Sparse Non-Negative Matrix Factorization with $L_{2,1}$ regularization for Multiplex Network Clustering

Mengge Chai, Long Zhang, Chao Gao, Xuelong Li, and Zhen Wang

***Abstract*—How to detect a consensus node cluster in multiplex networks and complex systems is of great significance for circuit system design and control. The robustness of sparsity constraints to the noise is often ignored in multiplex network clustering. In view of this, we propose a novel Sparse Collective Symmetric Non-negative Matrix Factorization with $L_{2,1}$ regularization (SCSNMF) to overcome the effect of noises on sparsity constraints. More specifically, SCSNMF first decomposes the unique low-dimensional representation of each layer, and then uses a collective way to merge these representations into a common representation where the community structure can be extracted. Especially, to avoid SCSNMF falling into the local optimum and maintain robustness, the $L_{2,1}$ regularization is added to all the decomposition processes to ensure the sparsity. Some experiments demonstrate the performance of proposed method in terms of the clustering accuracy and stability in real-world multiplex networks.**

***Index Terms*—Multiplex networks, community detection, symmetric non-negative matrix factorization.**

I. Introduction

COMPLEX networks are powerful tools for solving and analyzing complex systems and relationships [1]. Especially, the community detection is an important research branch of network science and is also widely used to analyze complex systems. For example, the failure area of the power grid can be predicted through the community structure [2] and the structural layout of the circuit system combined with the community can be more efficient [3]. A community structure usually refers to a group of nodes which are more tightly connected than the remaining ones in a network. Although two networks have the same global statistical characteristics, there might be significant differences in their community structures, which play different roles in networks. Therefore, identifying the community structures of networks provides a way to analyze the internal structure and function of complex systems.

Several algorithms have been proposed for detecting community structures on single-layer networks [4]–[7], which only contain one single type of connection between nodes. However, for practical systems, there might exist multiple relationships. Such systems include subsystems or layers of connectivity representing different modes of complexity. Therefore, the multiplex network clustering has recently attracted wide attentions [8]–[14]. Unlike the community structure in the single-layer network, multiplex communities capture the relations among nodes in different layers. Although the multiplex network has a consensus community structure, nodes in the consensus community have different types of interactions. However, it is difficult to integrate the data information in a multiplex network and reduce the impact of noises [8].

The sparse matrix is robust to noises [15] and the $L_{2,1}$ norm has the characteristic of limiting the sparsity of matrix. Therefore, there is no doubt that the $L_{2,1}$ regularization can benefit multiplex network clustering if added into the objective function. Taking Network Fusion for Composite Community Extraction (NF-CCE) [8] as an example, we propose a Sparse Collective Symmetric Non-negative Factorization with $L_{2,1}$ regularization (SCSNMF). More specifically, the $L_{2,1}$ regularization is introduced into every factorization process. In this way, our algorithm improves both the stability performance and the classification accuracy.

The rest of this paper is organized as follows. Sec. II summarizes the related work about multiplex network clustering. SCSNMF is formulated in Sec. III. Experiments and conclusions are presented in Sec. IV and Sec. V respectively.

II. Related work

Researches on community detection in multiplex networks can be broadly divided into matrix and tensor factorization method [8], [12], [13], information fusion method [9], [10], [14], and spectral clustering method [11]. Representative algorithms for each direction are described below.

Information Fusion Method: Similarity Network Fusion (SNF) [9], Graph-Based Multi-View Clustering (GMC) [10], and Feature Concatenation Multi-View Subspace Clustering (FCMSC) [14] are algorithms based on the information fusion. More specifically, in SNF algorithm, the information interacts between initial similarity matrix and the local affinity matrix. This fusion process can extract the local structure of graphs and improve the computational efficiency. GMC learns the data graph matrices of all views in a mutual reinforcement manner to generate a unified graph matrix, while multiple views are concatenated into a joint representation in FCMSC. However, information fusion methods are greatly affected by data noises.

Spectral Clustering: For spectral clustering method, the Spectral Clustering on Multi-Layer graphs (SC-ML) [11] is a representative algorithm. It defines a modified Laplacian matrix computed by every graph layer Laplacian matrix with a

Manuscript received xxx; revised xxxx. This work was supported in part by National Science Foundation under Grant... This brief was recommended by Associate Editor xxx. *(Corresponding author: xxxx)*

X. Li, C. Gao and L. Zhang are with the School of Computer Science, Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China (e-mail: li@nwpu.edu.cn; cgao@swu.edu.cn, and lzhang@mail.nwpu.edu.cn)

Z. Wang and M. Chai are with the School of Mechanical Engineering, Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China (e-mail: nkzhenwang@163.com and chaimengge@mail.nwpu.edu.cn).

distance constraint on the Grassmann manifold [16]. Finally, it utilizes spectral clustering to obtain network clusters. But SC-ML cannot cluster multiplex networks with isolated nodes.

Matrix and Tensor Factorization: A few matrix and tensor factorization methods have been proposed so far, such as Linked Matrix Factorization (LMF) [12], GraphFuse (GF) [13], and Collective Symmetric Non-negative Factorization (CSNMF) [8]. LMF decomposes the adjacency matrix of each layer into a standard low-dimensional matrix representing the common factor and different characteristic matrices. The GF algorithm considers the adjacency matrix of each layer of the multiplex network as a different slice of a three-way tensor and decomposes this tensor to obtain clusters. One drawback of the two algorithms mentioned above is that they cannot integrate the information of each layer well. CSNMF algorithm compensates for the shortcomings of all the above algorithms. CSNMF is a typical algorithm based on the NF-CCE framework. Through the adoption of this algorithm, a consensus low-dimensional representation can be derived accordingly, while this representation is able to maximize the common characteristics among all the low-dimensional representations for network layers. However, this algorithm is unable to guarantee the sparsity of the matrix in the decomposition process, and the clustering result is fluctuating.

In order to comprehensively avoid the defects of the above algorithms, the $L_{2,1}$ regularization is introduced into the NF-CCE framework, limiting the sparsity during the factorization.

III. METHOD

A. Problem Definition

Given an undirected and unweighted single-layer network G=(V, E), the vertex set is represented by V=$\{v_1, v_2, \ldots, v_n\}$ (n is the number of nodes in G, i.e., n=$|V|$) and the edge set can be defined as $E = \{(v_i, v_j)\}$ (m is the number of edges, i.e., m=$|V|$). The adjacency matrix A=$\{a_{ij}\}$ is made up of 0 and 1 whose element a_{ij}=1 if there is an edge between vertex v_i and v_j, 0 otherwise. For a given L-layered multiplex network M, it can be formulated with a set of adjacency matrices, e.g., $\{A_{(1)}, ..., A_{(L)}\}$ where $A_{(l)}$ is the adjacency matrix of the l^{th} layer. Community detection in multiplex networks aims to infer the consensus community assignment.

B. Algorithm Procedure

Sparse Collective Symmetric Non-negative Matrix Factorization (SCSNMF) with $L_{2,1}$ regularization (SCSNMF) uses the Network Fusion for Composite Community Extraction (NF-CCE) framework [8] to implement the multiplex network clustering which consists of two parts as plotted in Fig. 1.

Part 1. For each network layer, i, it is represented in a low-dimensional way by using Symmetric Non-negative Matrix Factorization (SNMF) [17]. That is, an $n \times n$ adjacency matrix $A_{(i)}$ can be converted to an $n \times k$ low-dimensional matrix $H_{(i)}$, under column orthonormality constraints, i.e., $H_{(i)}^T H_{(i)}$=I. And the $L_{2,1}$ regularization is introduced to ensure the structural sparsity of the reduced dimension matrix. The

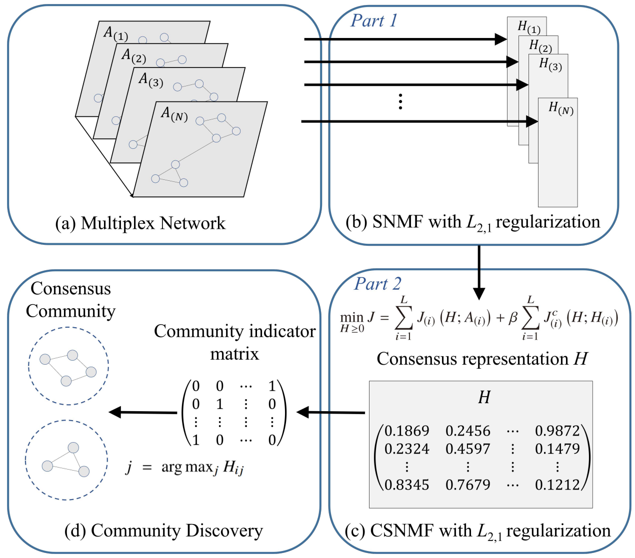

Fig. 1. Overview of the SCSNMF algorithm that comprises two parts.

low-dimensional matrix for each layer can be obtained by minimizing the reconstruction error J with $H_{(i)}^T H_{(i)} = I$

$$J = \min_{H_{(i)} \geq 0} \left\| A_{(i)} - H_{(i)} H_{(i)}^T \right\|_F^2 + \alpha \left\| H_{(i)} \right\|_{2,1} \tag{1}$$

where α is the regularization parameter, $\| \cdot \|_F$ denotes the Frobenius norm and $H_{(i)} \in \mathbb{R}_+^{n \times k}$ is the low-dimensional representation of the i^{th} network layer.

Part 2. To find a consensus low-dimensional representation H, this paper proposes a sparse collective SNMF method to decompose all adjacency matrices into a standard low-dimensional matrix. Moreover, it minimizes the projection distance [16] between the consensus representation H and every layer's representation $H_{(i)}$ to make them close enough on the Grassmann manifold [11]. The objective function contained in these two properties can be formulated as Eq. (2)

$$\min_{H \geq 0} J = \sum_{i=1}^{L} J_{(i)} \left(H; A_{(i)} \right) + \beta \sum_{i=1}^{L} J_{(i)}^c \left(H; H_{(i)} \right) \tag{2}$$

where β denotes the regularization parameter, $J_{(i)}$ represents the objective function of sparse collective SNMF, and $J_{(i)}^c$ means the projection distance on the Grassmann manifold.

Following are the details about the two terms:

1) Sparse collective SNMF: Different from SNMF, the collective SNMF minimizes the total error by using a common low-dimensional representation. Moreover, the collective SNMF also has sparsity constraint, which forms sparse collective SNMF finally. So the objective $J_{(i)}$ in Eq. (2) has the following expression with the regularization parameter γ

$$J_{(i)} \left(H; A_{(i)} \right) = \left\| A_{(i)} - HH^T \right\|_F^2 + \gamma \|H\|_{2,1} \tag{3}$$

$$s.t. H \geq 0, H^T H = I$$

2) Minimization of the projection distance on the Grassmann manifold: Dong et al. [11] propose a generic merging framework to make the consensus representative subspace close enough to all individual subspaces. So it is useful to minimize the projection distance [16] on the Grassmann manifold between these points mapped by the consensus matrix H and all the individual low-dimensional matrix $\left\{H_{(i)}\right\}_{i=1}^{L}$. Then the projection distance $J_{(i)}^{c}$ in Eq. (2) can be calculated

$$\begin{aligned} J_{(i)}^{c}\left(H, H_{(i)}\right) &= k - \operatorname{tr}\left(HH^{T} H_{(i)} H_{(i)}^{T}\right) \\ &= \left\|HH^{T} - H_{(i)} H_{(i)}^{T}\right\|_{F}^{2} \end{aligned} \tag{4}$$

where k denotes the column number of matrices.

In summary, the form of Eq.(2) can be rewritten as:

$$\begin{aligned} \min_{H \geq 0} J = &\sum_{i=1}^{L} \left\|A_{(i)} - HH^{T}\right\|_{F}^{2} + \gamma \|H\|_{2,1} \\ &+ \beta \sum_{i=1}^{L} \left\|HH^{T} - H_{(i)} H_{(i)}^{T}\right\|_{F}^{2} \end{aligned} \tag{5}$$

C. Optimization

We optimize the low-dimensional matrix using an alternating minimization method and following the Karush-Kuhn-Tucker (KKT) condition [18]. For *Part 1*, taking the derivative of Eq. (1) with respect to $H_{(i)}$ yields

$$\nabla_{H_{(i)}} J = -4A_{(i)} H_{(i)} + 4H_{(i)} H_{(i)}^{T} H_{(i)} + \alpha \Lambda_{(i)} H_{(i)} \tag{6}$$

where $\Lambda_{(i)}$ represents an $n \times n$ diagonal matrix as follows

$$\begin{pmatrix} \frac{1}{\|H_1\|_F} & & & \\ & \frac{1}{\|H_2\|_F} & & \\ & & \ddots & \\ & & & \frac{1}{\|H_n\|_F} \end{pmatrix} \tag{7}$$

where H_i indicates the i^{th} row of matrix $H_{(i)}$ and n is the number of rows in $H_{(i)}$.

The natural gradient [19] is proposed to solve the orthornormality constraint, which can be generated by

$$\widetilde{\nabla}_{H_{(i)}} J = \nabla_{H_{(i)}} J - H_{(i)} H_{(i)}^{T} \nabla_{H_{(i)}} J \tag{8}$$

where $\nabla_{H_{(i)}} J$ is the ordinary gradient calculated in Eq. (6).

In Eq. (8), $\widetilde{\nabla}_{H_{(i)}} J$ can be divided into two non-negative terms, i.e., $\widetilde{\nabla}_{H_{(i)}} J=\left[\tilde{\nabla}_{H_{(i)}} J\right]^{+}-\left[\tilde{\nabla}_{H_{(i)}} J\right]^{-}$. To resolve this, $\left[\tilde{\nabla}_{H_{(i)}} J\right]^{+}$ and $\left[\tilde{\nabla}_{H_{(i)}} J\right]^{-}$ can be described as

$$\left[\tilde{\nabla}_{H_{(i)}} J\right]^{-} = 4AH_{(i)} + \alpha H_{(i)} H_{(i)}^{T} \Lambda H_{(i)} \tag{9}$$

$$\left[\tilde{\nabla}_{H_{(i)}} J\right]^{+} = \alpha \Lambda H_{(i)} + 4H_{(i)} H_{(i)}^{T} A_{(i)} H_{(i)} \tag{10}$$

The update rule for $H_{(i)}$ shown in Eq. (11) can be acquired by the KKT condition

$$H_{(i)} \leftarrow H_{(i)} \frac{\left[\widetilde{\nabla}_{H_{(i)}} J\right]^{-}}{\left[\widetilde{\nabla}_{H_{(i)}} J\right]^{+}} \tag{11}$$

Similarly, the update rule for consensus matrix H can be expressed the same way shown above

$$H \leftarrow H \frac{\left[\widetilde{\nabla}_{H} J\right]^{-}}{\left[\widetilde{\nabla}_{H} J\right]^{+}} \tag{12}$$

$$\begin{aligned} \left[\widetilde{\nabla}_{H} J\right]^{-} = &\sum_{i=1}^{L} 4A_{(i)} H + \sum_{i=1}^{L} 2\beta H_{(i)} H_{(i)}^{T} H \\ &+ \sum_{i=1}^{L} \gamma HH^{T} \Sigma H \end{aligned} \tag{13}$$

$$\begin{aligned} \left[\widetilde{\nabla}_{H} J\right]^{+} = &\sum_{i=1}^{L} 4HH^{T} A_{(i)} H + \sum_{i=1}^{L} 2\beta HH^{T} H_{(i)} H_{(i)}^{T} H \\ &+ \sum_{i=1}^{L} \gamma \Sigma H \end{aligned} \tag{14}$$

where Σ can be calculated from H as in Eq. (7).

D. Clustering results

The community indicator matrix is usually applied to extract the node classification. It is denoted by $X \in \{0,1\}^{n \times k}$, where k is the number of communities. $X_{ij}=1$ if node i belongs to community j. This paper considers the case of non-overlapping communities, where a node can belong to only one community. Our algorithm approximates the consensus matrix H as the community indicator matrix. For each row i in the consensus matrix, we calculate $j = \arg\max_j H_{ij}$, which means that node i belongs to the j^{th} community.

IV. Experiments

Our algorithm is evaluated in a number of experiments. We demonstrate the higher clustering accuracy of SCSNMF compared with other seven algorithms in real-world multiplex networks. Meanwhile, the improvement of SCSNMF relative to CSNMF is shown in detail. Moreover, the parameter analysis is presented for examining the influence of parameters.

A. Evaluation Indexes and Datasets

Here we let $X=\{x_1, \ldots, x_k\}$ and $Y=\{y_1, \ldots, y_k\}$ indicate the ground truth clusters and the computed clusters respectively when the ground-truth community structure is known. Then we adopt Normalized Mutual Information (NMI) [20], Adjusted Rand Index (ARI) [20], and Purity [21] to evaluate the performance of algorithms. These evaluation measures are expressed as follows with a total number of nodes n:

- NMI can be calculated by Eq. (15) where $H(X)$ and $H(Y)$ represent the entropy of the ground truth and the computed clusters respectively, while $I(Y;C)$ is the mutual information between X and Y [20].

$$NMI(Y, C) = \frac{2 \times I(Y;C)}{H(Y) + H(C)} \tag{15}$$

- ARI is adopted from Rand Index (RI) [20]. Here let a represent the number of nodes which belongs to both X and Y while b indicates the contrary, RI can be defined as

$$\mathrm{RI} = \frac{2(a+b)}{n(n-1)} \tag{16}$$

Then ARI can be defined as Eq. (17) where $E[\cdot]$ represents the expected RI value

$$\mathrm{ARI} = \frac{\mathrm{RI} - E[\mathrm{RI}]}{\max(\mathrm{RI}) - E[\mathrm{RI}]} \tag{17}$$

- Purity refers to the proportion of nodes that are correctly classified with the expression [21]

$$\text{Purity}(X, Y) = \frac{1}{n} \sum_{k} \max_{j} \left| y_k \cap x_j \right| \tag{18}$$

NMI and Purity range from 0 to 1, while ARI is in the range $[-1, 1]$. The larger the value of these indicators, the higher the accuracy of clustering. In order to estimate the performance of SCSNMF method, seven real-world networks with known ground-truth division are used as shown in Table I.

TABLE I
EXPERIMENTAL MULTIPLEX NETWORKS USED FOR OUR COMPARISON

Network	Nodes	Layers	Clusters	Ref.
SND(o)	71	3	3	[22]
SND(s)	71	3	2	[22]
WBN	279	5	10	[23]
WTN	183	14	10	[24]
MPD	87	3	6	[25]
CoRA	1662	2	3	[26]
CiteSeer	3312	2	3	[26]

B. Clustering Accuracy

Table II shows the comparisons of our proposed algorithm with other methods in terms of NMI, ARI and Rurity on seven data-sets. The maximum value is highlighted by the bold. It is evident that the clustering accuracy of our algorithm surpasses other algorithms. In addition to our algorithm, CSNMF performs better than other algorithms in most cases. This is because the NF-CCE framework adopted by SCSNMF and CSNMF can fuse the information of each layer well and avoid the excessive influence of noise. In CoRA network, the NMI value of CSNMF is lower than GMC. Only in SND(o), the clustering performance of CSNMF is consistent with that of SCSNMF, while in the other six groups of comparison, the performance of CSNMF is inferior to that of SCSNMF. The cause of the gap between SCSNMF and CSNMF is that $L_{2,1}$ regularization is added in SCSNMF to optimize the NF-CCE framework. The $L_{2,1}$ regularization can limit the sparsity of matrices during the factorization. Overall, it is obvious that our algorithm has the higher clustering accuracy.

C. Stability Analysis

To fully evaluate the improvement of two algorithms with and without sparse contraints, i.e., SCSNMF and CSNMF, we run each algorithm 10 times and report the maximum, mean, and standard deviation of NMI. Here, the maximum and mean values are used to demonstrate the performance of improving the overall classification accuracy and capability of deriving the optimal solution, while the standard deviation is adopted to reflect the stability of algorithms. As shown in Table III, SCSNMF is more stable and of better average performance compared with CSNMF. In order to compare the results more intuitively, Fig. 2 is provided. As shown in Table II, the optimal performance of two algorithms in SND(o) network is consistent. However, from Fig. 2, it is clear that our results of each experiment are consistent whereas the results of CSNMF are fluctuating in SND(o). And in WBN network and CiteSeer network, the experimental results of CSNMF fluctuate considerably and even abnormal points exist. The reason for this gap is that the matrix sparsity is not guaranteed in the process of CSNMF decomposition. It is clear that SCSNMF is able to improve the performance of CSNMF, providing stable results with smaller variance.

TABLE II
COMPARISON OF CLUSTERING ACCURACY OF EIGHT ALGORITHMS ON REAL-WORLD MULTIPLEX NETWORKS

		SNF	LMF	GF	SC-ML	GMC	FCMSC	CSNMF	SCSNMF
SND(o)	NMI	0.64494	0.48506	0.67554	0.20638	0.59666	0.66217	**0.68156**	**0.68156**
	ARI	0.50223	0.34178	0.47772	0.17277	0.42821	0.46814	**0.49347**	**0.49347**
	Purity	0.94366	0.87324	0.94366	0.78873	0.92958	0.94366	**0.94367**	**0.94367**
SND(s)	NMI	0.06258	0.02506	0.04486	0.03012	0.05141	0.05277	0.09470	**0.10769**
	ARI	0.04661	0.02015	0.04465	0.02134	0.03414	0.05838	0.11149	**0.13224**
	Purity	0.61972	0.59155	0.61972	0.59121	0.60563	0.63380	0.67606	**0.69014**
WBN	NMI	0.36966	0.26373	0.36379	0.32469	0.24160	0.23651	0.40553	**0.42122**
	ARI	0.22088	0.06789	0.21124	0.17512	0.01224	0.16163	0.23387	**0.23786**
	Purity	0.53405	0.40860	0.49821	0.50896	0.38710	0.51366	0.52330	**0.54122**
MPD	NMI	0.35634	0.27418	0.31655	0.50781	0.45112	0.43871	0.51734	**0.52862**
	ARI	0.16881	0.08366	0.23548	0.42046	0.24767	0.27425	0.42362	**0.42531**
	Purity	0.60920	0.57471	0.63218	**0.72414**	0.66667	0.66667	**0.72414**	**0.72414**
CoRA	NMI	0.46418	0.21371	0.44798	0.51031	0.51935	0.46319	0.50591	**0.52146**
	ARI	0.35048	0.06727	0.27208	0.36986	0.37003	0.38611	0.37108	**0.37814**
	Purity	0.61449	0.46233	0.61301	0.62379	0.62761	0.63356	0.64041	**0.64384**
CiteSeer	NMI	0.17133	0.04690	0.08880	0.00479	0.02123	0.19447	0.20941	**0.21009**
	ARI	0.12822	0.00038	0.04392	0.00189	0.00044	0.19885	0.20422	**0.20755**
	Purity	0.35537	0.24758	0.34300	0.21649	0.22101	0.49124	0.50362	**0.52536**

TABLE III
THE MAXIMUM VALUE, MEAN VALUE AND STANDARD DEVIATION VALUE OF TEN EXPERIMENTS RESULTS

		maximum value	mean value	standard deviation
SND(o)	CSNMF	0.68156	0.65840	0.03004
	SCSNMF	0.68156	0.68156	0.00000
SND(s)	CSNMF	0.09470	0.05794	0.02232
	SCSNMF	0.10769	0.07000	0.01779
WTN	CSNMF	0.40553	0.28140	0.02045
	SCSNMF	0.42122	0.28940	0.01870
WBN	CSNMF	0.31936	0.37040	0.01877
	SCSNMF	0.32494	0.39710	0.01449
MPD	CSNMF	0.51734	0.43556	0.04125
	SCSNMF	0.52862	0.44243	0.03254
CoRA	CSNMF	0.50591	0.45248	0.03012
	SCSNMF	0.52146	0.46153	0.02919
CiteSeer	CSNMF	0.20941	0.18643	0.01380
	SCSNMF	0.21009	0.19604	0.00601

D. Parameter Sensitivity

Empirically, regularization parameters are always between 0 and 1, so SCSNMF sets its parameters (α, γ, β) in the range $[0, 1]$. The NMI results with respect to three parameters (α, γ, β) are shown in Fig. 3. Because the experiments' results are similar on these networks, this part only shows the WBN network results. As can be seen, SCSNMF is robust for the three parameters. Only when the parameter is less than 0.1, does a slight fluctuation exist in algorithm performance.

V. CONCLUSION

How to fuse the information of each layer of multiplex networks and reduce noise influence is an essential problem in multiplex network clustering. Although the non-negative

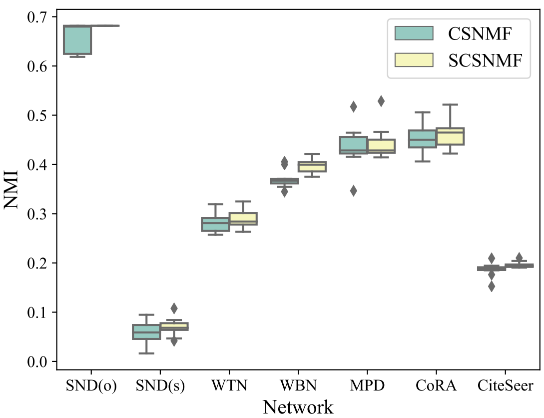

Fig. 2. The NMI boxplot of CSNMF and SCSNMF in ten experiments on real-world networks.

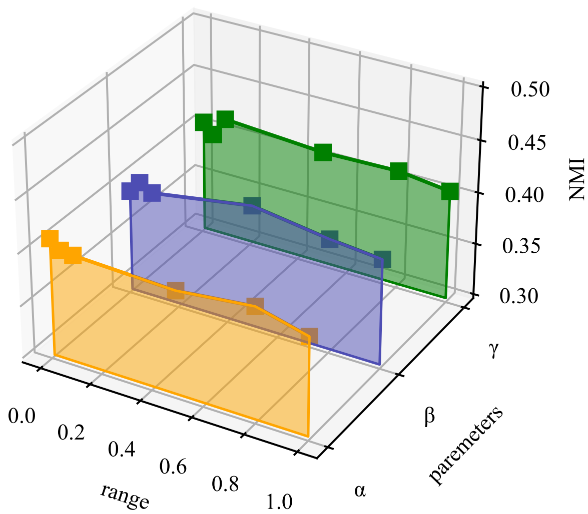

Fig. 3. The curve of NMI variation on WBN network by adjusting α, γ and β from 0.001 to 1.

matrix factorization is an excellent method for community detection, this method still needs to ensure the discretization and sparsity of matrix to play a better role. In this paper, SCSNMF uses the symmetric NMF to obtain low-dimensional representations from every layer firstly. Then, the consensus community structure is obtained through a common low-dimensional representation by collectively using symmetric NMF. Remarkably, the $L_{2,1}$ regularization is added to ensure the sparsity of low-dimensional representation matrices in every decomposition process. Our experiments indicate that our algorithm outperforms seven other clustering algorithms in clustering accuracy, stability and optimal solution.

References

[1] D. Zhao, L. Xu, S. Qin, G. Liu, and Z. Wang, "The feedback vertex set problem of multiplex networks," *IEEE Transactions on Circuits and Systems II: Express Briefs*, pp. 1–1, 2020.

[2] G. A. Pagani and M. Aiello, "The power grid as a complex network: a survey," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 11, pp. 2688–2700, 2013.

[3] C. Zhao, J. Zhao, C. Wu, X. Wang, F. Xue, and S. Lu, "Power grid partitioning based on functional community structure," *IEEE Access*, vol. 7, pp. 152 624–152 634, 2019.

[4] J. Leskovec, K. J. Lang, and M. Mahoney, "Empirical comparison of algorithms for network community detection," in *Proceedings of the 19th International Conference on World Wide Web*, 2010, pp. 631–640.

[5] A. Mahmood and M. Small, "Subspace based network community detection using sparse linear coding," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 801–812, 2015.

[6] L. Yang, X. Cao, D. Jin, X. Wang, and D. Meng, "A unified semi-supervised community detection framework using latent space graph regularization," *IEEE Transactions on Cybernetics*, vol. 45, no. 11, pp. 2585–2598, 2014.

[7] W. Luo, D. Zhang, H. Jiang, L. Ni, and Y. Hu, "Local community detection with the dynamic membership function," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 5, pp. 3136–3150, 2018.

[8] V. Gligorijević, Y. Panagakis, and S. Zafeiriou, "Non-negative matrix factorizations for multiplex network analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 4, pp. 928–940, 2018.

[9] B. Wang, A. M. Mezlini, F. Demir, M. Fiume, Z. Tu, M. Brudno, B. Haibe-Kains, and A. Goldenberg, "Similarity network fusion for aggregating data types on a genomic scale," *Nature Methods*, vol. 11, no. 3, p. 333, 2014.

[10] H. Wang, Y. Yang, and B. Liu, "GMC: Graph-based multi-view clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1116–1129, 2019.

[11] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov, "Clustering on multi-layer graphs via subspace analysis on grassmann manifolds," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 905–918, 2013.

[12] W. Tang, Z. Lu, and I. S. Dhillon, "Clustering with multiple graphs," in *2009 Ninth IEEE International Conference on Data Mining*. IEEE, 2009, pp. 1016–1021.

[13] E. E. Papalexakis, L. Akoglu, and D. Ience, "Do more views of a graph help? community detection and clustering in multi-graphs," in *Proceedings of the 16th International Conference on Information Fusion*. IEEE, 2013, pp. 899–905.

[14] Q. Zheng, J. Zhu, Z. Li, S. Pang, J. Wang, and Y. Li, "Feature concatenation multi-view subspace clustering," *Neurocomputing*, vol. 379, pp. 89–102, 2020.

[15] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2008.

[16] J. Hamm and D. D. Lee, "Grassmann discriminant analysis: a unifying view on subspace-based learning," in *Proceedings of the 25th International Conference on Machine Learning*, 2008, pp. 376–383.

[17] D. Kuang, C. Ding, and H. Park, "Symmetric nonnegative matrix factorization for graph clustering," in *Proceedings of the 2012 SIAM International Conference on Data Mining*. SIAM, 2012, pp. 106–117.

[18] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

[19] Y. Panagakis, C. Kotropoulos, and G. R. Arce, "Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 576–588, 2009.

[20] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge university press, 2008.

[21] Y. Zhao and G. Karypis, "Empirical and theoretical comparisons of selected criterion functions for document clustering," *Machine Learning*, vol. 55, no. 3, pp. 311–331, 2004.

[22] T. A. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock, "New specifications for exponential random graph models," *Sociological Methodology*, vol. 36, no. 1, pp. 99–153, 2006.

[23] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, "The structure of the nervous system of the nematode caenorhabditis elegans," *Philos Trans R Soc Lond B Biol Sci*, vol. 314, no. 1165, pp. 1–340, 1986.

[24] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, "Structural reducibility of multilayer networks," *Nature Communications*, vol. 6, no. 1, pp. 1–9, 2015.

[25] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov, "Clustering with multi-layer graphs: A spectral perspective," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5820–5831, 2012.

[26] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad, "Collective classification in network data," *AI Magazine*, vol. 29, no. 3, pp. 93–93, 2008.