally, we incorporate these three aspects into the objective function for joint-constrained learning, resulting in improved community detection. Our main contributions can be summarized as follows:

- Based on the NMF of orthogonal regular L1-norm sparsity constraint, a new community detection model based on non-negative matrix factorization is proposed. The proposed scheme simultaneously models the differences between communities, the similarity between nodes and the sparsity of community member matrix in directed networks in order to obtain the attributes of community structure to the greatest extent.
- An algorithm with convergence guarantee is proposed to optimize the model.
- Extensive experiments on synthetic and real data sets show that our proposed model has better performance on three metrics: jaccard similarity, normalized mutual information (nmi) and accuracy [15].

The rest of this article is organized as follows. We introduce related works in Sect. “[Related works](#)”, elaborate on the related issues of community detection in Sect. “[Problem description](#)”, and detail our proposed orthogonal regular L1-norm sparsity constraint non-negative matrix factorization model in Sect. “[Orthogonal regular and L1-norm sparse constrained NMF](#)”. Comprehensive experiments are performed to validate the effectiveness of the proposed scheme in Sect. “[Experiment and analysis](#)”, followed by conclusions in Sect. “[Conclusion](#)”.

## Related works

For the community networks existing in real life, how to mine the effective information and then identify the community to promote more practical applications, such as movie recom-

mendation, advertisement pushing, etc., is a basic problem of network analysis, and the process is also known as community detection. In recent years, various community detection methods have emerged, with significant attention given to those based on NMF [3, 5, 16–22]. These methods have gradually become a new direction in the field of community detection

NMF is a classic low-rank matrix decomposition model proposed by Lee et al. [23, 24]. The process involves decomposing a non-negative matrix into the product of two or more non-negative matrices. The goal is to find a non-negative base matrix and its corresponding non-negative coefficient matrix, which, when multiplied, approximates the original data matrix (i.e., the matrix before decomposition). NMF possesses a unique functionality, namely, inherent clustering capability. He et al. [25] demonstrated that NMF and its related improvements have similar effects to some classical clustering algorithms [26–28]. Community detection, fundamentally, is a clustering problem on complex networks. In addition to its clustering capability, NMF also has advantages such as interpretability. When using NMF for community detection, the adjacency matrix in the community network can be used as the feature matrix of NMF, and the decomposed results represent the community member matrix and the community feature matrix, respectively, which can be viewed as the probability value that the node belongs to the community in the community member matrix, so as to extract the relationship between the node and the community, which makes the results easier to understand and convince people, based on the above advantages, NMF is very suitable for community detection.

Most existing community detection methods based on NMF focus on enhancing the performance of the NMF algorithm to achieve better results in community detection. For example, Wang et al. [3] proposed symmetric non-negative matrix factorization (SNMF), asymmetric non-negative matrix factorization (ANMF) and joint non-negative matrix factorization (JNMF), respectively, for undirected networks, directed networks and composite networks to solve the problem of community discovery. The pairwise constrained symmetric non-negative matrix factorization method (PCSNMF) proposed by Shi et al. [29] considers the symmetric community structure of undirected networks, but also some pairwise constraints for basic information generation. Ye et al. [9] proposed homophilic positive non-negative matrix factorization (HPNMF), which models not only the topology of links but also takes into account the homogeneity of nodes in the network, providing a better reflection of the inherent structural properties of communities. Ye et al. [10] propose to learn an affinity matrix adaptively, which can capture the intrinsic similarity between nodes accurately, and therefore benefit the community detection results. Shi et al. [21] proposed a Bayesian non-negative

matrix factorization (NMF) method for adaptive community detection. In the decomposition process, the use of Bayesian methods allows not only for capturing the most appropriate number of communities in large networks through shrinkage but also for finding optimal thresholds for assigning nodes to communities in ambiguous situations. Tosyali et al. [30] proposed regularized asymmetric non-negative matrix factorization (RANMF) for directed network clustering based on the prior information of the network and the pairwise similarity of nodes. Zhang et al. [31] proposed homophilic non-negative matrix factorization (HNMF) to model bidirectional relationships between links and communities. From the community-to-link perspective, the method assumes that nodes with common communities have a higher probability of establishing links than nodes without common communities, applying a preference-based pairwise function. From the link-to-community perspective, the method assumes that linked nodes have similar community representations, introducing a novel network embedding-based community representation learning approach. Liu et al. [32] introduced a symmetric and graph-regularized non-negative matrix factorization (SGNMF) method. This approach incorporates multiple latent factors to enhance its representation learning capabilities and introduces regularization terms to account for the symmetry of undirected networks, ultimately improving community detection performance. Luo et al. [33] proposed a novel constrained fusion-induced symmetric non-negative matrix factorization (CFS) model. This model, designed for undirected networks, introduces a graph regularization factor that preserves the intrinsic geometry of the network's local invariance. This incorporation allows the proposed detector to effectively understand the community structure within the target network.

In summary, most existing community detection methods can achieve good results under certain conditions, especially when considering node attributes or labels as prior knowledge, which can effectively enhance detection accuracy. However, many models do not comprehensively consider the characteristics of the community's internal, external, and inherent properties. They only derive limited inherent properties of community structures from the community structure itself, without maximizing the extraction of network information. As a result, this can impact the effectiveness of community detection.

## Problem description

A community network can be represented as a graph  $G = (V, E)$ , where node set  $V = \{V_1, V_2, \dots, V_n\}$ ,  $V_i$  represents a node and  $n = |V|$  represents the number of nodes in the

community network, edge set  $E = \{e_{ij} | V_i \in V \cap V_j \in V\}$ ,  $e_{ij}$  represents the edge between nodes  $V_i$  and  $V_j$ ,  $m = |E|$  represents the number of edges in the network. Networks are usually divided into undirected networks and directed networks, there are many clustering methods for undirected networks, while there are relatively few studies on directed networks. Therefore, this article focuses on directed unweighted network clustering. In general, a directed network  $G$  can be described by an adjacency matrix  $A = [A_{ij}]^{n \times n}$ ,  $A_{ij}$  represents the relationship between node  $V_i$  and node  $V_j$ , when there is a connection between  $V_i$  to  $V_j$  (i.e.  $e_{ij} \in E$ )  $A_{ij} = 1$ , otherwise,  $A_{ij} = 0$ . Suppose the network  $G$  consists of  $k$  communities, and  $C$  denotes the community set of  $G$ , that is,  $C = \{C_i | C_i \neq \emptyset, 1 \leq i \leq k\}$ , where  $C_i$  represents the  $i$ th community and it is not empty. The purpose of community detection is to divide these nodes into  $k$  different groups according to the network topology, so that the number of edges within any specific group are maximized, while the number of edges across different groups are minimized. In this study, we focus on non-overlapping community detection, that is, the community set  $C$  should satisfy the condition  $C_i \cap C_j = \emptyset$  if  $i \neq j$ , which means that different communities  $C_i$  and  $C_j$  have no common nodes.

Recently, NMF has become the important method of community detection [5, 9, 16, 18–20, 34], which mainly has the following advantages:

- Better interpretability: given a network, after the non-negative matrix decomposition, a community member matrix will be obtained. Each element in the matrix can be understood as the probability or intensity that the node belongs to the corresponding community, which makes the results of community detection more interpretable.
- Convergence node-related information: NMF can integrate node-related information (such as node similarity information) as regularization constraints into the objective function, and jointly guide the iterative optimization of the objective function to improve the clustering performance.

In view of this, we adopt NMF for community detection. Specifically, the problem is defined as follows:

Given a directed and unweighted network  $G = (V, E)$ , using  $A$  to represent the adjacency matrix of this network, the individual nodes in the network can be divided into disjoint clusters by optimizing the following objective function:

$$\min \mathcal{L}(Z, H) = \|X - ZH\|_F^2, \quad \text{s.t. } Z \geq 0, H \geq 0, \quad (1)$$

where  $X \in R_+^{n \times m}$  is the original non-negative matrix,  $Z \in R_+^{n \times k}$  is the basis matrix,  $H \in R_+^{k \times m}$  is the coefficient matrix,

$k < \min\{n, m\}$ ,  $\|\bullet\|_F^2$  is the Frobenious norm, whose purpose is to find the optimal low-rank non-negative matrices  $Z$  and  $H$ , making  $ZH$  infinitely close to  $X$ .

When NMF is leveraged for community detection, the corresponding adjacency matrix  $A$  in the network will be used as the characteristic matrix for decomposition, that is,  $A \approx ZH$ , where  $H$  and  $Z$  represent the community characteristic matrix and the community member matrix, respectively. Furthermore,  $k$  represents the number of communities (clusters); and  $Z_{ij}$  represents the probability (strength) that the node  $V_i$  belongs to the community  $C_j$  (i.e.  $1 \leq j \leq k$ ).

The discussion above is the traditional NMF model for community detection. However, in the above model, the connectivity between communities is not considered, so Wang et al. [3] propose to integrate the information between communities into the objective function, and set  $H = CZ^T$  so that the original optimization problem is converted to the following optimization problem:

$$\min \mathcal{L}(Z, C) = \|A - ZCZ^T\|_F^2, \quad s.t \ Z \geq 0, C \geq 0, \quad (2)$$

where  $A \in R_+^{n \times n}$  is the adjacency matrix, including  $n$  nodes,  $Z \in R_+^{n \times k}$  is the community member matrix, storing the probability values of nodes belonging to communities, where  $Z_{ij}$  stores the probability of node  $V_i$  belonging to community  $C_j$ ,  $C \in R_+^{k \times k}$  is the cluster matrix representing the connectivity between two communities. For example, in a directed network, if the  $i$ th community points to the  $j$ th community, then  $C_{ij}$  is a non-zero value;  $Z$  and  $C$  are non-negative asymmetric matrices. In addition, researchers have also studied a variety of non-negative matrix factorization variants in this area. For example, Relative Pairwise Relationship constrained non-negative matrix factorisation (RPR-NMF) proposed by Jiang et al. [35].

A good clustering method should result in more similar nodes within communities and less community-to-community associations, i.e., the more pronounced the differences between communities. Nevertheless, the studies above only consider the network connections/edges while ignoring the similarity between nodes, that is, the tightness between nodes with similar features is often greater than that between nodes with different features. Therefore, in this study, we add similarity information to the objective function, that is, node similarity constraints. We study non-overlapping community detection, in order to determine the community to which the  $i$ th node belongs. We take the index angle of the maximum value in the  $i$ th row of the community member matrix  $Z$  as the community to which the  $i$ th node belongs. In order to obtain a better community member matrix  $Z$ , we will add a constraint to  $Z$  to produce only a few large values with most other values very small.

## Orthogonal regular and L1-norm sparse constrained NMF

We develop a new orthogonal regularized L1-norm sparse constrained non-negative matrix factorization model (ORSNMF), which considers the differences between communities, node similarity and how to obtain a better community membership matrix. In this study, we first model the above three aspects separately and then combine them into a unified model.

### Community difference modeling

Given a network, its topology contains rich information, therefore, it can serve as an essential starting point of community analysis. We know that orthogonality constraints ensure interpretability and maintain sparsity constraints to avoid some trivial solutions [36]. In practice, we hope that the vectors of cluster matrix  $C$  are different from each other. Since if the vectors are more orthogonal, the differences between communities are more significant, leading to better clustering results. Therefore, we add orthogonality constraints to cluster matrix  $C$  and integrate them into the objective function as follows:

$$\min \mathcal{L}_O = \|A - ZCZ^T\|_F^2 + \gamma \|C^T C - I\|_F^2, \quad s.t \ \gamma \geq 0, Z \geq 0, C \geq 0 \quad (3)$$

where  $A \in R_+^{n \times n}$  represents the adjacency matrix of the community network,  $n$  represents the total number of nodes in the community;  $Z \in R_+^{n \times k}$  represents the community membership matrix obtained after learning,  $k$  represents the number of communities, and  $Z_{ij}$  represents the probability value of node  $V_i$  belonging to community  $C_j$ ;  $C \in R_+^{k \times k}$  represents the community matrix, and  $C_{ij}$  represents the strength of the relationship between community  $C_i$  and community  $C_j$ ;  $Z^T$  represents the transpose of matrix  $Z$ ,  $\|\bullet\|_F^2$  represents the Frobenius norm,  $\gamma$  is an orthogonalization parameter used to balance the first error term and the sparsity of the second term,  $I$  represents the identity matrix, and matrices  $A$ ,  $Z$ , and  $C$  are all non-negative matrices.

### Node similarity modeling

In practice, we can observe that the relationship between a pair of nodes with similar characteristics is much stronger than that between a pair of nodes with different characteristics. Therefore, we consider adding a regularization term to the objective function to include the node similarity. The regularization is specified as follows:

$$\min \mathcal{L}_R = \frac{\lambda}{2} \sum_{i=1}^n \sum_{j=1}^n d(z_i, z_j) S_{ij}, \quad s.t \ z \geq 0, \quad (4)$$

where  $\lambda$  is the regularization parameter;  $S \in R^{n \times n}$  is the similarity matrix, which is a symmetric matrix;  $S_{ij}$  represents the similarity between node  $i$  and node  $j$ ;  $d(z_i, z_j)$  represents the distance between two nodes. In particular, since nodes in the same community are closer to each other, their distance should be smaller, while on the other hand, the distance between nodes from different communities should be larger. The commonly used method to represent the distance between two nodes is the Euclidean distance, which is calculated as follows:

$$d(z_i, z_j) = \|z_i - z_j\|^2. \quad (5)$$

Next, we will introduce three methods to calculate the similarity between nodes.

### Adjacency similarity

The simplest approach to calculate node similarity is the adjacency similarity, represented by  $S_{ij}$ , which is calculated as the number of associations between nodes  $V_i$  and  $V_j$ :

$$S = A + A^\top + I, \quad (6)$$

Since this study focuses on directed networks, the relationship between node  $V_i$  and node  $V_j$  needs to consider direction. Similarly, when calculating similarity, all relationships between two nodes need to be taken into account. Therefore, when using the adjacency matrix to calculate similarity,  $A + A^\top$  should be used. In addition, each node is strongly related to itself, so the identity matrix  $I$  is also included in the consideration.

### Katz centrality

In a network, the centrality of a node is used to measure its importance in the network. Katz centrality calculates the relative influence of a node in a network by measuring its direct neighbors (first-level neighbors) and the number of connections to all other nodes through these direct neighbors in the network. Katz centrality considers not only the contribution of the node's neighbors to it, but also the size of the contributing neighbors. In addition, a constant is added to represent the node itself. Therefore, Katz centrality is also often used as a similarity measure. Katz centrality is defined as

$$S = (I - \delta \cdot A)^{-1} \eta, \quad (7)$$

where  $\delta$  is the weight parameter, usually  $0 \leq \delta \leq 1$  [30], and  $\eta$  represents the constant of the node itself. In this article, the node's own constants are all set to 1 [30].

### Cosine similarity

Cosine similarity measures the similarity between two nodes by calculating the number of common neighbors between them. Specifically, it is expressed by dividing the number of common neighbors of node  $i$  and node  $j$  by the geometric mean of their degrees [37], which is calculated as follows:

$$S_{ij} = \frac{v_i^\top v_j}{\|v_i\| \cdot \|v_j\|}, \quad (8)$$

where  $v_i, v_j$  are the vectors corresponding to nodes  $V_i$  and  $V_j$  in the adjacency matrix  $A$ ,  $v_i^\top v_j$  represents the number of common neighbors between node  $V_i$  and node  $V_j$ , and  $\|v_i\|$  represents the geometric mean of the degrees of node  $V_i$ .

### Community membership matrix sparsity modeling

Given the community member matrix  $Z$  in non-overlapping community detection,  $Z_{ij}$  represents the probability (intensity) of node  $V_i$  belonging to community  $C_j$ . The method of assigning the  $i$ th node to a community is to take the index angle of the maximum value in the  $i$ th row of the community member matrix  $Z$  as the community to which the  $i$ th node belongs. To make this result better and more obvious, we will add  $L1$ -norm sparsity constraints to each row of the community membership matrix  $Z$  to ensure that only a few larger values are generated while all other values are very small so that each node captures the community to which it belongs.

The  $L1$ -norm sparsity constraint is the sum of the absolute values of all elements in the vector, which makes the algorithm tend to push the absolute values of some weights to zero during optimization, generating only a small number of larger values and achieving sparsity. This is because optimization algorithms like gradient descent, during the minimization of the objective function, apply gradients to each weight. The gradient of the  $L1$ -norm is non-differentiable at zero but constant at non-zero points. This implies that if the initial value of a weight is non-zero, the gradient will push it towards zero during the optimization process, leading to many small or zero values in the result, highlighting a few larger values. This makes the results more pronounced, interpretable, and robust. This is particularly useful for feature selection, as it retains only a few features most relevant to the target in the final model, pushing the weights corresponding to other features towards zero. In contrast, other sparse constraints like  $L2$ -norm sparse constraint is the square root of the sum of squares of all elements in the weight vector, encouraging weights to be distributed across all dimensions, attempting to make each feature contribute somewhat to the predicted value and preventing excessively large weights, thus aiding in preventing overfitting. For the result matrix in this article, the goal is to push some unimportant weights towards

zero, making nodes more likely to capture the community to which they belong, facilitating interpretability. Therefore, we apply  $L1$ -norm sparse constraint to the result matrix  $Z$ . Specifically, it is expressed as follows:

$$\min \mathcal{L}_S = \alpha \sum_{i=1}^n \|z_i\|_1^2, z \geq 0, \tag{9}$$

where  $Z_i$  represents the  $i$ th row of  $Z$ ,  $\|z_i\|_1$  represents the  $L1$ -norm applied on  $Z_i$ , which is the sum of the absolute values of each element on  $Z_i$ , and  $\alpha$  is the sparse parameter, which is used to balance the sparse term and the error between  $A$  and  $ZCZ^T$ .

**Orthogonal regular  $L1$ -norm sparse constrained non-negative matrix factorization model (ORSNMF)**

In summary, we consider the community difference model in Eq. (3), the node similarity model in Eq. (4) and the community member matrix sparsity model in Eq. (9) into the objective function to establish a unified model, so the overall objective function of our propose ORSNMF model is as follows:

$$\begin{aligned} \min_{Z \geq 0, C \geq 0} \mathcal{L} &= \mathcal{L}_O + \mathcal{L}_R + \mathcal{L}_S \\ &= \|A - ZCZ^T\|_F^2 + \gamma \|C^T C - I\|_F^2 \\ &\quad + \frac{\lambda}{2} \sum_{i=1}^n \sum_{j=1}^n d(z_i, z_j) S_{ij} + \alpha \sum_{i=1}^n \|z_i\|_1^2, \end{aligned} \tag{10}$$

By introducing the orthogonality constraint term  $\mathcal{L}_O$ , the regularization constraint term  $\mathcal{L}_R$ , and the  $L1$ -norm sparsity constraint term  $\mathcal{L}_S$ , we aim to fully capture the potential relationships among nodes in the network and the inherent properties of communities. These constraints are jointly iteratively learned to obtain improved result matrices.

**Update rules**

To optimize the objective function (10), for the  $\mathcal{L}_S$  in the target formula, we know that  $\|z_i\|_1^2$  is the absolute value of each element in the  $i$ th row of  $Z$  summed and then squared. The non-negative constraint characteristic of non-negative matrix decomposition makes every element in  $Z$  is non-negative value, so  $\|z_i\|_1^2$  is the sum and square of the elements in the  $i$ th row of  $Z$ . Consequently,  $\sum_{i=1}^n \|z_i\|_1^2$  is calculated as the sum of the elements of each row of  $Z$ , then squared, and finally summed up. Therefore, we can get  $\sum_{i=1}^n \|z_i\|_1^2 = tr(ZHZ^T)$ , where  $H \in 1^{k \times k}$  is a matrix of  $k \times k$  with all 1.

For the  $\mathcal{L}_R$  part in the target expression, it can be rewritten as

$$\begin{aligned} \min_{Z \geq 0} \mathcal{L}_R &= \frac{\lambda}{2} \sum_{i=1}^n \sum_{j=1}^n d(z_i, z_j) S_{ij} \\ &= \frac{\lambda}{2} \sum_{i=1}^n \sum_{j=1}^n \|z_i - z_j\|^2 S_{ij} \\ &= \frac{\lambda}{2} \sum_{i=1}^n \sum_{j=1}^n (z_i - z_j)^T (z_i - z_j) S_{ij} \\ &= \frac{\lambda}{2} \sum_{i=1}^n \sum_{j=1}^n (z_i^T z_i + z_j^T z_j) S_{ij} - \lambda \sum_{i=1}^n \sum_{j=1}^n z_i^T z_j S_{ij}, \end{aligned} \tag{11}$$

Since  $S$  is a symmetric matrix, Eq. (11) can be simplified as follows:

$$\begin{aligned} \min_{Z \geq 0} \mathcal{L}_R &= \lambda \left( \sum_{i=1}^n z_i^T z_i D_{ii} - \sum_{i=1}^n \sum_{j=1}^n z_i^T z_j S_{ij} \right) \\ &= \lambda tr(Z^T DZ) - \lambda tr(Z^T SZ) \\ &= \lambda tr(Z^T L_S Z), \end{aligned} \tag{12}$$

where  $tr(\cdot)$ , the trace of the matrix, equals to the sum of the elements of the main diagonal of the matrix, the Laplacian matrix  $L_S = D - S$ ,  $D$  is the diagonal matrix,  $D_{ii}$  represents the sum of values in the  $i$ th row of matrix  $S$ , i.e.  $D_{ii} = \sum_{j=1}^n S_{ij}$ . In this way, the regularization term in the similarity matrix  $S$  is integrated into the objective function to jointly guide the optimization objective function.

In summary, our objective function can be rewritten as

$$\begin{aligned} \min_{Z \geq 0, C \geq 0} \mathcal{L} &= \mathcal{L}_O + \mathcal{L}_R + \mathcal{L}_S \\ &= \|A - ZCZ^T\|_F^2 + \gamma \|C^T C - I\|_F^2 \\ &\quad + \lambda tr(Z^T DZ) - \lambda tr(Z^T SZ) + \alpha tr(ZHZ^T), \end{aligned} \tag{13}$$

The optimization problem in Eq. (13) is not simultaneously convex on the variables  $Z$  and  $C$ , therefore, finding the global minimum is difficult. Therefore, we use the multiplicative update rule to obtain the local optimal solution. Minimize the objective function in (13) using gradient descent, using  $\beta$  and  $\theta$  as Lagrangian multiplier numbers for constraints  $Z \geq 0$  and  $C \geq 0$ , Lagrangian  $\mathcal{L}$  is defined as

$$\begin{aligned} \mathcal{L} &= \|A - ZCZ^T\|_F^2 + \gamma \|C^T C - I\|_F^2 + \lambda tr(Z^T DZ) \\ &\quad - \lambda tr(Z^T SZ) + \alpha tr(ZHZ^T) + tr(\beta Z^T) + tr(\theta C^T) \\ &= tr(A^T A) - 2tr(A^T ZCZ^T) + tr(ZC^T Z^T ZCZ^T) \end{aligned}$$

$$\begin{aligned}
& +\gamma \text{tr}(C^4 - 2C^2 + I) + \lambda \text{tr}(Z^\top DZ) - \lambda \text{tr}(Z^\top SZ) \\
& +\alpha \text{tr}(ZH Z^\top) + \text{tr}(\beta Z^\top) + \text{tr}(\theta C^\top),
\end{aligned} \tag{14}$$

$\mathcal{L}$  calculates the partial derivative of  $Z$  and  $C$ , respectively:

$$\frac{\partial \mathcal{L}}{\partial C} = -2Z^\top AZ + 2Z^\top ZCZ^\top Z + 4\gamma C^3 - 4\gamma C + \theta, \tag{15}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial Z} &= 2(ZCZ^\top ZC^\top + ZC^\top Z^\top ZC) + 2\alpha ZH + \beta \\
&\quad - 2(A^\top ZC + AZC^\top) + 2\lambda(D^\top Z - S^\top Z),
\end{aligned} \tag{16}$$

According to the KKT condition, we know that  $\beta_{ir} z_{ir} = 0$  and  $\theta_{rj} C_{rj} = 0$ , so there are

$$\begin{aligned}
& [Z^\top ZCZ^\top Z]_{rj} c_{rj} + 2\gamma(C^3)_{rj} c_{rj} - 2\gamma(C)_{rj} c_{rj} \\
& \quad - (Z^\top AZ)_{rj} c_{rj} = 0,
\end{aligned} \tag{17}$$

$$\begin{aligned}
& (ZCZ^\top ZC^\top + ZC^\top Z^\top ZC)_{ir} z_{ir} + \lambda(D^\top Z - S^\top Z)_{ir} z_{ir} \\
& \quad + \alpha(ZH)_{ir} z_{ir} - (A^\top ZC + AZC^\top)_{ir} z_{ir} = 0,
\end{aligned} \tag{18}$$

Similar to the basic NMF, the multiplicative update rule of the objective function can be obtained:

$$c_{rj} \leftarrow c_{rj} \frac{[Z^\top AZ + 2\gamma C]_{rj}}{[Z^\top ZCZ^\top Z + 2\gamma C^3]_{rj}}, \tag{19}$$

$$z_{ir} \leftarrow z_{ir} \frac{[A^\top ZC + AZC^\top + \lambda S^\top Z]_{ir}}{[ZCZ^\top ZC^\top + ZC^\top Z^\top ZC + \lambda D^\top Z + \alpha ZH]_{ir}}, \tag{20}$$

The proof of convergence is shown in the appendix A.

### Overall procedure of ORSNMF algorithm

Given the adjacency matrix  $A$ , similarity matrix  $S$ , factorization rank (number of communities)  $k$  and the stop criteria of a directed network, we first use the modified non-negative double singular value decomposition (MNDSVD) [30] initialization to obtain the initial decomposition  $Z_0$  and  $C_0$ , then use the multiplication iteration update rule in Eqs. (19) and (20) to update  $Z$  and  $C$  until the stop criterion is met, and finally return to  $Z$  and  $C$ . The specific algorithm is shown in Algorithm 1.

### Experiment and analysis

In this section, we conduct experiments to demonstrate the effectiveness of the proposed algorithm for community detection in directed networks, which are done on both synthetic and real data sets. We compare the proposed

### Algorithm 1 ORSNMF algorithm

---

**Input:**  $A \in R^{n \times n}$ : adjacency matrix;  $S \in R^{n \times n}$ : similarity matrix;  $k$ : number of communities;  $\gamma \ \lambda \ \alpha$ : equilibrium parameter;  
**Output:** Cluster matrix  $C$ ; Community membership matrix  $Z$ ;  
1: Initialize  $Z_0$  and  $C_0$   
2: **while**  $\mathcal{L} \geq \varepsilon$  **do**  
3:   Update  $C$  according to equation(19)  
4:   Update  $Z$  according to equation(20)  
5: **end while**  
6: **return**  $C \ Z$

---

ORSNMF algorithm with SGNMF [32], CFS [33], SNCMF [38], HPNMF [9], RANMF [30], ANMF [3], Spectral clustering [39] and NCut [40] for community detection. All experiments were performed in Matlab.

### Comparative algorithms

- **ORSNMF:** ORSNMF is the model proposed in this study. It is abbreviated as ORSNMF in Table 1 and as ORSNMFR and ORSNMFM in Table 2, with the final letters R and M representing random initialization and MNDSVD initialization, respectively.
- **SGNMF:** Liu et al. [32] proposed a symmetry and graph-regularized non-negative matrix factorization (SGNMF) method, leveraging multiple latent factor matrices to represent a large-scale undirected network, thereby enhancing its representation learning ability. In Tables 1 and 2, it is abbreviated as SGNMF.
- **CFS:** Luo et al. [33] proposed a constraints fusion-induced symmetric non-negative matrix factorization (CFS) model, incorporating a symmetry-regularizer that preserves the symmetry of the learnt low-rank approximation to the adjacency matrix into the loss function, thus making the resultant detector well-aware of the target network's symmetry. In Tables 1 and 2, it is abbreviated as CFS.
- **SNCMF:** Yuan et al. [38] proposed a symmetric and non-negative constrained matrix factorization (SNCMF) community detection model based on undirected networks. This model introduces a graph regularization term to preserve the intrinsic geometric local invariance of the network, allowing the implemented detector to gain a comprehensive understanding of the community structure within the target network. In Tables 1 and 2, it is abbreviated as SNCMF.
- **HPNMF:** Ye et al. [9] proposed a homophily preserving NMF (HPNMF). This method models the network's link topology while also capturing the homogeneity of network nodes to better reflect the community structure. In Tables 1 and 2, it is abbreviated as HPNMF.
- **RANMF:** Tosyali et al. [30] proposed a regularized asymmetric non-negative matrix factorization (RANMF)



algorithm. In a given directed network, RANMF utilizes the pairwise similarity of nodes, guided by network prior information, to assign similar nodes to the same cluster. In Table 1, it is abbreviated as RANMF, and in Table 2, it is abbreviated as RANMFR and RANMFM, with the final letters *R* and *M* representing random initialization and MNDSVD initialization, respectively.

- ANMF: Wang et al. [3] proposed an asymmetric non-negative matrix factorization (ANMF) method for detecting communities in directed networks. Due to the asymmetry of the adjacency matrix and the weight matrix, the resulting matrix is not forcefully constrained. Instead, normalization of the result matrix is achieved by passing a diagonal matrix between the result matrix and the weight matrix. This approach aims to enhance the effectiveness of community detection. In Table 1, it is abbreviated as ANMF, and in Table 2, it is abbreviated as ANMFR and ANMFM, with the final letters *R* and *M* representing random initialization and MNDSVD initialization, respectively.
- Spect: Hespanha et al. [39] proposed a spectral decomposition-based graph partitioning algorithm, closely related to the Markov chain state aggregation algorithm introduced by Phillips and Kokotović [41]. This algorithm can be applied to the field of community detection to assess its effectiveness. In Tables 1 and 2, it is abbreviated as Spect.
- NCut: Shi et al. [40] proposed a method based on perceptual grouping. This method focuses not on local features of the problem but extracts information globally, subsequently introducing a normalized cut criterion. The aim is to measure the overall dissimilarity between different groups and the overall similarity within groups. This method can be applied to community detection and, to some extent, enhances the detection effectiveness. In Tables 1 and 2, it is abbreviated as Ncut.

### Data sets

We first compare the various clustering algorithms mentioned above on the LFR synthetic graphs. In the LFR synthetic graphs, the network topology complexity is controlled by the mixing parameter  $\mu$ , which controls the connection between communities. The larger the mixing parameter, the better the connectivity between communities, the more complex the network topology, the more difficult the community detection is.

In addition, we also selected the World Wide Knowledge Base (WebKB) data set, a real world data set, to test our proposed algorithm. This data set has more connections between communities, which increases the complexity of community detection and can better test our proposed algorithm. This

data set contains web page hyperlink information collected by four universities in Cornell, Wisconsin, Texas and Washington University. The specific meaning of the network is: nodes represent web pages, and directed edges represent links information between web pages. The web pages are divided into 5 categories, including students, courses, staff, projects, and teachers.

### Evaluation indicators

There are various methods for evaluating and comparing differences between algorithms [42]. In this study, to accurately assess the effectiveness of clustering algorithms, we adopted three evaluation metrics, namely jaccard similarity, nmi and accuracy [15].

The jaccard similarity is utilized here to compare the similarity between predicted results and true results. The similarity is assessed by calculating the ratio of the number of common elements in both the predicted and true results to the total number of elements in their union, as defined by Eq. (21):

$$j(PL, TL) = \frac{|PL \cap TL|}{|PL \cup TL|} = \frac{|PL \cap TL|}{|PL| + |TL| - |PL \cap TL|}. \quad (21)$$

where *PL* represents the predicted results, *TL* represents the true results, both *PL* and *TL* are  $1 \times n$  vectors, *n* represents the number of samples, *PL* and *TL* are used to store the predicted categories and true categories of each sample, respectively.  $|PL \cap TL|$  indicates the number of samples where the true results match the predicted results, and  $|PL \cup TL|$  represents the total number of different samples between the true results and the predicted results.

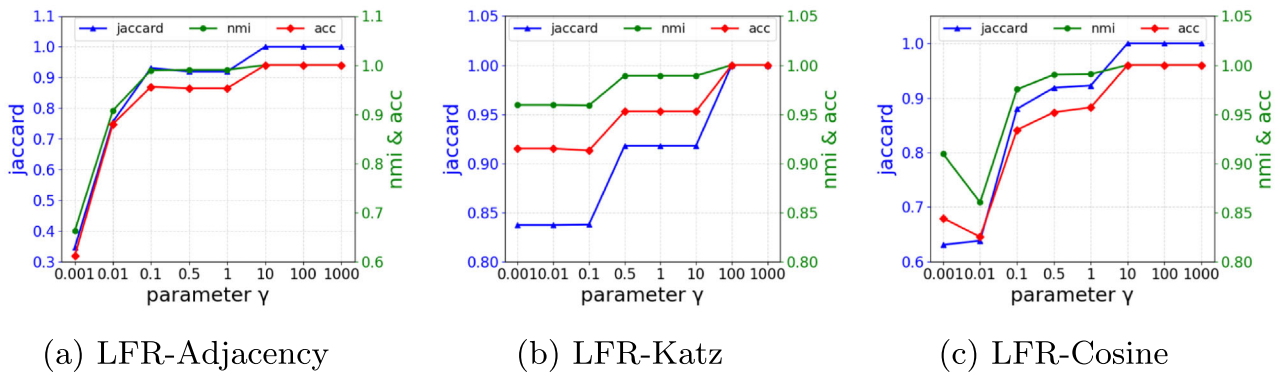
nmi is an external measure used to judge the quality of clustering. It is used to measure the similarity of two clustering results. The calculation is performed as Eq. (22):

$$nmi(C, C_T) = \frac{I(C, C_T)}{\sqrt{H(C)H(C_T)}}, \quad (22)$$

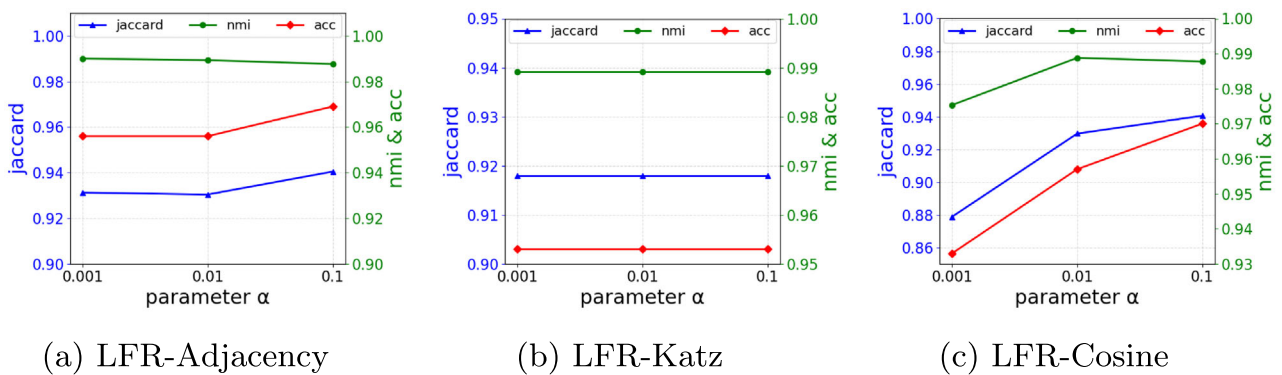
where *C* is a set of *k* clusters obtained after clustering, with each class represented as  $C = \{C_1, C_2, \dots, C_k\}$ , and containing several samples after clustering. *C<sub>T</sub>* represents the true class labels, with  $C_T = \{C_{T1}, C_{T2}, \dots, C_{Tk}\}$  and each *C<sub>Ti</sub>* containing all the samples belonging to that class.  $I(C_T, C)$  represents the mutual information between *C<sub>T</sub>* and *C*,  $H(C_T)$  and  $H(C)$ , respectively, represent the entropy of *C<sub>T</sub>* and *C*.

The accuracy is used to compare the obtained labels with the true labels provided by the original data, and is specifically defined as Eq. (23):

$$acc(TL, PL) = \sum_{i=1}^n \frac{\delta(TL_i, \text{map}(PL_i))}{n}, \quad (23)$$



**Fig. 1** On the LFR data set, with  $\alpha$  set to 0, the impact of the parameter  $\gamma$  in the ORSNMF model was analyzed by varying it from  $10^{-3}$  to  $10^3$ . **a–c** are the effect plots obtained using the adjacency matrix, the Katz similarity computation similarity matrix and the cosine similarity computation similarity matrix, respectively

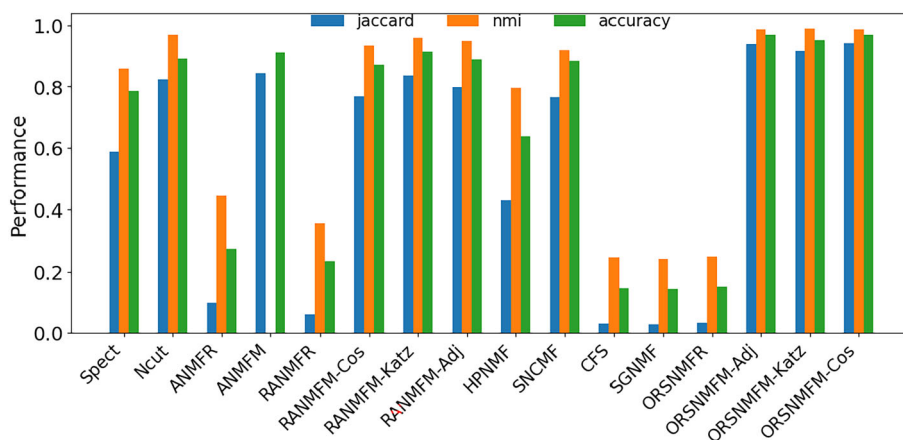


**Fig. 2** With  $\gamma$  fixed at the optimal value obtained in Fig. 1, the impact of the parameter  $\alpha$  in the ORSNMF model was analyzed by varying it from  $10^{-3}$  to  $10^{-1}$ . Three similarity calculation methods were experimented with, including **a–c**, representing the effectiveness of similarity matrices calculated using the Adjacency matrix, Katz, and Cosine similarity, respectively

**Table 1** Performance comparison on the LFR synthetic networks (bold numbers represent best results)

Algorithm	Initialization	Similarity matrix	jaccard	nmi	Accuracy
Spect	–	–	0.58682	0.8579	0.785
Ncut	–	–	0.82331	0.9688	0.891
ANMF	Rand	–	0.09925	0.4456	0.2753
ANMF	MNNDSD	–	0.84398	–	0.912
RANMF	Rand	–	0.06145	0.35804	0.2337
RANMF	MNNDSD	Cos	0.76843	0.93509	0.872
RANMF	MNNDSD	Katz	0.83701	0.95956	0.915
RANMF	MNNDSD	Adj	0.7979	0.94971	0.888
HPNMF	Rand	–	0.42922	0.79585	0.637
SNCMF	Rand	–	0.76603	0.91849	0.885
CFS	Rand	–	0.029744	0.24584	0.146
SGNMF	Rand	–	0.029105	0.24223	0.144
ORSNMF	Rand	–	0.03242	0.24982	0.15205
ORSNMF	MNNDSD	Adj	<b>0.94045</b>	<b>0.98764</b>	<b>0.969</b>
ORSNMF	MNNDSD	Katz	<b>0.9179</b>	<b>0.98918</b>	<b>0.953</b>
ORSNMF	MNNDSD	Cos	<b>0.94055</b>	<b>0.98775</b>	<b>0.97</b>

**Fig. 3** Visual comparison results of the conducted experiments on the LFR data set, corresponding to the visualized table 1



where  $PL_i$  is the label after clustering;  $TL_i$  is the ground truth label;  $n$  is the total number of data samples; and  $map(\cdot)$  represents the optimal redistribution of class labels to ensure the correctness of statistics. Generally, the optimal redistribution can be realized by the Hungarian algorithm, so as to solve the task (label) assignment problem in polynomial time.  $\delta$  represents the indicator function, which is defined as Eq. (24):

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

The larger the value of the above three evaluation indicators, the better the clustering performance.

### Experimental results

We conduct multiple experiments on the LFR networks data set and the World Wide Knowledge Base (WebKB) data set (Cornell, Wisconsin, Texas, and Washington). In addition to our proposed ORSNMF algorithm, there are also eight algorithms of SGNMF, CFS, SNCMF, HPNMF, RANMF, ANMF, Spectral clustering and NCut for comparative experiments.

#### LFR networks

For the LFR networks data set, we use  $\mu = 0.5$  [43],  $|V| = 1000$ ,  $|E| = 15,249$ ,  $k = 33$  to create a community network structure. We use Adjacency similarity, Katz centrality, and Cosine similarity as the similarity matrix of the LFR networks data set to test their performance. Using  $\lambda = 0.1$  [30], we set different values for the parameters  $\gamma$  and  $\alpha$  to test the algorithm. Inspired by the parameter setting method proposed by Ye et al. [9], we adopt the same method in this experiment to test the parameters in the range of  $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ , so that the parameters take values with better effect. We first evaluate the effect of  $\gamma$  on the model by fixing  $\alpha$  to 0, as shown in Fig. 1.

As can be seen from Fig. 1, no matter which method is used to calculate the similarity matrix, the three evaluation indexes of the model gradually increase with the increase of  $\gamma$  when  $\gamma \leq 10$ , and then tend to be stable. These results allow us to draw a conclusion: the algorithm is sensitive to the parameter  $\gamma$  to a certain extent. To obtain more effective results, it is necessary to consider community differences. We choose the better value among them as the final value of parameter  $\gamma$ . We make  $\gamma = 0.1$  when using adjacency similarity as the similarity matrix,  $\gamma = 0.5$  when selecting Katz centrality, and  $\gamma = 0.1$  when selecting Cosine similarity.

We then evaluate the effect of  $\alpha$  on clustering by fixing the value of  $\gamma$  (such that  $\gamma$  is equal to the figure of merit obtained above), the results are shown in Fig. 2.

Clustering performance drops sharply after  $\alpha = 0.1$ , so we only select data less than 0.1 to draw the result graphs. It can be seen from the above results that no matter which method is selected as the similarity matrix, the effect is the best when  $\alpha = 0.1$ . So under this data set, we take  $\alpha = 0.1$ .

In addition to using the above-obtained parameter values to carry out this algorithm experiment, a comparison experiment with SGNMF, CFS, SNCMF, HPNMF, RANMF, ANMF, Spectral clustering and NCut algorithms was also carried out. The specific results are shown in Table 1, and the corresponding visualization results are shown in Fig. 3.

It can be seen from Table 1 and Fig. 3 that no matter which similarity metric is used, our proposed algorithm achieves better results than other algorithms. Bold values indicate that the effect of this algorithm is better than that of other algorithms.

#### WebKB data set

In addition to comparative experiments on the LFR networks data set, we also conduct comparative experiments on the World Wide Knowledge Base (WebKB) data set (Cornell, Wisconsin, Texas, and Washington). Among them, Cornell contains 195 nodes and 304 directed edges. Texas contains

187 nodes and 328 directed edges. Wisconsin contains 265 nodes and 530 edges. Washington contains 230 nodes and 446 directed edges. Cosine similarity was used as the similarity matrix for the data sets collected by Cornell University and the University of Texas, Katz centrality was used as the similarity matrix for the data sets collected by the University of Washington, and the adjacency matrix was used as the similarity matrix for the data sets collected by the University of Wisconsin. The values of the parameters  $\lambda$ ,  $\gamma$  and  $\alpha$  are the same as the above processing methods, let  $\lambda = 0.1$ ,  $\alpha = 0$  to evaluate the influence of  $\gamma$  on it, and the results are shown in Fig. 4.

It can be concluded from Fig. 4 that for the Cornell data set we use  $\gamma = 0.001$ , Texas data set use  $\gamma = 0.001$ , Washington data set  $\gamma = 0.1$ , Wisconsin data set  $\gamma = 0.1$ . Now we fix the figure of merit  $\gamma$  obtained above to test the effect of  $\alpha$  on it, and the results are shown in Fig. 5.

It can be seen from Fig. 5 that the Cornell data set achieves a better value at  $\alpha = 10$ , and then it starts to fall, so we take  $\alpha = 10$ . The performance of the Texas data set increases slowly when  $\alpha = 0.1$ , but declines after  $\alpha = 0.1$ , so we take  $\alpha = 0.1$ . Washington data set has effects before  $\alpha = 0.1$ , but starts to decline after that, so we take  $\alpha = 0.1$ . When the adjacency matrix is used as the similarity matrix of the Wisconsin data set, performance starts to drop after  $\alpha = 0.1$ , so we take  $\alpha = 0.1$ .

After selecting all parameters, we will compare experiments with the eight algorithms of SGNMF, CFS, SNCMF, HPNMF, RANMF, ANMF, Spectral clustering and NCut. The results are shown in Table 2, and the corresponding visualization results are shown in Fig. 6. The bold values in Table 2 indicate the best performance. As can be seen from the table, the performance of our proposed algorithm is better than that of the other algorithms.

### Conclusion

This study addresses the community detection problem by proposing a new method, ORSNMF, within the fundamental framework of NMF. This method models the directed network topology, community distinctiveness, node homophily, and sparsity in the community membership matrix. We transform the objective of this model into an optimization problem, develop an efficient learning algorithm, and obtain a multiplicative update method to solve it. We conduct extensive experiments on both synthetic and real networks to demonstrate the superiority of the proposed model.

While the model proposed in this study demonstrates good performance, the primary focus is on directed, unweighted, and non-overlapping networks. In reality, overlapping and dynamic networks are prevalent. In community networks, samples or individuals often exhibit multiplicity, allowing

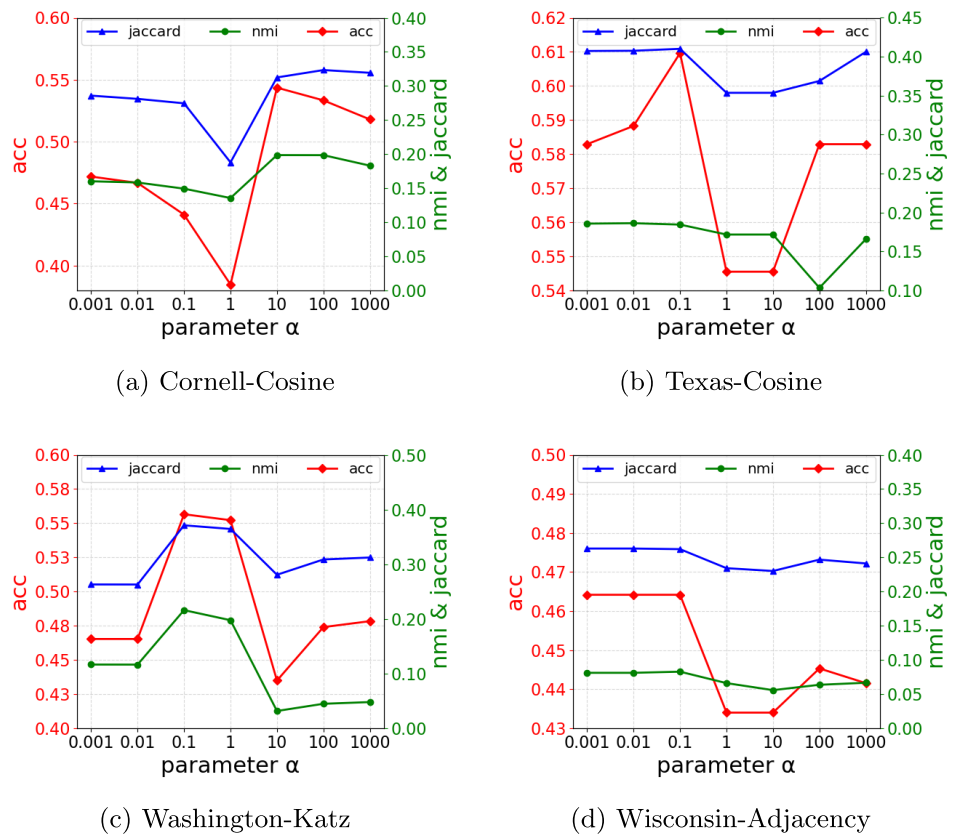
**Table 2** Performance comparison on the WebKB networks with ground-truth communities (bold numbers represent best results)

DS	Indicators	Spect	Neut	ANMFR	ANMFM	RANMFR	RANMFM	HPNMF	SNCMF	CFS	SGNMF	ORSNMF	ORSNMF
Cornell	jaccard	0.209	0.142	0.184	0.188	0.277	0.203	0.131	0.171	0.230	0.206	0.208	<b>0.312</b>
	nmi	0.053	0.064	0.144	0.097	0.168	0.127	0.042	0.043	0.078	0.048	0.143	<b>0.197</b>
Texas	Accuracy	0.359	0.318	0.368	0.354	0.451	0.379	0.272	0.308	0.410	0.338	0.386	<b>0.523</b>
	jaccard	0.312	0.156	0.213	0.348	0.344	<b>0.416</b>	0.15	0.322	0.324	0.174	0.304	0.378
Washington	nmi	0.049	0.031	0.147	0.225	0.173	0.194	0.036	0.116	0.109	0.042	0.169	<b>0.228</b>
	Accuracy	0.497	0.278	0.363	0.551	0.527	<b>0.594</b>	0.273	0.524	0.540	0.326	0.521	0.588
Wisconsin	jaccard	0.276	0.143	0.197	0.241	0.188	0.237	0.171	0.283	0.252	0.181	0.221	<b>0.371</b>
	nmi	0.048	0.024	0.138	0.100	0.150	0.102	0.023	0.134	0.044	0.019	0.151	<b>0.216</b>
Wisconsin	Accuracy	0.448	0.270	0.359	0.452	0.340	0.439	0.330	0.526	0.370	0.317	0.395	<b>0.557</b>
	jaccard	0.221	0.146	0.167	0.226	0.181	0.170	0.143	0.180	0.262	0.152	0.188	<b>0.262</b>
Wisconsin	nmi	0.064	0.023	0.067	0.076	0.067	0.082	0.022	0.051	0.055	0.044	0.052	<b>0.083</b>
	Accuracy	0.419	0.260	0.321	0.423	0.351	0.366	0.275	0.355	0.457	0.287	0.351	<b>0.464</b>

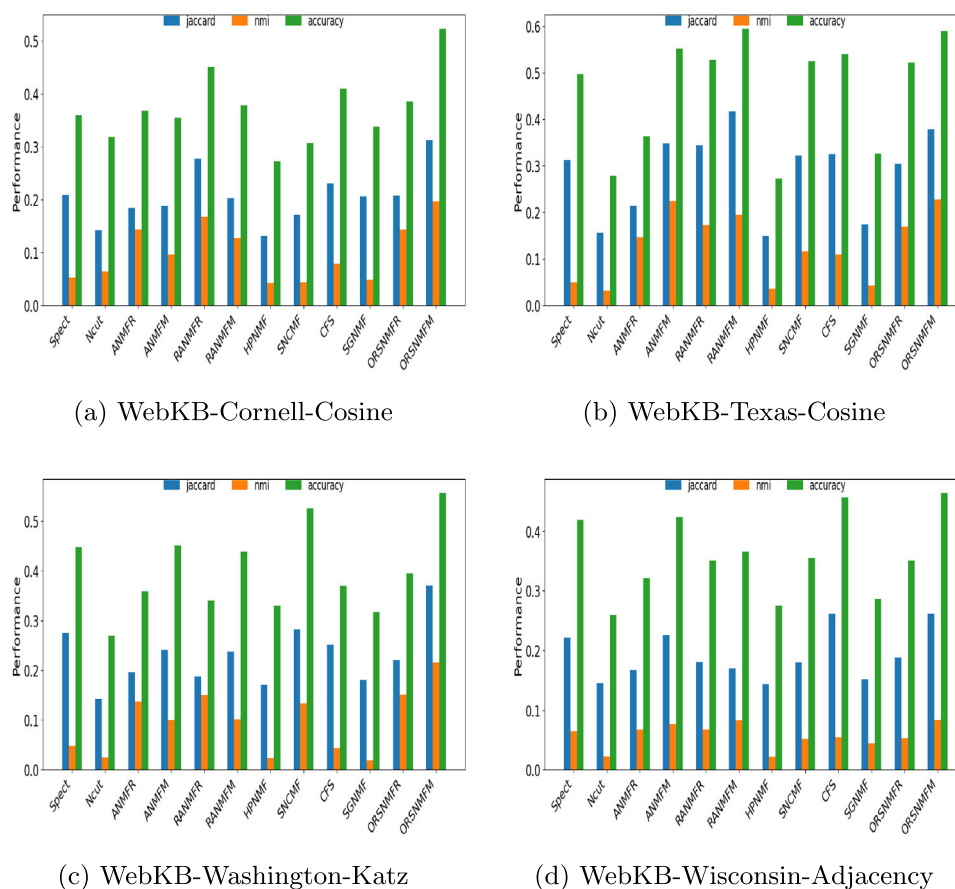
**Fig. 4 a–d**, respectively, depict the effectiveness of the ORSNMF model on the Cornell, Texas, Washington, and Wisconsin data sets, where  $\alpha$  is set to 0, and the parameter  $\gamma$  varies from  $10^{-3}$  to  $10^3$



**Fig. 5 a–d**, respectively, depict the effectiveness of the ORSNMF model on the Cornell, Texas, Washington, and Wisconsin data sets, with  $\gamma$  fixed at the optimal value obtained in Fig. 4, and the parameter  $\alpha$  varies from  $10^{-3}$  to  $10^3$



**Fig. 6** a–d, respectively, depict the visualized comparative experimental results on the Cornell, Texas, Washington, and Wisconsin data sets, corresponding to the visualized table 2



a sample to be assigned to multiple community categories. This phenomenon is known as overlapping networks, where, for example, an individual can simultaneously enjoy watching movies and playing basketball. Mechanically classifying such individuals solely into the movie-watching community would be overly simplistic. Therefore, detecting communities in overlapping networks becomes crucial. Due to the diversity in community networks, it becomes challenging to set a uniform number of attribution categories and probability threshold values for various community networks. In addition, with the passage of time, the structural attributes in the network are constantly changing, such as the number of citations of a paper will increase with the passage of time, and the relationship between users will be established and dissolved with the passage of time, so it can be seen that the dynamic network appears to be more general and applicable than the static network, and the community detection of the dynamic network not only can effectively delineate the members of the nodes of the network, but also can predict the development trend of the network. In future research, the emphasis will be on studying overlapping networks and dynamic networks. Furthermore, a variety of validation methods will be employed to highlight significant statistical differences

among different algorithms [42], demonstrating the superiority

**Acknowledgements** This work is supported by the Opening Project of Intelligent Policing Key Laboratory of Sichuan Province (Grant No. ZNJW2022KFZD002). The authors appreciate the reviewers for their helpful suggestions.

**Data availability** We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, and there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled “Multi-Constraint Non-negative Matrix Factorization for Community Detection: Orthogonal Regular Sparse constraint non-negative matrix factorization”

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix

Inspired by NMF, we can use the convergence proof of NMF to prove that the update steps of  $\mathcal{L}$  in Eqs. (19) and (20) are non-incremental. We also use the auxiliary function to prove the convergence of  $\mathcal{L}$ .

**Definition 1** Let  $G(c, c')$  be an auxiliary function of  $F(c)$ , then:

$$G(c, c') \geq F(c), \quad G(c, c) = F(c), \quad (25)$$

This helper function has the following lemma.

**Lemma 1** If  $G$  is an auxiliary function of  $F$ , then the objective function  $F$  is non-increasing under the following iteration rule,

$$c^{(t+1)} = \underset{c}{\operatorname{argmin}} G(c, c^{(t)}). \quad (26)$$

*Proof*

$$F(c^{(t+1)}) \leq G(c^{(t+1)}, c^{(t)}) \leq G(c^{(t)}, c^{(t)}) = F(c^{(t)}) \quad (27)$$

It is now shown that the update of  $C$  in Eq. (19) is equivalent to the update of Eq. (26) under the condition of the auxiliary function above.

Note in particular that  $\forall c_{ab} \in C$ ,  $F_{ab}$  represents an element in  $\mathcal{L}$  that is only related to  $c_{ab}$ . Now find the first derivative and the second derivative of the objective function with respect to  $c_{ab}$ , as follows:

$$F'_{ab} = \left( \frac{\partial \mathcal{L}}{\partial C} \right)_{ab} = \left[ 2Z^\top Z C Z^\top Z - 2Z^\top A Z + 4\gamma C^3 - 4\gamma C \right]_{ab} \quad (28)$$

$$F''_{ab} = \left[ 2Z^\top Z Z^\top Z + 4\gamma C^2 - 4\gamma \right]_{ab} \quad (29)$$

Since the nature of the update is elementwise, this is enough to show that each  $F_{ab}$  is non-incrementing under the update step of Eq. (19), so we have the following lemma.

**Lemma 2** The following function  $G(c, c^{(t)})$  is an auxiliary function of  $F_{ab}$ .

$$G(c, c^{(t)}) = F_{ab}(c^{(t)}) + F'_{ab}(c^{(t)})(c - c^{(t)}) + \frac{(Z^\top Z C Z^\top Z + 2\gamma C^3)_{ab}}{c^{(t)}} (c - c^{(t)})^2 \quad (30)$$

**Proof** Obviously  $G(c, c) = F_{ab}(c)$ , so we only need to prove  $G(c, c^{(t)}) \geq F_{ab}(c)$ . For this, we first consider the Taylor expansion formula of  $F_{ab}(c)$ :

$$F_{ab}(c) = F_{ab}(c^{(t)}) + F'_{ab}(c^{(t)})(c - c^{(t)}) + F''_{ab}(c^{(t)})(c - c^{(t)})^2 \quad (31)$$

Substituting Eq. (29) into Eq. (31) has:

$$F_{ab}(c) = F_{ab}(c^{(t)}) + F'_{ab}(c^{(t)})(c - c^{(t)}) + [Z^\top Z Z^\top Z + 2\gamma C^2 - 2\gamma]_{ab}(c - c^{(t)})^2 \quad (32)$$

Comparing Eq. (30) with Eq. (32), to show that  $G(c, c^{(t)}) \geq F_{ab}(c)$  is equivalent to proving:

$$\frac{(Z^\top Z C Z^\top Z + 2\gamma C^3)_{ab}}{c^{(t)}} \geq [Z^\top Z Z^\top Z + 2\gamma C^2 - 2\gamma]_{ab} \quad (33)$$

we know:

$$\begin{aligned} &(Z^\top Z C Z^\top Z + 2\gamma C^3)_{ab} \\ &= \sum_{q=1}^n \sum_{j=1}^r (Z^\top Z)_{aq} c_{qj}^{(t)} (Z^\top Z)_{jb} + 2\gamma \sum_{q=1}^n C_{aq}^2 c_{qb}^{(t)} \\ &\geq (Z^\top Z)_{ar} c_{ri}^{(t)} (Z^\top Z)_{ib} + 2\gamma C_{ar}^2 c_{rb}^{(t)} \\ &\geq (Z^\top Z)_{ar} c_{ri}^{(t)} (Z^\top Z)_{ib} + 2\gamma C_{ar}^2 c_{rb}^{(t)} - 2\gamma \end{aligned} \quad (34)$$

Therefore, the inequality  $G(c, c^{(t)}) \geq F_{ab}(c)$  holds.

Replacing  $G(c, c^{(t)})$  in Eq. (26) with Eq. (30) yields the following update rule:

$$\begin{aligned} c_{ab}^{(t+1)} &= c_{ab}^{(t)} - c_{ab}^{(t)} \frac{F'_{ab}(c_{ab}^{(t)})}{Z^\top Z C Z^\top Z + 2\gamma C^3} \\ &= c_{ab}^{(t)} \frac{[Z^\top A Z + 2\gamma C]_{ab}}{[Z^\top Z C Z^\top Z + 2\gamma C^3]_{ab}} \end{aligned} \quad (35)$$

It follows that  $F_{ab}$  is non-increasing under this update rule. It can also be proved that the update rule of  $Z$  is also equivalent, that is, the objective function (13) is non-increasing under the iteration rule (20).  $\square$

## References

- Zheng Z, Ye F, Li R-H, Ling G, Jin T (2017) Finding weighted k-truss communities in large networks. *Inf Sci* 417:344–360
- Wang X, Zhang Y, Zhang W, Lin X (2016) Efficient distance-aware influence maximization in geo-social networks. *IEEE Trans Knowl Data Eng* 29(3):599–612
- Wang F, Li T, Wang X, Zhu S, Ding C (2011) Community discovery using nonnegative matrix factorization. *Data Min Knowl Discov* 22(3):493–521
- Venkatesaramani R, Vorobeychik Y (2018) Community detection by information flow simulation. *arXiv preprint arXiv:1805.04920*

5. Sun B-J, Shen H, Gao J, Ouyang W, Cheng X (2017) A non-negative symmetric encoder-decoder approach for community detection. In: Proceedings of the 2017 ACM on conference on information and knowledge management, pp 597–606
6. Leskovec J, Lang KJ, Mahoney M (2010) Empirical comparison of algorithms for network community detection. In: Proceedings of the 19th international conference on world wide web, pp 631–640
7. He C, Fei X, Cheng Q, Li H, Hu Z, Tang Y (2021) A survey of community detection in complex networks using nonnegative matrix factorization. *IEEE Trans Comput Soc Syst* 9(2):440–457
8. Liu X, Wang W, He D, Jiao P, Jin D, Cannistraci CV (2017) Semi-supervised community detection based on non-negative matrix factorization with node popularity. *Inf Sci* 381:304–321
9. Ye F, Chen C, Wen Z, Zheng Z, Chen W, Zhou Y (2019) Homophily preserving community detection. *IEEE Trans Neural Netw Learn Syst* 31(8):2903–2915
10. Ye F, Li S, Lin Z, Chen C, Zheng Z (2018) Adaptive affinity learning for accurate community detection. In: 2018 IEEE international conference on data mining (ICDM). IEEE, pp 1374–1379
11. Yan C, Chang Z (2020) Modularized convex nonnegative matrix factorization for community detection in signed and unsigned networks. *Physica A Stat Mech Appl* 539:122904
12. He C, Liu H, Tang Y, Liu S, Fei X, Cheng Q, Li H (2021) Similarity preserving overlapping community detection in signed networks. *Future Gen Comput Syst* 116:275–290
13. Chunaev P (2020) Community detection in node-attributed social networks: a survey. *Comput Sci Rev* 37:100286
14. Guo T, Pan S, Zhu X, Zhang C (2018) Cfond: consensus factorization for co-clustering networked data. *IEEE Trans Knowl Data Eng* 31(4):706–719
15. Bothorel C, Cruz JD, Magnani M, Micenkova B (2015) Clustering attributed graphs: models, measures and methods. *Netw Sci* 3(3):408–444
16. Chen C, Zhu W, Peng B (2022) Differentiated graph regularized non-negative matrix factorization for semi-supervised community detection. *Physica A Stat Mech Appl* 604:127692
17. Chen Z, Li L, Peng H, Liu Y, Yang Y (2018) Attributed community mining using joint general non-negative matrix factorization with graph Laplacian. *Physica A Stat Mech Appl* 495:324–335
18. Lu H, Shen Z, Sang X, Zhao Q, Lu J (2020) Community detection method using improved density peak clustering and nonnegative matrix factorization. *Neurocomputing* 415:247–257
19. Jiao Pengfei, Wang W, Liu X, Dongxiao J, Di C (2017) Semi-supervised community detection based on non-negative matrix factorization with node popularity. *Inf Sci* 381:304–321
20. Jin H, Yu W, Li SJ (2019) Graph regularized nonnegative matrix tri-factorization for overlapping community detection. *Physica A Stat Mech Appl* 515:376–387
21. Shi X, Lu H, Jia G (2017) Adaptive overlapping community detection with bayesian nonnegative matrix factorization. In: International conference on database systems for advanced applications. Springer, Berlin, pp 339–353
22. Wang X, Cui P, Wang J, Pei J, Zhu W, Yang S (2017) Community preserving network embedding. In: Proceedings of the AAAI conference on artificial intelligence, vol 31
23. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755):788–791
24. Lee D, Seung HS (2000) Algorithms for non-negative matrix factorization. *Adv Neural Inf Process Syst* 13:556–562
25. Ding C, He X, Simon HD (2005) On the equivalence of nonnegative matrix factorization and spectral clustering. In: Proceedings of the 2005 SIAM international conference on data mining. SIAM, pp 606–610
26. Chen J, Zhao C, UlijiChen L (2020) Collaborative filtering recommendation algorithm based on user correlation and evolutionary clustering. *Complex Intell Syst* 6:147–156
27. Zhao F, Wang C, Liu H (2023) Differential evolution-based transfer rough clustering algorithm. *Complex Intell Syst* 9:5033–5047
28. Zhou W, Wang L, Han X, Parmar M, Li M (2023) A novel density deviation multi-peaks automatic clustering algorithm. *Complex Intell Syst* 9(1):177–211
29. Shi X, Lu H, He Y, He S (2015) Community detection in social network with pairwise constrained symmetric non-negative matrix factorization. In: Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015, pp 541–546
30. Tosyali A, Kim J, Choi J, Jeong MK (2019) Regularized asymmetric nonnegative matrix factorization for clustering in directed networks. *Pattern Recognit Lett* 125:750–757
31. Zhang H, Zhao T, King I, Lyu MR (2016) Modeling the homophily effect between links and communities for overlapping community detection. In: IJCAI, pp 3938–3944
32. Liu Z, Luo X (2023) A symmetry and graph regularized nonnegative matrix factorization model for community detection. arXiv preprint [arXiv:2302.12122](https://arxiv.org/abs/2302.12122)
33. Liu Z, Luo X (2023) A constraints fusion-induced symmetric non-negative matrix factorization approach for community detection. arXiv preprint [arXiv:2302.12114](https://arxiv.org/abs/2302.12114)
34. Dai X, Zhang K, Li J, Xiong J, Zhang N, Li H (2021) Robust semi-supervised non-negative matrix factorization for binary subspace learning. *Complex Intell Syst* 8:753–760
35. Jiang S, Kan L, Xu Y (2018) Relative pairwise relationship constrained non-negative matrix factorisation. *IEEE Trans Knowl Data Eng.* <https://doi.org/10.1109/TCSVT.2019.2892971>
36. Wu J, Chen B, Han T (2021) Two efficient algorithms for orthogonal nonnegative matrix factorization. *Math Probl Eng* 2021:1–13
37. Yang S (2013) Networks: an introduction by M. E. J. Newman. *J. Math. Sociol.* 37(4):250–251
38. Liu Z, Yuan G, Luo X (2022) Symmetry and nonnegativity-constrained matrix factorization for community detection. *IEEE/CAA J Autom Sin* 9(9):1691–1693
39. Hespanha JP (2004) An efficient matlab algorithm for graph partitioning. University of California, pp 1–8
40. Shi J, Malik J (2000) Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Mach Intell* 22(8):888–905
41. Phillips R, Kokotovic P (1981) A singular perturbation approach to modeling and control of Markov chains. *IEEE Trans Autom Control* 26(5):1087–1094
42. Derrac J, García S, Molina D, Herrera F (2011) A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm Evol Comput* 1(1):3–18
43. Lancichinetti A, Fortunato S (2009) Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys Rev E Stat Nonlinear Soft Matter Phys* 80(1):016118

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.