Title
Selecting the Number of Communities in Count-Weighted Networks

Author
Liu, Yucheng

Publication Date
2022

Peer reviewed|Thesis/dissertation

Selecting the Number of Communities in Count-Weighted Networks

By

YUCHENG LIU

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Xiaodong Li, Chair

Thomas Lee

Can Le

Committee in Charge

2022

To my family and my love.

Contents

Selecting the Number of Communities in Count-Weighted Networks

Abstract

This dissertation aims to address the problem of estimating the number of communities in count-weighted networks. We propose two methods from very different perspectives.

The first method we propose is a stepwise procedure, referred to as Stepwise Model-Assisted Spectral Thresholding (SMAST), selecting the number of communities in general count-weighted networks, which are common in the real-world. In the m-th step of the procedure, we first cluster the nodes into m groups with a certain spectral clustering method, and estimate the mean parameters of the general degree-corrected stochastic blockmodel (DCSBM). Then the adjacency matrix is normalized by the estimated degree-correction parameters as well as the solution to a small-scale matrix scaling problem. The eigenvalues of the resultant normalized adjacency matrix are then truncated with the threshold $(2+\epsilon)\sqrt{n}$ in magnitude, where ϵ is a prespecified small constant, e.g. 0.05. The procedure continues to the $(m+1)$-th step if the number of remaining significant eigenvalues is greater than m, and stops otherwise. A prominent feature of this method is that it is derived under the general DCSBM, and can be applied to general count-weighted networks. In theory, if SCORE is used for spectral clustering, SMAST is shown to be consistent in estimating the true number of communities under certain assumptions that hold for a broad class of count-weighted networks. An extension of the Nonsplitting Property of SCORE to the general DCSBM, and recent results on spectral radii of inhomogeneous random matrices, play essential roles in the analyses. Extensive numerical experiments on simulated and real networks have also been conducted to demonstrate the empirical effectiveness of SMAST.

The second method is a network rank selection method based on approximate risk estimation. We treat the problem of estimating the number of communities in a network as selecting the rank of the expected adjacency matrix. For each given rank r, the spectral estimate of the expected adjacency matrix is obtained by spectral truncation of the observed

adjacency matrix with rank r. The goodness of rank r is evaluated by quantifying the estimation error of the estimated expected adjacency matrix. By choosing the error measure based on the binomial deviance, following existing work in the literature on risk estimation under the Bernoulli model, we can estimate the optimism of the apparent error over the true error by the divergence of the estimator with respect to the observation. Moreover, the divergence formula of any spectral function can be calculated in closed-form, which provides an estimate of the true error as the apparent error plus the divergence. This framework can be straightforwardly extended to Poisson networks, and thus can be used to solve the problem of rank selection for count-weighted networks. The closed-form formula of the estimated true error is derived in this case as well. We have applied this method to several benchmark networks in the literature and found the performance comparable and even better than other state-of-the-art model-based approaches.

Acknowledgments

To begin with, I would like to express my deepest gratitude towards my advisor Prof. Xiaodong Li for his mentoring and continuous support during my years as a PhD student. Xiaodong is one of the most knowledgeable and responsible persons I have met and he has been a role model as a researcher to me. As an advisor, Xiaodong always provides ample time to spend with his students and is open to discuss various research topics. I have greatly expanded my knowledge base and grown interest in research. Without his effort and time, I could not have accomplished my works and become a more competent PhD student. I have learned from Xiaodong more than I could give him credit for here, and one of the most important lessons I want to mention is to be rigorous and responsible for my works, which I believe would be beneficial to my career and life.

Secondly, I am really greatful to the committee members of my qualifying exam and final exam: Prof. Thomas Lee, Prof. Can Le, Prof. Wolfgang Polonik, Prof. Prabir Burman and Prof. Luis Rademacher. I truly appreciate their kind assistance with the exams, encouragement and valuble feedbacks about my work.

Thirdly, I would like to express my appreciaton of the department of statistics and PhD program. It has been a fruitful 5-year-journey and it is fortunate for me to have experienced it. The department has offered its students the best resource to build foundation in statistics and explore statistical research. My apprecation goes to all the staff in the department, especially Sarah Driver, who has always been helping me with numerous issues patiently and promptly. Next, I would like to extend my sincere thanks to my fellow peer students, Xingmei Lou, Rui Hu, Satarupa Bhattacharjee, Alvaro Gajardo, for the valuable experience we had together. I am also thankful to my seniors, Tongyi Tang, Shitong Wei, Yi-wei Liu, Qin Ding and Junwen Yao for their help and experience shared.

Lastly, I wish to express the love that words cannot express to my family and my girlfriend. They have always been the strongest support in my life and I could not have undertaken this journey without them. To the girl I love, after these years especially the recent difficult ones, I know that we will start a happy life together and accompany each others for our lives.

CHAPTER 1

Introduction

In network analysis, community detection is regarding how to partition the vertices of a graph into clusters with similar connection patterns. This is an important problem with various applications in sociology, biology, and machine learning. The study of relevant models, methods, and theory has also attracted interest from researchers across a variety of fields including statistics, computer science, and physics; see, e.g., the survey of Fortunato [2010]. In order to derive community detection methods and analyze their theoretical properties, stochastic blockmodels (SBM) [Holland et al., 1983] and its extensions such as degree-corrected SBM (DCSBM) [Karrer and Newman, 2011] have been widely used in practice. Based on these stochastic blockmodels, plenty of community detection methods have been derived in the literature, such as likelihood-based methods [e.g., Amini et al., 2013; Bickel and Chen, 2009; Newman and Girvan, 2004; Zhao et al., 2012], spectral methods [e.g., Jin, 2015; Lei and Rinaldo, 2015; Rohe et al., 2011], and convex optimization methods [e.g., Abbe et al., 2015; Chen et al., 2018; Guédon and Vershynin, 2015].

Most proposed community detection methods in the literature require the knowledge of the number of communities. Take spectral community detection as an example, which usually carries out spectral clustering on some matrix derived from the observed network data, such as the adjacency matrix or the graph Laplacian. Clearly, the k-means step in spectral clustering requires the knowledge of the true number of communities, which is usually unknown and needs to be estimated from the network data. A variety of procedures and algorithms for the selection of the number of communities have already been proposed in the literature. Examples include likelihood methods and BIC [Daudin et al., 2008; Hu

et al., 2020; Latouche et al., 2012; Ma et al., 2021; Saldana et al., 2017; Wang and Bickel, 2017], cross-validation [Chen and Lei, 2018; Li et al., 2020], spectral methods [Le and Levina, 2015], goodness-of-fit [Jin et al., 2022], etc.

Note that the aforementioned methods are mostly proposed for unweighted (i.e. binary-edge) networks, while model selection for weighted networks is little-studied. In fact, community detection for weighted networks is important not only because they are common in practice, but also because weighted networks may present more refined community structures than unweighted networks because of more complicated connection patterns. This dissertation is thereby focused on the problem of selecting the number of communities in weighted networks. As a first step, we consider a special type of weighted network: the count-weighted network, which is common in practice. One well-known example is the word-word co-occurrence networks in natural language processing, where each edge-weight represents the number of co-occurrences of that pair of words in some context. In the literature of graph theory, count-weighted networks are also referred to as "multigraphs".

In this dissertation, we are proposing two methods to estimate the number of communities based on the adjacency matrix of a count-weighted network. Chapter 2 and 3 are devoted to explaining the two methods in detail, respectively. Simulation studies on synthetic and real-world networks are postponed to Chapter 4.

1.1. Selecting the Number of Communities in Count-Weighted Networks via Stepwise Model-Assisted Spectral Thresholding

Spectral methods are generally believed to be more robust to distributional assumptions than likelihood methods, but adaptive thresholds for spectral truncation are usually difficult to obtain. The goal of this part of the dissertation is to develop a procedure for selecting the number of communities consistently for a broad class of count-weighted networks. The proposed method is a procedure of stepwise spectral thresholding, in which adaptive thresholds can be obtained with the help of model fitting. We first consider the general

DCSBM with K communities, in which we model the weighted network only in its mean adjacency matrix rather than its whole distribution. In the m-th step of the procedure, where $m = 2, 3, \ldots$ ($m = 1$ is the trivial case) is the candidate number of communities, we first group the nodes into m distinct communities by using some spectral clustering method, e.g. SCORE proposed in Jin [2015]. Then we estimate the DCSBM parameters based on the estimated communities. Next, we aim to find diagonal matrix $\boldsymbol{D}$, such that when $m = K$, we can truncate the spectrum of the normalized adjacency matrix $\widetilde{\boldsymbol{A}} \coloneqq \boldsymbol{D}^{1/2}\boldsymbol{A}\boldsymbol{D}^{1/2}$ with an explicit threshold in order to reveal the true number of communities. The construction of $\boldsymbol{D}$ relies on estimated degree-correction parameters as well as solving a small-scale symmetric matrix scaling problem. The stopping criterion relies on the comparison between m and the remaining eigenvalues of $\widetilde{\boldsymbol{A}}$. The details will be explained in Chapter 2.

The idea of spectral truncation on normalized adjacency matrix is inspired by a recent work Landa et al. [2021], which studies how to reveal the rank of a Poisson data matrix by spectral truncation. The major difference is on how to normalize the data matrix. Applying their approach to count-weighted networks, the diagonal matrix $\boldsymbol{D}$ is constructed by requiring $\boldsymbol{D}\boldsymbol{A}\boldsymbol{D}$ to be doubly stochastic, i.e., solving a symmetric matrix scaling problem. This is indeed a very interesting idea, but also has several drawbacks. First, the existence of such $\boldsymbol{D}$ is not guaranteed. In fact, the classical result in guaranteeing the existence and uniqueness of matrix scaling requires that all the entries of the matrix are strictly positive [Sinkhorn, 1967]. Second, straightforward matrix scaling on $\boldsymbol{A}$ is a $O(n)$-scale problem, which will incur high computational complexity when the network is large. In contrast, the matrix scaling problem in each step of our procedure turns out to be a $O(m)$-scale problem, which is much easier to solve in computation. Of course, our method relies on estimating parameters in the DCSBM, so it involves spectral clustering in each step of the procedure. Once we have the diagonal matrix $\boldsymbol{D}$, the operator norm of $\boldsymbol{D}^{1/2}(\boldsymbol{A} - \mathbb{E}[\boldsymbol{A}])\boldsymbol{D}^{1/2}$ can be tightly and explicitly

controlled by recently established results on the spectral radii of inhomogeneous random matrix in Latała et al. [2018].

We establish the theoretical guarantees of our procedure under the general DCSBM with certain underdispersed and subexponential conditions. Technically speaking, our theoretical analysis is related to recent theoretical developments of stepwise model selection methods for unweighted networks [e.g. Jin et al., 2022; Ma et al., 2021]. In particular, we follow the analytical framework established in Jin et al. [2022] for the stepwise goodness-of-fit (StGoF) method, which avoids the analysis of the over-fitting case. On the other hand, our under-fitting case analysis relies on extending the Nonsplitting Property (NSP) shown in Jin et al. [2022] for unweighted networks to count-weighted networks, in order to reduce the number of possible realizations for the estimated DCSBM parameters, so that the probability union bound can be thereby applied. This is a crucial result proved in this dissertation.

1.2. Spectral Divergence-Based Rank Selection for Network Data

The second method we propose is a network rank selection method based on approximate risk estimation. Rather than imposing specific statistical models on the distribution of the network data, we only assume that it has independent edges. We treat the problem of estimating the number of communities in a network as selecting the rank of the expected adjacency matrix, which is true for various SBM used in the literature, such as the standard SBM, DCSBM, and mixed membership SBM.

For each given rank r, we consider the spectral estimate of the expected adjacency matrix, which is obtained by the spectral truncation of the observed adjacency matrix with the given rank. The goodness of rank r is evaluated by quantifying the estimation risk of the resultant estimate of the expected adjacency matrix. By choosing the error measures based on the binomial deviance, following existing work in the literature of risk estimation under Bernoulli models, we can estimate the optimism of the apparent error over the true error by the divergence of the estimator with respect to the observations. Moreover, this divergence

of spectral function can be calculated in closed-form by following existing results for singular value decomposition (SVD)-based spectral functions [Candès et al., 2013].

The above risk estimation-based approach to network rank selection is first proposed for standard Bernoulli networks with binary edges and symmetric entries. Interestingly, it can also be straightforwardly extended to solving the problem of rank selection for the Poisson network with count edges. We will give the closed-form formula of the estimated true error in this scenario, too.

This method is closely related to model selection, and here we briefly discuss the closely relevant ones. In linear models, for least squares estimate, if the error function is chosen as the squared error, the degree of freedom is determined by the number of features [Mallows, 2000]. This result can be further extended to estimation in parametric models, where both the estimator and error are likelihood-based [Akaike, 1974]. For the normal mean models, by choosing the squared error, the degree of freedom can be estimated by the divergence of general estimators [Stein, 1981], which further yields the famous Stein's unbiased risk estimation (SURE). On the other hand, for Bernoulli models and a broad class of error measures, it was shown in Efron [1986] that the penalty is determined by the sum of covariances between the estimates and the corresponding observations. This idea is also related to cross-validations in Efron et al. [2004]. Moreover, it is shown in Efron et al. [2004] that under Bernoulli models with general mean estimators, if the error measure is chosen as likelihood-based, the covariance penalty can be usually approximated by the divergence of the estimator with respect to the observations. This idea of approximation provides the essential idea to our proposed method of network rank selection.

Another related topic to our proposed method is the differentials of spectral functions. In fact, this is a well-studied area in the literature, see, e.g., Edelman [2005]; Lewis and Sendov [2001]; Papadopoulo and Lourakis [2000]. It is interesting to see that the divergence of SVD-based spectral functions on rectangular matrices has a very neat closed form formula

[Candès et al., 2013; Yuan, 2016]. By following the steps in Candès et al. [2013], where the arguments are based on the previous study of the Jacobians of spectral functions, we provide analogous divergence formulas for eigenvalue decomposition-based spectral functions on symmetric matrices.

A related model to our method is the rectangular data matrix with independent Poisson entries. One example in network analysis is the model proposed in Ball et al. [2011], in which overlapping K communities in a network are defined based on the edges rather than the nodes. In their work, the edges are assumed to have K colors according to the K communities, which are unobserved. For each pair of nodes i and j, the total number of edges between i and j satisfies $A_{ij} \overset{\text{indep}}{\sim} \text{Poisson}(\sum_{k=1}^{K} \theta_{ik}\theta_{jk})$, where θ_{ik} for $i \in [n]$ and $k \in [K]$ are unknown nonnegative parameters. Another example is nonnegative matrix factorization (NMF) proposed in Lee and Seung [1999] with successful applications in image processing and text mining, in which the count entries are assumed to be Poisson random variables, and the objective function proposed is subsequently derived from the Poisson likelihood. Risk estimation for SVD-based spectral estimators under this model has been studied in Bigot et al. [2017]. In fact, our choice of error measures is exactly the same as theirs. The major difference is on how to estimate the risk. An unbiased estimate of the discrepancy between the true and apparent errors was given in their work by using an identity on Poisson distribution identified in Hudson [1978]. However, this unbiased estimate is computationally expensive to calculate directly, so approximation based on the sum over random directional derivatives was proposed in their paper. In contrast, we use the divergence to approximate this discrepancy by following the ideas in Efron et al. [2004], for which a closed-form formula can be analytically obtained. In comparison to our method, the random approximation approach in Bigot et al. [2017] relies sensitively on the choice of the random directional derivative, and is usually more expensive in computation.

CHAPTER 2

Selecting the Number of Communities in Count-Weighted Networks via Stepwise Model-Assisted Spectral Thresholding

2.1. Methodology

Before delving into details, let's introduce some notations used in this chapter. Denote $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. We use $a \lesssim b$ or $b \gtrsim a$ to denote $a \leq Cb$ for a constant C. Denote $a \asymp b$ if $a \lesssim b$ and $b \gtrsim a$. For a vector $\boldsymbol{x} \in \mathbb{R}^n$, x_i and $x(i)$ denote its i-th entry. We use $\|\boldsymbol{x}\|_1$ and $\|\boldsymbol{x}\|_2$ to denote the vector's ℓ^1 and ℓ^2-norm, respectively. For a matrix $\boldsymbol{X} \in \mathbb{R}^{m \times n}$, X_{ij} and $X(i, j)$ denote its (i, j)-th entry. Denote $\lambda_k(\boldsymbol{X})$ as the $k-$th (in descending order of magnitudes) eigenvalue of a symmetric $n \times n$ matrix $\boldsymbol{X}$. Denote $\sigma_k(\boldsymbol{X})$ as the $k-$th (in descending order) singular value of a $n \times m$ matrix $\boldsymbol{X}$. Denote the Schatten p-norm of a symmetric matrix $\boldsymbol{X}$ as

$$\|\boldsymbol{X}\|_{S_p} = \left(\sum_i \left(\sum_j X_{ij}^2 \right)^{p/2} \right)^{1/p} = \left(\sum_i |\lambda_i(\boldsymbol{X})|^p \right)^{1/p},$$

and therefore the operator norm $\|\boldsymbol{X}\| = \|\boldsymbol{X}\|_{S_\infty}$. Denote the l^∞ norm of a random variable ξ as $\|\xi\|_\infty$.

As introduced in Chapter 1, the goal of the present chapter is to propose a generally applicable method of estimating the number of communities in a wide class of count-weighted networks. This section is intended to explain our stepwise procedure in detail. First, let's introduce the general DCSBM for count-weighted networks.

2.1.1. General DCSBM. The standard DCSBM proposed in Karrer and Newman [2011] can be extended to the general DCSBM for count-weighted networks by only modeling the mean structure, rather than the whole distribution of the network.

To be specific, let $\boldsymbol{A}$ be the symmetric adjacency matrix of the count-weighted network with n nodes, which belong to K separate communities. Denote by $\mathcal{N}_1, \ldots, \mathcal{N}_K$ the underlying communities, with respective cardinalities $n_1, \ldots, n_K$ and hence $n = n_1 + \cdots + n_K$. Denote by $\phi : [n] \to [K]$ the community membership function of nodes, such that $\phi(i) = k$ if and only if node i belongs to the community $\mathcal{N}_k$. We also use the vector $\boldsymbol{\pi}_i \in \mathbb{R}^K$ to represent the community belonging of node i by letting $\pi_i(k) = 1$ if $i \in \mathcal{N}_k$ and $\pi_i(k) = 0$ otherwise. Denote $\boldsymbol{\Pi} = [\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_n]^\top \in \mathbb{R}^{n \times K}$ as the community membership matrix.

Our general DCSBM only models the mean of the adjacency matrix $\boldsymbol{A}$, denoted as $\boldsymbol{M} \coloneqq \mathbb{E}[\boldsymbol{A}]$, to incorporate both the community structure and the heterogeneity of degrees. Note that here we allow the network to have self-loops for simplicity of analysis without loss of generality. Concretely, let $\boldsymbol{B}$ be a $K \times K$ symmetric matrix with positive entries and diagonal entries $B_{kk} = 1$ for $k = 1, \cdots, K$ for the sake of identifiability. Further, let $\theta_1, \ldots, \theta_n > 0$ be the degree-correction parameters. Denote $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^\top$ and $\boldsymbol{\Theta} = \operatorname{diag}(\theta_1, \ldots, \theta_n)$. Then the entries of the expected adjacency matrix are parameterized as

$$M_{ij} = \theta_i \theta_j B_{\phi(i)\phi(j)} = \theta_i \theta_j \boldsymbol{\pi}_i^\top \boldsymbol{B} \boldsymbol{\pi}_j, \quad 1 \le i \le j \le n. \tag{2.1}$$

In matrix form, the expected adjacency matrix can be represented as

$$\boldsymbol{M} = \boldsymbol{\Theta} \boldsymbol{\Pi} \boldsymbol{B} \boldsymbol{\Pi}^\top \boldsymbol{\Theta}. \tag{2.2}$$

A particular assumption made in our general DCSBM is the following "underdispersed" assumption:

$$\operatorname{Var}(A_{ij}) \le \mathbb{E}[A_{ij}] \quad 1 \le i \le j \le n.$$

This assumption holds for basic count distributions including the binomial and Poisson distributions.

2.1.2. Stepwise Clustering and DCSBM Parameter Estimation. Our method to estimate the number of communities is a stepwise method that relies on fitting the general DSCBM with the candidate number of communities $m = 1, 2, 3, \ldots$. In other words, for each $m \geq 2$ ($m = 1$ is the trivial case), we apply some standard community detection method to obtain m estimated communities, and then estimate the model parameters $\boldsymbol{\theta}$ and $\boldsymbol{B}$.

To be specific, suppose we apply some spectral clustering method, such as spectral clustering on ratios-of-eigenvectors (SCORE) given in Jin [2015] or regularized spectral clustering (RSC) in Joseph and Yu [2016], to obtain m estimated communities $\widehat{\mathcal{N}}_1^{(m)}, \ldots, \widehat{\mathcal{N}}_m^{(m)}$. Then, the DCSBM parameters, i.e. $\boldsymbol{\theta}$ and $\boldsymbol{B}$, can be directly estimated by pretending the estimated communities are the true communities. Here we review the formulas of such estimates with the notations given in Jin et al. [2022].

Let's first see how to represent $\boldsymbol{\theta}$ and $\boldsymbol{B}$ by the population adjacency matrix $\boldsymbol{M}$, the true communities $\mathcal{N}_1, \ldots, \mathcal{N}_K$, and the population degrees. Decompose $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = \boldsymbol{\theta}_1 + \cdots + \boldsymbol{\theta}_K$, where $\boldsymbol{\theta}_k \in \mathbb{R}^n$ for $k = 1, \ldots, K$ such that $\boldsymbol{\theta}_k(i) = \theta_i$ if $i \in \mathcal{N}_k$ and $\boldsymbol{\theta}_k(i) = 0$ otherwise. We can similarly decompose the n-dimensional all-one vector into

$$\mathbf{1}_n = \mathbf{1}_1 + \cdots + \mathbf{1}_K, \tag{2.3}$$

such that $\mathbf{1}_k(j) = 1$ if $j \in \mathcal{N}_k$ while $\mathbf{1}_k(j) = 0$ otherwise. It is easy to verify that for $1 \leq k, l \leq K$,

$$\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_l = B_{kl} \|\boldsymbol{\theta}_k\|_1 \|\boldsymbol{\theta}_l\|_1.$$

Since we assume that $B_{kk} = 1$ for $k = 1, \ldots, K$, the above equality implies that

$$\|\boldsymbol{\theta}_k\|_1 = \sqrt{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_k},$$

which further gives

$$B_{kl} = \frac{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_l}{\sqrt{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_k}\sqrt{\mathbf{1}_l^\top \boldsymbol{M} \mathbf{1}_l}}. \tag{2.4}$$

Denoting the degrees of the network as $d_i = \sum_{j=1}^n A_{ij}$, $i = 1, \cdots, n$. Then, for $i \in \mathcal{N}_k$, the population degree of node i is

$$\begin{aligned} d_i^* \coloneqq \mathbb{E}[d_i] &= \theta_i \left(B_{k1}\|\boldsymbol{\theta}_1\|_1 + B_{k2}\|\boldsymbol{\theta}_2\|_1 + \cdots + B_{kK}\|\boldsymbol{\theta}_K\|_1\right) \\ &= \theta_i \left(\frac{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_1}{\sqrt{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_k}} + \cdots + \frac{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_K}{\sqrt{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_k}} \right) = \theta_i \frac{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_n}{\sqrt{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_k}}. \end{aligned}$$

This implies that the degree-correction parameter θ_i can be expressed as

$$\theta_i = \frac{\sqrt{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_k}}{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_n} d_i^*. \tag{2.5}$$

With (2.4) and (2.5), we can obtain plug-in estimates of $\boldsymbol{B}$ and $\boldsymbol{\theta}$ by replacing the true communities $\mathcal{N}_1, \ldots, \mathcal{N}_K$ with the estimated communities $\widehat{\mathcal{N}}_1^{(m)}, \ldots, \widehat{\mathcal{N}}_m^{(m)}$, replacing $\boldsymbol{M}$ with $\boldsymbol{A}$, and replacing d_i^* with d_i. In analogy to the decomposition (2.3), we decompose the all-one vector to the sum of indicator vectors corresponding to the estimated communities: $\mathbf{1}_n = \hat{\mathbf{1}}_1^{(m)} + \cdots + \hat{\mathbf{1}}_m^{(m)}$, where for each $j = 1, \ldots, n$ and $k = 1, \ldots, m$, $\hat{\mathbf{1}}_k^{(m)}(j) = 1$ if $j \in \widehat{\mathcal{N}}_k^{(m)}$ and $\hat{\mathbf{1}}_k^{(m)}(j) = 0$ otherwise. Then the plug-in estimates are

$$\hat{\theta}_i^{(m)} \coloneqq \frac{\sqrt{(\hat{\mathbf{1}}_k^{(m)})^\top \boldsymbol{A} \hat{\mathbf{1}}_k^{(m)}}}{(\hat{\mathbf{1}}_k^{(m)})^\top \boldsymbol{A} \mathbf{1}_n} d_i, \quad \text{for } k = 1, \ldots, m \text{ and } i \in \widehat{\mathcal{N}}_k^{(m)}, \tag{2.6}$$

and

$$\widehat{B}_{kl}^{(m)} \coloneqq \frac{(\hat{\mathbf{1}}_k^{(m)})^\top \boldsymbol{A} \hat{\mathbf{1}}_l^{(m)}}{\sqrt{(\hat{\mathbf{1}}_k^{(m)})^\top \boldsymbol{A} \hat{\mathbf{1}}_k^{(m)}}\sqrt{(\hat{\mathbf{1}}_l^{(m)})^\top \boldsymbol{A} \hat{\mathbf{1}}_l^{(m)}}}, \quad \text{for } 1 \le k, l \le m. \tag{2.7}$$

Let's come back to spectral clustering. In our theory, we consider SCORE in particular. In SCORE, we first compute the m leading eigenvectors of $\boldsymbol{A}$: $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ corresponding to the m largest eigenvalues in magnitude. Then we construct an $n \times (m-1)$ matrix of entrywise ratios $\boldsymbol{R}^{(m)}$: $R^{(m)}(i,k) = u_{k+1}(i)/u_1(i)$ for $1 \le i \le n$ and $0 \le k \le m-1$. Finally, the rows of the ratio matrix $\boldsymbol{R}^{(m)}$ are clustered by the k-means algorithm assuming there are m clusters.

In the null case ($m = K$), the consistency of SCORE for unweighted networks has been well-studied in Jin [2015]; Jin et al. [2017]. More importantly, in the under-fitting case ($m < K$), it has been proved in Jin et al. [2022] that SCORE enjoys the Nonsplitting Property (NSP), i.e., the true communities are refinements of the estimated communities, for unweighted networks under some mild conditions. This property is essential for the theoretical analysis of the under-fitting case, and we will show in this article that SCORE enjoys NSP for a wide class of count-weighted networks under the general DCSBM.

It is noteworthy that we may replace SCORE with other spectral clustering methods. In fact, the regularized spectral clustering (RSC) proposed in Joseph and Yu [2016] has been shown to enjoy certain NSP in Ma et al. [2021] for unweighted networks. Extension of such result to count-weighted networks can be left as future work.

2.1.3. Matrix Scaling and Spectral Thresholding. We first explain the heuristic of matrix scaling and spectral thresholding in our method by assuming $\boldsymbol{\theta}$, $\boldsymbol{B}$ and the labeling function $\phi(\cdot)$ are known.[1] The key idea is to normalize $\boldsymbol{A}$ such that the true number of communities K can be uncovered through spectral thresholding on the normalized adjacency matrix. To be specific, for a diagonal matrix $\boldsymbol{\Xi} = \mathrm{diag}(\xi_1, \ldots, \xi_n) = \mathrm{diag}(\boldsymbol{\xi})$ whose diagonal entries are all positive, we normalize the adjacency matrix and decompose it into

$$\boldsymbol{\Xi}^{\frac{1}{2}}\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{\Xi}^{\frac{1}{2}} = \underbrace{\boldsymbol{\Xi}^{\frac{1}{2}}\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{M}\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{\Xi}^{\frac{1}{2}}}_{\boldsymbol{L}} + \underbrace{\boldsymbol{\Xi}^{\frac{1}{2}}\boldsymbol{\Theta}^{-\frac{1}{2}}(\boldsymbol{A}-\boldsymbol{M})\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{\Xi}^{\frac{1}{2}}}_{\boldsymbol{E}}.$$

[1] This actually implies that K is known, but of course we don't use this information explicitly in the heuristic of spectral thresholding.

We refer to $\boldsymbol{\xi}$ as the scaling factors. Note that $\text{rank}(\boldsymbol{L}) = \text{rank}(\boldsymbol{M}) = K$. Once we can tightly control the operator norm of $\boldsymbol{E}$, the rank of $\boldsymbol{L}$ is likely to be uncovered by a spectral thresholding on $\boldsymbol{\Xi}^{\frac{1}{2}}\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{\Xi}^{\frac{1}{2}}$. Therefore, the question is how to choose $\boldsymbol{\xi}$ in order to control the norm of $\boldsymbol{E}$ tightly and explicitly.

In order to control the norm of $\boldsymbol{E}$, noting that it has centered independent above diagonal entries, we plan to choose $\boldsymbol{\xi}$ such that the row sums of variances in $\boldsymbol{E}$ are all upper bounded by n. Then, by recently established results on the spectral radii of inhomogeneous random graphs, e.g. Theorem 4.8 in Latała et al. [2018], it is likely to have the bound $\|\boldsymbol{E}\| \leq (2+\epsilon)\sqrt{n}$, where ϵ is a small positive constant such as 0.05. Since $E_{ij} = (A_{ij} - M_{ij})\xi_i^{\frac{1}{2}}\xi_j^{\frac{1}{2}}\theta_i^{-\frac{1}{2}}\theta_j^{-\frac{1}{2}}$, the underdispersed assumption implies

$$\text{Var}(E_{ij}) \leq M_{ij}\xi_i\xi_j\theta_i^{-1}\theta_j^{-1} = (B_{\phi(i)\phi(j)}\theta_i\theta_j)\xi_i\xi_j\theta_i^{-1}\theta_j^{-1} = B_{\phi(i)\phi(j)}\xi_i\xi_j.$$

Then, we can simply require $\boldsymbol{\xi}$ to satisfy

$$\sum_{j=1}^{n} B_{\phi(i)\phi(j)}\xi_i\xi_j = n \quad \text{for } i = 1, \ldots, n. \tag{2.8}$$

In fact, the quadratic equations (2.8) is known to be the symmetric version of matrix scaling problem in the literature up to a global factor [Knight et al., 2014; Sinkhorn, 1967]. In particular, since all entries of $\left[B_{\phi(i)\phi(j)}\right]_{i,j=1}^{n}$ are positive, the existence and uniqueness for such positive $\boldsymbol{\xi}$ are guaranteed by Sinkhorn [1967].

In practice, without the knowledge of the true node labels and the true model parameters $\boldsymbol{\theta}$ and $\boldsymbol{B}$, the above heuristic leads to plug-in estimates of the population scaling factors $\boldsymbol{\xi}$. For each candidate number of communities m, we use some spectral clustering method, e.g. SCORE, to obtain the estimated communities $\widehat{\mathcal{N}}_1^{(m)}, \ldots, \widehat{\mathcal{N}}_m^{(m)}$, as well as the estimates $\hat{\boldsymbol{\theta}}^{(m)}$ and $\widehat{\boldsymbol{B}}^{(m)}$ in (2.6) and (2.7). Then, we define the estimated labeling mapping $\hat{\phi} : [n] \to [m]$ such that $\hat{\phi}(i) = k$ if and only if $i \in \widehat{\mathcal{N}}_k^{(m)}$. Then the sample scaling factors $\hat{\boldsymbol{\xi}}$ can be computed

in analogy to (2.8):

$$\sum_{j=1}^{n} \widehat{B}^{(m)}_{\hat{\phi}(i)\hat{\phi}(j)} \hat{\xi}^{(m)}_i \hat{\xi}^{(m)}_j = n \quad \text{for } i = 1, \ldots, n. \tag{2.9}$$

Finally, with the estimated degree-correction parameters $\hat{\boldsymbol{\theta}}^{(m)}$ and estimated scaling factors $\hat{\boldsymbol{\xi}}^{(m)}$, we calculate a normalized adjacency matrix: $\widetilde{\boldsymbol{A}}^{(m)} = \widehat{\boldsymbol{\Xi}}^{\frac{1}{2}} \widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}} \boldsymbol{A} \widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}} \widehat{\boldsymbol{\Xi}}^{\frac{1}{2}}$ where $\widehat{\boldsymbol{\Xi}}^{(m)} = \mathrm{diag}(\hat{\boldsymbol{\xi}}^{(m)})$, i.e.

$$\widetilde{A}^{(m)}_{ij} := \sqrt{\frac{\hat{\xi}^{(m)}_i \hat{\xi}^{(m)}_j}{\hat{\theta}^{(m)}_i \hat{\theta}^{(m)}_j}} A_{ij}. \tag{2.10}$$

Then we count the number of eigenvalues of $\widetilde{\boldsymbol{A}}^{(m)}$ that are greater than $(2+\epsilon)\sqrt{n}$ in magnitude. Based on our heuristic, in the case $m = K$, the number of such eigenvalues is likely to be less than or equal to m. Our stopping rule is the following: If we have at least $m+1$ eigenvalues with absolute values greater than $(2+\epsilon)\sqrt{n}$, m is determined to be less than K, and the stepwise procedure continues with candidate number of communities $m+1$; otherwise, we choose $\widehat{K} = m$, and stop the stepwise procedure.

Remark 1. Let's discuss briefly the computational issues of the matrix scaling step (2.9). At first glance, its computational cost seems to be high since the size of the problem is $O(n)$. It turns out that this scaling problem can be reduced to a much simpler weighted scaling problem as follows:

$$\sum_{l=1}^{m} \left| \widehat{\mathcal{N}}^{(m)}_l \right| \widehat{B}^{(m)}_{kl} \hat{a}^{(m)}_k \hat{a}^{(m)}_l = n \quad \text{for } k = 1, \ldots, m, \tag{2.11}$$

where $\hat{\xi}^{(m)}_i = \hat{a}^{(m)}_k$ if $i \in \widehat{\mathcal{N}}^{(m)}_k$ for $k = 1, \ldots, m$. Therefore, the size of the scaling problem is $O(m)$ instead of $O(n)$, which is much smaller in scale. We derive an algorithm to solve the weighted scaling problem (2.11) by following the ideas in Knight et al. [2014], and the details will be elaborated on in Chapter 4.

2.1.4. Summary of Stepwise Model-Assisted Spectral Thresholding (SMAST). Now we are ready to summarize our stepwise procedure to estimate the number of communities in count-weighted networks. For each $m = 1, 2, 3, \ldots$, our procedure implements the following steps:

(1) **Stepwise Clustering and DCSBM Parameters Estimation**: If $m = 1$, let $\widehat{\mathcal{N}}_1^{(1)} = \{1, \cdots, n\}$ be the only estimated community. If $m > 1$, implement some spectral clustering method, such as SCORE, to the adjacency matrix with the candidate number of communities equal to m. Denote the resulting clusters as $\widehat{\mathcal{N}}_1^{(m)}, \ldots, \widehat{\mathcal{N}}_m^{(m)}$. Fit the general DCSBM and obtain the estimates $\hat{\theta}_i^{(m)}$ for $i = 1, \ldots, n$ by (2.6) and $\widehat{B}_{kl}^{(m)}$ for $1 \le k, l \le m$ by (2.7). Denote $\widehat{\boldsymbol{\Theta}}^{(m)} = \mathrm{diag}(\hat{\theta}_1^{(m)}, \ldots, \hat{\theta}_n^{(m)})$.

(2) **Matrix Scaling**: Obtain the scaling factors $\hat{\boldsymbol{\xi}}^{(m)} \in \mathbb{R}^n$ as the solution to (2.9). Note that this can be achieved by solving a small-scale scaling problem (2.11).

(3) **Spectral Thresholding**: Normalize the adjacency matrix $\boldsymbol{A}$ to $\widetilde{\boldsymbol{A}}^{(m)}$ as in (2.10). Let the eigenvalues of $\widetilde{\boldsymbol{A}}^{(m)}$ be $\lambda_1(\widetilde{\boldsymbol{A}}^{(m)})$, ..., $\lambda_n(\widetilde{\boldsymbol{A}}^{(m)})$ such that $|\lambda_1(\widetilde{\boldsymbol{A}}^{(m)})| \ge \cdots \ge |\lambda_n(\widetilde{\boldsymbol{A}}^{(m)})|$. If $|\lambda_{m+1}(\widetilde{\boldsymbol{A}}^{(m)})| > (2+\epsilon)\sqrt{n}$ for some prespecified small constant ϵ, we continue the procedure with the candidate number of communities $m+1$; otherwise, we stop the iteration and obtain the estimated number of communities $\widehat{K} = m$.

2.2. Main Results

In this section, we will characterize the conditions for the general DCSBM described above, under which SMAST is guaranteed to be consistent in selecting the number of communities. Obviously, the consistency results consist of two parts: (1) the under-fitting case ($m < K$) and (2) the null case ($m = K$). In the underfitting case, we need to show that $|\lambda_{m+1}(\widetilde{\boldsymbol{A}}^{(m)})| > (2+\epsilon)\sqrt{n}$, so that there are at least $m+1$ eigenvalues of $\widetilde{\boldsymbol{A}}^{(m)}$ remained after the spectral thresholding step. In contrast, in the null case, we intend to show that $|\lambda_{m+1}(\widetilde{\boldsymbol{A}}^{(m)})| \le (2+\epsilon)\sqrt{n}$, so that there are at most m significant eigenvalues after spectral thresholding.

Let's first explain the intuition and rationale behind both the under-fitting and null cases. Without confusion in the context, we omit the superscript (m) in the estimates. Notice that the equality (2.10) implies that

$$\begin{aligned}
\widetilde{\boldsymbol{A}} &= \widehat{\boldsymbol{\Xi}}^{\frac{1}{2}}\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\boldsymbol{A}\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}} \\
&= \widehat{\boldsymbol{\Xi}}^{\frac{1}{2}}\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\boldsymbol{M}\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}} + \widehat{\boldsymbol{\Xi}}^{\frac{1}{2}}\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}(\boldsymbol{A}-\boldsymbol{M})\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}} \\
&= \widehat{\boldsymbol{\Xi}}^{\frac{1}{2}}\left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\boldsymbol{\Theta}^{\frac{1}{2}}\right)\left(\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{M}\boldsymbol{\Theta}^{-\frac{1}{2}}\right)\left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\boldsymbol{\Theta}^{\frac{1}{2}}\right)\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}} \\
&\quad + \left(\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}}\boldsymbol{\Xi}^{-\frac{1}{2}}\right)\left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\boldsymbol{\Theta}^{\frac{1}{2}}\right)\underbrace{\left(\boldsymbol{\Xi}^{\frac{1}{2}}\boldsymbol{\Theta}^{-\frac{1}{2}}(\boldsymbol{A}-\boldsymbol{M})\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{\Xi}^{\frac{1}{2}}\right)}_{\boldsymbol{E}}\left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\boldsymbol{\Theta}^{\frac{1}{2}}\right)\left(\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}}\boldsymbol{\Xi}^{-\frac{1}{2}}\right) \\
&:= \widetilde{\boldsymbol{A}}_1 + \widetilde{\boldsymbol{A}}_2. \qquad (2.12)
\end{aligned}$$

The above decomposition enables us to undertake analyses for both the null and under-fitting cases. In the under-fitting case, we have $m+1 \le K$, and $|\lambda_{m+1}(\widetilde{\boldsymbol{A}})|$ is thereby lower bounded by

$$\left|\lambda_{m+1}(\widetilde{\boldsymbol{A}})\right| \ge \left|\lambda_K(\widetilde{\boldsymbol{A}})\right| \ge \left|\lambda_K(\widetilde{\boldsymbol{A}}_1)\right| - \left\|\widetilde{\boldsymbol{A}}_2\right\|$$

where the last inequality follows from Weyl's inequality. We will show that under mild conditions of the underdispersed DCSBM, we have $\left|\lambda_K(\widetilde{\boldsymbol{A}}_1)\right| \gg \sqrt{n}$ and $\left\|\widetilde{\boldsymbol{A}}_2\right\| \lesssim \sqrt{n}$, which guarantees $\left|\lambda_{m+1}(\widetilde{\boldsymbol{A}})\right| > (2+\epsilon)\sqrt{n}$.

On the other hand, for the null case $m = K$, we intend to show $\left|\lambda_{K+1}(\widetilde{\boldsymbol{A}})\right| \le (2+\epsilon)\sqrt{n}$. It is clear that $\operatorname{rank}(\widetilde{\boldsymbol{A}}_1) = K$, which implies $\left|\lambda_{K+1}(\widetilde{\boldsymbol{A}}_1)\right| = 0$. Moreover, we will show that, under certain mild conditions for the underdispersed DCSBM, the matrix scaling introduced in Section 2.1.3 yields the upper bound $\left\|\widetilde{\boldsymbol{A}}_2\right\| \le (2+\epsilon)\sqrt{n}$ with high probability. The Weyl's inequality then gives $\left|\lambda_{K+1}(\widetilde{\boldsymbol{A}})\right| \le \left|\lambda_{K+1}(\widetilde{\boldsymbol{A}}_1)\right| + \left\|\widetilde{\boldsymbol{A}}_2\right\| \le (2+\epsilon)\sqrt{n}$.

In the following, we will introduce our main results that summarize the above intuitions. To this end, we first introduce a sequence of assumptions on the general DCSBM.

2.2.1. Assumptions. The assumptions imposed on the general DCSBM are captured by two constants c_0 and C_0. In particular, we use a unified c_0 to capture various lower bounds.

Assumption 1. We consider the general DCSBM described in Section 2.1.1 satisfying the following conditions

- [Entrywise positivity] The $K \times K$ matrix $\boldsymbol{B}$ is fixed, and its entries satisfy

$$B_{kk} = 1 \text{ for } k = 1, \dots, K, \quad c_0 \le B_{kl} \le 1 \text{ for } 1 \le k, l \le K. \tag{2.13}$$

- [Spectrum of $\boldsymbol{B}$] The eigenvalues of $\boldsymbol{B}$ are assumed to satisfy

$$\lambda_1(\boldsymbol{B}) > |\lambda_2(\boldsymbol{B})| \ge \cdots \ge |\lambda_K(\boldsymbol{B})| \ge c_0 > 0. \tag{2.14}$$

- [Balancedness] Denote $\theta_{\max} = \max\{\theta_1, \dots, \theta_n\}$, and $\theta_{\min} = \min\{\theta_1, \dots, \theta_n\}$. The following balancedness assumptions are satisfied:

$$\min_{1 \le k \le K} \frac{n_k}{n} \ge c_0 \quad \text{and} \quad \frac{\theta_{\min}}{\theta_{\max}} \ge c_0. \tag{2.15}$$

- [Sparseness] $\theta_{\min}$ is bounded by

$$\frac{1}{c_0} \ge \theta_{\max} \ge \theta_{\min} \ge C_0 \sqrt{\frac{\log^4 n}{n}}. \tag{2.16}$$

- [Underdispersion]

$$\mathrm{Var}(A_{ij}) \le M_{ij} \quad 1 \le i \le j \le n. \tag{2.17}$$

- [Bernstein condition] For any $i \le j$ and any integer $p \ge 2$, there holds

$$\mathbb{E}[|A_{ij} - M_{ij}|^p] \le C' \left(\frac{p!}{2}\right) R^{p-2} M_{ij}, \tag{2.18}$$

where C' and R are constants only depending on c_0.

Let's explain some intuitions and implications of the above assumptions. Assumption (2.13) means that the components of $\boldsymbol{B}$ are positive and on the order of $O(1)$, so the sparsity

of the network will be characterized by the degree correction parameters $\boldsymbol{\theta}$. As with Jin et al. [2017] and Jin et al. [2022], we assume the diagonal entries of $\boldsymbol{B}$ to be 1 for the sake of identifiability. Assumption (2.14) means that $\boldsymbol{B}$ is nondegenerate. Here it is worth highlighting that Assumption (2.13) implies that there is actually an explicit eigengap between the first and second largest eigenvalues of $\boldsymbol{B}$ (in magnitude), as shown in Lemma 2.4.2 introduced later. Assumption (2.15) means that the DCSBM is balanced to some extent in both the community sizes and degree correction parameters. Assumption (2.16) characterizes the sparsity of the network, so that the average degree of the nodes can be as low as $O(\log^4 n)$. The upper bound on $\theta_{\max}$ implies that $M_{ij} \leq 1/c_0^2$ for all i and j. Assumption (2.17) is the underdispersion condition introduced before, which holds for binomial and Poisson distributions. Assumption (2.18) is a particular Bernstein condition, which also holds for some distributions such as the Poisson distribution. In fact, for Poisson distribution we have the following lemma:

LEMMA 2.2.1. *Let $X \sim$ Poisson(λ), such that $\lambda \leq C(c_0)$ where $C(c_0)$ is a constant only depending on $c_0 > 0$. Then, for any integer $p \geq 2$, there holds*

$$\mathbb{E}[|X - \lambda|^p] \leq \left(\frac{p!}{2}\right) R(c_0)^{p-2}\lambda,$$

where $R(c_0)$ is a constant only depending on c_0.

PROOF. Since Poisson random variable is known to be discrete log-concave, by noting that $\mathrm{Var}(X) = \lambda$, this lemma can be directly obtained by Lemma 7.5 and Definition 1.2 of Schudy and Sviridenko [2011]. □

2.2.2. Consistency of SMAST. We now introduce our main results that guarantee the consistency of SMAST introduced in Section 2.1 in selecting the number of communities, provided Assumption 1 holds. Following the aforementioned outline, our main results consist of two parts: one for the null case and the other for the under-fitting case.

THEOREM 2.2.1 (Null Case). *Consider the general DCSBM such that Assumption 1 holds. For any fixed c_0, there exists a sufficiently large C_0, such that with probability $1 - O(n^{-3})$, the following event holds: Implementing SMAST with SCORE and candidate number of communities $m = K$, the obtained the normalized adjacency matrix $\widetilde{\boldsymbol{A}}^{(K)}$ defined in (2.10) satisfies $|\lambda_{K+1}(\widetilde{\boldsymbol{A}}^{(K)})| \leq 2.05\sqrt{n}$.*

THEOREM 2.2.2 (Under-Fitting Case). *Consider the general DCSBM such that Assumption 1 holds. For any fixed c_0, there exists a sufficiently large C_0, such that with probability $1 - O(n^{-3})$, the following event holds: For any candidate number of communities $1 \leq m < K$, implementing SMAST with SCORE, the obtained normalized adjacency matrix $\widetilde{\boldsymbol{A}}^{(m)}$ defined in (2.10) satisfies $|\lambda_{m+1}(\widetilde{\boldsymbol{A}}^{(m)})| > 2.05\sqrt{n}$.*

Obviously, combining Theorems 2.2.1 and 2.2.2 shows the consistency of SMAST in estimating K, the number of true communities. In Assumption (2.16), the sparsity assumption $\theta_{\min} \geq C_0\sqrt{\frac{\log^4 n}{n}}$ for might be suboptimal. In contrast, the corresponding assumption is typically $\theta_{\min} \geq C_0\sqrt{\frac{\log n}{n}}$ in the literature of unweighted network model selection [e.g. Jin et al., 2022]. The extra logarithm for the general DCSBM is majorly due to an application of the truncation technique in the proof. In fact, existing tight bounds for spectral radii of inhomogeneous random graphs recently established in the literature, e.g. Latała et al. [2018], require the entries in the random matrix to be uniformly bounded. This is why a truncation technique needs to be employed. It is an interesting open question to improve the logarithm factors for the general DCSBM in the future, but this may rely on improving the state-of-the-art tight bounds for spectral radii of inhomogeneous random graphs.

2.2.3. Nonsplitting Property. Our analysis relies critically on the properties of the estimates $\hat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{B}}$ and $\hat{\boldsymbol{\xi}}$ defined in (2.6), (2.7) and (2.9), respectively. In the null case, we can prove that SCORE is able to achieve exact recovery of the true communities with high probability, thus the concentration of $\hat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{B}}$ and $\hat{\boldsymbol{\xi}}$ around their population counterparts can

be established. The under-fitting case of $m < K$ is more difficult. As explained before, the Nonsplitting Property established in [Jin et al., 2022; Ma et al., 2021] is crucial in reducing the number of possible realizations of the estimates $\hat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{B}}$, so that the probability union bound can be applied.

Definition 1 (Nonsplitting Property [Jin et al., 2022]). Let the ground truth of communities in a network be $\mathcal{N}_1, \ldots, \mathcal{N}_K$. A stepwise community detection method is said to satisfy the Nonsplitting Property (NSP), if for each candidate number of communities $m = 1, \ldots, K$, the true communities $\mathcal{N}_1, \ldots, \mathcal{N}_K$ is a refinement of the estimated communities $\widehat{\mathcal{N}}_1^{(m)}, \ldots, \widehat{\mathcal{N}}_m^{(m)}$, i.e., for any $k = 1, \ldots, K$, there is exactly one $l = 1, \ldots, m$, such that $\mathcal{N}_k \cap \widehat{\mathcal{N}}_l^{(m)} \neq \emptyset$.

Intuitively speaking, the NSP implies that in the under-fitting case, each estimated community of $\widehat{\mathcal{N}}_1^{(m)}, \ldots, \widehat{\mathcal{N}}_m^{(m)}$ is a union of several true communities of $\mathcal{N}_1, \ldots, \mathcal{N}_K$. NSP of SCORE has been established in [Jin et al., 2022] for unweighted networks under standard DCSBM, and here we extend it to the general DCSBM for count-weighted networks.

Lemma 2.2.2. *Consider the general DCSBM satisfying Assumption 1. For any fixed c_0, there exists a sufficiently large C_0, such that with probability $1 - O(n^{-3})$, SCORE satisfies the NSP.*

The proof of this lemma is similar to the one established in Jin et al. [2022] for unweighted networks. A key component of the proof is the row-wise bounds of eigenvector perturbations for adjacency matrices from count-weighted networks. To make this extension, we follow the ideas in Jin et al. [2017], in which such row-wise bounds for unweighted networks are proved, and a crucial tool is a generic row-wise eigenvector perturbation bound given in Abbe et al. [2020]. The major difference is that we need to use the vector and matrix Bernstein inequalities with Bernstein condition (2.18). We will give a self-contained proof in Section 2.4.

2.3. Proofs of the Main Results

2.3.1. Preliminaries. We first cite two results in Landa [2020] for the sensitivity analysis of matrix scaling. Here we only state the results for symmetric matrix scaling with row sums equal to n.

LEMMA 2.3.1 (Lemma 2 in Landa [2020]). *Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix with positive entries. Then, there exists a unique positive vector $\boldsymbol{x} \in \mathbb{R}^n$ satisfying*

$$x_i \left(\sum_{j=1}^{n} A_{ij} x_j \right) = n \quad \text{for all } i = 1, \ldots, n.$$

Furthermore, denote $a_{\max} = \max_{1 \le i \le j \le n} A_{ij}$ and $a_{\min} = \min_{1 \le i \le j \le n} A_{ij}$. Then

$$\frac{\sqrt{a_{\min}}}{a_{\max}} \le x_i \le \frac{\sqrt{a_{\max}}}{a_{\min}}, \quad i = 1, \ldots, n.$$

LEMMA 2.3.2 (Lemma 9 in Landa [2020]). *Let $\widetilde{\boldsymbol{A}}$ be a symmetric matrix with positive entries. Denote $\tilde{a}_{\max} = \max_{1 \le i \le j \le n} \widetilde{A}_{ij}$ and $\tilde{a}_{\min} = \min_{1 \le i \le j \le n} \widetilde{A}_{ij}$. Suppose there is a constant $\epsilon \in (0, 1)$ and a positive vector $\boldsymbol{x} \in \mathbb{R}^n$, such that*

$$\left| \sum_{j=1}^{n} x_i \widetilde{A}_{ij} x_j - n \right| \le n\epsilon, \quad \text{for all } i = 1, \ldots, n.$$

Denote $x_{\min} = \min_{1 \le i \le n} x_i$. Then there exists a positive vector $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ such that

$$\sum_{j=1}^{n} \tilde{x}_i \widetilde{A}_{ij} \tilde{x}_j = n, \quad i = 1, \ldots, n$$

and

$$\left| \frac{\tilde{x}_i}{x_i} - 1 \right| \le \frac{\epsilon}{1 - \epsilon} + \frac{4\epsilon \sqrt{\tilde{a}_{\max}}}{\tilde{a}_{\min}^2 x_{\min}^3}, \quad i = 1, \ldots, n.$$

The following lemma provides a tight bound on the Schatten norms of inhomogeneous random matrices.

LEMMA 2.3.3 (Theorem 4.8 in Latała et al. [2018]). *Suppose $\boldsymbol{X}$ is a random matrix with independent and centered upper-triangular entries, and define the quantities*

$$\sigma_p := \left(\sum_i \left(\sum_j \mathbb{E}\, X_{ij}^2\right)^p\right)^{1/2p}, \quad \sigma_p^* := \left(\sum_{i,j} \|X_{ij}\|_\infty^{2p}\right)^{1/2p}.$$

Then for every $p \in \mathbb{N}$

$$\left(\mathbb{E}\, \|\boldsymbol{X}\|_{S_{2p}}^{2p}\right)^{1/2p} \leq 2\sigma_p + C\sqrt{p}\sigma_p^*,$$

where C is a universal constant. Moreover, we have

$$\mathbb{P}\left(\|\boldsymbol{X}\|_{S_{2p}} \geq 2\sigma_p + C\sqrt{p}\sigma_p^* + t\right) \leq \exp\left(-\frac{t^2}{C\max_{i,j}\|X_{ij}\|_\infty^2}\right)$$

for all $p \in \mathbb{N}$ and $t \geq 0$.

As a direct consequence of the previous lemma, the following corollary bounds the operator norm of inhomogeneous random matrices, denoted as Remark 4.12 in Latała et al. [2018].

COROLLARY 2.3.1. *Let $\boldsymbol{X}$ be as in Lemma 2.3.3. Define the quantities*

$$\sigma_\infty := \max_i \sqrt{\sum_j \mathbb{E}\, X_{ij}^2}, \quad \sigma_\infty^* := \max_{i,j} \|X_{ij}\|_\infty.$$

Then for every $0 \leq \epsilon \leq 1$ and $t \geq 0$, we have

$$\mathbb{P}\left(\|\boldsymbol{X}\| \geq 2(1+\epsilon)\sigma_\infty + t\right) \leq n \exp\left(-\frac{\epsilon t^2}{C\sigma_\infty^{*2}}\right)$$

where C is a universal constant.

PROOF. By Jensen's inequality and monotonicity of the Schatten norm

$$\left(\mathbb{E}\, \|\boldsymbol{X}\|\right)^{2p} \leq \mathbb{E}\left(\|\boldsymbol{X}\|^{2p}\right) \leq \mathbb{E}\, \|\boldsymbol{X}\|_{2p}^{2p},$$

Apply Lemma 2.3.3 with $p = \alpha \log n$ and $\alpha \geq 1$, we have

$$\mathbb{E}\,\|\boldsymbol{X}\| \leq 2e^{1/2\alpha}\sigma_\infty + Ce^{1/\alpha}\sigma_\infty^*\sqrt{\alpha \log n}.$$

Since $1 \leq e^{1/2\alpha} \leq 2$ for $\alpha \geq 1$, for every $0 \leq \epsilon \leq 1$, we have

$$\mathbb{E}\,\|\boldsymbol{X}\| \leq 2(1+\epsilon)\sigma_\infty + C_\epsilon\sigma_\infty^*\sqrt{\log n}$$

for a suitable constant C_ϵ depending on ϵ. Next, we follow arguments in Lemma 3.12 of Bandeira and Van Handel [2016] to derive the tail bound. By a form of Talagrand's concentration inequality,

$$\mathbb{P}\left(\|\boldsymbol{X}\| \geq \mathbb{E}\,\|\boldsymbol{X}\| + t\right) \leq \exp\left(-\frac{t^2}{C\sigma_\infty^{*2}}\right)$$

for all $t \geq 0$, where C is a universal constant. Combining the above inequalities, we have

$$\mathbb{P}\left(\|\boldsymbol{X}\| \geq 2\left(1+\epsilon\right)\sigma_\infty + C_\epsilon\sigma_\infty^*\sqrt{\log n} + t\right) \leq \mathbb{P}\left(\|\boldsymbol{X}\| \geq \mathbb{E}\,\|\boldsymbol{X}\| + t\right) \leq \exp\left(-\frac{t^2}{C\sigma_\infty^{*2}}\right)$$

for every $t \geq 0$. For $t \geq \sqrt{C \log n}$, we have

$$\begin{aligned}\mathbb{P}\left(\|\boldsymbol{X}\| \geq 2\left(1+\epsilon\right)\sigma_\infty + C'_\epsilon\sigma_\infty^* t\right) &\leq \mathbb{P}\left(\|\boldsymbol{X}\| \geq 2\left(1+\epsilon\right)\sigma_\infty + C_\epsilon\sigma_\infty^*\sqrt{\log n} + \sigma_\infty^* t\right) \\ &\leq \exp\left(-\frac{t^2}{C}\right)\end{aligned}$$

where C'_ϵ is chosen appropriately; while for $t \leq \sqrt{C \log n}$,

$$\mathbb{P}\left(\|\boldsymbol{X}\| \geq 2\left(1+\epsilon\right)\sigma_\infty + C'_\epsilon\sigma_\infty^* t\right) \leq 1 \leq n\exp\left(-\frac{t^2}{C}\right).$$

Combining the above two bounds completes the proof. □

2.3.2. Supporting Lemmas of Theorem 2.2.1. In the null case, by the NSP, we know that the estimated communities and the true communities give the same partitions of

the nodes. Without loss of generality, let $\widehat{\mathcal{N}}_k = \mathcal{N}_k$ for $k = 1, \ldots, K$. This also implies $\mathbf{1}_k = \hat{\mathbf{1}}_k$ for $k = 1, \ldots, K$.

We first show the concentration of $\hat{\theta}_i$ around θ_i for $i = 1, \ldots, n$ and $\widehat{B}_{kl}$ around B_{kl} for $1 \le k, l \le K$.

LEMMA 2.3.4. *Let $\epsilon > 0$ be any fixed small constant. If C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$, there hold*

$$1 - \epsilon \le \frac{\hat{\theta}_i}{\theta_i} \le 1 + \epsilon \quad \textit{for } i = 1, \ldots, n$$

and

$$1 - \epsilon \le \frac{\widehat{B}_{kl}}{B_{kl}} \le 1 + \epsilon \quad \textit{for } 1 \le k, l \le K.$$

PROOF. Recall that $d_i^* = \sum_{j=1}^n M_{ij}$, which implies $\sum_{j \in \mathcal{N}_k} d_j^* = \mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_n$. By the NSP, (2.6) can be rewritten as $\hat{\theta}_i = \frac{d_i}{\sum_{j \in \mathcal{N}_k} d_j} \sqrt{\mathbf{1}_k^\top \boldsymbol{A} \mathbf{1}_k}$. Combined with (2.4), (2.5) and (2.7), we get

$$\frac{\theta_i}{\hat{\theta}_i} = \frac{d_i^*}{d_i} \frac{\sum_{j \in \mathcal{N}_k} d_j}{\sum_{j \in \mathcal{N}_k} d_j^*} \sqrt{\frac{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_k}{\mathbf{1}_k^\top \boldsymbol{A} \mathbf{1}_k}}, \tag{2.19}$$

$$\frac{B_{kl}}{\widehat{B}_{kl}} = \frac{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_l}{\mathbf{1}_k^\top \boldsymbol{A} \mathbf{1}_l} \sqrt{\frac{\mathbf{1}_k^\top \boldsymbol{A} \mathbf{1}_k}{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_k}} \sqrt{\frac{\mathbf{1}_l^\top \boldsymbol{A} \mathbf{1}_l}{\mathbf{1}_l^\top \boldsymbol{M} \mathbf{1}_l}}. \tag{2.20}$$

Let's first study the concentration of $\frac{d_i}{d_i^*}$. Since d_i is the sum of independent random variables satisfying (2.18) with some constant R, by Lemma A.1.1, we have

$$\mathbb{P}\left(|d_i - d_i^*| \ge \epsilon' d_i^*\right) \le 2 \exp\left(-\frac{\epsilon'^2 d_i^*}{2(1 + R\epsilon')}\right) \tag{2.21}$$

for any fixed $\epsilon' > 0$. Note that $d_i^* \ge c_0 n \theta_{\min}^2 \ge C_0 c_0 \log^4 n$. Therefore, if C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$, we have

$$1 - \epsilon' \le \frac{d_i}{d_i^*} \le 1 + \epsilon' \quad \text{for } i = 1, \ldots, n.$$

Similarly, with probability $1 - O(n^{-3})$, we have $1 - \epsilon' \le \frac{\mathbf{1}_k^\top \boldsymbol{A} \mathbf{1}_k}{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_k} \le 1 + \epsilon'$ for all $k = 1, \dots, K$ and $1 - \epsilon' \le \frac{\mathbf{1}_k^\top \boldsymbol{A} \mathbf{1}_l}{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_l} \le 1 + \epsilon'$ for all $1 \le k < l \le K$. By choosing ϵ' sufficiently small, our claim is proved. □

Recall the population scaling parameters $\{\xi_i\}_{i=1}^n$ defined in (2.8), and the sample scaling parameters defined in (2.9). The following lemma shows the concentration of the sample scaling parameters around the corresponding population counterparts.

LEMMA 2.3.5. *Let $\epsilon > 0$ be any fixed small constant. If C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$, there hold*

$$1 - \epsilon \le \frac{\hat{\xi}_i}{\xi_i} \le 1 + \epsilon \quad \textit{for } i = 1, \dots, n.$$

PROOF. Assumption 1 states that $c_0 \le B_{kl} \le 1$ for all $1 \le k, l \le K$, thus the matrix $\boldsymbol{\Pi} \boldsymbol{B} \boldsymbol{\Pi}^\top$ is strictly positive. By Lemma 2.3.1, $\boldsymbol{\xi} = [\xi_1, \dots, \xi_n]^\top$ is the unique positive vector satisfying

$$\xi_i \left(\sum_{j=1}^n \left(\boldsymbol{\Pi} \boldsymbol{B} \boldsymbol{\Pi}^\top \right)_{ij} \xi_j \right) = \sum_{j=1}^n B_{\phi(i)\phi(j)} \xi_i \xi_j = n \quad \text{for } i = 1, \dots, n,$$

and its entries satisfy

$$c_0 \le \frac{\sqrt{\min_{k,l} B_{kl}}}{\max_{k,l} B_{kl}} \le \xi_i \le \frac{\sqrt{\max_{k,l} B_{kl}}}{\min_{k,l} B_{kl}} \le c_0^{-1} \quad \text{for } i = 1, \dots, n,$$

which also implies $\xi_{\min} = \min_{1 \le i \le n} \xi_i \ge c_0$. Fix $\epsilon' > 0$. By Lemma 2.3.4, when C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$,

$$1 - \epsilon' \le \frac{\widehat{B}_{kl}}{B_{kl}} \le 1 + \epsilon' \quad \text{for } 1 \le k, l \le K,$$

which implies

$$\begin{cases} \min_{1 \le k, l \le K} \widehat{B}_{kl} \ge (1 - \epsilon') \min_{1 \le k, l \le K} B_{kl} \ge c_0 (1 - \epsilon'), \\ \max_{1 \le k, l \le K} \widehat{B}_{kl} \le (1 + \epsilon') \max_{1 \le k, l \le K} B_{kl} \le (1 + \epsilon') / c_0. \end{cases}$$

Furthermore, there hold

$$1-\epsilon' \le \frac{\sum_{j=1}^n \widehat{B}_{\phi(i)\phi(j)}\xi_i\xi_j}{\sum_{j=1}^n B_{\phi(i)\phi(j)}\xi_i\xi_j} \le 1+\epsilon' \quad \text{for } i=1,\cdots,n,$$

and thereby imply

$$\left|\sum_{j=1}^n \widehat{B}_{\phi(i)\phi(j)}\xi_i\xi_j - n\right| \le n\epsilon' \quad \text{for } i=1,\cdots,n.$$

Recall that $\hat{\boldsymbol{\xi}}$ is the vector satisfying $\sum_{j=1}^n \widehat{B}_{\hat{\phi}(i)\hat{\phi}(j)}\hat{\xi}_i\hat{\xi}_j = n$. Since $\widehat{\mathcal{N}}_k = \mathcal{N}_k$ for $k=1,\ldots,K$, we have $\phi(i)=\hat{\phi}(i)$ for $i=1,\cdots,n$. Then by Lemma 2.3.2,

$$\left|\frac{\hat{\xi}_i}{\xi_i}-1\right| \le \frac{\epsilon'}{1-\epsilon'} + \frac{4\epsilon'\sqrt{1+\epsilon'}}{c_0^{11/2}(1-\epsilon')^2} \quad \text{for } i=1,\ldots,n.$$

For any $\epsilon>0$, by choosing ϵ' appropriately, we have $\left|\frac{\hat{\xi}_i}{\xi_i}-1\right| \le \epsilon$ for $i=1,\ldots,n$. □

Recall that $\boldsymbol{E} = \boldsymbol{\Xi}^{\frac{1}{2}}\boldsymbol{\Theta}^{-\frac{1}{2}}(\boldsymbol{A}-\boldsymbol{M})\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{\Xi}^{\frac{1}{2}}$. The following lemma gives a tight bound of $\|\boldsymbol{E}\|$.

LEMMA 2.3.6. *For any fixed $\epsilon>0$, by choosing a sufficiently large C_0 in Assumption 1, with probability $1-O(n^{-3})$, there holds*

$$\|\boldsymbol{E}\| \le (2+\epsilon)\sqrt{n}.$$

PROOF. Define $\widehat{E}_{ij} = E_{ij}1_{(|E_{ij}|<q)}$ and $\widetilde{E}_{ij} = \widehat{E}_{ij} - \mathbb{E}\,\widehat{E}_{ij}$, where $q=\frac{\sqrt{n}}{\log n}$. We first bound $\|\widetilde{\boldsymbol{E}}\|$, and then $\|\widehat{\boldsymbol{E}}\|$, and lastly $\|\boldsymbol{E}\|$.

Notice that $\widetilde{\boldsymbol{E}}$ is symmetric with independent and centered entries for $i\ge j$, and the entries are uniformly bounded. Define the quantities

$$\tilde{\sigma} := \max_i \sqrt{\sum_j \mathbb{E}\,\widetilde{E}_{ij}^2}, \quad \tilde{\sigma}_* := \max_{i,j}\|\widetilde{E}_{ij}\|_\infty.$$

Then apply Corollary 2.3.1 to $\widetilde{\boldsymbol{E}}$. For any $\epsilon, t > 0$, there holds

$$\mathbb{P}\left(\|\widetilde{\boldsymbol{E}}\| \geq 2(1+\epsilon)\tilde{\sigma} + t\right) \leq n \exp\left(-\frac{\epsilon t^2}{C\tilde{\sigma}_*^2}\right).$$

By letting $t = \sqrt{\frac{4C}{\epsilon}}\tilde{\sigma}_*\sqrt{\log n}$, with probability $1 - O(n^{-3})$, we have

$$\|\widetilde{\boldsymbol{E}}\| \leq 2(1+\epsilon)\tilde{\sigma} + C_\epsilon \tilde{\sigma}_* \sqrt{\log n}, \tag{2.22}$$

where C_ϵ is a constant depending on ϵ only. Note that $E_{ij} = (A_{ij} - M_{ij})\xi_i^{\frac{1}{2}}\xi_j^{\frac{1}{2}}\theta_i^{-\frac{1}{2}}\theta_j^{-\frac{1}{2}}$. Then by Assumption (2.17),

$$\mathbb{E}\, E_{ij}^2 = \mathrm{Var}(A_{ij})\xi_i\xi_j\theta_i^{-1}\theta_j^{-1} \leq M_{ij}\xi_i\xi_j\theta_i^{-1}\theta_j^{-1} = B_{\phi(i)\phi(j)}\xi_i\xi_j.$$

Also, notice that

$$\mathbb{E}\,\widetilde{E}_{ij}^2 = \mathrm{Var}\left(\widetilde{E}_{ij}\right) = \mathrm{Var}\left(\widehat{E}_{ij}\right) \leq \mathbb{E}\,\widehat{E}_{ij}^2 \leq \mathbb{E}\, E_{ij}^2.$$

Then the population scaling (2.8) gives $\sum_{j=1}^n \mathbb{E}\, E_{ij}^2 \leq n$ for $i = 1, \cdots, n$, which imply $\tilde{\sigma} \leq \sqrt{n}$.

Next, since $\|\widehat{E}_{ij}\|_\infty \leq q$, we have $\left|\mathbb{E}\,\widehat{E}_{ij}\right| \leq q$ and thus $\|\widetilde{E}_{ij}\|_\infty \leq \|\widehat{E}_{ij}\|_\infty + \left|\mathbb{E}\,\widehat{E}_{ij}\right| \leq 2q$, which implies $\tilde{\sigma}_* \leq 2q$. Then, from (2.22), we have

$$\|\widetilde{\boldsymbol{E}}\| \leq 2(1+\epsilon)\sqrt{n} + C_\epsilon q\sqrt{\log n}. \tag{2.23}$$

Since $q = \frac{\sqrt{n}}{\log n}$, when n is sufficiently large,

$$\|\widetilde{\boldsymbol{E}}\| \leq 2(1+\epsilon)\sqrt{n} + C_\epsilon\sqrt{\frac{n}{\log n}} \leq (2+3\epsilon)\sqrt{n}.$$

Next, we give an upper bound of $\|\widehat{\boldsymbol{E}}\|$. Since $\widetilde{\boldsymbol{E}} = \widehat{\boldsymbol{E}} - \mathbb{E}\,\widehat{\boldsymbol{E}}$, we have $\|\widehat{\boldsymbol{E}}\| \leq \|\widetilde{\boldsymbol{E}}\| + \|\mathbb{E}\,\widehat{\boldsymbol{E}}\|_F$. Notice that

$$\mathbb{E}\, E_{ij} 1_{(|E_{ij}| \geq q)} = \mathbb{E}\left(E_{ij} - \widehat{E}_{ij}\right) = -\mathbb{E}\,\widehat{E}_{ij}.$$

By Cauchy-Schwarz inequality,

$$(\mathbb{E}\,\widehat{E}_{ij})^2 = \left(\mathbb{E}\,E_{ij}1_{(|E_{ij}|\geq q)}\right)^2 \leq \left(\mathbb{E}\,E_{ij}^2\right)\mathbb{P}\left(|E_{ij}|\geq q\right).$$

Since $\mathbb{E}\,E_{ij}^2 \leq \xi_i\xi_j B_{\phi(i)\phi(j)}$, it suffices to control $\mathbb{P}\left(|E_{ij}|\geq q\right)$. Note that we have assumed $B_{kl} \leq 1$ for $1 \leq k,l \leq K$, and shown in the proof of Lemma 2.3.5 that $c_0 \leq \xi_i \leq c_0^{-1}$ for $i = 1,\ldots,n$. Since for all $1 \leq i \leq j \leq n$, A_{ij} satisfies (2.17) and (2.18), by Lemma A.1.1, for some constant R,

$$\begin{aligned}
\mathbb{P}(|E_{ij}|\geq q) &= \mathbb{P}\left(|A_{ij}-M_{ij}| \geq q\sqrt{\frac{\theta_i\theta_j}{\xi_i\xi_j}}\right)\\
&\leq 2\exp\left(-\frac{q^2\frac{\theta_i\theta_j}{\xi_i\xi_j}}{2\left(M_{ij}+Rq\sqrt{\frac{\theta_i\theta_j}{\xi_i\xi_j}}\right)}\right)\\
&\leq 2\exp\left(-\frac{c_0^2q^2}{2\left(1+c_0^{-1}Rq\theta_{\min}^{-1}\right)}\right)\\
&\leq 2\exp\left(-\frac{1}{4}\left(c_0^2q^2 \wedge c_0^3R^{-1}q\theta_{\min}\right)\right).
\end{aligned}$$

Since $\theta_{\min} \geq C_0\sqrt{\frac{\log^4 n}{n}}$ and $q = \frac{\sqrt{n}}{\log n}$, when n and C_0 are sufficiently large, we have $\mathbb{P}(|E_{ij}|\geq q) \leq 2n^{-5}$. Then

$$\begin{aligned}
\|\,\mathbb{E}\,\widehat{\boldsymbol{E}}\|_F^2 &= \sum_{1\leq i,j\leq n}(\mathbb{E}\,\widehat{E}_{ij})^2\\
&= \sum_{1\leq i,j\leq n}\left(\mathbb{E}\,E_{ij}1_{(|E_{ij}|\geq q)}\right)^2\\
&\leq \sum_{1\leq i,j\leq n}\left(\mathbb{E}\,E_{ij}^2\right)\mathbb{P}\left(|E_{ij}|\geq q\right)\\
&\leq 2n^{-5}\sum_{1\leq i,j\leq n}\xi_i\xi_jB_{\phi(i)\phi(j)}\\
&= 2n^{-3},
\end{aligned}$$

where the last equality is due to (2.8). Then

$$\|\widehat{\boldsymbol{E}}\| \le \|\widetilde{\boldsymbol{E}}\| + \|\mathbb{E}\,\widehat{\boldsymbol{E}}\|_F \le (2+4\epsilon)\sqrt{n}.$$

Finally, since $\mathbb{P}(|E_{ij}| \ge q) \le 2n^{-5}$, we have $\mathbb{P}(\widehat{\boldsymbol{E}} = \boldsymbol{E}) \ge 1 - 2n^{-3}$. This implies that with probability $1 - O(n^{-3})$, $\|\boldsymbol{E}\| = \|\widehat{\boldsymbol{E}}\| \le (2+4\epsilon)\sqrt{n}$. Replacing 4ϵ with ϵ completes the proof. □

2.3.3. Supporting Lemmas of Theorem 2.2.2. In the under-fitting case $m < K$, the NSP implies that the true communities $\mathcal{N}_1, \ldots, \mathcal{N}_K$ are refinements of the estimated communities $\widehat{\mathcal{N}}_1^{(m)}, \ldots, \widehat{\mathcal{N}}_m^{(m)}$. For each $k = 1, \ldots, m$, we assume that the number of true communities contained in $\widehat{\mathcal{N}}_k^{(m)}$ is $r_k \ge 1$, which implies that $r_1 + \cdots + r_m = K$. Then we can represent the estimated communities as

$$\widehat{\mathcal{N}}_k^{(m)} = \mathcal{N}_{h_{k1}} \cup \cdots \cup \mathcal{N}_{h_{kr_k}}, \quad k = 1, \ldots, m.$$

Here all indices h_{kj} for $k = 1, \ldots, m$ and $j = 1, \ldots, r_k$ are distinct over $1, \ldots, K$. We can also decompose

$$\hat{\mathbf{1}}_k^{(m)} = \mathbf{1}_{h_{k1}} + \cdots + \mathbf{1}_{h_{kr_k}}, \quad k = 1, \ldots, m.$$

LEMMA 2.3.7. *If C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$, there hold*

$$c_1 \le \frac{\hat{\theta}_i}{\theta_i} \le \frac{1}{c_1} \quad \textit{for } i = 1, \ldots, n,$$

and

$$c_1 \le \widehat{B}_{kl} \le \frac{1}{c_1} \quad \textit{for } 1 \le k < l \le m,$$

where $0 < c_1 < 1$ is a constant only depending on c_0 in Assumption 1.

PROOF. With the aforementioned notations, for $1 \le k, l \le m$, we have

$$\begin{aligned}\hat{\mathbf{1}}_k^\top \boldsymbol{M} \hat{\mathbf{1}}_l &= \sum_{a=1}^{r_k} \sum_{b=1}^{r_l} \mathbf{1}_{h_{ka}}^\top \boldsymbol{M} \mathbf{1}_{h_{lb}} \\ &= \sum_{a=1}^{r_k} \sum_{b=1}^{r_l} \boldsymbol{B}_{h_{ka} h_{lb}} \left\|\boldsymbol{\theta}_{h_{ka}}\right\|_1 \left\|\boldsymbol{\theta}_{h_{lb}}\right\|_1 .\end{aligned}$$

Thus by Assumption 1,

$$c_0 \left(\sum_{a=1}^{r_k} \left\|\boldsymbol{\theta}_{h_{ka}}\right\|_1\right)\left(\sum_{b=1}^{r_l} \left\|\boldsymbol{\theta}_{h_{lb}}\right\|_1\right) \le \hat{\mathbf{1}}_k^\top \boldsymbol{M} \hat{\mathbf{1}}_l \le \left(\sum_{a=1}^{r_k} \left\|\boldsymbol{\theta}_{h_{ka}}\right\|_1\right)\left(\sum_{b=1}^{r_l} \left\|\boldsymbol{\theta}_{h_{lb}}\right\|_1\right). \tag{2.24}$$

Recall that $d_i^* = \sum_{j=1}^n \boldsymbol{M}_{ij} = \theta_i \sum_{l=1}^K \boldsymbol{B}_{\phi(i)l} \|\boldsymbol{\theta}_l\|_1$, which implies

$$\sum_{j \in \mathcal{N}_k} d_j^* = \|\boldsymbol{\theta}_k\|_1 \left(\sum_{l=1}^K \boldsymbol{B}_{kl} \|\boldsymbol{\theta}_l\|_1\right),$$

and further implies

$$\sum_{j \in \widehat{\mathcal{N}}_k} d_j^* = \sum_{a=1}^{r_k} \|\boldsymbol{\theta}_{h_{ka}}\|_1 \left(\sum_{l=1}^K \boldsymbol{B}_{h_{ka} l} \|\boldsymbol{\theta}_l\|_1\right).$$

Thus

$$c_0 \|\boldsymbol{\theta}\|_1 \sum_{a=1}^{r_k} \|\boldsymbol{\theta}_{h_{ka}}\|_1 \le \sum_{j \in \widehat{\mathcal{N}}_k} d_j^* \le \|\boldsymbol{\theta}\|_1 \sum_{a=1}^{r_k} \|\boldsymbol{\theta}_{h_{ka}}\|_1. \tag{2.25}$$

From (2.5) and (2.6), $\theta_i = \frac{\sqrt{\mathbf{1}_{\phi(i)}^\top \boldsymbol{M} \mathbf{1}_{\phi(i)}}}{\sum_{j \in \mathcal{N}_{\phi(i)}} d_j^*} d_i^*$ and $\hat{\theta}_i = \frac{\sqrt{\hat{\mathbf{1}}_{\hat{\phi}(i)}^\top \boldsymbol{A} \hat{\mathbf{1}}_{\hat{\phi}(i)}}}{\sum_{j \in \widehat{\mathcal{N}}_{\hat{\phi}(i)}} d_j} d_i$, thus

$$\frac{\theta_i}{\hat{\theta}_i} = \frac{d_i^*}{d_i} \frac{\sum_{j \in \widehat{\mathcal{N}}_{\hat{\phi}(i)}} d_j}{\sum_{j \in \mathcal{N}_{\phi(i)}} d_j^*} \sqrt{\frac{\mathbf{1}_{\phi(i)}^\top \boldsymbol{M} \mathbf{1}_{\phi(i)}}{\hat{\mathbf{1}}_{\hat{\phi}(i)}^\top \boldsymbol{A} \hat{\mathbf{1}}_{\hat{\phi}(i)}}}. \tag{2.26}$$

Notice that by Lemma 2.3.4, for any $\epsilon > 0$, if C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$,

$$1 - \epsilon \leq \frac{d_i}{d_i^*} \leq 1 + \epsilon \quad \text{for } i = 1, \cdots, n,$$
$$1 - \epsilon \leq \frac{\mathbf{1}_k^\top \boldsymbol{A} \mathbf{1}_l}{\mathbf{1}_k^\top \boldsymbol{M} \mathbf{1}_l} \leq 1 + \epsilon \quad \text{for } 1 \leq k, l \leq K,$$

which further give

$$1 - \epsilon \leq \frac{\sum_{j \in \widehat{\mathcal{N}}_k} d_j}{\sum_{j \in \widehat{\mathcal{N}}_k} d_j^*} \leq 1 + \epsilon \quad \text{for } k = 1, \cdots, m, \tag{2.27}$$

$$1 - \epsilon \leq \frac{\hat{\mathbf{1}}_k^\top \boldsymbol{A} \hat{\mathbf{1}}_l}{\hat{\mathbf{1}}_k^\top \boldsymbol{M} \hat{\mathbf{1}}_l} \leq 1 + \epsilon \quad \text{for } 1 \leq k, l \leq m. \tag{2.28}$$

By (2.27) and (2.28), we have for any $\epsilon' > 0$, if C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$,

$$1 - \epsilon' \leq \frac{\hat{\theta}_i}{\frac{\sqrt{\hat{\mathbf{1}}_{\hat{\phi}(i)}^\top \boldsymbol{M} \hat{\mathbf{1}}_{\hat{\phi}(i)}}}{\sum_{j \in \widehat{\mathcal{N}}_{\hat{\phi}(i)}} d_j^*} d_i^*} \leq 1 + \epsilon',$$

which by (2.26) further means

$$1 - \epsilon' \leq \frac{\theta_i}{\hat{\theta}_i} \left(\frac{\sum_{j \in \widehat{\mathcal{N}}_{\hat{\phi}(i)}} d_j^*}{\sum_{j \in \mathcal{N}_{\phi(i)}} d_j^*} \sqrt{\frac{\mathbf{1}_{\phi(i)}^\top \boldsymbol{M} \mathbf{1}_{\phi(i)}}{\hat{\mathbf{1}}_{\hat{\phi}(i)}^\top \boldsymbol{M} \hat{\mathbf{1}}_{\hat{\phi}(i)}}} \right)^{-1} \leq 1 + \epsilon'.$$

Notice that

$$\begin{aligned} \frac{\sum_{j \in \widehat{\mathcal{N}}_{\hat{\phi}(i)}} d_j^*}{\sum_{j \in \mathcal{N}_{\phi(i)}} d_j^*} \sqrt{\frac{\mathbf{1}_{\phi(i)}^\top \boldsymbol{M} \mathbf{1}_{\phi(i)}}{\hat{\mathbf{1}}_{\hat{\phi}(i)}^\top \boldsymbol{M} \hat{\mathbf{1}}_{\hat{\phi}(i)}}} &= \frac{\sum_{j \in \widehat{\mathcal{N}}_{\hat{\phi}(i)}} d_j^*}{\|\boldsymbol{\theta}_{\phi(i)}\|_1 \left(\sum_{l=1}^K B_{\phi(i)l} \|\boldsymbol{\theta}_l\|_1 \right)} \sqrt{\frac{\|\boldsymbol{\theta}_{\phi(i)}\|_1^2}{\hat{\mathbf{1}}_{\hat{\phi}(i)}^\top \boldsymbol{M} \hat{\mathbf{1}}_{\hat{\phi}(i)}}} \\ &= \frac{\sum_{j \in \widehat{\mathcal{N}}_{\hat{\phi}(i)}} d_j^*}{\left(\sum_{l=1}^K B_{\phi(i)l} \|\boldsymbol{\theta}_l\|_1 \right) \sqrt{\hat{\mathbf{1}}_{\hat{\phi}(i)}^\top \boldsymbol{M} \hat{\mathbf{1}}_{\hat{\phi}(i)}}}. \end{aligned}$$

Combined with (2.24), (2.25) and $c_0\|\boldsymbol{\theta}\|_1 \le \sum_{l=1}^K B_{\phi(i)l}\|\boldsymbol{\theta}_l\|_1 \le \|\boldsymbol{\theta}\|_1$, we have

$$c_0 \le \frac{\sum_{j\in\widehat{\mathcal{N}}_{\hat{\phi}(i)}} d_j^*}{\sum_{j\in\mathcal{N}_{\phi(i)}} d_j^*}\sqrt{\frac{\mathbf{1}_{\phi(i)}^\top \boldsymbol{M}\mathbf{1}_{\phi(i)}}{\hat{\mathbf{1}}_{\hat{\phi}(i)}^\top \boldsymbol{M}\hat{\mathbf{1}}_{\hat{\phi}(i)}}} \le c_0^{-3/2},$$

which means for any $\epsilon' > 0$, if C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$,

$$c_0(1-\epsilon') \le \frac{\theta_i}{\hat{\theta}_i} \le c_0^{-3/2}(1+\epsilon') \quad \text{for } i = 1, \cdots, n.$$

Next we deal with $\widehat{B}_{kl}$. Recall the definition from (2.7):

$$\widehat{B}_{kl} = \frac{\hat{\mathbf{1}}_k^\top \boldsymbol{A}\hat{\mathbf{1}}_l}{\sqrt{\hat{\mathbf{1}}_k^\top \boldsymbol{A}\hat{\mathbf{1}}_k}\sqrt{\hat{\mathbf{1}}_l^\top \boldsymbol{A}\hat{\mathbf{1}}_l}}.$$

By (2.28), for any $\epsilon' > 0$, if C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$,

$$1-\epsilon' \le \frac{\widehat{B}_{kl}}{\frac{\hat{\mathbf{1}}_k^\top \boldsymbol{M}\hat{\mathbf{1}}_l}{\sqrt{\hat{\mathbf{1}}_k^\top \boldsymbol{M}\hat{\mathbf{1}}_k}\sqrt{\hat{\mathbf{1}}_l^\top \boldsymbol{M}\hat{\mathbf{1}}_l}}} \le 1+\epsilon',$$

which by (2.24) gives

$$c_0 \le \frac{\hat{\mathbf{1}}_k^\top \boldsymbol{M}\hat{\mathbf{1}}_l}{\sqrt{\hat{\mathbf{1}}_k^\top \boldsymbol{M}\hat{\mathbf{1}}_k}\sqrt{\hat{\mathbf{1}}_l^\top \boldsymbol{M}\hat{\mathbf{1}}_l}} \le c_0^{-1}.$$

Thus for any $\epsilon' > 0$, if C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$,

$$c_0(1-\epsilon') \le \widehat{B}_{kl} \le c_0^{-1}(1+\epsilon') \quad \text{for } 1 \le k < l \le m.$$

□

LEMMA 2.3.8. *If C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$, there hold*

$$c_2 \le \hat{\xi}_i \le \frac{1}{c_2} \quad \text{for } 1 \le i \le n,$$

where $c_2 < 1$ is a constant only depending on c_0 in Assumption 1.

PROOF. Notice that by Lemma 2.3.1, $\hat{\boldsymbol{\xi}}$ is the unique positive vector satisfying

$$\sum_{j=1}^{n} \widehat{B}_{\hat{\phi}(i)\hat{\phi}(j)} \hat{\xi}_i \hat{\xi}_j = n \quad \text{for } i = 1, \cdots, n,$$

and its entries satisfy

$$\frac{\sqrt{\min_{k,l} \widehat{B}_{kl}}}{\max_{k,l} \widehat{B}_{kl}} \le \hat{\xi}_i \le \frac{\sqrt{\max_{k,l} \widehat{B}_{kl}}}{\min_{k,l} \widehat{B}_{kl}} \quad \text{for } i = 1, \ldots, n.$$

By Lemma 2.3.7, if C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$,

$$c_1 \le \widehat{B}_{kl} \le \frac{1}{c_1} \quad \text{for } 1 \le k < l \le m$$

where c_1 is a constant only depending on c_0 in Assumption 1. Then there hold

$$c_1^{3/2} \le \hat{\xi}_i \le c_1^{-3/2} \quad \text{for } i = 1, \ldots, n.$$

□

LEMMA 2.3.9. *For any fixed $C_1 > 0$, if C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$, we have*

$$\left| \lambda_K \left(\widetilde{\boldsymbol{A}}_1 \right) \right| \ge C_1 (\log n)^2 \sqrt{n}.$$

PROOF. Recall that $\boldsymbol{M} = \boldsymbol{\Theta}\boldsymbol{\Pi}\boldsymbol{B}\boldsymbol{\Pi}^\top\boldsymbol{\Theta}$, therefore,

$$\boldsymbol{\Theta}^{-\frac{1}{2}} \boldsymbol{M} \boldsymbol{\Theta}^{-\frac{1}{2}} = \boldsymbol{\Theta}^{\frac{1}{2}} \boldsymbol{\Pi}\boldsymbol{B}\boldsymbol{\Pi}^\top \boldsymbol{\Theta}^{\frac{1}{2}}.$$

Observe that

$$\boldsymbol{\Theta}^{\frac{1}{2}}\boldsymbol{\Pi} = \left[\sqrt{\boldsymbol{\theta}}_1, \sqrt{\boldsymbol{\theta}_2}, \ldots, \sqrt{\boldsymbol{\theta}_K}\right].$$

Since $\{\sqrt{\boldsymbol{\theta}_k}\}_{k=1}^K$ are orthogonal vectors, we have

$$\sigma_K\left(\boldsymbol{\Theta}^{\frac{1}{2}}\boldsymbol{\Pi}\right) = \min_k \sqrt{\|\boldsymbol{\theta}_k\|_1},$$

thus

$$\left|\lambda_K\left(\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{M}\boldsymbol{\Theta}^{-\frac{1}{2}}\right)\right| \geq |\lambda_K(\boldsymbol{B})| \cdot \sigma_K\left(\boldsymbol{\Theta}^{\frac{1}{2}}\boldsymbol{\Pi}\right)^2 = |\lambda_K(\boldsymbol{B})| \min_k \|\boldsymbol{\theta}_k\|_1.$$

Combined with (2.14) and (2.15), we have

$$\left|\lambda_K\left(\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{M}\boldsymbol{\Theta}^{-\frac{1}{2}}\right)\right| \geq c_0^2 n\theta_{\min}.$$

By Lemma 2.3.7, if C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$, $c_1 \leq \frac{\theta_i}{\hat{\theta}_i} \leq \frac{1}{c_1}$ for $i = 1, \ldots, n$. Therefore,

$$\begin{aligned}\left|\lambda_K\left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\boldsymbol{M}\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\right)\right| &\geq c_1\left|\lambda_K\left(\boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{M}\boldsymbol{\Theta}^{-\frac{1}{2}}\right)\right| \\ &\geq c_0^2 c_1 n\theta_{\min} \\ &\geq c_0^2 c_1 C_0 (\log n)^2\sqrt{n}.\end{aligned}$$

Combined with Lemma 2.3.8, we have with probability $1 - O(n^{-3})$,

$$\begin{aligned}\left|\lambda_K\left(\widetilde{\boldsymbol{A}}_1\right)\right| &\geq c_2\left|\lambda_K\left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\boldsymbol{M}\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\right)\right| \\ &\geq c_0^2 c_1 c_2 C_0 (\log n)^2\sqrt{n},\end{aligned}$$

which proves the lemma. □

2.3.4. Proof of Theorem 2.2.1. Recall that $\widetilde{\boldsymbol{A}}_2 = \left(\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}}\boldsymbol{\Xi}^{-\frac{1}{2}}\right)\left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\boldsymbol{\Theta}^{\frac{1}{2}}\right)\boldsymbol{E}\left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}}\boldsymbol{\Theta}^{\frac{1}{2}}\right)\left(\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}}\boldsymbol{\Xi}^{-\frac{1}{2}}\right)$. By Lemma 2.3.4 and 2.3.5, for any fixed $\epsilon > 0$, if C_0 is sufficiently large, with probability

$1 - O(n^{-3})$, we have

$$1 - \epsilon \le \frac{\hat{\theta}_i}{\theta_i} \le 1 + \epsilon \quad \text{for } i = 1, \ldots, n,$$

$$1 - \epsilon \le \frac{\hat{\xi}_i}{\xi_i} \le 1 + \epsilon, \quad \text{for } i = 1, \ldots, n,$$

which gives $\left\| \left(\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}} \boldsymbol{\Xi}^{-\frac{1}{2}} \right) \left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}} \boldsymbol{\Theta}^{\frac{1}{2}} \right) \right\| \le 1 + \epsilon$. Next, by Lemma 2.3.6, for any fixed $\epsilon > 0$, by choosing a sufficiently large C_0, with probability $1 - O(n^{-3})$, there holds

$$\|\boldsymbol{E}\| \le (2 + \epsilon)\sqrt{n}. \tag{2.29}$$

Combining the results above, by choosing sufficiently small, we have with probability $1 - O(n^{-3})$,

$$\left\| \widetilde{\boldsymbol{A}}_2 \right\| \le 2.05\sqrt{n}$$

2.3.5. Proof of Theorem 2.2.2. By Theorem 2.2.2, for any fixed c_0, there exists a sufficiently large C_0 such that with probability $1 - O(n^{-3})$, SCORE satisfies the NSP. First, we show that, for any estimated community partition satisfying the NSP, with probability $1 - O(n^{-3})$, the obtained normalized adjacency matrix $\widetilde{\boldsymbol{A}}$ satisfies

$$\left| \lambda_{m+1}(\widetilde{\boldsymbol{A}}) \right| > 2.05\sqrt{n}. \tag{2.30}$$

By Lemma 2.3.9, for any fixed $C_1 > 0$, if C_0 in Assumption 1 is sufficiently large, with probability $1 - O(n^{-3})$, we have

$$\left| \lambda_K \left(\widetilde{\boldsymbol{A}}_1 \right) \right| \ge C_1 (\log n)^2 \sqrt{n}.$$

By Lemma 2.3.7 and 2.3.8, if C_0 is sufficiently large, with probability $1 - O(n^{-3})$, we have

$$c_1 \le \frac{\hat{\theta}_i}{\theta_i} \le \frac{1}{c_1} \quad \text{for } i = 1, \dots, n,$$
$$c_2 \le \hat{\xi}_i \le \frac{1}{c_2} \quad \text{for } i = 1, \dots, n,$$

where c_1, c_2 are constants only depending on c_0. Also $c_0^{\frac{3}{2}} \le \xi_i \le c_0^{-\frac{3}{2}}$ for $i = 1, \dots, n$. Then we have for some constant C depending on c_0

$$\left\| \left(\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}} \boldsymbol{\Xi}^{-\frac{1}{2}} \right) \left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}} \boldsymbol{\Theta}^{\frac{1}{2}} \right) \right\| \le C.$$

Combined with (2.29), we have with probability $1 - O(n^{-3})$,

$$\left\| \widetilde{\boldsymbol{A}}_2 \right\| \le \left\| \left(\widehat{\boldsymbol{\Xi}}^{\frac{1}{2}} \boldsymbol{\Xi}^{-\frac{1}{2}} \right) \left(\widehat{\boldsymbol{\Theta}}^{-\frac{1}{2}} \boldsymbol{\Theta}^{\frac{1}{2}} \right) \right\|^2 \|\boldsymbol{E}\| \lesssim \sqrt{n},$$

where $\lesssim$ only hides a constant depending on c_0. By Weyl's inequality,

$$\left| \lambda_{m+1}(\widetilde{\boldsymbol{A}}) \right| \ge \left| \lambda_K(\widetilde{\boldsymbol{A}}) \right| \ge \left| \lambda_K(\widetilde{\boldsymbol{A}}_1) \right| - \left\| \widetilde{\boldsymbol{A}}_2 \right\|.$$

By combining the results above, (2.30) is shown.

Notice that the possible number of realizations of the estimated community partitions $\widehat{\mathcal{N}}_1, \dots, \widehat{\mathcal{N}}_m$ satisfying the NSP is small. By the probability union bound, with probability $1 - O(n^{-3})$,

$$\left| \lambda_{m+1}(\widetilde{\boldsymbol{A}}) \right| > 2.05\sqrt{n}.$$

2.3.6. Note on the Assumptions Under the Bernoulli DCSBM. In Assumption 1, we assume $\theta_{\min} \ge C_0 \sqrt{\frac{\log^4 n}{n}}$ for technical reason as seen in the proof of Lemma 2.3.6. Here we show that under the Bernoulli DCSBM, to show the consistency of SMAST, we only need to assume $\theta_{\min} \ge C_0 \sqrt{\frac{\log n}{n}}$.

In Lemma 2.3.6, we apply Corollary 2.3.1 to the truncated matrix $\widetilde{\boldsymbol{E}}$ instead of $\boldsymbol{E}$ as under the general DCSBM, the entries of $\boldsymbol{E}$ are not necessarily bounded. The key difference

under the Bernoulli DCSBM is that the entries of $\boldsymbol{E}$ are uniformly bounded by 1. Thus we can obtain a tight bound of the operator norm of $\boldsymbol{E}$ by directly applying Corollary 2.3.1 to $\boldsymbol{E}$.

LEMMA 2.3.10. *Under the Bernoulli DCSBM, suppose Assumption 1 holds with (2.16) replaced by* $1 \geq \theta_{\max} \geq \theta_{\min} \geq C_0\sqrt{\frac{\log n}{n}}$. *For any fixed* $\epsilon > 0$, *by choosing a sufficiently large* C_0, *with probability* $1 - O(n^{-3})$, *there holds*

$$\|\boldsymbol{E}\| \leq (2+\epsilon)\sqrt{n}.$$

PROOF. Define the quantities

$$\tilde{\sigma} := \max_i \sqrt{\sum_j \mathbb{E}\, E_{ij}^2}, \quad \tilde{\sigma}_* := \max_{i,j} \|E_{ij}\|_\infty.$$

Then apply Corollary 2.3.1 to $\boldsymbol{E}$. For any $\epsilon, t > 0$, there holds

$$\mathbb{P}\left(\|\boldsymbol{E}\| \geq 2(1+\epsilon)\tilde{\sigma} + t\right) \leq n \exp\left(-\frac{\epsilon t^2}{C\tilde{\sigma}_*^2}\right).$$

By letting $t = \sqrt{\frac{4C}{\epsilon}}\tilde{\sigma}_*\sqrt{\log n}$, with probability $1 - O(n^{-3})$, we have

$$\|\boldsymbol{E}\| \leq 2(1+\epsilon)\tilde{\sigma} + C_\epsilon \tilde{\sigma}_* \sqrt{\log n} \tag{2.31}$$

where C_ϵ is a constant depending on ϵ only. Notice that

$$\begin{aligned}
\tilde{\sigma}^2 &= \max_i \sum_j \xi_i \xi_j (\theta_i \theta_j)^{-1} M_{ij}(1 - M_{ij}) \\
&\le \max_i \sum_j \xi_i \xi_j (\theta_i \theta_j)^{-1} M_{ij} \\
&\le \max_i \sum_j \xi_i \xi_j B_{\phi(i)\phi(j)} = n, \\
\tilde{\sigma}_* &= \max_{i,j} (\xi_i \xi_j)^{\frac{1}{2}} (\theta_i \theta_j)^{-\frac{1}{2}} \|A_{ij} - M_{ij}\|_\infty \\
&\le \max_{i,j} (\xi_i \xi_j)^{\frac{1}{2}} (\theta_i \theta_j)^{-\frac{1}{2}} \\
&\le c_0^{-1} \theta_{\min}^{-1} \le c_0^{-1} C_0^{-1} \sqrt{\frac{n}{\log n}}.
\end{aligned}$$

Plugged into (2.31), we have

$$\|\boldsymbol{E}\| \le 2\,(1+\epsilon)\,\sqrt{n} + C_\epsilon C_0^{-1}\sqrt{n} \le (2+3\epsilon)\sqrt{n}$$

provided that C_0 is sufficiently large. □

2.4. Proof of the Nonsplitting Property for SCORE

2.4.1. Population Eigenstructure. We first study the eigenvalues and eigenvectors of $\boldsymbol{M}$. The population adjacency matrix $\boldsymbol{M}$ is a rank-K matrix, with nonzero eigenvalues $\lambda_1^*, \lambda_2^*, \cdots, \lambda_K^*$ sorted in descending magnitude, and their corresponding unit-norm eigenvectors $\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_K^*$. Notice that by Perron's theorem [Horn and Johnson, 2012], λ_1^* is positive with multiplicity 1, and we can choose $\boldsymbol{u}_1^*$ such that all its entries are strictly positive. Denote $\boldsymbol{U}^* = [\boldsymbol{u}_2^*, \ldots, \boldsymbol{u}_K^*] \in \mathbb{R}^{n\times(K-1)}$ and $\big(\boldsymbol{U}_{i\cdot}^*\big)^\top$ as its i-th row. Also denote $\boldsymbol{\Lambda}^* = \mathrm{diag}(\lambda_2^*, \cdots, \lambda_K^*) \in \mathbb{R}^{(K-1)\times(K-1)}$.

We also define a diagonal matrix $\boldsymbol{H} \in \mathbb{R}^{K\times K}$ such that $H_{kk} = \|\boldsymbol{\theta}_k\|_2 / \|\boldsymbol{\theta}\|_2$. For each $1 \le k \le K$, let $\tilde{\lambda}_k^*$ be the k-th largest eigenvalue of $\boldsymbol{HBH}$ in magnitude, and let $\tilde{\boldsymbol{u}}_k^* \in \mathbb{R}^K$ be the associated (unit-norm) eigenvector. Similarly, by Perron's theorem, $\tilde{\lambda}_1^*$ is positive with

multiplicity 1, and the entries of $\tilde{\boldsymbol{u}}_1^*$ are strictly positive. Moreover, there is a one-to-one correspondence between $(\tilde{\lambda}_1^*, \ldots, \tilde{\lambda}_K^*)$ and $(\lambda_1^*, \ldots, \lambda_K^*)$, as well as the choice of $(\tilde{\boldsymbol{u}}_1^*, \ldots, \tilde{\boldsymbol{u}}_K^*)$ and the choice of $(\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_K^*)$, characterized by the following lemma.

LEMMA 2.4.1 (Lemma B.1 of Jin et al. [2022]). *Suppose Assumption 1 holds. Let $\lambda_k^*, \tilde{\lambda}_k^*, \boldsymbol{u}_k^*, \tilde{\boldsymbol{u}}_k^*$ be defined as above, then the following statements are true*

(1) $\lambda_k^* = \|\boldsymbol{\theta}\|_2^2 \tilde{\lambda}_k^*$ for $1 \le k \le K$.

(2) If $\tilde{\boldsymbol{u}}_k^*$ is an eigenvector of $\boldsymbol{HBH}$ corresponding to $\tilde{\lambda}_k^*$, then $\|\boldsymbol{\theta}\|_2^{-1} \boldsymbol{\Theta\Pi H}^{-1} \tilde{\boldsymbol{u}}_k^*$ is an eigenvector of $\boldsymbol{M}$ corresponding to λ_k^*, and conversely, if $\boldsymbol{u}_k^*$ is an eigenvector of $\boldsymbol{M}$ corresponding to λ_k^*, then $\|\boldsymbol{\theta}\|_2^{-1} \boldsymbol{H}^{-1} \boldsymbol{\Pi}^\top \boldsymbol{\Theta} \boldsymbol{u}_k^*$ is an eigenvector of $\boldsymbol{M}$ corresponding to $\tilde{\lambda}_k^*$.

Based on the previous Lemma, the following lemma, similar to Lemma B.1 of Jin et al. [2017], provides bounds of $\lambda_1^*, \cdots, \lambda_K^*$ under Assumption 1, which help us to prove the row-wise bounds for SCORE later. Notice that in this section, we use $C(c_0)$ to denote a constant depending on c_0.

LEMMA 2.4.2. *Under Assumption 1, the following statements are true*

(1) $\lambda_1^* \asymp \|\boldsymbol{\theta}\|_2^2$.

(2) $\lambda_1^* - |\lambda_2^*| \asymp \lambda_1^*$.

(3) $|\lambda_k^*| \asymp \|\boldsymbol{\theta}\|_2^2$, $2 \le k \le K$.

PROOF. Under Assumption 1, we can show that

$$C(c_0) \ge \lambda_1(\boldsymbol{B}) > |\lambda_2(\boldsymbol{B})| \ge \cdots \ge |\lambda_K(\boldsymbol{B})| \ge c_0 > 0,$$

$$C'(c_0) \le \|\boldsymbol{\theta}_k\|_2 / \|\boldsymbol{\theta}\|_2 \le C(c_0) \quad 1 \le k \le K.$$

Therefore, for each $1 \le k \le K$, by the definition of $\tilde{\lambda}_k^*$,

$$|\lambda_K(\boldsymbol{B})| \min_k \frac{\|\boldsymbol{\theta}_k\|_2^2}{\|\boldsymbol{\theta}\|_2^2} \le \left|\tilde{\lambda}_k^*\right| \le \lambda_1(\boldsymbol{B}) \max_k \frac{\|\boldsymbol{\theta}_k\|_2^2}{\|\boldsymbol{\theta}\|_2^2},$$

which by Lemma 2.4.1 means

$$C'(c_0)\|\boldsymbol{\theta}\|_2^2 \leq \left|\lambda_k^*\right| \leq C(c_0)\|\boldsymbol{\theta}\|_2^2.$$

Also by Theorem 1 of Han and Han [2019],

$$\lambda_1^* - |\lambda_2^*| \geq (1 - \tau(\boldsymbol{M})) \min_i \sum_{j=1}^n \boldsymbol{M}_{ij}$$

where

$$\tau(\boldsymbol{M}) = \frac{1 - \sqrt{\rho(\boldsymbol{M})}}{1 + \sqrt{\rho(\boldsymbol{M})}} \quad \text{and} \quad \rho(\boldsymbol{M}) = \min_{i,j,k,l} \frac{\boldsymbol{M}_{ik}\boldsymbol{M}_{jl}}{\boldsymbol{M}_{jk}\boldsymbol{M}_{il}}.$$

Notice that $\sum_{j=1}^n \boldsymbol{M}_{ij} \asymp \|\boldsymbol{\theta}\|_2^2$, thus we have for a constant $0 < c < 1$ depending on c_0,

$$\lambda_1^* - |\lambda_2^*| \geq c\lambda_1^*.$$

□

The following lemma from Jin et al. [2017] provides bounds for the entries of $\boldsymbol{u}_1^*$ and ℓ^2-norms of the rows of $\boldsymbol{U}^*$:

LEMMA 2.4.3 (Lemma B.2 of Jin et al. [2017]). *Under Assumption 1, the following statements are true*

(1) If we choose the sign of $\boldsymbol{u}_1^*$ such that $\sum_{i=1}^n u_1^*(i) > 0$, then the entries of $\boldsymbol{u}_1^*$ are positive satisfying $C^{-1}\theta_i/\|\boldsymbol{\theta}\|_2 \leq u_1^*(i) \leq C\theta_i/\|\boldsymbol{\theta}\|_2$, $1 \leq i \leq n$.

(2) $\|\boldsymbol{U}_{i\cdot}^*\|_2 \leq C\sqrt{K}\theta_i/\|\boldsymbol{\theta}\|_2$, $1 \leq i \leq n$.

2.4.2. Row-Wise Perturbation for SCORE. Let's come back to community detection with SCORE. Let $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_K|$ be the leading K eigenvalues of $\boldsymbol{A}$ in magnitude, with corresponding eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K$. Denote $\boldsymbol{U} = [\boldsymbol{u}_2, \ldots, \boldsymbol{u}_K] \in \mathbb{R}^{n\times(K-1)}$. Define $\boldsymbol{R}^{(K)}$ as an $n \times (K-1)$ matrix constructed from the eigenvector $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, ..., $\boldsymbol{u}_K$ by

taking entrywise ratios between $\boldsymbol{u}_2$,..., $\boldsymbol{u}_K$ and $\boldsymbol{u}_1$, i.e., $R^{(K)}(i,k) = u_{k+1}(i)/u_1(i)$ for $1 \le i \le n$ and $0 \le k \le K-1$. For any $2 \le m \le K$, let $\boldsymbol{R}^{(m)}$ be an $n \times (m-1)$ matrix consists of the first $m-1$ columns of $\boldsymbol{R}^{(K)}$. For any candidate number of clusters m, SCORE amounts to conducting k-means clustering on the row vectors of $\boldsymbol{R}^{(m)}$. Therefore, we need to study the statistical properties of the rows of $\boldsymbol{R}^{(m)}$.

As indicated in Jin et al. [2022], to establish the NSP for SCORE under the general DCSBM, one needs to show that there exists some $(K-1)\times(K-1)$ orthogonal matrix $\boldsymbol{Q}$, such that $\boldsymbol{R}^{(m)}$ is comparable to $\boldsymbol{R}^{*(m)}(\boldsymbol{Q})$. Then roughly speaking, the NSP of k-means clustering on the rows of $\boldsymbol{R}^{*(m)}(\boldsymbol{Q})$ can be extended to that of $\boldsymbol{R}^{(m)}$. Here we follow the notations in Jin et al. [2022] to construct $\boldsymbol{R}^{*(m)}(\boldsymbol{Q})$. Denote $\mathcal{O}_{K-1}$ as the space of $(K-1)\times(K-1)$ orthogonal matrices. For any $\boldsymbol{Q} \in \mathcal{O}_{K-1}$, and any $2 \le k \le K$, let $\boldsymbol{u}_k^*(\boldsymbol{Q})$ be the $(k-1)$-th column of $[\boldsymbol{u}_2^*, \ldots, \boldsymbol{u}_K^*]\boldsymbol{Q}$. Define $\boldsymbol{R}^{*(K)}(\boldsymbol{Q}) \in \mathbb{R}^{n\times(K-1)}$, whose $(k-1)$-th column is the entrywise ratio between $\boldsymbol{u}_k^*(\boldsymbol{Q})$ and $\boldsymbol{u}_1^*$. For any $2 \le m \le K$, let $\boldsymbol{R}^{*(m)}(\boldsymbol{Q}) \in \mathbb{R}^{n\times(m-1)}$ consist of the first $m-1$ columns of $\boldsymbol{R}^{*(K)}(\boldsymbol{Q})$.

We will investigate the properties of the rows of $\boldsymbol{R}^{*(m)}(\boldsymbol{Q})$ later. In this subsection we focus on the comparison between $\boldsymbol{R}^{(m)}$ and $\boldsymbol{R}^{*(m)}(\boldsymbol{Q})$. More specifically, we intend to show that, with high probability, there exists some $\boldsymbol{Q} \in \mathcal{O}_{K-1}$, such that the row-wise deviation between $\boldsymbol{R}^{(m)}$ and $\boldsymbol{R}^{*(m)}(\boldsymbol{Q})$ is well-controlled. Before showing the row-wise deviation bounds in Lemma 2.4.5, we first show a supporting lemma of eigenvector perturbation bounds.

LEMMA 2.4.4. *Under Assumption 1, with probability* $1-O(n^{-3})$*, the following statements are true:*

- *We can select* $\boldsymbol{u}_1$ *such that* $\|\boldsymbol{u}_1 - \boldsymbol{u}_1^*\|_\infty \le o(1/\sqrt{n})$.
- $\|\boldsymbol{U}\boldsymbol{Q}^\top - \boldsymbol{U}^*\|_{2\to\infty} \le o(1/\sqrt{n})$ *for some* $\boldsymbol{Q} \in \mathcal{O}_{K-1}$.

PROOF. Divide $\lambda_1^*, \cdots, \lambda_K^*$ into three groups: (1) λ_1^*, (2) positive values in $\lambda_2^*, \cdots, \lambda_K^*$, and (3) negative values in $\lambda_2^*, \cdots, \lambda_K^*$. We shall apply Theorem A.1.1 to all the three groups.

For succinctness, we only show in detail applying the theorem to group (2), and the proof for the other two groups is similar and thus omitted.

Denote K_1 as the number of eigenvalues in group (2). Define $\mathbf{\Lambda}_1^*$ as the diagonal matrix consisting of eigenvalues in group (2), and $\boldsymbol{U}_1^*$ as the matrix whose columns are the associated eigenvectors. Define the two matrices' empirical counterparts as $\mathbf{\Lambda}_1$ and $\boldsymbol{U}_1$. To show the second bullet point, we aim to first show with high probability that

(i) $\|\boldsymbol{U}_1\boldsymbol{Q}^\top - \boldsymbol{A}\boldsymbol{U}_1^*(\mathbf{\Lambda}_1^*)^{-1}\|_{2\to\infty} \le o(1/\sqrt{n})$ for some $\boldsymbol{Q} \in \mathcal{O}_{K_1-1}$.

(ii) $\|\boldsymbol{U}_1^* - \boldsymbol{A}\boldsymbol{U}_1^*(\mathbf{\Lambda}_1^*)^{-1}\|_{2\to\infty} \le o(1/\sqrt{n})$.

Proof of (i). To apply Theorem A.1.1, we need to determine γ and $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$, and then verify Assumption (A1)—(A4) as required by the theorem. Note that by Lemma 2.4.2, we have

$$\Delta^* = \min\{\lambda_1^* - \lambda_2^*, |\lambda_K^*|\} \ge C(c_0)n\theta_{\min}^2$$

and $\kappa \le C(c_0)$. We choose an appropriately large $C_1(c_0)$ and let

$$\gamma = \frac{C_1(c_0)}{\sqrt{\log n}}.$$

A1 This is trivial since

$$\|\boldsymbol{M}\|_{2\to\infty} \le c_0^{-2}\sqrt{n}\theta_{\min}^2 \le \Delta^*\gamma$$

when n is sufficiently large.

A2 This obviously holds.

A3 In the proof of A4, we show that $\varphi(\gamma) \le C(c_0)(\log n)^{-3/2}$, therefore $32\kappa\max\{\gamma, \varphi(\gamma)\} \le 1$ when n is sufficiently large.

Next, we show the spectral norm perturbation bound by applying the subexponential case of matrix Bernstein inequality (Theorem 6.2 of Tropp [2012], restated as Lemma A.1.4 in this dissertation). Let $\boldsymbol{X}^{ij} = (A_{ij} - M_{ij})(\boldsymbol{E}^{ij} + \boldsymbol{E}^{ji})$ for $i < j$, and $\boldsymbol{X}^{ii} = (A_{ii} - M_{ii})\boldsymbol{E}^{ii}$ for $i = 1 \cdots, n$, where $\boldsymbol{E}^{ij}$ is a $n \times n$ matrix with 1 on the (i, j)-th

entry and 0 elsewhere. Obviously, $\mathbb{E}[\boldsymbol{X}^{ij}] = \mathbf{0}$. Notice that

$$\mathbb{E}[(\boldsymbol{X}^{ij})^p] = \begin{cases} \mathbb{E}[(A_{ij} - M_{ij})^p](\boldsymbol{E}^{ij} + \boldsymbol{E}^{ji}) & \text{when } p \text{ is odd;} \\ \mathbb{E}[(A_{ij} - M_{ij})^p](\boldsymbol{E}^{ij} + \boldsymbol{E}^{ji}) & \text{when } p \text{ is even.} \end{cases}$$

Also note that $-(\boldsymbol{E}^{ii} + \boldsymbol{E}^{jj}) \preceq \boldsymbol{E}^{ij} + \boldsymbol{E}^{ji} \preceq \boldsymbol{E}^{ii} + \boldsymbol{E}^{jj}$. Then by Assumption (2.18), for integer $p \geq 2$, we have

$$\left|\mathbb{E}[(A_{ij} - M_{ij})^p]\right| \leq \mathbb{E}[|A_{ij} - M_{ij}|^p] \leq C' \left(\frac{p!}{2}\right) R^{p-2} M_{ij}$$

where C' and R only depend on c_0. Then

$$\mathbb{E}[(\boldsymbol{X}^{ij})^p] \preceq C' \frac{p!}{2} R^{p-2} M_{ij} (\boldsymbol{E}^{ii} + \boldsymbol{E}^{jj}) \text{ for } p = 2, 3, 4, \ldots.$$

Thus the conditions of Lemma A.1.4 are verified. Notice that $\sum_{i \leq j} \boldsymbol{X}^{ij} = \boldsymbol{A} - \boldsymbol{M}$. Denote

$$\sigma^2 = \left\| \sum_{i \leq j} C' M_{ij} (\boldsymbol{E}^{ii} + \boldsymbol{E}^{jj}) \right\| = C' \left(\max_i \sum_{j=1}^n M_{ij} \right) \lesssim n\theta_{\min}^2,$$

where $\lesssim$ only hides a constant depending on c_0. Then for all $t \geq 0$,

$$\mathbb{P}\left(\|\boldsymbol{A} - \boldsymbol{M}\| \geq t\right) \leq n \exp\left(-\frac{t^2}{2(\sigma^2 + Rt)}\right) \leq n \exp\left(-\frac{1}{4}\left(\frac{t^2}{\sigma^2} \wedge \frac{t}{R}\right)\right).$$

By the assumption $\theta_{\min} \geq C_0 \sqrt{\frac{\log^4 n}{n}}$, when n is sufficiently large, $n\theta_{\min}^2 \gg \log n$, so we can take $t = C(c_0)\theta_{\min}\sqrt{n \log n}$ for a sufficiently large $C(c_0)$. Then, with probability $1 - O(n^{-3})$,

$$\|\boldsymbol{A} - \boldsymbol{M}\| \leq C(c_0)\theta_{\min}\sqrt{n \log n}.$$

A4 By Lemma A.1.3, for any $1 \le i \le n$ and $\boldsymbol{W} \in \mathbb{R}^{n\times r}$, with probability at least $1 - n^{-4}$, there holds

$$\|(\boldsymbol{A} - \boldsymbol{M})_{i\cdot}\boldsymbol{W}\|_2 \le C_2(c_0) \max\left(\theta_{\min}\|\boldsymbol{W}\|_F\sqrt{\log n}, \|\boldsymbol{W}\|_{2\to\infty}(\log n)\right) \tag{2.32}$$

for a sufficiently large $C_2(c_0)$. Since $\Delta^* \ge C(c_0)n\theta_{\min}^2$, we have

$$\|(\boldsymbol{A} - \boldsymbol{M})_{i\cdot}\boldsymbol{W}\|_2 \le C_2(c_0)\Delta^*\|\boldsymbol{W}\|_{2\to\infty} \max\left(\frac{\sqrt{n\log n}}{n\theta_{\min}}\frac{\|\boldsymbol{W}\|_F}{\sqrt{n}\|\boldsymbol{W}\|_{2\to\infty}}, \frac{\log n}{n\theta_{\min}^2}\right).$$

Now we follow similar arguments as in the proof of Lemma 2.1 of Jin et al. [2017]. Define the quantities $t_1 = C_2(c_0)\,(n\theta_{\min})^{-1}\sqrt{n\log n}$ and $t_2 = C_2(c_0)\,(n\theta_{\min}^2)^{-1}\log n$. Define the function

$$\widetilde{\varphi}(x) = \min(t_1 x, t_2)\,.$$

Then we have for all $1 \le i \le n$, with probability at least $1 - n^{-4}$,

$$\|(\boldsymbol{A} - \boldsymbol{M})_{i\cdot}\boldsymbol{W}\|_2 \le \Delta^*\|\boldsymbol{W}\|_{2\to\infty}\widetilde{\varphi}\left(\frac{\|\boldsymbol{W}\|_F}{\sqrt{n}\|\boldsymbol{W}\|_{2\to\infty}}\right). \tag{2.33}$$

First, notice that $(\sqrt{n}\|\boldsymbol{W}\|_{2\to\infty})^{-1}\|\boldsymbol{W}\|_F \in \left[n^{-1/2}, 1\right]$. Then, observe that since $\theta_{\min} \ge C_0\sqrt{\frac{\log^4 n}{n}}$, when n is sufficiently large, we have $t_2/t_1 = (\sqrt{n}\theta_{\min})^{-1}\sqrt{\log n} > n^{-1/2}$ and $t_1 \gg t_2$. Thus when $x \in \left[n^{-1/2}, t_2/t_1\right]$, $t_1 x \le t_2$, i.e., $\widetilde{\varphi}(x) = t_2$; when $x \in (t_2/t_1, 1]$, $t_1 x > t_2$, i.e., $\widetilde{\varphi}(x) = t_1 x$. Therefore, we can construct $\varphi(\cdot)$ as:

$$\varphi(x) = \begin{cases} \sqrt{n}t_2 x & \text{for } 0 \le x \le n^{-1/2}; \\ t_2 & \text{for } n^{-1/2} < x \le t_2/t_1; \\ t_1 x & \text{for } t_2/t_1 < x \le 1; \\ t_1 & \text{for } x > 1. \end{cases}$$

Obviously, $\varphi(x)$ is continuous and non-decreasing in $\mathbb{R}_+$, with $\varphi(0) = 0$ and $\varphi(x)/x$ being non-increasing in $\mathbb{R}_+$. By (2.33) and $0 \le \widetilde{\varphi}(x) \le \varphi(x)$, we have with probability at least $1 - n^{-4}$,

$$\|(\boldsymbol{A} - \boldsymbol{M})_{i\cdot}\boldsymbol{W}\|_2 \le \Delta^* \|\boldsymbol{W}\|_{2\to\infty} \varphi\left(\frac{\|\boldsymbol{W}\|_F}{\sqrt{n}\|\boldsymbol{W}\|_{2\to\infty}}\right).$$

Based on the definition of t_1 and the assumption $\theta_{\min} \ge C_0\sqrt{\frac{\log^4 n}{n}}$, we have $t_1 \lesssim (\log n)^{-3/2}$ where $\lesssim$ only hides a constant depending on C_0 and c_0. Furthermore, since $\varphi(x) \le t_1$, we have $\varphi(\gamma) \le t_1 \lesssim (\log n)^{-3/2}$.

After verifying (A1)—(A4), we obtain the bound (A.1). Note that based on the definition of $\varphi(x)$, $\kappa(\kappa + \varphi(1)) \le C(c_0)$ and $\gamma + \varphi(\gamma) \le \frac{C(c_0)}{\sqrt{\log n}}$. Also by Lemma 2.4.3, $\|\boldsymbol{U}_1^*\|_{2\to\infty} \le \frac{C(c_0)}{\sqrt{n}}$. Since $\|\boldsymbol{M}\|_{2\to\infty} \le c_0^{-3}\sqrt{n}\theta_{\min}^2$ and $\Delta^* \ge C(c_0)n\theta_{\min}^2$, $\|\boldsymbol{M}\|_{2\to\infty}/\Delta^* \le \frac{C(c_0)}{\sqrt{n}}$. Finally, we obtain that with probability $1 - O(n^{-3})$,

$$\begin{aligned}\|\boldsymbol{U}_1\boldsymbol{Q} - \boldsymbol{A}\boldsymbol{U}_1^*(\boldsymbol{\Lambda}_1^*)^{-1}\|_{2\to\infty} &\le C\kappa(\kappa + \varphi(1))(\gamma + \varphi(\gamma))\|\boldsymbol{U}_1^*\|_{2\to\infty} + \gamma\|\boldsymbol{M}\|_{2\to\infty}/\Delta^* \\ &\lesssim \frac{1}{\sqrt{n\log n}},\end{aligned}$$

where $\lesssim$ only hides a constant depending on c_0 in Assumption 1.

Proof of (ii). Based on the fact $\boldsymbol{U}_1^* = \boldsymbol{M}\boldsymbol{U}_1^*(\boldsymbol{\Lambda}_1^*)^{-1}$, we have

$$\|\boldsymbol{U}_1^* - \boldsymbol{A}\boldsymbol{U}_1^*(\boldsymbol{\Lambda}_1^*)^{-1}\|_{2\to\infty} = \|(\boldsymbol{M} - \boldsymbol{A})\boldsymbol{U}_1^*(\boldsymbol{\Lambda}_1^*)^{-1}\|_{2\to\infty}.$$

By Lemma 2.4.3, $\left\|(\boldsymbol{U}_1^*)_{i\cdot}\right\|_2 \le C\theta_i/\|\boldsymbol{\theta}\|_2 \le \frac{C(c_0)}{\sqrt{n}}$, thus $\|\boldsymbol{U}_1^*\|_F \le C(c_0)$. Also since $\theta_{\min} \ge C_0\sqrt{\frac{\log^4 n}{n}}$, we have $(\sqrt{n}\theta_{\min})^{-1} \le C_0^{-1}(\log n)^{-2}$. Apply (2.32) with $\boldsymbol{W} = \boldsymbol{U}_1^*$ and combine the

above facts, we have with probability $1 - O(n^{-3})$,

$$\begin{aligned}
\|(\boldsymbol{A}-\boldsymbol{M})\boldsymbol{U}_1^*(\boldsymbol{\Lambda}_1^*)^{-1}\|_{2\to\infty} &\le \left(\max_{1\le i\le n}\|(\boldsymbol{A}-\boldsymbol{M})_{i\cdot}\boldsymbol{U}_1^*\|_2\right)\cdot\|(\boldsymbol{\Lambda}_1^*)^{-1}\| \\
&\le C(c_0)\max\left(\theta_{\min}\|\boldsymbol{U}_1^*\|_F\sqrt{\log n}, \|\boldsymbol{U}_1^*\|_{2\to\infty}(\log n)\right)\cdot(n\theta_{\min}^2)^{-1} \\
&\le C(c_0)\max\left((\sqrt{n}\theta_{\min})^{-1}\|\boldsymbol{U}_1^*\|_F\sqrt{\frac{\log n}{n}}, (\sqrt{n}\theta_{\min})^{-2}\|\boldsymbol{U}_1^*\|_{2\to\infty}(\log n)\right) \\
&\lesssim \frac{1}{\sqrt{n\log^3 n}},
\end{aligned}$$

where $\lesssim$ only hides a constant depending on C_0 and c_0 in Assumption 1.

With (i) and (ii), we have $\|\boldsymbol{U}_1\boldsymbol{Q}^\top - \boldsymbol{U}_1^*\|_{2\to\infty} \le o(1/\sqrt{n})$ for some $\boldsymbol{Q} \in \mathcal{O}_{K_1-1}$. For eigenvalues in group (3), we similarly define K_2, $\boldsymbol{\Lambda}_2^*$, $\boldsymbol{U}_2^*$, $\boldsymbol{\Lambda}_2$ and $\boldsymbol{U}_2$, then there holds $\|\boldsymbol{U}_2\boldsymbol{Q}^\top - \boldsymbol{U}_2^*\|_{2\to\infty} \le o(1/\sqrt{n})$ for some $\boldsymbol{Q} \in \mathcal{O}_{K_2-1}$. Combining these results yields the second bullet point: With probability $1 - O(n^{-3})$, we have

$$\|\boldsymbol{U}\boldsymbol{Q}^\top - \boldsymbol{U}^*\|_{2\to\infty} \le o(1/\sqrt{n}) \text{ for some } \boldsymbol{Q} \in \mathcal{O}_{K-1}.$$

To show the first bullet point, we apply Theorem A.1.1 to group (1) with $s = 0$ and $r = 1$. Note that by Lemma 2.4.2, we have

$$\Delta^* = \min\{\lambda_1^*, \lambda_1^* - \lambda_2^*\} \ge C(c_0)n\theta_{\min}^2$$

and $\kappa \le C(c_0)$. Following similar procedures, we can select $\boldsymbol{u}_1$ such that

(i) $\|\boldsymbol{u}_1 - \boldsymbol{A}\boldsymbol{u}_1^*/\lambda_1^*\|_\infty \le o(1/\sqrt{n})$.

(ii) $\|\boldsymbol{u}_1^* - \boldsymbol{A}\boldsymbol{u}_1^*/\lambda_1^*\|_\infty \le o(1/\sqrt{n})$.

Thus with probability $1 - O(n^{-3})$, we have

$$\|\boldsymbol{u}_1 - \boldsymbol{u}_1^*\|_\infty \le o(1/\sqrt{n}),$$

which proves the first bullet point.

□

LEMMA 2.4.5 (Row-wise deviation bound). *For* $2 \le m \le K$, *denote* $\left(\boldsymbol{r}_i^{(m)}\right)^\top$ *and* $\left(\boldsymbol{r}_i^{*(m)}(\boldsymbol{Q})\right)^\top$ *as the* i*-th row of* $\boldsymbol{R}^{(m)}$ *and* $\boldsymbol{R}^{*(m)}(\boldsymbol{Q})$, *respectively. Under Assumption 1, with probability* $1 - O(n^{-3})$, *there exists a* $(K-1)\times(K-1)$ *orthogonal matrix* $\boldsymbol{Q}$, *such that*

$$\|\boldsymbol{r}_i^{(m)} - \boldsymbol{r}_i^{*(m)}(\boldsymbol{Q})\|_2 \le \|\boldsymbol{r}_i^{(K)} - \boldsymbol{r}_i^{*(K)}(\boldsymbol{Q})\|_2 \le o(1) \tag{2.34}$$

for all $2 \le m \le K$ *and* $1 \le i \le n$.

PROOF. By Lemma 2.4.4, with probability $1 - O(n^{-3})$, we have

- We can select $\boldsymbol{u}_1$ such that $\|\boldsymbol{u}_1 - \boldsymbol{u}_1^*\|_\infty \le o(1/\sqrt{n})$.
- $\|\boldsymbol{U}\boldsymbol{Q}^\top - \boldsymbol{U}^*\|_{2\to\infty} \le o(1/\sqrt{n})$ for some $\boldsymbol{Q} \in \mathcal{O}_{K-1}$.

First, $\|\boldsymbol{r}_i^{(m)} - \boldsymbol{r}_i^{*(m)}(\boldsymbol{Q})\|_2 \le \|\boldsymbol{r}_i^{(K)} - \boldsymbol{r}_i^{*(K)}(\boldsymbol{Q})\|_2$ obviously holds. For convenience, write $\boldsymbol{r}_i = \boldsymbol{r}_i^{(K)}$, $\boldsymbol{r}_i^*(\boldsymbol{Q}) = \boldsymbol{r}_i^{*(K)}(\boldsymbol{Q})$ and $\boldsymbol{r}_i^* = \boldsymbol{r}_i^{*(K)}(\boldsymbol{I}_{K-1})$. Notice that by the definitions, $\boldsymbol{r}_i = \frac{1}{u_1(i)}\boldsymbol{U}_{i\cdot}$, $\boldsymbol{r}_i^* = \frac{1}{u_1^*(i)}\boldsymbol{U}_{i\cdot}^*$ and $\boldsymbol{r}_i^*(\boldsymbol{Q}) = \frac{1}{u_1^*(i)}\boldsymbol{Q}^\top\boldsymbol{U}_{i\cdot}^*$. Then we have

$$\|\boldsymbol{r}_i - \boldsymbol{r}_i^*(\boldsymbol{Q})\|_2 = \|\boldsymbol{r}_i - \boldsymbol{Q}^\top\boldsymbol{r}_i^*\|_2 = \|\boldsymbol{Q}\boldsymbol{r}_i - \boldsymbol{r}_i^*\|_2$$

and

$$\begin{aligned}
\boldsymbol{Q}\boldsymbol{r}_i - \boldsymbol{r}_i^* &= \frac{1}{u_1(i)}\boldsymbol{Q}\boldsymbol{U}_{i\cdot} - \frac{1}{u_1^*(i)}\boldsymbol{U}_{i\cdot}^* \\
&= \frac{1}{u_1(i)}\left(\boldsymbol{Q}\boldsymbol{U}_{i\cdot} - \boldsymbol{U}_{i\cdot}^*\right) + \left(\frac{1}{u_1(i)} - \frac{1}{u_1^*(i)}\right)\boldsymbol{U}_{i\cdot}^* \\
&= \frac{1}{u_1(i)}\left(\boldsymbol{Q}\boldsymbol{U}_{i\cdot} - \boldsymbol{U}_{i\cdot}^*\right) + \frac{u_1^*(i) - u_1(i)}{u_1(i)}\boldsymbol{r}_i^*.
\end{aligned}$$

By Lemma 2.4.3, $\|\boldsymbol{U}_{i\cdot}^*\|_2 \le \frac{C(c_0)}{\sqrt{n}}$, and $\frac{C'(c_0)}{\sqrt{n}} \le u_1^*(i) \le \frac{C(c_0)}{\sqrt{n}}$ for $1 \le i \le n$. Thus we have $\|\boldsymbol{r}_i^*\|_2 \le C(c_0)$. Also since $\|\boldsymbol{u}_1 - \boldsymbol{u}_1^*\|_\infty \le o(1/\sqrt{n})$, $u_1(i) \ge \frac{C(c_0)}{\sqrt{n}}$ for $1 \le i \le n$. Plug the above

facts into the previous equation and we get

$$\begin{aligned}\max_{1\le i\le n}\|\boldsymbol{Q}\boldsymbol{r}_i-\boldsymbol{r}_i^*\|_2 &\lesssim \max_{1\le i\le n}\sqrt{n}\left(\|\boldsymbol{Q}\boldsymbol{U}_{i\cdot}-\boldsymbol{U}_{i\cdot}^*\|_2+C(c_0)|u_1^*(i)-u_1(i)|\right)\\ &\le \sqrt{n}\left(\|\boldsymbol{U}\boldsymbol{Q}^\top-\boldsymbol{U}^*\|_{2\to\infty}+C(c_0)\|\boldsymbol{u}_1-\boldsymbol{u}_1^*\|_\infty\right)\\ &\le o(1).\end{aligned}$$

Here $\lesssim$ only hides a constant depending on c_0. □

2.4.3. Proof of Theorem 2.2.2. Before proving the NSP in Theorem 2.2.2, it remains to study the geometry underlying the rows of $\boldsymbol{R}^{*(m)}(\boldsymbol{Q})$ for any $\boldsymbol{Q}\in\mathcal{O}_{K-1}$. Recall that $\tilde{\boldsymbol{u}}_k^*$ is the associated eigenvector of the k-th largest eigenvalue of $\boldsymbol{HBH}$ in magnitude. For any $\boldsymbol{Q}\in\mathcal{O}_{K-1}$, and any $2\le k\le K$, let $\tilde{\boldsymbol{u}}_k^*(\boldsymbol{Q})$ be the $(k-1)$th column of $[\tilde{\boldsymbol{u}}_2^*,\ldots,\tilde{\boldsymbol{u}}_K^*]\boldsymbol{Q}$. Define $\widetilde{\boldsymbol{R}}^{*(K)}(\boldsymbol{Q})\in\mathbb{R}^{K\times(K-1)}$, whose $(k-1)$th column is the entrywise ratio between $\tilde{\boldsymbol{u}}_k^*(\boldsymbol{Q})$ and $\tilde{\boldsymbol{u}}_1^*$. For any $2\le m\le K$, $\widetilde{\boldsymbol{R}}^{*(m)}(\boldsymbol{Q})\in\mathbb{R}^{K\times(m-1)}$ consists of the first $m-1$ columns of $\widetilde{\boldsymbol{R}}^{*(K)}(\boldsymbol{Q})$. For each $1\le k\le K$, denote $\left(\tilde{\boldsymbol{r}}_k^{*(m)}(\boldsymbol{Q})\right)^\top$ as the k-th row of $\widetilde{\boldsymbol{R}}^{*(m)}(\boldsymbol{Q})$.

The following lemma is a result from Jin et al. [2022], which characterizes the relationship between the rows of $\boldsymbol{R}^{*(m)}(\boldsymbol{Q})$ and $\widetilde{\boldsymbol{R}}^{*(m)}(\boldsymbol{Q})$.

LEMMA 2.4.6 (Lemma 4.2 of Jin et al. [2022]). *Under Assumption 1, for any $\boldsymbol{Q}\in\mathcal{O}_{K-1}$ and for each $i\in[n]$ and $k\in[K]$, we have $\boldsymbol{r}_i^{*(m)}(\boldsymbol{Q})=\tilde{\boldsymbol{r}}_k^{*(m)}(\boldsymbol{Q})$ if $i\in\mathcal{N}_k$.*

Next, we introduce the following concept from Jin et al. [2022] which defines a metric for the relative positions of the K cluster centers (rows of $\widetilde{\boldsymbol{R}}^{*(m)}(\boldsymbol{Q})$).

DEFINITION 2 (Definition 4.1 of Jin et al. [2022]). Fixing $K>1$ and $1<m\le K$, consider a $K\times(m-1)$ matrix $\boldsymbol{U}=[\boldsymbol{u}_1,\ldots,\boldsymbol{u}_K]^\top$. First, let $d_K(\boldsymbol{U})$ be the minimum pairwise distance of all K rows. Second, let $\boldsymbol{u}_k$ and $\boldsymbol{u}_l$ $(k<l)$ be the pair that satisfies $\|\boldsymbol{u}_k-\boldsymbol{u}_l\|=d_K(\boldsymbol{U})$ (if this holds for multiple pairs, pick the first pair in the lexicographical order). Remove row

l from the matrix $\boldsymbol{U}$ and let $d_{K-1}(\boldsymbol{U})$ be the minimum pairwise distance for the remaining $(K-1)$ rows. Repeat this step and define $d_{K-2}(\boldsymbol{U})$, ..., $d_2(\boldsymbol{U})$ recursively.

The following lemma provides a uniform lower bound for $d_m(\widetilde{\boldsymbol{R}}^{*(m)}(\boldsymbol{Q}))$.

LEMMA 2.4.7 (Lemma 4.3 of Jin et al. [2022]). *Under Assumption 1, for each $2 \le m \le K$, there exists a constant $C_m > 0$ which may depend on m, such that*

$$d_m(\widetilde{\boldsymbol{R}}^{*(m)}(\boldsymbol{Q})) > C_m$$

for any $\boldsymbol{Q} \in \mathcal{O}_{K-1}$.

Now, we introduce the new k-means theorem for SCORE proposed in Jin et al. [2022].

LEMMA 2.4.8 (Theorem 4.1 of Jin et al. [2022]). *Fix $1 < m \le K$ and let n be sufficiently large. Consider the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ that take only K values in $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K$. Write $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K]^\top$. Let $F_k = \{1 \le i \le n : \boldsymbol{x}_i = \boldsymbol{y}_k\}$ for $1 \le k \le K$. Suppose for some constants $0 < \alpha_0 < 1$ and $\beta_0 > 0$, $\min_{1\le k\le K} |F_k| \ge \alpha_0 n$ and $\max_{1\le k\le K} \|\boldsymbol{y}_k\|_2 \le \beta_0 d_m(\boldsymbol{Y})$. We apply the k-means clustering to a set of n points $\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_n$ assuming $\le m$ clusters, and denote by $\widehat{S}_1, \ldots, \widehat{S}_m$ the obtained clusters. There exists a constant $c > 0$, which only depends on (α_0, β_0, m), such that, if $\max_{1\le i\le n} \|\hat{\boldsymbol{x}}_i - \boldsymbol{x}_i\|_2 \le c d_m(\boldsymbol{Y})$, then $\#\{1 \le j \le m : \widehat{S}_j \cap F_k \ne \phi\} = 1$ for each $1 \le k \le K$.*

With the above results, we are ready to establish the NSP for SCORE under the general DCSBM. By Lemma 2.4.5, with probability $1 - O(n^{-3})$, there exists a $(K-1) \times (K-1)$ orthogonal matrix $\boldsymbol{Q}$, such that for each $2 \le m \le K$,

$$\max_{1\le i\le n} \|\boldsymbol{r}_i^{(m)} - \boldsymbol{r}_i^{*(m)}(\boldsymbol{Q})\|_2 \le o(1). \tag{2.35}$$

To prove Theorem 2.2.2, we apply Lemma 2.4.8 with $\boldsymbol{Y} = \widetilde{\boldsymbol{R}}^{*(m)}(\boldsymbol{Q})$, $\boldsymbol{x}_i = r_i^{*(m)}(\boldsymbol{Q})$, $\hat{\boldsymbol{x}}_i = r_i^{(m)}$ and $F_k = \mathcal{N}_k$. By Assumption 1, $\min_{1\le k\le K} |F_k| \ge c_0 n$. As shown in the proof

of Lemma 2.4.5, $\|\boldsymbol{x}_i\|_2 \le \|r_i^*(\boldsymbol{Q})\|_2 = \|r_i^*\|_2 \le C(c_0)$. Combined with Lemma 2.4.6, we have $\max_{1\le k\le K}\|\boldsymbol{y}_k\|_2 \le C(c_0)$, which by Lemma 2.4.7 further means $\max_{1\le k\le K}\|\boldsymbol{y}_k\|_2 \le C(c_0)d_m(\boldsymbol{Y})$.

Also by (2.35), we have $\max_{1\le i\le n}\|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|_2 \le c\cdot d_m(\boldsymbol{Y})$ for a sufficiently small constant c. The claim follows by applying Lemma 2.4.8.

CHAPTER 3

Spectral Divergence-Based Rank Selection for Network Data

3.1. Methodology

Consider an undirected binary-edge network with independent edges and no self-loop, in which n nodes are represented by $1, \ldots, n$. In this case, the entries of the adjacency matrix are distributed as

$$A_{ij} \overset{\text{indep}}{\sim} \text{Bernoulli}(M_{ij}), \quad 1 \le i < j \le n. \tag{3.1}$$

In addition, $A_{ji} = A_{ij}$ and $A_{ii} = 0$ so $\boldsymbol{A}$ is a symmetric matrix. Here the expected adjacency matrix $\boldsymbol{M} = (M_{ij})_{\{1 \le i,j \le n\}}$ is a $n \times n$ symmetric matrix whose entries are between 0 and 1. In the literature, the mean matrix $\boldsymbol{M}$ is sometimes referred to as the graphon (e.g. Gao et al. [2015]), but we simply refer to it as the expected adjacency matrix. Since there is no self-loop, the random graph does not rely on the diagonal entries of $\boldsymbol{M}$. Therefore, the diagonal entries of $\boldsymbol{M}$ can be defined arbitrarily for the convenience of discussions without loss of generality.

Our assumption is that there are K communities within the observed network, and we want to select K accurately. In various widely used generative models to characterize the communities in a network, the number of communities is exactly the same as the rank of the expected adjacency matrix (with its diagonal entries appropriately chosen). In SBM, the nodes are assumed to belong to K communities, and the nodes i and j are connected with probability $M_{ij} = B_{\phi(i)\phi(j)}$, where $\phi(\cdot)$ is the labeling mapping. It is straightforward to verify that the rank of $\boldsymbol{M}$ is K. SBM has various extensions and the most important one of them is DCSBM, in which $M_{ij} = \theta_i \theta_j B_{\phi(i)\phi(j)}$ with degree-correction parameters $\theta_1, \cdots, \theta_n$; another

one is the Overlapping Continuous Community Assignment Model (OCCAM) [Zhang et al., 2014], in which $M_{ij} = \theta_i \theta_j \boldsymbol{\pi}_i^\top \boldsymbol{B} \boldsymbol{\pi}_j$, and the K dimensional vectors $\boldsymbol{\pi}_i$ represents the propensities of node i to the K groups. Again, it is clear that in all the above models, the rank of $\boldsymbol{M}$ is K. Motivated by these examples, we treat the problem of estimating the number of communities as a rank selection problem.

Instead of solving the rank selection problem directly, let's consider a related problem: For any candidate rank r, if we first perform spectral decomposition for the adjacency matrix, denoted as $\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top$ with $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n| > 0$, and then propose to estimate the expected adjacency matrix $\boldsymbol{M}$ through spectral truncation

$$\widehat{\boldsymbol{M}}_r = \sum_{i=1}^{r} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top,$$

then, can we evaluate how close this estimate is to the ground truth $\boldsymbol{M}$? In other words, is there a way to estimate the discrepancy between $\widehat{\boldsymbol{M}}_r$ and $\boldsymbol{M}$ from the data? If this is feasible, we can simply choose the rank r such that the estimated discrepancy is the smallest.

Classical methods such as cross-validation are usually computationally expensive, and might be sensitive to the choice of hyperparameters. Instead, we will apply the risk estimation framework proposed in Efron et al. [2004]. In particular, one interesting result we use from Efron et al. [2004] is that for general mean estimators of independent Bernoulli variables, as long as the error measures are derived from the binomial deviation, the optimism of the apparent error over the true error can be approximately estimated by the divergence of the estimator with respect to the data. Although this approximation still lacks serious theoretical justification, empirically we have found that this idea is often effective and efficient when applied to real-world data.

Now let's introduce the risk estimation framework proposed in Efron et al. [2004]. For a mean estimator $\widehat{\boldsymbol{M}}$, we first choose an error measure $q(\widehat{\boldsymbol{M}}, \boldsymbol{A})$ based on the binomial deviance to quantify the discrepancy between the estimator and the original data matrix, and refer

to it as the *apparent error*. This apparent error cannot be used directly for model selection, since the data is used twice: one for the calculation of the estimator, and the other for the evaluation of such estimator. In contrast, we define a companion true error $Q(\widehat{\boldsymbol{M}}, \boldsymbol{M})$ to quantify the discrepancy between $\widehat{\boldsymbol{M}}$ and $\boldsymbol{M}$. We will explain later in detail how to define the true error $Q(\cdot, \cdot)$ based on the apparent error $q(\cdot, \cdot)$.

Ideally, we should use $Q(\widehat{\boldsymbol{M}}, \boldsymbol{M})$ to quantify the discrepancy between $\widehat{\boldsymbol{M}}$ and $\boldsymbol{M}$. However, this is usually not applicable since $\boldsymbol{M}$ is unknown. Then, the key question is how to estimate the true error based on the apparent error. In fact, the true error $Q(\widehat{\boldsymbol{M}}, \boldsymbol{M})$ is usually higher than the apparent error, and the difference is called optimism, i.e.,

$$Q(\widehat{\boldsymbol{M}}, \boldsymbol{M}) = q(\widehat{\boldsymbol{M}}, \boldsymbol{A}) + \text{optimism}.$$

For a general class of estimators and a wide class of error measures, it has been shown in Efron [1986] that the optimism can be estimated by summing up the covariances $\operatorname{cov}(\widehat{M}_{ij}, A_{ij})$. Furthermore, if the error measures are chosen as the likelihood-based deviances, this covariance penalty can be approximated by the divergence of the estimator with respect to the data [Efron et al., 2004]. On the other hand, for SVD-based spectral estimators, closed-form divergence formulas have been explicitly derived in the literature [Candès et al., 2013; Yuan, 2016]. A crucial idea of our method is to integrate these two ideas in order to provide a closed-form approximate estimator for $Q(\widehat{\boldsymbol{M}}, \boldsymbol{M})$.

Before we discuss the risk estimation method under the independent Bernoulli model in Efron [1986] and Efron et al. [2004], let's first review the Stein's unbiased risk estimation (SURE) framework for the normal mean estimation problem [Stein, 1981].

3.1.1. Stein's Unbiased Risk Estimation for the Normal Model. The problem of normal mean estimation is as follows: Suppose that our observation satisfies the stochastic model $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$, where $\boldsymbol{\mu} \in \mathbb{R}^n$ is the mean vector, while $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$ represents the *i.i.d.* normal noises. From the perspective of prediction, the standard linear regression model

with *i.i.d.* normal noises is a special case of the above general model, where an additional design matrix $\boldsymbol{X}$ is also incorporated. The expectation of $\boldsymbol{y}$, i.e. $\boldsymbol{\mu}$, can be estimated by methods such as James-Stein shrinkage [James and Stein, 1961], wavelet shrinkage [Donoho and Johnstone, 1995] or Lasso [Tibshirani, 1996; Zou et al., 2007]. Sometimes the estimate depends on some tuning parameter λ, then one approach to selecting λ is based on the evaluation of the estimate. Again, this involves the choice of the apparent error $q(\hat{\boldsymbol{\mu}}, \boldsymbol{y})$, the true error $Q(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu})$, and the derivation of the estimated true error $\widehat{Q}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu})$.

Let $\hat{\boldsymbol{\mu}} = g(\boldsymbol{y})$ be an estimate of $\boldsymbol{\mu}$ with some almost differentiable function $g : \mathbb{R}^n \to \mathbb{R}^n$. The apparent error between the observed vector and the estimated mean is simply defined as the squared error $q(\hat{\boldsymbol{\mu}}, \boldsymbol{y}) = \|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2$. Similarly, we also define the true error between $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}$ as the squared error

$$\begin{aligned} Q(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) &= \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 = \|g(\boldsymbol{y}) - \boldsymbol{y} + \boldsymbol{\epsilon}\|^2 \\ &= \|g(\boldsymbol{y}) - \boldsymbol{y}\|^2 + \|\boldsymbol{\epsilon}\|^2 + 2\boldsymbol{\epsilon}^\top [g(\boldsymbol{y}) - \boldsymbol{y}] \end{aligned}$$

Denote g_i as the coordinate function of g. Denote $\nabla g_i(\boldsymbol{y}) = \left(\frac{\partial g_i}{\partial y_1}, \cdots, \frac{\partial g_i}{\partial y_n}\right)^\top$ and $\mathrm{div}(g) = \sum_{i=1}^n \left.\frac{\partial g_i}{\partial y_i}\right|_{\boldsymbol{y}}$. The Stein's lemma [Stein, 1981] shows that as long as g is almost differentiable with $\mathbb{E}\,\|\nabla g_i(\boldsymbol{y})\|_2 < \infty$ for all i, one has $\mathbb{E}[\boldsymbol{\epsilon}^\top g(\boldsymbol{y})] = \mathbb{E}[\sigma^2 \mathrm{div}(g)]$. This implies

$$\begin{aligned} \mathbb{E}\,\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 &= \mathbb{E}\,\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2 + n\sigma^2 + 2\,\mathbb{E}\left[\boldsymbol{\epsilon}^\top [g(\boldsymbol{y}) - \boldsymbol{y}]\right] \\ &= \mathbb{E}\left[\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2 - n\sigma^2 + 2\sigma^2 \mathrm{div}(g)\right]. \end{aligned}$$

Then an unbiased estimate of $\mathbb{E}\,\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2$ is by SURE $= \|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2 - n\sigma^2 + 2\sigma^2 \mathrm{div}(g)$.

Through this chapter, for simplicity, we do not differentiate between "estimated true error" and "estimated risk". In other words, we treat the estimated risk as an estimate of the true error. Then we obtain a relationship between the estimated and apparent errors by

$$\widehat{Q}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) = q(\hat{\boldsymbol{\mu}}, \boldsymbol{y}) - n\sigma^2 + 2\sigma^2 \mathrm{div}(g).$$

3.1.2. Extension to the Bernoulli Model. Now, we give a short review of the extension of the SURE framework to the Bernoulli model proposed in Efron et al. [2004]. This is straightforwardly related to the problem of network rank selection.

Suppose the observations follow $y_i \overset{\text{indep}}{\sim} \text{Bernoulli}(\mu_i)$, for $i = 1, \ldots, n$. The problem is similar to that in the normal model: estimate $\boldsymbol{\mu}$ from $\boldsymbol{y}$ and evaluate this estimate. For a given estimator $\hat{\boldsymbol{\mu}}$, one method to choose the true error between $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}$ as well as the apparent error between $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{y}$ in Efron et al. [2004] is based on (half of) the binomial deviance:

$$Q(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) = \sum_{i=1}^{n} [\mu_i(-\log \hat{\mu}_i) + (1 - \mu_i)(-\log(1 - \hat{\mu}_i))]$$

and

$$q(\hat{\boldsymbol{\mu}}, \boldsymbol{y}) = \sum_{i=1}^{n} [y_i(-\log \hat{\mu}_i) + (1 - y_i)(-\log(1 - \hat{\mu}_i))] .$$

As $\boldsymbol{\mu}$ is not available, the goal is again to obtain an estimated true error $\widehat{Q}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu})$ based on the apparent error $q(\hat{\boldsymbol{\mu}}, \boldsymbol{y})$. Notice that

$$Q(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) - q(\hat{\boldsymbol{\mu}}, \boldsymbol{y}) = \sum_{i=1}^{n} (y_i - \mu_i)[\log \hat{\mu}_i - \log(1 - \hat{\mu}_i)].$$

By denoting $\hat{\lambda}_i = \log \hat{\mu}_i - \log(1 - \hat{\mu}_i)$, there holds

$$\mathbb{E}[Q(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) - q(\hat{\boldsymbol{\mu}}, \boldsymbol{y})] = \sum_{i=1}^{n} \mathbb{E}[(y_i - \mu_i)\hat{\lambda}_i] = \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left[(y_i - \mu_i)\hat{\lambda}_i | \boldsymbol{y}_{(i)}\right]\right],$$

where $\boldsymbol{y}_{(i)} = (y_1, y_2, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)^\top$. Notice that $\hat{\lambda}_i$ is a function of $\boldsymbol{y}$ thus a function of $(y_i, \boldsymbol{y}_{(i)})$. Furthermore, given y_i and $\boldsymbol{y}_{(i)}$ are independent, we have

$$\mathbb{E}\left[(y_i - \mu_i)\hat{\lambda}_i \middle| \boldsymbol{y}_{(i)}\right] = \mu_i(1 - \mu_i)[\hat{\lambda}_i(1, \boldsymbol{y}_{(i)}) - \hat{\lambda}_i(0, \boldsymbol{y}_{(i)})].$$

Similar to the discussion in Efron et al. [2004], this identity can be further approximated by partial derivatives:

$$\begin{aligned}\mathbb{E}\left[(y_i-\mu_i)\hat{\lambda}_i\middle|\boldsymbol{y}_{(i)}\right] &\approx \mu_i(1-\mu_i)\left.\frac{\partial\hat{\lambda}_i}{\partial\widehat{\mu}_i}\right|_{\widehat{\mu}_i}\left[\hat{\mu}_i(1,\boldsymbol{y}_{(i)})-\hat{\mu}_i(0,\boldsymbol{y}_{(i)})\right]\\ &= \mu_i(1-\mu_i)\frac{1}{\widehat{\mu}_i(1-\widehat{\mu}_i)}\left[\hat{\mu}_i(1,\boldsymbol{y}_{(i)})-\hat{\mu}_i(0,\boldsymbol{y}_{(i)})\right] \approx \left.\frac{\partial\hat{\mu}_i}{\partial y_i}\right|_{\boldsymbol{y}}.\end{aligned}$$

The above approximation leads to the following estimated error for the Bernoulli model

$$\widehat{Q}(\hat{\boldsymbol{\mu}},\boldsymbol{\mu}) = q(\hat{\boldsymbol{\mu}},\boldsymbol{y}) + \sum_{i=1}^{n}\left.\frac{\partial\hat{\mu}_i}{\partial y_i}\right|_{\boldsymbol{y}}. \tag{3.2}$$

We want to emphasize that the approximation suggested in Efron et al. [2004] is $\widehat{Q}(\boldsymbol{\mu},\hat{\boldsymbol{\mu}}) = q(\boldsymbol{y},\hat{\boldsymbol{\mu}}) + \sum_{i=1}^{n}\left.\frac{\partial\hat{\mu}_i}{\partial y_i}\right|_{\hat{\boldsymbol{\mu}}}$. However, based on the above approximation, the approximate risk estimation (3.2) seems to be more natural and straightforward. The study on the distinction of these two approximations can be left as future work.

3.1.3. Application to Binary Networks. Let's now discuss how to apply the aforementioned risk estimation method to the problem of rank selection in the network model (3.1). As mentioned before, for a given rank r, we estimate $\boldsymbol{M}$ by truncating the spectral decomposition $\boldsymbol{A} = \sum_{i=1}^{n}\lambda_i\boldsymbol{u}_i\boldsymbol{u}_i^\top$ into $\widehat{\boldsymbol{M}}_r = \sum_{i=1}^{r}\lambda_i\boldsymbol{u}_i\boldsymbol{u}_i^\top$.

For an undirected binary network, i.e. Bernoulli network, and a mean estimator $\widehat{\boldsymbol{M}}$, we define the true and apparent errors as in Section 3.1.2:

$$Q(\widehat{\boldsymbol{M}},\boldsymbol{M}) = \sum_{1\le i<j\le n}(-\log\widehat{M}_{ij})M_{ij} + (-\log(1-\widehat{M}_{ij}))(1-M_{ij}) \tag{3.3}$$

and

$$q(\widehat{\boldsymbol{M}},\boldsymbol{A}) = \sum_{1\le i<j\le n}(-\log\widehat{M}_{ij})A_{ij} + (-\log(1-\widehat{M}_{ij}))(1-A_{ij}). \tag{3.4}$$

We can then obtain the formula for the estimated error from (3.2):

$$\widehat{Q}(\widehat{\boldsymbol{M}}, \boldsymbol{M}) = q(\widehat{\boldsymbol{M}}, \boldsymbol{A}) + \sum_{1 \le i < j \le n} \frac{\partial \widehat{M}_{ij}}{\partial A_{ij}}. \tag{3.5}$$

Note that entries of the spectral truncation $\widehat{\boldsymbol{M}}_r$ might be below 0 or above 1, in which case the true error $Q(\widehat{\boldsymbol{M}}_r, \boldsymbol{M})$, the apparent error $q(\widehat{\boldsymbol{M}}_r, \boldsymbol{A})$, and the estimated error $\widehat{Q}(\widehat{\boldsymbol{M}}_r, \boldsymbol{M})$ are ill-defined. Therefore, the entries $(\widehat{M}_r)_{ij}$ are clipped into $\psi((\widehat{M}_r)_{ij})$ by some clipping function ψ with range $[\delta, 1-\delta]$. In practice, we can set $\delta = 10^{-6}$.

The remaining task is to give a closed-form formula of the divergence term in the estimated error (3.5). The mean estimator we are considering is the spectral truncation estimator. In fact, Jacobians of general spectral functions have been widely studied in the literature, see, e.g., Deledalle et al. [2012]; Edelman [2005]; Lewis and Sendov [2001]; Papadopoulo and Lourakis [2000]. Moreover, the divergence formula of SVD-based spectral functions has been explicitly calculated in Candès et al. [2013]; Yuan [2016] with distinct mathematical arguments. Interested readers are referred to these papers. Here, we only restate the divergence formula for SVD-based spectral truncation with rank r: For an $n_1 \times n_2$ ($n_1 \le n_2$) matrix $\boldsymbol{Y}$ with SVD $\boldsymbol{Y} := \sum_{i=1}^{n_1} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ where $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_r > \sigma_{r+1} \ge \ldots \ge \sigma_{n_1} \ge 0$, the divergence formula for the rank r SVD spectral truncation $\widehat{\boldsymbol{M}} := \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ is

$$\sum_{1 \le i \le n_1} \sum_{1 \le j \le n_2} \frac{\partial \widehat{M}_{ij}}{\partial Y_{ij}} = (n_1 + n_2 - r)r + 2 \sum_{k=1}^{r} \sum_{l=r+1}^{n_1} \frac{\sigma_l^2}{\sigma_k^2 - \sigma_l^2}.$$

However, to the best of our knowledge, an explicit divergence formula of spectral functions for symmetric matrices has not been explicitly given in the literature. Recall the spectral decomposition of the adjacency matrix as $\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^\top$ and denote U_{ik} as the (i,k)-th entry of $\boldsymbol{U}$. By following the steps in Candès et al. [2013], we can obtain the following

divergence formula of spectral truncation estimator $\widehat{\boldsymbol{M}}_r$:

$$\sum_{i<j}\frac{\partial(\widehat{M}_r)_{ij}}{\partial A_{ij}} = \frac{r(2n-r+1)}{2} + \sum_{k=1}^{r}\sum_{l=r+1}^{n}\frac{\lambda_l}{\lambda_k-\lambda_l} - 2\sum_{k=1}^{r-1}\sum_{l=k+1}^{r}\sum_{i=1}^{n}(U_{ik}U_{il})^2$$
$$- 2\sum_{k=1}^{r}\sum_{l=r+1}^{n}\frac{\lambda_k}{\lambda_k-\lambda_l}\sum_{i=1}^{n}(U_{ik}U_{il})^2 - \sum_{k=1}^{r}\sum_{i=1}^{n}(U_{ik})^4. \tag{3.6}$$

We will give a complete derivation procedure in Section 3.3.

3.2. Extension to Count-Weighted Networks

The above framework for binary networks can be extended to count-weighted networks, i.e., each pair of nodes can be connected with multiple edges. Count-weighted networks usually arise when each edge represents some co-occurrence between two nodes. In this chapter, we assume that each count edge follows Poisson distribution, and the edges between different pairs of nodes are independent, then the count-weighted adjacency matrix $\boldsymbol{A}$ has independent (strict) upper triangular entries with the distribution

$$A_{ij} \overset{\text{indep}}{\sim} \text{Poisson}(M_{ij}), \quad 1 \le i < j \le n, \tag{3.7}$$

where $\boldsymbol{M} = (M_{ij})_{(1\le i,j\le n)}$ is a symmetric matrix with nonnegative entries.

For count-weighted networks, can we still estimate the number of communities by estimating the rank of $\boldsymbol{M}$? We justify this perspective by recalling the K-color model proposed in [Ball et al., 2011], in which potential overlapping communities in a network based on edges rather than nodes, are assumed to come from K groups (or colors). Moreover, for a pair of nodes and each given color of edge, the number of their edges of that color is assumed to follow Poisson distribution, and all these Poisson random variables are assumed to be independent. Mathematically, each node $i \in [n]$ and each color $k \in [K]$ are connected with some parameter θ_{ik}. For each pair of nodes i and j, the number of their edges with the color k is assumed to follow the distribution Poisson$(\theta_{ik}\theta_{jk})$. Given the fact that the sum of

independent Poisson random variables is still a Poisson random variable, the total number of edges between i and j satisfies $A_{ij} \overset{\text{indep}}{\sim} \text{Poisson}(\sum_{k=1}^{K} \theta_{ik}\theta_{jk})$. Compare this with the above general model (3.7), we have $M_{ij} = \sum_{k=1}^{K} \theta_{ik}\theta_{jk}$ and it is straightforward to verify that the rank of $\boldsymbol{M}$ is K. This example serves as a reason why we still resort to rank selection for count-weighted networks.

As with the case of binary network, for each candidate rank r, with the spectral decomposition of the adjacency matrix $\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top$, we estimate the expected adjacency matrix by the spectral truncation estimator $\widehat{\boldsymbol{M}}_r = \sum_{i=1}^{r} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top$. In order to quantify the discrepancy between $\widehat{\boldsymbol{M}}_r$ and $\boldsymbol{M}$, we need to define the true error $Q(\widehat{\boldsymbol{M}}_r, \boldsymbol{M})$ and the apparent error $q(\widehat{\boldsymbol{M}}_r, \boldsymbol{A})$, and then find a reasonable estimated error $\widehat{Q}(\widehat{\boldsymbol{M}}_r, \boldsymbol{M})$.

To this end, let's still resort to Efron et al. [2004] for the risk estimation framework of the independent Poisson model. In fact, the Poisson case is quite similar to the Bernoulli case. Suppose we have independent Poisson observations

$$y_i \overset{\text{indep}}{\sim} \text{Poisson}(\mu_i), \quad i = 1, \ldots, n.$$

Again, for any estimator $\hat{\boldsymbol{\mu}}$, based on (half of) the Poisson deviance, the apparent error can be defined as $\sum_{i=1}^{n} [y_i(\log y_i - 1) - y_i \log \hat{\mu}_i + \hat{\mu}_i]$. Notice that our goal is to use the estimated error to compare the accuracy of estimators, for which $\sum_{i=1}^{n} [y_i(\log y_i - 1)]$ keeps unchanged. Therefore, we can consider a simpler apparent error

$$q(\hat{\boldsymbol{\mu}}, \boldsymbol{y}) = \sum_{i=1}^{n} (-y_i \log \hat{\mu}_i + \hat{\mu}_i)$$

and the true error is consequently

$$Q(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) = \sum_{i=1}^{n} \left(-\mu_i \log \hat{\mu}_i + \hat{\mu}_i\right).$$

As with the Bernoulli case,

$$\mathbb{E}\left[Q(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) - q\left(\hat{\boldsymbol{\mu}}, \boldsymbol{y}\right)\right] = \sum_{i=1}^{n} \mathbb{E}\left[(y_i - \mu_i)\log\hat{\mu}_i\right] = \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left[(y_i - \mu_i)\log\hat{\mu}_i | \boldsymbol{y}_{(i)}\right]\right].$$

With the following Taylor expansion

$$\log\hat{\mu}_i(y_i, \boldsymbol{y}_{(i)}) \approx \log\hat{\mu}_i(\mu_i, \boldsymbol{y}_{(i)}) + \frac{1}{\hat{\mu}_i(\mu_i, \boldsymbol{y}_{(i)})}\frac{\partial\hat{\mu}_i}{\partial y_i}(\mu_i, \boldsymbol{y}_{(i)})(y_i - \mu_i),$$

we have the approximation

$$\begin{aligned}\mathbb{E}[(y_i - \mu_i)\log\hat{\mu}_i | \boldsymbol{y}_{(i)}] &\approx \mathbb{E}\left[(y_i - \mu_i)\left(\log\hat{\mu}_i(\mu_i, \boldsymbol{y}_{(i)}) + \frac{1}{\hat{\mu}_i(\mu_i, \boldsymbol{y}_{(i)})}\frac{\partial\hat{\mu}_i}{\partial y_i}(\mu_i, \boldsymbol{y}_{(i)})(y_i - \mu_i)\right)\middle| \boldsymbol{y}_{(i)}\right] \\ &= \frac{\mu_i}{\hat{\mu}_i(\mu_i, \boldsymbol{y}_{(i)})}\frac{\partial\hat{\mu}_i}{\partial y_i}(\mu_i, \boldsymbol{y}_{(i)}) \approx \frac{\partial\hat{\mu}_i}{\partial y_i}(\mu_i, \boldsymbol{y}_{(i)}) \approx \left.\frac{\partial\hat{\mu}_i}{\partial y_i}\right|_{\boldsymbol{y}}.\end{aligned}$$

Consequently, the estimated error of the Poisson model is the same as (3.2) in the form

$$\widehat{Q}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) = q(\hat{\boldsymbol{\mu}}, \boldsymbol{y}) + \sum_{i=1}^{n}\left.\frac{\partial\hat{\mu}_i}{\partial y_i}\right|_{\boldsymbol{y}}.$$

Applying the above result to the model of count-weighted networks (3.7), we have the true and apparent errors defined for a mean estimator $\widehat{\boldsymbol{M}}$ as

$$Q(\widehat{\boldsymbol{M}}, \boldsymbol{M}) = \sum_{1\le i<j\le n}\left[-M_{ij}\log\widehat{M}_{ij} + \widehat{M}_{ij}\right]$$

and

$$q(\widehat{\boldsymbol{M}}, \boldsymbol{A}) = \sum_{1\le i<j\le n}[-A_{ij}\log\widehat{M}_{ij} + \widehat{M}_{ij}].$$

The estimated error is defined in the same form as (3.5), i.e.,

$$\widehat{Q}(\widehat{\boldsymbol{M}}, \boldsymbol{M}) = q(\widehat{\boldsymbol{M}}, \boldsymbol{A}) + \sum_{1\le i<j\le n}\frac{\partial\widehat{M}_{ij}}{\partial A_{ij}}.$$

Notice that this approximate risk estimation is different from the one in Bigot et al. [2017], in which an unbiased estimate of $\mathbb{E}\left[Q(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) - q\left(\hat{\boldsymbol{\mu}}, \boldsymbol{y}\right)\right]$ is given by using an identity proved in Hudson [1978]. This unbiased estimate is computationally expensive, and an approximation by the sum of random directional derivatives has been proposed in Bigot et al. [2017]. In contrast, our approach follows the ideas in Efron et al. [2004], and the resultant estimated error has the same form as in the Bernoulli case. In the next subsection, we will introduce the risk estimation method in Bigot et al. [2017] for symmetric Poisson networks.

3.2.1. Symmetric Poisson Network GSURE. We first introduce the following lemma, which is a variant of Lemma 7 in Bigot et al. [2017] for symmetric Poisson networks.

LEMMA 3.2.1. *Let $\mathbb{R}_n^{upp} \subset \mathbb{R}^{n\times n}$ be the space of $n \times n$ strictly upper triangular matrices. Denote $\boldsymbol{A}^{upp}$ as the upper triangular matrix consisting of the entries of $\boldsymbol{A}$ above the diagonal. Let $f : \mathbb{R}_n^{upp} \to \mathbb{R}^{n\times n}$ such that $\widehat{\boldsymbol{M}} = f(\boldsymbol{A}^{upp})$. For $1 \le i < j \le n$, denote $f_{ij} : \mathbb{R}_n^{upp} \to \mathbb{R}$ which makes $f_{ij}(\boldsymbol{A}^{upp})$ the (i, j)-th entry of the matrix $f(\boldsymbol{A}^{upp})$. Define $\boldsymbol{E}^{ij}$ as the $n \times n$ matrix with the (i, j)-th entry being 1 and the rest being 0. Then we have*

$$\mathbb{E}\left[\sum_{1\le i<j\le n} M_{ij} f_{ij}(\boldsymbol{A}^{upp})\right] = \mathbb{E}\left[\sum_{1\le i<j\le n} A_{ij} f_{ij}(\boldsymbol{A}^{upp} - \boldsymbol{E}^{ij})\right].$$

Define the mean-squared error as

$$\mathrm{MSE}(\widehat{\boldsymbol{M}}, \boldsymbol{M}) = \sum_{1\le i<j\le n} \mathbb{E}(\widehat{M}_{ij} - M_{ij})^2 = \sum_{1\le i<j\le n} \mathbb{E}\left[\widehat{M}_{ij}^2 - 2M_{ij}\widehat{M}_{ij} + M_{ij}^2\right].$$

By Lemma 3.2.1, the following quantity

$$\mathrm{PURE}(\widehat{\boldsymbol{M}}, \boldsymbol{M}) = \sum_{1\le i<j\le n} \widehat{M}_{ij}^2 - 2A_{ij} f_{ij}(\boldsymbol{A}^{\mathrm{upp}} - \boldsymbol{E}^{ij})$$

unbiasedly estimates $\mathrm{MSE}(\widehat{\boldsymbol{M}}, \boldsymbol{M}) - \sum_{1\le i<j\le n} M_{ij}^2$.

On the other hand, define the KL risk as

$$\mathrm{MKLA}(\widehat{\boldsymbol{M}}, \boldsymbol{M}) = \sum_{1\le i<j\le n} \mathbb{E}\left[\widehat{M}_{ij}^2 - M_{ij} - M_{ij}\log\frac{\widehat{M}_{ij}}{M_{ij}}\right].$$

By Lemma 3.2.1, the following quantity

$$\mathrm{PUKLA}(\widehat{\boldsymbol{M}}, \boldsymbol{M}) = \sum_{1\le i<j\le n} \widehat{M}_{ij}^2 - 2A_{ij}\log\left(f_{ij}(\boldsymbol{A}^{\mathrm{upp}} - \boldsymbol{E}^{ij})\right)$$

is an unbiased estimator of $\mathrm{MKLA}(\widehat{\boldsymbol{M}}, \boldsymbol{M}) + \sum_{1\le i<j\le n} M_{ij}^2 - M_{ij}\log M_{ij}$.

Let $\boldsymbol{\Delta} \in \mathbb{R}_n^{\mathrm{upp}}$ such that for $i < j$, $\Delta_{ij} \in \{-1, 1\}$ is Bernoulli distributed with $p = 0.5$. To avoid computation of $f_{ij}(\boldsymbol{A}^{\mathrm{upp}} - \boldsymbol{E}^{ij})$ for all $1 \le i < j \le n$ in $\mathrm{MKLA}(\widehat{\boldsymbol{M}}, \boldsymbol{M})$ and $\mathrm{PUKLA}(\widehat{\boldsymbol{M}}, \boldsymbol{M})$, the following approximation can be used:

$$\sum_{1\le i<j\le n} A_{ij} f_{ij}(\boldsymbol{A}^{\mathrm{upp}} - \boldsymbol{E}^{ij}) \approx \sum_{1\le i<j\le n} A_{ij}\left[f_{ij}(\boldsymbol{A}^{\mathrm{upp}}) - \Delta_{ij}\left(\mathrm{d}f[\boldsymbol{\Delta}]\right)_{ij}\right],$$
$$\sum_{1\le i<j\le n} A_{ij}\log\left(f_{ij}(\boldsymbol{A}^{\mathrm{upp}} - \boldsymbol{E}^{ij})\right) \approx \sum_{1\le i<j\le n} A_{ij}\log\left[f_{ij}(\boldsymbol{A}^{\mathrm{upp}}) - \Delta_{ij}\left(\mathrm{d}f[\boldsymbol{\Delta}]\right)_{ij}\right],$$

where $\mathrm{d}f[\boldsymbol{\Delta}]$ is the differential of $f(\cdot)$ with input $\boldsymbol{\Delta}$ and can be calculated as in Section 3.3.

3.3. Derivation of Spectral Divergence Formula

Our rank selection method relies crucially on the estimated error (3.5), in which the divergence term has a closed-form formula as shown in (3.6). This section aims to present a complete derivation of this formula. As mentioned earlier, the divergence formula has been derived for SVD-based spectral functions of rectangular matrices in Candès et al. [2013] and Yuan [2016]. Compared to Yuan [2016], the approach in Candès et al. [2013] follows the literature of Jacobians of spectral functions and gives the divergence formula for general SVD-based spectral functions. We can derive (3.6) by similar arguments.

Recall the spectral decomposition of the adjacency matrix as

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top,$$

where $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n] \in \mathbb{R}^{n\times n}$ is orthogonal, and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. Consider a general spectral estimator

$$\widehat{\boldsymbol{M}} = \sum_{i=1}^{n} g_i(\lambda_i) \boldsymbol{u}_i \boldsymbol{u}_i^\top := \boldsymbol{U} g(\boldsymbol{\Lambda}) \boldsymbol{U}^\top.$$

Let $\mathbb{R}_n^{\mathrm{upp}} \subset \mathbb{R}^{n\times n}$ be the space of $n \times n$ strictly upper triangular matrices. Denote $\boldsymbol{A}^{\mathrm{upp}}$ as the upper triangular matrix consisting of the entries of $\boldsymbol{A}$ above the diagonal. Then we have

$$\boldsymbol{A} = \boldsymbol{A}^{\mathrm{upp}} + (\boldsymbol{A}^{\mathrm{upp}})^\top.$$

We also denote

$$\widehat{\boldsymbol{M}} = f(\boldsymbol{A}^{\mathrm{upp}}),$$

where f is determined by g. The differential of f at $\boldsymbol{A}^{\mathrm{upp}}$, $\mathrm{d}f|_{\boldsymbol{A}^{\mathrm{upp}}}$, is defined to be the linear mapping from $\mathbb{R}_n^{\mathrm{upp}} \to \mathbb{R}^{n\times n}$ such that for any $\boldsymbol{\Delta} \in \mathbb{R}_n^{\mathrm{upp}}$,

$$\lim_{\boldsymbol{\Delta}\to\boldsymbol{0}} \frac{\|f(\boldsymbol{A}^{\mathrm{upp}} + \boldsymbol{\Delta}) - f(\boldsymbol{A}^{\mathrm{upp}}) - \mathrm{d}f|_{\boldsymbol{A}^{\mathrm{upp}}}[\boldsymbol{\Delta}]\|_F}{\|\boldsymbol{\Delta}\|_F} = 0.$$

Let $\mathrm{d}\boldsymbol{U}|_{\boldsymbol{A}^{\mathrm{upp}}}$ and $\mathrm{d}(g \circ \boldsymbol{\Lambda})|_{\boldsymbol{A}^{\mathrm{upp}}}$ be linear mappings defined similarly. For any $\boldsymbol{\Delta} \in \mathbb{R}_n^{\mathrm{upp}}$, denote $\boldsymbol{\Omega}[\boldsymbol{\Delta}] = \boldsymbol{U}^\top \mathrm{d}\boldsymbol{U}[\boldsymbol{\Delta}]$. By taking the differential on both sides of $\boldsymbol{I}_n = \boldsymbol{U}^\top\boldsymbol{U}$, we get

$$\boldsymbol{0} = \boldsymbol{U}^\top \mathrm{d}\boldsymbol{U}[\boldsymbol{\Delta}] + \mathrm{d}\boldsymbol{U}[\boldsymbol{\Delta}]^\top \boldsymbol{U} = \boldsymbol{\Omega}[\boldsymbol{\Delta}] + \boldsymbol{\Omega}[\boldsymbol{\Delta}]^\top,$$

which means $\boldsymbol{\Omega}[\boldsymbol{\Delta}]$ is skew-symmetric. Furthermore, by taking the differential on both sides of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$, we have

$$\boldsymbol{\Delta} + \boldsymbol{\Delta}^\top = \mathrm{d}\boldsymbol{U}[\boldsymbol{\Delta}]\boldsymbol{\Lambda}\boldsymbol{U}^\top + \boldsymbol{U}\mathrm{d}\boldsymbol{\Lambda}[\boldsymbol{\Delta}]\boldsymbol{U}^\top + \boldsymbol{U}\boldsymbol{\Lambda}\mathrm{d}\boldsymbol{U}[\boldsymbol{\Delta}]^\top,$$

which gives

$$\begin{aligned} \boldsymbol{U}^\top(\boldsymbol{\Delta}+\boldsymbol{\Delta}^\top)\boldsymbol{U} &= \boldsymbol{\Omega}[\boldsymbol{\Delta}]\boldsymbol{\Lambda} + \mathrm{d}\boldsymbol{\Lambda}[\boldsymbol{\Delta}] + \boldsymbol{\Lambda}\boldsymbol{\Omega}^\top[\boldsymbol{\Delta}] \\ &= \mathrm{d}\boldsymbol{\Lambda}[\boldsymbol{\Delta}] + (\boldsymbol{\Omega}[\boldsymbol{\Delta}]\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{\Omega}[\boldsymbol{\Delta}])\,. \end{aligned} \tag{3.8}$$

Notice that $\boldsymbol{\Lambda}$ is diagonal and $\boldsymbol{\Omega}[\boldsymbol{\Delta}]$ is skew-symmetric, thus $\mathrm{d}\boldsymbol{\Lambda}[\boldsymbol{\Delta}]$ is a diagonal matrix while the diagonal elements of $\boldsymbol{\Omega}[\boldsymbol{\Delta}]\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}\boldsymbol{\Omega}[\boldsymbol{\Delta}]$ are zero. By comparing the diagonal and off-diagonal entries of (3.8) separately, we have

$$\mathrm{d}\lambda_k[\boldsymbol{\Delta}] = [\boldsymbol{U}^\top(\boldsymbol{\Delta}+\boldsymbol{\Delta}^\top)\boldsymbol{U}]_{kk}, \quad \text{for } k = 1,\dots,n$$

and

$$\Omega_{kl}[\boldsymbol{\Delta}] = \frac{1}{\lambda_l - \lambda_k}[\boldsymbol{U}^\top(\boldsymbol{\Delta}+\boldsymbol{\Delta}^\top)\boldsymbol{U}]_{kl}, \quad \text{for } k \neq l.$$

By taking differentials on both sides of $f(\boldsymbol{A}^{\mathrm{upp}}) = \boldsymbol{U}g(\boldsymbol{\Lambda})\boldsymbol{U}^\top$, we get

$$\mathrm{d}f[\boldsymbol{\Delta}] = \mathrm{d}\boldsymbol{U}[\boldsymbol{\Delta}]g(\boldsymbol{\Lambda})\boldsymbol{U}^\top + \boldsymbol{U}\mathrm{d}(g\circ\boldsymbol{\Lambda})[\boldsymbol{\Delta}]\boldsymbol{U}^\top + \boldsymbol{U}g(\boldsymbol{\Lambda})\mathrm{d}\boldsymbol{U}[\boldsymbol{\Delta}]^\intercal.$$

Then, by multiplying $\boldsymbol{U}^\top$ and $\boldsymbol{U}$ on left and right sides, respectively, we get

$$\begin{aligned} \boldsymbol{U}^\top\mathrm{d}f[\boldsymbol{\Delta}]\boldsymbol{U} &= \boldsymbol{\Omega}[\boldsymbol{\Delta}]g(\boldsymbol{\Lambda}) + \mathrm{d}(g\circ\boldsymbol{\Lambda})[\boldsymbol{\Delta}] + g(\boldsymbol{\Lambda})\boldsymbol{\Omega}[\boldsymbol{\Delta}]^\top \\ &= \boldsymbol{\Omega}[\boldsymbol{\Delta}]g(\boldsymbol{\Lambda}) + \mathrm{d}g_{\boldsymbol{\Lambda}}[\mathrm{d}\boldsymbol{\Lambda}_{\boldsymbol{A}^{\mathrm{upp}}}[\boldsymbol{\Delta}]] - g(\boldsymbol{\Lambda})\boldsymbol{\Omega}[\boldsymbol{\Delta}]. \end{aligned} \tag{3.9}$$

Notice that the diagonal elements of $\boldsymbol{\Omega}[\boldsymbol{\Delta}]g(\boldsymbol{\Lambda})$ and $g(\boldsymbol{\Lambda})\boldsymbol{\Omega}[\boldsymbol{\Delta}]$ are zero, while $\mathrm{d}g_{\boldsymbol{\Lambda}}[\mathrm{d}\boldsymbol{\Lambda}_{\boldsymbol{A}^{\mathrm{upp}}}[\boldsymbol{\Delta}]]$ is a diagonal matrix. By comparing diagonal and off-diagonal entries of (3.9) separately, we have

$$\begin{aligned} (\boldsymbol{U}^\top\mathrm{d}f[\boldsymbol{\Delta}]\boldsymbol{U})_{kk} &= g_k'(\lambda_k)\mathrm{d}\lambda_k[\boldsymbol{\Delta}] \\ &= g_k'(\lambda_k)[\boldsymbol{U}^\top(\boldsymbol{\Delta}+\boldsymbol{\Delta}^\top)\boldsymbol{U}]_{kk}, \quad \text{for } k = 1,\dots,n \end{aligned}$$

and

$$(\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{\Delta}]\boldsymbol{U})_{kl} = (g_k(\lambda_k) - g_l(\lambda_l))\Omega_{kl}[\boldsymbol{\Delta}]$$
$$= \frac{g_k(\lambda_k) - g_l(\lambda_l)}{\lambda_k - \lambda_l}[\boldsymbol{U}^\top(\boldsymbol{\Delta} + \boldsymbol{\Delta}^\top)\boldsymbol{U}]_{kl}, \quad \text{for } k \neq l.$$

Thus we have the complete expression of $\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{\Delta}]\boldsymbol{U}$ for any $\boldsymbol{\Delta} \in \mathbb{R}_n^{\mathrm{upp}}$.

For any $i < j$, let $\boldsymbol{E}^{ij}$ be an $n \times n$ matrix whose (i, j)-th entry is 1 while all other entries are 0. Denote $\boldsymbol{L}^{ij} = \boldsymbol{U}^\top \boldsymbol{E}^{ij} \boldsymbol{U}$, then its (k, l)-th entry is $L_{kl}^{ij} = U_{ik}U_{jl}$. Since $\{\boldsymbol{E}^{ij}\}_{1 \le i < j \le n}$ is the canonical basis of $\mathbb{R}_n^{\mathrm{upp}}$, we have

$$\sum_{1 \le i < j \le n} \frac{\partial \widehat{M}_{ij}}{\partial A_{ij}} = \sum_{1 \le i < j \le n} \langle \boldsymbol{E}^{ij}, \mathrm{d}f[\boldsymbol{E}^{ij}] \rangle$$
$$= \sum_{1 \le i < j \le n} \langle \boldsymbol{L}^{ij}, \boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{E}^{ij}]\boldsymbol{U} \rangle$$
$$= \underbrace{\sum_{1 \le i < j \le n} \sum_{k \neq l} L_{kl}^{ij} \cdot (\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{E}^{ij}]\boldsymbol{U})_{kl}}_{S_1} + \underbrace{\sum_{1 \le i < j \le n} \sum_{k=1}^{n} L_{kk}^{ij} \cdot (\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{E}^{ij}]\boldsymbol{U})_{kk}}_{S_2}.$$

By plugging the previous results of differentials into S_1 and S_2, after some tedious calculation, one obtains

$$
\begin{aligned}
S_1 &= \sum_{1\le i<j\le n}\sum_{k\neq l} L_{kl}^{ij}\cdot\frac{g_k(\lambda_k)-g_l(\lambda_l)}{\lambda_k-\lambda_l}[\boldsymbol{U}^\intercal(\boldsymbol{E}^{ij}+\boldsymbol{E}^{ji})\boldsymbol{U}]_{kl}\\
&= \sum_{i<j}\left\{\sum_{k<l}\frac{g_k(\lambda_k)-g_l(\lambda_l)}{\lambda_k-\lambda_l}(L_{kl}^{ij}+L_{lk}^{ij})\cdot L_{kl}^{ij}+\sum_{k>l}\frac{g_k(\lambda_k)-g_l(\lambda_l)}{\lambda_k-\lambda_l}(L_{kl}^{ij}+L_{lk}^{ij})\cdot L_{kl}^{ij}\right\}\\
&= \sum_{i<j}\left\{\sum_{k<l}\frac{g_k(\lambda_k)-g_l(\lambda_l)}{\lambda_k-\lambda_l}(L_{kl}^{ij}+L_{lk}^{ij})\cdot L_{kl}^{ij}+\sum_{k<l}\frac{g_k(\lambda_k)-g_l(\lambda_l)}{\lambda_k-\lambda_l}(L_{lk}^{ij}+L_{kl}^{ij})\cdot L_{lk}^{ij}\right\}\\
&= \sum_{k<l}\frac{g_k(\lambda_k)-g_l(\lambda_l)}{\lambda_k-\lambda_l}\sum_{i<j}(L_{kl}^{ij}+L_{lk}^{ij})^2\\
&= \sum_{k<l}\frac{g_k(\lambda_k)-g_l(\lambda_l)}{\lambda_k-\lambda_l}\underbrace{\sum_{i<j}(U_{ik}U_{jl}+U_{il}U_{jk})^2}_{V_1(k,l)}
\end{aligned}
$$

and

$$
S_2 = \sum_{1\le i<j\le n}\sum_{k=1}^{n} L_{kk}^{ij}\cdot 2L_{kk}^{ij}g_k'(\lambda_k) = \sum_{k=1}^{n} g_k'(\lambda_k)\sum_{1\le i<j\le n} 2(L_{kk}^{ij})^2 = \sum_{k=1}^{n} g_k'(\lambda_k)\underbrace{\sum_{1\le i<j\le n} 2(U_{ik}U_{jk})^2}_{V_2(k)}.
$$

Since $\boldsymbol{U}$ is orthogonal, for each $k < l$, $V_1(k,l)$ can be simplified as

$$\begin{aligned}
V_1(k,l) &= \sum_{1\le i<j\le n} U_{ik}^2 U_{jl}^2 + U_{il}^2 U_{jk}^2 + 2U_{ik}U_{il}U_{jk}U_{jl} \\
&= \sum_{i\neq j} U_{ik}^2 U_{jl}^2 + \sum_{i\neq j} U_{ik}U_{il}U_{jk}U_{jl} \\
&= \sum_{i=1}^n U_{ik}^2 \left(\sum_{j\neq i} U_{jl}^2\right) + \sum_{i=1}^n U_{ik}U_{il}\left(\sum_{j\neq i} U_{jk}U_{jl}\right) \\
&= \sum_{i=1}^n U_{ik}^2 (1 - U_{il}^2) + \sum_{i=1}^n U_{ik}U_{il}\big(0 - U_{ik}U_{il}\big) \\
&= \sum_{i=1}^n (U_{ik}^2 - U_{ik}^2 U_{il}^2) - \sum_{i=1}^n U_{ik}^2 U_{il}^2 \\
&= 1 - 2\sum_{i=1}^n (U_{ik}U_{il})^2.
\end{aligned}$$

Similarly, for each $k = 1, \cdots, n$, $V_2(k)$ can be simplified as

$$V_2(k) = \sum_{1\le i<j\le n} 2U_{ik}^2 U_{jk}^2 = \sum_{i\neq j} U_{ik}^2 U_{jk}^2 = \sum_{i=1}^n U_{ik}^2 \sum_{j\neq i} U_{jk}^2 = \sum_{i=1}^n U_{ik}^2 \big(1 - U_{ik}^2\big) = 1 - \sum_{i=1}^n (U_{ik})^4.$$

Therefore we have

$$\begin{aligned}
S_1 &= \sum_{k<l} \frac{g_k(\lambda_k) - g_l(\lambda_l)}{\lambda_k - \lambda_l}\left[1 - 2\sum_{i=1}^n (U_{ik}U_{il})^2\right], \\
S_2 &= \sum_{k=1}^n g_k'(\lambda_k)\left[1 - \sum_{i=1}^n (U_{ik})^4\right].
\end{aligned}$$

Putting everything together, the divergence is

$$\sum_{1\le i<j\le n} \frac{\partial \widehat{M}_{ij}}{\partial A_{ij}} = \sum_{k<l} \frac{g_k(\lambda_k) - g_l(\lambda_l)}{\lambda_k - \lambda_l}\left[1 - 2\sum_{i=1}^n (U_{ik}U_{il})^2\right] + \sum_{k=1}^n g_k'(\lambda_k)\left[1 - \sum_{i=1}^n (U_{ik})^4\right]. \tag{3.10}$$

If the spectral estimator is the spectral truncation estimator with rank r as we suggested previously, i.e., $\widehat{\boldsymbol{M}}_r = \sum_{i=1}^r \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top$, we further have $g(\lambda_k) = \lambda_k$ if $k \le r$ while $g(\lambda_k) = 0$ if

$k > r$. In this case, both S_1 and S_2 can be further simplified as

$$\begin{aligned} S_1 =& \frac{r(2n-r-1)}{2} + \sum_{k=1}^{r}\sum_{l=r+1}^{n}\frac{\lambda_l}{\lambda_k-\lambda_l} - 2\sum_{k=1}^{r-1}\sum_{l=k+1}^{r}\sum_{i=1}^{n}(U_{ik}U_{il})^2 \\ & - 2\sum_{k=1}^{r}\sum_{l=r+1}^{n}\frac{\lambda_k}{\lambda_k-\lambda_l}\sum_{i=1}^{n}(U_{ik}U_{il})^2 \end{aligned}$$

and $S_2 = r - \sum_{k=1}^{r}\sum_{i=1}^{n}(U_{ik})^4$, which give the divergence

$$\begin{aligned} \sum_{1\le i<j\le n}\frac{\partial(\widehat{M}_r)_{ij}}{\partial A_{ij}} =& \frac{r(2n-r+1)}{2} + \sum_{k=1}^{r}\sum_{l=r+1}^{n}\frac{\lambda_l}{\lambda_k-\lambda_l} - 2\sum_{k=1}^{r-1}\sum_{l=k+1}^{r}\sum_{i=1}^{n}(U_{ik}U_{il})^2 \\ & - 2\sum_{k=1}^{r}\sum_{l=r+1}^{n}\frac{\lambda_k}{\lambda_k-\lambda_l}\sum_{i=1}^{n}(U_{ik}U_{il})^2 - \sum_{k=1}^{r}\sum_{i=1}^{n}(U_{ik})^4. \end{aligned} \tag{3.11}$$

CHAPTER 4

Experiments

In this chapter, the methods introduced in Chapter 2 and 3 are referred to as "SMAST" and "GSURE", respectively.

4.1. Implementation Details of SMAST

First, we address the computational issue mentioned in Chapter 2. When solving (2.9) in practice, in order to preserve the symmetry in the scaled matrix, we use the iterative scaling algorithm introduced in Knight et al. [2014]. However, as shown in Chapter 2, the equivalent weighted scaling problem (2.11) is much simpler to solve. Therefore, the following algorithm is proposed based on the algorithm in Theorem 3.1 of Knight et al. [2014]. By Theorem 3.1 of Knight et al. [2014], $\lim_{s\to\infty} \sum_{l=1}^{m} \left|\widehat{\mathcal{N}}_l\right| F_{kl}^{(s)} a_k^{(s)} a_l^{(s)} = 1$ for $k = 1, \cdots, m$. Therefore, the output $\hat{\boldsymbol{\xi}}$ is the unique scaling vector solving (2.9). In the experiments, we set the tolerance $\delta = 10^{-8}$.

Algorithm 1 Symmetric Matrix Scaling

Input: Estimated communities $\widehat{\mathcal{N}}_1, \cdots, \widehat{\mathcal{N}}_m$, the $m \times m$ matrix $\widehat{\boldsymbol{B}}$, tolerance $\delta > 0$.

Initialize: $\boldsymbol{F}^{(0)} = \widehat{\boldsymbol{B}}$, $\boldsymbol{a}^{(0)} = \mathbf{1}_m$, $s = 0$.

while $\max_k \left|\sum_{l=1}^{m} \left|\widehat{\mathcal{N}}_l\right| F_{kl}^{(s)} - 1\right| \geq \delta$, **do**:

- $a^{(s+1)}(k) = \left(\sum_{l=1}^{m} \left|\widehat{\mathcal{N}}_l\right| F_{kl}^{(s)}\right)^{-1/2}$ for $k = 1, \cdots, m$.
- $\boldsymbol{F}^{(s+1)} = \mathrm{diag}(\boldsymbol{a}^{(s+1)}) \boldsymbol{F}^{(s)} \mathrm{diag}(\boldsymbol{a}^{(s+1)})$.
- Update $s = s + 1$.

Output: $\hat{\boldsymbol{\xi}} \in \mathbf{R}^n$ such that $\hat{\xi}(i) = \sqrt{n} \cdot a^{(s)}(k)$ if $i \in \widehat{\mathcal{N}}_k$ for $k = 1, \ldots, m$.

Notice that for SMAST, the DCSBM parameter estimates $\hat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{B}}$ need to be well-defined and positive. We have shown this holds theoretically with high probability, however

in all the simulations, in order to obtain valid estimators, we calculate the estimators in (2.6) and (2.7) as

$$\hat{\theta}_i = \frac{\sqrt{\hat{\mathbf{1}}_k^\top \boldsymbol{A} \hat{\mathbf{1}}_k \vee 1}}{\hat{\mathbf{1}}_k^\top \boldsymbol{A} \mathbf{1}_n \vee 1} \cdot (d_i \vee 1), \quad \text{for } k = 1, \ldots, m \text{ and } i \in \widehat{\mathcal{N}}_k \tag{4.1}$$

and

$$\widehat{B}_{kl} = \frac{\hat{\mathbf{1}}_k^\top \boldsymbol{A} \hat{\mathbf{1}}_l \vee 1}{\sqrt{\hat{\mathbf{1}}_k^\top \boldsymbol{A} \hat{\mathbf{1}}_k \vee 1} \sqrt{\hat{\mathbf{1}}_l^\top \boldsymbol{A} \hat{\mathbf{1}}_l \vee 1}}, \quad \text{for } 1 \le k, l \le m. \tag{4.2}$$

4.2. Synthetic Networks

This section studies the performance of SMAST in selecting the number of communities in synthetic count-weighted networks. To compare its performance, we consider two penalized likelihood methods: the corrected BIC (CBIC) in Hu et al. [2020] and integrated classification likelihood (ICL) method in Daudin et al. [2008]. As discussed in Section 6.2 of Hu et al. [2020], the likelihood is based on Poisson distribution so the methods naturally apply to count-weighted networks from the general DCSBM.

Denote $\boldsymbol{z} \in [K]^n$ as the node label vector. For CBIC, the penalized log-likelihood function with m communities in Hu et al. [2020] is:

$$\mathrm{CBIC}(m) = \max_{\boldsymbol{z}} \sup_{\boldsymbol{B}} \log f(\boldsymbol{A}|\boldsymbol{z}, \boldsymbol{B}) - \left[\lambda n \log m + \frac{m(m+1)}{2} \log n \right],$$

where the tuning parameter $\lambda = 1$ as suggested by the experiments in Hu et al. [2020]. On the other hand, the penalized log-likelihood for ICL is:

$$\mathrm{ICL}(m) = \max_{\boldsymbol{z}} \sup_{\boldsymbol{B}} \log f(\boldsymbol{A}|\boldsymbol{z}, \boldsymbol{B}) - \left[\sum_{k=1}^{m} n_k \log \left(\frac{n}{n_k} \right) + \frac{m(m+2)}{2} \log n \right].$$

Since it is intractable to optimize the log-likelihood over all possible community assignments, we use the label vector obtained by SCORE or the regularized spectral clustering (RSC) algorithm [Amini et al., 2013; Joseph and Yu, 2016] to compute the log-likelihood.

Since SMAST also involves spectral clustering, although we use SCORE for theory, in simulations we implement with both SCORE and RSC in order to be aligned with CBIC and ICL for comparison. Notice that for RSC, the regularization applied is $0.25 \times \bar{d}/n$, where $\bar{d}$ is the average node degree of the adjacency.

Now we discuss how to generate the expected adjacency matrix of DCSBM. The generating mechanism of the model parameters is similar to the one in Section 6.2 of Hu et al. [2020]. Recall the expected adjacency matrix of DCSBM is parameterized as

$$M_{ij} = \theta_i \theta_j \boldsymbol{B}_{\phi(i)\phi(j)}, \quad 1 \le i \le j \le n.$$

The entries of $\boldsymbol{B}$ are set to $B_{kl} = \rho\,(1 + r \times \mathbf{1}(k = l))$. We consider the following combinations of (ρ, r): $(0.04, 4)$, $(0.06, 3)$ and $(0.12, 2)$. Let $K = 2, \cdots, 6$. The block sizes are set according to the sequence $(50, 100, 150, 50, 100, 150)$. For example, if $K = 2$, $(n_1, n_2) = (50, 100)$; if $K = 3$, $(n_1, n_2, n_3) = (50, 100, 150)$. The degree-correction parameters are *i.i.d.* generated from the following distribution:

$$\begin{cases} \text{Uniform}(0.6, 1.4), & \text{with probability } 0.8; \\ 7/11, & \text{with probability } 0.1; \\ 15/11, & \text{with probability } 0.1. \end{cases}$$

After generating $\boldsymbol{M}$, we consider the following simulations to generate $\boldsymbol{A}$:

- **Simulation 1:** Generate $A_{ij} \overset{ind}{\sim} \text{Poisson}(M_{ij})$ for $1 \le i < j \le n$.
- **Simulation 2:** Generate $A_{ij} \overset{ind}{\sim} \text{Binomial}(5, M_{ij}/5)$ for $1 \le i < j \le n$.
- **Simulation 3:** Generate $A_{ij} \overset{ind}{\sim} \text{NB}(5, M_{ij}/5)$ for $1 \le i < j \le n$, where $\text{NB}(n, p)$ is the negative binomial distribution with n trials and success probability p in each trial.

Notice that in Simulation 3, Assumption (2.17) is slightly violated, and SMAST is applied with $2.1\sqrt{n}$ as the spectral threshold.

For each setting, we generate synthetic networks $n_{\max} = 200$ times and apply the aforementioned methods to estimate the true rank K. Then the accuracy rates ("Rate"), defined as the number of times a method correctly estimates K divided by $n_{\max}$, are recorded. We also record the average estimated number of communities for each setting of the $n_{\max}$ simulated networks, denoted as "Mean" in the tables. For CBIC and ICL, a prespecified range for $\widehat{K}$ is required, i.e., $\widehat{K} \in \{1, 2, \cdots, K_{\max}\}$, and we let $K_{\max} = 15$.

Table 4.1 to 4.3 display results from Simulation 1; Table 4.4 to 4.6 display results from Simulation 2; Table 4.7 to 4.9 display results from Simulation 3.

		SMAST		CBIC		ICL	
		Rate	Mean	Rate	Mean	Rate	Mean
SCORE	$K = 2$	0.965	2.04	0.965	2.025	0.965	2.025
	$K = 3$	0.675	3.36	0.385	3.20	0.39	3.22
	$K = 4$	0.925	4.075	0.29	3.50	0.315	3.53
	$K = 5$	0.95	5.04	0.32	4.485	0.32	4.445
	$K = 6$	0.97	5.98	0.19	5.055	0.20	5.015
RSC	$K = 2$	1.00	2.00	1.00	2.00	1.00	2.00
	$K = 3$	0.995	3.005	1.00	3.00	1.00	3.00
	$K = 4$	1.00	4.00	0.685	3.685	0.925	3.925
	$K = 5$	0.98	4.98	0.555	4.555	0.805	4.805
	$K = 6$	0.75	5.75	0.275	5.275	0.555	5.555

TABLE 4.1. Comparison of SMAST, CBIC and ICL for Simulation 1 (Poisson): $\rho = 0.04$, $r = 4$.

		SMAST		CBIC		ICL	
		Rate	Mean	Rate	Mean	Rate	Mean
SCORE	$K = 2$	0.99	2.01	0.995	2.005	0.995	2.005
	$K = 3$	0.87	3.13	0.49	3.185	0.49	3.20
	$K = 4$	0.975	4.015	0.125	3.205	0.135	3.235
	$K = 5$	0.97	4.98	0.20	4.285	0.215	4.295
	$K = 6$	0.735	5.745	0.21	5.11	0.215	5.10
RSC	$K = 2$	0.995	2.005	1.00	2.00	1.00	2.00
	$K = 3$	0.995	3.005	1.00	3.00	1.00	3.00
	$K = 4$	1.00	4.00	0.855	3.855	0.97	3.97
	$K = 5$	0.88	4.90	0.415	4.415	0.70	4.70
	$K = 6$	0.53	5.53	0.18	5.18	0.40	5.41

TABLE 4.2. Comparison of SMAST, CBIC and ICL for Simulation 1 (Poisson): $\rho = 0.06$, $r = 3$.

4.3. Real-World Networks

4.3.1. Binary Networks. In this subsection, we apply the two methods SMAST and GSURE to the following real-world networks: Football [Girvan and Newman, 2002], Political books [Newman, 2006], Dolphins [Lusseau et al., 2003], Karate [Zachary, 1977], Political blogs [Adamic and Glance, 2005], UK faculty [Nepusz et al., 2008]. The datasets can be downloaded from `http://www-personal.umich.edu/~mejn/netdata/`. The UKfaculty network has an asymmetric adjacency matrix (directed), so it is symmetrized by ignoring the edge directions. All these datasets are from binary-edge networks, and have been intensively studied in the literature by methods estimating the number of communities in binary networks. The true number of communities K of each network has been widely discussed in the literature, and some networks have multiple true K proposed. For example, the Polbooks network is suggested by Le and Levina [2015] to have 3 communities, but can also be modeled by the degree corrected mixed-membership (DCMM) model with two communities

		SMAST		CBIC		ICL	
		Rate	Mean	Rate	Mean	Rate	Mean
SCORE	$K = 2$	0.97	2.035	1.00	2.00	1.00	2.00
	$K = 3$	0.955	3.045	0.825	3.13	0.825	3.13
	$K = 4$	0.96	4.01	0.195	3.25	0.22	3.265
	$K = 5$	0.915	4.925	0.095	4.195	0.095	4.195
	$K = 6$	0.52	5.53	0.21	5.29	0.215	5.26
RSC	$K = 2$	0.97	2.03	1.00	2.00	1.00	2.00
	$K = 3$	0.985	3.015	1.00	3.00	1.00	3.00
	$K = 4$	0.97	4.01	0.685	3.685	0.855	3.855
	$K = 5$	0.88	4.89	0.59	4.59	0.78	4.78
	$K = 6$	0.315	5.31	0.235	5.245	0.35	5.36

TABLE 4.3. Comparison of SMAST, CBIC and ICL for Simulation 1 (Poisson): $\rho = 0.12$, $r = 2$.

[Jin et al., 2017], which makes $K = 2$ also reasonable. For a discussion of the true K of each network and a summary of other methods' performance, one can refer to Section 6 of Jin et al. [2022]. In this section, for methods requiring a range of $\widehat{K} = 3$, we set the maximum candidate number of communities $K_{\max} = 15$. Table 4.10 displays the summary statistics of the networks and the estimated number of communities by SMAST and GSURE.

For the Polblogs dataset, we apply SMAST and GSURE to the regularized adjacency matrix $\boldsymbol{A}_\tau = \boldsymbol{A} + \tau \boldsymbol{J}$ where $\boldsymbol{J}$ is the $n \times n$ matrix of 1's and τ is a tuning parameter. The value of τ tested ranges from 0.05 to 0.5. SMAST gives the estimated number of communities $\widehat{K} = 3$ for $\tau = 0.05, 0.1, 0.15$ and $\widehat{K} = 2$ for $\tau \in [0.2, 0.5]$; while GSURE gives $\widehat{K} = 3$ for all τ's.

4.3.2. *Les Misèrable* Network. In this subsection, we study the count-weighted *Les Misèrable* network compiled in Knuth [1993], also analyzed in the literature of network analysis, see, e.g., Ball et al. [2011]; Newman and Girvan [2004]; Newman and Reinert

		SMAST		CBIC		ICL	
		Rate	Mean	Rate	Mean	Rate	Mean
SCORE	$K = 2$	0.985	2.015	0.985	2.015	0.985	2.015
	$K = 3$	0.80	3.215	0.45	3.125	0.45	3.105
	$K = 4$	0.98	4.02	0.19	3.32	0.245	3.34
	$K = 5$	0.98	4.99	0.205	4.46	0.23	4.455
	$K = 6$	0.935	5.935	0.215	5.11	0.225	5.095
RSC	$K = 2$	1.00	2.00	1.00	2.00	1.00	2.00
	$K = 3$	1.00	3.00	1.00	3.00	1.00	3.00
	$K = 4$	0.99	3.99	0.68	3.68	0.945	3.945
	$K = 5$	0.925	4.925	0.495	4.495	0.81	4.81
	$K = 6$	0.675	5.675	0.325	5.325	0.625	5.625

TABLE 4.4. Comparison of SMAST, CBIC and ICL for Simulation 2 (Binomial): $\rho = 0.04$, $r = 4$.

[2016]. In this network, any two characters (nodes) are connected by a count-weighted edge representing the number of co-occurrences between the pair in the same chapter of the book. There are six groups of characters corresponding to major subplots of the story. The estimated number of communities by some model-based approach in Newman and Reinert [2016] is 6.

A naive way to estimate the number of communities in this network is to treat it as unweighted and apply certain method for binary networks. We want to investigate the difference between the estimated numbers of communities when the network is treated as unweighted and weighted. For the unweighted case, we apply CBIC, ICL, as well as two popular methods for unweighted networks: the Bethe Hessian matrix (BH) and stepwise goodness-of-fit (StGoF) to the unweighted adjacency matrix to obtain the estimated number of communities $\widehat{K}$, as shown in Table 4.11. Notice that we use SCORE to cluster the nodes

		SMAST		CBIC		ICL	
		Rate	Mean	Rate	Mean	Rate	Mean
SCORE	$K = 2$	0.975	2.025	0.99	2.01	0.99	2.01
	$K = 3$	0.975	3.025	0.645	3.135	0.645	3.12
	$K = 4$	1.00	4.00	0.17	3.255	0.185	3.235
	$K = 5$	0.96	4.96	0.195	4.235	0.21	4.275
	$K = 6$	0.66	5.66	0.21	5.18	0.205	5.17
RSC	$K = 2$	1.00	2.00	1.00	2.00	1.00	2.00
	$K = 3$	1.00	3.00	1.00	3.00	1.00	3.00
	$K = 4$	0.985	3.985	0.64	3.64	0.865	3.865
	$K = 5$	0.84	4.84	0.435	4.435	0.715	4.715
	$K = 6$	0.33	5.33	0.19	5.19	0.38	5.38

TABLE 4.5. Comparison of SMAST, CBIC and ICL for Simulation 2 (Binomial): $\rho = 0.06$, $r = 3$.

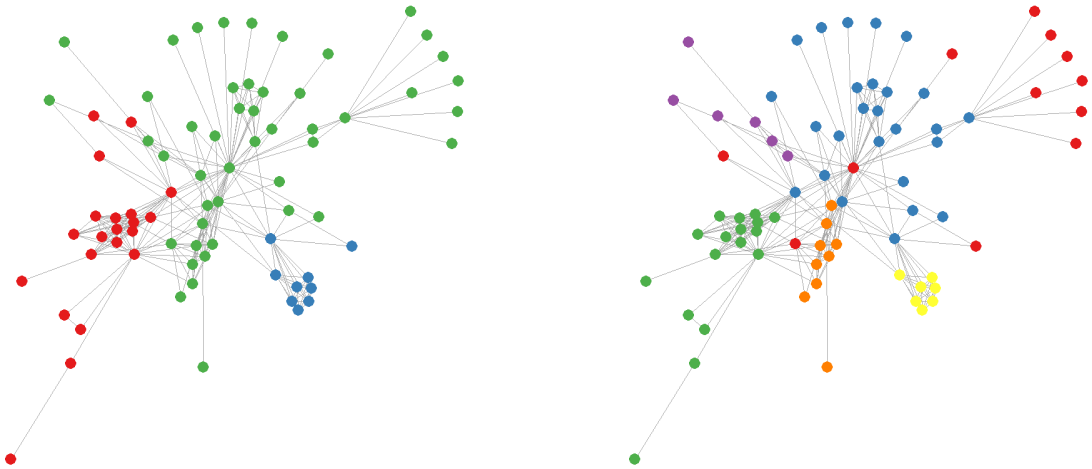
whenever necessary for the methods. For StGoF, the algorithm fails to stop, so we choose $\widehat{K} = 3$ as the number of communities minimizing the test statistic.

For the weighted case, we apply CBIC and ICL to the weighted adjacency while SMAST is applied to the regularized weighted adjacency with tuning parameter τ. Table 4.12 shows the estimated number of communities of the three methods with SCORE and RSC as the clustering methods.

In Figure 4.1, we plot the estimated node clusters by applying SCORE to the unweighted adjacency with 3 communities and to the weighted adjacency with 6 communities.

		SMAST		CBIC		ICL	
		Rate	Mean	Rate	Mean	Rate	Mean
SCORE	$K = 2$	1.00	2.00	1.00	2.00	1.00	2.00
	$K = 3$	1.00	3.00	0.915	3.065	0.915	3.065
	$K = 4$	0.99	3.99	0.30	3.38	0.32	3.39
	$K = 5$	0.84	4.84	0.205	4.31	0.23	4.34
	$K = 6$	0.255	5.255	0.115	5.135	0.125	5.145
RSC	$K = 2$	1.00	2.00	1.00	2.00	1.00	2.00
	$K = 3$	1.00	3.00	1.00	3.00	1.00	3.00
	$K = 4$	0.985	3.985	0.665	3.665	0.87	3.87
	$K = 5$	0.775	4.775	0.615	4.615	0.825	4.825
	$K = 6$	0.25	5.25	0.25	5.25	0.44	5.44

TABLE 4.6. Comparison of SMAST, CBIC and ICL for Simulation 2 (Binomial): $\rho = 0.12$, $r = 2$.



(a) Unweighted with 3 communities (b) Weighted with 6 communities

FIGURE 4.1. Estimated node clusters in Les Misèrable by applying SCORE to the unweighted adjacency with 3 communities and to the weighted adjacency with 6 communities.

		SMAST		CBIC		ICL	
		Rate	Mean	Rate	Mean	Rate	Mean
SCORE	$K = 2$	0.955	2.045	0.945	2.05	0.945	2.05
	$K = 3$	0.695	3.335	0.335	3.15	0.335	3.195
	$K = 4$	0.96	4.04	0.37	3.565	0.41	3.605
	$K = 5$	0.96	5.04	0.245	4.605	0.245	4.61
	$K = 6$	0.965	5.965	0.245	5.27	0.25	5.30
RSC	$K = 2$	0.995	2.005	1.00	2.00	1.00	2.00
	$K = 3$	0.995	3.005	1.00	3.00	1.00	3.00
	$K = 4$	1.00	4.00	0.80	3.80	0.97	3.97
	$K = 5$	0.955	4.955	0.715	4.715	0.87	4.87
	$K = 6$	0.725	5.725	0.435	5.435	0.685	5.685

TABLE 4.7. Comparison of SMAST, CBIC and ICL for Simulation 3 (Negative Binomial): $\rho = 0.04$, $r = 4$.

		SMAST		CBIC		ICL	
		Rate	Mean	Rate	Mean	Rate	Mean
SCORE	$K = 2$	0.97	2.03	0.995	1.995	0.995	1.995
	$K = 3$	0.905	3.10	0.44	3.23	0.44	3.195
	$K = 4$	0.96	4.03	0.23	3.395	0.24	3.39
	$K = 5$	0.96	4.96	0.205	4.415	0.20	4.43
	$K = 6$	0.87	5.87	0.245	5.39	0.26	5.40
RSC	$K = 2$	0.99	2.015	1.00	2.00	1.00	2.00
	$K = 3$	0.99	3.01	1.00	3.00	1.00	3.00
	$K = 4$	0.995	3.995	0.805	3.805	0.935	3.935
	$K = 5$	0.92	4.92	0.725	4.725	0.91	4.91
	$K = 6$	0.56	5.56	0.345	5.355	0.635	5.645

TABLE 4.8. Comparison of SMAST, CBIC and ICL for Simulation 3 (Negative Binomial): $\rho = 0.07$, $r = 3$.

		SMAST		CBIC		ICL	
		Rate	Mean	Rate	Mean	Rate	Mean
SCORE	$K = 2$	0.955	2.045	1.00	2.00	1.00	2.00
	$K = 3$	0.95	3.05	0.78	3.21	0.78	3.21
	$K = 4$	0.98	4.02	0.23	3.475	0.235	3.51
	$K = 5$	0.985	5.005	0.215	4.33	0.215	4.33
	$K = 6$	0.785	5.785	0.16	5.335	0.175	5.335
RSC	$K = 2$	0.91	2.10	1.00	2.00	1.00	2.00
	$K = 3$	0.965	3.035	0.995	3.005	0.995	3.005
	$K = 4$	0.945	4.055	0.93	3.93	0.985	3.985
	$K = 5$	0.95	5.00	0.815	4.815	0.94	4.94
	$K = 6$	0.69	5.69	0.495	5.495	0.67	5.67

TABLE 4.9. Comparison of SMAST, CBIC and ICL for Simulation 3 (Negative Binomial): $\rho = 0.12$, $r = 2$.

Name	n	K	$\bar{d}$	SMAST	GSURE
Football	115	11	10.66	8	10
Polbooks	105	2,3	8.4	3	3
Dolphins	62	2,4	5.13	2	2
Karate	34	2	4.59	1	2
UKfaculty	81	4	14.25	4	4
Polblogs	1222	2	27.36	2, 3	3

TABLE 4.10. Estimated K in real-world networks.

	BH	CBIC	ICL	StGoF
$\widehat{K}$	4	3	3	3^*

TABLE 4.11. Estimated K in the unweighted adjacency of Les Misèrable.

	SMAST ($\tau = 0.05$)	SMAST ($\tau = 0.1$)	SMAST ($\tau = 0.25$)	SMAST ($\tau = 0.5$)	CBIC	ICL
SCORE	7	6	7	6	7	7
RSC	7	7	6	6	7	5

TABLE 4.12. Estimated K by applying SMAST to the regularized weighted adjacency of Les Misèrable with different τ values and CBIC, ICL to the original weighted adjacency.

APPENDIX A

Appendix for Chapter 2

A.1. Preliminaries of Section 2.4

The following lemma provides a tail bound for the sum of independent random variables satisfying the Bernstein condition.

LEMMA A.1.1 (Bernstein's Inequality, Corollary 2.11 of Boucheron et al. [2013]). *Let $X_1, \cdots, X_n$ be independent real-valued random variables. Assume that there exist positive numbers v and c such that $\sum_{i=1}^n \mathbb{E}[X_i^2] \le v$ and*

$$\sum_{i=1}^n \mathbb{E}\,|X_i|^q \le \frac{q!}{2} c^{q-2} v \quad \text{for all intergers } q \ge 3.$$

Then for all $t > 0$,

$$\mathbb{P}\left(\sum_{i=1}^n (X_i - \mathbb{E}\,X_i) \ge t\right) \le \exp\left(-\frac{t^2}{2(v + ct)}\right).$$

Notice that this lemma can be viewed as a special case of the following vector Bernstein-type inequality.

LEMMA A.1.2 (Vector Bernstein-type inequality, Theorem 2.5 of Bosq [2000]). *If $\{\boldsymbol{X}_i\}_{i=1}^n$ are independent random vectors in a separable Hilbert space (where the norm is denoted by $\|\cdot\|$) with $\mathbb{E}[\boldsymbol{X}_i] = \mathbf{0}$ and*

$$\sum_{i=1}^n \mathbb{E}\,\|\boldsymbol{X}_i\|^p \le \frac{p!}{2}\sigma^2 R^{p-2}, \quad p = 2, 3, 4, \ldots$$

Then,

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \boldsymbol{X}_i\right\| \ge t\right) \le 2\exp\left(-\frac{t^2}{2(\sigma^2 + Rt)}\right), \quad \text{for all } t > 0.$$

The following lemma is an application of the vector Bernstein-type inequality supporting the proof of Lemma 2.4.4.

LEMMA A.1.3. *Let $X_1, \dots, X_n$ be independent random variables satisfying (2.18). Denote $\lambda_i = \mathbb{E}(X_i)$, and suppose $\lambda_{\max} = \max_{1 \le i \le n} \lambda_i \le C(c_0)$. Let $\boldsymbol{w}_1, \dots, \boldsymbol{w}_n \in \mathbb{R}^d$ be fixed vectors. Then for all $1 \le i \le n$, with probability at least $1 - n^{-4}$,*

$$\left\| \sum_{i=1}^{n} (X_i - \lambda_i) \boldsymbol{w}_i \right\|_2 \le C(c_0) \max \left(\sqrt{\lambda_{\max}} \|\boldsymbol{W}\|_F \sqrt{\log n}, \|\boldsymbol{W}\|_{2\to\infty} (\log n) \right),$$

where $\boldsymbol{W}^\top = [\boldsymbol{w}_1, \dots, \boldsymbol{w}_n]$.

PROOF. By (2.18), there holds

$$\begin{aligned}
\sum_{i=1}^{n} \mathbb{E} \|(X_i - \lambda_i) \boldsymbol{w}_i\|_2^p &= \sum_{i=1}^{n} (\mathbb{E} |X_i - \lambda_i|^p) \|\boldsymbol{w}_i\|_2^p \\
&\le \sum_{i=1}^{n} C' \left(\frac{p!}{2} \right) R^{p-2} \lambda_i \|\boldsymbol{w}_i\|_2^p \\
&\le \sum_{i=1}^{n} C' \left(\frac{p!}{2} \right) R^{p-2} \lambda_{\max} \|\boldsymbol{w}_i\|_2^2 \|\boldsymbol{W}\|_{2\to\infty}^{p-2} \\
&= \left(\frac{p!}{2} \right) (R \|\boldsymbol{W}\|_{2\to\infty})^{p-2} \left(C' \lambda_{\max} \|\boldsymbol{W}\|_F^2 \right)
\end{aligned}$$

for $p = 2, 3, 4, \dots$, where R and C' only depend on c_0. Then by Lemma A.1.2, for any $t > 0$,

$$\mathbb{P} \left(\left\| \sum_{i=1}^{n} (X_i - \lambda_i) \boldsymbol{w}_i \right\|_2 \ge t \right) \le 2 \exp \left(- \frac{t^2}{C' \lambda_{\max} \|\boldsymbol{W}\|_F^2 + (R \|\boldsymbol{W}\|_{2\to\infty}) t} \right).$$

We can choose

$$t = C(c_0) \max \left(\sqrt{\lambda_{\max}} \|\boldsymbol{W}\|_F \sqrt{\log n}, \ \|\boldsymbol{W}\|_{2\to\infty} (\log n) \right)$$

for some sufficiently large $C(c_0)$, and then prove the result. □

The following lemma is the subexponential case of matrix Bernstein inequality.

LEMMA A.1.4 (Theorem 6.2 of Tropp [2012]). *Consider a finite sequence* $\{\boldsymbol{X}_k\}$ *of independent, random, symmetric matrices with dimension* $\boldsymbol{d}$. *Assume that*

$$\mathbb{E}[\boldsymbol{X}_k] = \mathbf{0} \text{ and } \mathbb{E}[\boldsymbol{X}_k^p] \preceq \frac{p!}{2} R^{p-2} \boldsymbol{A}_k^2 \text{ for } p = 2, 3, 4, \ldots.$$

Compute the variance parameter

$$\sigma^2 := \left\| \sum_k \boldsymbol{A}_k^2 \right\|.$$

Then, the following chain of inequalities holds for all $t \geq 0$:

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_k \boldsymbol{X}_k\right) \geq t\right) \leq d \exp\left(-\frac{t^2}{2(\sigma^2 + Rt)}\right) \leq \begin{cases} d \exp\left(-t^2/4\sigma^2\right) & \text{for } t \leq \sigma^2/R; \\ d \exp\left(-t/4R\right) & \text{for } t \geq \sigma^2/R. \end{cases}$$

The following theorem from Abbe et al. [2020] provides perturbation bound for eigenspaces which is crucial to the proof of Theorem 2.2.2.

THEOREM A.1.1 ([Abbe et al., 2020]). *Suppose* $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ *is a symmetric random matrix, and let* $\boldsymbol{A}^* = \mathbb{E}[\boldsymbol{A}]$. *Denote the eigenvalues of* $\boldsymbol{A}$ *by* $\lambda_1 \geq \cdots \geq \lambda_n$, *and their associated eigenvectors by* $\{\boldsymbol{u}_j\}_{j=1}^n$. *Analogously for* $\boldsymbol{A}^*$, *the eigenvalues and eigenvectors are denoted by* $\lambda_1^* \geq \cdots \geq \lambda_n^*$ *and* $\{\boldsymbol{u}_j^*\}_{j=1}^n$. *For convenience, we also define* $\lambda_0 = \lambda_0^* = \infty$ *and* $\lambda_{n+1} = \lambda_{n+1}^* = -\infty$. *Note that we allow eigenvalues to be identical, so some eigenvectors may be defined up to rotations.*

Suppose r *and* s *are two integers satisfying* $1 \leq r \leq n$ *and* $0 \leq s \leq n - r$. *Let* $\boldsymbol{U} = [\boldsymbol{u}_{s+1}, \ldots, \boldsymbol{u}_{s+r}] \in \mathbb{R}^{n \times r}$, $\boldsymbol{U}^* = [\boldsymbol{u}_{s+1}^*, \ldots, \boldsymbol{u}_{s+r}^*] \in \mathbb{R}^{n \times r}$ *and* $\boldsymbol{\Lambda}^* = diag(\lambda_{s+1}^*, \ldots, \lambda_{s+r}^*) \in \mathbb{R}^{r \times r}$. *Define*

$$\Delta^* = (\lambda_s^* - \lambda_{s+1}^*) \wedge (\lambda_{s+r}^* - \lambda_{s+r+1}^*) \wedge \min_{1 \leq i \leq r} |\lambda_{s+i}^*|$$

and

$$\kappa = \max_{1 \leq i \leq r} |\lambda_{s+i}^*| / \Delta^*.$$

Suppose for some $\gamma \geq 0$, *the following assumptions hold:*

A1 (Incoherence) $\|\boldsymbol{A}^*\|_{2\to\infty} \le \gamma\Delta^*$.

A2 (Row- and column-wise independence) For any $m \in [n]$*, the entries in the* m*-th row and column of* $\boldsymbol{A}$ *are independent with other entries: namely,* $\{A_{ij}, i = m \text{ or } j = m\}$ *are independent of* $\{A_{ij} : i \ne m, j \ne m\}$.

A3 (Spectral norm concentration) $32\kappa \max\{\gamma, \varphi(\gamma)\} \le 1$ *and for some* $\delta_0 \in (0,1)$,

$$\mathbb{P}\left(\|\boldsymbol{A} - \boldsymbol{A}^*\| \le \gamma\Delta^*\right) \ge 1 - \delta_0.$$

A4 (Row concentration) Suppose $\varphi(x)$ *is continuous and non-decreasing in* $\mathbb{R}_+$ *with* $\varphi(0) = 0$, $\varphi(x)/x$ *is non-increasing in* $\mathbb{R}_+$, *and* $\delta_1 \in (0,1)$. *For any* $i \in [n]$ *and* $\boldsymbol{W} \in \mathbb{R}^{n\times r}$,

$$\mathbb{P}\left(\|(\boldsymbol{A} - \boldsymbol{A}^*)_{i\cdot}\boldsymbol{W}\|_2 \le \Delta^*\|\boldsymbol{W}\|_{2\to\infty}\varphi\left(\frac{\|\boldsymbol{W}\|_F}{\sqrt{n}\|\boldsymbol{W}\|_{2\to\infty}}\right)\right) \ge 1 - \frac{\delta_1}{n}.$$

Under Assumptions (A1)—(A4), with probability at least $1 - \delta_0 - 2\delta_1$, *there exists an orthogonal matrix* $\boldsymbol{Q}$ *such that*

$$\|\boldsymbol{U}\boldsymbol{Q} - \boldsymbol{A}\boldsymbol{U}^*(\boldsymbol{\Lambda}^*)^{-1}\|_{2\to\infty} \lesssim \kappa(\kappa + \varphi(1))(\gamma + \varphi(\gamma))\|\boldsymbol{U}^*\|_{2\to\infty} + \gamma\|\boldsymbol{A}^*\|_{2\to\infty}/\Delta^*. \tag{A.1}$$

Here the notation $\lesssim$ *only hides absolute constants.*

APPENDIX B

Appendix for Chapter 3

B.1. Extension to Pairwise Comparison Networks

The rank selection framework in Chapter 3 can be extended to pairwise comparison networks as well. Let's consider the Nonparametric Bradley-Terry model. Suppose for each pair of players $i < j$ from n players, they play k_{ij} games. Furthermore, assume that the k_{ij} games have independent and identically distributed outcomes, and the probability that player i beats j is $\widetilde{M}_{ij}$. Let the number of games in which i actually wins j be $\widetilde{A}_{ij}$, then we have

$$\widetilde{A}_{ij} \overset{\text{ind}}{\sim} \text{Binomial}(k_{ij}, \widetilde{M}_{ij}), \quad 1 \le i < j \le n.$$

Note that here we assume all games in the tournament are independent, and $\widetilde{A}_{ij}$'s are thereby independent. We assume there is no draw in this tournament, so for each $i < j$, the number of games that player j beats i is $\widetilde{A}_{ji} = k_{ij} - \widetilde{A}_{ij}$. Let $k_{ji} = k_{ij}$. By further denoting $\widetilde{M}_{ji} = 1 - \widetilde{M}_{ij}$, we have $\widetilde{A}_{ji} \overset{\text{ind}}{\sim} \text{Binomial}(k_{ji}, \widetilde{M}_{ji})$ for each $i < j$.

Let's now transform both $\widetilde{\boldsymbol{M}}$ and $\widetilde{\boldsymbol{A}}$ into skew-symmetric matrices. For each $i \neq j$, let $M_{ij} = 2\widetilde{M}_{ij} - 1$ and $A_{ij} = (2/k_{ij})\widetilde{A}_{ij} - 1$. Straightforward calculation gives $M_{ij} = -M_{ji}$ and $A_{ij} = -A_{ji}$. Also notice that $\mathbb{E}[A_{ij}] = M_{ij}$. By letting $M_{ii} = 0$ and $A_{ii} = 0$ for all $i = 1, \ldots, n$, we know both $\boldsymbol{A}$ and $\boldsymbol{M}$ are skew-symmetric. The fact $\mathbb{E}[\boldsymbol{A}] = \boldsymbol{M}$ implies that the transformed data matrix $\boldsymbol{A}$ is the empirical version of $\boldsymbol{M}$, and we aim to estimate the rank of $\boldsymbol{M}$ by studying the spectrum of $\boldsymbol{A}$.

Consider the Youla decomposition of $\boldsymbol{A}$:

$$\boldsymbol{A} = \sum_{k=1}^{\lfloor n/2 \rfloor} \lambda_k \left(\boldsymbol{\phi}_k \boldsymbol{\psi}_k^\top - \boldsymbol{\psi}_k \boldsymbol{\phi}_k^\top \right),$$

where $\lambda_1 > \ldots > \lambda_{\lfloor n/2 \rfloor} > 0$, and $\boldsymbol{\phi}_1, \boldsymbol{\psi}_1, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\phi}_{\lfloor n/2 \rfloor}, \boldsymbol{\psi}_{\lfloor n/2 \rfloor}$ are orthonormal. The Youla decomposition can also be represented in the matrix form $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top$, where

$$\boldsymbol{U} = [\boldsymbol{\phi}_1, \boldsymbol{\psi}_1, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\phi}_{\lfloor n/2 \rfloor}, \boldsymbol{\psi}_{\lfloor n/2 \rfloor}]$$

is an orthonormal basis matrix, and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0 & \lambda_1 & 0 & 0 & \ldots & 0 & 0 \\ -\lambda_1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & \ldots & 0 & 0 \\ 0 & 0 & -\lambda_2 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & \lambda_{\lfloor n/2 \rfloor} \\ 0 & 0 & 0 & 0 & \ldots & -\lambda_{\lfloor n/2 \rfloor} & 0 \end{bmatrix}. \tag{B.1}$$

Given $\boldsymbol{M}$ is also skew-symmetric, we know its rank must be even. Therefore, we can just consider the candidate rank r to be even. We propose to estimate $\boldsymbol{M}$ by the truncated Youla decomposition

$$\widehat{\boldsymbol{M}} = \sum_{k=1}^{r/2} \lambda_k \left(\boldsymbol{\phi}_k \boldsymbol{\psi}_k^\top - \boldsymbol{\psi}_k \boldsymbol{\phi}_k^\top \right).$$

Again, the question is to measure the discrepancy between $\widehat{\boldsymbol{M}}$ and $\boldsymbol{M}$, which relies on defining the apparent error $q(\widehat{\boldsymbol{M}}, \boldsymbol{A})$, the true error $Q(\widehat{\boldsymbol{M}}, \boldsymbol{M})$, and the the estimated error $\widehat{Q}(\widehat{\boldsymbol{M}}, \boldsymbol{M})$. Define $\widehat{\widetilde{M}}_{ij} = \frac{1+\widehat{M}_{ij}}{2}$. Then $\widehat{\widetilde{\boldsymbol{M}}}$ is an estimate of $\widetilde{\boldsymbol{M}}$. Suppose we know how to define the true error $\widetilde{Q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{M}})$ and apparent error $\tilde{q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{A}})$, and assume that we know to how to derive an estimated error $\widehat{Q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{M}})$. Then we can achieve the goal by letting $Q(\widehat{\boldsymbol{M}}, \boldsymbol{M}) \coloneqq \widetilde{Q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{M}})$, $q(\widehat{\boldsymbol{M}}, \boldsymbol{A}) \coloneqq \tilde{q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{A}})$ and $\widehat{Q}(\widehat{\boldsymbol{M}}, \boldsymbol{M}) \coloneqq \widehat{Q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{M}})$.

Our approach to risk estimation for binomial data is quite similar to the Bernoulli case. Assume that $y_i \overset{\text{ind}}{\sim} \text{Binomial}(k_i, \mu_i)$ for $i = 1, \ldots, n$. With some estimator $\hat{\boldsymbol{\mu}} = f(\boldsymbol{y})$, the

apparent error between $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{y}$ can be defined as

$$q(\hat{\boldsymbol{\mu}}, \boldsymbol{y}) = \sum_{i=1}^{n} \left[-y_i \log(\hat{\mu}_i) - (k_i - y_i) \log(1 - \hat{\mu}_i) \right],$$

and the corresponding true error is

$$Q(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) = \sum_{i=1}^{n} \left[-(k_i \mu_i) \log(\hat{\mu}_i) - k_i (1 - \mu_i) \log(1 - \hat{\mu}_i) \right].$$

As with the Bernoulli and Poisson cases,

$$\mathbb{E}\left[Q(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) - q\left(\hat{\boldsymbol{\mu}}, \boldsymbol{y}\right) \right] = \sum_{i=1}^{n} \mathbb{E}\left[(y_i - k_i \mu_i) \log \frac{\hat{\mu}_i}{1 - \hat{\mu}_i} \right] = \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left[(y_i - k_i \mu_i) \log \frac{\hat{\mu}_i}{1 - \hat{\mu}_i} \middle| \boldsymbol{y}_{(i)} \right] \right].$$

Taylor expansion gives

$$\begin{aligned} \log \frac{\hat{\mu}_i(y_i, \boldsymbol{y}_{(i)})}{1 - \hat{\mu}_i(y_i, \boldsymbol{y}_{(i)})} \approx & \log \frac{\hat{\mu}_i(k_i \mu_i, \boldsymbol{y}_{(i)})}{1 - \hat{\mu}_i(k_i \mu_i, \boldsymbol{y}_{(i)})} \\ & + \frac{1}{\hat{\mu}_i(k_i \mu_i, \boldsymbol{y}_{(i)}) \left[1 - \hat{\mu}_i(k_i \mu_i, \boldsymbol{y}_{(i)}) \right]} \frac{\partial \hat{\mu}_i}{\partial y_i}(k_i \mu_i, \boldsymbol{y}_{(i)})(y_i - k_i \mu_i). \end{aligned}$$

Thus we have the approximation

$$\begin{aligned} & \mathbb{E}\left[(y_i - k_i \mu_i) \log \frac{\hat{\mu}_i}{1 - \hat{\mu}_i} \middle| \boldsymbol{y}_{(i)} \right] \\ \approx & \mathbb{E}\left[(y_i - k_i \mu_i) \left(\log \frac{\hat{\mu}_i(k_i \mu_i, \boldsymbol{y}_{(i)})}{1 - \hat{\mu}_i(k_i \mu_i, \boldsymbol{y}_{(i)})} \right. \right. \\ & \left. \left. + \frac{1}{\hat{\mu}_i(k_i \mu_i, \boldsymbol{y}_{(i)}) \left[1 - \hat{\mu}_i(k_i \mu_i, \boldsymbol{y}_{(i)}) \right]} \frac{\partial \hat{\mu}_i}{\partial y_i}(k_i \mu_i, \boldsymbol{y}_{(i)})(y_i - k_i \mu_i) \right) \middle| \boldsymbol{y}_{(i)} \right] \\ = & \frac{k_i \mu_i (1 - \mu_i)}{\hat{\mu}_i(k_i \mu_i, \boldsymbol{y}_{(i)}) \left[1 - \hat{\mu}_i(k_i \mu_i, \boldsymbol{y}_{(i)}) \right]} \frac{\partial \hat{\mu}_i}{\partial y_i}(\mu_i, \boldsymbol{y}_{(i)}) \\ \approx & k_i \frac{\partial \hat{\mu}_i}{\partial y_i}(\mu_i, \boldsymbol{y}_{(i)}) \approx k_i \left. \frac{\partial \hat{\mu}_i}{\partial y_i} \right|_{\boldsymbol{y}}. \end{aligned}$$

We then define the estimated error as

$$\widehat{Q}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) = q(\hat{\boldsymbol{\mu}}, \boldsymbol{y}) + \sum_{i=1}^{n} k_i \frac{\partial \hat{\mu}_i}{\partial y_i}\bigg|_{\boldsymbol{y}}. \tag{B.2}$$

Let's come back to the problem of estimating the true error $\widetilde{Q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{M}})$. By the above results, we can define the true error $\widetilde{Q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{M}})$ and the apparent error $\tilde{q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{A}})$ as

$$\widetilde{Q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{M}}) = \sum_{1 \le i < j \le n} (-\log \widehat{\widetilde{M}}_{ij}) \left(k_{ij} \widetilde{M}_{ij} \right) + (-\log(1 - \widehat{\widetilde{M}}_{ij})) k_{ij} (1 - \widetilde{M}_{ij})$$

and

$$\tilde{q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{A}}) = \sum_{1 \le i < j \le n} (-\log \widehat{\widetilde{M}}_{ij}) \widetilde{A}_{ij} + (-\log(1 - \widehat{\widetilde{M}}_{ij}))(k_{ij} - \widetilde{A}_{ij}).$$

Then, the true error can be approximately estimated by

$$\widehat{Q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{M}}) = q(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{A}}) + \sum_{1 \le i < j \le n} k_{ij} \frac{\partial \widehat{\widetilde{M}}_{ij}}{\partial \widetilde{A}_{ij}}.$$

By plugging in $\widetilde{A}_{ij} = k_{ij} \left(\frac{1 + A_{ij}}{2} \right)$, $\widetilde{M}_{ij} = \frac{1 + M_{ij}}{2}$, and $\widehat{\widetilde{M}}_{ij} = \frac{1 + \widehat{M}_{ij}}{2}$, we have

$$\begin{aligned} Q(\widehat{\boldsymbol{M}}, \boldsymbol{M}) &\coloneqq \widetilde{Q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{M}}) \\ &= \sum_{1 \le i < j \le n} k_{ij} \left[\left(-\log \frac{1 + \widehat{M}_{ij}}{2} \right) \frac{1 + M_{ij}}{2} + \left(-\log \frac{1 - \widehat{M}_{ij}}{2} \right) \frac{1 - M_{ij}}{2} \right] \end{aligned}$$

and

$$\begin{aligned} q(\widehat{\boldsymbol{M}}, \boldsymbol{A}) &\coloneqq \tilde{q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{A}}) \\ &= \sum_{1 \le i < j \le n} k_{ij} \left[\left(-\log \frac{1 + \widehat{M}_{ij}}{2} \right) \frac{1 + A_{ij}}{2} + \left(-\log \frac{1 - \widehat{M}_{ij}}{2} \right) \frac{1 - A_{ij}}{2} \right]. \end{aligned}$$

Since $\frac{\partial \widehat{\widetilde{M}}_{ij}}{\partial \widetilde{A}_{ij}} = \frac{1}{k_{ij}} \frac{\partial \widehat{M}_{ij}}{\partial A_{ij}}$, we have $\sum_{1\le i<j\le n} k_{ij} \frac{\partial \widehat{\widetilde{M}}_{ij}}{\partial \widetilde{A}_{ij}} = \sum_{1\le i<j\le n} \frac{\partial \widehat{M}_{ij}}{\partial A_{ij}}$. Therefore, we have the following estimated error

$$\widehat{Q}(\widehat{\boldsymbol{M}}, \boldsymbol{M}) \coloneqq \widehat{Q}(\widehat{\widetilde{\boldsymbol{M}}}, \widetilde{\boldsymbol{M}}) = q(\widehat{\boldsymbol{M}}, \boldsymbol{A}) + \sum_{1\le i<j\le n} \frac{\partial \widehat{M}_{ij}}{\partial A_{ij}}.$$

Furthermore, since $\widehat{\boldsymbol{M}}$ is the truncated Youla decomposition, we can calculate the divergence as

$$\sum_{1\le i<j\le n} \frac{\partial \widehat{M}_{ij}}{\partial A_{ij}} = \left(n - \frac{r}{2} - \frac{1}{2}\right) r + \sum_{k=1}^{r/2} \sum_{l=r/2+1}^{\lfloor n/2 \rfloor} \frac{\lambda_l^2}{\lambda_k^2 - \lambda_l^2}.$$

The complete derivation of this formula is given in the next section.

B.2. Divergence Formula of Spectral Denoising for Skew-Symmetric Matrices

For simplicity, we only consider the case when n is even. Recall that

$$\boldsymbol{A} = \sum_{k=1}^{n/2} \lambda_k \left(\boldsymbol{\phi}_k \boldsymbol{\psi}_k^\top - \boldsymbol{\psi}_k \boldsymbol{\phi}_k^\top\right) = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^\top,$$

where $\boldsymbol{U} = [\boldsymbol{\phi}_1, \boldsymbol{\psi}_1, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\phi}_{n/2}, \boldsymbol{\psi}_{n/2}]$ consists of orthonormal columns, and $\boldsymbol{\Lambda}$ is a tridiagonal matrix as defined in (B.1).

As with the case of symmetric matrices, denote by $\mathbb{R}_n^{\mathrm{upp}} \subset \mathbb{R}^{n\times n}$ the space of $n \times n$ strictly upper triangular matrices. Denote $\boldsymbol{A}^{\mathrm{upp}}$ as the upper triangular matrix consisting of the entries of $\boldsymbol{A}$ above the diagonal. Then we have

$$\boldsymbol{A} = \boldsymbol{A}^{\mathrm{upp}} - (\boldsymbol{A}^{\mathrm{upp}})^\top.$$

Beyond spectral truncation, consider the general spectral estimator

$$\widehat{\boldsymbol{M}} = \sum_{k=1}^{n/2} g_k(\lambda_k) \left(\boldsymbol{\phi}_k \boldsymbol{\psi}_k^\top - \boldsymbol{\psi}_k \boldsymbol{\phi}_k^\top\right) \coloneqq \boldsymbol{U} g\left(\boldsymbol{\Lambda}\right) \boldsymbol{U}^\top \coloneqq f(\boldsymbol{A}^{\mathrm{upp}}),$$

where f is determined by g. At any point $\boldsymbol{A}^{\mathrm{upp}}$, the differentials $\mathrm{d}f|_{\boldsymbol{A}^{\mathrm{upp}}}(\boldsymbol{A}^{\mathrm{upp}})$, $\mathrm{d}\boldsymbol{U}|_{\boldsymbol{A}^{\mathrm{upp}}}(\boldsymbol{A}^{\mathrm{upp}})$, $\mathrm{d}\boldsymbol{\Lambda}|_{\boldsymbol{A}^{\mathrm{upp}}}(\boldsymbol{A}^{\mathrm{upp}})$, etc, are similarly defined as in Section 3.3.

As with the symmetric case, for any $\boldsymbol{\Lambda} \in \mathbb{R}_n^{\mathrm{upp}}$, denote $\boldsymbol{\Omega}[\boldsymbol{\Lambda}] = \boldsymbol{U}^\top \mathrm{d}\boldsymbol{U}[\boldsymbol{\Lambda}]$. By taking the differential on both sides of $\boldsymbol{I}_n = \boldsymbol{U}^\top \boldsymbol{U}$, we have that $\boldsymbol{\Omega}[\boldsymbol{\Lambda}] = \boldsymbol{U}^\top \mathrm{d}\boldsymbol{U}[\boldsymbol{\Lambda}]$ is skew-symmetric. Furthermore, by taking differentials on both sides of

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top = \boldsymbol{A}^{\mathrm{upp}} - (\boldsymbol{A}^{\mathrm{upp}})^\top,$$

we have

$$\boldsymbol{\Lambda} - \boldsymbol{\Lambda}^\top = \mathrm{d}\boldsymbol{U}[\boldsymbol{\Lambda}]\boldsymbol{\Lambda}\boldsymbol{U}^\top + \boldsymbol{U}\mathrm{d}\boldsymbol{\Lambda}[\boldsymbol{\Lambda}]\boldsymbol{U}^\top + \boldsymbol{U}\boldsymbol{\Lambda}\mathrm{d}\boldsymbol{U}[\boldsymbol{\Lambda}]^\top,$$

which implies that

$$\begin{aligned} \boldsymbol{U}^\top(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}^\top)\boldsymbol{U} &= \boldsymbol{\Omega}[\boldsymbol{\Lambda}]\boldsymbol{\Lambda} + \mathrm{d}\boldsymbol{\Lambda}[\boldsymbol{\Lambda}] + \boldsymbol{\Lambda}\boldsymbol{\Omega}[\boldsymbol{\Lambda}]^\top \\ &= \mathrm{d}\boldsymbol{\Lambda}[\boldsymbol{\Lambda}] + (\boldsymbol{\Omega}[\boldsymbol{\Lambda}]\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{\Omega}[\boldsymbol{\Lambda}])\,. \end{aligned} \tag{B.3}$$

For a skew-symmetric matrix $\boldsymbol{B} \in \mathbb{R}^{n\times n}$, denote the $(2k-1, 2k) \times (2l-1, 2l)$ block of size 2×2 in the matrix as $\boldsymbol{B}_{(k,l)}$, for all $k, l \in \{1, \ldots, n/2\}$. Since $\boldsymbol{\Omega}[\boldsymbol{\Lambda}]$ is skew-symmetric and $\boldsymbol{\Lambda}$ is of the form (B.1), it is easy to verify that for any $k = 1, \ldots, n/2$, the $(2k-1, 2k)\times(2k-1, 2k)$ diagonal block of $\boldsymbol{\Omega}[\boldsymbol{\Lambda}]\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{\Omega}[\boldsymbol{\Lambda}]$ is $\boldsymbol{0}$. On the other hand, $\mathrm{d}\boldsymbol{\Lambda}[\boldsymbol{\Lambda}]$ is a complementary block-diagonal matrix. Therefore, the $\{(2k-1, 2k) \times (2k-1, 2k) : k = 1, \ldots, n/2\}$ diagonal blocks of $\boldsymbol{U}^\top(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}^\top)\boldsymbol{U}$ are uniquely determined by $\mathrm{d}\boldsymbol{\Lambda}[\boldsymbol{\Lambda}]$ and off-diagonal blocks are uniquely determined by $\boldsymbol{\Omega}[\boldsymbol{\Lambda}]\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{\Omega}[\boldsymbol{\Lambda}]$.

To be concrete, for each $k = 1, \ldots, n/2$, the $(2k-1, 2k) \times (2k-1, 2k)$ diagonal block of $\boldsymbol{U}^\top(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}^\top)\boldsymbol{U}$ is

$$(\boldsymbol{U}^\top(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}^\top)\boldsymbol{U})_{(k,k)} = \mathrm{d}\lambda_k[\boldsymbol{\Lambda}] \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{B.4}$$

while for each $k, l \in \{1, \ldots, n/2\}$ and $k \neq l$, the $(2k-1, 2k) \times (2l-1, 2l)$ off-diagonal block of $\boldsymbol{U}^\top(\boldsymbol{\Delta} - \boldsymbol{\Delta}^\top)\boldsymbol{U}$ can be calculated as

$$(\boldsymbol{U}^\top(\boldsymbol{\Delta} - \boldsymbol{\Delta}^\top)\boldsymbol{U})_{(k,l)} = \begin{bmatrix} -\boldsymbol{\Omega}_{2k-1,2l}[\boldsymbol{\Delta}]\lambda_l - \boldsymbol{\Omega}_{2k,2l-1}[\boldsymbol{\Delta}]\lambda_k & \boldsymbol{\Omega}_{2k-1,2l-1}[\boldsymbol{\Delta}]\lambda_l - \boldsymbol{\Omega}_{2k,2l}[\boldsymbol{\Delta}]\lambda_k \\ -\boldsymbol{\Omega}_{2k,2l}[\boldsymbol{\Delta}]\lambda_l + \boldsymbol{\Omega}_{2k-1,2l-1}[\boldsymbol{\Delta}]\lambda_k & \boldsymbol{\Omega}_{2k,2l-1}[\boldsymbol{\Delta}]\lambda_l + \boldsymbol{\Omega}_{2k-1,2l}[\boldsymbol{\Delta}]\lambda_k \end{bmatrix}.$$

For notational simplicity, denote

$$(\boldsymbol{U}^\top(\boldsymbol{\Delta} - \boldsymbol{\Delta}^\top)\boldsymbol{U})_{(k,l)} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}. \tag{B.5}$$

Then solving for the $(2k-1, 2k) \times (2l-1, 2l)$ block of $\boldsymbol{\Omega}[\boldsymbol{\Delta}]$ yields

$$(\boldsymbol{\Omega}[\boldsymbol{\Delta}])_{(k,l)} = \frac{1}{\lambda_k^2 - \lambda_l^2} \begin{bmatrix} \gamma\lambda_k - \beta\lambda_l & \delta\lambda_k + \alpha\lambda_l \\ -\alpha\lambda_k - \delta\lambda_l & -\beta\lambda_k + \gamma\lambda_l \end{bmatrix}.$$

Let's now move to the differentials of $f(\boldsymbol{A}^{\text{upp}})$. The equality $\boldsymbol{U}g(\boldsymbol{\Lambda})\boldsymbol{U}^\top = f(\boldsymbol{A}^{\text{upp}})$ gives

$$\mathrm{d}f[\boldsymbol{\Delta}] = \mathrm{d}\boldsymbol{U}[\boldsymbol{\Delta}]g(\boldsymbol{\Lambda})\boldsymbol{U}^\top + \boldsymbol{U}\mathrm{d}(g \circ \boldsymbol{\Lambda})[\boldsymbol{\Delta}]\boldsymbol{U}^\top + \boldsymbol{U}g(\boldsymbol{\Lambda})\mathrm{d}\boldsymbol{U}[\boldsymbol{\Delta}]^\top,$$

which further implies

$$\begin{aligned} \boldsymbol{U}^\top\mathrm{d}f[\boldsymbol{\Delta}]\boldsymbol{U} &= \boldsymbol{\Omega}[\boldsymbol{\Delta}]g(\boldsymbol{\Lambda}) + \mathrm{d}(g \circ \boldsymbol{\Lambda})[\boldsymbol{\Delta}] + g(\boldsymbol{\Lambda})\boldsymbol{\Omega}[\boldsymbol{\Delta}]^\top \\ &= \mathrm{d}(g \circ \boldsymbol{\Lambda})[\boldsymbol{\Delta}] + (\boldsymbol{\Omega}[\boldsymbol{\Delta}]g(\boldsymbol{\Lambda}) - g(\boldsymbol{\Lambda})\boldsymbol{\Omega}[\boldsymbol{\Delta}]). \end{aligned} \tag{B.6}$$

As with (B.3), the $\{(2k-1, 2k) \times (2k-1, 2k) : k = 1, \ldots, \lfloor n/2 \rfloor\}$ diagonal blocks of $\boldsymbol{U}^\top\mathrm{d}f[\boldsymbol{\Delta}]\boldsymbol{U}$ are uniquely determined by $\mathrm{d}(g \circ \boldsymbol{\Lambda})[\boldsymbol{\Delta}]$ and the off-diagonal blocks are uniquely determined

by $\boldsymbol{\Omega}[\boldsymbol{\Delta}]g(\boldsymbol{\Lambda}) - g(\boldsymbol{\Lambda})\boldsymbol{\Omega}[\boldsymbol{\Delta}]$. To be concrete, for each $k = 1, \ldots, n/2$,

$$\begin{aligned}(\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{\Delta}]\boldsymbol{U})_{(k,k)} &= (\mathrm{d}(g \circ \boldsymbol{\Lambda})[\boldsymbol{\Delta}])_{(k,k)} \\ &= g_k'(\lambda_k)\mathrm{d}\lambda_k[\boldsymbol{\Delta}] \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \\ &= g_k'(\lambda_k)(\boldsymbol{U}^\top(\boldsymbol{\Delta} - \boldsymbol{\Delta}^\top)\boldsymbol{U})_{(k,k)}, \end{aligned} \tag{B.7}$$

while for each $k, l \in \{1, \ldots, n/2\}$ and $k \neq l$, the $(2k-1, 2k) \times (2l-1, 2l)$ block of $\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{\Delta}]\boldsymbol{U}$ is

$$\begin{aligned} &(\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{\Delta}]\boldsymbol{U})_{(k,l)} \\ =&(\boldsymbol{\Omega}[\boldsymbol{\Delta}]g(\boldsymbol{\Lambda}) - g(\boldsymbol{\Lambda})\boldsymbol{\Omega}[\boldsymbol{\Delta}])_{(k,l)} \\ =&\begin{bmatrix} -\boldsymbol{\Omega}_{2k-1,2l}[\boldsymbol{\Delta}]g_l(\lambda_l) - \boldsymbol{\Omega}_{2k,2l-1}[\boldsymbol{\Delta}]g_k(\lambda_k) & \boldsymbol{\Omega}_{2k-1,2l-1}[\boldsymbol{\Delta}]g_l(\lambda_l) - \boldsymbol{\Omega}_{2k,2l}[\boldsymbol{\Delta}]g_k(\lambda_k) \\ -\boldsymbol{\Omega}_{2k,2l}[\boldsymbol{\Delta}]g_l(\lambda_l) + \boldsymbol{\Omega}_{2k-1,2l-1}[\boldsymbol{\Delta}]g_k(\lambda_k) & \boldsymbol{\Omega}_{2k,2l-1}[\boldsymbol{\Delta}]g_l(\lambda_l) + \boldsymbol{\Omega}_{2k-1,2l}[\boldsymbol{\Delta}]g_k(\lambda_k) \end{bmatrix} \\ =&\frac{1}{\lambda_l^2 - \lambda_k^2}\begin{bmatrix} \alpha \cdot p(\lambda_k, \lambda_l) + \delta \cdot q(\lambda_k, \lambda_l) & \beta \cdot p(\lambda_k, \lambda_l) - \gamma \cdot q(\lambda_k, \lambda_l) \\ \gamma \cdot p(\lambda_k, \lambda_l) - \beta \cdot q(\lambda_k, \lambda_l) & \delta \cdot p(\lambda_k, \lambda_l) + \alpha \cdot q(\lambda_k, \lambda_l) \end{bmatrix} \\ =&\frac{p(\lambda_k, \lambda_l)}{\lambda_k^2 - \lambda_l^2}\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} + \frac{q(\lambda_k, \lambda_l)}{\lambda_k^2 - \lambda_l^2}\begin{bmatrix} \delta & -\gamma \\ -\beta & \alpha \end{bmatrix}, \end{aligned} \tag{B.8}$$

where $\alpha, \beta, \gamma, \delta$ are defined in (B.5) and

$$\begin{cases} p(\lambda_k, \lambda_l) \coloneqq \lambda_k g_k(\lambda_k) - \lambda_l g_l(\lambda_l), \\ q(\lambda_k, \lambda_l) \coloneqq \lambda_l g_k(\lambda_k) - \lambda_k g_l(\lambda_l). \end{cases}$$

For any $i < j$, we still denote $\boldsymbol{E}^{ij}$ as the $n \times n$ matrix whose the (i, j)-th entry is 1 while all other entries are 0, and $\boldsymbol{L}^{ij} = \boldsymbol{U}^\top \boldsymbol{E}^{ij} \boldsymbol{U}$. Recall that $\boldsymbol{U} = [\boldsymbol{\phi}_1, \boldsymbol{\psi}_1, \boldsymbol{\phi}_2, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\phi}_{n/2}, \boldsymbol{\psi}_{n/2}]$,

which implies the $(2k-1,2k)\times(2l-1,2l)$ block of $\boldsymbol{L}^{ij}$ is

$$\boldsymbol{L}^{ij}_{(k,l)} = \begin{bmatrix} \phi_{ik}\phi_{jl} & \phi_{ik}\psi_{jl} \\ \psi_{ik}\phi_{jl} & \psi_{ik}\psi_{jl} \end{bmatrix}.$$

Since $\{\boldsymbol{E}^{ij}\}_{1\le i<j\le n}$ is the canonical basis for $\mathbb{R}_n^{\mathrm{upp}}$, we have

$$\begin{aligned}
\sum_{1\le i<j\le n} \frac{\partial \widehat{M}_{ij}}{\partial A_{ij}} &= \sum_{1\le i<j\le n} \langle \boldsymbol{E}^{ij}, \mathrm{d}f[\boldsymbol{E}^{ij}]\rangle \\
&= \sum_{1\le i<j\le n} \langle \boldsymbol{L}^{ij}, \boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{E}^{ij}]\boldsymbol{U}\rangle \\
&= \underbrace{\sum_{1\le i<j\le n}\sum_{k=1}^{n/2}\sum_{l\neq k} \langle \boldsymbol{L}^{ij}_{(k,l)}, (\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{E}^{ij}]\boldsymbol{U})_{(k,l)}\rangle}_{S_1} \\
&\quad + \underbrace{\sum_{1\le i<j\le n}\sum_{k=1}^{n/2} \langle \boldsymbol{L}^{ij}_{(k,k)}, (\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{E}^{ij}]\boldsymbol{U})_{(k,k)}\rangle}_{S_2}.
\end{aligned}$$

Note that for each $k,l\in\{1,\dots,n/2\}$,

$$\begin{aligned}
(\boldsymbol{U}^\top(\boldsymbol{E}^{ij}-\boldsymbol{E}^{ji})\boldsymbol{U})_{(k,l)} &= \begin{bmatrix} L^{ij}_{2k-1,2l-1} - L^{ji}_{2k-1,2l-1} & L^{ij}_{2k-1,2l} - L^{ji}_{2k-1,2l} \\ L^{ij}_{2k,2l-1} - L^{ji}_{2k,2l-1} & L^{ij}_{2k,2l} - L^{ji}_{2k,2l} \end{bmatrix} \\
&= \begin{bmatrix} \phi_{ik}\phi_{jl} - \phi_{jk}\phi_{il} & \phi_{ik}\psi_{jl} - \phi_{jk}\psi_{il} \\ \psi_{ik}\phi_{jl} - \psi_{jk}\phi_{il} & \psi_{ik}\psi_{jl} - \psi_{jk}\psi_{il} \end{bmatrix}.
\end{aligned}$$

Therefore, plugging $\boldsymbol{\Delta} = \boldsymbol{E}^{ij}$ into (B.8) gives

$$\begin{aligned}
(\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{E}^{ij}]\boldsymbol{U})_{(k,l)} =& \frac{p(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2} \begin{bmatrix} \phi_{ik}\phi_{jl} - \phi_{jk}\phi_{il} & \phi_{ik}\psi_{jl} - \phi_{jk}\psi_{il} \\ \psi_{ik}\phi_{jl} - \psi_{jk}\phi_{il} & \psi_{ik}\psi_{jl} - \psi_{jk}\psi_{il} \end{bmatrix} \\
&+ \frac{q(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2} \begin{bmatrix} \psi_{ik}\psi_{jl} - \psi_{jk}\psi_{il} & -(\psi_{ik}\phi_{jl} - \psi_{jk}\phi_{il}) \\ -(\phi_{ik}\psi_{jl} - \phi_{jk}\psi_{il}) & \phi_{ik}\phi_{jl} - \phi_{jk}\phi_{il} \end{bmatrix},
\end{aligned} \tag{B.9}$$

and plugging $\mathbf{\Delta} = \boldsymbol{E}^{ij}$ into (B.7) gives

$$(\boldsymbol{U}^\intercal \mathrm{d}f[\boldsymbol{E}^{ij}]\boldsymbol{U})_{(k,k)} = g_k'(\lambda_k)(\phi_{ik}\psi_{jk} - \phi_{jk}\psi_{ik}) \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{B.10}$$

By plugging (B.9) and (B.10) into S_1 and S_2, after some tedious calculation, we obtain

$$\begin{aligned}
S_1 = \sum_{1\le i<j\le n} \sum_{k=1}^{n/2} \sum_{l\neq k} \frac{p(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2} \Big(& \phi_{ik}\phi_{jl}(\phi_{ik}\phi_{jl} - \phi_{jk}\phi_{il}) + \phi_{ik}\psi_{jl}(\phi_{ik}\psi_{jl} - \phi_{jk}\psi_{il}) \\
& + \psi_{ik}\phi_{jl}(\psi_{ik}\phi_{jl} - \psi_{jk}\phi_{il}) + \psi_{ik}\psi_{jl}(\psi_{ik}\psi_{jl} - \psi_{jk}\psi_{il}) \Big) \\
& + \frac{q(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2} \Big(\phi_{ik}\phi_{jl}(\psi_{ik}\psi_{jl} - \psi_{jk}\psi_{il}) - \phi_{ik}\psi_{jl}(\psi_{ik}\phi_{jl} - \psi_{jk}\phi_{il}) \\
& - \psi_{ik}\phi_{jl}(\phi_{ik}\psi_{jl} - \phi_{jk}\psi_{il}) + \psi_{ik}\psi_{jl}(\phi_{ik}\phi_{jl} - \phi_{jk}\phi_{il}) \Big) \\
= \sum_{1\le i<j\le n} \sum_{k<l} \Bigg[& \frac{p(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2} \Big((\phi_{ik}\phi_{jl} - \phi_{jk}\phi_{il})^2 + (\phi_{ik}\psi_{jl} - \phi_{jk}\psi_{il})^2 + (\psi_{ik}\phi_{jl} - \psi_{jk}\phi_{il})^2 \\
& + (\psi_{ik}\psi_{jl} - \psi_{jk}\psi_{il})^2 \Big) + 2\frac{q(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2} \Big(\phi_{ik}\phi_{il}\psi_{jk}\psi_{jl} + \phi_{jk}\phi_{jl}\psi_{ik}\psi_{il} \\
& - \phi_{ik}\psi_{il}\phi_{jl}\psi_{jk} - \phi_{jk}\psi_{jl}\phi_{il}\psi_{ik} \Big) \Bigg] \\
= \sum_{k<l} \Bigg\{ & \frac{p(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2} \sum_{i<j} \Big((\phi_{ik}\phi_{jl} - \phi_{jk}\phi_{il})^2 + (\phi_{ik}\psi_{jl} - \phi_{jk}\psi_{il})^2 + (\psi_{ik}\phi_{jl} - \psi_{jk}\phi_{il})^2 \\
& + (\psi_{ik}\psi_{jl} - \psi_{jk}\psi_{il})^2 \Big) + 2\frac{q(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2} \sum_{i<j} \Big(\phi_{ik}\phi_{il}\psi_{jk}\psi_{jl} + \phi_{jk}\phi_{jl}\psi_{ik}\psi_{il} \\
& - \phi_{ik}\psi_{il}\phi_{jl}\psi_{jk} - \phi_{jk}\psi_{jl}\phi_{il}\psi_{ik} \Big) \Bigg\}
\end{aligned}$$

(Continued from previous page)

$$
\begin{aligned}
S_1 = &\sum_{k<l}\left\{\frac{p(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2}\sum_{i\neq j}\Big(\phi_{ik}^2\phi_{jl}^2-\phi_{ik}\phi_{jl}\phi_{jk}\phi_{il}+\phi_{ik}^2\psi_{jl}^2-\phi_{ik}\psi_{jl}\phi_{jk}\psi_{il}+\psi_{ik}^2\phi_{jl}^2-\psi_{ik}\phi_{jl}\psi_{jk}\phi_{il}\right.\\
&\left.+\psi_{ik}^2\psi_{jl}^2-\psi_{ik}\psi_{jl}\psi_{jk}\psi_{il}\Big)+2\frac{q(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2}\sum_{i\neq j}\Big(\phi_{ik}\phi_{il}\psi_{jk}\psi_{jl}-\phi_{ik}\psi_{il}\phi_{jl}\psi_{jk}\Big)\right\}\\
=&\sum_{k<l}\left\{\frac{p(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2}\sum_{i=1}^{n}\Big(\phi_{ik}^2(1-\phi_{il}^2)-\phi_{ik}\phi_{il}(-\phi_{ik}\phi_{il})+\phi_{ik}^2(1-\psi_{il}^2)-\phi_{ik}\psi_{il}(-\phi_{ik}\psi_{il})\right.\\
&+\psi_{ik}^2(1-\phi_{il}^2)-\psi_{ik}\phi_{il}(-\psi_{ik}\phi_{il})+\psi_{ik}^2(1-\psi_{il}^2)-\psi_{ik}\psi_{il}(-\psi_{ik}\psi_{il})\Big)\\
&\left.+2\frac{q(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2}\sum_{i=1}^{n}\phi_{ik}\phi_{il}(-\psi_{ik}\psi_{il})-\phi_{ik}\psi_{il}(-\phi_{il}\psi_{ik})\right\}\\
=&\sum_{k<l}\frac{p(\lambda_k,\lambda_l)}{\lambda_k^2-\lambda_l^2}\sum_{i=1}^{n}\Big(\phi_{ik}^2+\phi_{ik}^2+\psi_{ik}^2+\psi_{ik}^2\Big)\\
=&4\sum_{1\le k<l\le\frac{n}{2}}\frac{\lambda_k g_k(\lambda_k)-\lambda_l g_l(\lambda_l)}{\lambda_k^2-\lambda_l^2},
\end{aligned}
$$

and

$$
\begin{aligned}
S_2 &= \sum_{1\le i<j\le n}\sum_{k=1}^{n/2}\langle L_{(k,k)}^{ij},(\boldsymbol{U}^\top \mathrm{d}f[\boldsymbol{E}^{ij}]\boldsymbol{U})_{(k,k)}\rangle\\
&= \sum_{1\le i<j\le n}\sum_{k=1}^{n/2} g_k'(\lambda_k)(\phi_{ik}\psi_{jk}-\phi_{jk}\psi_{ik})^2\\
&= \sum_{k=1}^{n/2} g_k'(\lambda_k)\sum_{i\neq j}\Big(\phi_{ik}^2\psi_{jk}^2-\phi_{ik}\phi_{jk}\psi_{ik}\psi_{jk}\Big)\\
&= \sum_{k=1}^{n/2} g_k'(\lambda_k)\sum_{i=1}^{n}\Big(\phi_{ik}^2(1-\psi_{ik}^2)-\phi_{ik}\psi_{ik}(-\phi_{ik}\psi_{ik})\Big)\\
&= \sum_{k=1}^{n/2} g_k'(\lambda_k).
\end{aligned}
$$

If the spectral estimation is the truncated Youla decomposition with rank r such that $\widehat{\boldsymbol{M}} = \sum_{k=1}^{r/2} \lambda_k \left(\boldsymbol{\phi}_k \boldsymbol{\psi}_k^\top - \boldsymbol{\psi}_k \boldsymbol{\phi}_k^\top\right)$, the divergence will be

$$\begin{aligned}\sum_{1\le i<j\le n} \frac{\partial \widehat{M}_{ij}}{\partial A_{ij}} &= \frac{r^2}{2} - \frac{r}{2} + 4 \sum_{k=1}^{r/2} \sum_{l=r/2+1}^{n/2} \frac{\lambda_k^2}{\lambda_k^2 - \lambda_l^2} \\ &= \left(n - \frac{r}{2} - \frac{1}{2}\right) r + 4 \sum_{k=1}^{r/2} \sum_{l=r/2+1}^{n/2} \frac{\lambda_l^2}{\lambda_k^2 - \lambda_l^2}.\end{aligned}$$

Bibliography

Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2015.

Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of statistics*, 48(3):1452, 2020.

Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43, 2005.

Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

Arash A Amini, Aiyou Chen, Peter J Bickel, and Elizaveta Levina. Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41(4): 2097–2122, 2013.

Brian Ball, Brian Karrer, and Mark EJ Newman. Efficient and principled method for detecting communities in networks. *Physical Review E*, 84(3):036103, 2011.

Afonso S Bandeira and Ramon Van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506, 2016.

Peter J Bickel and Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50): 21068–21073, 2009.

Jérémie Bigot, Charles Deledalle, and Delphine Féral. Generalized sure for optimal shrinkage of singular values in low-rank matrix denoising. *Journal of Machine Learning Research*, 18(137):1–50, 2017.

Denis Bosq. *Stochastic Processes and Random Variables in Function Spaces*, pages 15–42. Springer New York, New York, NY, 2000.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

E. J. Candès, C. A. Sing-Long, and J. D. Trzasko. Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Transactions on Signal Processing*, 61(19): 4643–4657, Oct 2013. ISSN 1053-587X. doi: 10.1109/TSP.2013.2270464.

Kehui Chen and Jing Lei. Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association*, 113(521):241–251, 2018.

Yudong Chen, Xiaodong Li, and Jiaming Xu. Convexified modularity maximization for degree-corrected stochastic block models. *Ann. Statist.*, 46(4):1573–1602, 08 2018. doi: 10.1214/17-AOS1595.

J-J Daudin, Franck Picard, and Stéphane Robin. A mixture model for random graphs. *Statistics and computing*, 18(2):173–183, 2008.

Charles-Alban Deledalle, Samuel Vaiter, Gabriel Peyré, Jalal Fadili, and Charles Dossal. Risk estimation for matrix recovery with spectral regularization. *arXiv preprint arXiv:1205.1482*, 2012.

David L Donoho and Iain M Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association*, 90(432):1200–1224, 1995.

A Edelman. Finite random matrix theory. jacobians of matrix transforms (without wedge products)(2005), 2005.

Bradley Efron. How biased is the apparent error rate of a prediction rule? *Journal of the American Statistical Association*, 81(394):461–470, 1986. ISSN 01621459.

Bradley Efron, Prabir Burman, L. Denby, J. M. Landwehr, C. L. Mallows, Xiaotong Shen, Hsin-Cheng Huang, Jianming Ye, Jimmy Ye, and Chunming Zhang. The estimation of prediction error: Covariance penalties and cross-validation [with comments, rejoinder]. *Journal of the American Statistical Association*, 99(467):619–642, 2004. ISSN 01621459.

Santo Fortunato. Community detection in graphs. *Physics Reports*, 3(486):75–174, 2010.

Chao Gao, Yu Lu, and Harrison H Zhou. Rate-optimal graphon estimation. *The Annals of Statistics*, 43(6):2624–2652, 2015.

Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

Olivier Guédon and Roman Vershynin. Community detection in sparse networks via grothendieck's inequality. *Probability Theory and Related Fields*, pages 1–25, 2015.

Wendi Han and Guangyue Han. A new proof of hopf's inequality using a complex extension of the hilbert metric, 2019. URL https://arxiv.org/abs/1906.04875.

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

Jianwei Hu, Hong Qin, Ting Yan, and Yunpeng Zhao. Corrected bayesian information criterion for stochastic block models. *Journal of the American Statistical Association*, 115 (532):1771–1783, 2020.

H Malcolm Hudson. A natural identity for exponential families with applications in multiparameter estimation. *The Annals of Statistics*, 6(3):473–484, 1978.

W. James and Charles Stein. Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 361–379, Berkeley, Calif., 1961. University of California

Press.

Jiashun Jin. Fast community detection by score. *The Annals of Statistics*, 43(1):57–89, 2015.

Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. Estimating network memberships by simplex vertex hunting. *arXiv preprint arXiv:1708.07852*, 2017.

Jiashun Jin, Zheng Tracy Ke, Shengming Luo, and Minzhe Wang. Optimal estimation of the number of communities. *Journal of the American Statistical Association*, (just-accepted): 1–41, 2022.

Antony Joseph and Bin Yu. Impact of regularization on spectral clustering. *The Annals of Statistics*, 44(4):1765–1791, 2016.

Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1):016107, 2011.

Philip A Knight, Daniel Ruiz, and Bora Uçar. A symmetry preserving algorithm for matrix scaling. *SIAM journal on Matrix Analysis and Applications*, 35(3):931–955, 2014.

Donald Ervin Knuth. *The Stanford GraphBase: a platform for combinatorial computing*. AcM Press New York, 1993.

Boris Landa. Scaling positive random matrices: concentration and asymptotic convergence. *arXiv preprint arXiv:2012.06393*, 2020.

Boris Landa, Thomas TCK Zhang, and Yuval Kluger. Biwhitening reveals the rank of a count matrix. *arXiv preprint arXiv:2103.13840*, 2021.

Rafał Latała, Ramon van Handel, and Pierre Youssef. The dimension-free structure of nonhomogeneous random matrices. *Inventiones mathematicae*, 214(3):1031–1080, 2018.

Pierre Latouche, Etienne Birmele, and Christophe Ambroise. Variational bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, 12(1):93–115, 2012.

Can M Le and Elizaveta Levina. Estimating the number of communities in networks by spectral methods. *arXiv preprint arXiv:1507.00827*, 2015.

Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

Adrian S Lewis and Hristo S Sendov. Twice differentiable spectral functions. *SIAM Journal on Matrix Analysis and Applications*, 23(2):368–386, 2001.

Tianxi Li, Elizaveta Levina, and Ji Zhu. Network cross-validation by edge sampling. *Biometrika*, 107(2):257–276, 2020.

David Lusseau, Karsten Schneider, Oliver J Boisseau, Patti Haase, Elisabeth Slooten, and Steve M Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003.

Shujie Ma, Liangjun Su, and Yichong Zhang. Determining the number of communities in degree-corrected stochastic block models. *Journal of Machine Learning Research*, 22(69): 1–63, 2021.

Colin L Mallows. Some comments on cp. *Technometrics*, 42(1):87–94, 2000.

Tamás Nepusz, Andrea Petróczi, László Négyessy, and Fülöp Bazsó. Fuzzy communities and the concept of bridgeness in complex networks. *Physical Review E*, 77(1):016107, 2008.

Mark EJ Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104, 2006.

Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

Mark EJ Newman and Gesine Reinert. Estimating the number of communities in a network. *Physical review letters*, 117(7):078301, 2016.

Théodore Papadopoulo and Manolis IA Lourakis. Estimating the jacobian of the singular value decomposition: Theory and applications. In *European Conference on Computer*

Vision, pages 554–570. Springer, 2000.

Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

D Franco Saldana, Yi Yu, and Yang Feng. How many communities are there? *Journal of Computational and Graphical Statistics*, 26(1):171–181, 2017.

Warren Schudy and Maxim Sviridenko. Bernstein-like concentration and moment inequalities for polynomials of independent random variables: multilinear case. *arXiv preprint arXiv:1109.5193*, 2011.

Richard Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, 74(4):402–405, 1967.

Charles M. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9(6):1135–1151, 11 1981. doi: 10.1214/aos/1176345632.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. ISSN 00359246.

Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

YX Rachel Wang and Peter J Bickel. Likelihood-based model selection for stochastic block models. *The Annals of Statistics*, 45(2):500–528, 2017.

Ming Yuan. Degrees of freedom in low rank matrix estimation. *Science China Mathematics*, 59(12):2485–2502, Dec 2016. ISSN 1869-1862. doi: 10.1007/s11425-016-0426-8.

Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473, 1977.

Yuan Zhang, Elizaveta Levina, and Ji Zhu. Detecting overlapping communities in networks using spectral methods. *arXiv preprint arXiv:1412.3432*, 2014.

Yunpeng Zhao, Elizaveta Levina, and Ji Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):

2266–2292, 2012.

Hui Zou, Trevor Hastie, and Robert Tibshirani. On the degrees of freedom of the lasso. *Ann. Statist.*, 35(5):2173–2192, 10 2007. doi: 10.1214/009053607000000127.