Robust asymmetric nonnegative matrix factorization for community detection in directed network

Kangmin Li[a], Yongguang Yu[a], Wei Hu[a,*], Yibing Cui[a], Jiting Li[b]

[a]*School of Mathematics and Statistics, Beijing Jiaotong University, Beijing, 100044, P.R.China*
[b]*Academy of Military Sciences, Beijing, 100071, P.R.China*

Abstract

Community detection is a fundamental problem in social network analysis, and as such, it has recently attracted a great deal of attention, with numerous community detection methods being developed. In recent times, a significant number of methods based on nonnegative matrix factorization (NMF) have been proposed, primarily due to the enhanced interpretability and integration of this approach. However, the potential of using the $l_{2,1}$ norm to enhance robustness to noise and outliers remains underutilized in current research on asymmetric NMF-based community detection algorithms. In this paper, we propose a novel robust asymmetric NMF model for community detection in undirected and directed networks, which is related to an edge-weighted stochastic block model (SBM). The approach works by minimizing an objective function which is equivalent to maximizing the likelihood function given by SBM to perform clustering, and we weaken the constraints to a regular term, allowing greater centrality for some nodes in the network. The optimization algorithm for the proposed model is given, and its convergence properties are also discussed. Extensive experimentation has been conducted in both real-world and synthetic networks to demonstrate the superiority of the proposed model.

Keywords: Community detection, Nonnegative matrix factorization, Directed graph, Clustering, Stochastic block model

*Corresponding author

Email addresses: 23111500@bjtu.edu.cn (Kangmin Li), ygyu@bjtu.edu.cn (Yongguang Yu), huwei@bjtu.edu.cn (Wei Hu), ybcui@bjtu.edu.cn (Yibing Cui), lijiting_1993@126.com (Jiting Li)

1. Introduction

Community detection represents a fundamental problem in social network analysis (Girvan and Newman (2002); Newman and Girvan (2004)), and has recently attracted considerable attention, coupled with the release of many community detection methods. In the context of complex networks, a community can be defined as comprising a set of nodes that engage in actions that are highly correlated with one another and with the actions of other nodes within the same community (Coscia et al. (2011)). The prevailing definition of network proximity is based on the observation that the density of links exhibits significant variation between different regions of the network. Therefore, the objective of a community detection algorithm is to divide a network into multiple distinct parts, with nodes within the same part exhibiting a high degree of connectivity (that is, a large number of links between them) and nodes in different groups exhibiting a low degree of connectivity.

The identification of communities (or clusters) within directed networks represents a significant challenge, with numerous important applications across a diverse range of domains. However, the issue of graph clustering has been predominantly addressed and examined in the context of undirected networks (Malliaros and Vazirgiannis (2013)). Numerous disparate algorithms have been proposed for undirected networks, with contributions from the fields of computer science, statistical physics, and biology (Fortunato (2010)). Nevertheless, a considerable proportion of graph data in a range of applications is inherently directed. Consequently, it is valuable to integrate all available information during the clustering process, including the directionality of the edges.

Nonnegative matrix factorization (NMF, Lee and Seung (1999)) is an effective data representation technology that has recently attracted considerable interest. Ding et al. (2005) highlighted the close relationship among NMF-based algorithms, k-means and graph partitioning methods. Due to the high interpretability as well as the integration of this approach, many NMF-based methods have been proposed recently. For undirected cases, Sun et al. (2017) proposed a nonnegative symmetric encoder–decoder (NNSED) method. Ye et al. (2019) proposed a homophily preserving NMF (HPNMF), which considered not only the link topology but also the node homophily. Liu and Luo (2023) considered a constraint fusion-induced symmetric (CFS) model. Moreover, research has been conducted on the subject of deep matrix factorization (Ye et al. (2018); Zhang and Zhou (2020); Al-sharoa and Rahahleh (2023); Hajiveiseh et al. (2024)).

Wang et al. (2011) introduced an asymmetric nonnegative matrix factorization (ANMF) approach for community detection in directed networks. Scholars have extended ANMF-based approach by incorporating regularization terms to better capture community structures (e.g., RANMF suggested by Tosyali et al. (2019)). However, such regularization introduces trade-offs: while improving generalization, it compromises interpretability, increases hyperparameter sensitivity, and raises computational costs.

In this paper, we proposed a robust asymmetric NMF model for clustering in directed network. Inspired by Zhang et al. (2018), we start by setting a particular stochastic block model (SBM) and weaken its maximum likelihood to an unconstrained NMF problem. Additionally, we develop multiplicative updating rules to address the proposed objective function, and present a proof of convergence.

The main contributions of this paper are as follows:

- The proposed model represents a novel approach to learn community structure in directed graph through the extension of the asymmetric NMF structure. It aims to maximize the likelihood function of a denoising process on a stochastic block model. Additionally, the proposed model avoids excessive use of regularization, thereby maintaining interpretability.

- The proposed approach is a combination of asymmetric NMF and robust NMF. In the form of the formula, a $l_{2,1}$-norm term is added to the objective function of the asymmetric NMF, which enhances the robustness against noise.

- The proposed approach is evaluated on a range of synthetic and real-world networks to assess its capacity to identify the community structure of the underlying network. Additionally, the proposed algorithm is benchmarked against existing state-of-the-art algorithms, with performance assessed using a suite of evaluation metrics.

The following is the structure of this paper: Section 3 provides an overview of the fundamental concepts and theoretical foundation. Section 4 presents the proposed robust asymmetric NMF model and optimization algorithm.

Section 5 presents the results of experiments carried out to demonstrate the efficiency of the approach. Section 6 presents the results and recommendations for further work.

2. Related Works

Here, we present a review of some NMF-based methods, especially ANMF-based methods, proposed for the task of community detection in directed networks. To detect community structure of directed graph, Wang et al. (2011) proposed asymmetric NMF (ANMF). The adjacency matrix is used to represent the directed network, which is used directly as input and approximated as a product of nonnegative matrices ($\mathbf{A} \approx \mathbf{XSX}^{\mathrm{T}}$). Frobenius norm is adopted to measure the distance between the adjacency matrix and its tri-factorization, which gives the objective of shallow ANMF.

Wang et al. (2017) proposed a modularized NMF model (M-NMF). By extending the matrix form of modularity (Newman (2006)), M-NMF incorporated the community structure into the embedding of the network. Similarly, by applying the extended modularity matrix formulation as a regularization term to the ANMF model, Yan and Chang (2019) presented a modularized tri-factor NMF model (MtriNMF).

Tosyali et al. (2019) proposed regularized asymmetric NMF (RANMF) method, incorporating prior similarity information to ANMF as an additional regularization term. The basic idea of the proposed method is to consider matrix decomposition as a reduction in dimensionality of the original data, while expecting similarity between nodes to be maintained in the reduced data. This approach allows for good compatibility that permits the definition of the similarity matrix to be selected in accordance with the specific characteristics of the network in question. The regularization term has also been adopted in other recent studies on community detection for directed networks, including RAsNMF (Abdollahi et al. (2020)), ORASNMF (Chen et al. (2024)).

In addition, it is worth noting that Kong et al. (2011) suggested a robust NMF (RNMF) approach, which employed the $l_{2,1}$ norm instead of the Frobenius norm to enhance its robustness to noise and outliers, which has almost the same computational cost as standard NMF. Based on that, He et al. (2020) introduced a deep robust NMF (DRNMF) approach.

Despite its advantages, the use of the $l_{2,1}$ norm for noise and outlier robustness in NMF-based community detection has received limited attention.

To the best of our knowledge, no research has yet addressed the combination of robust NMF and asymmetric NMF.

3. Premilinaries

3.1. Notations

In this paper, matrices are denoted by bold uppercase letters. $\mathbf{M}_{m\times p}$ is used to denote a matrix with m rows and p columns. The i-th row vector, the j-th column vector and (i,j)-th element of the matrix are represented by $\mathbf{M}_{i\cdot}$, $\mathbf{M}_{\cdot j}$ and $\mathbf{M}_{ij}$ respectively. $tr(\mathbf{M})$, $\mathbf{M}^{\mathrm{T}}$ represents the trace and transpose of the matrix. $\mathbf{1}$ represents a unit matrix, with every entry equals to one.

3.2. Nonnegative Matrix Factorization (NMF)

The basic NMF approach represents a generic low-rank matrix decomposition methodology that is primarily concerned with the analysis of nonnegative data matrices. For a nonnegative matrix $\mathbf{B} \in \mathbb{R}^{+}_{m\times p}$, the objective of NMF is to identify a pair of nonnegative matrices $\mathbf{X} \in \mathbb{R}^{+}_{m\times k}$ and $\mathbf{Y} \in \mathbb{R}^{+}_{k\times p}$ whose product provides an approximation of the original matrix, where $r < \min\{m, p\}$ is predetermined. NMF can be generally formulated as

$$\min_{\mathbf{X}\geq 0,\mathbf{Y}\geq 0} ||\mathbf{B} - \mathbf{XY}||_F^2 \tag{1}$$

where Frobenius norm is defined as $||\mathbf{M}_{m\times p}||_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{p} \mathbf{M}_{ij}^2$.

3.2.1. Asymmetric Nonnegative Matrix Factorization (ANMF)

In detecting k communities of a directed network, its adjacency matrix $\mathbf{A}_{n\times n}$ can be used to represent the directed graph, where n denotes the total number of nodes in the network. The adjacency matrix $\mathbf{A}$ is a binary matrix, with its element $\mathbf{A}_{ij}$ denotes if there is a directed edge from node i to node j. By replacing the matrix $\mathbf{Y}$ in the NMF with $\mathbf{SX}^{\mathrm{T}}$, asymmetric matrix factorization is formulated as

$$\min_{\mathbf{X}\geq 0,\mathbf{S}\geq 0} ||\mathbf{A} - \mathbf{XSX}^{\mathrm{T}}||_F^2 \tag{2}$$

where the factorization rank k denotes the number of communities to be clustered. In Eq.(2), $\mathbf{X} \in \mathbb{R}^{+}_{n\times k}$ obtained by ANMF represents the membership matrix, with element $\mathbf{X}_{ir}$ representing a probability value of node i belonging to community r; $\mathbf{S} \in \mathbb{R}^{+}_{k\times k}$ represents a community matrix, with element S_{rs} representing relationship intensity between communities r and s.

3.2.2. Robust Nonnegative Matrix Factorization (RNMF)

The robust NMF approach (Kong et al. (2011)) adopts minimization of the norm $l_{2,1}$ instead of the Frobenius norm, increasing the robustness of the objective function to noisy and outlying data. RNMF is expressed as

$$\min_{\mathbf{X}\geq 0,\mathbf{Y}\geq 0} ||\mathbf{B}-\mathbf{X}\mathbf{Y}||_{2,1} \tag{3}$$

where $l_{2,1}$ norm is defined as $||\mathbf{M}_{m\times p}||_{2,1} = \sum_{i=1}^{m}\sqrt{\sum_{j=1}^{p}\mathbf{M}_{ij}^2}$.

4. Proposed Method

Zhang et al. (2018) examines the relationship between the SBM and NMF. In the study, the equivalence of likelihood maximization of a normally distributed edge-weighted SBM (Aicher et al. (2015)) and constrained NMF is given. Specifically, for n nodes with k community labels, the edge-weight between nodes i, j respectively from communities r, s is modeled as a random variable with normal distribution $N(\mu_{rs}, \sigma^2)$. Clustering of a network with adjacency matrix $\mathbf{A}$ by edge-weighted SBM is to maximize the likelihood function

$$p(\mathbf{A}|\mu,\sigma) = \prod_{i\in r,j\in s} p(\mathbf{A}_{ij}|\mu_{rs},\sigma) = \prod_{i\in r,j\in s}\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(\mathbf{A}_{ij}-\mu_{rs})^2}{2\sigma^2}\} \tag{4}$$

Taking into account the membership matrix $\mathbf{X}$ with $\sum_j \mathbf{X}_{ij} = 1$, the expectation is given by $\mu_{rs} = (\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}})_{ij}$. Maximizing Eq.(4) is equivalent to a constrained asymmetric matrix factorization

$$\min_{\mathbf{X},\mathbf{S}\geq 0} ||\mathbf{A}-\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}||_F^2 \quad s.t. \sum_{j=1}^{n}\mathbf{X}_{ij} = 1\ ,\quad i = 1,2,\ldots,n \tag{5}$$

It should be noted that the network under examination may not convey edge weight information, since adjacency matrix $\mathbf{A}$ is a binary matrix. Nevertheless, the direct application of the adjacency matrix as a weighting mechanism for connected edges, while somewhat negating edge-to-edge differences, does not run counter to the clustering objective.

In short, ANMF employs maximum likelihood estimation of a simple probabilistic generative model.

$$\mathbf{A}_{ij} = (\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}})_{ij} + \epsilon_{ij}\ ,\quad \epsilon_{ij}\sim N(0,\sigma^2) \tag{6}$$

Similarly, robust NMF considers the model with Laplacian noise instead, as mentioned in Kong et al. (2011).

Here, a two-step noise addition process is examined. The model is provided in a simple format as follows:

$$\begin{aligned} \mathbf{A}_{ij} = \mathbf{W}_{ij} + \epsilon_{1ij} \quad &, \quad \epsilon_{1i\cdot} \sim \exp\{-||\epsilon_{1i\cdot}||_2/b\} \\ \mathbf{W}_{ij} = (\mathbf{XSX}^{\mathrm{T}})_{ij} + \epsilon_{2ij} \quad &, \quad \epsilon_{2ij} \sim N(0, \sigma^2) \end{aligned} \tag{7}$$

Using $\mathbf{XSX}^{\mathrm{T}}$ as the edge weight expectation matrix, in turn two types of noise ϵ_2, ϵ_1 with different distributions are added. Each row in the noise matrix ϵ_1 follows a Laplacian distribution with a scalar parameter b, similar to RNMF.

The maximum data log likelihood can be written as

$$\begin{aligned} \max \log p(\mathbf{A}, \mathbf{W}|\mathbf{X}, \mathbf{S}, \sigma, b) &= \max \left(\log p(\mathbf{A}|\mathbf{W}, b) + \log p(\mathbf{W}|\mathbf{X}, \mathbf{S}, \sigma)\right) \\ &= \max \left(-\sum_i \frac{||\mathbf{A}_{i\cdot} - \mathbf{W}_{i\cdot}||_2}{b} - \sum_{i,j} \frac{(\mathbf{W} - \mathbf{XSX}^{\mathrm{T}})_{ij}^2}{2\sigma^2} \right) \\ &= \max -\frac{1}{b}||\mathbf{A} - \mathbf{W}||_{2,1} - \frac{1}{2\sigma^2}||\mathbf{W} - \mathbf{XSX}^{\mathrm{T}}||_F^2 \end{aligned} \tag{8}$$

Given an input adjacency matrix $\mathbf{A}$, the learning objective is to minimize

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{X},\mathbf{S}\geq 0} \alpha||\mathbf{A} - \mathbf{W}||_{2,1} + ||\mathbf{W} - \mathbf{XSX}^{\mathrm{T}}||_F^2 \\ &s.t. \sum_{j=1}^{n} \mathbf{X}_{ij} = 1, \ i = 1, 2, \ldots, n \end{aligned} \tag{9}$$

where $\alpha > 0$ is a parameter that regulates the intensity weight of the noise. The constraints on the matrix row sums are defined according to the Eq.(5).

From the formulation given in Eq.(9), it can be observed that the matrix we actually decompose non-negatively is $\mathbf{W}$ rather than the input matrix $\mathbf{A}$, where $\mathbf{W}$ is linked to $\mathbf{A}$ through the $l_{2,1}$ norm. The $l_{2,1}$ norm excels in feature selection and noise-robust modeling through its unique dual operation (Argyriou et al. (2006)): the l_2 norm maintains feature group integrity while the l_1 norm selects between groups. This combined approach preserves meaningful feature relationships within groups while eliminating irrelevant ones, making it ideal for structured data analysis.

Furthermore, the construction of this $l_{2,1}$ paradigm term is based, on the one hand, on the assumption that the input adjacency matrix may contain incorrectly connected edges (cf. the link prediction problem), and on the other hand, it protects the privacy of the original network topology information, which is the motivation behind the design of the dual noise structure.

In this paper, we transform the constrained optimization problem Eq.(9) into an unconstrained one. Our NMF model is defined as follows

$$\min_{\mathbf{W},\mathbf{X},\mathbf{S}\geq 0} \alpha||\mathbf{A}-\mathbf{W}||_{2,1} + ||\mathbf{W}-\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}||_F^2 + \lambda \sum_{i=1}^{n}(\sum_{j=1}^{k}\mathbf{X}_{ij}-1)^2 \tag{10}$$

where $\lambda > 0$ is a sparse parameter. One rationale for this approach is that it offers a simplified optimization problem which may be more easily solved. Furthermore, it hypothesizes that some nodes in the network have a higher probability (or edge weight for SBM) of connecting edges to all other nodes, implying greater centrality.

Algorithm 1 shows the overall procedure of our proposed method. Given the adjacency matrix $\mathbf{A}$ of a network and the factorization rank k (i.e. the number of communities), a modified nonnegative double singular value decomposition (Tosyali et al. (2019)) procedure is used first to initialize $\mathbf{X}$ and $\mathbf{S}$, as suggested in (Chen et al. (2024)), while $\mathbf{W}$ is initialized as $\mathbf{1}_{n\times n}$. Subsequently, a multiplication iteration update rule is employed, with a detailed explanation provided in Section 5. Finally, it returns the membership matrix $\mathbf{X}$ to obtain a clustering result.

Algorithm 1 Proposed Nonnegative Matrix Factorization

Input: $\mathbf{A}$: the adjacency matrix of the network; n: the number of nodes in the network; k: factorization rank;

Output: $\mathbf{X}$: the membership matrix;

1: Initialize $\mathbf{W} \leftarrow \mathbf{1}_{n\times n}$;
2: Initialize $\mathbf{X}, \mathbf{S} \leftarrow \mathrm{MNNDSVD}(\mathbf{A}, k)$;
3: **while** condition **do**
4: Updating $\mathbf{W}, \mathbf{X}, \mathbf{S}$ according to Eq.(13);
5: **end while**

5. Optimization Solution

In this section, the solution to the proposed method is presented in detail. The objective function Eq.(10) can be written as

$$
\begin{aligned}
\mathcal{L}(\mathbf{W}, \mathbf{X}, \mathbf{S}) &= \alpha||\mathbf{A} - \mathbf{W}||_{2,1} + ||\mathbf{W} - \mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}||_F^2 + \lambda \sum_{i=1}^{n} (\sum_{j=1}^{k} \mathbf{X}_{ij} - 1)^2 \\
&= \alpha tr((\mathbf{A} - \mathbf{W})\mathbf{Q}(\mathbf{A} - \mathbf{W})^{\mathrm{T}}) \\
&+ tr(\mathbf{W}\mathbf{W}^{\mathrm{T}}) - 2tr(\mathbf{W}\mathbf{X}\mathbf{S}^{\mathbf{T}}\mathbf{X}^{\mathrm{T}}) + tr(\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}) \\
&+ \lambda tr(\mathbf{X}\mathbf{1}_{k\times k}\mathbf{X}^{\mathrm{T}}) - 2\lambda tr(\mathbf{X}\mathbf{1}_{k\times n}) + \lambda n
\end{aligned}
\tag{11}
$$

where $\mathbf{Q}$ is a diagnal matrix with elements

$$
\mathbf{Q}_{ii} = (\sum_{j=1}^{n} (\mathbf{A}_{ij} - \mathbf{W}_{ij})^2)^{-\frac{1}{2}} \ , \quad i = 1, 2, \ldots, n
\tag{12}
$$

Then, matrices $\mathbf{W}, \mathbf{X}, \mathbf{S}$ are solved using the following multiplicative update rule by iterated

$$
\begin{aligned}
\mathbf{W}_{ij} &\leftarrow \mathbf{W}_{ij} \cdot \left(\frac{(\alpha\mathbf{A}\mathbf{Q} + \mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}})_{ij}}{(\alpha\mathbf{W}\mathbf{Q} + \mathbf{W})_{ij}} \right)^{\frac{1}{2}} \\
\mathbf{X}_{il} &\leftarrow \mathbf{X}_{il} \cdot \left(\frac{(\mathbf{W}^{\mathrm{T}}\mathbf{X}\mathbf{S} + \mathbf{W}\mathbf{X}\mathbf{S}^{\mathrm{T}})_{il} + \lambda}{(\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S}^{\mathrm{T}} + \mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S} + \lambda\mathbf{X}\mathbf{1}_{k\times k})_{il}} \right)^{\frac{1}{4}} \\
\mathbf{S}_{lj} &\leftarrow \mathbf{S}_{lj} \cdot \frac{(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})_{lj}}{(\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\mathbf{X})_{lj}}
\end{aligned}
\tag{13}
$$

Theorem 1. *The objective function in Eq.(10) is non-increasing under the update rules in (13).*

The proof of the above theorem is divided into two parts, *correctness* and *convergence*. The *correctness* of the update rules is illustrated in the following theorem, see Appendix A.

Theorem 2. *If the iterative update rule (13) converges, the final solution satisfies the Karush–Kuhn–Tucker (KKT) optimality condition.*

We use the auxiliary function approach to prove the *convergence* of (13), following (Lee and Seung (2000)). The exact proof process is similar to that used in (Wang et al. (2011)), as described in Appendix B.

5.1. Complexity calculation

In this subsection, we will investigate the computational complexity of the algorithm. As shown in Algorithm 1, the proposed method is mainly composed of step 2 and step 4. For step 2, the NNDSVD procedure computes the approximation of a matrix factorized by two nonnegative matrices with $\mathcal{O}(n^2k)$. For step 4 in the while loop, the update of each matrix $\mathbf{W}, \mathbf{X}, \mathbf{S}$ takes $\mathcal{O}(n^2k)$ time in each iteration. Therefore, the total time complexity of Algorithm 1 is of order $\mathcal{O}(tn^2k)$, where t is the number of iterations.

6. Experiment Results

This section presents the results of an empirical evaluation of the effectiveness of our proposed model, in comparison with several state-of-the-art methods. To evaluate the performance of the proposed model, numerical experiments have been conducted on eight real-world directed networks.

6.1. Datasets

In order to ascertain the efficacy of the proposed algorithm, we conducted experiments on several real-world directed networks. These networks include citation networks and a communication network. The specific information of the real networks is exhibited in Table 2. RC represents the actual number of communities present in the networks.

Datasets		NMF	SNMF	ANMF	DANMF	SDNMF	**MsSNMF**	**MsANMF**
Texas	NMI	0.1470	0.1549	0.1457	0.1344	0.0932	0.1819	**0.2184**
	ARI	0.1834	0.1012	0.1125	0.1993	0.1464	0.2189	**0.2683**
Cornell	NMI	0.1339	0.1628	0.1452	0.0980	0.1056	0.1541	**0.1726**
	ARI	0.0519	0.0278	0.0944	0.0940	0.0732	0.0895	**0.0950**
Wisconsin	NMI	0.0788	0.0635	0.0680	0.0899	0.0668	0.0786	**0.0967**
	ARI	0.0294	0.0512	0.0401	0.0538	**0.1073**	0.0649	0.0781

Table 1: NMI and ARI results on real-world networks

Datasets	Nodes	Edges	RC
Texas	187	309	5
Cornell	195	301	5
Wisconsin	197	502	5

Table 2: Real-world network details.

Datasets		BTLSC	GSNMF	DNBNMF	SDNMF	MvDGNMF	DAutoED-ONMF	**proposed**
Texas	NMI	**0.3121**	0.1823	0.2388	0.1681	0.1527	0.1490	0.2664
	ACC	**0.6673**	0.5379	0.6039	0.4743	0.4562	0.4954	0.6339
Cornell	NMI	0.1767	0.1461	0.1284	0.1428	0.1266	0.1573	**0.3235**
	ACC	0.5179	0.3760	0.4205	0.3602	0.3416	0.4169	**0.5792**
Wisconsin	NMI	0.1357	0.1564	0.1006	0.0901	0.0948	0.1051	**0.2900**
	ACC	**0.5458**	0.4932	0.4924	0.4469	0.4483	0.4573	0.5139
Cora	NMI	0.3204	0.2170	0.2021	0.2464	0.2833	0.2518	**0.3421**
	ACC	0.4878	0.4184	0.3504	0.4504	0.4308	0.4341	**0.5126**

Table 3: NMI and ACC results on real-world attributed networks

Several synthetic networks[1] have also been adopted to test the performance of the proposed approach. Girvan–Newman benchmark (GNB, Newman and Girvan (2004)) networks are generated with communities of equal size, where each node has a fixed number of internal, z_i, and external, z_e, edges. Planted partition benchmark (PPB, Condon and Karp (2001)) networks are generated with variable cluster sizes, internal edge probability p_i and external edge probability p_e. The characteristics and ground truth of synthetic networks under testing are presented in Table 4.

Network	Nodes	RC	community boundary	z_i	z_e	p_i	p_e
GNB1	200	8	[0, 25, 50, 75, 100, 125, 150, 175, 200]	20	25	-	-
GNB2	200	8	[0, 25, 50, 75, 100, 125, 150, 175, 200]	20	20	-	-
GNB3	200	8	[0, 25, 50, 75, 100, 125, 150, 175, 200]	15	20	-	-
GNB4	200	8	[0, 25, 50, 75, 100, 125, 150, 175, 200]	15	15	-	-
PPB1	300	10	[0, 30, 50, 70, 100, 140, 170, 200, 240, 260, 300]	-	-	0.7	0.4
PPB2	300	5	[0, 30, 70, 140, 260, 300]	-	-	0.7	0.4
PPB3	300	10	[0, 30, 50, 70, 100, 140, 170, 200, 240, 260, 300]	-	-	0.7	0.2
PPB4	300	5	[0, 30, 70, 140, 260, 300]	-	-	0.7	0.2

Table 4: Synthetic network details.

[1]https://github.com/mmitalidis/ComDetTB/tree/master/Graphs

6.2. Baseline Methods

This paper adopts the following eight baseline methods for comparison. These methods include algorithms based on NMF models, four of which are based on asymmetric ones, and two non-NMF-based methods.

- NMF (Lee and Seung (1999)): The originally proposed NMF.
- PNMF (Yuan and Oja (2005)): Projective NMF is a variant of NMF that learns sparse, spatially localized, part-based subspace representations.
- ONMF (Ding et al. (2006)): Orthogonal NMF model constrains the latent matrix to be orthogonal.
- ANMF (Wang et al. (2011)): The originally proposed asymmetric NMF model.
- RANMF (Tosyali et al. (2019)): Regularized asymmetric NMF model.
- MtriNMF (Yan and Chang (2019)): ANMF-based method using modularity matrix formulation as regularization term.
- RAsNMF (Abdollahi et al. (2020)): Regularized asymmetric semi-NMF model introduces a clustering approach by relaxing the nonnegativity constraints, inspired by Ding et al. (2008).
- SGNMF (Liu et al. (2023)): Symmetry and graph-regularized NMF model introduces a symmetry regularization term which implies the equality constraint between its multiple latent factor matrices.
- HEA (Ghouchan Nezhad Noor Nia et al. (2022)): A Louvain-based algorithm called High-entropy Alloy detects communities by enhanced particle swarm optimization.
- GEE (Shen et al. (2023)): Simultaneous vertex embedding and cluster detection using a normalized single-point graph encoder and a rank-based clustering scale metric.

For methods based on asymmetric NMF (including ANMF, RANMF, RAsNMF and the proposed NMF), the MNNDSVD approach (Tosyali et al. (2019)) is adopted for initialization of matrices. The randomly generated matrices on the scale (0, 0.5) are used to initialize the other methods. The iteration number is set to 500 for all methods. Each parameter is tuned in the range of $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$. The similarity matrix defaults to the cosine similarity matrix. For methods based on non-NMF baseline methods (HEA[2], GEE[3]), we followed the experimental setups described in their original publications to ensure a faithful reproduction of their reported performance.

6.3. Evaluation Metrics

To evaluate how effective algorithms are at clustering when correct labels exist, three metrics are used to measure performance, consisting of Jaccard similarity, normalization mutual information (NMI) and clustering accuracy (ACC). The values of these metrics are normalized between $[0, 1]$, with the value of 1 indicating an optimal detected structure. These metrics were also adopted to measure clustering results in several recent works (Tosyali et al. (2019), Hajiveiseh et al. (2024)).

6.4. Community Detection Results

This section shows the performance of the proposed and compared methods detecting community structure on the seven directed graphs. As demonstrated in Tables 5 - 6, the results are presented with the optimal performance highlighted in bold and the second-best performance indicated by underlining.[4]

It is noted that the proposed method outperforms comparison methods on *Dolphin*, *Cornell*, *Washington* networks, and ranks at least second in specific evaluation indicators on the remaining networks. In a study of 24 different cases, the proposed method demonstrated the highest performance in 13 cases and the second-best performance in 6 of the remaining 11 cases compared to all other methods. When all three evaluation metrics are taken into account, the performance of our algorithm is comparable to, if not better than, the other baseline methods. Even on the most challenging *Email* dataset, our algorithm still demonstrates robust competitiveness, with its overall performance comparable to (if not superior to) any existing baseline

[2]https://github.com/rghoochannejad/HEAs-Community-Detection

[3]http://github.com/cshen6/GraphEmd

[4]Due to ACC's strict requirement of matching cluster quantities, it was omitted from HEA's evaluation, as the algorithm dynamically adjusts the number of communities.

method. Notably, the proposed method exhibits robust performance even on the large-scale network *BlogCatalog*, demonstrating its scalability and effectiveness. In addition, the HEA algorithm's variable number of clusters leads to strong NMI performance on datasets *Wisconsin* and *Wiki*, yet its Jaccard scores consistently trail those of NMF-based methods. The experimental results also demonstrate that GEE underperforms most NMF-based algorithms in community detection tasks on these datasets.

In comparison to the other methods based on ANMF, the proposed one significantly better than ANMF and RAsNMF, slightly better than RANMF. These methods, initialized by MNNDSVD, are highly effective on the testing directed networks, as shown in experimental results. It again validates that this initialization procedure can help the algorithm to converge on better solutions. Therefore, in the experiments in Section 6.6, we will continue to focus on comparing these ANMF-based methods initialized by MNNDSVD.

Method	Dolphin	Texas	Cornell	Wisconsin	Washington	Email	Wiki	BlogCatalog
NMF	0.8900	0.3206	0.2233	0.2243	0.2992	0.3237	0.1145	0.1653
PNMF	0.8900	0.3170	0.2169	0.2398	0.2826	0.3304	0.1200	0.1795
ONMF	0.8900	0.3034	0.2212	0.2240	0.2809	0.3667	0.1227	0.1830
ANMF	0.8900	0.3731	0.2981	0.2190	0.3041	0.3581	0.1220	0.1873
RANMF	0.8900	**0.4195**	0.3166	0.2667	0.3835	0.3716	**0.1275**	0.1872
MtriNMF	0.8900	0.3920	0.3158	0.2534	0.3091	0.3691	0.1213	0.1873
RAsNMF	0.8407	0.4046	0.1917	0.2343	0.3011	0.1397	0.1165	0.1826
SGNMF	**0.9430**	0.3783	0.2697	0.2782	0.3549	**0.3807**	0.1232	0.1818
HEA	0.8920	0.2458	0.1979	0.2018	0.2259	0.0433	0.0548	0.1603
GEE	0.7421	0.2870	0.1969	0.2001	0.2166	0.2697	0.1019	0.1619
proposed	**0.9430**	0.3964	**0.3458**	**0.2998**	**0.4039**	0.3758	0.1237	**0.1948**

Table 5: Jaccard results on real-world datasets.

Method	Dolphin	Texas	Cornell	Wisconsin	Washington	Email	Wiki	BlogCatalog
NMF	0.9677	0.5187	0.4308	0.4189	0.5391	0.5701	0.3356	0.3728
PNMF	0.9677	0.5080	0.4051	0.4264	0.5174	0.5801	0.3422	0.2242
ONMF	0.9677	0.4920	0.4103	0.4151	0.5087	0.6070	0.3185	0.3811
ANMF	0.9677	0.5722	0.4821	0.0506	0.5087	0.6219	0.3414	0.3832
RANMF	0.9677	**0.6043**	0.5282	**0.5094**	0.5957	**0.6338**	**0.3692**	0.3832
MtriNMF	0.9677	**0.6043**	0.5179	0.4566	0.5304	0.6209	0.3397	0.3832
RAsNMF	0.9516	0.5989	0.3538	0.4151	0.4739	0.4010	0.3347	0.3915
SGNMF	**0.9839**	0.5722	0.4308	0.4717	0.5826	0.6159	0.3426	0.3768
GEE	0.9032	0.4920	0.3692	0.3660	0.4304	0.5035	0.2732	0.3264
proposed	**0.9839**	**0.6043**	**0.5641**	0.5057	**0.6478**	0.6189	0.3534	**0.4443**

Table 6: Accuracy results on real-world datasets.

6.5. Parameter Sensitivity

This section is concerned with an analysis of the impact of hyperparameters α and λ on the performance of the model. The two parameters utilized in the origin model, respectively, correspond to the intensity weight of the noise and the regularization term. The candidate sets of α and λ are set as $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$ and $\{0, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$, respectively. Fig.1 shows the Jaccard, NMI and Accuracy of the proposed method on World Wide Knowledge Base (WebKB) datasets (including *Texas*, *Cornell*, *Wisconsin*, *Washington*) with the different tested values. In the figure, a darker color represents a higher value of the metric, indicating a better clustering result. According to the results, the proposed model is sensitive to the parameter pair, and the values of α and λ are data dependent. In cases where real-world datasets do not have ground-truth labels, appropriate parameter values can be selected according to the evaluated data sets with similar characteristics. It is also feasible to use metrics that do not require real labels, Modularity (Newman (2004)), for example, to evaluate clustering results obtained using different parameters.

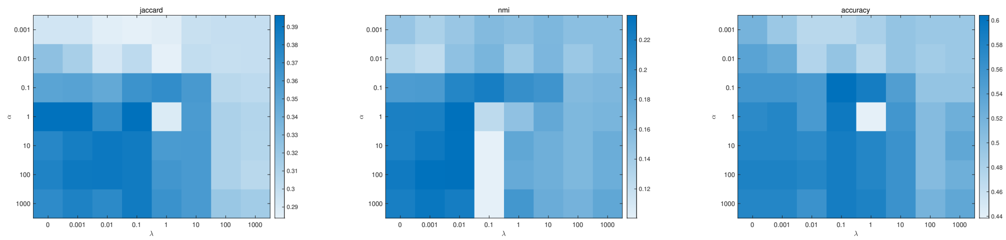
6.6. Comparison with Variants

In this section, three variants are designed to facilitate an in-depth examination of the proposed model. Each variant employs an alternative extra strategy, and experiments are employed on synthetic networks. As a basis for comparison, asymmetric NMF methods were selected. Variant-1 adopts the second-order proximity matrix (Yang et al. (2015)) as input instead of the adjacency matrix, as suggested in Wang et al. (2017) and Al-sharoa and Rahahleh (2023). To be specific, the second-order proximity matrix $\mathbf{K}$ of a network with adjacency matrix $\mathbf{A}$ is defined as follows:

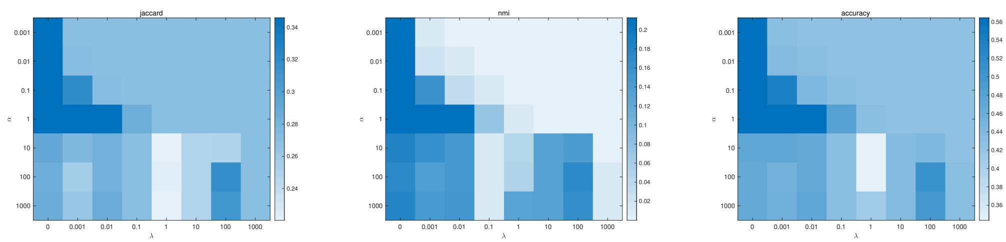
$$\mathbf{K} = \frac{\mathbf{J} + \mathbf{J}^2}{2}, \qquad \mathbf{J}_{ij} = \frac{\mathbf{A}_{ij}}{\sum_{j=1}^{n} \mathbf{A}_{ij}} \tag{14}$$

where $\mathbf{J}_{ij}$ is called the first-order proximity matrix. Variant-2 fixes the matrix $\mathbf{W}$ equal to the input matrix $\mathbf{A}$ and no longer updates it, which is equivalent to optimizing the following objective function

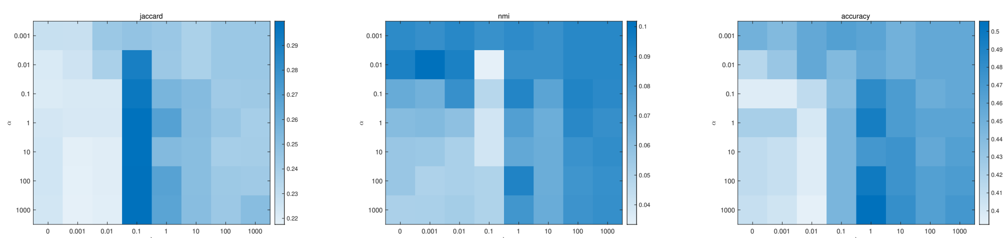
$$\min_{\mathbf{X},\mathbf{S}\geq 0} ||\mathbf{A} - \mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}||_F^2 + \lambda \sum_{i=1}^{n} (\sum_{j=1}^{k} \mathbf{X}_{ij} - 1)^2 \tag{15}$$



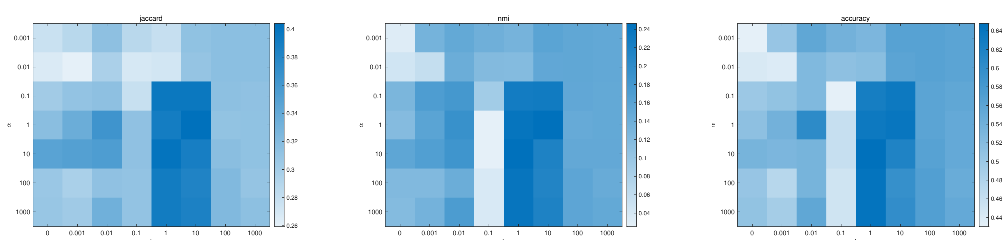
(a) Texas



(b) Cornell



(c) Wisconsin



(d) Washington

Figure 1: Parameter analysis (in terms of Jaccard, NMI, and Accuracy) on α and λ

Formally, it is asymmetric NMF with a regular term. Variant-3 uses the Frobenius norm to measure the difference between the matrices $\mathbf{A}$ and $\mathbf{W}$, and solves the following optimization problem

$$\min_{\mathbf{W},\mathbf{X},\mathbf{S}\geq 0} \alpha||\mathbf{A}-\mathbf{W}||_F^2 + ||\mathbf{W}-\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}||_F^2 + \lambda \sum_{i=1}^{n}(\sum_{j=1}^{k}\mathbf{X}_{ij}-1)^2 \tag{16}$$

with updating matrix $\mathbf{W}$ by iterating

$$\mathbf{W}_{ij} \leftarrow \mathbf{W}_{ij} \cdot \left(\frac{(\alpha\mathbf{A}+\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}})_{ij}}{(\alpha\mathbf{W}+\mathbf{W})_{ij}}\right)^{\frac{1}{2}} \tag{17}$$

instead. The experimental results are presented in Table 7.

It can be found that variant-1 significantly outperforms the other methods, obtaining optimal results on five of these networks, indicating that edge-weighted by using second-order proximity matrix helps detect community structure by learning higher-order connection information. For networks with the same community labels, variant-1 maintains good performance as the number of external edges z_e for GNB (or the probability of external edges p_e for PPB) increases, indicating its strong robustness under different noise levels. In addition, the proposed algorithm also outperforms variant-2 and variant-3 on each of the synthetic networks, suggesting that the original objective function and the modelling of the noise are more appropriate.

7. Conclusion

In this paper, a robust asymmetric NMF approach is proposed for the community detection task in directed networks based on a rigorous statistical interpretation that clustering by NMF is equivalent to likelihood maximization of a stochastic block model (Zhang et al. (2018)). By modelling the input matrix (suggested to be the adjacency or second-proximity matrix) as generated by a two-step noise addition process, the objective function is given with two parameters to be tuned. Extensive experimentation has been conducted in both real-world and synthetic networks to demonstrate the superiority of the proposed model.

For future work, the proposed approach will be considered to combine with the deep NMF method, to use complex hierarchical information for clustering (Hajiveiseh et al. (2024)). Additionally, we also consider applying the NMF method to other problems, such as the link prediction problem (Lv et al. (2022); Tang and Wang (2022)).

	Metric	ANMF	RANMF	RAsNMF	**proposed**	variant-1	variant-2	variant-3
GNB1	Jaccard	0.1866	0.1882	0.1058	0.7756	**1.0000**	0.7756	0.7634
	NMI	0.3488	0.3153	0.0675	0.9426	**1.0000**	0.9426	0.9289
	Accuracy	0.3250	0.2750	0.2167	0.8750	**1.0000**	0.8750	0.8750
GNB2	Jaccard	0.2456	0.3287	0.0930	0.7306	**1.0000**	0.7050	0.7087
	NMI	0.4668	0.6361	0.0846	0.9281	**1.0000**	0.9201	0.9211
	Accuracy	0.3750	0.4583	0.2333	0.8500	**1.0000**	0.8167	0.8250
GNB3	Jaccard	0.2248	0.2763	0.0989	0.5370	**1.0000**	0.5389	0.5341
	NMI	0.4539	0.5630	0.0510	0.8100	**1.0000**	0.8066	0.8080
	Accuracy	0.3417	0.3750	0.2167	0.7167	**1.0000**	0.7167	0.7167
GNB4	Jaccard	0.7887	0.7887	0.0977	**1.0000**	**1.0000**	**1.0000**	**1.0000**
	NMI	0.9565	0.9565	0.2066	**1.0000**	**1.0000**	**1.0000**	**1.0000**
	Accuracy	0.8750	0.8750	0.3250	**1.0000**	**1.0000**	**1.0000**	**1.0000**
PPB1	Jaccard	0.0612	0.1037	0.0909	0.1049	**0.2221**	0.0968	0.1047
	NMI	0.0662	0.0000	0.0339	0.0845	**0.3757**	0.0430	0.0566
	Accuracy	0.1800	0.1333	0.1800	0.2000	**0.4667**	0.1667	0.1800
PPB2	Jaccard	0.1489	0.2575	0.2066	0.5464	**0.6430**	0.5383	0.5362
	NMI	0.0212	0.0000	0.0177	0.6270	**0.6818**	0.6047	0.6174
	Accuracy	0.3100	0.4000	0.3433	0.6667	**0.8267**	0.6700	0.6667
PPB3	Jaccard	0.1428	0.1214	0.0874	0.5530	**0.8457**	0.5505	0.5505
	NMI	0.2483	0.1968	0.0230	0.7988	**0.9466**	0.7986	0.7986
	Accuracy	0.3000	0.2667	0.1600	0.7033	**0.9200**	0.7000	0.7000
PPB4	Jaccard	0.4968	0.8573	0.2040	0.8760	**1.0000**	0.7174	0.7161
	NMI	0.6270	0.8919	0.0193	0.9230	**1.0000**	0.7939	0.7939
	Accuracy	0.6333	0.9167	0.3533	0.8667	**1.0000**	0.8033	0.8033

Table 7: Experimental results on synthetic networks.

Acknowledgment

This work was supported by the the National Key R&D Program of China under Grant 2023YFC2605604, the National Natural Science Foundation of China under Grants 62173027.

Appendix A. Proof of correctness

By introducing the Lagrangian multipliers $\beta_1, \beta_2, \beta_3$ for the nonnegativity of matrices $\mathbf{W}, \mathbf{X}, \mathbf{S}$, the Lagrangian function can be constructed as

$$\begin{aligned} L(\mathbf{W}, \mathbf{X}, \mathbf{S}) &= \mathcal{L}(\mathbf{W}, \mathbf{X}, \mathbf{S}) - tr(\beta_1 \mathbf{W}^{\mathrm{T}}) - tr(\beta_2 \mathbf{X}^{\mathrm{T}}) - tr(\beta_3 \mathbf{S}^{\mathrm{T}}) \\ &= \alpha tr((\mathbf{A} - \mathbf{W})\mathbf{Q}(\mathbf{A} - \mathbf{W})^{\mathrm{T}}) \\ &+ tr(\mathbf{W}\mathbf{W}^{\mathrm{T}}) - 2tr(\mathbf{W}\mathbf{X}\mathbf{S}^{\mathbf{T}}\mathbf{X}^{\mathrm{T}}) + tr(\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}) \\ &+ \lambda tr(\mathbf{X}\mathbf{1}_{k\times k}\mathbf{X}^{\mathrm{T}}) - 2\lambda tr(\mathbf{X}\mathbf{1}_{k\times n}) + \lambda n \\ &- tr(\beta_1 \mathbf{W}^{\mathrm{T}}) - tr(\beta_2 \mathbf{X}^{\mathrm{T}}) - tr(\beta_3 \mathbf{S}^{\mathrm{T}}) \end{aligned} \tag{A.1}$$

where $\mathcal{L}$ is the objective function of the proposed NMF, given in (11). Then we have

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{W}} &= 2(-\alpha\mathbf{A}\mathbf{Q} + \alpha\mathbf{W}\mathbf{Q} + \mathbf{W} - \mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}) - \beta_1 \\ \frac{\partial L}{\partial \mathbf{X}} &= 2(-\mathbf{W}\mathbf{X}\mathbf{S}^{\mathrm{T}} - \mathbf{W}^{\mathrm{T}}\mathbf{X}\mathbf{S} + \mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S}^{\mathrm{T}} + \mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S} + \lambda\mathbf{X}\mathbf{1}_{k\times k} - \lambda\mathbf{1}_{n\times k}) - \beta_2 \\ \frac{\partial L}{\partial \mathbf{S}} &= 2(-\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X} + \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\mathbf{S}) - \beta_3 \end{aligned} \tag{A.2}$$

Let the above equations take 0, and follow the KKT complementary slackness condition,

$$\begin{aligned} \beta_{1ij}\mathbf{W}_{ij} &= 2(-\alpha\mathbf{A}\mathbf{Q} + \alpha\mathbf{W}\mathbf{Q} + \mathbf{W} - \mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}})_{ij}\mathbf{W}_{ij} = 0 \\ \beta_{2il}\mathbf{X}_{il} &= 2(-\mathbf{W}\mathbf{X}\mathbf{S}^{\mathrm{T}} - \mathbf{W}^{\mathrm{T}}\mathbf{X}\mathbf{S} + \mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S}^{\mathrm{T}} + \mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S} + \lambda\mathbf{X}\mathbf{1}_{k\times k} - \lambda\mathbf{1}_{n\times k})_{il}\mathbf{X}_{il} = 0 \\ \beta_{3lj}\mathbf{S}_{lj} &= 2(-\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X} + \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\mathbf{S})_{lj}\mathbf{S}_{lj} = 0 \end{aligned} \tag{A.3}$$

It can be seen that the update rule (13) satisfied KKT conditions. Furthermore, the final matrices $\mathbf{W}, \mathbf{X}, \mathbf{S}$ would be nonnegative, since these matrices are all nonnegative during the updating process. In summary, the correctness of the algorithm is proved.

Appendix B. Proof of convergence

Definition 1. *A function $Q(x,\tilde{x})$ is an auxiliary function of $F(x)$, if the following condition holds*

$$Q(x,\tilde{x}) \geq F(x), \quad Q(x,x) = F(x) \tag{B.1}$$

If Q is the auxiliary function of F, then F is non-increasing under the update rule

$$x^{(t+1)} = \arg\min_{x} Q(x, x^{(t)}) \tag{B.2}$$

since $F(x^{(t+1)}) \leq Q(x^{(t+1)}, x^{(t)}) \leq Q(x^{(t)}, x^{(t)}) = F(x^{(t)})$.

Fixing $\mathbf{X}$ and $\mathbf{S}$ in the objective function $\mathcal{L}$ given by (11), matrix $\mathbf{W}$ can be updated by minimizing F_1 with

$$\begin{aligned}
F_1(\mathbf{W}) &= -2\alpha tr(\mathbf{W}\mathbf{Q}\mathbf{A}^{\mathrm{T}}) + \alpha tr(\mathbf{W}\mathbf{Q}\mathbf{W}^{\mathrm{T}}) + tr(\mathbf{W}\mathbf{W}^{\mathrm{T}}) - 2tr(\mathbf{W}\mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}) \\
&\leq -2\alpha tr(\tilde{\mathbf{W}}\mathbf{Q}\mathbf{A}^{\mathrm{T}} + \mathbf{P}\mathbf{Q}\mathbf{A}^{\mathrm{T}}) + \alpha tr(\mathbf{V}\mathbf{Q}\tilde{\mathbf{W}}^{\mathrm{T}}) + tr(\mathbf{V}\tilde{\mathbf{W}}^{\mathrm{T}}) \\
&\quad - 2tr(\tilde{\mathbf{W}}\mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}} + \mathbf{P}\mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}) \\
&\overset{\text{def}}{=} G_1(\mathbf{W}, \tilde{\mathbf{W}})
\end{aligned} \tag{B.3}$$

where $(\mathbf{P}_{n\times n})_{ij} = \tilde{\mathbf{W}}_{ij} \ln \mathbf{W}_{ij}/\tilde{\mathbf{W}}_{ij}$ and $(\mathbf{V}_{n\times n})_{ij} = \mathbf{W}_{ij}^2/\tilde{\mathbf{W}}_{ij}$ and the equality holds when $\mathbf{W} = \tilde{\mathbf{W}}$, shows that G_1 satisfied the condition to be an auxiliary function of F_1. Then update rule for $\mathbf{W}$ in (13) can be obtained by KKT condition

$$\frac{\partial G_1}{\partial \mathbf{W}_{ij}} = 2\frac{\mathbf{W}_{ij}}{\tilde{\mathbf{W}}_{ij}}\left(\alpha\tilde{\mathbf{W}}\mathbf{Q} + \tilde{\mathbf{W}}\right)_{ij} - 2\frac{\tilde{\mathbf{W}}_{ij}}{\mathbf{W}_{ij}}\left(\alpha\mathbf{A}\mathbf{Q} + \mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\right)_{ij} = 0 \tag{B.4}$$

Fixing $\mathbf{W}$ and $\mathbf{S}$, $\mathbf{X}$ can be updated by minimizing

$$\begin{aligned}
F_2(\mathbf{X}) &= -2tr(\mathbf{W}\mathbf{X}\mathbf{S}^{\mathbf{T}}\mathbf{X}^{\mathrm{T}}) + tr(\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}) + \lambda tr(\mathbf{X}\mathbf{1}_{k\times k}\mathbf{X}^{\mathrm{T}}) - 2\lambda tr(\mathbf{X}\mathbf{1}_{k\times n}) \\
&\leq -2tr(\mathbf{W}\tilde{\mathbf{X}}\mathbf{S}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}} + \mathbf{W}\mathbf{Z}\mathbf{S}^{\mathrm{T}}\tilde{\mathbf{X}}^{\mathrm{T}} + \mathbf{W}\tilde{\mathbf{X}}\mathbf{S}^{\mathrm{T}}\tilde{\mathbf{X}}^{\mathrm{T}}) \\
&\quad + \frac{1}{2}tr(\mathbf{Y}\mathbf{S}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\mathbf{S}^{\mathrm{T}}\tilde{\mathbf{X}}^{\mathrm{T}} + \mathbf{Y}\mathbf{S}^{\mathrm{T}}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\mathbf{S}\tilde{\mathbf{X}}^{\mathrm{T}}) \\
&\quad + \frac{\lambda}{2}tr(\mathbf{Y}\mathbf{1}_{k\times k}\tilde{\mathbf{X}}^{\mathrm{T}} + \tilde{\mathbf{X}}\mathbf{1}_{k\times k}\tilde{\mathbf{X}}^{\mathrm{T}}) \\
&\quad - 2\lambda tr(\mathbf{Z}\mathbf{1}_{k\times n} + \tilde{\mathbf{X}}\mathbf{1}_{k\times n}) \\
&\overset{\text{def}}{=} G_2(\mathbf{X}, \tilde{\mathbf{X}})
\end{aligned} \tag{B.5}$$

where $(\mathbf{Y}_{n\times k})_{il} = \mathbf{X}_{il}^4/\tilde{\mathbf{X}}_{il}^3$ and $(\mathbf{Z}_{n\times k})_{il} = \tilde{\mathbf{X}}_{il}\ln \mathbf{X}_{il}/\tilde{\mathbf{X}}_{il}$, the equality holds when $\mathbf{X} = \tilde{\mathbf{X}}$. G_2 is an auxiliary function of F_2, then the KKT condition

$$\begin{aligned}\frac{\partial G_2}{\partial \mathbf{X}_{il}} =& 2\frac{\mathbf{X}_{il}^3}{\tilde{\mathbf{X}}_{il}^3}\left(\tilde{\mathbf{X}}\mathbf{S}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\mathbf{S}^{\mathrm{T}} + \tilde{\mathbf{X}}\mathbf{S}^{\mathrm{T}}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\mathbf{S} + \lambda\tilde{\mathbf{X}}\mathbf{1}_{k\times k}\right)_{il} \\ &-2\frac{\tilde{\mathbf{X}}_{il}}{\mathbf{X}_{il}}\left(\mathbf{W}\tilde{\mathbf{X}}\mathbf{S}^{\mathrm{T}} + \mathbf{W}^{\mathrm{T}}\tilde{\mathbf{X}}\mathbf{S} + \lambda\mathbf{1}_{n\times k}\right)_{il} = 0\end{aligned} \tag{B.6}$$

gives the update rule for $\mathbf{X}$ in (13).

Fixing $\mathbf{W}$ and $\mathbf{X}$, the objective function can be fomulated as

$$\begin{aligned}F_3(\mathbf{S}) &= -2tr(\mathbf{W}\mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}) + tr(\mathbf{X}\mathbf{S}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}) \\ &\leq -2tr(\mathbf{W}\mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}) + tr(\mathbf{X}\tilde{\mathbf{S}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{T}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}) \\ &\stackrel{\text{def}}{=} G_3(\mathbf{S}, \tilde{\mathbf{S}})\end{aligned} \tag{B.7}$$

where $(\mathbf{T}_{k\times k})_{lj} = \mathbf{S}_{lj}^2/\tilde{\mathbf{S}}_{lj}$. Let G_3 be the auxiliary function of F_3, then update rule for $\mathbf{S}$ in (13) can be obtained by KKT condition

$$\frac{\partial G_3}{\partial \mathbf{S}_{lj}} = 2\frac{\mathbf{S}_{lj}}{\tilde{\mathbf{S}}_{lj}}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\tilde{\mathbf{S}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)_{lj} - 2\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)_{lj} = 0 \tag{B.8}$$

Therefore, we have

$$\begin{aligned}\mathcal{L}(\mathbf{W}^{(0)}, \mathbf{X}^{(0)}, \mathbf{S}^{(0)}) &\geq \mathcal{L}(\mathbf{W}^{(1)}, \mathbf{X}^{(0)}, \mathbf{S}^{(0)}) \geq \mathcal{L}(\mathbf{W}^{(1)}, \mathbf{X}^{(1)}, \mathbf{S}^{(0)}) \\ &\geq \mathcal{L}(\mathbf{W}^{(1)}, \mathbf{X}^{(1)}, \mathbf{S}^{(1)}) \geq \dots\end{aligned} \tag{B.9}$$

Therefore the objective function $\mathcal{L}$ is monotonically decreasing. Since $\mathcal{L}$ is bounded below, the convergence of the update rule is proved.

References

Abdollahi, R., Seyedi, S.A., Noorimehr, M.R., 2020. Asymmetric semi-nonnegative matrix factorization for directed graph clustering, in: 2020 10th International Conference on Computer and Knowledge Engineering (ICCKE), IEEE. pp. 323–328.

Aicher, C., Jacobs, A.Z., Clauset, A., 2015. Learning latent block structure in weighted networks. Journal of Complex Networks 3, 221–248.

Al-sharoa, E., Rahahleh, B., 2023. Community detection in networks through a deep robust auto-encoder nonnegative matrix factorization. Engineering Applications of Artificial Intelligence 118, 105657.

Argyriou, A., Evgeniou, T., Pontil, M., 2006. Multi-task feature learning. Advances in Neural Information Processing Systems 19.

Chen, Z., Xiao, Q., Leng, T., Zhang, Z., Pan, D., Liu, Y., Li, X., 2024. Multi-constraint non-negative matrix factorization for community detection: orthogonal regular sparse constraint non-negative matrix factorization. Complex & Intelligent Systems , 1–16.

Condon, A., Karp, R.M., 2001. Algorithms for graph partitioning on the planted partition model. Random Structures & Algorithms 18, 116–140.

Coscia, M., Giannotti, F., Pedreschi, D., 2011. A classification for community discovery methods in complex networks. Statistical Analysis and Data Mining: The ASA Data Science Journal 4, 512–546.

Ding, C., He, X., Simon, H.D., 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.

Ding, C., Li, T., Peng, W., Park, H., 2006. Orthogonal nonnegative matrix t-factorizations for clustering, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 126–135.

Ding, C.H., Li, T., Jordan, M.I., 2008. Convex and semi-nonnegative matrix factorizations. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 45–55.

Fortunato, S., 2010. Community detection in graphs. Physics Reports 486, 75–174.

Ghouchan Nezhad Noor Nia, R., Jalali, M., Mail, M., Ivanisenko, Y., Kübel, C., 2022. Machine learning approach to community detection in a high-entropy alloy interaction network. ACS Omega 7, 12978–12992.

Girvan, M., Newman, M.E., 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99, 7821–7826.

Hajiveiseh, A., Seyedi, S.A., Tab, F.A., 2024. Deep asymmetric nonnegative matrix factorization for graph clustering. Pattern Recognition 148, 110179.

He, C., Liu, H., Tang, Y., Fei, X., Li, H., Zhang, Q., 2020. Network embedding using deep robust nonnegative matrix factorization. IEEE Access 8, 85441–85453.

Kong, D., Ding, C., Huang, H., 2011. Robust nonnegative matrix factorization using l21-norm, in: Proceedings of the 20th ACM international Conference on Information and Knowledge Management, pp. 673–682.

Lee, D., Seung, H.S., 2000. Algorithms for non-negative matrix factorization. Advances in Neural Information Processing Systems 13.

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791.

Liu, Z., Luo, X., 2023. A constraints fusion-induced symmetric nonnegative matrix factorization approach for community detection. arXiv preprint arXiv:2302.12114 .

Liu, Z., Luo, X., Zhou, M., 2023. Symmetry and graph bi-regularized non-negative matrix factorization for precise community detection. IEEE Transactions on Automation Science and Engineering 21, 1406–1420.

Lv, L., Bardou, D., Hu, P., Liu, Y., Yu, G., 2022. Graph regularized nonnegative matrix factorization for link prediction in directed temporal networks using pagerank centrality. Chaos, Solitons & Fractals 159, 112107.

Malliaros, F.D., Vazirgiannis, M., 2013. Clustering and community detection in directed networks: A survey. Physics Reports 533, 95–142.

Newman, M.E., 2004. Fast algorithm for detecting community structure in networks. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics 69, 066133.

Newman, M.E., 2006. Modularity and community structure in networks. Proceedings of the National Academy of Sciences 103, 8577–8582.

Newman, M.E., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical Review E 69, 026113.

Shen, C., Park, Y., Priebe, C.E., 2023. Graph encoder ensemble for simultaneous vertex embedding and community detection, in: Proceedings of the 2023 2nd International Conference on Algorithms, Data Mining, and Information Technology, pp. 13–18.

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.

Tang, M., Wang, W., 2022. Cold-start link prediction integrating community information via multi-nonnegative matrix factorization. Chaos, Solitons & Fractals 162, 112421.

Tosyali, A., Kim, J., Choi, J., Jeong, M.K., 2019. Regularized asymmetric nonnegative matrix factorization for clustering in directed networks. Pattern Recognition Letters 125, 750–757.

Wang, F., Li, T., Wang, X., Zhu, S., Ding, C., 2011. Community discovery using nonnegative matrix factorization. Data Mining and Knowledge Discovery 22, 493–521.

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, pp. 203–209.

Yan, C., Chang, Z., 2019. Modularized tri-factor nonnegative matrix factorization for community detection enhancement. Physica A: Statistical Mechanics and its Applications 533, 122050.

Yang, C., Liu, Z., Zhao, D., Sun, M., Chang, E.Y., 2015. Network representation learning with rich text information., in: International Joint Conferences on Artificial Intelligence, pp. 2111–2117.

Ye, F., Chen, C., Wen, Z., Zheng, Z., Chen, W., Zhou, Y., 2019. Homophily preserving community detection. IEEE Transactions on Neural Networks and Learning Systems 31, 2903–2915.

Ye, F., Chen, C., Zheng, Z., 2018. Deep autoencoder-like nonnegative matrix factorization for community detection, in: Proceedings of the 27th ACM

International Conference on Information and Knowledge Management, pp. 1393–1402.

Yuan, Z., Oja, E., 2005. Projective nonnegative matrix factorization for image compression and feature extraction, in: Image Analysis: 14th Scandinavian Conference, SCIA 2005, Joensuu, Finland, June 19-22, 2005. Proceedings 14, Springer. pp. 333–342.

Zhang, M., Zhou, Z., 2020. Structural deep nonnegative matrix factorization for community detection. Applied Soft Computing 97, 106846.

Zhang, Z.Y., Gai, Y., Wang, Y.F., Cheng, H.M., Liu, X., 2018. On equivalence of likelihood maximization of stochastic block model and constrained nonnegative matrix factorization. Physica A: Statistical Mechanics and its Applications 503, 687–697.