Article

http://justc.ustc.edu.cn

Received: December 29, 2024; Accepted: April 27, 2025

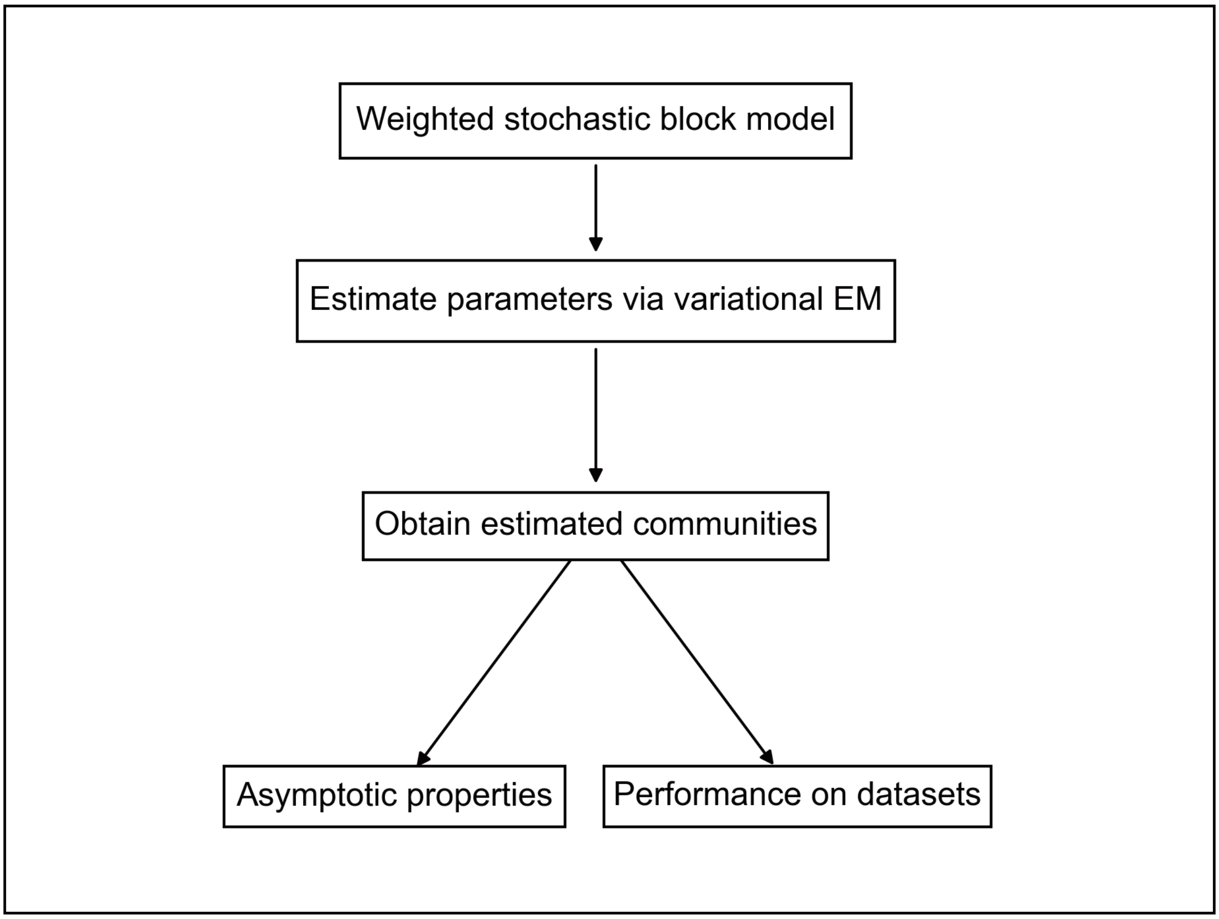
A variational approach to community detection in weighted networks

Lei Liu[1], Jingnan Zhang[1] ✉, and Chengye Li[2]

[1]*Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China;*
[2]*School of Data Science, City University of Hong Kong, Hong Kong 999077, China*

✉Correspondence: Jingnan Zhang, E-mail: jnzhang@ustc.edu.cn

Graphical abstract



The WSBM framework for model inference and verification.

Public summary

■ Propose a weighted stochastic block model for weighted networks.

■ Develop a variational EM algorithm to estimate parameters and detect communities.

■ Establish asymptotic consistency for both parameter estimation and community detection.

Citation: Liu L, Zhang J N, Li C Y. A variational approach to community detection in weighted networks. *JUSTC*, **2026**, 56(2): 0203. CSTR: 32290.14.JUSTC-2024-0164. DOI: 10.52396/JUSTC-2024-0164

JUSTC
Article
http://justc.ustc.edu.cn
Received: December 29, 2024; Accepted: April 27, 2025

A variational approach to community detection in weighted networks

Lei Liu[1], Jingnan Zhang[1] ✉, and Chengye Li[2]

[1]*Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China;*
[2]*School of Data Science, City University of Hong Kong, Hong Kong 999077, China*

✉Correspondence: Jingnan Zhang, E-mail: jnzhang@ustc.edu.cn

Cite This: *JUSTC*, **2026**, 56(2): 0203 (12pp)   Read Online

Abstract: Community detection is an essential aspect in studying network structures. Most community detection methods focus on binary networks, which ignore valuable information in the edge weights. In this study, we propose a novel weighted stochastic block model (WSBM). In the proposed model, we employ the variational EM algorithm to estimate the model parameters, and then extract the community structure of the weighted network. The detection accuracy and efficiency of the proposed algorithm are demonstrated through both synthetic and real examples. The theoretical results of the proposed method in terms of the asymptotic consistency of estimating the model parameters and community structure are established. The proposed method provides an effective way to detect communities in weighted networks, with promising potential for practical applications.

Keywords: weighted network; variational EM; yeast experimental networks; asymptotic consistency

CLC number: O212.4   **Document code:** A

2020 Mathematics Subject Classification: 05C80; 62F12; 62F30

1 Introduction

Network data have become one of the most common data structures in recent years and appear in various domains, including biology, ecology, sociology, and information science[1]. Community structure in networks refers to groups of nodes with significantly higher intra-group connectivity and similar properties compared to lower connectivity between groups. Community detection identifies patterns and associations between nodes in a network[2], helping to understand its structure and the information it contains. This process simplifies the functional analysis of various structures and helps us better leverage the information within the network for more effective decision-making and analysis[3–5]. For example, in social network analysis, community detection reveals the nature of social interactions and helps to better understand business models and disease transmission[6].

In the literature, various efforts have been devoted to developing effective community detection methods and many approaches have been proposed. These approaches can be loosely classified into three categories[7]. The first category is algorithm-based methods, which use similarity measures between nodes for network partitioning, such as the hierarchical clustering method[8,9]. The second category is criterion-based methods, which optimize a given criterion to find the most accurate community structure[10]. Examples include the modularity criterion[11] and the cut criterion[12]. The third category is model-based methods, which assume that the network follows a model with a community structure, such as the stochastic block model[13] and latent space model[14].

However, most community detection methods focus on binary networks, which only consider the presence of edges between nodes. Many real-world networks are weighted networks, such as the transmission probabilities of different modalities in epidemiological networks[15] and the interaction frequencies between different nodes in social networks[16]. When community detection methods are applied to weighted networks, some approaches are inherently unable to utilize weight information. For example, the method in Ref. [1] primarily relies on shortest path calculations between nodes, which effectively neglects the role of edge weights in defining community structure. Some attempts have been made to extend classical unweighted network algorithms to weighted networks, such as the weighted modularity approach[17], which incorporates weight matrices into modularity calculations. However, this method lacks a theoretical framework. Precision graph partitioning[18], which takes edge weights into account when optimizing intracommunity and intercommunity connection strengths, suffers from high computational complexity. Additionally, weighted spectral clustering[19] has been explored, but it struggles to adapt effectively to sparse networks. Multiple methods have been proposed for community detection in weighted networks, such as nonnegative matrix factorization[20], which decomposes the nonnegative adjacency matrix into two matrices to derive community assignments.

In this study, we introduce a novel weighted stochastic block model for weighted networks. Additionally, we propose a parameter estimation method based on the variational

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203

EM algorithm. The community assignments of nodes are derived from the parameter estimation of the model. Notably, we establish the asymptotic consistency of the proposed method in terms of both parameter estimation and community detection. To assess the proposed method, we conduct experiments on both synthetic and real-world weighted network datasets and compare it with four existing community detection methods. The results indicate that the proposed method demonstrates superior performance.

The rest of the paper is organized as follows. Section 2 introduces the proposed weighted stochastic block model. Section 3 develops an efficient parameter estimation method based on the variational EM algorithm. Section 4 establishes the theoretical results. Section 5 conducts experiments on simulated and real examples. Section 6 concludes the paper with a brief summary, and all technical proofs are given in the Appendix.

2 Weighted stochastic block model

The stochastic block model (SBM) assumes that nodes are partitioned into communities, and the probability of an edge existing between two nodes depends solely on their community memberships[13]. The weighted stochastic block model (WSBM) extends the SBM by introducing weighted edges. In this model, the distribution of edge weights between two nodes follows a probabilistic model that depends solely on the community memberships of the nodes.

Consider a network $\mathcal{G}$ with n nodes, and let $\boldsymbol{A}=(a_{ij})_{n\times n}$ denote its adjacency matrix, where a_{ij} denotes the edge weight between nodes i and j. Under the WSBM model, we assume that the distribution of a_{ij} is

$$a_{ij}\sim\mathcal{N}(\mu_{g(i),g(j)},\sigma^2_{g(i),g(j)}), \tag{1}$$

where $g(i)$ and $g(j)$ represent the communities of nodes i and j, respectively, $\mu_{g(i),g(j)}$ and $\sigma^2{}_{g(i),g(j)}$ represent the mean and variance of a normal distribution. The distribution of a_{ij} clearly depends only on the community memberships of i and j. The community membership matrix is denoted as $\boldsymbol{C}=(c_{ik})_{n\times K}$, where $c_{ik}=1$ if $g(i)=k$ and 0 otherwise. Each row of $\boldsymbol{C}$ is independently modeled with a multinomial distribution $\boldsymbol{C}_i\sim\mathcal{M}(1;\boldsymbol{\alpha})$, for $i\in[n]$, where $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_K)$ with $\alpha_k>0$ and $\sum_{k=1}^{K}\alpha_k=1$.

Given $g(i)=k$ and $g(j)=l$, let $\boldsymbol{\Delta}_{kl}$ denote (μ_{kl},σ^2_{kl}). Then, the density function of a_{ij} under the WSBM model is $f_{\Delta_{kl}}(a_{ij})=\dfrac{1}{\sqrt{2\pi\sigma^2_{kl}}}\exp\left(\dfrac{(a_{ij}-\mu_{kl})^2}{-2\sigma^2_{kl}}\right)$. Thus, the log-likelihood function of $(\boldsymbol{A},\boldsymbol{C})$ is

$$\log P(\boldsymbol{A},\boldsymbol{C})=\log P(\boldsymbol{C})+\log P(\boldsymbol{A}|\boldsymbol{C})=\sum_{i=1}^{n}\sum_{k=1}^{K}c_{ik}\log(\alpha_k)+\sum_{i<j}^{n}\sum_{k,l=1}^{K}c_{ik}c_{jl}\log(f_{\Delta_{kl}}(a_{ij})). \tag{2}$$

Finally, the log-likelihood function of $\boldsymbol{A}$ is $\log P(\boldsymbol{A})=\log\left(\sum_{\boldsymbol{C}}P(\boldsymbol{A},\boldsymbol{C})\right)$.

However, the community membership matrix $\boldsymbol{C}$ has K^n possible values, making the optimization of the log-likelihood function of a computationally challenging. Furthermore, the traditional EM algorithm[21] also cannot be applied to solve the optimization problem because $\log P(\boldsymbol{C}|\boldsymbol{A})$ is intractable. Therefore, we employ the variational EM algorithm[22] to address this issue, which considers a lower bound on $\log P(\boldsymbol{A})$ and simplifies the computation by restricting the conditional distribution of $\boldsymbol{C}$ given $\boldsymbol{A}$ to be a factorized form. The detailed steps of the proposed approach are presented in the next section.

3 Variational EM algorithm

To address the limitations encountered with the traditional EM algorithm, we consider a lower bound for $\log P(\boldsymbol{A})$, denote $\boldsymbol{\Theta}=(\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\Sigma})$, and

$$\mathcal{L}(R_A(\boldsymbol{C}),\boldsymbol{\Theta})=\mathcal{H}(R_A(\boldsymbol{C}))+\sum_{\boldsymbol{C}}R_A(\boldsymbol{C})\log(P(\boldsymbol{A},\boldsymbol{C};\boldsymbol{\Theta})), \tag{3}$$

where $R_A(\boldsymbol{C})$ stands for some distributions on $\boldsymbol{C}$ and $\mathcal{H}(R_A(\boldsymbol{C}))$ is its entropy. The maximum value of $\mathcal{L}(R_A(\boldsymbol{C}),\boldsymbol{\Theta})$ is $\log P(\boldsymbol{A})$, which is attained when $R_A(\boldsymbol{C})=P(\boldsymbol{C}|\boldsymbol{A})$. Therefore, Eq. (3) provides a lower bound for $\log P(\boldsymbol{A})$.

Recall that $P(\boldsymbol{C}|\boldsymbol{A})$ is not tractable, thus we restrict the distribution of $R_A(\boldsymbol{C})$ to the factorization distribution, denoted as $R_A(\boldsymbol{C},\boldsymbol{\beta})$,

$$R_A(\boldsymbol{C},\boldsymbol{\beta})=\prod_{i=1}^{n}h(c_i,\boldsymbol{\beta}_i), \tag{4}$$

where h denotes the multinomial distribution function and $\boldsymbol{\beta}_i=(\beta_{i1},\cdots,\beta_{iK})\in[0,1]^K$ satisfying $\boldsymbol{\beta}\mathbf{1}_K=\mathbf{1}_n$. Write $\mathcal{L}(R_A(\boldsymbol{C},\boldsymbol{\beta}),\boldsymbol{\Theta})$ simply as $\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Theta})$.

Note that

$$\mathcal{H}(R_A(\boldsymbol{C}))=-\sum_{i=1}^{n}\sum_{k=1}^{K}\beta_{ik}\log(\beta_{ik}).$$

Moreover, it follows from Eqs. (2) and (4) that

$$\sum_{\boldsymbol{C}}R_A(\boldsymbol{C})\log(P(\boldsymbol{A},\boldsymbol{C};\boldsymbol{\Theta}))=\sum_{i=1}^{n}\sum_{k=1}^{K}\beta_{ik}\log(\alpha_k)+\sum_{i<j}^{n}\sum_{k,l=1}^{K}\beta_{ik}\beta_{jl}\log\left(\frac{1}{\sqrt{2\pi\sigma^2_{kl}}}\exp\left(\frac{(a_{ij}-\mu_{kl})^2}{-2\sigma^2_{kl}}\right)\right).$$

By combining these two components, the objective function $\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Theta})$ can be reformulated as

$$\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Theta})=-\sum_{i=1}^{n}\sum_{k=1}^{K}\beta_{ik}\log(\beta_{ik})+\sum_{i=1}^{n}\sum_{k=1}^{K}\beta_{ik}\log(\alpha_k)+\sum_{i<j}^{n}\sum_{k,l=1}^{K}\beta_{ik}\beta_{jl}\log\left(\frac{1}{\sqrt{2\pi\sigma^2_{kl}}}\exp\left(\frac{(a_{ij}-\mu_{kl})^2}{-2\sigma^2_{kl}}\right)\right). \tag{5}$$

Our goal is to obtain the estimates $(\hat{\boldsymbol{\beta}}^v,\hat{\boldsymbol{\Theta}}^v)$, where $(\hat{\boldsymbol{\beta}}^v,\hat{\boldsymbol{\Theta}}^v)=\underset{\boldsymbol{\beta},\boldsymbol{\Theta}:\boldsymbol{\beta}\mathbf{1}_K=\mathbf{1}_n}{\operatorname{argmax}}\ \mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Theta})$.

We iteratively update $\boldsymbol{\Theta}$ and $\boldsymbol{\beta}$ until convergence to solve the optimization problem. Specifically, given $\boldsymbol{\beta}^{(t)}$ at the tth step, we address the following two sub-optimization problems separately,

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203

$$\boldsymbol{\Theta}^{(t+1)} = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\ \mathcal{L}(\boldsymbol{\beta}^{(t)},\boldsymbol{\Theta}), \tag{6}$$

$$\boldsymbol{\beta}^{(t+1)} = \underset{\boldsymbol{\beta}:\boldsymbol{\beta}\mathbf{1}_K=\mathbf{1}_n}{\operatorname{argmax}}\ \mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Theta}^{(t+1)}). \tag{7}$$

In Eq. (6), $\boldsymbol{\Theta} = (\boldsymbol{\alpha},\boldsymbol{\Delta})$. Therefore, we update $\boldsymbol{\alpha}$ and $\boldsymbol{\Delta}$ separately. First, we update $\boldsymbol{\alpha}$ by

$$\boldsymbol{\alpha}^{(t+1)} = \underset{\boldsymbol{\alpha}}{\operatorname{argmax}}\ \sum_{i=1}^{n}\sum_{k=1}^{K}\beta_{ik}^{(t)}\log(\alpha_k).$$

Then we have that for each $k \in [K]$,

$$\alpha_k^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\beta_{ik}^{(t)}.$$

Then, we update $\boldsymbol{\Delta} = (\boldsymbol{\mu},\boldsymbol{\Sigma})$ by

$$(\mu_{kl}^{(t+1)},\sigma_{kl}^{2\ (t+1)}) = \underset{\mu_{kl},\sigma_{kl}^2}{\operatorname{argmax}}\ \sum_{i<j}^{n}\sum_{k,l=1}^{K}\beta_{ik}^{(t)}\beta_{jl}^{(t)}\log\left(\frac{1}{\sqrt{2\pi\sigma_{kl}^2}}\exp\left(\frac{(a_{ij}-\mu_{kl})^2}{-2\sigma_{kl}^2}\right)\right),$$

and we have that for each $k,l \in [K]$,

$$\mu_{kl}^{(t+1)} = \frac{\sum_{i\neq j}^{n}\beta_{ik}^{(t)}\beta_{jl}^{(t)}a_{ij}}{\sum_{i\neq j}^{n}\beta_{ik}^{(t)}\beta_{jl}^{(t)}},$$

$$\sigma_{kl}^{2\ (t+1)} = \frac{\sum_{i\neq j}^{n}\beta_{ik}^{(t)}\beta_{jl}^{(t)}(a_{ij}-\mu_{kl}^{(t+1)})^2}{\sum_{i\neq j}^{n}\beta_{ik}^{(t)}\beta_{jl}^{(t)}}.$$

Next, we update $\boldsymbol{\beta}$ via Lagrange method,

$$\mathcal{L}_\lambda(\boldsymbol{\beta},\boldsymbol{\Theta}^{(t+1)}) = \mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Theta}^{(t+1)}) + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{\beta}\mathbf{1}_K - \mathbf{1}_n),$$

where $\boldsymbol{\lambda} = (\lambda_1,\cdots,\lambda_n)$ is the Lagrange multiplier. Let

$$\frac{\partial\mathcal{L}_\lambda(\boldsymbol{\beta},\boldsymbol{\Theta}^{(t+1)})}{\partial\boldsymbol{\beta}_i} = 0,\qquad \frac{\partial\mathcal{L}_\lambda(\boldsymbol{\beta},\boldsymbol{\Theta}^{(t+1)})}{\partial\lambda_i} = 0.$$

The partial derivative of β_{ik} with respect to $\mathcal{L}_\lambda(\boldsymbol{\beta},\boldsymbol{\Theta}^{(t+1)})$ is

$$\frac{\partial\mathcal{L}_\lambda(\boldsymbol{\beta},\boldsymbol{\Theta}^{(t+1)})}{\partial\beta_{ik}} = -\log\beta_{ik} - 1 + \log\alpha_k^{(t+1)} + \sum_{j\neq i}^{n}\sum_{l=1}^{K}\beta_{jl}\left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma_{kl}^{2\ (t+1)} - \frac{1}{2}\frac{(a_{ij}-\mu_{kl}^{(t+1)})^2}{\sigma_{kl}^{2\ (t+1)}}\right) + \lambda_i. \tag{8}$$

The partial derivative of λ_i with respect to $\mathcal{L}_\lambda(\boldsymbol{\beta},\boldsymbol{\Theta}^{(t+1)})$ is

$$\frac{\partial\mathcal{L}_\lambda(\boldsymbol{\beta},\boldsymbol{\Theta}^{(t+1)})}{\partial\lambda_i} = \sum_{k=1}^{K}\beta_{ik} - 1. \tag{9}$$

It then follows for each $i \in [n]$ and $k \in [K]$ that

$$\beta_{ik}^{(t+1)} \propto \alpha_k^{(t+1)}\prod_{j\neq i}^{n}\prod_{l=1}^{K}\left[\frac{1}{\sqrt{2\pi\sigma_{kl}^{2\ (t+1)}}}\exp\left(\frac{(a_{ij}-\mu_{kl}^{(t+1)})^2}{-2\sigma_{kl}^{2\ (t+1)}}\right)\right]^{\beta_{jl}^{(t+1)}}.$$

We then use the final estimated $\hat{\boldsymbol{\beta}}$ to obtain the community assignments. Specifically, for each $i \in [n]$, let $\hat{g}(i) = k$ if $k = \arg\max_{k\in[K]}\hat{\beta}_{ik}$. We denote the final parameters estimated by the variational EM algorithm as $(\hat{\boldsymbol{\beta}}^v,\hat{\boldsymbol{\Theta}}^v)$, where $\hat{\boldsymbol{\Theta}}^v = (\hat{\boldsymbol{\alpha}}^v,\hat{\boldsymbol{\Delta}}^v)$, and $\hat{\boldsymbol{\Delta}}^v = (\hat{\boldsymbol{\mu}}^v,\hat{\boldsymbol{\Sigma}}^v)$. The details of the proposed algorithm can be summarized in Algorithm 1.

The time complexity of the WSBM method is $O(T_m K^2 n^2)$, where T_m is the maximum number of iterations. Among the competing algorithms, the time complexities of spectral clustering, Louvain modularity method, and nonnegative matrix factorization are $O(n^3)$, $O(n\log n)$, and $O(T_m K n^2)$, respectively. Although the Louvain modularity method has lower time complexity, it is important to note that it is not specifically designed for community detection in weighted networks. Therefore, its community detection accuracies are very low as shown in the numerical examples. Moreover, the time complexity of the proposed method is comparable to the nonnegative matrix factorization method since K is usually very small in real-world applications.

Algorithm 1: WSBM parameter estimation.

Require: Observed adjacency matrix $\boldsymbol{A}$, max iterations T_m.

1: Initialize: $\boldsymbol{\beta}^{(0)},\boldsymbol{\Theta}^{(0)} = (\boldsymbol{\alpha}^{(0)},\boldsymbol{\mu}^{(0)},\boldsymbol{\Sigma}^{(0)})$, and set $t = 0$.

2: $\alpha_k^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\beta_{ik}^{(t)}$, $1 \le k \le K$;

3: $\mu_{kl}^{(t+1)} = \dfrac{\sum_{i\neq j}^{n}\beta_{ik}^{(t)}\beta_{jl}^{(t)}a_{ij}}{\sum_{i\neq j}^{n}\beta_{ik}^{(t)}\beta_{jl}^{(t)}}$, $1 \le k,l \le K$;

4: $\sigma_{kl}^{2\ (t+1)} = \dfrac{\sum_{i\neq j}^{n}\beta_{ik}^{(t)}\beta_{jl}^{(t)}(a_{ij}-\mu_{kl}^{(t+1)})^2}{\sum_{i\neq j}^{n}\beta_{ik}^{(t)}\beta_{jl}^{(t)}}$, $1 \le k,l \le K$;

5: $\beta_{ik}^{(t+1)} \propto \alpha_k^{(t+1)}\prod_{j\neq i}^{n}\prod_{l=1}^{K}\left[\dfrac{1}{\sqrt{2\pi\sigma_{kl}^2(t+1)}}\exp\left(\dfrac{\left(a_{ij}-\mu_{kl}^{(t+1)}\right)^2}{-2\sigma_{kl}^{2\ (t+1)}}\right)\right]^{\beta_{jl}^{(t+1)}}$, $1 \le i \le n, 1 \le k \le K$;

6: $\beta_{ik}^{(t+1)} = \beta_{ik}^{(t+1)}/\sum_{k=1}^{K}\beta_{ik}^{(t+1)}$, $1 \le i \le n, 1 \le k \le K$;

7: **Repeat** Steps 2–6 with $t \leftarrow t+1$ **until** $t > T_m$.

Ensure: $\hat{\boldsymbol{\alpha}}^v = \boldsymbol{\alpha}^{(t)}$, $\hat{\boldsymbol{\mu}}^v = \boldsymbol{\mu}^{(t)}$, $\hat{\boldsymbol{\Sigma}}^v = \boldsymbol{\Sigma}^{(t)}$, $\hat{\boldsymbol{\beta}}^v = \boldsymbol{\beta}^{(t)}$.

4 Theoretical results

In this section, we establish the asymptotic consistency of the proposed method in estimating the model parameters and recovering the community structure. To proceed, we first clarify some notation. Let the adjacency matrix $\boldsymbol{A}$ follow the weighted stochastic block model. Specifically, each a_{ij}

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203

satisfies Model (1), where the true parameters are denoted by $\boldsymbol{\Delta}^*=(\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*)$, with $\boldsymbol{\Delta}^*_{kl}=(\mu^*_{kl},\sigma^{*2}_{kl})$. The true membership matrix is $\boldsymbol{c}^*=(c^*_{ik})_{n\times K}$, where $c^*_{ik}\in\{0,1\}$ indicates whether the ith node belongs to the kth community. $N_k(\boldsymbol{c}^*)=\sum_{i=1}^n c^*_{ik}$, denotes the number of nodes in the kth community, for $k\in[K]$. The limit of the ratio $N_k(\boldsymbol{c}^*)/n$ exists and is denoted by $\alpha_k{}^*$. We denote the maximum likelihood estimate of $\log P(\boldsymbol{A})$ as $\hat{\boldsymbol{\Theta}}=(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\Delta}})$. In the following, we write $x_n\prec y_n$ if $x_n=o(y_n)$ and $x_n\preceq y_n$ if $x_n=O(y_n)$.

Define $\psi_{kl}(\boldsymbol{\Delta}_{k'l'})=E_{\boldsymbol{\Delta}^*}\left(\log f_{\boldsymbol{\Delta}_{k'l'}}(\boldsymbol{a}_{ij})\right)$ when $c^*_{ik}=c^*_{jl}=1$. It is then clear that $\psi_{kl}(\boldsymbol{\Delta}^*_{kl})=E_{\boldsymbol{\Delta}^*}\left(\log f_{\boldsymbol{\Delta}^*_{kl}}(a_{ij})\right)$. The estimation error of $\hat{\boldsymbol{\Delta}}$ and $\hat{\boldsymbol{\Delta}}^v$ is measured by

$$d(\boldsymbol{\Delta},\boldsymbol{\Delta}^*)=\min_{\sigma}\sum_{k,l}\left(\psi_{kl}(\boldsymbol{\Delta}^*_{kl})-\psi_{kl}(\boldsymbol{\Delta}_{\sigma(k)\sigma(l)})\right),$$

where σ denotes any permutation on $[K]$. It is easy to see that $d(\boldsymbol{\Delta},\boldsymbol{\Delta}^*)$ is a variant of the KL divergence used to measure the difference between the estimated and true parameters. Similar measurements have been considered in Ref. [23].

4.1 Asymptotic consistency in estimating the model parameter

In this section, we first study the asymptotic property of the proposed method in estimating the model parameters $\boldsymbol{\Delta}^*$. The following result holds.

Lemma 1. For any $1\le i<j\le n$ with $c^*_{ik}=c^*_{jl}=1$, we have

$$E_{\boldsymbol{\Delta}^*}\left(\log f_{\boldsymbol{\Delta}_{k'l'}}(a_{ij})\right)\le E_{\boldsymbol{\Delta}^*}\left(\log f_{\boldsymbol{\Delta}^*_{kl}}(a_{ij})\right), \tag{10}$$

for any $\boldsymbol{\Delta}$ and $k',l'\in[K]$, and the equality holds if and only if $\boldsymbol{\Delta}_{k'l'}=\boldsymbol{\Delta}^*_{kl}$.

Lemma 1 shows that the log-likelihood function $\log f_{\boldsymbol{\Delta}_{k'l'}}(a_{ij})$ is maximized at $\boldsymbol{\Delta}^*_{kl}$ in expectation. Thus, it is an appropriate loss function for estimating $\boldsymbol{\Delta}^*_{kl}$. Furthermore, Lemma 1 implies that $d(\boldsymbol{\Delta},\boldsymbol{\Delta}^*)\ge 0$ and the equality holds if and only if $\boldsymbol{\Delta}=\boldsymbol{\Delta}^*$. Therefore $d(\boldsymbol{\Delta},\boldsymbol{\Delta}^*)\ge 0$ is a well-defined measurement.

Lemma 2. Let $N_k(\boldsymbol{C})=\sum_{i=1}^n c_{ik}$, where $\boldsymbol{c}_i=(c_{i1},\cdots,c_{iK})$ is independently and identically sampled from $\mathcal{M}(1,\boldsymbol{\alpha})$. Then for any $\gamma>0$ and $k\in[K]$, it holds true as $n\to\infty$ that $\sqrt{\dfrac{n}{(\log n)^{1+\gamma}}}\left|\dfrac{N_k(\boldsymbol{C})}{n}-\alpha_k\right|\to 0$, a.s.

Lemma 2 shows that as the number of n increases, the relative frequency of each component of $\boldsymbol{C}$, $N_k(\boldsymbol{C})/n$, converges almost surely to α_k. Furthermore, the convergence rate is determined by $\sqrt{n/(\log n)^{1+\gamma}}$. To proceed, the following technical assumptions are made.

Assumption 1. For any $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$, there exists $\gamma>0$ and a positive sequence $\{T_n\}_{n=1}^\infty$ such that $\min_k\alpha_k\ge(\log n)^{-\gamma/4}$ and $T_n{}^{-1}\le\sigma^2_{kl}\le T_n$ for all $1\le k,l\le K$.

Assumption 2. For any $\boldsymbol{\Delta}$, $\sum_{l=1}^K\left|\boldsymbol{\Delta}_{k_1l}-\boldsymbol{\Delta}_{k_2l}\right|_2>0$ when $1\le k_1\ne k_2\le K$.

Assumption 1 assumes a lower bound on $\boldsymbol{\alpha}$ and a bound on $\boldsymbol{\Sigma}$, which means that no community can degenerate too fast and the variance of edge weights should be neither too large nor too small. Note that Assumption 1 is weaker than the condition in Ref. [24], where $\min_k\alpha_k$ is required to be larger than some positive constant. Assumption 2 ensures that the community structure is distinguishable in $\boldsymbol{\Delta}$ in the sense that there exists $l\in[K]$ such that $\boldsymbol{\Delta}_{k_1l}$ and $\boldsymbol{\Delta}_{k_2l}$ are not equal. Let $l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c})=\sum_{i<j}\log(P_{\boldsymbol{\Delta}}(a_{ij}\mid\boldsymbol{C}=\boldsymbol{c}))$, then the following result holds.

Proposition 1. Suppose Assumption 1 holds and $\epsilon_n\succ T_n n^{-1/2}(\log n)^{1/2}$, then when n is sufficiently large, it holds true that

$$P_{\boldsymbol{\Delta}^*}\left(\sup_{(\boldsymbol{\Delta},\boldsymbol{c})}\frac{2}{n(n-1)}\left|l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c})-E_{\boldsymbol{\Delta}^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c}))\right|\ge\epsilon_n\right)\le\left(\frac{K}{n}\right)^n. \tag{11}$$

Proposition 1 demonstrates the asymptotic convergence of $l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c})$ to its expected value $E_{\boldsymbol{\Delta}^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c}))$, which is an important intermediate result for the following theorem.

Theorem 1. Suppose Assumptions 1 and 2 hold. If $n^{1/2}(\log n)^{-1/2-\gamma/2}\max_k\left|N_k(c^*)/n-\alpha^*_k\right|\to 0$ and $T_n n^{-1/2}\cdot(\log n)^{(1+\gamma)/2}\prec 1$ for some $\gamma>0$, then for any $\epsilon>0$, it holds true that $P_{\boldsymbol{\Delta}^*}(d(\hat{\boldsymbol{\Delta}},\boldsymbol{\Delta}^*)\ge\epsilon)\to 0$ almost surely as $n\to\infty$.

Theorem 1 shows that under some mild conditions, the estimator $\hat{\boldsymbol{\Delta}}$ converges to the true parameter $\boldsymbol{\Delta}^*$ in probability. Note that T_n is allowed to diverge to infinity at a rate not faster than $(\log n)^{-(1+\gamma)/2}n^{1/2}$, indicating strong variance for each edge. Recall that $\hat{\boldsymbol{\Delta}}$ is not attainable and our final estimate is $\boldsymbol{\Delta}^v$. Next, we establish the asymptotic consistency of the variational estimate $\boldsymbol{\Delta}^v$.

Theorem 2. Suppose all assumptions in Theorem 1 hold. For any $\epsilon>0$, it holds true that $P_{\boldsymbol{\Delta}^*}(d(\hat{\boldsymbol{\Delta}}^v,\boldsymbol{\Delta}^*)\ge\epsilon)\to 0$ almost surely as $n\to\infty$.

Theorem 2 establishes the asymptotic consistency of the variational estimator $\hat{\boldsymbol{\Delta}}^v$. This theoretical result provides a foundation for the application of the WSBM model and variational EM algorithm, ensuring that even with computational simplifications, the variational estimate still effectively approximates the true value, making it widely applicable in practice.

4.2 Asymptotic consistency in estimating community structure

To establish the asymptotic consistency of community detection, the following assumption is added.

Assumption 3. For any $1\le i<j\le n$, if $c^*_{ik}=c^*_{jl}=1$, then for any $\Delta^*_{k'l'}\ne\Delta^*_{kl}$, $-E_{\boldsymbol{\Delta}^*}\left(\log\dfrac{f_{\Delta^*_{k'l'}}(a_{ij})}{f_{\Delta^*_{kl}}(a_{ij})}\right)\ge(\log n)^{-\gamma/4}$.

It follows from Lemma 1 that $-E_{\boldsymbol{\Delta}^*}\left(\log\dfrac{f_{\Delta^*_{k'l'}}(a_{ij})}{f_{\Delta^*_{kl}}(a_{ij})}\right)$ is always positive if $c^*_{ik}=c^*_{jl}=1$ and $\boldsymbol{\Delta}^*_{k'l'}\ne\boldsymbol{\Delta}^*_{kl}$. Thus, Assumption 3 further ensures that it cannot vanish too fast. Then we have the following results.

Proposition 2. Suppose all assumptions in Theorem 1 and Assumption 3 hold. For any $\epsilon>0$, it holds true that $P_{\boldsymbol{\Delta}^*}(P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}\ne\boldsymbol{c}^*\mid\boldsymbol{A})\ge\epsilon)\to 0$ almost surely as $n\to\infty$.

Theorem 3. Suppose all assumptions in Theorem 1 and Assumption 3 hold. The following results hold as $n\to\infty$:

(a) $P_{\boldsymbol{\Delta}^*}(\tilde{\boldsymbol{c}}\ne\boldsymbol{c}^*)\to 0$, where $\tilde{\boldsymbol{c}}=\operatorname{argmax}_c P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}=\boldsymbol{c}\mid\boldsymbol{A})$;

(b) $P_{\boldsymbol{\Delta}^*}\left(KL\left(R(\boldsymbol{C},\hat{\boldsymbol{\beta}}(\boldsymbol{\Theta}^*)),P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}|\boldsymbol{A})\right)\ge\epsilon\right)\to 0$, where $\hat{\boldsymbol{\beta}}(\boldsymbol{\Theta}^*)=\operatorname*{argmax}_{\beta_{ik}\in[0,1];\sum_k\beta_{ik}=1}\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Theta}^*)$ and $R(\boldsymbol{C},\hat{\boldsymbol{\beta}}(\boldsymbol{\Theta}^*))=\prod_{i=1}^n h_{\hat{\beta}_i(\boldsymbol{\Theta}^*)}(\boldsymbol{C}_i)$.

Proposition 2 paves a bridge to establish Theorem 3. In

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, 2026, 56(2): 0203

Theorem 3, (a) first shows that the maximizer of $P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}|\boldsymbol{A})$ exactly recovers the true community structure $\boldsymbol{c}^*$ asymptotically. Although directly optimizing $P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}|\boldsymbol{A})$ is computationally intractable, (b) further establishes that it can be well approximated by $R(\boldsymbol{C},\hat{\boldsymbol{\beta}}(\boldsymbol{\Theta}^*))$ when $\boldsymbol{\Theta}^*$ is known. This result ensures that, even in the absence of an explicit posterior distribution, we can approximate the true community structure effectively. When $\boldsymbol{\Theta}^*$ is unknown, it is necessary to quantify the behavior of $\hat{\boldsymbol{\Delta}}^v$ as specified in Assumption 3, which is much more involved and additional assumptions should be needed. Fortunately, it is shown in our numerical experiments that the performance of the proposed method in terms of community detection is satisfactory due to the fact that $\hat{\boldsymbol{\Delta}}^v$ approximates $\boldsymbol{\Delta}^*$ well enough under Theorem 2.

5 Numerical experiments

In this section, we examine the performance of the proposed WSBM model in both the synthetic dataset and the yeast experimental dataset, and compare the WSBM method with four popular community detection methods in the literature. These methods include the spectral clustering (SC) method[25], which estimates communities through spectral decomposition and the K-means algorithm; the threshold spectral clustering (TSC) method[25], which is similar to spectral clustering but sets a threshold to convert the weighted adjacency matrix into an unweighted 0-1 matrix; the Louvain modularity (LM) method[11], which utilizes iterative optimization of modularity to identify community structure; and the nonnegative matrix factorization (NMF)[20], which factorizes the network's adjacency matrix or other related matrices into two nonnegative matrices, reveals the underlying community structure. The performance of each method is evaluated by the community detection error[26], which is defined as

$$\operatorname{error}(g,\hat{g}) = \frac{2}{n(n-1)} \sum_{i<j} I(\hat{g}(i)=\hat{g}(j), g(i)\neq g(j)) + I(g(i)=g(j), \hat{g}(i)\neq\hat{g}(j)),$$

where $I(\cdot)$ is the indicator function, $\{g(1),\cdots,g(n)\}$ and $\{\hat{g}(1),\cdots,\hat{g}(n)\}$ denote the true and estimated community labels, respectively. It is clear that $\operatorname{error}(g,\hat{g})$ represents the probability of divergence between g and $\hat{g}$ and is invariant with respect to label permutation.

5.1 Synthetic networks

Let $n=1000$ and $K=4$. For each node $i\in[n]$, we generate $\boldsymbol{C}_i \overset{\text{iid}}{\sim} \mathcal{M}(1;\boldsymbol{\alpha})$. Next, we generate $\boldsymbol{\mu}=(\mu_{kl})_{K\times K}$ and $\boldsymbol{\Sigma}=(\sigma^2_{kl})_{K\times K}$. Specifically, let $\mu_{kl}\sim U(-d_m,d_m)$, and $\sigma^2_{kl}=d^2_v$. Finally, the adjacency matrix $\boldsymbol{A}=(a_{ij})_{n\times n}$ is generated with $a_{ij}\sim N(\mu_{g(i)g(j)},\sigma^2_{g(i)g(j)})$.

We use the parameter d_m to control the range of all communities. As d_m increases, nodes in different communities can be assigned into a wider space. The variance of nodes within the same communities is controlled by d_v. As d_v decreases, nodes in one community will be more concentrated around its center. Therefore, it is expected that the task of community detection will become easier as d_m increases or d_v decreases. We evaluate all competing methods in these two aspects in the following two examples.

Example 1. In this example, we set $\boldsymbol{\alpha}=(0.25, 0.25, 0.25, 0.25)$, $d_m\in\{0.3, 0.4, 0.5, 0.6\}$ with $d_v\in\{0.6, 0.8\}$. The averaged community detection errors of all methods over 50 independent replications and their standard errors are plotted in Fig. 1.

Example 2. In this example, we set $\boldsymbol{\alpha}=(0.25, 0.25, 0.25, 0.25)$, $d_v\in\{0.4, 0.6, 0.8, 1.0\}$ with $d_m\in\{0.4, 0.5\}$. The averaged community detection errors of all methods over 50 independent replications and their standard errors are plotted in Fig. 2.

From Figs. 1 and 2, it is clear that the proposed method has the superior performance over four competitors in all cases. TSC and LM do not perform well with error rates exceeding 20% due to the fact that these two methods are designed for binary networks. SC and NMF outperform both TSC and LM, but are still worse than WSBM. It is interesting to note that as d_m increases the error rates of almost all methods decrease. This is because communities will be more separated from each other when d_m increases, which reduces the difficulty of the community detection. Similarly, as d_v increases, the error rates of almost all methods increase. This is because the nodes within the community are more dispersed as d_v increases, and thus it becomes more difficult to detect communities.

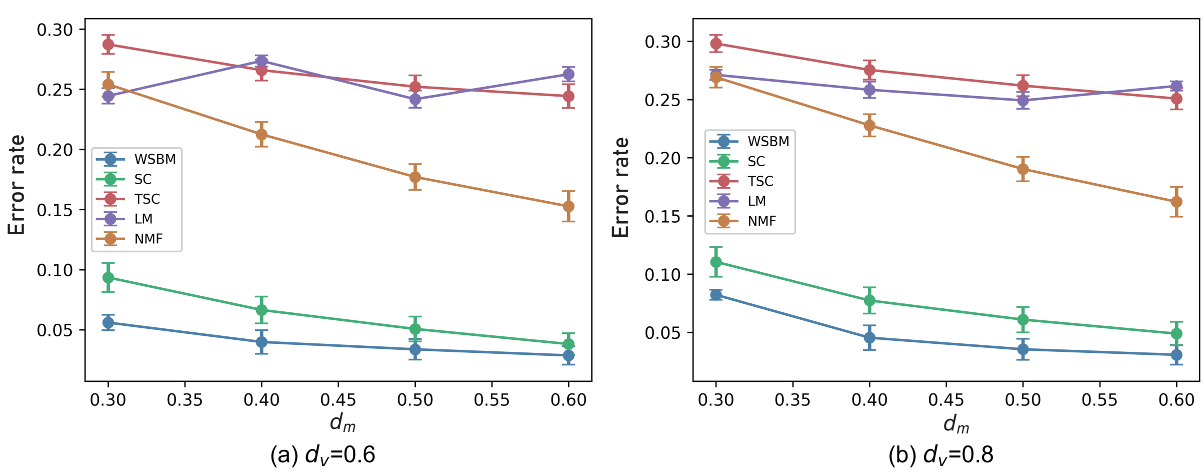


Fig. 1. The averaged community detection errors of all methods over 50 independent replications and its standard errors in Example 1.

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203

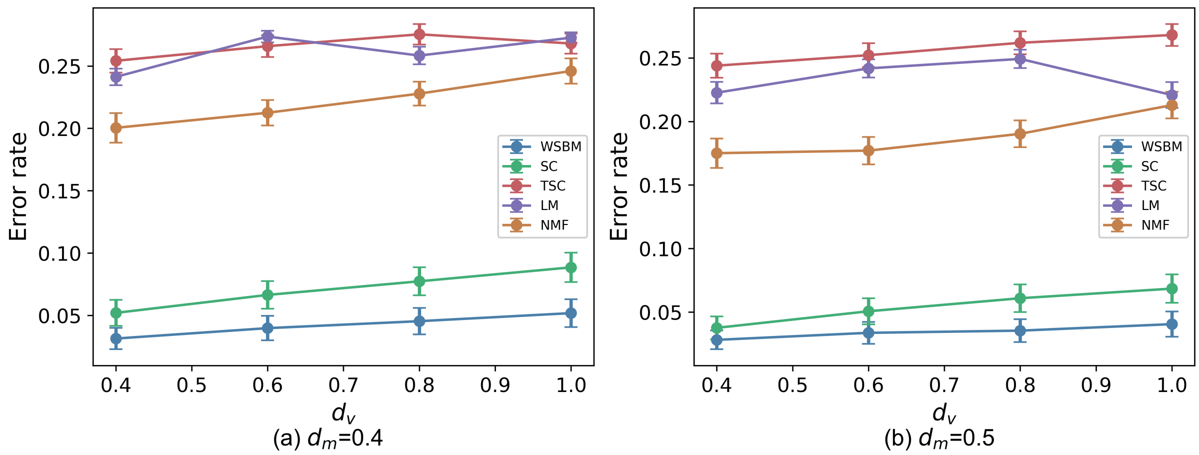


Fig. 2. The averaged community detection errors of all methods over 50 independent replications and its standard errors in Example 2.

Example 3. In this example, we set $\boldsymbol{\alpha} = (0.1, 0.2, 0.3, 0.4)$, and $d_m \in \{0.3, 0.4, 0.5, 0.6\}$ with $d_v \in \{0.6, 0.8\}$. The averaged community detection errors of all methods over 50 independent replications and their standard errors are plotted in Fig. 3.

Examples 1 and 2 assess the performance of different methods for community detection under varying parameters in balanced communities, whereas Example 3 examines their performance in imbalanced communities. From Fig. 3, it is clear that the proposed method has the superior performance compared to four competitors in all cases. TSC and LM do not perform well with error rates exceeding 20%. SC and NMF outperform both TSC and LM but are still worse than WSBM. Meanwhile, as d_m increases, the error rates of almost all methods decrease, because communities become more distinct.

5.2 Real application

In this section, we study the yeast experimental dataset using the proposed method and compare its performance with four competitors. Ideker et al.[27] studied the genes involved in the galactose utilisation (GAL) process in Saccharomyces cerevisiae, where gene similarity reflects the strength of gene interactions. Yeung et al.[28] then conducted experiments on yeast, measuring gene expression responses to 20 systematic perturbations through various genetic and environmental manipulations, with four replications of each perturbation. The experimental dataset is publicly available at http://www.expression.washington.edu/public. For our analysis, we select a subset of 205 genes from this dataset, which exhibit a community structure that can be classified into four nonoverlapping communities[29]. These four gene communities largely split with the functional categories of Saccharomyces cerevisiae, including protein metabolism and modification, carbohydrate metabolism and catabolism, metabolism and transport of nucleobases, nucleosides, nucleotides, and nucleic acids.

The dataset is processed as follows. Denote gene $i \in [205]$ as node i, with the gene expression response of gene i under 20 systematic perturbations denoted by the vector $\boldsymbol{x}_i \in \mathbb{R}^{20}$. The similarity between genes i and j is defined as $a_{ij} = \exp\left\{-\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|_2\right\}$. This implies that a larger difference in gene expression responses corresponds to lower similarity between the two genes. Since the yeast experiment was repeated four times for each perturbation, we define the adjacency matrix be $\boldsymbol{A} = (\boldsymbol{A}_1 + \boldsymbol{A}_2 + \boldsymbol{A}_3 + \boldsymbol{A}_4)/4$, where $\boldsymbol{A}_i$ represents the adjacency matrix in the ith experiment.

Next, we plot the largest 50 singular values of the adjacency matrix in Fig. 4 and a gap between the fourth and fifth singular values can be observed. This also suggests that there should be four communities[30,31]. If K is set to 3, we find the

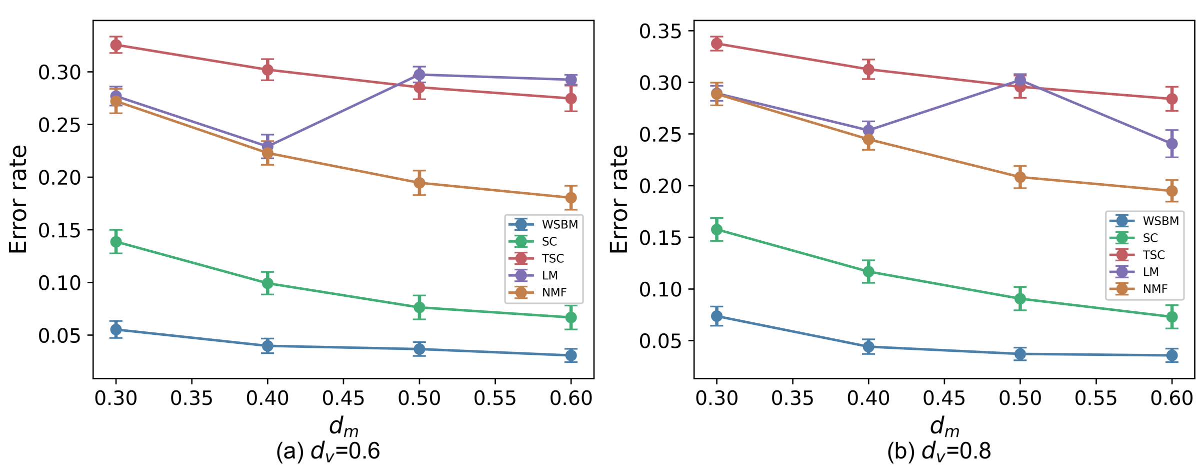


Fig. 3. The averaged community detection errors of all methods over 50 independent replications and its standard errors in Example 3.

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203

true second and fourth communities are merged into one community. If K is set to 5, we find the true first community is split into more than one community. Therefore, we choose $K = 4$ for these data.

We apply all methods to this yeast dataset with $K = 4$, and the community detection errors of all methods are reported in Table 1. As shown in Table 1, WSBM outperforms all competing methods, with an improvement of 20.3% compared with SC, 48.2% compared with TSC, 52.2% compared with LM, and 38.4% compared with NMF. This demonstrates the power of the proposed method in detecting communities in weighted networks.

6 Conclusions

In this paper, we propose a novel weighted stochastic block model (WSBM) for weighted networks. The variational EM algorithm is employed to estimate the model parameters, and recover the community structure of the network. Moreover, we establish the asymptotic consistency of the proposed method in terms of both parameter estimation and community detection. We evaluate the proposed method on both simulated datasets and the yeast experimental dataset, comparing its performance with four existing community detection algorithms. The superior performance of the proposed model demonstrates the strength of the proposed method in detecting communities in weighted networks.

In summary, this study provides an effective framework for community detection in weighted networks, offering improved accuracy and interpretability in revealing latent network structures. However, the current analysis is limited to single-layer weighted networks. Future work could extend the proposed method to weighted multilayer networks with inter-layer dependencies and relax the Gaussian assumption of the adjacency matrix into more flexible modeling frameworks.

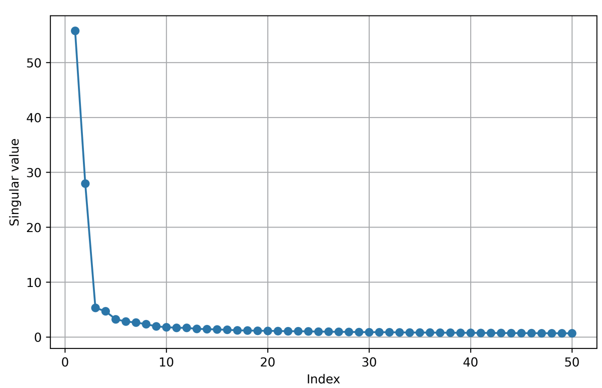


Fig. 4. Top 50 singular values of the adjacency matrix.

Table 1. Community detection errors of various methods in yeast experimental dataset.

WSBM	SC	TSC	LM	NMF
0.0665	0.0834	0.1285	0.139	0.1079

Acknowledgements

This work was supported by USTC Research Funds of the Double First-Class Initiative (YD2040002020) and the National Natural Science Foundation of China (12301388).

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Lei Liu is currently a Ph.D. candidate at the School of Management, University of Science and Technology of China. Her research mainly focuses on community detection and network analysis.

Jingnan Zhang is currently an Associate Professor at the University of Science and Technology of China. He received his Ph.D. degree in Statistics from City University of Hong Kong in 2022. His research mainly focuses on statistical machine learning and network data analysis.

References

[1] Girvan M, Newman M E J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, **2002**, *99* (12): 7821–7826.

[2] Yang J, McAuley J, Leskovec J. Community detection in networks with node attributes. In: 2013 IEEE 13th International Conference on Data Mining. Dallas, TX, USA: IEEE, **2013**: 1151–1156.

[3] Cantini L, Medico E, Fortunato S, et al. Detection of gene communities in multi-networks reveals cancer drivers. *Scientific Reports*, **2015**, *5* (1): 17386.

[4] Pothen A. Graph partitioning algorithms with applications to scientific computing. In: Parallel Numerical Algorithms. Dordrecht, Netherlands: Springer, **1997**: 323–368.

[5] Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, **1998**, *30*: 107–117.

[6] Bedi P, Sharma C. Community detection in social networks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **2016**, *6* (3): 115–135.

[7] Zhao Y. A survey on theoretical advances of community detection in networks. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2017**, *9* (5): e1403.

[8] MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. Berkeley, CA, USA: University of California Press, **1967**: 281–297.

[9] Roux M. A comparative study of divisive hierarchical clustering algorithms. ArXiv: 1506.08977, **2015**.

[10] Von Luxburg U. A tutorial on spectral clustering. *Statistics and Computing*, **2007**, *17* (4): 395–416.

[11] Newman M E. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, **2006**, *103* (23): 8577–8582.

[12] Shi J, Malik J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2000**, *22* (8): 888–905.

[13] Holland P W, Laskey K B, Leinhardt S. Stochastic blockmodels: First steps. *Social Networks*, **1983**, *5* (2): 109–137.

[14] Hoff P D, Raftery A E, Handcock M S. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, **2002**, *97* (460): 1090–1098.

[15] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203

networks. *Physical Review Letters*, **2001**, *86* (14): 3200–3203.
[16] Jin J, Pan L, Wang C, et al. A center-based community detection method in weighted networks. In: 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence. Boca Raton, FL, USA: IEEE, **2011**: 513–518.
[17] Haq N F, Moradi M, Wang Z J. Community structure detection from networks with weighted modularity. *Pattern Recognition Letters*, **2019**, *122*: 14–22.
[18] Hager W W, Phan D T, Zhang H. An exact algorithm for graph partitioning. *Mathematical Programming*, **2013**, *137* (1): 531–556.
[19] Wang T-S, Lin H-T, Wang P. Weighted-spectral clustering algorithm for detecting community structures in complex networks. *Artificial Intelligence Review*, **2017**, *47*: 463–483.
[20] Wang F, Li T, Wang X, et al. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, **2011**, *22*: 493–521.
[21] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the *EM* algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **1977**, *39* (1): 1–22.
[22] Mariadassou M, Robin S, Vacher C. Uncovering latent structure in valued graphs: a variational approach. *Annals of Applied Statistics*, **2010**, *4* (2): 715–742.
[23] Zhang J, Li C, Wang J. A stochastic block Ising model for multi-layer networks with inter-layer dependence. *Biometrics*, **2023**, *79* (4): 3564–3573.
[24] Celisse A, Daudin J-J, Pierre L. Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics*, **2012**, *6*: 1847–1899.
[25] Xu M, Jog V, Loh P-L. Optimal rates for community estimation in the weighted stochastic block model. *The Annals of Statistics*, **2020**, *48* (1): 183–204.
[26] Wang J. Consistent selection of the number of clusters via crossvalidation. *Biometrika*, **2010**, *97* (4): 893–904.
[27] Ideker T, Thorsson V, Ranish J A, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, **2001**, *292* (5518): 929–934.
[28] Yeung K Y, Medvedovic M, Bumgarner R E. Clustering gene-expression data with repeated measurements. *Genome Biology*, **2003**, *4*: 1–17.
[29] Li C. Community detection on multi-layer networks with inter-layer dependence. Thesis. Hong Kong, China: City University of Hong Kong, **2022**.
[30] Ji P, Jin J. Coauthorship and citation networks for statisticians. *The Annals of Applied Statistics*, **2016**, *10* (4): 1779–1812.
[31] Zhang J, He X, Wang J. Directed community detection with network embedding. *Journal of the American Statistical Association*, **2022**, *117* (540): 1809–1819.
[32] Hoeffding W. Probability inequalities for sums of bounded random variables. In: The Collected Works of Wassily Hoeffding. New York: Springer, **1994**: 409–426.
[33] Boucheron S, Lugosi G, Massart P. Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford, UK: Oxford University Press, **2013**.
[34] Massart P. Concentration Inequalities and Model Selection. Berlin, Heidelberg: Springer-Verlag, **2007**.

Appendix

Proof of Lemma 1. Let $c^*_{ik}=c^*_{jl}=c_{ik'}=c_{jl'}=1$, $\boldsymbol{\Delta}_{k'l'}=(\mu_{k'l'},\sigma^2_{k'l'})$, $\boldsymbol{\Delta}^*_{kl}=(\mu^*_{kl},\sigma^{*2}_{kl})$, then

$$\begin{aligned}E_{\boldsymbol{\Delta}^*}\left(\log f_{\boldsymbol{\Delta}_{k'l'}}(a_{ij})\right)=&E_{\boldsymbol{\Delta}^*}\left(-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log\sigma^2_{k'l'}-\frac{(a_{ij}-\mu_{k'l'})^2}{2\sigma^2_{k'l'}}\right)=\\&-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log\sigma^2_{k'l'}-\frac{1}{2\sigma^2_{k'l'}}E_{\boldsymbol{\Delta}^*}(a_{ij}-\mu_{k'l'})^2=-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log\sigma^2_{k'l'}-\frac{(\mu^*_{kl}-\mu_{k'l'})^2+\sigma^{*2}_{kl}}{2\sigma^2_{k'l'}}.\end{aligned}\tag{A1}$$

The derivation shows that Eq. (A1) attains its maximum value if and only if $(\mu_{k'l'},\sigma^2_{k'l'})=(\mu^*_{kl},\sigma^{*2}_{kl})$.

Proof of Lemma 2. For $\epsilon\geq 0$, by Hoeffding's inequality[32], when n is large enough there are

$$P\left(\sqrt{\frac{n}{(\log n)^{1+\gamma}}}\left|\frac{N_k(\boldsymbol{C})}{n}-\alpha_k\right|\geq\epsilon\right)=P\left(\left|\frac{N_k(\boldsymbol{C})}{n}-\alpha_k\right|\geq\epsilon\sqrt{\frac{(\log n)^{1+\gamma}}{n}}\right)\leq 2\exp\{-2\epsilon^2(\log n)^{1+\gamma}\}=2n^{-2\epsilon^2(\log n)^{\gamma}}\leq 2n^{-2},$$

then

$$\sum_{n=1}^{\infty}P\left(\sqrt{\frac{n}{(\log n)^{1+\gamma}}}\left|\frac{N_k(\boldsymbol{C})}{n}-\alpha_k\right|\geq\epsilon\right)\leq\sum_{n=1}^{\infty}\frac{2}{n^2}<\infty.$$

By Borel–Cantelli lemma,

$$P\left(\limsup_{n\to\infty}\left\{\sqrt{\frac{n}{(\log n)^{1+\gamma}}}\left|\frac{N_k(\boldsymbol{C})}{n}-\alpha_k\right|\geq\epsilon\right\}\right)=0.$$

Proof of Proposition 1. By Assumption 1,

$$\left|\log\left(f_{\boldsymbol{\Delta}_{kl}}(a_{ij})\right)\right|\leq\left|\frac{(a_{ij}-\mu_{kl})^2}{2\sigma^2_{kl}}\right|+\frac{1}{2}\max_{k,l}\left|\log(2\pi)+\log(\sigma^2_{kl})\right|\leq\frac{(a_{ij}-\mu_{kl})^2}{2\sigma^2_{kl}}+\frac{1}{2}\log(2\pi)+\frac{1}{2}\max_{k,l}\left|\log(\sigma^2_{kl})\right|\leq B_1T_n,$$

where the second inequality follows from the monotonicity of the composite logarithmic-inverse proportional function and B_1 is some positive constant. It is further shown that

$$\left|\log\left(P_{\boldsymbol{\Delta}}(a_{ij}\mid\boldsymbol{C}=\boldsymbol{c})\right)\right|=\sum_{k,l=1}^{K}c_{ik}c_{jl}\left|\log\left(f_{\boldsymbol{\Delta}_{kl}}(a_{ij})\right)\right|\leq B_1T_n.\tag{A2}$$

Let $q(a_{ij};\boldsymbol{\Delta})=\log\left(P_{\boldsymbol{\Delta}}(a_{ij}\mid\boldsymbol{C}=\boldsymbol{c})\right)-E_{\boldsymbol{\Delta}^*}\left(\log\left(P_{\boldsymbol{\Delta}}(a_{ij}\mid\boldsymbol{C}=\boldsymbol{c})\right)\right)$, then by the triangle inequality, for $1\leq i<j\leq n$, we have

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203

$$\left|q(a_{ij};\boldsymbol{\Delta})\right| \le 2B_1 T_n.$$

Let $\boldsymbol{A}'$ be an independent replication of $\boldsymbol{A}$, and let ϵ_{ij}'s be independent Rademacher random variables[33], then

$$E_{\Delta^*}\left(\sup_{\Delta}\left|\sum_{i<j} q(a_{ij};\boldsymbol{\Delta})\right|\right) \le E_{\Delta^*}\left(\sup_{\Delta}\left|\sum_{i<j}\left\{\log\left(P_{\Delta}(a_{ij}\mid \boldsymbol{C}=\boldsymbol{c})\right)-\log\left(P_{\Delta}\left(a'_{ij}\mid \boldsymbol{C}=\boldsymbol{c}\right)\right)\right\}\right|\right) \le$$

$$2E_{\Delta^*}\left(\sup_{\Delta} E_{\epsilon}\left|\sum_{i<j}\epsilon_{ij}\log\left(P_{\Delta}(a_{ij}\mid \boldsymbol{C}=\boldsymbol{c})\right)\right|\right) \le 2E_{\Delta^*}\left(\sup_{\Delta}\sqrt{V_{\epsilon}\left(\sum_{i<j}\epsilon_{ij}\log\left(P_{\Delta}(a_{ij}\mid \boldsymbol{C}=\boldsymbol{c})\right)\right)}\right) \le B_1\sqrt{2n(n-1)}T_n,$$

where the first two inequalities follow from the standard symmetrization argument, the third inequality follows from Jensen's inequality, and the last inequality follows from (A2).

Moreover, for $1 \le i < j \le n$, there are $E_{\Delta^*}(q(a_{ij};\boldsymbol{\Delta})) = 0$ and $V_{\Delta^*}(q(a_{ij};\boldsymbol{\Delta})) \le 4B_1^2T_n^2$. Since $|l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c}) - E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c}))| = \left|\sum_{i<j} q(a_{ij};\boldsymbol{\Delta})\right|$, and $q(a_{ij};\boldsymbol{\Delta})$ satisfies Talagrand's concentration inequality[34], then for any $x > 0$, we have

$$P_{\Delta^*}\left(\sup_{\Delta}\frac{2}{n(n-1)}\left|\sum_{i<j}q(a_{ij};\boldsymbol{\Delta})\right| \ge \frac{2B_1\sqrt{2}T_n}{\sqrt{n(n-1)}} + 8B_1T_n\sqrt{\frac{x}{n(n-1)}} + \frac{8B_1T_n}{n(n-1)}x\right) \le \mathrm{e}^{-x}.$$

Because $\epsilon_n \succ T_n n^{-1/2}(\log n)^{1/2}$, let $x = n\log n$,

$$P_{\Delta^*}\left(\sup_{(\Delta,\boldsymbol{c})}\frac{2}{n(n-1)}\left|\sum_{i<j}q(a_{ij};\boldsymbol{\Delta})\right| \ge \epsilon_n\right) \le \sum_{\boldsymbol{c}} P_{\Delta^*}\left(\sup_{\Delta}\frac{2}{n(n-1)}\left|\sum_{i<j}q(a_{ij};\boldsymbol{\Delta})\right| \ge \epsilon_n\right) \le \sum_{\boldsymbol{c}}\mathrm{e}^{-n\log n} \le \left(\frac{K}{n}\right)^n,$$

where the last inequality is obtained by the fact that $\boldsymbol{C}$ has K^n different values.

Proof of Theorem 1. Let's start with a few definitions.

Let $\mathcal{M} = \{\boldsymbol{M} = (M_{kl})_{k,l=1}^K : M_{kl} \in [0,1], \sum_{l=1}^K M_{kl} = 1, M_{kl} = n_l/N_k(\boldsymbol{c}^*),\ n_l \in \mathbb{Z}\}$. For any $\boldsymbol{\Theta} = (\boldsymbol{\alpha},\boldsymbol{\Delta})$, definite $\mathbb{H}_n(\boldsymbol{\alpha},\boldsymbol{\Delta}) = \dfrac{2}{n(n-1)}\cdot \log P_{\boldsymbol{\Theta}}(\boldsymbol{A})$, and

$$\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*) = \max_{\boldsymbol{M}\in\mathcal{M}}\left\{\sum_{k,l}\alpha_k^*\alpha_l^*\sum_{k',l'}M_{kk'}M_{ll'}\psi_{kl}(\boldsymbol{\Delta}_{k'l'})\right\}.$$

Define $\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{M}) = \sum_{k,l}\alpha_k^*\alpha_l^*\sum_{k',l'}M_{kk'}M_{ll'}\psi_{kl}(\boldsymbol{\Delta}_{k'l'})$. Then $\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*) = \max_{\boldsymbol{M}\in\mathcal{M}}\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{M})$.

Let $\hat{\boldsymbol{M}}(\boldsymbol{\Delta}) = \arg\max_{\boldsymbol{M}\in\mathcal{M}}\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{M})$, and denote $\hat{\boldsymbol{M}} = \hat{\boldsymbol{M}}(\boldsymbol{\Delta}), \hat{\boldsymbol{M}}^* = \hat{\boldsymbol{M}}(\boldsymbol{\Delta}^*)$. Then we have

$$\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*) - \mathbb{H}(\boldsymbol{\Delta}^*,\boldsymbol{c}^*) = \mathbb{H}(\boldsymbol{\Delta},\hat{\boldsymbol{M}}(\boldsymbol{\Delta})) - \mathbb{H}\left(\boldsymbol{\Delta}^*,\hat{\boldsymbol{M}}^*\right).$$

Noting that $\hat{\boldsymbol{M}}(\boldsymbol{\Delta}^*) = \arg\max_{\boldsymbol{M}\in\mathcal{M}}\mathbb{H}(\boldsymbol{\Delta}^*,\boldsymbol{M})$, so $\mathbb{H}\left(\boldsymbol{\Delta}^*,\hat{\boldsymbol{M}}^*\right) - \mathbb{H}(\boldsymbol{\Delta}^*,\boldsymbol{I}_K) \ge 0$. However, it follows from Lemma 1 that

$$\mathbb{H}\left(\boldsymbol{\Delta}^*,\hat{\boldsymbol{M}}^*\right) - \mathbb{H}(\boldsymbol{\Delta}^*,\boldsymbol{I}_K) = \sum_{k,l}\alpha_k^*\alpha_l^*\sum_{k',l'}\hat{M}_{kk'}^*\hat{M}_{ll'}^*\left\{\psi_{kl}(\boldsymbol{\Delta}_{k'l'}^*) - \psi_{kl}(\boldsymbol{\Delta}_{kl}^*)\right\} \le 0,$$

which implies that $\mathbb{H}\left(\boldsymbol{\Delta}^*,\hat{\boldsymbol{M}}^*\right) = \mathbb{H}(\boldsymbol{\Delta}^*,\boldsymbol{I}_K)$. Then

$$\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*) - \mathbb{H}(\boldsymbol{\Delta}^*,\boldsymbol{c}^*) = \mathbb{H}(\boldsymbol{\Delta},\hat{\boldsymbol{M}}(\boldsymbol{\Delta})) - \mathbb{H}(\boldsymbol{\Delta}^*,\boldsymbol{I}_K) = \sum_{k,l}\alpha_k^*\alpha_l^*\left[\sum_{k',l'}\hat{M}_{kk'}\hat{M}_{ll'}\left\{\psi_{kl}(\boldsymbol{\Delta}_{k'l'}) - \psi_{kl}(\boldsymbol{\Delta}_{kl}^*)\right\}\right] \le 0,$$

where $\hat{M}_{kk'}$ and $\hat{M}_{ll'}$ denote the corresponding entries of $\hat{\boldsymbol{M}}(\boldsymbol{\Delta})$.

By Lemma 1, for any $k,k',l,l' \in [K]$, it holds that $\psi_{kl}(\boldsymbol{\Delta}_{k'l'}) \le \psi_{kl}(\boldsymbol{\Delta}_{kl}^*)$. Therefore, $\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*) = \mathbb{H}(\boldsymbol{\Delta}^*,\boldsymbol{c}^*)$ if and only if for any $k,k',l,l' \in [K]$, we have $\hat{M}_{kk'}^*\hat{M}_{ll'}^*(\psi_{kl}(\boldsymbol{\Delta}_{k'l'}) - \psi_{kl}(\boldsymbol{\Delta}_{kl}^*)) = 0$. However, by the definition of $\mathcal{M}$, Lemma 1, and Assumption 2, we have $\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*) < \mathbb{H}(\boldsymbol{\Delta}^*,\boldsymbol{c}^*)$ for any $\boldsymbol{\Delta}$ with $d(\boldsymbol{\Delta},\boldsymbol{\Delta}^*) > 0$. Therefore, it follows from the continuity of $\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*)$ that $\sup_{d(\Delta,\Delta^*)\ge\epsilon}\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*) < \mathbb{H}(\boldsymbol{\Delta}^*,\boldsymbol{c}^*)$ for any given $\epsilon > 0$.

By Theorem 3.4 in Ref. [24], to prove the convergence of Theorem 1, it is sufficient to show that

$$P_{\Delta^*}\left(\sup_{(\alpha,\Delta)}|\mathbb{H}_n(\boldsymbol{\alpha},\boldsymbol{\Delta}) - \mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*)| \ge \epsilon\right) \to 0 \text{ almost surely as } n\to\infty. \tag{A3}$$

Denote $\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}} l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c})$ and $\tilde{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}} E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c}))$,

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203

$$|\mathbb{H}_n(\boldsymbol{\alpha},\boldsymbol{\Delta})-\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*)|\le\left|\mathbb{H}_n(\boldsymbol{\alpha},\boldsymbol{\Delta})-\frac{2}{n(n-1)}l(\boldsymbol{A};\boldsymbol{\Delta},\hat{\boldsymbol{c}})\right|+\frac{2}{n(n-1)}|l(\boldsymbol{A};\boldsymbol{\Delta},\hat{\boldsymbol{c}})-E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\tilde{\boldsymbol{c}}))|+$$
$$\left|\frac{2}{n(n-1)}E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\tilde{\boldsymbol{c}}))-\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*)\right|\triangleq|I_1|+\frac{2}{n(n-1)}|I_2|+|I_3|.$$

In the following, we analyze $|I_1|,|I_2|$, and $|I_3|$ individually.

Firstly, $I_1=\mathbb{H}_n(\boldsymbol{\alpha},\boldsymbol{\Delta})-\frac{2}{n(n-1)}l(\boldsymbol{A};\boldsymbol{\Delta},\hat{\boldsymbol{c}})=\frac{2}{n(n-1)}(\log P_{\boldsymbol{\Theta}}(\boldsymbol{A})-l(\boldsymbol{A};\boldsymbol{\Delta},\hat{\boldsymbol{c}}))$.

Let $\boldsymbol{\Pi}=\left\{\boldsymbol{\beta}\in[0,1]^{n\times K};\sum_{k=1}^K\beta_{ik}=1\ \ \forall i\in[n]\right\}$, and $\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}(\boldsymbol{\Theta})=\operatorname{argmax}_{\boldsymbol{\beta}\in\boldsymbol{\Pi}}\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Theta})$. It then follows from Lemma F.2 in Ref. [24] that

$$\mathcal{L}(\hat{\boldsymbol{c}},\boldsymbol{\Theta})\le\mathcal{L}(\hat{\boldsymbol{\beta}},\boldsymbol{\Theta})\le\log P_{\boldsymbol{\Theta}}(\boldsymbol{A})\le l(\boldsymbol{A};\boldsymbol{\Delta},\hat{\boldsymbol{c}}).\tag{A4}$$

Meanwhile, $\mathcal{L}(\hat{\boldsymbol{c}},\boldsymbol{\Theta})-l(\boldsymbol{A};\boldsymbol{\Delta},\hat{\boldsymbol{c}})=\sum_{i,k}\hat{c}_{ik}\log\alpha_k$, then we have

$$|I_1|=\frac{2}{n(n-1)}\left|\log P_{\boldsymbol{\Theta}}(\boldsymbol{A})-l(\boldsymbol{A};\boldsymbol{\Delta},\hat{\boldsymbol{c}})\right|\le\frac{2}{n(n-1)}\left|\sum_{i,k}\hat{c}_{ik}\log\alpha_k\right|\le\frac{\log(\log n)}{n-1}\prec 1,$$

where the second inequality follows from Assumption 1 that $(\log n)^{-\delta/4}\le\min_k\alpha_k$.

Secondly, $I_2=l(\boldsymbol{A};\boldsymbol{\Delta},\hat{\boldsymbol{c}})-E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\tilde{\boldsymbol{c}}))$.

Simple algebra yields that

$$\begin{aligned}\frac{2}{n(n-1)}|I_2|=&\frac{2}{n(n-1)}I_2I(I_2>0)+\frac{2}{n(n-1)}(-I_2)I(I_2\le0)\le\\&\frac{2}{n(n-1)}|l(\boldsymbol{A};\boldsymbol{\Delta},\hat{\boldsymbol{c}})-E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\hat{\boldsymbol{c}}))|I(I_2>0)+\frac{2}{n(n-1)}|l(\boldsymbol{A};\boldsymbol{\Delta},\tilde{\boldsymbol{c}})-E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\tilde{\boldsymbol{c}}))|I(I_2\le0)\le\\&\frac{2}{n(n-1)}\sup_{\boldsymbol{c}}|l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c})-E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c}))|,\end{aligned}$$

where the first inequality follows from the definition of $\hat{c}$ and $\tilde{c}$.

Finally, $I_3=\frac{2}{n(n-1)}E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\tilde{\boldsymbol{c}}))-\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*)$.

For $|I_3|$, note that

$$\begin{aligned}E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\boldsymbol{c}))=&\sum_{i<j}E_{\Delta^*}(\log(P_{\Delta}(a_{ij}\mid\boldsymbol{C}=\boldsymbol{c})))=\sum_{i<j}\sum_{k,l,k',l'}\left(c^*_{ik}c^*_{jl}c_{ik'}c_{jl'}\psi_{kl}(\boldsymbol{\Delta}_{k'l'})\right)=\\&\frac{1}{2}\sum_{k,l,k',l'}N_{kk'}(\boldsymbol{c})N_{ll'}(\boldsymbol{c})\psi_{kl}(\boldsymbol{\Delta}_{k'l'})-\frac{1}{2}\sum_{k,k'}N_{kk'}(\boldsymbol{c})\psi_{kl}(\boldsymbol{\Delta}_{k'k'}),\end{aligned}$$

where $N_{kk'}(\boldsymbol{c})=\sum_{i=1}^n c^*_{ik}c_{ik'}$ for $1\le k,k'\le K$. Further, define $\boldsymbol{M}_{\boldsymbol{c}}=(M_{\boldsymbol{c},kk'})_{k,k'}$ with $M_{\boldsymbol{c},kk'}=N_{kk'}(\boldsymbol{c})/N_k(\boldsymbol{c}^*)$. Then $\boldsymbol{M}_{\boldsymbol{c}}\in\mathcal{S}$, and we re-write $E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\tilde{\boldsymbol{c}}))$ as $\Phi(\boldsymbol{\Delta},\boldsymbol{M}_{\tilde{c}})$. We have

$$\begin{aligned}|I_3|=&\left|\frac{2}{n(n-1)}E_{\Delta^*}(l(\boldsymbol{A};\boldsymbol{\Delta},\tilde{\boldsymbol{c}}))-\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*)\right|=\left|\frac{2}{n(n-1)}\Phi(\boldsymbol{\Delta},\boldsymbol{M}_{\tilde{c}})-\mathbb{H}(\boldsymbol{\Delta},\hat{\boldsymbol{M}}(\boldsymbol{\Delta}))\right|\le\sup_{\boldsymbol{M}\in\mathcal{M}}\left|\frac{2}{n(n-1)}\Phi(\boldsymbol{\Delta},\boldsymbol{M})-\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{M})\right|\le\\&\sup_{\boldsymbol{M}\in\mathcal{M}}\left|\sum_{k,l,k',l'}\left\{\frac{N_k(\boldsymbol{c}^*)N_l(\boldsymbol{c}^*)}{n(n-1)}-\alpha^*_k\alpha^*_l\right\}M_{kk'}M_{ll'}\psi_{kl}(\boldsymbol{\Delta}_{k'l'})\right|+\sup_{\boldsymbol{M}\in\mathcal{M}}\left|\sum_{k,k'}\frac{N_k(\boldsymbol{c}^*)}{n(n-1)}M_{kk'}\psi_{kl}(\boldsymbol{\Delta}_{k'l'})\right|\le\\&B_1T_n\left(\left|\sum_{k,l}\left\{\frac{N_k(\boldsymbol{c}^*)N_l(\boldsymbol{c}^*)}{n(n-1)}-\alpha^*_k\alpha^*_l\right\}\right|+\frac{1}{n-1}\right)\le B_1n^{1/2}(\log n)^{-1/2-\delta/2}\left(\left|\sum_{k,l}\left(\frac{N_k(\boldsymbol{c}^*)N_l(\boldsymbol{c}^*)}{n(n-1)}-\alpha^*_k\alpha^*_l\right)\right|+\frac{1}{n-1}\right)\le\\&B_1n^{1/2}(\log n)^{-1/2-\delta/2}\left(2\sum_k\left|\frac{N_k(\boldsymbol{c}^*)}{n}-\alpha^*_k\right|+\frac{2}{n-1}\right),\end{aligned}$$

where the second inequality follows from (A2) and the definition of ψ_{kl} that

$$|\psi_{kl}(\Delta_{k'l'})|\le E_{\Delta^*,c^*}\left(\left|f_{\Delta_{k'l'}}(a_{ij})\right|\right)\le B_1T_n.$$

It then follows from the assumption for $\boldsymbol{\alpha}^*$ that $|I_3|\prec 1$, and

$$P_{\Delta^*}\left(\sup_{(\alpha,\Delta)}|\mathbb{H}_n(\boldsymbol{\alpha},\boldsymbol{\Delta})-\mathbb{H}(\boldsymbol{\Delta},\boldsymbol{c}^*)|\ge\epsilon\right)\le P_{\Delta^*}\left(\sup_{\Delta}\frac{2}{n(n-1)}|I_2|\ge\epsilon/2\right)\to 0,$$

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203

for any $\epsilon > 0$, where the convergence follows from Proposition 1. This completes the proof of Theorem 1.

Proof of Theorem 2. Let $\boldsymbol{\Pi} = \left\{\boldsymbol{\beta} \in [0,1]^{n\times K}; \sum_{k=1}^{K} \beta_{ik} = 1\ \forall i \in [n]\right\}$, and $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\boldsymbol{\Theta}) = \operatorname{argmax}_{\boldsymbol{\beta}\in\boldsymbol{\Pi}} \mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Theta})$. Then $\hat{\boldsymbol{\beta}}^{v} = \hat{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\Theta}}^{v}\right)$. For each $\boldsymbol{\Theta}$, define $\mathbb{J}_n(\boldsymbol{\alpha},\boldsymbol{\Delta}) = \dfrac{2}{n(n-1)}\mathcal{L}(\hat{\boldsymbol{\beta}},\boldsymbol{\Theta})$. It follows from (A4) that

$$\left|\mathcal{L}(\hat{\boldsymbol{\beta}},\boldsymbol{\Theta}) - \log P_{\boldsymbol{\theta}}(\boldsymbol{A})\right| \le n \max_k \left|\log \alpha_k\right| \le n \log\log n.$$

Therefore,

$$\sup_{(\boldsymbol{\alpha},\boldsymbol{\Delta})} |\mathbb{J}_n(\boldsymbol{\alpha},\boldsymbol{\Delta}) - \mathbb{H}_n(\boldsymbol{\alpha},\boldsymbol{\Delta})| = \frac{2}{n(n-1)} \sup_{\boldsymbol{\Theta}}\left|\mathcal{L}(\hat{\boldsymbol{\beta}},\boldsymbol{\Theta}) - \log P_{\boldsymbol{\theta}}(\boldsymbol{A})\right| \le \frac{\log\log n}{n} \prec 1. \tag{A5}$$

The desired result follows immediately from (A5) and Theorem 3.4 in Ref. [24].

Proof of Proposition 2. Using the boundedness of the probability we get

$$P_{\boldsymbol{\Delta}^*}(P_{\boldsymbol{\Theta}^*}(\boldsymbol{C} \ne \boldsymbol{c}^* \mid \boldsymbol{A}) \ge \epsilon) \le P_{\boldsymbol{\Delta}^*}\left(\frac{P_{\boldsymbol{\Theta}^*}(\boldsymbol{C} \ne \boldsymbol{c}^* \mid \boldsymbol{A})}{P_{\boldsymbol{\Theta}^*}(\boldsymbol{C} = \boldsymbol{c}^* \mid \boldsymbol{A})} \ge \epsilon\right) = P_{\boldsymbol{\Delta}^*}\left(\sum_{\boldsymbol{c}\ne\boldsymbol{c}^*}\frac{P_{\boldsymbol{\Theta}^*}(\boldsymbol{C} = \boldsymbol{c} \mid \boldsymbol{A})}{P_{\boldsymbol{\Theta}^*}(\boldsymbol{C} = \boldsymbol{c}^* \mid \boldsymbol{A})} \ge \epsilon\right).$$

Simple algebra yields that

$$\log\frac{P_{\boldsymbol{\Theta}^*}(\boldsymbol{C} = \boldsymbol{c} \mid \boldsymbol{A})}{P_{\boldsymbol{\Theta}^*}(\boldsymbol{C} = \boldsymbol{c}^* \mid \boldsymbol{A})} = \sum_{i<j}\log\frac{P_{\boldsymbol{\Delta}^*}(a_{ij} \mid \boldsymbol{C} = \boldsymbol{c})}{P_{\boldsymbol{\Delta}^*}(a_{ij} \mid \boldsymbol{C} = \boldsymbol{c}^*)} + \log\frac{P_{\boldsymbol{\alpha}^*}(\boldsymbol{C} = \boldsymbol{c})}{P_{\boldsymbol{\alpha}^*}(\boldsymbol{C} = \boldsymbol{c}^*)} \triangleq J_1.$$

Then we have

$$P_{\boldsymbol{\Delta}^*}\left(\sum_{\boldsymbol{c}\ne\boldsymbol{c}^*}\frac{P_{\boldsymbol{\Theta}^*}(\boldsymbol{C} = \boldsymbol{c} \mid \boldsymbol{A})}{P_{\boldsymbol{\Theta}^*}(\boldsymbol{C} = \boldsymbol{c}^* \mid \boldsymbol{A})} \ge \epsilon\right) = P_{\boldsymbol{\Delta}^*}\left(\sum_{\boldsymbol{c}\ne\boldsymbol{c}^*} \mathrm{e}^{J_1} \ge \epsilon\right).$$

Let $\boldsymbol{g} = (g_1,\cdots,g_n)$ with $g_i = k$ if $c_{ik} = 1$, and $\boldsymbol{g}^* = (g_1^*,\cdots,g_n^*)$ with $g_i^* = k$ if $c_{ik}^* = 1$. Then, for any $\epsilon > 0$, when n is sufficiently large, we have

$$P_{\boldsymbol{\Delta}^*}\left(\sum_{\boldsymbol{c}\ne\boldsymbol{c}^*} \mathrm{e}^{J_1} \ge \epsilon\right) = P_{\boldsymbol{\Delta}^*}\left(\sum_{r=1}^{n}\sum_{\|\boldsymbol{g}-\boldsymbol{g}^*\|_0 = r} \mathrm{e}^{J_1} \ge \epsilon\right) \le \sum_{r=1}^{n}\sum_{\|\boldsymbol{g}-\boldsymbol{g}^*\|_0 = r} P_{\boldsymbol{\Delta}^*}\left(J_1 \ge \log\frac{\epsilon}{n^{r+1}(K-1)^r}\right) \le \sum_{r=1}^{n}\sum_{\|\boldsymbol{g}-\boldsymbol{g}^*\|_0 = r} P_{\boldsymbol{\Delta}^*}(J_1 \ge -4r\log n).$$

For any $\boldsymbol{g}$ with $\|\boldsymbol{g}-\boldsymbol{g}^*\|_0 = r > 0$, let $\mathcal{D}_r^*(\boldsymbol{g},\boldsymbol{g}^*) = \left\{(i,j) : i<j \text{ and } \boldsymbol{\Delta}^*_{g_i g_j} \ne \boldsymbol{\Delta}^*_{g_i^* g_j^*}\right\}$. Then,

$$\begin{aligned} P_{\boldsymbol{\Delta}^*}(J_1 \ge -4r\log n) =& P_{\boldsymbol{\Delta}^*}\left(\sum_{i<j}\log\frac{P_{\boldsymbol{\Delta}^*}(a_{ij} \mid \boldsymbol{C} = \boldsymbol{c})}{P_{\boldsymbol{\Delta}^*}(a_{ij} \mid \boldsymbol{C} = \boldsymbol{c}^*)} + \log\frac{P_{\boldsymbol{\alpha}^*}(\boldsymbol{C} = \boldsymbol{c})}{P_{\boldsymbol{\alpha}^*}(\boldsymbol{C} = \boldsymbol{c}^*)} \ge -4r\log n\right) = \\ & P_{\boldsymbol{\Delta}^*}\left(\sum_{\substack{(i,j)\in\\ \mathcal{D}_r^*(\boldsymbol{g},\boldsymbol{g}^*)}}\left(\log\frac{P_{\boldsymbol{\Delta}^*}(a_{ij} \mid \boldsymbol{C} = \boldsymbol{c})}{P_{\boldsymbol{\Delta}^*}(a_{ij} \mid \boldsymbol{C} = \boldsymbol{c}^*)} - E_{\boldsymbol{\Delta}^*}\left(\log\frac{P_{\boldsymbol{\Delta}^*}(a_{ij} \mid \boldsymbol{C} = \boldsymbol{c})}{P_{\boldsymbol{\Delta}^*}(a_{ij} \mid \boldsymbol{C} = \boldsymbol{c}^*)}\right)\right) \ge s\right), \end{aligned}$$

where

$$s = -4r\log n - \sum_{(i,j)\in\mathcal{D}_r^*(\boldsymbol{g},\boldsymbol{g}^*)} E_{\boldsymbol{\Delta}^*}\left(\log\frac{P_{\boldsymbol{\Delta}^*}(a_{ij} \mid \boldsymbol{C} = \boldsymbol{c})}{P_{\boldsymbol{\Delta}^*}(a_{ij} \mid \boldsymbol{C} = \boldsymbol{c}^*)}\right) - \sum_{i=1}^{n}\log\frac{\alpha^*_{g_i}}{\alpha^*_{g_i^*}}.$$

With Assumption 1 and Lemma 2, it follows that for some positive integer B_2, when n is sufficiently large, $\alpha_k \ge B_2(\log n)^{-\gamma/4}$ and $N_k(\boldsymbol{c}^*) \ge B_2 n(\log n)^{-\gamma/4}/2$, it then follows from Proposition B.4 in Ref. [24] that $\left|\mathcal{D}_r^*(\boldsymbol{g},\boldsymbol{g}^*)\right| \ge B_2^2 nr(\log n)^{-\gamma/2}/8$. Then, as $n \to \infty$,

$$\frac{4r\log n}{\left|\mathcal{D}_r^*(\boldsymbol{g},\boldsymbol{g}^*)\right|} \le \frac{2^5(\log n)^{1+\gamma/2}}{B_2^2 n} \to 0$$

and

$$\left|\frac{1}{\left|\mathcal{D}_r^*(\boldsymbol{g},\boldsymbol{g}^*)\right|}\sum_{i=1}^{n}\log\frac{\alpha^*_{g_i}}{\alpha^*_{g_i^*}}\right| \le \frac{r\log\left(B_2^{-1}(\log n)^{\gamma/4}\right)}{\left|\mathcal{D}_r^*(\boldsymbol{g},\boldsymbol{g}^*)\right|} \le \frac{2^3\log\left(B_2^{-1}(\log n)^{\gamma/4}\right)(\log n)^{\gamma/2}}{B_2^2 n} \longrightarrow 0.$$

Note that we have

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203

$$-E_{\Delta^*}\left(\log\frac{P_{\Delta^*}(a_{ij}\mid \boldsymbol{C}=\boldsymbol{c})}{P_{\Delta^*}(a_{ij}\mid \boldsymbol{C}=\boldsymbol{c}^*)}\right)\geq B_3(\log n)^{-\gamma/4}$$

for any $(i,j)\notin \mathcal{D}_r^*(\boldsymbol{g},\boldsymbol{g}^*)$ and some positive constant B_3 from Assumption 3. Therefore, $s\geq 2^{-1}B_3(\log n)^{-\gamma/4}\left|\mathcal{D}_r^*(\boldsymbol{g},\boldsymbol{g}^*)\right|>0$.

It follows from (A2) that $\left|\log(P_{\Delta}(a_{ij}\mid \boldsymbol{C}=\boldsymbol{c}))\right|\leq B_1T_n$ for any Δ and $\boldsymbol{c}$, thus we have $\left|\log\dfrac{P_{\Delta^*}(a_{ij}\mid \boldsymbol{C}=\boldsymbol{c})}{P_{\Delta^*}(a_{ij}\mid \boldsymbol{C}=\boldsymbol{c}^*)}\right|\leq 2B_1T_n$, and

$$P_{\Delta^*}(J_1\geq -4r\log n)\leq \exp\left\{-\frac{s^2/\left|\mathcal{D}_r^*(\boldsymbol{g},\boldsymbol{g}^*)\right|}{2^3B_1^2T_n^2}\right\}\leq \exp\left\{-\frac{B_3^2B_2^2nr(\log n)^{-\gamma}}{2^5B_1^2T_n^2}\right\},$$

following Hoeffding's inequality.

Therefore, let $B_4=\dfrac{B_3^2B_2^2}{2^5B_1^2}$, when n is large enough, we have

$$P_{\Delta^*}\left(\sum_{\boldsymbol{c}\neq\boldsymbol{c}^*}\mathrm{e}^{J_1}\geq\epsilon\right)\leq\sum_{r=1}^{n}\sum_{\|\boldsymbol{g}-\boldsymbol{g}^*\|_0=r}\exp\{-B_4nr(\log n)^{-\gamma}T_n^{-2}\}=\sum_{r=1}^{n}\binom{n}{r}(K-1)^r\exp\{-B_4nr(\log n)^{-\gamma}T_n^{-2}\}=$$
$$[1+(K-1)\exp\{-B_4n(\log n)^{-\gamma}T_n^{-2}\}]^n-1\leq\exp[n(K-1)\exp\{-B_4n(\log n)^{-\gamma}T_n^{-2}\}]-1\leq$$
$$2nK\exp\{-\log nB_4n(\log n)^{-1-\gamma}T_n^{-2}\}=2Kn^{-B_4n(\log n)^{-1-\gamma}T_n^{-2}+1},$$

where the last inequality follows from the fact that $\mathrm{e}^x-1\leq 2x$ when $x>0$ is sufficiently close to 0. The desired result follows immediately from $T_nn^{-1/2}(\log n)^{(1+\gamma)/2}\prec 1$.

Proof of Theorem 3. Proposition 2 shows that

$$P_{\Delta^*}(\tilde{\boldsymbol{c}}\neq\boldsymbol{c}^*)=P_{\Delta^*}(\{\exists\boldsymbol{c}\neq\boldsymbol{c}^*\text{ s.t. }P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}=\boldsymbol{c}^*\mid\boldsymbol{A})<P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}=\boldsymbol{c}\mid\boldsymbol{A})\})\leq P_{\Delta^*}\left(P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}=\boldsymbol{c}^*\mid\boldsymbol{A})<\sum_{\boldsymbol{c}\neq\boldsymbol{c}^*}P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}=\boldsymbol{c}\mid\boldsymbol{A})\right)\leq$$
$$P_{\Delta^*}(P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}\neq\boldsymbol{c}^*\mid\boldsymbol{A})\geq P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}=\boldsymbol{c}^*\mid\boldsymbol{A}))=P_{\Delta^*}(P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}\neq\boldsymbol{c}^*\mid\boldsymbol{A})\geq 1/2)\to 0.$$

Define $\delta_{\boldsymbol{c}^*}(\boldsymbol{c})=\prod_{i\in[n],k\in[K]}\{I(c_{ik}=c_{ik}^*)\}^{c_{ik}}$. It then follows from the definition of $\hat{\boldsymbol{\beta}}(\boldsymbol{\Theta}^*)=\operatorname{argmax}\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Theta}^*)$ that

$$\mathrm{KL}\left(R\left(\boldsymbol{C},\hat{\boldsymbol{\beta}}(\boldsymbol{\Theta}^*)\right),P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}\mid\boldsymbol{A})\right)=\min_{\boldsymbol{\beta}\in\boldsymbol{\Pi}}\mathrm{KL}(R(\boldsymbol{C},\boldsymbol{\beta}),P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}\mid\boldsymbol{A}))\leq\mathrm{KL}(\delta_{\boldsymbol{c}^*}(\boldsymbol{C}),P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}\mid\boldsymbol{A}))=$$
$$\sum_{\boldsymbol{c}}\delta_{\boldsymbol{c}^*}(\boldsymbol{c})\log\frac{\delta_{\boldsymbol{c}^*}(\boldsymbol{c})}{P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}=\boldsymbol{c}\mid\boldsymbol{A})}=-\log P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}=\boldsymbol{c}^*\mid\boldsymbol{A}).$$

Therefore, for any $\epsilon>0$, it holds true that

$$P_{\Delta^*}\left(\mathrm{KL}\left(R\left(\boldsymbol{C},\hat{\boldsymbol{\beta}}(\boldsymbol{\Theta}^*)\right),P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}\mid\boldsymbol{A})\right)\geq\epsilon\right)\leq P_{\Delta^*}(-\log P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}=\boldsymbol{c}^*\mid\boldsymbol{A})\geq\epsilon)=$$
$$P_{\Delta^*}(P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}=\boldsymbol{c}^*\mid\boldsymbol{A})\leq\mathrm{e}^{-\epsilon})=P_{\Delta^*}(P_{\boldsymbol{\Theta}^*}(\boldsymbol{C}\neq\boldsymbol{c}^*\mid\boldsymbol{A})\geq 1-\mathrm{e}^{-\epsilon}).$$

This completes the proof of Theorem 3.

CSTR: 32290.14.JUSTC-2024-0164
DOI: 10.52396/JUSTC-2024-0164
JUSTC, **2026**, 56(2): 0203