Optimal network membership estimation under severe degree heterogeneity

Zheng Tracy Ke and Jingming Wang
Harvard University

Abstract

Real networks often have severe degree heterogeneity. We are interested in studying the effect of degree heterogeneity on estimation of the underlying community structure. We consider the degree-corrected mixed membership model (DCMM) for a symmetric network with $n$ nodes and $K$ communities, where each node $i$ has a degree parameter $\theta_i$ and a mixed membership vector $\pi_i$. The level of degree heterogeneity is captured by $F_n(\cdot)$ – the empirical distribution associated with $n$ (scaled) degree parameters. We first show that the optimal rate of convergence for the $\ell^1$-loss of estimating $\pi_i$’s depends on an integral with respect to $F_n(\cdot)$. We call a method optimally adaptive to degree heterogeneity (in short, optimally adaptive) if it attains the optimal rate for arbitrary $F_n(\cdot)$. Unfortunately, none of the existing methods satisfy this requirement. We propose a new spectral method that is optimally adaptive, the core idea behind which is using a pre-PCA normalization to yield the optimal signal-to-noise ratio simultaneously at all entries of each leading empirical eigenvector. As one technical contribution, we derive a new row-wise large-deviation bound for eigenvectors of the regularized graph Laplacian.

Keywords. DCMM, entry-wise eigenvector analysis, graph Laplacian, least-favorable configuration, leave-one-out, random matrix theory, SCORE, vertex hunting.

AMS 2010 subject classification. 62C20, 62H30, 91C20, 62P25.
1 Introduction

In the analysis of large social network data, mixed membership estimation is a problem of great interest Airoldi et al. (2008). Let $A \in \mathbb{R}^{n \times n}$ be the adjacency matrix of an undirected network with $n$ nodes, where

$$A(i, j) = \begin{cases} 1, & \text{if nodes } i \text{ and } j \text{ have an edge,} \\ 0, & \text{otherwise.} \end{cases} 1 \leq i, j \leq n.$$ 

The diagonals of $A$ are zero since we do not allow for self-edges. Suppose the network has $K$ perceivable communities $C_1, C_2, \ldots, C_K$. For each node $1 \leq i \leq n$, there is a Probability Mass Function (PMF) $\pi_i = (\pi_i(1), \pi_i(2), \ldots, \pi_i(K))' \in \mathbb{R}^K$ such that

$$\pi_i(k) \text{ is the weight that node } i \text{ puts on } C_k, \quad 1 \leq k \leq K.$$ 

We call node $i$ a pure node if $\pi_i$ is degenerate (i.e., one entry is 1 and the other entries are 0) and a mixed node otherwise. The goal is to estimate these membership vectors $\pi_1, \pi_2, \ldots, \pi_n$ from the data matrix $A$. This problem has found applications in learning research interests of statisticians from co-citation networks Ji et al. (2022) and understanding developmental brain disorders from gene co-expression networks Liu et al. (2018).

Various methods have been proposed for mixed membership estimation and inference. The Bayesian approach (Airoldi et al., 2008) puts a Dirichlet prior on $\pi_i$’s and uses variational inference to get the posteriors. The spectral approach (Jin et al., 2017; Zhang et al., 2020) estimates $\pi_i$’s from the leading eigenvectors of $A$; for example, Jin et al. (2017) discovered a simplex structure in the spectral domain and transformed membership estimation to a simplex vertex hunting problem. Fan et al. (2022) considered testing $\pi_i = \pi_j$ for two given nodes $i$ and $j$ and proposed eigenvector-based test statistics that have tractable null distributions. Despite these progresses in the literature, the optimal rate of mixed membership estimation still remains unknown. In this paper, we study the optimal rate of mixed membership estimation and propose a rate-optimal spectral method.

1.1 The DCMM model and the optimal rate of mixed membership estimation

We adopt the degree-corrected mixed membership (DCMM) model Zhang et al. (2020); Jin et al. (2017, 2021). It assumes that

$$\{A(i, j) : 1 \leq i < j \leq n\} \text{ are independent Bernoulli variables.} \quad (1.1)$$

For a symmetric non-negative matrix $P \in \mathbb{R}^{K,K}$ that models the community structure and a positive vector $\theta = (\theta_1, \theta_2, \ldots, \theta_n)'$ that contains the degree parameters,

$$\mathbb{P}(A(i, j) = 1) = \theta_i \theta_j \cdot \pi_i' P \pi_j, \quad 1 \leq i < j \leq n. \quad (1.2)$$

To ensure model identifiability, we follow Jin et al. (2017) to assume

$$P \text{ is non-singular and have unit diagonals.} \quad (1.3)$$
Write $\Pi = [\pi_1, \pi_2, \ldots, \pi_n]'$ and $\Theta = \text{diag}(\theta_1, \theta_2, \ldots, \theta_n)$. We can write the DCMM model in the matrix form:

$$A = \Omega - \text{diag}(\Omega) + W,$$

where $\Omega = \Theta \Pi \Pi' \Theta$ and $W = A - \mathbb{E}A$.

A nice feature of DCMM is its flexibility to accommodate degree heterogeneity. The level of degree heterogeneity is characterized by the cumulative distribution function (CDF):

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1\{\bar{\theta}_i \leq t\},$$

where $\bar{\theta} = \frac{1}{n} \sum_{i=1}^{n} \theta_i$. (1.4)

The well-known mixed-membership stochastic block model (MMSBM) [Airoldi et al. (2008)] is a special case where $F_n(\cdot)$ is a point mass at 1. In the case of moderate degree heterogeneity, all $\theta_i$’s are at the same order, so $F_n(\cdot)$ has a compact support bounded below from zero.

One of our main discoveries is that the optimal error rate of mixed membership estimation depends on $F_n(\cdot)$ in a subtle way. Given any estimator $\hat{\Pi} = [\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_n]'$, we measure its performance by the average $\ell^1$-error:

$$\mathcal{L}(\hat{\Pi}, \Pi) = \min_T \left\{ \frac{1}{n} \sum_{i=1}^{n} \|T \hat{\pi}_i - \pi_i\|_1 \right\},$$

where the minimum is over all permutations of columns of $\hat{\Pi}$. For any vector $\theta$, let $Q_n(\theta)$ be a collection of $(\Pi, P)$ that satisfy some regularity conditions (see Section 3 for details).

We show that, up to a logarithmic factor of $n$,

$$\inf_{\hat{\Pi}} \mathbb{E}\mathcal{L}(\hat{\Pi}, \Pi) \asymp \int \min\left\{ \frac{\text{err}_n}{\bar{\theta}^2 \wedge 1}, 1 \right\} dF_n(t), \quad \text{with} \quad \text{err}_n = \frac{K \sqrt{K}}{\delta_n \sqrt{n \bar{\theta}^2}}. \quad (1.6)$$

Here, $\text{err}_n$ is the baseline rate, in which expression $n \bar{\theta}^2$ is the order of average node degree and $\delta_n$ is the order of the minimum eigenvalue of $P$. We are interested in the asymptotic regime of $\text{err}_n \to 0$. The optimal rate is in terms of an integral with respect to $F_n(\cdot)$. Since $F_n(\cdot)$ is a discrete distribution, this integral is always well defined. Under moderate degree heterogeneity, the support of $F_n(\cdot)$ is bounded above and below from 0, so the optimal rate is the same as the baseline rate. However, under severe degree heterogeneity, the optimal rate can be slower than the baseline rate. For example, when $\theta_i$’s are independently drawn from a Gamma distribution with a shape parameter $\alpha$, the optimal rate is $\text{err}_n^{\text{min}(1,2\alpha)}$, which is slower than the baseline rate when $\alpha < 1/2$. More examples are given in Section 3.

### 1.2 A spectral method that is optimally adaptive

The optimal rate in (1.6) depends on the degree heterogeneity through a CDF $F_n(\cdot)$. We say that a method $\hat{\Pi}$ is optimally adaptive if it attains the optimal rate for arbitrary $F_n(\cdot)$. Unfortunately, none of the existing methods is optimally adaptive. The spectral method
in Jin et al. (2017) only attains the optimal rate under moderate degree heterogeneity (i.e., when the support of $F_n(\cdot)$ is bounded above and below from zero); under severe degree heterogeneity, its rate of convergence does not match with the expression in (1.6). We note that ‘optimal adaptivity’ is a strong requirement. It essentially needs that the error rate at each $\tilde{\pi}_i$ has an ‘optimal’ dependence on $\theta_i$, simultaneously for all nodes $i$.

In this paper, we propose a new spectral method, Mixed-SCORE-Laplacian. Our method first applies a ‘pre-PCA normalization’ on the adjacency matrix to obtain

$$\tilde{A} = MAM,$$

where $M$ is a diagonal matrix with positive diagonals. (1.7)

It aims to re-balance the signal-to-noise ratios (SNRs) of the $n$ entries in each leading eigenvector of $\tilde{A}$. Without the pre-PCA normalization, a high-degree node will bring in large noise to every entry of an eigenvector and decreases the SNRs at those entries associated with low-degree nodes. The role of $M$ is to properly down-weight (up-weight) the contributions of high-degree (low-degree) nodes in PCA. We choose $M$ in a way such that for every leading eigenvector of $\tilde{A}$, the SNR at each entry has an ‘optimal’ dependence on $\theta_i$, simultaneously for all $i$. Our careful eigenvector analysis suggests that a satisfactory choice is

$$M_{ii} = (d_i + \bar{d})^{-1/2},$$

where $d_i$ is the degree of node $i$ and $\bar{d}$ is the average node degree. The resulting $\tilde{A}$ happens to be the regularized graph Laplacian. Next, we apply the SCORE normalization Jin (2015) to the leading eigenvectors of $\tilde{A}$ and discover that there exists a low-dimensional simplex geometry associated with the normalized eigenvectors. We then estimate $\pi_i$ by taking advantage of this simplex geometry. It gives rise to a polynomial-time algorithm for estimating $\Pi$.

The pre-PCA normalization is one of the main contributions of our method. While it coincides with the classical Laplacian normalization, it does not mean that we simply took an ad-hoc combination of graph Laplacian with the spectral approach to mixed membership estimation. We in fact started from a general pre-PCA normalization as in (1.7) and pointed out that it serves to adjust the SNRs in leading eigenvectors. We then used careful large-deviation analysis of eigenvectors to identify the correct choice of $M$ as in (1.8). Last, we prove that this $M$ indeed yields the optimal rate of mixed membership estimation under arbitrary degree heterogeneity. Without our insights and analysis, it is unknown that (a) what the optimal rate is, (b) whether there is an $M$ that attains the optimal rate, and (c) whether graph Laplacian is the correct $M$.

In theory, we show that Mixed-SCORE-Laplacian is optimally adaptive, i.e., it attains the rate in (1.6), up to a logarithmic factor of $n$, for quite arbitrary $F_n(\cdot)$. To obtain the targeted error rate, especially under severe degree heterogeneity, we need sharp entry-wise large-deviation bounds for leading eigenvectors of the regularized graph Laplacian. As a main technical contribution, we derive such large-deviation bounds by extending the leave-one-out approach Abbe et al. (2020) of eigenvector analysis to random matrices with weakly dependent entries.
1.3 Connections to the literature

Mixed-SCORE-Laplacian can be viewed as a variant of the Mixed-SCORE algorithm in Jin et al. (2017). However, the main contributions of two papers are orthogonal. Jin et al. (2017) applied the SCORE normalization Jin (2015) to eigenvectors of the adjacency matrix, and discovered a simplex geometry in the spectral domain that enables estimation of $\Pi$ from the eigenvectors. Their primary focus is to reveal an explicit connection between eigenvectors and the target quantity $\Pi$. Our primary focus is to improve the entry-wise signal-to-noise ratios in the eigenvectors and to attain the optimal error rate in (1.6). Compared with the orthodox Mixed-SCORE, our method has several non-trivial modifications, including the pre-PCA normalization in (1.7)-(1.8) and proper trimming of low-degree nodes (see Section 2 for details). These modifications are inspired by eigenvector analysis and can significantly improve the error rate of Mixed-SCORE under severe degree heterogeneity.

Our proposed pre-PCA normalization coincides with the use of regularized graph Laplacian for community detection Rohe et al. (2011); Qin and Rohe (2013); Jin et al. (2022). However, we study mixed membership estimation, which is a more sophisticated problem. Furthermore, our pre-PCA normalization is motivated by re-balancing the entry-wise SNRs in leading empirical eigenvectors, in hopes of matching with the optimal rate in (1.6) for arbitrary $F_n(\cdot)$. It happens that the correct normalization is the regularized graph Laplacian. In Rohe et al. (2011); Qin and Rohe (2013); Jin et al. (2022), the main purpose of using the regularized graph Laplacian is to improve the bound for the spectral norm of a Wigner-type noise matrix. Therefore, the problems, motivations and theoretical analysis are all different.

Our study is also connected to the recent interests of entry-wise eigenvector analysis of random graphs (Abbe et al., 2020; Erdős et al., 2013; Fan et al., 2022; Jin et al., 2017; Mao et al., 2021; Tang and Priebe, 2018). Most of these works studied eigenvectors of the adjacency matrix. A major technical difference is that the upper triangular entries of the adjacency matrix are independent, but this does not hold for the regularized graph Laplacian. It prevents us from applying the leave-one-out argument in Abbe et al. (2020). We need a more sophisticated leave-one-out argument to deal with the dependence among entries of the regularized graph Laplacian (see Section 4). Tang and Priebe (2018) studied eigenvectors of the regularized graph Laplacian for a network model with no degree heterogeneity and obtained bounds for the maximum $\ell^2$-norm error over all $n$ rows of $\hat{\Xi}$ ($\hat{\Xi}$ is the matrix consisting of the first $K$ eigenvectors). This is however insufficient for our purpose, as we work on a model with (severe) degree heterogeneity and need different bounds for different rows of $\hat{\Xi}$.

Community detection (Chen et al., 2018; Jin, 2015; Jin et al., 2022; Lei and Rinaldo, 2015; Ma et al., 2020; Zhang and Zhou, 2016) is a related problem. It assumes that $\pi_i$’s are degenerate and aims to cluster nodes into $K$ non-overlapping communities. For community detection, the loss function is the clustering error, and its optimal rate of convergence has an exponential dependence on $\theta_i$’s Zhang and Zhou (2016); Gao et al. (2018). However, for mixed membership estimation, the loss function is the $\ell^1$-loss, and the optimal rate in
is a polynomial of \( \theta_i \)'s.

The remaining of this paper is organized as follows. In Section 2, we describe the Mixed-SCORE-Laplacian algorithm and explain the rationale behind it. In Section 3, we present the main theoretical results, including the entry-wise eigenvector analysis, rate of convergence of Mixed-SCORE-Laplacian, a matching lower bound, and the extension to other loss functions. Section 4 describes the proof ideas of the entry-wise large-deviation bounds for eigenvectors. Section 5 provides the least-favorable configurations and proofs of lower bounds. Section 6 contains simulation results. Section 7 concludes the paper with discussions. Proofs of secondary lemmas are relegated to the supplementary material.

Notations. Throughout this paper, we use the notation \( C, C_i \) for \( i \in \mathbb{Z}^+ \) and \( c \) to represent generic constants independent of dimension \( n \), which may vary from line to line. For any two sequences \( a_n \) and \( b_n \), \( a_n \asymp b_n \) means there is a constant \( C > 1 \) such that \( C^{-1}b_n \leq a_n \leq Cb_n \); \( a_n \gtrsim b_n \) means there exist a constant \( c > 0 \) such that \( a_n \geq cb_n \). For arbitrary matrix \( A \), we denote by \( A(i) \) the \( i \)-th row of \( A \), \( A(i,j) \) or \( A_{ij} \) the \((i,j)\)-th entry of \( A \). We write \( a \vee b \) for \( \max\{a,b\} \) and \( a \wedge b \) for \( \min\{a,b\} \) for any \( a, b \). We adopt the convention \( \{e_i\}_{i=1}^n \) for the standard basis of \( \mathbb{R}^n \). We use \( \|\cdot\| \) to denote the Euclidean norm for a vector or operator norm for a matrix and use \( \|\cdot\|_\infty \) for the \( \ell_\infty \) norm for either a vector or a matrix.

2 A new spectral algorithm

In Section 2.1, we explain the idea of pre-PCA normalization. In Section 2.2, we describe the Mixed-SCORE-Laplacian algorithm.

2.1 Improving the accuracy of PCA under degree heterogeneity

In our model,

\[
A = \Omega - \text{diag}(\Omega) + W = \text{‘main signal’ + ‘secondary signal’ + ‘noise’},
\]

(2.1)

where \( \Omega = \Theta \Pi \Pi' \Theta \) and \( W = A - \mathbb{E}A \). Our goal is to find an optimal spectral approach to estimating \( \Pi = [\pi_1, \pi_2, \ldots, \pi_n]' \). We assume that there is a constant \( 0 < c_0 < 1 \) such that \( \Omega_{ij} \leq c_0 \) for all \( 1 \leq i, j \leq n \). This is a mild condition that holds for most networks. Note that

\[
\mathbb{E}[A_{ij}] = \Omega_{ij}, \quad \text{Var}(W_{ij}) = \Omega_{ij}(1 - \Omega_{ij}) \asymp \Omega_{ij}, \quad \text{where } \Omega_{ij} = \theta_i \theta_j \pi_i' P \pi_j.
\]

Due to severe degree heterogeneity, \( \theta_i \) may have different magnitudes. Therefore, the means (and also the variances) of different entries of \( A \) may have different magnitudes. In such a case, a direct use of ordinary Principal Component Analysis (PCA) may produce undesirable results (Jin et al., 2022), so we wish to combine PCA with proper normalizations.

A possible approach is to combine PCA with a pre-PCA normalization: for a positive diagonal matrix \( M = \text{diag}(m_1, \ldots, m_n) \) to be determined, we multiply \( M \) on both sides of
A and then apply PCA to $MAM$; note that $MAM = M\Omega M - M\text{diag}(\Omega)M + MW M$. The main problem here is that, we can not find an $M$ that can properly normalize the ‘main signal’ matrix $\Omega$ and the ‘noise’ matrix $W$ simultaneously. For example, if we take $M = \Theta^{-1}$, then $(M\Omega M)_{ij} = \pi_i P \pi_j$ and $\text{Var}((MW M)_{ij}) \approx \pi_i P \pi_j$, $1 \leq i, j \leq n, i \neq j$. In this case, the degree heterogeneity effect is removed in the ‘signal’ but still presents in the ‘noise’. A similar claim can be drawn if we take $M = \Theta^{-1/2}$, where $\text{Var}((MW M)_{ij}) \approx \theta_i \theta_j - 1 \pi_i P \pi_j$, $1 \leq i, j \leq n, i \neq j$.

To fix the problem, we use the strategy of two normalizations – a pre-PCA normalization and a post-PCA normalization. The pre-PCA normalization is as above, and the post-PCA normalization is conducted on the leading eigenvectors. Let $\hat{\lambda}_k$ be the $k$-th largest (in magnitude) eigenvalue of $MAM$ and let $\hat{\xi}_k$ be the corresponding eigenvector, for $1 \leq k \leq K$. Write $\hat{\Xi} = [\hat{\xi}_1, \ldots, \hat{\xi}_K]$. Recall that $M\Omega M$ is the ‘main signal’ in $MAM$. Let $\lambda_k$ and $\xi_k$ be the $k$th largest eigenvalue and the corresponding eigenvector of $M\Omega M$. Write $\Xi = [\xi_1, \xi_2, \ldots, \xi_K]$. Then,

$$\hat{\Xi} = \Xi + \text{“noise”}. \quad (2.2)$$

Since $M\Omega M$ is a low-rank matrix, those non-leading eigenvectors $MAM$ can only be driven by noise. PCA removes all non-leading eigenvectors and leads to a dramatic noise reduction. As a result, the SNR in (2.2) is much higher than (2.1). To some extent, we can view that any normalization on $\hat{\Xi}$ is mainly on the ‘signal’ part $\Xi$. Now, in the two-normalization strategy, the ‘signal’ part is affected by both pre-PCA and post-PCA normalizations, but the ‘noise’ part is (almost) only affected by the pre-PCA normalization. This makes it possible to have different normalization effects on the ‘signal’ and ‘noise’.

The SCORE normalization ([Jin, 2015]) is a post-PCA normalization approach that aims to reduce the degree heterogeneity effect in eigenvectors. Given $\hat{\xi}_1, \ldots, \hat{\xi}_K$, it constructs an $n \times (K - 1)$ matrix $\hat{R}$ by $\hat{R}(i,k) = \hat{\xi}_{k+1}(i)/\hat{\xi}_1(i)$, $1 \leq k \leq K - 1, 1 \leq i \leq n$. [Jin et al., 2017] applied this normalization on eigenvectors of the adjacency matrix $A$ and discovered an interesting simplex structure associated with the rows of $\hat{R}$; they further used this simplex structure to develop an algorithm for estimating $\Pi$, which we call the orthodox Mixed-SCORE (OMS). OMS is a one-normalization approach that has no pre-PCA normalization. We now aim to combine it with a proper pre-PCA normalization. This pre-PCA normalization must satisfy:

- It is compatible with the post-PCA normalization by SCORE (because these two normalizations will both take effect on the ‘signal’).
- It simultaneously optimizes the SNR at each row of $\hat{R}$.

The first requirement is always satisfied, because we can write $M\Omega M = M(\Theta \Pi \Pi' \Theta)M = \tilde{\Theta} \Pi \Pi' \tilde{\Theta}$, where $\tilde{\Theta} = M\Theta$. The matrix $M\Omega M$ has a similar structure as $\Omega$, where $\tilde{\theta}_i = m_i\theta_i$ is the ‘auxiliary degree parameters’ of node $i$. It can be shown that the main ideas of OMS (e.g., the post-PCA normalization and the post-PCA simplex geometry) can be extended to this case. What remains is to find a proper $M$ such that the second requirement is satisfied.
We consider $M = \Theta^{-b}$, where $b > 0$ is a constant. Without loss of generality, we
assume $K$ is finite and $\frac{\lambda_2}{\lambda_1}, \ldots, \frac{\lambda_K}{\lambda_1}$ are distinct constants (these conditions help simplify the
illustration of main ideas, but they are not needed in our theory). We measure the SNR at
the $i$th row of $\hat{R}$ by (below, SD denotes the standard deviation)

$$\text{SNR}_i(\hat{R}) = [\xi_1(i)]^{-1} \max_{1 \leq k \leq K-1} \text{SD}(\hat{\xi}_{k+1}(i)).$$

By linear algebra analysis, we can show that, $\lambda_k \approx \sum_{i=1}^n \theta_i^{2-2b}$ and $|\xi_{k+1}(i)| = O(\lambda_1^{-1/2} \theta_i^{1-b})$. Moreover, we approximate $\hat{\xi}_{k+1}$ by its first-order approximation: The definition of eigenvectors implies $\hat{\xi}_{k+1} = \hat{\lambda}_{k+1} M A M \xi_{k+1}$; under mild conditions, $\hat{\lambda}_{k+1} \approx \lambda_{k+1}$ and $\hat{\xi}_{k+1} \approx \xi_{k+1}$; therefore, we have $\hat{\xi}_{k+1} \approx \lambda_{k+1}^{-1} M A M \xi_{k+1} \equiv \hat{\xi}_{k+1}^*$. Each entry of $\hat{\xi}_{k+1}^*$ is a weighted sum of
independent Bernoulli variables, whose variance can be calculated explicitly. Our calculations suggest that

$$\text{SNR}_i(\hat{R}) \approx \frac{\left(\sum_{j=1}^n \theta_j^{3-4b}\right)^{1/2}}{\sqrt{\theta_i} \sum_{j=1}^n \theta_j^{2-2b}} = \frac{a_n}{\sqrt{n \theta_i}}, \quad \text{where } a_n = \frac{\int t^{3-4b} dF_n(t)}{\int t^{2-2b} dF_n(t)}. \quad (2.3)$$

Here, $1/\sqrt{n \theta_i}$ is the intrinsic order of SNR, and $a_n$ captures the degree heterogeneity
effect, which does not depend on $i$. We note that $F_n(\cdot)$ is self-normalized, where by (1.4), \int tF_n(t) = 1. Therefore, if we choose $b = 1/2$, then $a_n$ is always equal to 1 and will not be
affected by degree heterogeneity! In contrast, if we choose $b = 0$, then $a_n$ will be heavily
influenced by those large $\theta_i$, as $\int t^2 dF_n(t) \gg \int t^2 dF_n(t)$ is possible; if we choose $b = 1$, then $a_n$ will be heavily influenced by those small $\theta_i$, as $\int t^{-1} dF_n(t)$ may grow with $n$.

We have seen that $M = \Theta^{-1/2}$ is the best choice. Moreover, if we take $M = J \Theta^{-1/2}$
for a positive diagonal matrix $J$ such that $(\max_i J_{ii})/(\min_i J_{ii}) \leq C$, the same conclusion holds. In practice, we do not know $\theta_i$ but we observe $d_i$, the degree of node $i$. Under mild
regularity conditions, $\mathbb{E}[d_i] \approx \theta_i \cdot n \bar{\theta}$. It inspires the choice of

$$M^* = \text{diag}(d_1, d_2, \ldots, d_n)^{-1/2}. \quad (2.4)$$

The calculation in (2.3) assumes that $M$ is a non-stochastic matrix. In this stochastic $M^*$,
for a low-degree node $i$, the noise dominates in $M^*$. We thereby add a regularization and use

$$M = [\text{diag}(d_1, d_2, \ldots, d_n) + \lambda_n I_n]^{-1/2}. \quad (2.5)$$

In theory, it suffices if $\lambda_n$ is at the order of $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i$. The resulting $M A M$ happens to be the regularized graph Laplacian. However, we did not start from an ad-hoc combination of graph Laplacian and OMS. We used careful analysis of row-wise SNRs for $\hat{R}$ to come to discover that graph Laplacian is the correct normalization. Such insights are new.

### 2.2 The Mixed-SCORE-Laplacian algorithm

For a constant $\tau > 0$, we consider the regularized graph Laplacian matrix:

$$L = H^{-1/2} AH^{-1/2}, \quad \text{where } H = \text{diag}(d_1, d_2, \ldots, d_n) + \tau \bar{d} \cdot I_n. \quad (2.6)$$
Our analysis in Section 2.1 has suggested that this is the correct pre-PCA normalization. Let \( \hat{\lambda}_1, \ldots, \hat{\lambda}_K \) be the \( K \) largest eigenvalues (in magnitude) of \( L \), and let \( \hat{\xi}_1, \ldots, \hat{\xi}_K \in \mathbb{R}^n \) be the associated eigenvectors. The SCORE normalization \cite{jin2015} is a post-PCA normalization that constructs an \( n \times (K - 1) \) matrix \( \hat{R} \) by

\[
\hat{R}(i, k) = \frac{\hat{\xi}_{k+1}(i)}{\hat{\xi}_1(i)}, \quad 1 \leq i \leq n, 1 \leq k \leq K - 1.
\] (2.7)

By Perron’s theorem, as long as the network is connected, \( \hat{\xi}_1 \) is a strictly positive vector \cite{jin2015}. Therefore, \( \hat{R} \) is always well-defined. Let \( \hat{r}_1', \hat{r}_2', \ldots, \hat{r}_n' \) denote the rows of \( \hat{R} \). The next lemma introduces a population counterpart of \( \hat{R} \) and shows that there is a simplex structure associated with the rows of \( \hat{R} \).

**Lemma 2.1** (The simplex geometry). Consider a DCMM model, where each community \( k \) has at least one pure node (i.e., \( \pi_i = \delta_k \)). Let \( H_0 = \bar{E}H \) and \( L_0 = H_0^{-1/2} \Omega H_0^{-1/2} \). Let \( \lambda_k \) be the \( k \)th largest eigenvalue (in magnitude) of \( L_0 \), and let \( \hat{\xi}_1 \) be the corresponding eigenvector. Then, \( \hat{\xi}_1 \) is a strictly positive vector. Consider the matrix \( \hat{R} \in \mathbb{R}^{n \times (K - 1)} \), where \( \hat{R}(i, k) = \hat{\xi}_{k+1}(i)/\hat{\xi}_1(i), \) \( 1 \leq i \leq n, 1 \leq k \leq K - 1 \). Write \( \hat{R} = [r_1', r_2', \ldots, r_n']' \).

- There exists a simplex \( S \subset \mathbb{R}^{K-1} \) with \( K \) vertices \( v_1, v_2, \ldots, v_K \), such that \( r_1, r_2, \ldots, r_n \) are contained in \( S \). If node \( i \) is a pure node, \( r_i \) is on one vertex of this simplex; if node \( i \) is a mixed node, \( r_i \) is in the interior of the simplex (it can be on an edge or a face, but not on any of the vertices).

- Each \( r_i \) is a convex combination of the \( K \) vertices, \( r_i = \sum_{k=1}^{K} w_i(k) v_k \). The combination coefficient vector is \( w_i = \| \pi_i \circ b_1 \|^{-1} (\pi_i \circ b_1) \), where \( \circ \) is the Hardarmart product, \( b_1(k) = 1/\sqrt{\lambda_1 + v_k' \text{diag}(\lambda_2, \ldots, \lambda_K) v_k} \).

Motivated by Lemma 2.1, we estimate the \( K \) vertices of \( S \) from \( \hat{r}_1, \ldots, \hat{r}_n \). Let

\[
\hat{S}_n(c) := \{1 \leq i \leq n : d_i \hat{\delta}^2_n \geq cK^3 \log(n)\}, \quad \text{where } \hat{\delta}_n = \min\{\sqrt{K(\hat{\lambda}_1 - \hat{\lambda}_2)}, K|\hat{\lambda}_K|\}, \quad (2.8)
\]

and

\[
\hat{S}_n^*(c, \gamma) = \hat{S}_n(c) \cap \{1 \leq i \leq n : d_i \geq \gamma \hat{d}_i\}. \quad (2.9)
\]

Here, \( c > 0 \) and \( \gamma \in (0, 1) \) are tuning parameters. We apply the successive projection algorithm \cite{araujo2001} on \( \{\hat{r}_i : i \in \hat{S}_n^*(c, \gamma)\} \) to estimate the \( K \) vertices of the simplex (details are in Step 3 below). Let \( \hat{v}_1, \ldots, \hat{v}_K \) be the estimated vertices. By the second bullet point of Lemma 2.1, \( \hat{r}_i \approx \sum_{k=1}^{K} w_i(k) \hat{v}_k \), so we can estimate \( w_i \) from \( \hat{r}_i \) and \( \hat{v}_1, \ldots, \hat{v}_K \) (see Step 4 below). Once \( \hat{w}_i \) is available, we can estimate \( b_i \) from \( \hat{\lambda}_k \)'s and \( \hat{v}_k \)'s, and then use \( (\hat{b}_1, \hat{w}_i) \) to construct \( \hat{\pi}_i \), following the second bullet point of Lemma 2.1 (details are in Step 4 below).

**Mixed-SCORE-Laplacian.** Input: \( K, A, c \in (0, 1), \) and \( \gamma \in (0, 1) \). Output: \( \hat{\Pi} \).

1. Let \( L \) be the graph Laplacian in (2.6) (with \( \tau = 1 \)). Let \( \hat{\lambda}_k \) be the \( k \)th largest eigenvalue (in magnitude) of \( L \), and let \( \hat{\xi}_1 \) be the associated eigenvector, \( 1 \leq k \leq K \). Obtain the matrix \( \hat{R} \) as in (2.7). Denote its rows by \( \hat{r}_1, \hat{r}_2, \ldots, \hat{r}_n \).
2. Let \( \hat{S}_n(c) \) be as in (2.8). For any \( i \notin \hat{S}_n(c) \), set \( \hat{\pi}_i = K^{-1}1_K \).

3. Let \( \hat{S}^*_n(c, \gamma) \) be as in (2.9). Run the successive projection algorithm on \( \{\hat{r}_i : i \in \hat{S}^*_n(c, \gamma)\} \):
   - Find \( i_1 = \arg\max_{i \in \hat{S}^*_n(c, \gamma)} \|\hat{r}_i\| \). Let \( \hat{v}_1 = \hat{r}_{i_1} \).
   - For each \( 2 \leq k \leq K \), obtain \( \hat{v}_k \) from \( \hat{v}_1, \ldots, \hat{v}_{K-1} \) as follows: Let \( \mathcal{P} = Y(Y'Y)^{-1}Y \), where \( Y \) is an \( n \times K \) matrix whose \( k \)th row is \( (1, \hat{r}_i)' \). Find \( i_k = \arg\max_{i \in \hat{S}^*_n(c, \gamma)} \|(I_{K-1} - \mathcal{P})x_i\| \). Let \( \hat{v}_k = \hat{r}_{i_k} \).

4. For each \( i \in \hat{S}_n(c) \), solve \( \hat{w}_i \in \mathbb{R}^K \) from the linear equation set: \( \sum_{k=1}^{K} \hat{w}_i(k)\hat{v}_k = \hat{r}_i \) and \( \sum_{k=1}^{K} \hat{w}_i(k) = 1 \). Obtain \( \hat{b}_1 \in \mathbb{R}^K \) from \( \hat{b}_1(k) = [\hat{\lambda}_1 + \hat{v}_k^t\hat{\lambda}_1\hat{v}_k]^{-1/2} \), for \( 1 \leq k \leq K \), where \( \hat{\Lambda} = \text{diag}(\hat{\lambda}_2, \ldots, \hat{\lambda}_K) \). Let \( \hat{\pi}^*_i \in \mathbb{R}^K \) be such that \( \hat{\pi}^*_i(k) = \max\{\hat{w}_i(k)/\hat{b}_1(k), 0\} \), \( 1 \leq k \leq K \). Output \( \hat{\pi}_i = \hat{\pi}^*_i/\|\hat{\pi}^*_i\|_1 \), for each \( i \in \hat{S}_n(c) \).

By default, we set the two tuning parameters as \( c = 0.1 \) and \( \gamma = 0.05 \).

Compared with the OMS algorithm in Jin et al. (2017), the above algorithm not only is equipped with a pre-PCA normalization but also carefully trims off low-degree nodes. In Step 2, it removes those \( \hat{r}_i \) with \( i \notin \hat{S}_n(c) \). For these nodes, the ‘noise’ in \( \hat{r}_i \) is too high, and it is impossible to get a better estimate of \( \pi_i \) than random guessing. In Step 3, we further remove those \( \hat{r}_i \)'s with \( i \in \hat{S}_n(c) \setminus \hat{S}^*_n(c, \gamma) \). We still estimate \( \pi_i \) from these \( \hat{r}_i \), but we do not use them for estimating vertices. This can be viewed as denoising before vertex hunting. Ke and Jin (2021) pointed out that a direct use of the successive projection algorithm can be unsatisfactory, especially when the noise level is high or there are outliers, and they recommended to add a denoising sub-step. In our problem, the noise level at \( \hat{r}_i \) is monotone increasing with \( d_i \), so we denoise by filtering out low-degree nodes.

3 Main results

In Section 3.1, we conduct entry-wise eigenvector analysis for the regularized graph Laplacian. In Section 3.2, we study the error rate of Mixed-Score-Laplacian. In Section 3.3, we provide a matching lower bound. In Section 3.4, we extend the upper/lower bound results to a weighted \( \ell^1 \)-loss.

Consider the DCMM model (1.1)-(1.3). Let \( D_\theta \in \mathbb{R}^{n \times n} \) be a positive diagonal matrix with \( \theta(i, i) = (e_i - \frac{1}{n}1_n)'[\Omega - \text{diag}(\Omega)]1_n, 1 \leq i \leq n \). Define
\[
G := K \cdot \Pi' \Theta D_\theta^{-1} \Theta \Pi \in \mathbb{R}^{K \times K}.
\]

For a constant \( c_1 > 0 \), we assume
\[
\|G\| \leq c_1, \quad \|G^{-1}\| \leq c_1, \quad \min_{1 \leq k \leq K} \left\{ \sum_{i=1}^{n} \theta_i \pi_i(k) \right\} \geq c_1 \|\theta\|_1. \tag{3.1}
\]
Let \( \lambda_k(PG) \) denote its \( k \)-th largest eigenvalue (in magnitude), \( 1 \leq k \leq K \). Since \( PG \) is a nonnegative matrix, by Perron’s theorem, \( \lambda_1(PG) > 0 \). For a constant \( c_2 \in (0, 1) \) and some \( \alpha_n \in [1, K] \) and \( \beta_n \in (0, 1) \), we assume

\[
\lambda_1(PG) \geq \alpha_n, \quad |\lambda_K(PG)| \geq \beta_n, \quad \max_{k \neq 1} \lambda_k(PG) \leq \min \{(1 - c_2)\lambda_1(PG), c_2^{-1}\sqrt{K}\}. \quad (3.2)
\]

Let \( \eta_1 > 0 \) be the leading right eigenvector of \( PG \). For a constant \( c_3 > 0 \), we assume

\[
\min_{1 \leq k \leq K} \eta_1(k) > 0, \quad \text{and} \quad \frac{\min_{1 \leq k \leq K} \eta_1(k)}{\max_{1 \leq k \leq K} \eta_1(k)} \geq c_3. \quad (3.3)
\]

Last, for a constant \( c_4 \in (0, 1) \), we assume that

\[
\{1 \leq i \leq n : \pi_i(k) = 1, \ \theta_i \geq c_4 \bar{\theta} \} \neq \emptyset, \quad \text{for each } 1 \leq k \leq K. \quad (3.4)
\]

These regularity conditions are mild. Condition (3.1) is about ‘balance’ of communities. The third inequality in (3.1) controls the degree balance. To understand the first two inequalities, consider a case where all \( \pi_i \)'s are degenerate; then, \( G \) is a diagonal matrix, whose \( k \)-th diagonal entry is approximately \( K \) times the fraction of nodes in community \( k \), so the first two inequalities in (3.1) translate to that the sizes of \( K \) communities are at the same order. In Condition (3.2), the first two inequalities are not assumptions but just specifying \( \alpha_n \) and \( \beta_n \) as the respective orders of \( \lambda_1(PG) \) and \( |\lambda_K(PG)| \). The third inequality is a mild eigengap condition. Take for example the case of \( G = I_K \) (i.e., no mixed membership, equal community size) and \( P = (1 - b)I_K + b1_K1_K' \). It can be shown that \( \lambda_1(PG) \approx K \) and \( \lambda_k(PG) = 1 - b \) for all \( 2 \leq k \leq K \), so the third inequality in (3.2) holds. Condition (3.3) is also mild. Note that \( PG \) is a nonnegative matrix. By Perron’s theorem (Horn and Johnson, 1985), this condition holds if either \( P \) or \( G \) is irreducible. In the aforementioned example, \( \eta_1 = \frac{1}{\sqrt{K}}1_K \) and both requirements in (3.3) are satisfied. Condition (3.4) requires that each community has at least one pure node whose degree parameter is properly large, which is a mild assumption.

### 3.1 Entrywise eigenvector analysis of the regularized graph Laplacian

Recall that the regularized graph Laplacian matrix \( L \) is as in (2.6) and that \( \hat{\xi}_1, \ldots, \hat{\xi}_K \) are the first \( K \) eigenvectors of \( L \). Write \( \hat{\Xi}_1 = [\hat{\xi}_2, \ldots, \hat{\xi}_K] \) and \( \hat{\Xi} = [\hat{\xi}_1, \hat{\Xi}_1] \). Define

\[
L_0 = H_0^{-1/2}\Omega H_0^{-1/2}, \quad \text{where} \quad H_0 = \mathbb{E}H. \quad (3.5)
\]

Then, \( L_0 \) is the population counterpart of \( L \). Let \( \lambda_1, \ldots, \lambda_K \) be the \( K \) largest eigenvalues (in magnitude) of \( L_0 \), and let \( \xi_1, \ldots, \xi_K \in \mathbb{R}^n \) be the corresponding eigenvectors. Similarly, we write \( \Xi_1 = [\xi_2, \ldots, \xi_K] \) and \( \Xi = [\xi_1, \Xi_1] \).
Theorem 3.1. Consider the DCMM model in (1.1)-(1.3), where (3.1)-(3.3) are satisfied. Suppose $K^3 \log(n)/(n\theta^2 \beta_n^2) \to 0$ as $n \to \infty$. With probability $1 - o(n^{-3})$, there exists $\omega \in \{1, -1\}$ and an orthogonal matrix $O_1 \in \mathbb{R}^{K-1,K-1}$ such that

$$|\omega \xi_1(i) - \xi_1(i)| \leq C \left( K^3 \theta_i \log(n) \right)^{1/2} \left( 1 + \frac{\log(n)}{n\theta_i} \right),$$

$$\|\hat{e}_i'(\hat{\Xi}_1O_1 - \Xi_1)\| \leq C \left( K^3 \theta_i \log(n) \right)^{1/3} \left( 1 + \frac{\log(n)}{n\theta_i} \right),$$

simultaneously for all $1 \leq i \leq n$, where $C > 0$ is a constant that only depends on $(c_1, c_2, c_3)$ in (3.1)-(3.3) and $\tau$ in (2.6).

As a corollary of Theorem 3.1 we can obtain a large-deviation bound for each row of the matrix $\hat{R}$ in (2.7). Define $R \in \mathbb{R}^{n \times (K-1)}$ by

$$R(i,k) = \xi_{k+1}(i)/\xi_1(i), \quad 1 \leq i \leq n, 1 \leq k \leq K - 1.$$

This is a population counterpart of $\hat{R}$. Denote by $r'_1, r'_2, \ldots, r'_n$ the rows of $R$. Let

$$\delta_n = \min\{K^{-1/2} \alpha_n, \beta_n\},$$

where $\alpha_n$ and $\beta_n$ are as in (3.2). In many cases (e.g., $K$ is bounded, or $K \to \infty$ and $\lambda_1(P) \geq \beta_n \sqrt{K}$), we can show that $\delta_n \asymp \beta_n \asymp |\lambda_{\min}(P)|$, where $\lambda_{\min}(P)$ is the minimum eigenvalue (in magnitude) of $P$. Hence, $\delta_n$ captures the order of the minimum eigenvalue of $P$.

Corollary 3.1. Consider the DCMM model in (1.1)-(1.3), where (3.1)-(3.3) are satisfied. Suppose $K^3 \log(n)/(n\theta^2 \delta_n^2) \to 0$ as $n \to \infty$. For any constant $c_0 > 0$, define

$$S_n(c_0) := \{1 \leq i \leq n : n\theta_i \delta_n^2 \geq c_0 K^3 \log(n)\}.$$

With probability $1 - o(n^{-3})$, there exists an orthogonal matrix $O_1 \in \mathbb{R}^{K-1,K-1}$ such that, simultaneously for $i \in S_n(c_0)$,

$$\|O_1 \hat{r}_i - r_i\| \leq C \sqrt{\frac{K^3 \log n}{n\theta_i \delta_n^2}}.$$

The statement holds for any $c_0 > 0$, except that the constant $C$ will depend on $c_0$.

Corollary 3.1 only considers nodes in $S_n(c_0)$. For a node $i \notin S_n(c_0)$, its degree is so small that $\hat{r}_i$ is too noisy to contain useful information of $\pi_i$. In Mixed-SCORE-Laplacian, the set $S_n(c_0)$ is estimated by $\hat{S}_n(c)$, and those $\hat{r}_i$'s with $i \notin \hat{S}_n(c)$ are discarded.
3.2 The error rate of Mixed-SCORE-Laplacian

We first study the node-wise errors.

**Theorem 3.2.** Consider the DCMM model in (1.1)-(1.3), where (3.1)-(3.3) are satisfied, and additionally, (3.4) holds. Suppose 
\[ K^3 \log(n)/(n\theta^2\delta_n^2) \to 0 \text{ as } n \to \infty. \]
Let \( \hat{\Pi} \) be the estimator from Mixed-SCORE-Laplacian, where the tuning parameters are such that 
\[ c > 0 \quad \text{and} \quad 0 < \gamma < c^4 \]
(here \( c^4 \) is the same as in (3.4)). Then, with probability \( 1 - o(n^{-3}) \), there exists a permutation \( T \) on \( \{1, 2, \ldots, K\} \), such that
\[ \|T\hat{\pi}_i - \pi_i\|_1 \leq C \min \left\{ \sqrt{K^3 \log n/n\theta^2}, 1 \right\}, \]
(3.12)
simultaneously for all \( 1 \leq i \leq n \).

We compare Theorem 3.2 with the node-wise error bounds in Jin et al. (2017) for the orthodox Mixed-SCORE (OMS) algorithm: With high probability, there exists a permutation \( T \) such that
\[ \max_{1 \leq i \leq n} \|T\hat{\pi}_{\text{OMS}}_i - \pi_i\|_1 \leq \sqrt{\theta^3 \max \{\theta_{i+1} \theta_{i+1}\} \cdot C \sqrt{K^3 \log(n)} / n\theta^2 \delta_n^2}. \]
(3.13)
This bound is for the maximum node-wise error, but it does not give the \( \theta_i \)-dependent bounds for individual nodes. The reason is that the eigenvector analysis in Jin et al. (2017) only gave a bound for \( \|\hat{\Xi}O - \Xi\|_{2 \to \infty} \), which translates to a bound for \( \max_i \{\|O_i\hat{\pi}_i - r_i\|\} \).

Furthermore, even if we set \( \theta_i = \theta_{\text{min}} \) in (3.12) and compare two bounds, the bound in (3.13) still has an additional factor that is larger than 1. The main reason is that our new algorithm has a pre-PCA normalization and trimming of low-degree nodes.

We then ensemble the node-wise errors to get upper bounds for the \( \ell^1 \)-loss in (1.5). Write
\[ \text{err}_n = K^{3/2}\delta_n^{-1}(n\theta^2)^{-1/2}. \]
(3.14)
Recall that \( F_n(\cdot) \) as in (1.4). The next corollary is proved in the supplementary material.

**Corollary 3.2.** Suppose conditions of Theorem 3.2 hold. Let \( \hat{\Pi} \) be the estimator from Mixed-SCORE-Laplacian, \( \mathcal{L}(\hat{\Pi}, \Pi) \) be the \( \ell^1 \)-loss in (1.5), and \( \text{err}_n \) be as in (3.14). Then,
\[ \mathbb{E}\mathcal{L}(\hat{\Pi}, \Pi) \leq C \sqrt{\log(n)} \int \min \left\{ \frac{\text{err}_n}{\sqrt{t \wedge 1}}, 1 \right\} dF_n(t), \]
(3.15)

3.3 A matching lower bound

We aim to develop a lower bound argument that is specific to \( \theta \). To get such a strong lower bound, we need a technical condition (such a condition is not needed for the upper bound argument in Section 3.2).
Definition 3.1. Fix constants $\gamma > 0$ and $a_0 \in (0,1)$. Recall that $\text{err}_n$ is as in (3.14). Let $\mathcal{G}_n(\gamma,a_0)$ be the collection of $\theta \in \mathbb{R}^n$ such that there exists $c_n > 0$ satisfying that $\gamma c_n \geq \text{err}^2_n$, $F_n(c_n) \leq 1 - a_0$, and $\int_{c_n}^{\infty} \frac{1}{\sqrt{1 + t}} dF_n(t) \geq a_0 \int_{\text{err}_n^2}^{\infty} \frac{1}{\sqrt{1 + t}} dF_n(t)$.

This condition excludes those $F_n(\cdot)$ that have ill behavior in the neighborhood of 0. It is needed in the construction of the least-favorable configuration (see Section 5). In Section E.5 of the supplementary material, we show that the requirements in Definition 3.1 are satisfied if $\theta_i$’s are i.i.d. drawn from $\kappa_n F(\cdot)$, where $\kappa_n > 0$ is a scalar and $F(\cdot)$ is a fixed, finite-mean distribution satisfying one of the following conditions: (i) $F(\cdot)$ is a discrete distribution; (ii) $F(\cdot)$ is a continuous distribution with support in $[c,\infty)$, for some $c > 0$. (iii) $F(\cdot)$ is a continuous distribution supported in $(0,\infty)$, and its density $f(t)$ satisfies that $\lim_{t \to \infty} t^b f(t) = C$, for some $b \neq 1/2$ and $C > 0$.

Theorem 3.3. Fix constants $c_1 \cdots c_4$, $\gamma$ and $a_0 < 1$. Given $(n,K)$ and $\theta \in \mathcal{G}_n(\gamma,a_0)$, let $\mathcal{Q}_n(\theta)$ be the collection of $(\Pi,P)$ such that (3.1)-(3.4) are satisfied. There exists a constant $C > 0$ such that, for all sufficiently large $n$,

$$\inf_n \sup_{(\Pi,P) \in \mathcal{Q}_n(\theta)} \mathbb{E}[\hat{\mathcal{L}}(\hat{\Pi},\Pi)] \geq C \int \min \{ \frac{\text{err}_n}{\sqrt{t} \wedge 1}, 1 \} dF_n(t).$$

(3.16)

By Corollary 3.2 and Theorem 3.3, Mixed-SCORE-Laplacian attains the optimal rate up to a logarithmic factor of $n$. Furthermore, this is true for every $\theta \in \mathcal{G}_n(\gamma,a_0)$, hence, Mixed-SCORE-Laplacian is optimally adaptive.

Example 1 (Order of $\text{err}_n$). Let $G_n$ be a distribution on the standard simplex of $\mathbb{R}^K$ such that $\mathbb{E}_{G_n}[\pi] = \frac{1}{K} 1_K$ and $\lambda_{\min}(\mathbb{E}_{G_n}[\pi \pi^T]) \geq \frac{c}{K}$, for a constant $c > 0$. For $a_n, b_n \in (0,1)$, suppose

$$P = (1 - b_n) I_K + b_n 1_K 1_K^T, \quad \pi_i \overset{iid}{\sim} G_n, \quad \theta_i = \sqrt{a_n} \cdot \eta_i, \quad \eta_i \overset{iid}{\sim} F_n.$$

It can be shown that $\lambda_1(PG) \asymp K, |\lambda_K(PG)| \asymp 1 - b_n$ and $\delta_n \asymp 1 - b_n$. Therefore,

$$\text{err}_n = \frac{K \sqrt{K}}{\sqrt{n a_n (1 - b_n)^2}}.$$

Here, $a_n$ captures network sparsity, and $1 - b_n$ captures the dissimilarity of communities.

Example 2 (Order of the optimal rate). The baseline rate $\text{err}_n$ does not depend on the degree heterogeneity characterized by $F_n(\cdot)$, but the optimal rate does. We consider several examples of $F_n(\cdot)$. Note that by (1.4), the mean of $F_n(\cdot)$ is always equal to 1.

- Uniform distribution: $F_n = U([1 - \epsilon, 1 + \epsilon])$, for a constant $\epsilon \in (0,1)$.
- Pareto distribution: $F_n = \frac{a_{\min}^{a - 1}}{a} \text{Pareto}(c,a)$, where $c > 0$ is the scale parameter, $a > 1$ is the shape parameter, and the support of the Pareto distribution is $[c_{\min}, \infty]$.
- Gamma distribution: $F_n = \frac{\beta}{\alpha} \text{Gamma}(\alpha, \beta)$, where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the rate parameter.
• Mixture distribution: \( F_n = \sum_{\ell=1}^L \epsilon_\ell \delta_{x_\ell} \), where \( x_1 < x_2 < \ldots < x_L \), \( \delta_x \) is a point mass at \( x \), and \( \sum_{\ell=1}^L \epsilon_\ell x_\ell = 1 \). As \( n \to \infty \), \( L \) is fixed, each \( x_\ell \) can depend on \( n \), but \( x_1 \gg \text{err}^2_n \).

For all above cases, the requirements in Definition 3.1 are satisfied, so that the lower bound in Theorem 3.3 applies. By direct calculations, the optimal rate is

\[
\begin{cases}
\text{err}_n, & \text{for Uniform, Pareto, and Gamma with } \alpha > 1/2, \\
\text{err}_n^{2\alpha}, & \text{for Gamma with } \alpha < 1/2, \\
\text{err}_n \max_{1 \leq \ell \leq L} \left\{ \frac{\epsilon_\ell}{\sqrt{x_\ell}} \right\}, & \text{for Mixture}.
\end{cases}
\]

Here, Pareto is an example where there are a small number of extremely large \( \theta_i \). These \( \theta_i \)'s affect the baseline rate \( \text{err}_n \) by changing \( \bar{\theta} \), and the optimal rate is the same as the baseline rate. Gamma is an example where a considerable fraction of nodes have very small \( \theta_i \). They have a negligible effect on \( \bar{\theta} \), but they can affect the optimal rate. When \( \alpha < 1/2 \), the optimal rate is slower than the baseline rate.

3.4 Extension to other loss functions

We define a general loss function:

\[
L(\hat{\Pi}, \Pi; p, q) = \min_T \left\{ \left( \frac{1}{n} \sum_{i=1}^n \frac{\theta_i}{\bar{\theta}} \| T \hat{\pi}_i - \pi_i \|_q^q \right)^{1/q} \right\}, \quad \text{for } p \geq 0 \text{ and } q \geq 1.
\] (3.17)

The \( \ell^1 \)-loss in (1.5) is a special case with \( p = 0 \) and \( q = 1 \). When \( p > 0 \), the estimation errors are re-weighted by degree parameters. In many real applications, the \( \pi_i \) of high-degree nodes are more interesting, so this general loss metric is relevant. We are particularly interested in the case of \( p = 1/2 \) and \( q = 1 \). We call \( L^w(\hat{\Pi}, \Pi) \equiv L(\hat{\Pi}, \Pi; 1/2, 1) \) the weighted \( \ell^1 \)-loss.

Corollary 3.3. Suppose conditions of Theorem 3.2 hold. Let \( \hat{\Pi} \) be the estimator from Mixed-SCORE-Laplacian and \( L(\hat{\Pi}, \Pi; p, q) \) be the loss metric in (3.17), for \( p \geq 0 \) and \( q \geq 1 \). Then,

\[
L(\hat{\Pi}, \Pi; p, q) \leq C \sqrt{\log(p)} \left( \int t^{p} \min\left\{ \frac{\text{err}_n^q}{(t \wedge 1)^{q/2}}, 1 \right\} dF_n(t) \right)^{1/q}.
\] (3.18)

Furthermore, in the special case of \( p = 1/2 \) and \( q = 1 \),

\[
\mathbb{E} L^w(\hat{\Pi}, \Pi) \leq C \sqrt{\log(n)} \text{err}_n.
\] (3.19)

For the loss metric \( L^w(\hat{\Pi}, \Pi) \), we provide a matching lower bound as follows. It suggests that the optimal rate for this loss metric does not depend on degree heterogeneity and that Mixed-SCORE-Laplacian is still optimally adaptive under this loss metric.

Theorem 3.4. Fix constants \( c_1-c_4 \). Given \( (n, K) \) and \( \theta \in \mathbb{R}^n \) such that \( F_n(\text{err}_n^2) \leq \tilde{c} \), for a constant \( \tilde{c} \in (0,1) \). Let \( Q_n(\theta) \) be the collection of \( (\Pi, P) \) satisfying (3.1)-(3.4). There exists a constant \( C > 0 \) such that, for all sufficiently large \( n \),

\[
\inf_{\Pi} \sup_{(\Pi, P) \in Q_n(\theta)} \mathbb{E} L^w(\hat{\Pi}, \Pi) \geq C \text{err}_n.
\] (3.20)
4 Proof ideas for entrywise eigenvector analysis

To study the error rates of Mixed-SCORE-Laplacian, the key is to derive the row-wise large-deviation bounds for eigenvectors in Theorem 3.1. In this section, we explain the proof ideas for this theorem.

There are two possible approaches for entry-wise eigenvector analysis: The first is using a Taylor expansion of the matrix resolvent [Erdős et al. (2013)]. Take the analysis of $\hat{\xi}_1$ for example. The key is to derive a large-deviation bound for $|\epsilon'_k(L - L_0)^k\xi_1|$, for $1 \leq k \leq O(\log(n))$. When degree heterogeneity is severe and when the network is sparse (so that the errors in $\|H - H_0\|$ may have a non-negligible effect), it is unclear how to conduct such large-deviation analysis. The second is using the “leave-one-out” idea in [Abbe et al. (2020)] to create a proxy of $\hat{\xi}_1$ by zeroing out the $i$th row and column of $L$. However, the leave-out approach in [Abbe et al. (2020)] only works for the adjacency matrix. For $L$, if we zero out its $i$th row & column and let $\xi_1$ be the leading eigenvector of the resulting matrix, then $\xi_1$ is not independent of the $i$th row and column of $L$, making the proofs in [Abbe et al. (2020)] ineligible.

We still want to use the leave-one-out idea, but our way of “leaving out” is more complicated than that in [Abbe et al. (2020)]. Without loss of generality, we set $\tau = 1$ in (2.6). Write $W = A - \Xi A$. Note that $A = \Omega - \text{diag}(\Omega) + W$. Fix $1 \leq i \leq n$. Let $W^{(i)}$ be the matrix obtained by zeroing-out the $i$-th row and column of $W$. Let

$$A^{(i)} = \Omega - \text{diag}(\Omega) + W^{(i)}.$$  

We similarly define a “leave-out” proxy of $H$. By direct calculations, $H(j, j) = \sum_s A(j, s) + \frac{1}{n} \sum_{(s, \ell)} A(s, \ell)$ Each $H(j, j)$ depends on the $i$th row and column of $W$. We define $\hat{H}^{(i)}$ by completely removing the effect of the $i$th row and column of $W$:

$$\hat{H}^{(i)}(j, j) = \begin{cases} \sum_{s \neq i} A(j, s) + \frac{1}{n} \sum_{s \neq i, \ell \neq i} A(s, \ell) + \sum_{s \neq i} \Omega(i, s), & \text{for } j \neq i, \\ \sum_{s \neq i} \Omega(j, s) + \frac{1}{n} \sum_{s \neq i, \ell \neq i} A(s, \ell) + \sum_{s \neq i} \Omega(i, s), & \text{for } j = i. \end{cases}$$  \hspace{1cm} (4.1)

Recall that

$$L = H^{-\frac{1}{2}} AH^{-\frac{1}{2}}, \quad L_0 = H_0^{-\frac{1}{2}} \Omega H_0^{-\frac{1}{2}}.$$  

We now define two intermediate matrices:

$$\bar{L}^{(i)} := (\hat{H}^{(i)})^{-\frac{1}{2}} \Omega (\hat{H}^{(i)})^{-\frac{1}{2}}, \quad \bar{T}^{(i)} := (\hat{H}^{(i)})^{-\frac{1}{2}} A^{(i)} (\hat{H}^{(i)})^{-\frac{1}{2}}.$$  \hspace{1cm} (4.2)

For $\bar{L}^{(i)}$, we define $(\hat{\lambda}_k^{(i)}, \hat{\xi}_1^{(i)}, \hat{\xi}_k^{(i)})$ similarly as before, and for $\bar{T}^{(i)}$, we define $(\hat{\lambda}_k^{(i)}, \hat{\xi}_1^{(i)}, \hat{\xi}_k^{(i)})$ similarly as before. So far, we have created two collections of proxy eigenvectors:

- $\hat{\xi}_1^{(i)}, \ldots, \hat{\xi}_k^{(i)}$: They are from replacing $H_0$ with $\hat{H}^{(i)}$ in $L_0$, followed by eigen-decomposition. We use $\hat{\xi}_k^{(i)}$ as a ‘proxy’ to $\xi_k$. Compared with $\xi_k$, $\hat{\xi}_k^{(i)}$ is more convenient for subsequent analysis.

- $\hat{\xi}_1^{(i)}, \ldots, \hat{\xi}_k^{(i)}$: They are obtained by first replacing $(A, H)$ with $(A^{(i)}, \hat{H}^{(i)})$ in $L$ and then conducting eigen-decomposition. Since $\hat{\xi}_k^{(i)}$ is independent of the $i$th row and
the $i$th column of $W$, it will serve as a key intermediate quantity to connect $\hat{\xi}_k$ and $\hat{\xi}_k$.

Below, we prove the claim for $\hat{\xi}_1(i)$ in Theorem 3.1. The claims for $\hat{X}_1$ can be proved similarly, which proof is relegated to the supplementary material.

### 4.1 Proof of the first claim in Theorem 3.1

First, we give a lemma about the difference between $\xi_1$ and $\hat{\xi}_1(i)$. Note that both $L_0$ and $\tilde{L}^{(i)}$ are low rank, and they are different only up to multiplication of diagonal matrices. This allows us to get a tractable form of $\hat{\xi}_1(i) - \xi_1$. The following lemma is proved in the supplementary material.

**Lemma 4.1.** Suppose the conditions of Theorem 3.1 hold. We pick the sign of $\xi_1$ such that $\xi_1(1) \geq 0$. For each $1 \leq i \leq n$, we pick the sign of $\hat{\xi}_1(i)$ such that $\hat{\xi}_1(i)(1) \geq 0$. With probability $1 - o(n^{-3})$, simultaneously for all $1 \leq i, j \leq n$,

$$
|\hat{\xi}_1(i)(j) - \xi_1(j)| \leq CK^2 \kappa_j \left( \sqrt{\frac{\theta}{n}} \wedge 1 \right), \quad \text{where} \quad \kappa_j := \sqrt{\frac{\log n}{n\theta^2}} \sqrt{\frac{\theta_j}{n\theta}}. \quad (4.3)
$$

Next, we study the difference between $\tilde{\xi}_1(i)$ and $\hat{\xi}_1(i)$. From $L_0$ to $\tilde{L}^{(i)}$, the perturbation is only from multiplication of diagonal matrices. However, from $\tilde{L}^{(i)}$ to $L$, the perturbation involves adding the noise matrix $W$, and we are unable to mimic the proof of Lemma 4.1 any more: If we do so, the bound we get will again depend on $\hat{\xi}_1 - \tilde{\xi}_1(i)$, which creates a circular reasoning. Our strategy is to use $\tilde{\xi}_1(i)$ to help escape from the circular reasoning.

**Lemma 4.2.** Under the assumptions in Theorem 3.1. With probability $1 - o(n^{-3})$, there exists some $w \in \{1, -1\}$ such that, simultaneously for all $1 \leq i \leq n$,

$$
|w\hat{\xi}_1(i) - \tilde{\xi}_1(i)(i)| \leq CK^2 \alpha_n^{-1} \kappa_i + CK \alpha_n^{-1} |e'_i \Delta \hat{\xi}_1|, \quad (4.4)
$$

$$
|e'_i \Delta \hat{\xi}_1| \leq |e'_i \Delta \tilde{\xi}_1(i)| + |e'_i \Delta (\tilde{\xi}_1(i) - \hat{\xi}_1(i))| + \frac{C}{\sqrt{n\theta^2}} \|w\hat{\xi}_1 - \tilde{\xi}_1(i)\|, \quad (4.5)
$$

where $\kappa_i := \sqrt{\frac{\log n}{n\theta^2}} \alpha_n \sqrt{\frac{\theta}{n\theta}}$, $\Delta \equiv \Delta(i) := (\hat{H}^{(i)})^{-1/2} W H^{-1/2}$, and we fix the choice of $\tilde{\xi}_1(i)$ such that $\text{sgn}(\tilde{\xi}_1(i)) = 1$.

The proof of Lemma 4.2 is quite involved, which is relegated to the supplementary material. Write $\tilde{\xi}_1(i) = \tilde{\xi}_1, \hat{\xi}_1(i) = \hat{\xi}_1$ and $\hat{H}^{(i)} = \hat{H}$ for short. The role of Lemma 4.2 is to reduce the analysis of $|\hat{\xi}_1(i) - \tilde{\xi}_1(i)|$ to the analyses of $|e'_i \Delta \hat{\xi}_1|, |e'_i \Delta (\tilde{\xi}_1(i) - \hat{\xi}_1(i))|$ and $\|\hat{\xi}_1 - \tilde{\xi}_1\|$. Among the three quantities, the first two are relatively easy to bound, because $e'_i \Delta \approx e'_i H^{-1/2} W H^{-1/2}$, which is independent of $(\tilde{\xi}_1, \hat{\xi}_1)$ (if we put aside the effect of $\hat{H}$, which is relatively easy to control); therefore, we can apply large-deviation inequalities. What causes a trouble is the third quantity $\|\hat{\xi}_1 - \tilde{\xi}_1\|$. We need another key technical lemma:
Lemma 4.3. Under the assumptions in Lemma 4.2, with probability $1 - o(n^{-3})$, for the same $w \in \{1, -1\}$ in Lemma 4.2, simultaneously for all $1 \leq i \leq n$,

$$
|e'_i \Delta \tilde{\xi}^{(i)}_1| \leq C\tilde{\kappa}_i, \quad \text{where} \quad \tilde{\kappa}_i := \frac{1}{n\theta} \sqrt{\frac{\log n}{n\theta^2}} \sqrt{n\theta_1 \vee \log n}, \quad (4.6)
$$

$$
|e_i \Delta (\tilde{\xi}^{(i)}_1 - \xi^{(i)}_1)| \leq C\tilde{\kappa}_i \left(1 + n\tilde{\theta} \|(\tilde{H}^{(i)})^{-1/2}(w\tilde{\xi}_1 - \tilde{\xi}^{(i)}_1)\|_{\infty}\right) + \frac{C\log(n)}{n\theta^2} \|w\tilde{\xi}_1 - \tilde{\xi}^{(i)}_1\|, \quad (4.7)
$$

$$
\|w\tilde{\xi}_1 - \xi^{(i)}_1\| \leq \frac{CK\alpha_n}{\tilde{\kappa}_i} \left[1 + n\tilde{\theta} \|(\tilde{H}^{(i)})^{-1/2}(w\tilde{\xi}_1 - \tilde{\xi}^{(i)}_1)\|_{\infty}\right] + \frac{CK\sqrt{\log n}}{n\theta^2\alpha_n^2} \|w\tilde{\xi}_1 - \xi^{(i)}_1\|. \quad (4.8)
$$

The proof of Lemma 4.3 is quite complicated and tedious, which we relegate to the supplementary material. This lemma connects each term on the right hand side of (4.4)-(4.5) to $\|(\tilde{H}^{(i)})^{-1/2}(w\tilde{\xi}_1 - \tilde{\xi}^{(i)}_1)\|_{\infty}$ and $|\tilde{\xi}_1(i) - \xi^{(i)}_1(i)|$.

We now use Lemmas 4.1-4.3 to show the claim. In Lemma 4.2, although the vector $\tilde{\xi}^{(i)}_1$ depends on $i$, the number $w \in \{\pm 1\}$ is shared by all $1 \leq i \leq n$; similarly, the $w$ in Lemma 4.3 is also shared by all $1 \leq i \leq n$. Therefore, we can assume $w = 1$ in all claims, without loss of generality. When there is no confusion, we write $\tilde{\xi}^{(i)}_1 = \tilde{\xi}_1$, $\tilde{\xi}^{(i)}_1 = \tilde{\xi}_1$ and $\tilde{H}^{(i)} = \tilde{H}$ for short. We plug (4.6)-(4.8) into (4.4)-(4.5) and note that $\kappa_i \leq \tilde{\kappa}_i$. It gives

$$
|\tilde{\xi}_1(i) - \tilde{\xi}_1(i)| \leq CK^2\alpha_n^{-1}\tilde{\kappa}_i + CK\alpha_n^{-1}\tilde{\kappa}_i n\tilde{\theta} \|\tilde{H}^{-1/2}(\tilde{\xi}_1 - \tilde{\xi}_1)\|_{\infty} + \frac{CK^2\sqrt{\log n}}{n\theta^2\alpha_n^2} \|\tilde{\xi}_1(i) - \tilde{\xi}_1(i)\|. \quad (4.9)
$$

Since $K^2\sqrt{\log n} \ll n\theta^2\alpha_n^2$, we can rearrange the above inequality to get

$$
|\tilde{\xi}_1(i) - \tilde{\xi}_1(i)| \leq CK^2\alpha_n^{-1}\tilde{\kappa}_i + CK\alpha_n^{-1}\tilde{\kappa}_i n\tilde{\theta} \|\tilde{H}^{-1/2}(\tilde{\xi}_1 - \tilde{\xi}_1)\|_{\infty}. \quad (4.9)
$$

We further apply Lemma 4.1 and note that $\kappa_j \leq \tilde{\kappa}_j$ for all $j$. It follows that

$$
|\tilde{\xi}_1(i) - \xi_1(i)| \leq |\tilde{\xi}_1(i) - \tilde{\xi}_1(i)| + |\tilde{\xi}_1(i) - \xi_1(i)|
\leq CK^2\alpha_n^{-1}\tilde{\kappa}_i + CK\alpha_n^{-1}\tilde{\kappa}_i n\tilde{\theta} \|\tilde{H}^{-1/2}(\tilde{\xi}_1 - \xi_1)\|_{\infty}
\leq CK^2\alpha_n^{-1}\tilde{\kappa}_i + CK\alpha_n^{-1}\tilde{\kappa}_i n\tilde{\theta} \|H_0^{-1/2}(\hat{\xi}_1 - \xi_1)\|_{\infty}
\leq CK^2\alpha_n^{-1}\tilde{\kappa}_i + CK\alpha_n^{-1}\tilde{\kappa}_i n\tilde{\theta} \left(\|H_0^{-1/2}(\hat{\xi}_1 - \xi_1)\|_{\infty} + \|H_0^{-1/2}(\hat{\xi}_1 - \xi_1)\|_{\infty}\right).
$$

To bound the second term in the bracket, we apply Lemma 4.1 again and note that $H_0(j, j) \approx n\tilde{\theta} (\theta \vee \theta_j)$ (see (B.4)) and $\kappa_j[H_0(j, j)]^{-\frac{1}{2}} \leq \tilde{\kappa}_j[H_0(j, j)]^{-\frac{1}{2}} \leq \sqrt{\log n \frac{1}{n\theta^2}} \frac{1}{n\theta}$. It gives

$$
\|H_0^{-1/2}(\hat{\xi}_1 - \xi_1)\|_{\infty} \leq C\sqrt{\frac{\log n}{n\theta^2}} \frac{1}{n\theta}.
$$

Combining the above gives

$$
|\hat{\xi}_1(i) - \xi_1(i)| \leq CK^2\alpha_n^{-1}\tilde{\kappa}_i + CK\alpha_n^{-1}\tilde{\kappa}_i n\tilde{\theta} \|H_0^{-1/2}(\hat{\xi}_1 - \xi_1)\|_{\infty}. \quad (4.10)
$$
We multiply $H_0^{-\frac{1}{2}}(i, i)$ on both sides and use $\kappa_j[H_0(j, j)]^{-\frac{1}{2}} \leq \sqrt{\frac{\log n}{n\theta^2}} \frac{1}{n\theta}$ again. It gives

$$|e_i'H_0^{-\frac{1}{2}}(\hat{\xi}_1 - \xi_1)| \leq CK^2 \frac{\log n}{n\theta^2\alpha_2^2} \frac{1}{n\theta} + CK \frac{\log n}{n\theta^2\alpha_2^2} \|H_0^{-\frac{1}{2}}(\hat{\xi}_1 - \xi_1)\|_\infty.$$  

This is true for every $1 \leq i \leq n$. Since $\hat{\xi}_1 - \xi_1$ does not depend on $i$, we can bound $\|H_0^{-\frac{1}{2}}(\hat{\xi}_1 - \xi_1)\|_\infty$ by taking the maximum over $i$. Then, $\|H_0^{-\frac{1}{2}}(\hat{\xi}_1 - \xi_1)\|_\infty$ appears on both hand sides. Since $K \sqrt{\frac{\log n}{n\theta^2\alpha_2^2}} = o(1)$, we can re-arrange the terms to get

$$\|H_0^{-\frac{1}{2}}(\hat{\xi}_1 - \xi_1)\|_\infty \leq CK^2 \frac{\log n}{n\theta^2\alpha_2^2} \frac{1}{n\theta} \leq CK^2 \frac{1}{n\theta}. \quad (4.11)$$

Plugging (4.11) into (4.10) gives

$$|\hat{\xi}_1(i) - \xi_1(i)| \leq CK^2 \alpha_n^{-1}\kappa_i. \quad (4.12)$$

The claim (3.6) follows immediately by plugging in the definition of $\kappa_i$ in Lemma 4.3.

## 5 Least-favorable configurations and proof of the lower bound

The key of proving the lower bound arguments in Theorems 3.3-3.4 is to carefully construct the least-favorable configurations (LFC). The LFC for these two theorems are different. We start from the less complicated one, the LFC for the weighted $\ell^1$-loss $L^w(\hat{\Pi}, \Pi)$, and then modify it to construct the LFC for the standard $\ell^1$-loss $L(\hat{\Pi}, \Pi)$. The following notation is useful.

**Definition 5.1.** Given $(n, K, \alpha_n, \beta_n)$, $\theta \in \mathbb{R}^n$ and $P \in \mathbb{R}^{K \times K}$, let $Q_n(K, \theta, P, \alpha_n, \beta_n)$ denote the collection of eligible membership matrices $\Pi$ such that (3.1)-(3.4) are satisfied.

First, we construct the LFC for proving the lower bound in Theorem 3.4. We take a special form of $P$,

$$P^* = \beta_n I_K + (1 - \beta_n)1_K^11_K', \quad (5.1)$$

and construct a collection of $\Pi$. We need a well-known lemma (e.g., see Tsybakov (2009) for a proof):

**Lemma 5.1 (Varshamov-Gilbert bound for packing numbers).** For any $s \geq 8$, there exist $J \geq 2^s/8$ and $\omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(J)} \in \{0, 1\}^s$ such that $\omega^{(0)} = 0_s$ and $\|\omega^{(j)} - \omega^{(k)}\|_1 \geq s/8$, for all $0 \leq j < k \leq J$.  

19
In Theorem 3.4, we assume $F_n(\text{err}_n) \leq \tilde{c}$, for a constant $\tilde{c} \in (0, 1)$. Let $c = \frac{1+\tilde{c}}{2} \in (0, 1)$.

Let $n_1 = \lfloor K^{-1} cn \rfloor$ and $n_0 = n - Kn_1$. We set

$$\Pi^* = \left( \frac{1}{K} 1_K, \ldots, \frac{1}{K} 1_K, e_1, \ldots, e_1, \ldots, e_K, \ldots, e_K \right)'.$$  \hfill (5.2)

Without loss of generality, we can assume that those $\theta_i$’s corresponding to the pure nodes in $\Pi^*$ contains the top $\lfloor (c-\tilde{c})n \rfloor$ degrees and they are evenly assigned to different communities such that the average degrees of the pure nodes in different communities are of the same order; we can also assume that the first $n_0$ $\theta_i$’s satisfy $\theta_i/\bar{\theta} \geq \text{err}_n^2$, by the assumption of $F_n(\text{err}_n^2) \leq \tilde{c}$. Note that we can always find a permutation to achieve such $\theta$ and reconstruct $\Pi^*$ correspondingly. Let $m = \lfloor n_0/2 \rfloor$ and $r = \lfloor K/2 \rfloor$. We apply Lemma 5.1 to $s = mr$ to get $\omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(J)}$, where $J \geq 2^{(mr/8)}$. We re-arrange each $\omega^{(j)}$ to an $m \times r$ matrix, denoted as $H^{(j)}$, and then construct $\Gamma^{(j)} \in \mathbb{R}^{n \times K}$ whose nonzero entries only appear in the top left $(2m) \times (2r)$ block:

$$\Gamma^{(j)} = \begin{bmatrix} H^{(j)} & -H^{(j)} & 0_{m \times 1} \\ -H^{(j)} & H^{(j)} & 0_{m \times 1} \\ 0_{1 \times r} & 0_{1 \times r} & 0_{1 \times 1} \end{bmatrix}, \quad 0 \leq j \leq J. \hfill (5.3)$$

In (5.3), if $K$ is an even number, then the last column (consisting of zero entries) disappears; similarly, if $n_0$ is an even number, then the last row above the dashed line (consisting of zero entries) disappears. Let $\Theta = \text{diag}(\theta_1, \theta_2, \ldots, \theta_n)$. We construct $\Pi^{(0)}, \Pi^{(1)}, \ldots, \Pi^{(J)}$ by

$$\Pi^{(j)} = \Pi^* + \gamma_n \Theta^{-\frac{1}{2}} \Gamma^{(j)}, \quad \text{where} \quad \gamma_n = c_0 K^\frac{1}{2} (n\tilde{\theta} \beta_n^2)^{-\frac{1}{2}}, \quad \text{for} \ 0 \leq j \leq J, \hfill (5.4)$$

where $c_0 > 0$ is a properly small constant. The following theorem is proved in the supplementary material.

**Theorem 5.1.** Fix $c_1$-$c_4$ in (3.4)-(3.6) and $\tilde{c}$ in Theorem 3.4. Given any $(n, K, \alpha_n, \beta_n)$ and $\theta \in \mathbb{R}^n$ such that $F_n(\text{err}_n^2) \leq \tilde{c}$, let $P^*$ be as in (5.1), and construct $\Pi^{(0)}, \Pi^{(1)}, \ldots, \Pi^{(J)}$ as in (5.2)-(5.4). When $c_0$ in (5.4) is properly small, the following statements are true.

- For any constant $c_5 > 0$, let $Q_n(\theta, P^*) = Q_n(K, \theta, P^*, c_5 K, c_5 \beta_n)$ (see Definition 5.1). There exists a properly small $c_5$ such that $\Pi^{(j)}$ is contained in $Q_n(\theta, P^*)$, for $0 \leq j \leq J$.

- There exists a constant $C_1 > 0$ such that $\mathcal{L}^w(\Pi^{(j)}, \Pi^{(k)}) \geq C_1 \text{err}_n$, for all $0 \leq j < k \leq J$.

- Let $\mathcal{P}_j$ be the probability measure of a DCMM model with $(\theta, P^*, \Pi^{(j)})$ and let $\text{KL}(\cdot, \cdot)$ denote the Kullback-Leibler divergence. There exists a constant $\epsilon_1 \in (0, 1/8)$ such that $\sum_{1 \leq j \leq J} \text{KL}(\mathcal{P}_j, \mathcal{P}_0) \leq (1/8 - \epsilon_1) J \log(J)$.

Furthermore, $\inf_{\hat{\Pi}} \sup_{\Pi \in Q_n(\theta, P^*)} \mathbb{E} \mathcal{L}^w(\hat{\Pi}, \Pi) \geq C \text{err}_n$. 

20
Theorem 3.4 follows immediately from Theorem 5.1.

Next, we construct the LFC for proving the lower bound in Theorem 3.3. We still take $P^*$ in (5.1) and construct a collection of $\Pi$. Compared with the previous case, the targeted lower bound now depends on $F_n(\cdot)$, so that the construction is more sophisticated. We separate two cases according to whether the following holds:

$$\int_0^{\text{err}_n} dF(t) \leq C \int_{\text{err}_n}^{\infty} \frac{\text{err}_n}{\sqrt{t} \wedge 1} dF_n(t). \quad (5.5)$$

To understand (5.5), note that $\theta_i/\hat{\theta} \leq \text{err}_n$ is equivalent to $n\hat{\theta}i\hat{\theta}_n^2 \leq K^3 \log(n)$. For such a node $i$, the best estimator is the naive estimator $\pi_i^{\text{naive}} = 1/K$. In (5.5), the left hand side is the total contribution of these nodes in $L(\hat{\Pi}, \Pi)$, and the right hand side is the contribution of remaining nodes. Therefore, (5.5) guarantees that the rate of convergence of the unweighted $\ell^1$-loss is driven by those nodes for which we can indeed construct non-trivial estimators of $\pi_i$ from data. When (5.5) is violated, the lower bound can be proved by similar techniques but simpler least-favorable configurations. The details of this case is relegated in the supplementary materials and below we will focus on the case that (5.5) holds. Note that all examples in the end of Section 3.2 satisfy (5.5).

We need a technical lemma about the property of $F_n(\cdot)$ that satisfies the requirement in Theorem 3.3. It is proved in the supplementary material.

**Lemma 5.2.** Fix $\gamma > 0$ and $a_0 \in (0, 1)$. Given any $\theta \in \mathcal{G}_n(\gamma, a_0)$ (see Definition 3.1), recall that $F_n(\cdot)$ is the empirical distribution associated with $\eta_i = \theta_i/\hat{\theta}$, $1 \leq i \leq n$. Let $F_n$ be the empirical distribution associated with $\tilde{\eta}_i = \eta_i \wedge 1$. For any $c > 0$ and $\epsilon \in (0, 1)$, define

$$\tau_n(c, \epsilon) = \inf \left\{ t > 0 : \int_{\text{err}_n}^{t} d\tilde{F}(t) = (1 - \epsilon) \int_{\text{err}_n}^{c} d\tilde{F}(t) \right\}.$$ 

If $F_n(\cdot)$ satisfies (5.5), then there exists a number $c_n > \text{err}_n^2$ and a constant $\tilde{a}_0 \in (0, 1)$ such that $F_n(c_n) \leq 1 - \tilde{a}_0$ and

$$\int_{\tau_n(c_n, 1/8)}^{c_n} \frac{1}{\sqrt{t} \wedge 1} dF_n(t) \geq \tilde{a}_0 \int_{\text{err}_n^2}^{\infty} \frac{1}{\sqrt{t} \wedge 1} dF_n(t). \quad (5.6)$$

We now construct a collection of $\Pi$ using $c_n$ in Lemma 5.2. We re-order $\theta_i$’s such that

$$\theta_(1) \leq \theta_(2) \leq \ldots \leq \theta_(n). \quad (5.7)$$

From the way $\tilde{\eta}_i$’s are defined, this ordering also implies that $\tilde{\eta}(1) \leq \tilde{\eta}(2) \leq \ldots \leq \tilde{\eta}(n)$. Let $c_n$ be as in Lemma 5.2. Define

$$s_n = \min \{1 \leq i \leq n : \tilde{\eta}(i) \geq \text{err}_n^2\}, \quad n_0 = \min \{1 \leq i \leq n : \tilde{\eta}(i) \geq c_n\} - s_n.$$ 

It follows from the definition of $\tilde{F}_n$ that $n_0$ is approximately the total number of $\tilde{\eta}_i$’s such that $\text{err}_n^2 \leq \tilde{\eta}_i < c_n$. By definition of $\tau_n(c_n, 1/8)$, we have $\tilde{\eta}_i[\tau_n(\eta_i, 8) + s_n] \leq \tau_n(c_n, 1/8)$. Combining these claims with (5.6) gives

$$\frac{1}{n} \sum_{\tau_n(\eta_i, 8) + s_n \leq n_0} \frac{1}{\sqrt{\eta(i) \wedge 1}} \gtrsim \tilde{a}_0 \int_{\text{err}_n^2}^{\infty} \frac{1}{\sqrt{t} \wedge 1} dF_n(t).$$
We multiply $\text{err}_n$ on both hand sides. By the condition \((5.5)\), we have $\text{err}_n \int_{\text{err}_n^2}^{\infty} \frac{1}{\sqrt{\lambda t}} dF_n(t) \geq C^{-1} \int \min \{ \frac{\text{err}_n}{\sqrt{\lambda t}}, 1 \} dF_n(t)$, which yields a lower bound for the right hand side. For the left hand side, we plug in $\eta = \theta \land \bar{\theta}$. It follows that

$$\sqrt{\frac{K^3}{n\bar{\theta}^2_n}} \cdot \frac{1}{n} \sum_{i=0}^{n_0/8} \frac{1}{\sqrt{\theta(i) \land \bar{\theta}}} \geq \int \min \{ \frac{\text{err}_n}{\sqrt{t \land 1}}, 1 \} dF_n(t).$$

Let $\mathcal{M}_0$ be the index set of nodes that are are ordered between $s_n$ and $s_n + n_0$ in \((5.7)\). Let $\gamma_n = c_0 K^{\frac{1}{2}} (n\bar{\theta}^2_n)^{-\frac{1}{2}}$. The above implies that

$$\gamma_n \inf_{\mathcal{M} \subset \mathcal{M}_0, |M| \geq n_0/8} \left\{ \frac{1}{n} \sum_{i \in M} \frac{1}{\sqrt{\theta_i \land \bar{\theta}}} \right\} \geq K^{-1} \int \min \{ \frac{\text{err}_n}{\sqrt{t \land 1}}, 1 \} dF_n(t). \quad (5.8)$$

The set $\mathcal{M}_0$ plays a key role in the construction of the least-favorable configurations. We now re-arrange nodes by putting nodes in $\mathcal{M}_0$ as the first $n_0$ nodes, with the last $n-n_0$ nodes ordered in a way such that the average degrees of the pure nodes in different communities of $\Pi^*$ are of the same order (such an ordering always exists). After node re-arrangement, we construct $\Pi^*$ and $\Gamma^{(0)}, \Gamma^{(1)}, \ldots, \Gamma^{(J)}$ in the same way as in \((5.2)-(5.3)\). Let $\hat{\theta}_i = \theta_i \land \bar{\theta}$ and $\hat{\Theta} = \text{diag}(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)$. Let

$$\Pi^{(j)} = \Pi^* + \gamma_n \hat{\Theta}^{-\frac{1}{2}} \Gamma^{(j)}, \quad 0 \leq j \leq J, \quad (5.9)$$

where $\gamma_n = c_0 K^{\frac{1}{2}} (n\bar{\theta}^2_n)^{-\frac{1}{2}}$ is the same as in \((5.8)\). The following theorem is an analog of Theorem \ref{thm:lower_bound_theta} for the unweighted $\ell^1$-loss and is proved in the supplementary material.

**Theorem 5.2.** Fix $c_1 - c_4$ in \((3.1)-(3.4)\) and $(\gamma, a_0)$ in Theorem \ref{thm:lower_bound_theta}. Given $(n, K, \alpha_n, \beta_n)$ and $\theta \in \mathcal{G}_n(\gamma, a_0)$, let $P^*$ be as in \((5.1)\), and construct $\Pi^{(0)}, \Pi^{(1)}, \ldots, \Pi^{(J)}$ as in \((5.9)\). When $c_0$ in \((5.4)\) is properly small, the following statements are true.

- For any constant $c_5 > 0$, let $Q_n(\theta, P^*) = Q_n(K, \theta, P^*, c_5 K, c_5 \beta_n)$. There exists a properly small $c_5$ such that $\Pi^{(j)}$ is contained in $Q_n(\theta, P^*)$, for $0 \leq j \leq J$.

- There exists a constant $C_2 > 0$ such that $\mathcal{L}(\Pi^{(j)}, \Pi^{(k)}) \geq C_2 \int \min \{ \frac{\text{err}_n}{\sqrt{\lambda t}}, 1 \} dF_n(t)$, for all $0 \leq j < k \leq J$.

- Let $\mathcal{P}_j$ be the probability measure of a DCMM model with $(\theta, P^*, \Pi^{(j)})$ and let $\text{KL}(\cdot, \cdot)$ denote the Kullback-Leibler divergence. There exists a constant $\epsilon_2 \in (0, 1/8)$ such that $\sum_{1 \leq j \leq J} \text{KL}(\mathcal{P}_j, \mathcal{P}_0) \leq (1/8 - \epsilon_2) J \log(J)$.

Furthermore, $\inf_{\Pi \in Q_n(\theta, P^*)} \mathbb{E} L(\hat{\Pi}, \Pi) \geq C \int \min \{ \frac{\text{err}_n}{\sqrt{\lambda t}}, 1 \} dF_n(t)n$.

**Theorem 3.3** follows immediately from Theorem 5.2.

**Remark:** In Theorems \ref{thm:lower_bound_theta} \ref{thm:lower_bound_theta_specific}, we fix $P = P^*$ and prove the lower bounds by taking supreme over a class of $\Pi$. Such lower bounds are not only $\theta$-specific but also $P$-specific, and they are stronger than the $\theta$-specific lower bounds in Theorems \ref{thm:lower_bound_theta_specific} \ref{thm:lower_bound_theta_specific}. In Section \ref{sec:lower_bounds}, of the supplementary material, we show that we can prove such $P$-specific lower bounds for an arbitrary $P$ if one of the following holds as $n \to \infty$: (a) $(K, P)$ are fixed; (b) $(K, P)$ can depend on $n$, but $K \leq C$ and $P_1 K \propto 1_K$; (c) $(K, P)$ can depend on $n$, and $K$ can be unbounded, but $P_1 K \propto 1_K$ and $|\lambda_2(P)| \leq C \beta_n = o(1)$. 

22
6 Simulations

We conduct two experiments. In Experiment 1, we compare the performance of Mixed-SCORE-Laplacian with the orthodox Mixed-SCORE [Jin et al. (2017)] that has no pre-PCA normalization. In Experiment 2, we investigate the node-wise errors of Mixed-SCORE-Laplacian and study its relationship with $\theta_i$.

In both experiments, we fix $K = 2$ and $P = \beta_n I_2 + (1 - \beta_n)1_2 1_2'$. The degree parameters $\theta = (\theta_1, \ldots, \theta_n)'$ are generated as follows: We first draw $\theta^0_1, \theta^0_2, \ldots, \theta^0_n$ independently from a fixed distribution $F$; then, let $\theta_i = b_n \cdot \theta^0_i / \|\theta^0\|$, so that the $\ell^2$-norm of $\theta$ is equal to $b_n$. Write $\text{SNR} = b_n(1 - P(1,2)) = b_n \cdot \beta_n$. This quantity controls the signal-to-noise ratio. In each experiment, we consider two choices of $F$: Uniform([0.3, 5]), which corresponds to moderate degree heterogeneity, and Pareto(10, 0.3) (10 is the scale parameter and 0.3 is the shape parameter), which corresponds to severe degree heterogeneity. The membership matrix $\Pi = [\pi_1, \pi_2, \ldots, \pi_n]'$ is generated as follows: for the first and second 15% of nodes, we set $\pi_i = (1, 0)'$ and $\pi_i = (0, 1)'$, respectively; for the remaining 70% of nodes, $\pi_i = (t_i, 1 - t_i)$, where $t_i$’s are independently drawn from Uniform(0, 1).

![Figure 1: Comparison of Mixed-SCORE-Laplacian (solid curves) and orthodox Mixed-SCORE (dashed curves). In both panels, the x-axis is $\|\theta\|$, and the y-axis is either the unweighted $\ell^1$-loss $\mathcal{L}(\hat{\Pi}, \Pi)$ (blue curves) or the weighted $\ell^1$-loss $\mathcal{L}^w(\hat{\Pi}, \Pi)$ (red curves), averaged over 100 repetitions. The left panel corresponds to moderate degree heterogeneity, where $\theta_i$’s are drawn from a uniform distribution. The right panel corresponds to severe degree heterogeneity, where $\theta_i$’s are drawn from a Pareto distribution. $(n, K) = (2000, 2).$](image)

**Experiment 1:** Comparison with the orthodox Mixed-SCORE. We compare the numerical performances of Mixed-SCORE-Laplacian (MSL) and orthodox Mixed-SCORE (OMS) [Jin et al. (2017)]. A major difference between two methods is that OMS uses $M = I_n$ in (1.7), while our proposed MSL uses $M = [\text{diag}(d_1, \ldots, d_n) + \hat{d}I_n]^{-1/2}$. Fix $(n, K) = (2000, 2)$. We consider two sub-experiments: In Experiment 1.1, the fixed distribution $F$ is Uniform([0.3, 5]); in Experiment 1.2, $F$ is Pareto(10, 0.3). We let $b_n = \|\theta\|$ range in {5, 6, ..., 12} in Experiment 1.1 and range in {5, 5.5, 6, 6.5, ..., 8} in Experiment 1.2. As
Figure 2: $\theta_i$’s are generated from $\mathcal{U}([0.3, 5])$ and Pareto$(10, 0.3)$, respectively for the left and the right. In both plots, the slope of the red lines is $-1/2$. The results are based on 100 repetitions.

As $b_n$ varies, we change $\beta_n$ accordingly by fixing SNR $= b_n \cdot \beta_n = 4.5$. We report both the unweighted $\ell^1$-loss $\mathcal{L}(\Pi, \Pi)$ and the weighted $\ell^1$-loss $\mathcal{L}^w(\Pi, \Pi)$, over 100 repetitions. The results are displayed in Figure 1.

We first look at the unweighted $\ell^1$-loss (blue curves). When degree heterogeneity is moderate (left panel of Figure 1), MSL and OMS have similar performances. This is as expected, because both methods can attain the optimal rate under moderate degree heterogeneity. When degree heterogeneity is severe (right panel of Figure 1), MSL significantly outperforms OMS. This is consistent with our theory, where the rate of MSL is optimal in this case but the rate of OMS is non-optimal. These numerical results also confirm the advantage of using a pre-PCA normalization. We then look at the weighted $\ell^1$-loss (red curves), where the conclusions are similar. Note that for each method, the magnitudes of two loss metrics are close under moderate degree heterogeneity (left panel) and quite different under severe degree heterogeneity (right panel). This is also consistent with the definitions of two loss metrics.

**Experiment 2: Node-wise errors of Mixed-SCORE-Laplacian**

We investigate the errors of Mixed-SCORE-Laplacian at individual $\hat{\pi}_i$’s and study its relationship with $\theta_i$. Theorem 3.2 claims that the error at $\hat{\pi}_i$ is approximately proportional to $(\theta_i \wedge \bar{\theta})^{-1/2}$. To verify this result, we plot log($\|\hat{\pi}_i - \pi_i\|_1$) versus log($\theta_i$), for $\theta_i \leq \bar{\theta}$. We expect the scatter plot to fit a straight light with a slope of $-1/2$. Fix $(n, K) = (2000, 2)$. We first generate $\theta$ and $\Pi$ as in Experiment 1. We then fix $(\theta, \Pi)$, generate 100 networks, and compute the average of $\|\hat{\pi}_i - \pi_i\|_1$ over these 100 repetitions, for each $1 \leq i \leq n$. Similarly as in Experiment 1, we consider two experiments: In Experiment 2.1, the fixed distribution $F$ is Uniform([0.3, 5]); in Experiment 2.2, $F$ is Pareto$(10, 0.3)$. We fix $b_n = \|\theta\| = 26$ and SNR $= b_n \cdot \beta_n = 23$. The results are displayed in Figure 2. For both moderate degree heterogeneity (left panel) and severe degree heterogeneity (right panel), the plot fits reasonably well a straight line with
slope $-1/2$. This verifies the claim of Theorem 3.2. Interestingly, when degree heterogeneity is more severe, the empirical evidence of Theorem 3.2 is even stronger.

7 Discussions

Many real networks have severe degree heterogeneity. It motivates the study of network models equipped with node-wise degree parameters. When the target quantity is the underlying community structure, the large number of degree parameters are nuisance. There was little understanding to how these nuisance parameters affect the inference of community structure. We give a rigorous answer in the context of mixed membership estimation. Under the DCMM model, we show that the $n$ degree parameters $\theta_1, \theta_2, \ldots, \theta_n$ affect the optimal rate of mixed membership estimation through $\bar{\theta}$ and $F_n(\cdot)$, where $\bar{\theta}$ captures the overall sparsity and $F_n(\cdot)$ captures the degree heterogeneity. We identify both examples where the degree heterogeneity does or does not alter the optimal rate.

A desirable mixed membership estimation method should be optimally adaptive to degree heterogeneity. The OMS Jin et al. (2017) is a spectral algorithm. It attains the optimal rate under moderate degree heterogeneity but not severe degree heterogeneity. We combine OMS with a pre-PCA normalization by the regularized graph Laplacian, and show that the new algorithm achieves the optimal rate under quite arbitrary degree heterogeneity. On the one hand, it is surprising that such a simple variant of OMS can be optimally adaptive. On the other hand, it requires a lot of non-trivial insights and analysis to come to this conclusion: (a) We first derive the optimal rate, so that we have a benchmark to access the optimality of any method. (b) We point out that a pre-PCA normalization like (1.7) is promising to tackle severe degree heterogeneity. (c) We calculate the entry-wise SNRs of empirical eigenvectors and find out that $M \propto \Theta^{-1/2}$ is the ideal choice, which motivates the use of graph Laplacian. (d) With many technical efforts (especially the large-deviation analysis of eigenvectors in Section 4), we rigorously prove that the resulting spectral algorithm is optimally adaptive. Our contribution is far more than a naive combination of OMS and graph Laplacian. Without our insights and analysis, it was even unclear whether there exists an $M$ that is optimal, not to say whether graph Laplacian is the optimal $M$.

As a byproduct of our analysis, we provide a new row-wise large-deviation bound for the leading eigenvectors of the regularized graph Laplacian. Most existing results of entrywise eigenvector analysis (e.g., Abbe et al. (2020)) focus on the case of no degree heterogeneity or moderate degree heterogeneity, and study the adjacency matrix. We provide the first entrywise eigenvector analysis that simultaneously (i) deals with graph Laplacian, (ii) allows for severe degree heterogeneity, and (iii) yields the $\theta_i$-dependent bound for each entry. We believe this result is of independent interest.

References

Abbe, E., J. Fan, K. Wang, and Y. Zhong (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* 48(3), 1452–1474.
Airoldi, E., D. Blei, S. Fienberg, and E. Xing (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* 9, 1981–2014.

Araújo, M. C. U., T. C. B. Saldanha, R. K. H. Galvao, T. Yoneyama, H. C. Chame, and V. Visani (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems* 57(2), 65–73.

Bandeira, A. S. and R. Van Handel (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.* 44(4), 2479–2506.

Chen, Y., X. Li, and J. Xu (2018). Convexified modularity maximization for degree-corrected stochastic block models. *Ann. Statist.* 46(4), 1573–1602.

Erdős, L., A. Knowles, H.-T. Yau, and J. Yin (2013). Spectral statistics of erdős–rényi graphs i: Local semicircle law. *Ann. Probab.* 41(3B), 2279–2375.

Fan, J., Y. Fan, X. Han, and J. Lv (2022). SIMPLE: Statistical inference on membership profiles in large networks. *J. R. Stat. Soc. Ser. B.* 84(2), 630–653.

Gao, C., Z. Ma, A. Y. Zhang, and H. H. Zhou (2018). Community detection in degree-corrected block models. *Ann. Statist.* 46(5), 2153–2185.

Horn, R. and C. Johnson (1985). *Matrix Analysis*. Cambridge University Press.

Ji, P., J. Jin, Z. T. Ke, and W. Li (2022). Co-citation and co-authorship networks of statisticians. *J. Bus. Econom. Statist.* 40, 469–485.

Jin, J. (2015). Fast community detection by score. *Ann. Statist.* 43(1), 57–89.

Jin, J., Z. T. Ke, and S. Luo (2017). Estimating network memberships by simplex vertex hunting. *arXiv:1708.07852*.

Jin, J., Z. T. Ke, and S. Luo (2021). Optimal adaptivity of signed-polygon statistics for network testing. *Ann. Statist.* 49(6), 3408–3433.

Jin, J., Z. T. Ke, and S. Luo (2022). Improvements on SCORE, especially for weak signals. *Sankhya A* 84(1), 127–162.

Ke, Z. T. and J. Jin (2021). The SCORE normalization, especially for highly heterogeneous network and text data. *Manuscript*.

Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* 43(1), 215–237.

Liu, F., D. Choi, L. Xie, and K. Roeder (2018). Global spectral clustering in dynamic networks. *Proc. Natl. Acad. Sci.* 115(5), 927–932.

Ma, Z., Z. Ma, and H. Yuan (2020). Universal latent space model fitting for large networks with edge covariates. *J. Mach. Learn. Res.* 21(4), 1–67.

Mao, X., P. Sarkar, and D. Chakrabarti (2021). Estimating mixed memberships with sharp eigenvector deviations. *J. Amer. Statist. Assoc.* 116(536), 1928–1940.

Qin, T. and K. Rohe (2013). Regularized spectral clustering under the degree-corrected stochastic block-model. In *Adv. Neural Inf Process. Syst.*., pp. 3120–3128.
Rohe, K., S. Chatterjee, and B. Yu (2011, 08). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* 39(4), 1878–1915.

Tang, M. and C. E. Priebe (2018). Limit theorems for eigenvectors of the normalized laplacian for random graphs. *Ann. Statist.* 46(5), 2360–2415.

Tsybakov, A. B. (2009). Introduction to nonparametric estimation. revised and extended from the 2004 french original. translated by vladimir zaiats.

Zhang, A. Y. and H. H. Zhou (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* 44(5), 2252–2280.

Zhang, Y., E. Levina, and J. Zhu (2020). Detecting overlapping communities in networks using spectral methods. *SIAM J. Math. Data Sci.* 2(2), 265–283.
A Proof of Lemma 2.1

Recall that $L_0 = H_0^{-\frac{1}{2}} \Theta \Pi (P \Pi' \Theta H_0^{-\frac{1}{2}})$. Under the condition that each community has at least one pure node, $L_0$ has a rank $K$. It follows that $L_0$ has the same column space as $H_0^{-\frac{1}{2}} \Theta \Pi$. Meanwhile, $\Xi = [\xi_1, \xi_2, \ldots, \xi_K]$ also has the same column space as $L_0$. Therefore, there exists a non-singular matrix $B \in \mathbb{R}^{K \times K}$ such that

$$\Xi = H_0^{-\frac{1}{2}} \Theta \Pi B.$$  

Write $B = [b_1, b_2, \ldots, b_K]$. Define $v_1, v_2, \ldots, v_K \in \mathbb{R}^{K-1}$ by $v_k(\ell) = b_{\ell+1}(k)/b_1(k)$, for $1 \leq k \leq K, 1 \leq \ell \leq K-1$. Write $V = [v_1, v_2, \ldots, v_K]' \in \mathbb{R}^{K \times (K-1)}$. It follows that

$$B = \text{diag}(b_1)[1_K, V].$$

By definition of $R$, $[1_n, R] = [\text{diag}(\xi_1)]^{-1} \Xi$. It follows that

$$[1_n, R] = [\text{diag}(\xi_1)]^{-1} \Xi = [\text{diag}(\xi_1)]^{-1} H_0^{-\frac{1}{2}} \Theta \Pi B = [\text{diag}(\xi_1)]^{-1} H_0^{-\frac{1}{2}} \Theta \Pi \text{diag}(b_1)[1_K, V].$$

Define $W = [\text{diag}(\xi_1)]^{-1} H_0^{-\frac{1}{2}} \Theta \Pi \text{diag}(b_1)$. The above equation implies that $1_n = W 1_K$ and $R = WV$. Denote by $w'_i$ the $i$th row of $V$. It follows that $w'_i 1_K = 1$ and $r_i = \sum_{k=1}^{K} w_i(k) v_k$. Furthermore, under the condition (3.3), we can show that both $\xi_1$ and $b_1$ are strictly positive vectors; the proof is similar to the proof of Lemma B.4 of [Jin et al. (2017)], which we omit. It suggests that $W$ is also a nonnegative matrix. Combining the above, each $r_i$ is a convex combination of $v_1, v_2, \ldots, v_K$. This proves the simplex structure.

We now derive the connection between $w_i$ and $\pi_i$. Write $\alpha_i = \xi_1^{-1}(i) H_0^{-\frac{1}{2}}(i, i) \theta_i$. Then, $w'_i = \alpha_i \cdot \pi'_i \text{diag}(b_1) = \alpha_i \cdot (\pi_i \circ b_1)$. Since $\|w_i\|_1 = 1$, we immediately have $\alpha_i = 1/\|\pi_i \circ b_1\|_1$. This proves that $w_i = \frac{1}{\|\pi_i \circ b_1\|_1} (\pi_i \circ b_1)$. To get the expression of $B_1$, we notice that

$$\Lambda = \Xi L_0 \Xi = (H_0^{-\frac{1}{2}} \Theta \Pi B)'(H_0^{-\frac{1}{2}} \Theta \Pi P \Pi' \Theta H_0^{-\frac{1}{2}})(H_0^{-\frac{1}{2}} \Theta \Pi B) = B'(\Pi' \Theta D_\theta^{-1} \Theta \Pi)P(\Pi' \Theta D_\theta^{-1} \Theta \Pi)B' = K^{-2} \cdot B' G P G B,$$

where $D_\theta$ and $G$ are as defined in Section 3 and we note that $D_\theta$ is actually $H_0$. Moreover, $G = K \cdot (H_0^{-\frac{1}{2}} \Theta \Pi)'(H_0^{-\frac{1}{2}} \Theta \Pi) = K \cdot (\Xi B^{-1})'(\Xi B^{-1}) = K \cdot (BB')^{-1}$. It follows that

$$B' \Lambda B' = K^{-2} \cdot BB' G P G B B = P.$$

Write $\Lambda = \text{diag}(\lambda_1, \Lambda_1)$, where $\Lambda_1 = \text{diag}(\lambda_2, \ldots, \lambda_K)$. Also, recall that $B = \text{diag}(b_1)[1_K, V]$. We plug them into the above expression to get

$$P = \text{diag}(b_1)[1_K, V] \begin{bmatrix} \lambda_1 \\ \Lambda_1 \end{bmatrix} \begin{bmatrix} 1_K' \\ V' \end{bmatrix} \text{diag}(b_1).$$

It follows that $P(k, k) = b_1(k) \cdot [\lambda_1 + v_k' \Lambda_1 v_k] \cdot b_1(k)$. The identifiability condition (1.3) says that $P(k, k) = 1$. Therefore, $b_1(k) = 1/\sqrt{\lambda_1 + v_k' \Lambda_1 v_k}$.  

28
B Auxiliary lemmas on regularized graph Laplacian

Denote
\[S_1 = \{1 \leq i \leq n : \theta_i \geq \bar{\theta}\}, \quad S_2 = \{1 \leq i \leq n : \theta_i < \bar{\theta}\}.\] (B.1)

B.1 Properties of \(L_0\)

Recall that \(L_0 = H_0^{-1/2} \Omega H_0^{-1/2}\). We state the following three lemmas on its spectrum properties and also estimates of the degree regularization matrix \(H_0\).

**Lemma B.1.** Under the conditions of Theorem 3.1,
\[\lambda_1 > 0, \quad \lambda_1 \asymp K^{-1} \lambda_1(\Omega), \quad |\lambda_K| \asymp K^{-1} \lambda_K(\Omega), \quad \lambda_1 - \max_{2 \leq k \leq K} |\lambda_k| \geq c \lambda_1.\] (B.2)

**Lemma B.2.** Under the conditions of Theorem 3.1,
\[\xi_1(i) \asymp \frac{1}{\sqrt{n}} \begin{cases} \sqrt{\frac{\theta_i}{\bar{\theta}}}, & i \in S_1, \\ \theta_i/\bar{\theta}, & i \in S_2, \end{cases}, \quad \|\Xi(i)\| \leq C \frac{\sqrt{K}}{\sqrt{n}} \begin{cases} \sqrt{\frac{\theta_i}{\bar{\theta}}}, & i \in S_1, \\ \theta_i/\bar{\theta}, & i \in S_2, \end{cases},\] (B.3)

and
\[H_0(i,i) \asymp \begin{cases} n\theta_i\bar{\theta}, & i \in S_1, \\ n\theta^2, & i \in S_2. \end{cases}\] (B.4)

**Lemma B.3.** Under the conditions of Theorem 3.1 with probability \(1 - o(n^{-3})\),
\[\|I_n - H_0^{-1}H\| \leq \frac{C}{\sqrt{n} \bar{\theta}^2}, \quad \|H_0^{-1/2}(A - \Omega)H_0^{-1/2}\| \leq \frac{C}{\sqrt{n} \bar{\theta}^2}.\] (B.5)

The proof of Lemma B.1 is straightforward by noting that \(D_\theta = H_0^{-1}\) (see the definition of \(D_\theta\) in Section 3) and therefore \(L_0\) share the same eigenvalues as \(K^{-1}PG\). Immediately, one can conclude (B.2) from (3.2). In the sequel, we show the proof of the Lemmas B.2 and B.3. Before that, we introduce the Bernstein inequality which we will use frequently to bound sum of independent Bernoulli entries.

**Theorem B.1** (Bernstein inequality). Let \(X_1, \ldots, X_n\) be independent zero-mean random variables. Suppose that \(|X_i| \leq M\) almost surely, for all \(i\). Then for all positive \(t\),
\[\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq 2 \exp\left(-\frac{t^2/2}{\sigma^2 + Mt/3}\right),\]
with \(\sigma^2 := \sum_{i=1}^{n} \mathbb{E}(X_i^2)\). In particular, taking \(t = C(\sigma \sqrt{\log(n)} + M \log(n))\) for properly large \(C\), then
\[\left|\sum_{i=1}^{n} X_i\right| \leq C(\sigma \sqrt{\log(n)} + M \log(n)) \quad \text{with probability } 1 - o(n^{-5}).\]
Proof of Lemma B.2. First, we show (B.4). Uniformly for all $1 \leq i \leq n$,

$$\mathbb{E}d_i = \theta_i \sum_{j \neq i} \theta_j \pi_j' P \pi_i \geq \theta_i \sum_{j \neq i} \theta_j \sum_k \pi_j(k) \pi_i(k) \geq c_1 n \theta_i \bar{\theta}$$  \hspace{1cm} (B.6)

by the last inequality in (3.1). On the other hand, $\pi_j' P \pi_i \leq \max_{t,s} P(t, s)$ for all $i, j$, then $\mathbb{E}d_i = \theta_i \sum_{j \neq i} \theta_j \pi_j' P \pi_i \leq c n \theta_i \bar{\theta}$ for all $i$. As a result, $\mathbb{E}d_i \asymp n \theta_i \bar{\theta}$, and further

$$H_0(i, i) = \mathbb{E}d_i + \mathbb{E}\bar{d} \asymp \begin{cases} n \theta_i \bar{\theta}, & i \in S_1, \\ n \bar{\theta}^2, & i \in S_2. \end{cases}$$

This completes the proof of (B.4). Next, we turn to prove (B.3). By the definition $L_0 = H_0^{-\frac{1}{2}} \Omega H_0^{-\frac{1}{2}}$, there exists a non-singular matrix $B \in \mathbb{R}^{K \times K}$ satisfying

$$\Xi = H_0^{-\frac{1}{2}} \Theta B, \quad BB' = (\Pi' \Theta H_0^{-1} \Theta \Pi)^{-1}.$$ 

Using (3.1) and $H_0 = D_\theta$, one gets $\|BB'\| \leq Kc$, $\lambda_{\min}(BB') \geq Kc^{-1}$. Write $B = (b_1, \cdots, b_K)$. We have $Kc^{-1} \leq \|b_i\|^2 \leq Kc$ for $1 \leq i \leq K$. Taking the $i$-th row of $\Xi$,

$$\|\Xi(i)\| = \frac{\theta_i}{\sqrt{H_0(i, i)}} \|\pi_i' B\| \leq C \sqrt{\frac{\theta_i}{n \bar{\theta}}} \|\pi_i\| \|BB'\|^\frac{1}{2} \leq C \sqrt{\bar{K}} \sqrt{\frac{\theta_i}{n \bar{\theta}}}.$$ 

For the leading eigenvector $\xi_1$, we have

$$\xi_1(i) = \frac{\theta_i}{\sqrt{H_0(i, i)}} \pi_i' b_1$$

It follows from $L_0 \Xi = \Xi \Lambda$ that $H_0^{-\frac{1}{2}} \Theta \Pi \Pi' \Theta H_0^{-1} \Theta \Pi B = H_0^{-\frac{1}{2}} \Theta \Pi B \Lambda$, which implies that $\Pi \Pi' \Theta H_0^{-1} \Theta \Pi B = \Lambda$. As a consequence, $b_1$ is the first right singular vector of $\Pi \Pi' \Theta H_0^{-1} \Theta \Pi$, and equivalently, the first right singular vector of $PG$. Using (3.3), we easily conclude that $b_1(k) > 0, b_1(k) > 1$ for all $1 \leq k \leq K$. Then, $\pi_i' b_1 \approx 1$ for all $1 \leq i \leq n$, and the entrywise estimate of $\xi_1$ simply follows from (B.4). \hfill \Box

Proof of Lemma B.3. Recall the definition of $H_0, H$. We write

$$\frac{H(i, i)}{H_0(i, i)} - 1 = \frac{d_i - \mathbb{E}d_i + \bar{d} - \mathbb{E}\bar{d}}{H_0(i, i)}, \quad d_i - \mathbb{E}d_i = \sum_{j \neq i} A_{ij} - \mathbb{E}A_{ij}.$$ 

By (B.4), we easily see that $H_0(i, i) \asymp n \bar{\theta}(\theta_i \vee \bar{\theta})$. What remains is to estimate the numerator, or $d_i - \mathbb{E}d_i$ for all $1 \leq i \leq n$. This actually can be done with the help of Bernstein inequality. Applying the Bernstein inequality (Theorem B.1) to $d_i - \mathbb{E}d_i$, we see that

$$\mathbb{P}\left(\left|\sum_{j \neq i} A_{ij} - \mathbb{E}A_{ij}\right| \geq t\right) \leq 2 \exp\left(-\frac{1}{2} \frac{t^2}{\sum_{j \neq i} \text{var}A_{ij} + \frac{1}{3} Mt}\right)$$

30
where $M = \sup_j |A_{ij} - \mathbb{E}A_{ij}| \leq 2$. Moreover, we have the crude bound

$$\sum_{j \neq i} \text{var}A_{ij} \leq c n\theta_i \bar{\theta}.$$ 

Taking $t = C\sqrt{\log(n)/n\theta_i \vee \log(n)}$, it gives that $|\sum_{j \neq i} A_{ij} - \mathbb{E}A_{ij}| \leq C\sqrt{\log(n)(n\theta_i \vee \log(n))}$ with probability $1 - o(n^{-5})$. Consider all $i$’s together, one gets

$$P\left(\bigcup_{i=1}^{n} \left\{ \left| \sum_{j \neq i} A_{ij} - \mathbb{E}A_{ij} \right| \geq \sqrt{n\theta_i \bar{\theta} \vee \log(n)} \right\} \right) \leq cn^{-4},$$

This, combined with $H_0(i, i) \approx n\bar{\theta}(\theta_i \vee \bar{\theta})$, implies that

$$\left| 1 - H(i, i)/H_0(i, i) \right| \leq \frac{|d_i - \mathbb{E}d_i|}{H_0(i, i)} + \frac{1}{n} \sum_{j=1}^{n} \left| \frac{d_j - \mathbb{E}d_j}{H_0(i, i)} \right| \leq C\sqrt{\frac{\log n}{n\theta^2}} + C\sqrt{\frac{\log n}{n\theta^2} \sum_{j=1}^{n} \max \left\{ \sqrt{\frac{\theta_j}{\bar{\theta}}}, \sqrt{\frac{\log n}{n\theta^2}} \right\}},$$

$$\leq C\sqrt{\frac{\log n}{n\theta^2}} \left( 1 + \frac{1}{n} \sum_{j=1}^{n} \sqrt{\frac{\theta_j}{\bar{\theta}}} + \sqrt{\frac{\log n}{n\theta^2}} \right) \leq C\sqrt{\frac{\log n}{n\theta^2}}$$

with probability $1 - o(n^{-3})$ uniformly for all $1 \leq i \leq n$. Here in the last step, we used the Cauchy-Schwarz inequality $\sum_{j=1}^{n} \sqrt{\theta_j} \leq \sqrt{n} \sum_{j=1}^{n} \theta_j = n\sqrt{\bar{\theta}}$. This finished the first estimate of (B.5). Now, we proceed to the second estimate. We crudely bound $\|H_0^{-1/2}(A - \Omega)H_0^{-1/2}\| \leq \|H_0^{-1/2}WH_0^{-1/2}\| + \|H_0^{-1/2}\text{diag}(\Omega)H_0^{-1/2}\|$. First it is easy to get the bound

$$\|H_0^{-1/2}\text{diag}(\Omega)H_0^{-1/2}\| \leq C \max_{1 \leq i \leq n} \frac{\theta_i^2 \pi_i'P\pi_i}{H_0(i, i)} \leq \frac{C}{n\bar{\theta}} \leq \frac{C}{\sqrt{n\theta^2}}.$$ 

Next, we apply the non-asymptotic bounds for random matrices in [Bandeira and Van Handel (2016)] to bound the operator norm of $\hat{W} := H_0^{-1/2}WH_0^{-1/2}$. Note that $\hat{W}$ is a symmetric random matrix with independent upper triangular entries. Using Corollary 3.12 of [Bandeira and Van Handel (2016)] with Remark 3.13, we bound

$$P(\|\hat{W}\| \geq C\tilde{\sigma} + t) \leq ne^{-t^2/c\tilde{\sigma}^2}$$

for some constant $C, c > 0$, with

$$\tilde{\sigma} = \max_i \sqrt{\sum_j \mathbb{E}\hat{W}(i, j)^2} \leq 1/\sqrt{n\theta^2}, \quad \tilde{\sigma}_* = \max_{i,j} \|\hat{W}(i, j)\|_{\infty} \leq C/n\bar{\theta}^2.$$ 

Then one just need to take $t = c/\sqrt{n\theta^2}$ for properly large $c$ and use the assumption $n\theta^2 \gg \log(n)$. It follows that $\|\hat{W}\| \leq C/\sqrt{n\theta^2}$ with probability $1 - o(n^{-3})$. We thus complete the proof of Lemma B.3. \qed
B.2 Properties of $\tilde{L}^{(i)}$

In this section, for an arbitrary fixed index $i$ and the intermediate matrix $\tilde{L}^{(i)}$, we collect the spectrum properties of $\tilde{L}^{(i)}$ and estimate on $\tilde{H}^{(i)}$ in the lemmas below. Let $E$ be the event that Lemma $[B.3]$ holds.

**Lemma B.4.** Under the conditions in Theorem $[3.1]$ Over the event $E$, for any fixed $1 \leq i \leq n$, the eigenvalues $\tilde{\lambda}_1^{(i)}, \ldots, \tilde{\lambda}_K^{(i)}$ of $\tilde{L}^{(i)}$ satisfy

$$
\tilde{\lambda}_1^{(i)} > 0, \quad \tilde{\lambda}_1^{(i)} \approx K^{-1} \lambda_1(PG), \quad [\tilde{\lambda}_K^{(i)}] \approx K^{-1} |\lambda_K(PG)|, \quad \tilde{\lambda}_1^{(i)} - \max_{2 \leq k \leq K} |\tilde{\lambda}_k^{(i)}| \geq C^{-1} \tilde{\lambda}_1^{(i)};
$$

and for the associated eigenvectors,

$$\xi^{(i)}_1(j) \geq \frac{1}{\sqrt{n}} \left\{ \sqrt{\theta_j/\bar{\theta}}, \quad j \in S_1, \right\} \quad \|\tilde{\xi}^{(i)}(j)\| \leq \frac{C \sqrt{K}}{\sqrt{n}} \left\{ \sqrt{\theta_j/\bar{\theta}}, \quad j \in S_1, \right\} \quad \|\tilde{\xi}^{(i)}(j)\| \leq \frac{C \sqrt{K}}{\sqrt{n}} \left\{ \sqrt{\theta_j/\bar{\theta}}, \quad j \in S_2. \right\} \quad \text{(B.7)}
$$

**Lemma B.5.** Under the conditions of Theorem $[3.1]$ Over the event $E$, for any fixed $1 \leq i \leq n$ and $\tilde{H}^{(i)}$,

$$
\|I_n - H_0^{-1} \tilde{H}^{(i)}\| \leq \frac{C \log(n)}{\sqrt{n \bar{\theta}^2}}, \quad \|I_n - (\tilde{H}^{(i)})^{-1}\| \leq \frac{C \log(n)}{\sqrt{n \bar{\theta}^2}}. \quad \text{(B.9)}
$$

Besides the above lemma, by Theorem $[B.1]$ and after elementary computations, we also have that for each $1 \leq j \leq n, j \neq i$, over the event $E$,

$$
|\tilde{H}^{(i)}(j, j) - H(j, j)| = \left| - A(j, i) - \frac{1}{n} \left( \sum_{s \neq i} A(i, s) + \sum_{s \neq i} A(i, s) - \mathbb{E} A(i, s) \right) \right|
$$

$$
\leq A(j, i) + C \theta_i \bar{\theta} + \frac{C \log(n)}{n}; \quad \text{(B.10)}
$$

and

$$
|\tilde{H}^{(i)}(i, i) - H(i, i)| \leq C \sqrt{n \theta_i \log n + C \log(n)}. \quad \text{(B.11)}
$$

Applying (B.10) and (B.11) with Lemma $[B.3]$, it is easy to deduce the estimates in Lemma $[B.5]$. To show the eigen-properties of $\tilde{L}$ in Lemma $[B.4]$ one only need to rely on the estimate

$$
\|\tilde{L}^{(i)} - L\| \approx \|I_n - H_0^{-1} \tilde{H}^{(i)}\| |\lambda_1(L_0)| \leq C \sqrt{\frac{\log(n)}{n \bar{\theta}^2}} \gg |\lambda_K|
$$

under the assumption of Theorem $[3.1]$ then (B.7) can be derived simply by further applying Lemma $[B.1]$. Moreover, (B.8) follows from Lemmas $[4.1, C.1]$ and $[B.2]$. Thereby, we omit the proofs of Lemmas $[B.4, B.5]$. We comment here that the proof of Lemmas $[4.1, C.1]$ only depends on the lemmas in Section $[B.1]$, i.e., the properties of $L_0$, not the properties of $\tilde{L}^{(i)}$. There is no circular logic for the lemmas presenting in this subsection.
C Entrywise eigenvector analysis

Here we show the complete proof of Theorem 3.1 and a brief proof of Corollary 3.1 in our manuscript. In Sections C.1-C.3, we state the proofs of key lemmas for proving (3.6), while the claim of (3.6) is already presented in the manuscript. Section C.4 collects the proof of the second claim in Theorem 3.1 (i.e., (3.7)) which provides the entry-wise estimates for the 2- to $K$-th eigenvectors. Similarly to the proof of the first claim in Theorem 3.1 (i.e., (3.6)), we introduce three key lemmas, Lemmas C.1-C.3, counterpart to Lemmas 4.1-4.3. The proofs of Lemmas C.1-C.3 are provided correspondingly in Section C.5-C.7. In the end of this section, we briefly state the proof of Corollary 3.1 based on Theorem 3.1.

C.1 Proof of Lemma 4.1

In this subsection, we show the proof of Lemma 4.1 using the eigen-properties of $L_0$ in Section B.1.

Fix the index $i$, we study the perturbation from $L_0 = H_0^{-1/2} \Omega H_0^{-1/2}$ to $\tilde{L}^{(i)} = (\tilde{H}^{(i)})^{-1/2} \Omega (\tilde{H}^{(i)})^{-1/2}$. By definition,

$$L_0 = H_0^{-1/2} \Omega H_0^{-1/2} = \sum_{k=1}^K \lambda_k \xi_k \xi_k^t, \quad (\tilde{H}^{(i)})^{-1/2} \Omega (\tilde{H}^{(i)})^{-1/2} \tilde{\xi}_1^{(i)} = \tilde{\lambda}_1^{(i)} \tilde{\xi}_1^{(i)}.$$  

Write $\tilde{Y} \equiv \tilde{Y}^{(i)} := H_0^{1/2} (\tilde{H}^{(i)})^{-1/2}$. Then, we have

$$\tilde{Y} \left( \sum_{k=1}^K \lambda_k \xi_k \xi_k^t \right) \tilde{Y}^{(i)} = \tilde{\lambda}_1^{(i)} \tilde{\xi}_1^{(i)}.$$  

It follows that, for each $1 \leq j \leq n$,

$$\frac{1}{\tilde{Y}(j,j)} \tilde{\xi}_1^{(i)}(j) = \frac{\lambda_1(\xi_1^t \tilde{Y}^{(i)} \tilde{\xi}_1^{(i)})}{\tilde{\lambda}_1^{(i)}} \xi_1(j) + \sum_{k=2}^K \frac{\lambda_k(\xi_1^t \tilde{Y}^{(i)} \tilde{\xi}_1^{(i)})}{\tilde{\lambda}_1^{(i)}} \xi_k(j). \quad (C.1)$$

As a result,

$$|\tilde{\xi}_1^{(i)}(j) - \xi_1(j)| \leq \left| \frac{1}{\tilde{Y}(j,j)} - 1 \right| |\tilde{\xi}_1^{(i)}(j)| + \frac{|\lambda_1(\xi_1^t \tilde{Y}^{(i)} \tilde{\xi}_1^{(i)})|}{\tilde{\lambda}_1^{(i)}} - 1 \left| |\xi_1(j)| + \sum_{k=2}^K \frac{|\lambda_k(\xi_1^t \tilde{Y}^{(i)} \tilde{\xi}_1^{(i)})|}{\tilde{\lambda}_1^{(i)}} \right| \xi_k(j). \quad (C.2)$$

By Lemma B.1, $\|L_0\| \leq CK^{-1}\lambda_1(PG)$. And using the first estimate in (B.9), it is easy to conclude that

$$\|\tilde{Y} - I_n\| = \|I_n - (H_0^{-1/2} \tilde{H}^{(i)})^{-1/2}\| \leq C\sqrt{\log(n)} \sqrt{n\theta^2}, \quad (C.3)$$

over the event $E$ where Lemma B.3 holds. As a result, we have $\|\tilde{L}^{(i)} - L_0\| \leq CK^{-1}\lambda_1(PG)\|\tilde{Y} - I_n\|$. Using Weyl’s inequality, we then see that

$$\max_{1 \leq k \leq K} |\tilde{\lambda}_1^{(i)} - \lambda_k| \leq \|\tilde{L}^{(i)} - L_0\| \leq CK^{-1}\lambda_1(PG)\|\tilde{Y} - I_n\| \leq C\|\tilde{Y} - I_n\|$$

33
since $\lambda_1(PG) \leq CK$ under our model assumption. Furthermore, by Lemma [B.1] the eigen-gap between the largest eigenvalue and the other nonzero eigenvalues of $L_0$ is at the order $K^{-1}\lambda_1(PG)$. Hence, the eigengap between $\lambda_1$ and $\tilde{\lambda}_{2}^{(i)}, \cdots, \tilde{\lambda}_{K}^{(i)}$ is still of the order $K^{-1}\lambda_1(PG)$. It follows from the sin-theta theorem that

$$|\xi_1^{(i)} \tilde{\gamma}_1^{(i)} - 1| \leq |\xi_1^{(i)} \tilde{\gamma}_1^{(i)} - 1| + \|\xi_1^{(i)} (\tilde{Y} - I_n) \tilde{\xi}_1^{(i)}\|$$

$$\leq C(K\lambda_1^{-1}(PG)\|\tilde{L}^{(i)} - L_0\|)^2 + \|\tilde{Y} - I_n\| \leq C\|\tilde{Y} - I_n\|.$$ 

Here $\text{sgn}(\xi_1^{(i)} \tilde{\gamma}_1^{(i)}) = 1$ since we fix our choices of $\xi_1, \tilde{\xi}_1^{(i)}$ with positive first components and they are both from the positive matrices. Then this will be claimed by Perron’s theorem.

Using Cauchy-Schwarz inequality, we bound $\sum_{k=2}^{K} |\xi_k^{(i)} \tilde{Y} \tilde{\xi}_1^{(i)}| |\xi_k(j)| \leq \|\Xi_{1}^{(i)} \tilde{\xi}_1^{(i)}\| \|\Xi_1(j)\|$. And by sine-theta theorem,

$$\|\Xi_{1}^{(i)} \tilde{\gamma}_1^{(i)}\| \leq \|(\tilde{\xi}_1^{(i)})^\prime \Xi_1\| + \|\tilde{Y} - I_n\|$$

$$\leq C\left(K\lambda_1^{-1}(PG)\|\tilde{L}^{(i)} - L_0\| + \|\tilde{Y} - I_n\|\right)$$

$$\leq C\|\tilde{Y} - I_n\|.$$ 

Plugging the above estimates, we have

$$|\tilde{\xi}_1^{(i)}(j) - \xi_1(j)| \leq C\|\tilde{Y} - I_n\| |\tilde{\xi}_1^{(i)}(j)| + C\|\tilde{Y} - I_n\| \|\Xi(j)\|$$

$$\leq C\|\tilde{Y} - I_n\| |\tilde{\xi}_1^{(i)}(j) - \xi_1(j)| + C\|\tilde{Y} - I_n\| \|\Xi(j)\|.$$ 

Since $\|\tilde{Y} - I_n\| = o(1)$ over the event $E$, rearranging the terms gives

$$|\tilde{\xi}_1^{(i)}(j) - \xi_1(j)| \leq C\|\tilde{Y} - I_n\| \|\Xi(j)\|, \quad \text{for all } 1 \leq t \leq n.$$ (C.4)

We plug (C.3) into (C.4) and use the bound for $\|\Xi(j)\|$ in (B.3). It follows that over the event $E$, for all $1 \leq j \leq n$,

$$|\tilde{\xi}_1^{(i)}(j) - \xi_1(j)| \leq C\sqrt{K} \sqrt{\frac{\log(n)}{n\theta^2}} \sqrt{\frac{\theta_j}{n\bar{\theta}}} \left(\sqrt{\frac{\theta_j}{\bar{\theta}}} \wedge 1\right).$$ (C.5)

Then, consider all $i$’s together, we conclude (4.3) with probability $1 - o(n^{-3})$ simultaneously for all $1 \leq i, j \leq n$.

### C.2 Proof of Lemma 4.2

In this subsection, we state the proof of Lemma 4.2 which heavily relies on the eigen-properties of $\tilde{L}^{(i)}$ in Section B.2.

Fix the index $i$, we first show (4.5) which is based on the decomposition

$$w \hat{\xi}_1 = \tilde{\xi}_1^{(i)} + (\tilde{\xi}_1^{(i)} - \xi_1^{(i)}) + (w \hat{\xi}_1 - \tilde{\xi}_1^{(i)})$$

34
where \( w = \text{sgn}(\xi_1^i \dot{\xi}_1) \) will be claimed later. It is not hard to derive
\[
|e_i \Delta (w \dot{\xi}_1 - \bar{\xi}_1^i)| \leq \Delta \|w \dot{\xi}_1 - \bar{\xi}_1^i\| \leq \|H_0^{1/2} W H_0^{-1/2} H^{1/2} H^{-1/2}\| \|H_0^{1/2} (\bar{\bar{H}}(i))^{-1/2}\| \|w \dot{\xi}_1 - \bar{\xi}_1^i\| \\
\leq \frac{C}{\sqrt{n \theta^2}} \|w \dot{\xi}_1 - \bar{\xi}_1^i\|
\]
over the event \( E \), in light of Lemmas \( \text{B.3} \) and \( \text{B.5} \). We thus end up with \( (4.5) \). We now turn to prove \( (4.4) \). We study the perturbation from \( \bar{L}(i) = (\bar{H}(i))^{-1/2} \Omega (\bar{H}(i))^{-1/2} \) to \( L = H^{-1/2} A H^{-1/2} \). Write \( X \equiv X(i) := (\bar{H}(i))^{1/2} H^{-1/2} \). We can rewrite
\[
L = X (\bar{H}(i))^{-1/2} A (\bar{H}(i))^{-1/2} X = X \bar{L}(i) X - X (\bar{H}(i))^{-1/2} \text{diag} (\Omega) (\bar{H}(i))^{-1/2} X + X \Delta
\]
with \( \Delta = (\bar{H}(i))^{-1/2} W H^{-1/2} \). By definition, \( \bar{L}(i) = (\bar{H}(i))^{-1/2} \Omega (\bar{H}(i))^{-1/2} = \sum_{k=1}^{K} \bar{\lambda}_k^i \bar{\xi}_k^i (\bar{\xi}_k^i)' \) and \( H^{-1/2} A H^{-1/2} \bar{\xi}_1 = \bar{\lambda}_1 \bar{\xi}_1 \). It follows that
\[
\sum_{k=1}^{K} \bar{\lambda}_k^i (\bar{\xi}_k^i X \bar{\xi}_k^i) - X (\bar{H}(i))^{-1/2} \text{diag} (\Omega) (\bar{H}(i))^{-1/2} X \bar{\xi}_1 + X \Delta \bar{\xi}_1 = \bar{\lambda}_1 \bar{\xi}_1.
\]
As a result,
\[
\hat{\xi}_1(i) = \frac{\bar{\lambda}_1^i (\bar{\xi}_1^i X \bar{\xi}_1^i)}{\bar{\lambda}_1} X(i, i) \bar{\xi}_1^i(i) + \sum_{k=2}^{K} \frac{\bar{\lambda}_k^i (\bar{\xi}_k^i X \bar{\xi}_k^i)}{\bar{\lambda}_1} X(i, i) \bar{\xi}_k^i(i) \\
+ \frac{X^2(i, i) \Omega(i, i)}{\bar{\lambda}_1 \bar{H}(i)(i, i)} \bar{\xi}_1^i(i) + \frac{X(i, i)}{\bar{\lambda}_1} e_i \Delta \bar{\xi}_1(i).
\]
By Lemma \( \text{B.5} \) it is easy to deduce that
\[
\|X - I_n\| \leq C \frac{\sqrt{\log n}}{\sqrt{n \theta^2}}.
\]
Since \( \text{B.7} \), by Weyl’s inequality,
\[
\max_{1 \leq k \leq K} \{ |\bar{\lambda}_k - \bar{\lambda}_k^i| \} \leq C \|L - \bar{L}(i)\|
\]
and over the event \( E \),
\[
\|L - \bar{L}(i)\| \leq C \|X - I_n\| H^{-1/2} A H^{-1/2} + \|H^{-1/2} H_0^{-1/2} A - \Omega H_0^{-1/2}\| \\
\leq C \|X - I_n\| H_0^{-1/2} \Omega H_0^{-1/2} + \|H^{-1/2} H_0^{-1/2} A - \Omega H_0^{-1/2}\| \\
\leq C \frac{K^{-1} \lambda_1(PG) \sqrt{\log(n)}}{\sqrt{n \theta^2}} + \frac{1}{\sqrt{n \theta^2}} \ll |\bar{\lambda}_k^i|
\]
since Lemmas \( \text{B.1}, \text{B.3} \) and \( \text{B.7} \) with the condition \( K \beta_n^{-1} \sqrt{\log(n)} / \sqrt{n \theta^2} \ll 1 \). Therefore, \( \bar{\lambda}_1, \ldots, \bar{\lambda}_K \) share the same asymptotics as \( \bar{\lambda}_1^i, \ldots, \bar{\lambda}_K^i \). The eigengap between \( \bar{\lambda}_1 \) and \( \bar{\lambda}_2 \), \ldots, \( \bar{\lambda}_K \) is \( K^{-1} \lambda_1(PG) \). It follows from the sin-theta theorem that
\[
|\xi_1^i X \bar{\xi}_1^i - w(i)| \leq \|X - I_n\| + C \left( K \lambda_1^{-1}(PG) \|L - \bar{L}(i)\| \right)^2
\]
(8)
for some $w(i) \in \{1, -1\}$. In particular,
\[
\|\hat{\xi}_i X(\xi_2^{(i)}, \ldots, \xi_K^{(i)})\| \leq \|X - I_n\| + CK\lambda_1^{-1}(PG)\|L - \tilde{L}^{(i)}\|. \tag{C.9}
\]
We can actually claim that $w(i) \equiv w := \text{sgn}(\xi_1^{(i)} \hat{\xi}_1)$. To see this, we derive $|\hat{\xi}_1\xi_1 + w(i)| > 
2 - |\hat{\xi}_1\xi_1 - w(i) > 2 - |\hat{\xi}_1\xi_1^{(i)} - w(i)| - \|\xi_1^{(i)} - \xi_1\| = 2 - o(1)$ for all $1 \leq i \leq n$; in addition, $|\hat{\xi}_1\xi_1 - w| \leq K\|L - L_0\| = o(1)$. Then by these two inequalities, it is easy to obtain the
claim. In the sequel, we directly write $w$ instead of $w(i)$. We can further derive
\[
\left| \frac{X(i,i)^2\Omega(i,i)}{\hat{\lambda}_1\hat{H}^{(i)}(i,i)} \xi_1(i) \right| \leq \frac{CK\theta^2}{n\theta(\theta \vee \theta_i)} \frac{\kappa_i}{\hat{\lambda}_1(PG)\sqrt{\log(n)}} \leq \frac{CK\lambda_1^{-1}(PG)}{\sqrt{\log(n)}} \left(\sqrt{\frac{\theta}{\theta_i}} \wedge 1\right),
\]
by the estimate $\tilde{H}^{(i)}(i,i) \asymp n\theta(\theta \vee \theta_i)$ following from \[C.4\] and the first estimate in \[B.9\],
with $\kappa_i = \frac{\sqrt{\log(n)}\theta}{n\theta_i} \cdot \frac{\sqrt{\theta}}{\sqrt{\theta_i}}$. Then, plugging the above bound and \[C.7\]-\[C.9\] into \[C.6\] gives
\[
|w\hat{\xi}_1(i) - \xi_1^{(i)}(i)| \leq C\kappa_i \left(\sqrt{\frac{\theta_i}{\theta}} \wedge 1\right) \left(\sqrt{K} + \frac{K}{\hat{\lambda}_1(PG)\sqrt{\log(n)}}\right) + CK\lambda_1^{-1}(PG)\|L - \tilde{L}^{(i)}\| \cdot \|\tilde{\Xi}^{(i)}(i)\| + CK\lambda_1^{-1}(PG)|e_i^1\Delta\hat{\xi}_1|
\leq C\kappa_i \left(\sqrt{\frac{\theta_i}{\theta}} \wedge 1\right) \left(\sqrt{K} + K^{\frac{1}{2}}\lambda_1^{-1}(PG)/\sqrt{\log(n)}\right) + CK\lambda_1^{-1}(PG)|e_i^1\Delta\hat{\xi}_1| \tag{C.10}
\]
where in the last step, we plugged in \[C.7\] and the bound of $\|\tilde{\Xi}^{(i)}(i)\|$ in \[B.8\]. Since the
assumption $\lambda_1(PG) \geq \alpha_n$, it gives that over the event $E$,
\[
|w\hat{\xi}_1(i) - \xi_1^{(i)}(i)| \leq C\sqrt{K}\kappa_i \left(\sqrt{\frac{\theta_i}{\theta}} \wedge 1\right) \left(1 + \frac{K}{\alpha_n\sqrt{\log(n)}}\right) + CK\alpha_1^{-1}|e_i^1\Delta\hat{\xi}_1|.
\leq CK^{\frac{1}{2}}\alpha_1^{-1}\kappa_i + CK\alpha_1^{-1}|e_i^1\Delta\hat{\xi}_1|.
\]
This concludes our proof by considering all $i$'s together.

### C.3 Proof of Lemma 4.3

In this section, we prove Lemma 4.3. We separate the proofs into three parts corresponding to the three estimates \[(4.6)\]-\[(4.8)\].

#### C.3.1 Proof of \[(4.6)\]

For any fixed $i$, recall that $X \equiv X(i) := (\tilde{H}^{(i)})^{1/2}H^{-1/2}$. We rewrite $\Delta \equiv \Delta(i) = (\tilde{H}^{(i)})^{-\frac{1}{2}}W(\tilde{H}^{(i)})^{-\frac{1}{2}}X$. It follows that
\[
e_i^1\Delta\tilde{\xi}_1 = \frac{W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}X\tilde{\xi}_1^{(i)}}{\sqrt{\tilde{H}^{(i)}(i,i)}} = \frac{W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{\xi}_1^{(i)}}{\sqrt{\tilde{H}^{(i)}(i,i)}} + \frac{W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(X - I_n)\tilde{\xi}_1^{(i)}}{\sqrt{\tilde{H}^{(i)}(i,i)}} \tag{C.11}
\]
First, we study the term \( |W(i)(\tilde{H}^{(i)})^{-1/2}\tilde{\xi}_1^{(i)}| \). Write

\[
W(i)(\tilde{H}^{(i)})^{-1/2}\tilde{\xi}_1^{(i)} = \sum_{1 \leq j \leq n; j \neq i} \frac{W(i, j)}{\sqrt{\tilde{H}^{(i)}(j, j)}} \tilde{\xi}_1^{(i)}(j).
\]

In the sequel, we only consider the randomness of \( W(i) \). Note that the mean is 0. The variance is bounded by (up to some constant \( C \))

\[
\sum_{j \neq i} \frac{\theta_i \theta_j}{n \theta_j} (\tilde{\xi}_1^{(i)}(j))^2 \leq \sum_{j} \frac{\theta_i \theta_j}{n \theta_j} (\tilde{\xi}_1^{(i)}(j))^2 \leq \frac{\theta_i}{n \theta}.
\]

Recall the definition of index sets \( S_1, S_2 \) in (B.1). Each term in the sum is bounded by

\[
\frac{|\tilde{\xi}_1^{(i)}(j)|}{\sqrt{\tilde{H}^{(i)}(j, j)}} \leq \frac{C}{n \theta} \begin{cases} 1, & j \in S_1, \\ \theta_j / \bar{\theta}, & j \in S_2. \end{cases}
\]

Applying Theorem B.1 one see that over the event \( E \),

\[
|W(i)(\tilde{H}^{(i)})^{-1/2}\tilde{\xi}_1^{(i)}| \leq C \sqrt{\frac{\theta_i \log n}{n \theta}} + C \frac{\log n}{n \theta}
\]

Hence, over the event \( E \),

\[
\frac{W(i)(\tilde{H}^{(i)})^{-1/2}\tilde{\xi}_1^{(i)}}{\sqrt{\tilde{H}^{(i)}(i, i)}} \leq C \sqrt{\frac{\log(n)}{n \theta}} \left(1 + \frac{\sqrt{\theta_i}}{\sqrt{\theta}} + \sqrt{\frac{\log(n)}{n \theta^2}} \right) \leq C \tilde{\kappa}_i \tag{C.12}
\]

by using the estimate \( \tilde{H}^{(i)}(i, i) \approx n \theta (\bar{\theta} \vee \theta) \) and the definition of \( \tilde{\kappa}_i \) in (4.6). Next, we study the term \( |W(i)(\tilde{H}^{(i)})^{-1/2}(X - I_n)\tilde{\xi}_1^{(i)}| \). Over the event \( E \), by (B.10) and (B.11), we have

\[
|X(j, j) - 1| \leq C \frac{|H(j, j) - \tilde{H}^{(i)}(j, j)|}{\tilde{H}^{(i)}(j, j)} \leq C \frac{A(i, j) + \theta_i \bar{\theta} + \log(n)/n}{\tilde{H}^{(i)}(j, j)}, \quad j \neq i; \tag{C.13}
\]

It follows that

\[
|W(i)(\tilde{H}^{(i)})^{-1/2}(X - I_n)\tilde{\xi}_1^{(i)}| = \left| \sum_{1 \leq j \leq n; j \neq i} W(i, j) \frac{|X(j, j) - 1|^2 \tilde{\xi}_1^{(i)}(j)}{\sqrt{\tilde{H}^{(i)}(j, j)}} \right|
\]

\[
\leq \sum_{1 \leq j \leq n; j \neq i} |W(i, j)| \frac{|X(j, j) - 1| \tilde{\xi}_1^{(i)}(j)}{\sqrt{\tilde{H}^{(i)}(j, j)}}
\]

\[
\leq C \sum_{1 \leq j \leq n; j \neq i} (A(i, j) + \theta_i \bar{\theta} + \log(n)/n) \frac{|\tilde{\xi}_1^{(i)}(j)|}{[\tilde{H}^{(i)}(j, j)]^{3/2}},
\]

37
where in the last line we have used the fact that \(|W(i, j)| ≤ 1\). We shall apply Bernstein’s inequality. The mean is bounded by (up to some constant \(C\))

\[
\sum_{j \neq i} \frac{\theta_i \bar{\theta}_j + \theta_i \bar{\theta} + \log(n)/n}{(n\bar{\theta}(\bar{\theta}_j \lor \bar{\theta}))^{3/2}} \cdot \sqrt{\theta_j} \leq \sum_{j} \frac{\theta_i + \log(n)/(n\bar{\theta})}{(n\bar{\theta})^2} \leq \frac{\theta_i}{n\bar{\theta}^2} \left(1 + \frac{\log(n)}{n\bar{\theta}_i}\right).
\]

The variance is bounded by (up to some constant \(C\))

\[
\sum_{j \neq i} \frac{\theta_i \theta_j}{(n\bar{\theta}(\bar{\theta}_j \lor \bar{\theta}))^{3}} (\xi_1^{(i)}(j))^2 \leq \frac{1}{(n\bar{\theta})^2} \sum_{j} \frac{\theta_i \theta_j}{n\theta_j \bar{\theta}} (\xi_1^{(i)}(j))^2 \leq \frac{1}{(n\bar{\theta})^2} \cdot \frac{\theta_i}{n\bar{\theta}}.
\]

Each individual term is bounded by (up to some constant \(C\))

\[
\frac{|\tilde{\xi}_1^{(i)}(j)|}{[H(i,j)]^{3/2}} \leq C \frac{1}{n\bar{\theta}} \begin{cases} 1/(n\theta_j \bar{\theta}), & j \in S_1, \\ \theta_j/(n\bar{\theta}^3), & j \in S_2, \end{cases} \leq \frac{C}{n\bar{\theta}} \cdot \frac{1}{n\bar{\theta}^2}.
\]

We then have

\[
|W(i)(\bar{H}^{(i)})^{-\frac{1}{2}}(X - I_n)\tilde{\xi}_1^{(i)}| \leq C \frac{\theta_i}{n\bar{\theta}^2} \left(1 + \frac{\log(n)}{n\bar{\theta}_i}\right)
\]

As a result, over the event \(E\),

\[
\frac{W(i)(\bar{H}^{(i)})^{-\frac{1}{2}}(X - I_n)\tilde{\xi}_1^{(i)}}{\sqrt{H(i,i)}} \leq C \frac{\theta_i}{n\bar{\theta}^2} \left(1 + \frac{\theta_i}{\sqrt{n\bar{\theta}}} + \frac{\log(n)}{n\theta_i} \sqrt{n\bar{\theta}_i} \sqrt{\log(n)}\right) \leq C \frac{\tilde{\kappa}_i}{\sqrt{n\bar{\theta}^2 \log(n)}}
\]

(C.14)

We plug (C.12) and (C.14) into (C.11), and consider all \(i\)’s over the event \(E\), then we conclude the proof of (4.6).

### C.3.2 Proof of (4.7)

Similarly to (C.11), we have

\[
|e^i \Delta(\tilde{\xi}_1^{(i)} - \bar{\xi}_1^{(i)})| \leq \frac{|W(i)(\bar{H}^{(i)})^{-1/2}(\bar{\xi}_1^{(i)} - \bar{\xi}_1^{(i)})|}{\sqrt{H(i,i)}} + \frac{C|W(i)(\bar{H}^{(i)})^{-1/2}(X - I_n)(\bar{\xi}_1^{(i)} - \bar{\xi}_1^{(i)})|}{\sqrt{H(i,i)}}
\]

(C.15)

We first study the term \(|W(i)(\bar{H}^{(i)})^{-1/2}(\bar{\xi}_1^{(i)} - \bar{\xi}_1^{(i)})|\). Write

\[
W(i)(\bar{H}^{(i)})^{-1/2}(\bar{\xi}_1^{(i)} - \bar{\xi}_1^{(i)}) = \sum_{1 \leq j \leq n; j \neq i} \frac{W(i,j)(\xi_1^{(i)}(j) - \bar{\xi}_1^{(i)}(j))}{\sqrt{H(i,j)}}.
\]
We shall apply Bernstein’s inequality since \((\tilde{H}^{(i)})^{-1/2}(\tilde{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})\) is independent of \(W(i)\). The variance is bounded by (up to some constant)
\[
\sum_{j \neq i} \frac{\theta_i \theta_j}{n \theta_j \wedge \theta} \left( \tilde{\xi}_1^{(i)}(j) - \tilde{\xi}_1^{(i)}(j) \right)^2 \leq \sum_{j} \frac{\theta_i \theta_j}{n \theta_j \theta} \left( \tilde{\xi}_1^{(i)}(j) - \tilde{\xi}_1^{(i)}(j) \right)^2 \leq \frac{2\theta_i}{n \theta}
\]
Each individual term is bounded by \(\| (\tilde{H}^{(i)})^{-1/2}(\tilde{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})\|_\infty\). As a result,
\[
|W(i)(\tilde{H}^{(i)})^{-1/2}(\tilde{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})| \\
\leq C \frac{\sqrt{\theta_i \log(n)}}{\sqrt{n \theta i}} + C \log(n) \| (\tilde{H}^{(i)})^{-1/2}(\tilde{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})\|_\infty \\
\leq C \frac{\sqrt{\theta_i \log(n)}}{\sqrt{n \theta i}} + C \log(n) \| (\tilde{H}^{(i)})^{-1/2}(w\tilde{\xi}_1 - \tilde{\xi}_1^{(i)})\|_\infty + C \log(n) \| (\tilde{H}^{(i)})^{-1/2}(\tilde{\xi}_1^{(i)} - w\tilde{\xi}_1)\| \\
\leq C \frac{\sqrt{\theta_i \log(n)}}{\sqrt{n \theta i}} + C \log(n) \| (\tilde{H}^{(i)})^{-1/2}(w\tilde{\xi}_1 - \tilde{\xi}_1^{(i)})\|_\infty + \frac{C \log(n)}{\sqrt{n \theta i^2}} \| \tilde{\xi}_1^{(i)} - w\tilde{\xi}_1 \|.
\]
Further,
\[
\frac{|W(i)(\tilde{H}^{(i)})^{-1/2}(\tilde{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})|}{\sqrt{\tilde{H}^{(i)}(i, i)}} \leq C \kappa_i + C \sqrt{n \theta \log(n)} \| (\tilde{H}^{(i)})^{-1/2}(w\tilde{\xi}_1 - \tilde{\xi}_1^{(i)})\|_\infty + \frac{C \log(n)}{\sqrt{n \theta i^2}} \| \tilde{\xi}_1^{(i)} - w\tilde{\xi}_1 \|
\]
by the definition of \(\kappa_i = \frac{1}{\theta_i} \sqrt{\frac{\log(n)}{n \theta i} \sqrt{n \theta i} \vee \log(n)}\).

We then study the term \(|W(i)(\tilde{H}^{(i)})^{-1/2}(X - I_n)(\tilde{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})|\). On the event \(E\), recall (C.13). It follows that
\[
|W(i)(\tilde{H}^{(i)})^{-1/2}(X - I_n)(\tilde{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})| = \left| \sum_{1 \leq j \leq n; j \neq i} W(i, j) \left[ \frac{\left| X(j, j) \right| - 1}{\tilde{H}^{(i)}(j, j)} \right] \right| \\
\leq \sum_{1 \leq j \leq n; j \neq i} |W(i, j)| \left| \frac{\left| X(j, j) \right| - 1}{\tilde{H}^{(i)}(j, j)} \right| \\
\leq C \sum_{1 \leq j \leq n; j \neq i} \frac{A(i, j) + \theta_i \bar{\theta} + \log(n) / n}{\tilde{H}^{(i)}(j, j)} \left| \frac{\tilde{\xi}_1^{(i)}(j) - \tilde{\xi}_1^{(i)}(j)}{\sqrt{\tilde{H}^{(i)}(j, j)}} \right|.
\]
We now decompose the RHS above as
\[
\sum_{1 \leq j \leq n; j \neq i} \frac{A(i, j) + \theta_i \bar{\theta} + \log(n) / n}{\tilde{H}^{(i)}(j, j)} \left| \tilde{\xi}_1^{(i)}(j) - w\tilde{\xi}_1(j) \right| + \sum_{1 \leq j \leq n; j \neq i} \frac{A(i, j) + \theta_i \bar{\theta} + \log(n) / n}{\tilde{H}^{(i)}(j, j)} \left| w\tilde{\xi}_1(j) - \tilde{\xi}_1^{(i)}(j) \right|
\]
and we bound the sub-terms separately as below.

\[
\sum_{1 \leq j \leq n: j \neq i} \frac{A(i, j) + \theta \bar{\theta} + \frac{\log(n)}{n}}{H(i, j)^3} \left| \tilde{\xi}_1^{(i)}(j) - w\hat{\xi}_1(j) \right| \leq \left( \sum_{1 \leq j \leq n: j \neq i} \frac{(A(i, j) + \theta \bar{\theta} + \frac{\log(n)}{n})^2}{H(i, j)^3} \right)^{\frac{1}{2}} \| \tilde{\xi}_1^{(i)} - w\hat{\xi}_1 \|
\]

by Cauchy-Schwarz inequality, and

\[
\sum_{1 \leq j \leq n: j \neq i} \frac{A(i, j) + \theta \bar{\theta} + \frac{\log(n)}{n}}{H(i, j)^3} \left| w\hat{\xi}_1(j) - \tilde{\xi}_1^{(i)}(j) \right| \leq \sum_{1 \leq j \leq n: j \neq i} \frac{A(i, j) + \theta \bar{\theta} + \frac{\log(n)}{n}}{H(i, j)^3} \| (\tilde{H}(i)^{-\frac{1}{2}}(\xi_1^{(i)} - w\hat{\xi}_1)) \|_\infty.
\]

Applying Theorem B.1 we can have the estimates

\[
\left( \sum_{1 \leq j \leq n: j \neq i} \frac{(A(i, j) + \theta \bar{\theta} + \frac{\log(n)}{n})^2}{H(i, j)^3} \right)^{\frac{1}{2}} \leq C \left( \frac{n \bar{\theta} \theta_i + \log n}{(n \theta^2)^3} \right)^{\frac{1}{2}} \leq C \left( \frac{1}{n \theta^2} \sqrt{\frac{\theta_i}{\bar{\theta}}} + \frac{1}{n \theta^2} \sqrt{\frac{\log n}{n \theta^2}} \right)
\]

and

\[
\sum_{1 \leq j \leq n: j \neq i} \frac{A(i, j) + \theta \bar{\theta} + \frac{\log(n)}{n}}{H(i, j)^3} \leq C \frac{\theta_i}{\bar{\theta}} + C \frac{\log n}{n \theta^2}
\]

over the event \(E\). We thus conclude that

\[
\left| W(i)(\tilde{H}(i)^{-1/2}(X - I_n)(\tilde{\xi}_1^{(i)} - \xi_1^{(i)}) \right|
\]

\[
\leq C \left( \frac{\theta_i}{\bar{\theta}} + \frac{\log n}{n \theta^2} \right) \| (\tilde{H}(i)^{-\frac{1}{2}}(\xi_1^{(i)} - w\hat{\xi}_1)) \|_\infty + C \left( \frac{1}{n \theta^2} \sqrt{\frac{\theta_i}{\bar{\theta}}} + \frac{1}{n \theta^2} \sqrt{\frac{\log n}{n \theta^2}} \right) \| \tilde{\xi}_1^{(i)} - w\hat{\xi}_1 \|
\]

Further with \(\tilde{H}(i)^{i, i} \approx n \bar{\theta}(\theta_i \vee \bar{\theta})\), we have

\[
\sqrt{\tilde{H}(i, i)} \left| W(i)(\tilde{H}(i)^{-1/2}(X - I_n)(\tilde{\xi}_1^{(i)} - \xi_1^{(i)}) \right|
\]

\[
\leq C \tilde{\kappa}_i n \bar{\theta} \| (\tilde{H}(i)^{-1/2}(w\hat{\xi}_1 - \xi_1^{(i)})) \|_\infty + C(n \theta^2)^{-\frac{1}{2}} \| \xi_1^{(i)} - w\hat{\xi}_1 \|
\]

over the event \(E\). Now, plugging (C.16) and (C.21) into (C.15) and combining all \(i\)'s, we thus finish the proof of (4.7).

**C.3.3 Proof of (4.8)**

Note that \(\xi_1^{(i)}\) is the first eigenvector of \((\tilde{H}(i)^{-1/2} \tilde{A}(i)(\tilde{H}(i)^{-1/2})\). The eigen-gap between \(\tilde{\lambda}_1^{(i)}\) and \(|\tilde{\lambda}_2^{(i)}|\) is of order \(K^{-1} \lambda_1(PG)\) in light of Weyl’s inequality

\[
\max_i |\tilde{\lambda}_1^{(i)} - \tilde{\lambda}_2^{(i)}| \leq \| (\tilde{H}(i)^{-1/2}(\tilde{A}(i) - \Omega)(\tilde{H}(i)^{-1/2}) \| \leq C \sqrt{\frac{\log n}{n \theta^2}}
\]
and $K^{-1}\lambda_1(PG) \gg \sqrt{\log n/n^2}$. Similarly, the eigengap between $\hat{\xi}_1$ and $|\hat{\lambda}_2(i)|$ is of order $K^{-1}\lambda_1(PG)$. We claim that for $w = \text{sgn}(\hat{\xi}_1^{(i)} \xi_1^{(i)}) \equiv \text{sgn}(\hat{\xi}_1^{(i)} \xi_1^{(i)} + w) \geq C$. To see this, we derive

\[ |\hat{\xi}_1^{(i)} + w| \geq 2 - |\hat{\xi}_1^{(i)} - w| - |\xi_1^{(i)} - \hat{\xi}_1^{(i)}| \geq 2 - o(1) - \|\xi_1^{(i)} - \hat{\xi}_1^{(i)}\| \]

\[ \geq 2 - \sqrt{2}\|\xi_1^{(i)} - \hat{\xi}_1^{(i)}\| - o(1) \]

\[ = 2 - o(1) \]

due to $|\langle \hat{\xi}_1^{(i)} \xi_1^{(i)} \rangle - 1| \leq K\lambda_1^{-1}(PG)\|\bar{L}^{(i)} - \bar{Z}^{(i)}\| = o(1)$ by sine-theta theorem. This indicates that the sin-theta theorem applied to $\hat{\xi}_1$ and $\xi_1^{(i)}$ should involve $w$ instead of $-w$. More precisely,

\[ \|\hat{\xi}_1^{(i)} - w\hat{\xi}_1\| \leq K\lambda_1^{-1}(PG)\left(\left(\hat{H}^{(i)}\right)^{-1/2}\bar{A}^{(i)}\left(\hat{H}^{(i)}\right)^{-1/2} - H^{-1/2}A H^{-1/2}\right)\hat{\xi}_1 \]

Recall $X = \left(\hat{H}^{(i)}\right)^{1/2}H^{-1/2}$. It is seen that

\[ \tilde{\Delta} = \left(\hat{H}^{(i)}\right)^{-1/2}(\bar{A}^{(i)} - A)(\hat{H}^{(i)})^{-1/2} + ((\hat{H}^{(i)})^{-1/2}A(\hat{H}^{(i)})^{-1/2} - H^{-1/2}AH^{-1/2}) \]

\[ = -\left(\hat{H}^{(i)}\right)^{-1/2}(e_i W(i) + W(i)e'_i)(\hat{H}^{(i)})^{-1/2} + (\hat{H}^{(i)})^{-1/2}A((\hat{H}^{(i)})^{-1/2} - H^{-1/2}) + ((\hat{H}^{(i)})^{-1/2} - H^{-1/2})AH^{-1/2} \]

\[ = -\left(\hat{H}^{(i)}\right)^{-1/2}(e_i W(i) + W(i)e'_i)(\hat{H}^{(i)})^{-1/2} + (\hat{H}^{(i)})^{-1/2}A(\hat{H}^{(i)})^{-1/2}(I_n - X) + (X^{-1} - I_n)H^{-1/2}AH^{-1/2}. \]

By definition, $H^{-1/2}AH^{-1/2}\hat{\xi}_1 = \hat{\lambda}_1\hat{\xi}_1$. It follows that

\[ \tilde{\Delta}\hat{\xi}_1 = -\left(\hat{H}^{(i)}\right)^{-1/2}(e_i W(i) + W(i)e'_i)(\hat{H}^{(i)})^{-1/2}\hat{\xi}_1 + (\hat{H}^{(i)})^{-1/2}A(\hat{H}^{(i)})^{-1/2}(I_n - X)\hat{\xi}_1 + \hat{\lambda}_1(X^{-1} - I_n)\hat{\xi}_1. \]

As a result,

\[ \|\hat{\xi}_1^{(i)} - w\hat{\xi}_1\| \leq K\lambda_1^{-1}(PG)\|W(i)(\hat{H}^{(i)})^{-1/2}\hat{\xi}_1\| + K\lambda_1^{-1}(PG)\|((\hat{H}^{(i)})^{-1/2}W(i))\| \cdot |\hat{\xi}_1(i)| + C\|(I_n - X)\hat{\xi}_1\| \]

\[ \leq K\lambda_1^{-1}(PG)\|W(i)(\hat{H}^{(i)})^{-1/2}\hat{\xi}_1\| + \frac{CK\lambda_1^{-1}(PG)}{\sqrt{n^2/2}}|\hat{\xi}_1(i)| + C\|(I_n - X)\hat{\xi}_1\|, \]

\[ \leq C\lambda_1^{-1}(PG)\left(\frac{|W(i)(\hat{H}^{(i)})^{-1/2}\hat{\xi}_1|}{\sqrt{\hat{H}^{(i)}(i,i)}} + \frac{1}{\sqrt{n^2}}|\hat{\xi}_1(i)| + \frac{1}{\sqrt{n^2}}|w\hat{\xi}_1(i) - \hat{\xi}_1(i)|\right) + C\|(I_n - X)\hat{\xi}_1\|, \]

\[ \leq C\lambda_1^{-1}(PG)\left(\frac{|W(i)(\hat{H}^{(i)})^{-1/2}\hat{\xi}_1|}{\sqrt{\hat{H}^{(i)}(i,i)}} + \frac{\kappa_i}{\sqrt{\log(n)}} + \frac{1}{\sqrt{n^2}}|w\hat{\xi}_1(i) - \hat{\xi}_1(i)|\right) + C\|(I_n - X)\hat{\xi}_1\|, \]

where in the first line we have used $\|((\hat{H}^{(i)})^{-1/2}A(\hat{H}^{(i)})^{-1/2}\| \leq C\lambda_1^{-1}(PG)$ and $\|X - I_n\| \leq 1/2$, in the second line we have used the estimate

\[ (\hat{H}^{(i)}(i,i))^{-1/2}W(i)i'^{1/2} \leq \sqrt{W(i)(\hat{H}^{(i)})^{-1/2}W(i)} \]
Below, we bound that in (C.12) and (C.16), we have seen that the first two terms are bounded by up to some constant. Combining the above gives and in the last line we have used (B.3).

We consider the first term in (C.23). Note that

\[ |\tilde{\kappa}_i (1 + n\tilde{\theta}) (H(\hat{\xi}_1)^{-\frac{1}{2}} (w \hat{\xi}_1 - \tilde{\xi}_1(i))\|_{\infty}) + \frac{\log(n)}{n\tilde{\theta}} \|\tilde{\xi}_1(i) - w \hat{\xi}_1\| \]

up to some constant. Combining the above gives

\[ \frac{|W(i)(\tilde{H}(\hat{\xi}_1)^{-\frac{1}{2}} |}{\tilde{H}(\hat{\xi}_1)} \leq C\tilde{\kappa}_i (1 + n\tilde{\theta}) (H(\hat{\xi}_1)^{-\frac{1}{2}} (w \hat{\xi}_1 - \tilde{\xi}_1(i))\|_{\infty}) + \frac{1}{n\tilde{\theta}^2} \|\tilde{\xi}_1(i) - w \hat{\xi}_1\|. \quad \text{(C.24)} \]

We plug it into (C.23) and move all terms of \(\|\tilde{\xi}_1 - w \hat{\xi}_1\|\) to the left hand side. It follows that

\[ \|\tilde{\xi}_1 - w \hat{\xi}_1\| \leq C K \lambda_1^{-1}(PG) \left( \tilde{\kappa}_i (1 + n\tilde{\theta}) (H(\hat{\xi}_1)^{-\frac{1}{2}} (w \hat{\xi}_1 - \tilde{\xi}_1(i))\|_{\infty}) + \frac{|w \hat{\xi}_1(i) - \tilde{\xi}_1(i)|}{\sqrt{n\tilde{\theta}}^2} \right) + C\|(I_n - X)\hat{\xi}_1\| \quad \text{(C.25)} \]

Below, we bound \(\|\hat{\xi}_1(i)\|\). Note that

\[ \|\hat{\xi}_1(i)\| \leq \|\hat{\xi}_1(i)\| + \|\hat{\xi}_1(i)\|, \quad \text{(C.26)} \]

where

\[ \|\hat{\xi}_1(i)\| \leq \sum_{j=1}^{n} |X(j, j) - 1|^2 (\tilde{\xi}_1(i))^2 =: (J_1) \]

\[ \|\hat{\xi}_1(i)\| \leq (\tilde{H}(\hat{\xi}_1)^{-\frac{1}{2}} (w \hat{\xi}_1 - \tilde{\xi}_1(i))\|_{\infty}^2 \sum_{j=1}^{n} |X(j, j) - 1|^2 \tilde{H}(\hat{\xi}_1(i))^2 \]

\[ =: \|\tilde{H}(\hat{\xi}_1)^{-\frac{1}{2}} (w \hat{\xi}_1 - \tilde{\xi}_1(i))\|_{\infty}^2 \cdot (J_2). \]

Recall the bound of \(|X(j, j) - 1|\) in (C.13). It follows that over the event \(E\),

\[ (J_1) \leq C \sum_{j=1}^{n} \frac{|A(i, j) + \theta \hat{\theta} + \frac{\log(n)}{n}|^2 (\tilde{\xi}_1(i))^2}{\tilde{H}(\hat{\xi}_1(i))^2} \leq C \sum_{j=1}^{n} \left( A(i, j) + \theta \hat{\theta} + \frac{\log(n)}{n} \right) \left( \frac{(\tilde{\xi}_1(i))^2}{\tilde{H}(\hat{\xi}_1(i))^2} \right)^2 \]
(J_2) \leq C \sum_{j=1}^{n} \dfrac{|A(i, j) + \theta_i \tilde{\theta} + \dfrac{\log(n)}{n}|^2}{\tilde{H}^{(i)}(j, j)} \leq C \sum_{j=1}^{n} \left( A(i, j) + \theta_i \tilde{\theta} + \dfrac{\log(n)}{n} \right) \frac{1}{\tilde{H}^{(i)}(j, j)}

where we again use the fact that \( A(i, j) \in \{0, 1\} \). We shall bound the two terms similarly, using the Bernstein’s inequality (Theorem B.1). For \( (J_1) \),

- The mean is bounded by (up to some constant)
  \[
  \sum_{j=1}^{n} \left( \theta_i \theta_j + \theta_i \tilde{\theta} + \dfrac{\log(n)}{n} \right) \frac{1}{(n\theta)^3(\theta \vee \theta_j)} \leq \frac{1}{n\theta^2} \cdot \dfrac{n\theta_i + \log(n)}{(n\theta)^2};
  \]

- The variance is bounded by (up to some constant)
  \[
  \sum_{j=1}^{n} \frac{\theta_i \theta_j}{(n\theta)^6(\theta \vee \theta_j)^2} \leq \frac{1}{(n\theta)^2} \cdot \dfrac{\theta_i}{(n\theta)^2};
  \]

- Each individual term is bounded by (up to some constant)
  \[
  \dfrac{|\tilde{\xi}_i^{(i)}(j)|}{|\tilde{H}^{(i)}(j, j)|^2} \leq \dfrac{1}{n\theta^2} \cdot \dfrac{1}{(n\theta)^2}.
  \]

We then have

\[
(J_1) \leq C \dfrac{\theta_i}{n\theta^2} \left( 1 + \dfrac{\log(n)}{n\theta_i} \right) \quad \text{(C.27)}
\]

For \( (J_2) \),

- The mean is bounded by (up to some constant)
  \[
  \sum_{j=1}^{n} \left( \theta_i \theta_j + \theta_i \tilde{\theta} + \dfrac{\log(n)}{n} \right) \frac{1}{n\theta(\theta \vee \theta_j)} \leq \dfrac{\theta_i}{\theta} + \dfrac{\log(n)}{n\theta^2};
  \]

- The variance is bounded by (up to some constant)
  \[
  \sum_{j=1}^{n} \frac{\theta_i \theta_j}{(n\theta)^2(\theta \vee \theta_j)^2} \leq \frac{1}{n\theta^2} \cdot \dfrac{\theta_i}{\theta};
  \]

- Each individual term is bounded by (up to some constant)
  \[
  \dfrac{1}{\tilde{H}(j, j)} \leq \dfrac{C}{n\theta^2}.
  \]

It follows that

\[
(J_2) \leq \dfrac{C\theta_i}{\theta} + \dfrac{C\log(n)}{n\theta^2} \quad \text{(C.28)}
\]

Plugging \( \text{(C.27)}-\text{(C.28)} \) into \( \text{(C.26)} \), we find out that

\[
||\textbf{I}_n - X|| \tilde{\xi}_i|| \leq C \dfrac{\tilde{\kappa}_i}{\sqrt{\log(n)}} \left( 1 + n\theta \right) ||(\tilde{H}^{(i)})^{-\frac{1}{2}}(w\tilde{\xi}_1 - \tilde{\zeta}_1^{(i)})||_{\infty} \quad \text{(C.29)}
\]

We plug \( \text{(C.29)} \) into \( \text{(C.25)} \), together with the assumption \( \lambda_1(PG) \geq \alpha_n \), to get

\[
||\tilde{\xi}_1^{(i)} - w\tilde{\xi}_1|| \leq C \dfrac{K}{\alpha_n} \tilde{\kappa}_i \left( 1 + n\theta \right) ||(\tilde{H}^{(i)})^{-\frac{1}{2}}(w\tilde{\xi}_1 - \tilde{\zeta}_1^{(i)})||_{\infty} + \dfrac{K}{\alpha_n \sqrt{n\theta^2}} |w\tilde{\xi}_1(i) - \tilde{\zeta}_1^{(i)}(i)| \quad \text{(C.30)}
\]

over the event \( E \), which proved \( \text{(4.8)} \) by considering all \( i \)'s altogether.
C.4 Proof of the second claim in Theorem 3.1

In this section, we show the proof of (3.7). Similarly to the proof of (3.6), we streamline the proof into the following lemmas. In addition to the notations in the end of Section 1 below we will use \( \| \cdot \|_{2 \to \infty} \) to denote the matrix \( 2 \to \infty \) norm. Specifically, for any matrix \( A \) of dimension \( n \times m \), \( x \) represents column vector of dimension \( m \), \( \| A \|_{2 \to \infty} := \max_{\| x \|_1} \| Ax \|_\infty = \max_i \| A(i) \|_\infty \).

Lemma C.1. Suppose the assumptions in Theorem 3.1 hold. Recall \( \kappa_t := \sqrt{\log n \over n^2} \sqrt{\bar{\theta}_t \over \theta} \) for \( 1 \leq t \leq n \). With probability \( 1 - o(n^{-3}) \), simultaneously for \( 1 \leq i, t \leq n \),

\[
\| \hat{\Xi}_1^{(i)}(t)O_2^{(i)} - \Xi_1(t) \| \leq CK^{2} \beta_n^{-1} \kappa_t \left( 1 \wedge \sqrt{\bar{\theta}_t \over \theta} \right),
\]

(C.31)

for some orthogonal matrices \( O_2^{(i)} \in \mathbb{R}^{K^{-1},K^{-1}} \).

Lemma C.2. Under the assumptions in Theorem 3.1. With probability \( 1 - o(n^{-3}) \), simultaneously for \( 1 \leq i \leq n \),

\[
\| \hat{\Xi}_1(i) - \hat{\Xi}_1^{(i)}(i)O_3^{(i)} \| \leq CK^{2} \beta_n^{-1} \kappa_i + CK \beta_n^{-1} \| e'_i \Delta \hat{\Xi}_1 \|,
\]

(C.32)

\[
\| e'_i \Delta \hat{\Xi}_1 \| \leq \| e'_i \Delta \hat{\Xi}_1^{(i)} \| + \| e'_i \Delta (\hat{\Xi}_1 - \hat{\Xi}_1^{(i)}O_4^{(i)}) \| + C \sqrt{n \theta^{-2}} \| \hat{\Xi}_1 - \hat{\Xi}_1^{(i)}O_5^{(i)} \|,
\]

(C.33)

for some orthogonal matrices \( O_4^{(i)} \), \( O_5^{(i)} \in \mathbb{R}^{K^{-1},K^{-1}} \) and \( O_3^{(i)} := O_4^{(i)}O_5^{(i)} \), where \( \Delta \equiv \Delta(i) := (\bar{H}(i))^{-1/2}WH^{-1/2} \) for short.

Lemma C.3. Under the assumptions in Lemma C.2. Recall the notation of \( \bar{\kappa}_i \) in (4.6) for \( 1 \leq i \leq n \). With probability \( 1 - o(n^{-3}) \), simultaneously for \( 1 \leq i \leq n \),

\[
\| e'_i \Delta \hat{\Xi}_1^{(i)} \| \leq CK^{2} \bar{\kappa}_i
\]

(C.34)

\[
\| e'_i \Delta (\hat{\Xi}_1 - \hat{\Xi}_1^{(i)}O_4^{(i)}) \| \leq CK^{2} \left( \bar{\kappa}_i \left( 1 + n\bar{\theta} \| (\bar{H}(i))^{-1/2} \hat{\Xi}_1 - \hat{\Xi}_1^{(i)}O_3^{(i)} \|_{2 \to \infty} \right) + \log n \over n^2 \theta \| \hat{\Xi}_1 - \hat{\Xi}_1^{(i)}O_5^{(i)} \| \right),
\]

(C.35)

\[
\| \hat{\Xi}_1 - \hat{\Xi}_1^{(i)}O_5^{(i)} \| \leq CK^{2} \beta_n^{-1} \bar{\kappa}_i \left( 1 + n\bar{\theta} \| H^{-1/2} (\hat{\Xi}_1 - \hat{\Xi}_1^{(i)}O_3^{(i)}) \|_{2 \to \infty} \right) + CK \beta_n^{-1} \sqrt{n \theta^{-2}} \| \hat{\Xi}_1(i) - \hat{\Xi}_1^{(i)}(i)O_5^{(i)} \|.
\]

(C.36)

In the sequel, we will prove the second claim in Theorem 3.1 (i.e., (3.7)) based on the above lemmas. The proofs of the lemmas are postponed to the next three subsections.

Proof of (3.7). Plugging Lemma C.3 into (C.33), we first have with probability \( 1 - o(n^{-3}) \), simultaneously for all \( 1 \leq i \leq n \),

\[
\| e'_i \Delta \hat{\Xi}_1 \| \leq CK^{2} \left( \bar{\kappa}_i \left( 1 + n\bar{\theta} \| (\bar{H}(i))^{-1/2} \hat{\Xi}_1 - \hat{\Xi}_1^{(i)}O_3^{(i)} \|_{2 \to \infty} \right) + \log n \over n^2 \theta \| \hat{\Xi}_1(i) - \hat{\Xi}_1^{(i)}(i)O_3^{(i)} \| \right)
\]

(C.36)
which, further substituted to \((C.32)\), implies that
\[
\|\hat{\Xi}(i) - \tilde{\Xi}(i)O_3(i)\| \leq CK^2\beta_n^{-1}\tilde{\kappa}_i(1 + n\bar{\theta})(\bar{H}(i))^{-\frac{1}{2}}(\hat{\Xi}(i) - \tilde{\Xi}(i)O_3(i))\|_{2\to\infty} + CK^3\log n \frac{1}{\sqrt{\log n}}\|\hat{\Xi}(i) - \tilde{\Xi}(i)O_3(i)\|.
\]

Since \(K^3\beta_n^{-2}\log n/n\bar{\theta}^2 = o(1)\), we then arrive at
\[
\|\hat{\Xi}(i) - \tilde{\Xi}(i)O_3(i)\| \leq CK^2\beta_n^{-1}\tilde{\kappa}_i(1 + n\bar{\theta})(\bar{H}(i))^{-\frac{1}{2}}(\hat{\Xi}(i) - \tilde{\Xi}(i)O_3(i))\|_{2\to\infty}.
\]

Set \(\hat{O}_1 \equiv O_1(i) := (O_3(i)^o)O_3(i)\). Using Lemma \([C.1]\) and let \(t = i\), we will see that
\[
\|\hat{\Xi}(i) - \Xi(i)\hat{O}_1\| \leq \|\hat{\Xi}(i) - \tilde{\Xi}(i)O_3(i)\| + \|\tilde{\Xi}(i)O_3(i) - \Xi(i)\|
\leq CK^2\beta_n^{-1}\tilde{\kappa}_i(1 + n\bar{\theta})(\bar{H}(i))^{-\frac{1}{2}}(\hat{\Xi}(i) - \tilde{\Xi}(i)O_3(i))\|_{2\to\infty}
\]

over the event \(E\). Suppose that \(\hat{\Xi}(\Xi_1)\) has the singular value decomposition (SVD) \(\hat{\Xi}(\Xi_1) = U'\cos \Theta V\), we define \(O_1 = \text{sgn}(\hat{\Xi}_1\Xi_1) : = U'V\). Using sine-theta theorem, we can derive
\[
\|O_1 - \hat{O}_1\| \leq \|\hat{\Xi}(\Xi_1 - O_1)\| + \|\hat{\Xi}(\Xi_1 - \hat{O}_1)\|
\leq \|\hat{\Xi}(\Xi_1 - O_1)\| + \|\hat{\Xi}(\Xi_1 - \tilde{\Xi}(i)O_3)\| + \|\tilde{\Xi}(i)O_3 - O_2(i)\|
\leq CK\beta_n^{-1}\left(\|O_1 - \hat{O}_1\| + \|\hat{\Xi}(\Xi_1 - O_1)\| + \|\hat{\Xi}(\Xi_1 - \tilde{\Xi}(i)O_3)\|\right)
\leq CK\beta_n^{-1}\sqrt{\frac{\log n}{n\bar{\theta}^2}},
\]

by which, we will obtain
\[
\|\hat{\Xi}(i) - \Xi(i)O_1\| \leq \|\hat{\Xi}(i) - \Xi(i)\hat{O}_1\| + \|\Xi(i)\| \cdot \|O_1 - \hat{O}_1\|
\leq CK^2\beta_n^{-1}\tilde{\kappa}_i(1 + n\bar{\theta})(\bar{H}(i))^{-\frac{1}{2}}(\hat{\Xi}(i) - \tilde{\Xi}(i)O_3(i))\|_{2\to\infty}
\]
\[(C.37)\]

and
\[
\|H_0^{-\frac{1}{2}}\Xi(1 - \hat{O}_1)\|_{2\to\infty} \leq \|H_0^{-\frac{1}{2}}\Xi(1 - \hat{O}_1)\|_{2\to\infty} \leq CK^2\beta_n^{-1}\frac{1}{n\bar{\theta}}\sqrt{\frac{\log n}{n\bar{\theta}^2}}
\]
\[(C.38)\]

Here to obtain the above two inequalities, we used the second estimate of \((B.3)\).

Applying Lemma \([C.1]\) again together with \((B.9), (C.38)\), it is easy to deduce that
\[
\|(\bar{H}(i))^{-\frac{1}{2}}(\hat{\Xi}(1 - \Xi_1O_1))\|_{2\to\infty}
\leq C\|H_0^{-\frac{1}{2}}(\hat{\Xi}(1 - \Xi_1O_1))\|_{2\to\infty} + C\|H_0^{-\frac{1}{2}}(\Xi_1O_2(i)) - \Xi_1(i)O_3(i)\|_{2\to\infty}
\leq C\|H_0^{-\frac{1}{2}}(\hat{\Xi}(1 - \Xi_1O_1))\|_{2\to\infty} + C\|H_0^{-\frac{1}{2}}(\Xi_1O_2(i))\|_{2\to\infty} + C\|H_0^{-\frac{1}{2}}(\Xi_1O_2(i))\|_{2\to\infty}
\leq C\|H_0^{-\frac{1}{2}}(\hat{\Xi}(1 - \Xi_1O_1))\|_{2\to\infty} + \frac{CK^2\beta_n^{-1}}{n\bar{\theta}}\sqrt{\frac{\log n}{n\bar{\theta}^2}}
\]

45
Thereby, according to the condition $K^3\beta_n^{-2}\log n/n\theta^2 = o(1)$, (C.37) can further improved to
\[
\|\hat{\Xi}_1(i) - \Xi_1(i)O'_1\| \leq CK^2\beta_n^{-1}\kappa_i(1 + n\theta\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1 - \Xi_1O'_1)\|_{2\to\infty}).
\] (C.39)

Next, we multiply both sides of the above inequality by $H_0^{-\frac{1}{2}}(i, i)$ and take the maximum over $i$ since $\hat{\Xi}_1 - \Xi O'_1$ is independent of $i$, it yields that,
\[
\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1 - \Xi_1O'_1)\|_{2\to\infty} = \max_i \|e_i' H_0^{-\frac{1}{2}} (\hat{\Xi}_1 O_1 - \Xi_1)\|
\[
\leq CK^2\beta_n^{-1}(n\theta)^{-1}\sqrt{\frac{\log n}{n\theta^2}} + o(\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1 - \Xi_1O'_1)\|_{2\to\infty})
\] (C.40)

Rearranging both sides of (C.40), we can conclude that
\[
\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1 - \Xi_1O'_1)\|_{2\to\infty} \leq CK^2\beta_n^{-1}(n\theta)^{-1}\sqrt{\frac{\log n}{n\theta^2}}.
\]

which, further substituted into (C.39), yields (3.7) due to the condition $K^2\beta_n^{-2}\log n/n\theta^2 = o(1)$.

\[\square\]

\section*{C.5 Proof of Lemma C.1}

We state the proof of Lemma C.1 which is quite similar to Lemma A.1 with additional attention to the non-commutative multiplication of matrices. Fix the index $i$, we start with the perturbation from $L_0$ to $\tilde{L}^{(i)}$.

\[
\hat{\Xi}_1^{(i)}\hat{\Lambda}_1^{(i)} = \tilde{L}^{(i)}\hat{\Xi}_1^{(i)} = (H_0^\frac{1}{2}(\hat{H}^{(i)})^{-\frac{1}{2}})L_0(H_0^\frac{1}{2}(\hat{H}^{(i)})^{-\frac{1}{2}})\hat{\Xi}_1^{(i)} = \hat{Y}\lambda_1\xi_1\hat{\Xi}_1^{(i)} + \hat{Y}\Xi_1\lambda_1\Xi_1\hat{Y}\hat{\Xi}_1^{(i)}
\]

by recalling the definition $\hat{Y} = H_0^\frac{1}{2}(\hat{H}^{(i)})^{-\frac{1}{2}}$. Then, for each $1 \leq t \leq n$

\[
\tilde{\Xi}_1^{(i)}(t) = \hat{Y}(t, t)\lambda_1\xi_1(t)\xi_1'\hat{\Xi}_1^{(i)}(\hat{\Lambda}_1^{(i)})^{-1} + \hat{Y}(t, t)\Xi_1(t)\lambda_1\Xi_1\hat{Y}\hat{\Xi}_1^{(i)}(\hat{\Lambda}_1^{(i)})^{-1}.
\] (C.41)

Recall (B.7), over the event $E$, we first crudely bound $\|\lambda_1(\hat{\Lambda}_1^{(i)})^{-1}\|$ by $\beta_n^{-1}\lambda_1(PG)$. Then, using the estimate (C.3), we can crudely bound the first term on the RHS of (C.41) by

\[
\|\hat{Y}(t, t)\lambda_1\xi_1(t)\xi_1'\hat{\Xi}_1^{(i)}(\hat{\Lambda}_1^{(i)})^{-1}\| \leq C\beta_n^{-1}\lambda_1(PG)\left(\|\hat{Y} - I_n\|\|\xi_1(t)\| + \|\xi_1'\hat{\Xi}_1^{(i)}\|\|\xi_1(t)\|\right)
\]

\[
\leq C\beta_n^{-1}\lambda_1(PG)\kappa_i(1 \wedge \sqrt{\frac{\theta}{\theta}})
\] (C.42)

over the event $E$, where we used the first estimate in Lemma B.2 and sin-theta theorem for $\|\xi_1'\hat{\Xi}_1^{(i)}\|$ that

\[
\|\xi_1'\hat{\Xi}_1^{(i)}\| \leq CK\lambda_1^{-1}(PG)\|\tilde{L}^{(i)} - L_0\| \leq C\|\hat{Y} - I_n\|.
\]

46
For the second term on the RHS of (C.41), we have
\[
\| \tilde{Y}(t, t) \Xi_1(t) \Lambda_1 \Xi_1' \tilde{Y} \Xi_1(i) (\tilde{\Lambda}_1(i))^{-1} - \Xi_1(t) \Lambda_1 \Xi_1' \tilde{Y} \Xi_1(i) (\tilde{\Lambda}_1(i))^{-1} \| \leq C\beta_n^{-1} \lambda_1(PG) \| \tilde{Y} - I_n \| \| \Xi_1(t) \|
\]
and
\[
\Xi_1(t) \Lambda_1 \Xi_1' \tilde{Y} \Xi_1(i) (\tilde{\Lambda}_1(i))^{-1} = \Xi_1(t) \Xi_1' \tilde{L}_0 \Xi_1(i) (\tilde{\Lambda}_1(i))^{-1} = \Xi_1(t) \Xi_1' \Xi_1 + \Xi_1(t) \Xi_1' (L_0 - \tilde{L}(i)) \Xi_1(i) (\tilde{\Lambda}_1(i))^{-1}.
\]
By singular value decomposition (SVD), we write \( \Xi_1' \Xi_1(i) = U \cos \Theta V' \) for some orthogonal matrices \( U, V \) and diagonal matrix \( \cos \Theta \) all of which are \( i \)-dependent. Setting \( O_2(i) = (\text{sgn}(\Xi_1' \Xi_1(i)))' := VU' \) which is an orthogonal matrix, then we obtain that
\[
\| \Xi_1'( \tilde{Y} \Xi_1(i) - O_2(i))' \| \leq C(K \beta_n^{-1} \| \tilde{L}(i) - L_0 \|)^2 \leq C K \beta_n^{-1} \| \tilde{L}(i) - L_0 \|. \tag{C.43}
\]
Here we used the fact that \( K \beta_n^{-1} \| \tilde{L}(i) - L_0 \| \leq C K \beta_n^{-1} \| \tilde{Y} - I_n \| \leq K \beta_n^{-1} \sqrt{\log n}/\sqrt{n \theta^2} = o(1) \). Further we crudely bound
\[
\| \Xi_1(t) \Xi_1'(L_0 - \tilde{L}(i)) \Xi_1(i) (\tilde{\Lambda}_1(i))^{-1} \| \leq C K \beta_n^{-1} \| \tilde{L}(i) - L_0 \| \| \Xi_1(t) \|.
\]
Hence,
\[
\| \tilde{Y}(t, t) \Xi_1(t) \Lambda_1 \Xi_1' \tilde{Y} \Xi_1(i) (\tilde{\Lambda}_1(i))^{-1} - \Xi_1(t)(O_2(i))' \| \leq C \beta_n^{-1} \lambda_1(PG) \| \tilde{Y} - I_n \| \| \Xi_1(t) \|
\leq C \sqrt{K} \beta_n^{-1} \lambda_1(PG) \kappa_i \left( 1 \wedge \sqrt{\frac{\theta_i}{\theta}} \right) \tag{C.44}
\]
over the event \( E \).

Plugging in (C.42) and (C.44) back to (C.41), we simply conclude that
\[
\| \hat{\Xi}_1(i)(t) - \Xi_1(t)O_2 \| \leq C \sqrt{K} \beta_n^{-1} \lambda_1(PG) \kappa_i \left( 1 \wedge \sqrt{\frac{\theta_i}{\theta}} \right)
\]
over the event \( E \). Combining all \( i \)'s together and noting \( P(E) = 1 - o(n^{-3}) \). This finished the proof of Lemma [C.1] by further noticing that \( \lambda_1(PG) = C K \).

### C.6 Proof of Lemma [C.2]

In this section, we prove Lemma [C.2].

Let us fix the index \( i \). The proof of (C.33) is straightforward by the decomposition
\[
\hat{\Xi}_1 = \hat{\Xi}_1(i)O_4(i)O_5(i) + (\hat{\Xi}_1(i) - \hat{\Xi}_1(i)O_4(i))O_5(i) + \hat{\Xi}_1(i)O_5(i)
\]
where the two orthogonal matrices \( O_4(i), O_5(i) \) will be specified later. We further bound
\[
\| \epsilon_i \Delta (\hat{\Xi}_1 - \Xi_1(i)O_5(i)) \| \leq \frac{1}{\sqrt{H(i)(i, i)}} \| W(i)(\bar{H}(i))^{-\frac{1}{2}} \| \| X \| \| \hat{\Xi}_1 - \Xi_1(i)O_5(i) \|
\]
\[ \leq \frac{C}{n^{\theta^2}} \| \hat{X}_1 - \Xi_1 O_5^{(i)} \| \]

over the event \( E \), by writing \( \Delta = (\hat{H}^{(i)})^{-1/2}W(\hat{H}^{(i)})^{-1/2}X \) and using the fact \( \|X\| \leq C \) and \( \|W(i)(\hat{H}^{(i)})^{-1/2}\| \leq \sqrt{\theta_i/\theta} \vee \sqrt{\log n/n^{\theta^2}} \) over the event \( E \). This together with the trivial identities \( \|e_i^j\Delta e_i^j(\hat{X}_1O_4^{(i)}O_5^{(i)}\| = \|e_i^j\Delta(\Xi_1 - \hat{X}_1O_4^{(i)})O_5^{(i)}\| = \|e_i^j\Delta(\Xi_1 - \hat{X}_1O_4^{(i)})\| \) implies (C.33).

We then turn to show (C.32). Note that

\[ \hat{X}_1 \lambda_1 = L \hat{X}_1 = X(\hat{H}^{(i)})^{-1/2}A(\hat{H}^{(i)})^{-1/2}X \hat{X}_1 = X(\hat{H}^{(i)})^{-1/2}\Omega(\hat{H}^{(i)})^{-1/2}X \hat{X}_1 + X(\hat{H}^{(i)})^{-1/2}(A - \Omega)(\hat{H}^{(i)})^{-1/2}X \hat{X}_1 \]

by the notation \( X = (\hat{H}^{(i)})^{1/2}H^{-1/2} \). Then,

\[ \hat{X}_1(i) = X(i, i)\hat{\lambda}_1^{(i)}(\hat{\xi}_1^{(i)}(i) \hat{\xi}_1^{(i)})^t X \hat{X}_1 \hat{A}_1^{-1} + X(i, i)\hat{\xi}_1^{(i)}(i)\hat{\lambda}_1^{(i)}(\hat{\xi}_1^{(i)})^t X \hat{X}_1 \hat{A}_1^{-1} + X(i, i)e_i^j(\hat{H}^{(i)})^{-1/2}(A - \Omega)(\hat{H}^{(i)})^{-1/2}X \hat{X}_1 \hat{A}_1^{-1}. \]  

(C.45)

Recall the estimate \( \|X - I_n\| \leq C\sqrt{\log n/\sqrt{n^{\theta^2}}} \) following from Lemma B.5 and the properties of eigenvalues and eigenvectors of \( \hat{L}^{(i)} \) in Lemma B.4. Then, for the first term on the RHS of (C.45), we have

\[ \|X(i, i)\hat{\lambda}_1^{(i)}(\hat{\xi}_1^{(i)}(i) \hat{\xi}_1^{(i)})^t X \hat{X}_1 \hat{A}_1^{-1}\| \leq CK^{-1}\lambda_1 (PG)\| \hat{A}_1^{-1}\|\| \hat{\xi}_1^{(i)}(i)\| \left(\|X - I_n\| + \|((\hat{\xi}_1^{(i)}))^t \hat{X}_1\|\right). \]

(C.46)

Recall that \( \hat{\lambda}_j \)’s for \( 1 \leq j \leq K \) share the same asymptotic as \( \lambda_j \)’s in (B.2) over the event \( E \). By sin-theta theorem and (C.7), we have the bound

\[ \|(\hat{\xi}_1^{(i)})^t \hat{X}_1\| \leq CK\lambda_1^{-1}(PG)\|L - \hat{L}^{(i)}\| \leq C\left(\sqrt{\frac{\log n}{n^{\theta^2}}} + \frac{K\lambda_1^{-1}(PG)}{\sqrt{n^{\theta^2}}}\right). \]

(C.47)

Thus, plugging (C.47), (B.8) together with \( \|X - I_n\| \leq C\sqrt{\log n/\sqrt{n^{\theta^2}}} \) into (C.46), we arrive at

\[ \|X(i, i)\hat{\lambda}_1^{(i)}(\hat{\xi}_1^{(i)}(i) \hat{\xi}_1^{(i)})^t X \hat{X}_1 \hat{A}_1^{-1}\| \leq CK\beta^{-1}_n \kappa_i \left(1 \wedge \sqrt{\frac{\theta_i}{\theta}}\right), \]

(C.48)

where we used the trivial bound \( \lambda_1 (PG) \leq CK \).

To estimate the other two term in (C.45), we need the assistance of \( \Xi_1 \), the eigenspace of \((\hat{H}^{(i)})^{-1/2} \hat{A}^{(i)}(\hat{H}^{(i)})^{-1/2} \), which is counterpart to \( \hat{X}_1 \) and \( \hat{X}_1 \). Recall that \( \hat{A}^{(i)} = \Omega + \hat{W}^{(i)} - \text{diag}(\Omega) \) where \( \hat{W}^{(i)} \) is obtained by zeroing-out \( i \)-th row and column of \( W \). Similarly to (C.43), we can then claim that there exists an orthogonal matrix \( O_4^{(i)} \) by sin-theta theorem such that

\[ \|\Xi_1 - \Xi_1 O_4^{(i)}\| \leq CK\beta^{-1}_n \| (\hat{H}^{(i)})^{-1/2} \hat{A}^{(i)}(\hat{H}^{(i)})^{-1/2} - (\hat{H}^{(i)})^{-1/2} \Omega(\hat{H}^{(i)})^{-1/2} \| \leq \frac{CK\beta^{-1}_n}{\sqrt{n^{\theta^2}}} \]

(C.49)
over the event $E$, where $O_4^i = \text{sgn}((\hat{\Xi}_i^{(i)})'\hat{\Xi}_i^{(i)})$. We will also need an orthogonal matrix $O_5 \equiv O_5(i) := \text{sgn}((\hat{\Xi}_i^{(i)})'\hat{\Xi}_i)$. Again by sin-theta theorem,

$$
\|((\hat{\Xi}_i^{(i)})'\hat{\Xi}_i - O_5)\|^\frac{1}{2} \leq CK \beta_n^{-1}\|(\hat{H}^{(i)})^{-\frac{1}{2}} A^{(i)}(\hat{H}^{(i)})^{-\frac{1}{2}} - H^{-\frac{1}{2}} A H^{-\frac{1}{2}}\| \\
\leq C \beta_n^{-1} \lambda_1(PG) \|X - I_n\| + CK \beta_n^{-1}\|(\hat{H}^{(i)})^{-\frac{1}{2}} (A^{(i)} - A)(\hat{H}^{(i)})^{-\frac{1}{2}}\| \\
\leq CK \beta_n^{-1} \sqrt{\frac{\log n}{n \theta^2}}.
$$

(C.50)

Here we used $\|(\hat{H}^{(i)})^{-\frac{1}{2}} A(\hat{H}^{(i)})^{-\frac{1}{2}}\| \prec \|(\hat{H}^{(i)})^{-\frac{1}{2}} \Omega(\hat{H}^{(i)})^{-\frac{1}{2}}\| \prec K^{-1} \lambda_1(PG)$ to get $K$ canceled for the first term of second line above. We then introduce the shorthand notation $O_3^i = O_4^i O_5^i$. And for the second term on the RHS of (C.45), similarly to (C.44), we get

$$
\|X(i,i)\hat{\Xi}_i^{(i)(i)}\hat{\Lambda}^{(i)}(\hat{\Xi}_i^{(i)})'X\hat{\Xi}_i^{(i)(i)} - \hat{\Xi}_i^{(i)(i)}O_3^i\| \\
\leq C(\beta_n^{-1} \lambda_1(PG) \|X - I_n\| + K \beta_n^{-1} \|L - \hat{L}^{(i)}\| + \|(\hat{\Xi}_i^{(i)})'\hat{\Xi}_i^{(i)} - O_3^i\|) \|\hat{\Xi}_i^{(i)(i)}\| \\
\leq CK \beta_n^{-1} \kappa_i + C\|(\hat{\Xi}_i^{(i)})'\hat{\Xi}_i^{(i)} - O_3^i\|\|\hat{\Xi}_i^{(i)(i)}\| \\
$$

(C.51)

over the event $E$, where we recall that $\kappa_i = \sqrt{\log n/n \theta^2} \cdot \sqrt{\theta_i/n \theta}$. Moreover, we have

$$
\|(\hat{\Xi}_i^{(i)})'\hat{\Xi}_i - O_3^i\| \leq \|\Xi_i^{(i)} - \hat{\Xi}_i^{(i)}\| + \|(\Xi_i^{(i)})'\hat{\Xi}_i - O_5^i\| \\
$$

which with (C.49), (C.50) and (B.8) leads to

$$
\|X(i,i)\hat{\Xi}_i^{(i)(i)}\hat{\Lambda}^{(i)}(\hat{\Xi}_i^{(i)})'Y\hat{\Xi}_i^{(i)(i)} - \hat{\Xi}_i^{(i)(i)}O_3^i\| \leq CK \beta_n^{-1} \kappa_i
$$

(C.52)

Combining (C.48) and (C.52) back into (C.45), we get

$$
\|\hat{\Xi}_i^{(i)} - \Xi_i^{(i)}(i)\| \leq CK \beta_n^{-1} \kappa_i + \|X(i,i)e'_i(\hat{H}^{(i)})^{-\frac{1}{2}}(A - \Omega)(\hat{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_i^{(i)(i)}\| \\
$$

(C.53)

In the sequel, we proceed to the second on the RHS above. First, using the trivial bound $|X(i,i)| \leq 2$ and $\|\hat{\Lambda}_i\|^{-1} \leq K \beta_n^{-1}$, we have

$$
\|X(i,i)e'_i(\hat{H}^{(i)})^{-\frac{1}{2}}(A - \Omega)(\hat{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_i^{(i)(i)}\| \\
\leq CK \beta_n^{-1}\|e'_i(\hat{H}^{(i)})^{-\frac{1}{2}}(A - \Omega)(\hat{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_i\| \\
\leq CK \beta_n^{-1}\left(\|e'_i(\hat{H}^{(i)})^{-\frac{1}{2}}W(\hat{H}^{(i)})^{-\frac{1}{2}}X\Xi_1\| + \|e'_i(\hat{H}^{(i)})^{-\frac{1}{2}}\text{diag}(\Omega)(\hat{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1\|\right)
$$

We can simply get the bound

$$
\|e'_i(\hat{H}^{(i)})^{-\frac{1}{2}}\text{diag}(\Omega)(\hat{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1\| = \|(\hat{H}^{(i)})^{-\frac{1}{2}}(i,i)\Omega(i,i)(X(i,i)\hat{\Xi}_1^{(i)}(i))\| \\
\leq C \frac{\theta_i^2}{n \theta (\theta \lor \sqrt{\theta_i})} \|\hat{\Xi}_1^{(i)}\| \\
\leq \frac{\sqrt{K}}{\sqrt{\log(n)}} \kappa_i \left(1 \lor \sqrt{\frac{\theta_i}{\theta}}\right)
$$

This leads to

$$
\|X(i,i)e'_i(\hat{H}^{(i)})^{-\frac{1}{2}}(A - \Omega)(\hat{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_i^{(i)(i)}\| \leq CK \beta_n^{-1} \kappa_i + CK \beta_n^{-1}\|e'_i\Delta\hat{\Xi}_1\|
$$

(C.54)

over the event $E$ satisfying $\mathbb{P}(E) = 1 - o(n^{-3})$. Combining (C.54) and (C.53) and considering all $i$'s, we then conclude the proof of (C.32).
C.7 Proof of Lemma C.3

The proof of Lemma C.3 is rather complicated. We will show the three claims (i.e., (C.34)–(C.36)) separately in the following three parts.

C.7.1 Proof of (C.34)

Write \( \Delta = (\tilde{H}^{(i)})^{-\frac{1}{2}} W(\tilde{H}^{(i)})^{-\frac{1}{2}} X \), we first crudely have
\[
\| e'_i \Delta \tilde{\xi}^{(i)}_1 \| \leq \| e'_i (\tilde{H}^{(i)})^{-\frac{1}{2}} W(\tilde{H}^{(i)})^{-\frac{1}{2}} \tilde{\xi}^{(i)}_1 \| + \| e'_i (\tilde{H}^{(i)})^{-\frac{1}{2}} W(\tilde{H}^{(i)})^{-\frac{1}{2}} (X - I_n) \tilde{\xi}^{(i)}_1 \| \quad \text{(C.55)}
\]
We start with the first term on the RHS of (C.55).
\[
\| e'_i (\tilde{H}^{(i)})^{-\frac{1}{2}} W(\tilde{H}^{(i)})^{-\frac{1}{2}} \tilde{\xi}^{(i)}_1 \| = \left\| \frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}} W(i) (\tilde{H}^{(i)})^{-\frac{1}{2}} \tilde{\xi}^{(i)}_1 \right\| = \left\| \frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}} \sum_{t=1}^{n} W(i,t) \tilde{\xi}^{(i)}_1 (t) \right\|.
\]

Thanks to the independence between \( \tilde{\xi}^{(i)}_1 \) and \( W(i) \), we can estimate \( \sum_{t=1}^{n} \frac{W(i,t) \tilde{\xi}^{(i)}_1 (t)}{\sqrt{\tilde{H}^{(i)}(t,t)}} \) componentwisely by Bernstein inequality with respect to the randomness of \( W(i) \). For each \( 2 \leq p \leq K \), we can bound the variance of \( \sum_{t=1}^{n} \frac{W(i,t) \tilde{\xi}^{(i)}_1 (t)}{\sqrt{\tilde{H}^{(i)}(t,t)}} \) by
\[
\text{var} \left( \sum_{t=1}^{n} \frac{W(i,t) \tilde{\xi}^{(i)}_1 (t)}{\sqrt{\tilde{H}^{(i)}(t,t)}} \right) = \sum_{t=1}^{n} \frac{\theta_i \theta_t}{\tilde{\xi}^{(i)}_1 (t)} (\tilde{\xi}^{(i)}_1 (t))^2 \leq C \sum_{t=1}^{n} \frac{\theta_i \theta_t}{n \theta_i \theta_t \lor \theta} (\tilde{\xi}^{(i)}_1 (t))^2 \leq C \frac{\theta_i}{n \theta}
\]
Each individual summand can be bounded by \( C/n \theta \) over the event \( E \). As a result,
\[
\left\| \sum_{t=1}^{n} \frac{W(i,t) \tilde{\xi}^{(i)}_1 (t)}{\sqrt{\tilde{H}^{(i)}(t,t)}} \right\| \leq C \left( \sqrt{\frac{\theta_i \log n}{n \theta}} + \frac{\log n}{n \theta} \right) \leq C \sqrt{\frac{\log n}{n \theta}} \sqrt{n \theta_i \lor \log n}
\]
Further with \( \tilde{H}^{(i)}(i,i) \approx n \theta_i \lor \theta \), we finally conclude that
\[
\| e'_i (\tilde{H}^{(i)})^{-\frac{1}{2}} W(\tilde{H}^{(i)})^{-\frac{1}{2}} \tilde{\xi}^{(i)}_1 \| = \left\| \frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}} \sum_{t=1}^{n} W(i,t) \tilde{\xi}^{(i)}_1 (t) \right\| \leq C K^{\frac{1}{2}} \tilde{\kappa}_i. \quad \text{(C.56)}
\]

Next, regarding the term \( \| e'_i (\tilde{H}^{(i)})^{-\frac{1}{2}} W(\tilde{H}^{(i)})^{-\frac{1}{2}} (X - I_n) \tilde{\xi}^{(i)}_1 \| \), using the estimate (C.13), we can derive
\[
\| W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}} (X - I_n) \tilde{\xi}^{(i)}_1 \| \leq C \sum_{t=1, t \neq i}^{n} \left| W(i,t) \right| \left| A(i,t) + \theta_i \tilde{\theta} + \frac{\log(n)}{n} \right| \left\| \tilde{\xi}^{(i)}_1 (t) \right\| \leq C \sum_{t=1, t \neq i}^{n} \left| W(i,t) \right| \left| A(i,t) + \theta_i \tilde{\theta} + \frac{\log(n)}{n} \right| \left\| \tilde{\xi}^{(i)}_1 (t) \right\| \leq C \sqrt{K} \tilde{\kappa}_i \quad \text{(C.57)}
\]
where the last step is analogous to how we get (C.14) by Bernstein’s inequality and one can refer to the details in Section C.3.1. Combining (C.56) and (C.57) into (C.55), and considering all \( i \)’s, we thus conclude (C.34).
C.7.2 Proof of (C.35)

The proof is similar to the proof of (4.7) in Section C.3.2. First, by definition, we bound

\[ \|e'_i \Delta(\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)})\| \leq \|e'_i (\tilde{H}^{(i)})^{-\frac{1}{2}} W(\tilde{H}^{(i)})^{-\frac{1}{2}} (\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)})\| \]

\[ + \|e'_i (\tilde{H}^{(i)})^{-\frac{1}{2}} W(\tilde{H}^{(i)})^{-\frac{1}{2}} (X - I_n)(\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)})\|. \quad (C.58) \]

We rewrite the first term on the RHS by

\[ \|e'_i (\tilde{H}^{(i)})^{-\frac{1}{2}} W(\tilde{H}^{(i)})^{-\frac{1}{2}} (\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)})\| = \left\| \frac{1}{\sqrt{\tilde{H}^{(i)}(i, i)}} W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}} (\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)}) \right\|. \]

According to the definition of \( \Xi^{(i)}_1, \Xi^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)} \) is also independent of \( W(i) \). Then, analogously to the previous section, restricted to the randomness of \( W(i) \), we bound the variance of each component of \( W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}} (\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)}) \) by

\[ \sum_{t=1}^{n} \theta_i \theta_t \left( \frac{(\tilde{\Xi}^{(i)}_1(t) - \tilde{\Xi}^{(i)}_1(t)O_4^{(i)}) e_p}{\sqrt{\tilde{H}^{(i)}(t, t)}} \right)^2 \leq \frac{C \theta_i}{n \bar{\theta}}. \]

Here to obtain the RHS upper bound, we used an elementary derivation

\[ \sum_t (\tilde{\Xi}^{(i)}_1(t)O_4^{(i)}) e_p^2 = e'_p (O_4^{(i)})'(\tilde{\Xi}^{(i)}_1) e_p = e'_p (O_4^{(i)})'(\tilde{\Xi}^{(i)}_1) e_p = 1. \]

There is some ambiguity over the dimension of \( e_p \) and \( e_t \). \( e_p \) shall be of dimension \( K - 2 \) while \( e_t \) is of dimension \( n \). Further, each summand in the \( p \)-th component of \( W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}} (\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)}) \) is bounded by \( C \| (\tilde{H}^{(i)})^{-\frac{1}{2}} (\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)})e_p \|_{\infty} \). We further have

\[ \sqrt{\sum_{p=2}^{K} \| (\tilde{H}^{(i)})^{-\frac{1}{2}} (\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)})e_p \|_{\infty}^2} \leq \sqrt{K} \| (\tilde{H}^{(i)})^{-\frac{1}{2}} (\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)}) \|_{2 \rightarrow \infty} + \frac{C \sqrt{K}}{\sqrt{n \bar{\theta}^2}} \| \tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)} \| \]

Thus, over the event \( E \),

\[ \|e'_i (\tilde{H}^{(i)})^{-\frac{1}{2}} W(\tilde{H}^{(i)})^{-\frac{1}{2}} (\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)})\| \leq C \sqrt{K} \kappa_i + C \sqrt{K} \frac{\log n}{\sqrt{n \bar{\theta} (\theta \vee \theta_i)}} \| (\tilde{H}^{(i)})^{-\frac{1}{2}} (\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)} O_5^{(i)}) \|_{2 \rightarrow \infty} + \frac{C \sqrt{K} \log n}{n \bar{\theta}^2} \| \tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_5^{(i)} \| \quad (C.59) \]

Next, for the second term of (C.58), using the estimate (C.13), we have

\[ \frac{\| W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}} (X - I_n)(\tilde{\Xi}^{(i)}_1 - \tilde{\Xi}^{(i)}_1 O_4^{(i)})\|}{\sqrt{\tilde{H}^{(i)}(i, i)}} \]
\[
\leq \frac{C}{\sqrt{H^{(i)}(i, i)}} \sum_{t=1, t \neq i}^{n} \frac{|W(i, t)|}{H^{(i)}(t, t)} \frac{A(i, t) + \theta_i \tilde{\theta} + \log(n)/n}{H^{(i)}(t, t)} \frac{\|\Xi_1(t) - \tilde{\Xi}_1^{(i)}(t)O_4^{(i)}\|}{\sqrt{H^{(i)}(t, t)}}
\]

Similarly to the derivations of upper bounds of (C.17) and (C.18), we bound the two sums on the RHS of (C.60) corresponding to the two terms in the parenthesis separately as follows:

\[
\leq \frac{1}{\sqrt{H^{(i)}(i, i)}} \sum_{t=1, t \neq i}^{n} \frac{A(i, t) + \theta_i \tilde{\theta} + \log(n)/n}{H^{(i)}(t, t)} \|((H^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)}))_{2 \to \infty}
\]

and

\[
\leq \frac{1}{\sqrt{H^{(i)}(i, i)}} \sum_{t=1, t \neq i}^{n} \frac{A(i, t) + \theta_i \tilde{\theta} + \log(n)/n}{H^{(i)}(t, t)} \|((H^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)}))_{2 \to \infty}
\]

over the event E, in which we applied (C.20) and (C.19). We plug the above two estimates into (C.60) and conclude that over the event E,

\[
\|W(i)(\hat{H}^{(i)})^{-\frac{1}{2}}(X - I_n)(\Xi_1^{(i)} - \tilde{\Xi}_1^{(i)}O_4^{(i)})\| \leq C \kappa_i \eta \|((\hat{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)}))_{2 \to \infty} + \sqrt{K(n\theta^2)^{-\frac{3}{2}}\|\hat{\Xi}_1 - \Xi_1^{(i)}O_5^{(i)}\|}
\]

This, together with (C.59), concludes the proof of (C.35) for fixed i, by the fact that \(\log(n)/\sqrt{n\theta^2} \leq \kappa_i\). Combining all i’s and the fact \(\mathbb{P}(E) = 1 - o(n^{-3})\), we finish the proof.

C.7.3 Proof of (C.36)

By sin-theta theorem and the fact that the eigen-gap is of the order \(O(K^{-1} \beta_n)\) in light of Weyl’s inequality (see (C.22)), analogously to (C.23), we first have

\[
\|\hat{\Xi}_1 - \Xi_1^{(i)}O_5^{(i)}\|
\]
\[ \leq K\beta_n^{-1}\|H^{-\frac{1}{2}}A^{-\frac{1}{2}} - (\bar{H}^{(i)})^{-\frac{1}{2}}A(\bar{H}^{(i)})^{-\frac{1}{2}}\| \]
\[ \leq K\beta_n^{-1}\| (I_n - X^{-1})H^{-\frac{1}{2}}A^{-\frac{1}{2}} + (\bar{H}^{(i)})^{-\frac{1}{2}}A(\bar{H}^{(i)})^{-\frac{1}{2}}(X - I_n)\| + \|(\bar{H}^{(i)})^{-\frac{1}{2}}(A - \bar{A}^{(i)})\| \]
\[ \leq CK\beta_n^{-1}\| (X - I_n)\| + \|(\bar{H}^{(i)})^{-\frac{1}{2}}A(\bar{H}^{(i)})^{-\frac{1}{2}}(X - I_n)\| + \|(\bar{H}^{(i)})^{-\frac{1}{2}}(e_iW + W(i'e_i'))\| . \]

(C.61)

We start with a simple derivation,
\[ \|(\bar{H}^{(i)})^{-\frac{1}{2}}A(\bar{H}^{(i)})^{-\frac{1}{2}}(X - I_n)\| \]
\[ \leq \|(H^{(i)})^{-\frac{1}{2}}A(\bar{H}^{(i)})^{-\frac{1}{2}}|| (X - I_n)\| + \|(\bar{H}^{(i)})^{-\frac{1}{2}}(A - \Omega)(\bar{H}^{(i)})^{-\frac{1}{2}}|| (X - I_n)\| \]
\[ \leq CK^{-1}\lambda_1(PG)\| (X - I_n)\|. \]

Second, we have
\[ \|(\bar{H}^{(i)})^{-\frac{1}{2}}W(i')\| = \bar{H}^{(i)}(i,i)\| \bar{X}_1(i)(\bar{H}^{(i)})^{-\frac{1}{2}}W(i')\| \]
\[ \leq \frac{C}{\sqrt{\theta^2}}\| \bar{X}_1(i) \| \]
\[ \leq \frac{C\sqrt{K}\kappa_i}{\sqrt{\log n}} + \frac{C}{\sqrt{\theta^2}}\| \bar{X}_1(i) - \bar{X}_1(i)\|. \]

where we decomposed \( \bar{X}_1(i) \) as \( \bar{X}_1(i)O_3^{(i)} + \bar{X}_1(i) - \bar{X}_1(i)O_3^{(i)} \) and employed (B.8) in the last step. Thus, we further bound the RHS of (C.61) as
\[ \| \bar{X}_1 - \bar{X}_1O_3^{(i)} \| \leq C\beta_n^{-1}\lambda_1(PG)\| (X - I_n)\| + CK\beta_n^{-1}\| \bar{H}^{(i)}(i,i)\| \]
\[ + \frac{CK\beta_n^{-1}\kappa_i}{\sqrt{\log n}} + \frac{CK\beta_n^{-1}}{\sqrt{\theta^2}}\| \bar{X}_1(i) - \bar{X}_1(i)\|. \]

(C.62)

In the sequel, we analyze the first two terms on the RHS above. For \( \| (X - I_n)\| \), similarly to (C.26), we decompose \( \bar{X}_1 \) and get that
\[ \| (X - I_n)\| \leq \| (X - I_n)\| + \| (X - I_n)(\bar{X}_1 - \bar{X}_1O_3^{(i)})\| \]

Then, one just copy the derivations for the two terms in (C.26) with \( \xi_1^{(i)} \), \( \bar{X}_1 \) and \( w \) replaced by \( O_3^{(i)} \) to get
\[ \| (X - I_n)\| \leq C\sum_{j=1}^n \left( \frac{A(i,j) + \theta_i\bar{\theta} + \frac{\log(n)}{n}}{H^{(i)}(j,j)} \right)^2 \leq \frac{CK\kappa_i^2}{\log(n)} \]
\[ \| (X - I_n)(\bar{X}_1 - \bar{X}_1O_3^{(i)})\| \leq C\| (\bar{H}^{(i)})^{-1/2}(\bar{X}_1 - \bar{X}_1O_3^{(i)})\|_2 \]
\[ \leq C\| (\bar{H}^{(i)})^{-1/2}(\bar{X}_1 - \bar{X}_1O_3^{(i)})\|_2 \cdot \frac{n\theta_i + \log(n)}{n\theta^2} \]

53
over the event $E$. More detailed steps can be referred to derivations from (C.26)-(C.28). We thereby arrive at

$$
\| (X - I_n) \hat{\Xi}_1 \| \leq C \sqrt{K} \tilde{\kappa}_i \left( 1 + n\bar{\theta} \| (\hat{H}(i))^{-\frac{1}{2}} (\hat{\Xi}_1 - \hat{\Xi}(i) O_3) \|_{2\rightarrow\infty} \right)
$$

(C.63)

Now we turn to study the term $\| \hat{H}(i, i)^{-\frac{1}{2}} W(i) (\hat{H}(i))^{-\frac{1}{2}} \hat{\Xi}_1 \|$. Using (C.56), (C.59), we can deduce that

$$
\| \hat{H}(i, i)^{-\frac{1}{2}} W(i) (\hat{H}(i))^{-\frac{1}{2}} \hat{\Xi}_1 \| \leq \| \hat{H}(i, i)^{-\frac{1}{2}} W(i) (\hat{H}(i))^{-\frac{1}{2}} \hat{\Xi}_1 \| + \| \hat{H}(i, i)^{-\frac{1}{2}} W(i) (\hat{H}(i))^{-\frac{1}{2}} (\hat{\Xi}_1 - \hat{\Xi}(i) O_3) \|
$$

$$
+ C \sqrt{K} \tilde{\kappa}_i \left( 1 + n\bar{\theta} \| (\hat{H}(i))^{-\frac{1}{2}} (\hat{\Xi}_1 - \hat{\Xi}(i) O_3) \|_{2\rightarrow\infty} \right)
$$

$$
+ C \left( \sqrt{K} \log \frac{n}{n\bar{\theta}^2} + \frac{1}{\sqrt{n\bar{\theta}^2}} \right) \| \hat{\Xi}_1 - \Xi(i) O_3 \|
$$

(C.64)

over the event $E$. Combining (C.63) and (C.64) into (C.62) and putting all terms equipped with factor $\| \hat{\Xi}_1 - \Xi(i) O_3 \|$ to the LHS, under the condition $K^3 \beta_n^{-2} \log(n)/n\bar{\theta}^2 = o(1)$ and $\lambda_1(PG) \leq CK$, we finally see that

$$
\| \hat{\Xi}_1 - \Xi(i) O_3 \| \leq CK^2 \beta_n^{-1} \tilde{\kappa}_i \left( 1 + n\bar{\theta} \| (\hat{H}(i))^{-\frac{1}{2}} (\hat{\Xi}_1 - \hat{\Xi}(i) O_3) \|_{2\rightarrow\infty} \right) + \frac{CK\beta_n^{-1}}{\sqrt{n\bar{\theta}^2}} \| \hat{\Xi}_1(i) - \Xi(i) O_3 \|
$$

over the event $E$. Thus we complete the proof by considering all $i$’s.

C.8 Proof of Corollary 3.1

Fix the choice of $\hat{\xi}_1$ such that $w = 1$ in (3.6). Choose the orthogonal matrix $O_1$ appeared in Theorem 3.1. By definition,

$$
\| O_1^T \hat{r}_i - r_i \| = \| e_i' (\hat{\Xi}_1 O_1 / \hat{\xi}_1(i) - \Xi / \xi_1(i)) \| \leq \| e_i' (\hat{\Xi}_1 O_1 - \Xi) / \hat{\xi}_1(i) \| + \| \Xi(i) \| \left| \frac{1}{\hat{\xi}_1(i)} - \frac{1}{\xi_1(i)} \right|
$$

Employing Theorem 3.1 with Lemma B.2, for $i \in S_n(c_0)$, we have

$$
\| e_i' (\hat{\Xi}_1 O_1 - \Xi) / \hat{\xi}_1(i) \| \leq C \left\| e_i' (\hat{\Xi}_1 O_1 - \Xi) / \xi_1(i) \right\| \leq C \sqrt{K^3 \log(n) / n\bar{\theta}^2} \leq C \sqrt{K^3 \log(n) / n\bar{\theta}^2} \delta_n^2
$$

and

$$
\| \Xi(i) \| \left| \frac{1}{\hat{\xi}_1(i)} - \frac{1}{\xi_1(i)} \right| \leq C \left\| \Xi(i) \right\| \left| \frac{\hat{\xi}_1(i) - \xi_1(i)}{\xi_1(i)} \right| \leq C \sqrt{K^3 \log(n) / n\bar{\theta}^2} \delta_n^2
$$

with probability $1 - o(n^{-3})$ simultaneously for $i \in S_n(c_0)$. Combining the above inequalities, we immediately get (3.11) simultaneously for $i \in S_n(c_0)$, with probability $1 - o(n^{-3})$.  

54
D Rate of Mixed-SCORE-Laplacian

We prove the error rate of Mixed-SCORE-Laplacian in this Section. More specifically, we show the proof of Theorem 3.2 in Section D.1, and we briefly state the proofs of Corollary 3.2 and 3.3 which are simple consequences of Theorem 3.2 in Section D.2.

D.1 Proof of Theorem 3.2

We only focus on \( \Phi \) by (3.2) in Section D.2. In Section D.1, we show the proof of Theorem 3.2 in Section D.1; And we briefly state the proofs of Corollary 3.2 and 3.3, which are simple consequences of Theorem 3.2 in Section D.2.

We prove the error rate of Mixed-SCORE-Laplacian in this Section. More specifically, we prove the error rate of Mixed-SCORE-Laplacian in this Section. The estimation error is then trivially bounded by some constant. Recall the definition, for \( i \in \hat{S}_n(c) \),

\[
\hat{\pi}_i^*(k) = \max\{\hat{w}_i(k)/\hat{b}_1(k), 0\}, \quad \hat{\pi}_i = \hat{\pi}_i^*/\|\hat{\pi}_i^*\|_1
\]

and correspondingly in the oracle case, \( \pi_i = \pi_i^*/\|\pi_i^*\|_1, \pi_i^* = [\text{diag}(b_1)]^{-1}w_i \). We shall study errors of \( \hat{w}_i \)’s and \( \hat{b}_1 \) compared to \( w_i \)’s and \( b_1 \) separately.

We first study \( \hat{w}_i \)’s. Thanks to the choice of a variant of successive projection as our vertex hunting algorithm, referring to Lemma 3.1 of Jin et al. (2017), it is easy to deduce that

\[
\|P\hat{V}O_1 - V\|_{2\to\infty} \leq C \max_{i \in \hat{S}_n(c)} \|O_1\hat{r}_i - r_i\| \leq \sqrt{\frac{K^3\log n}{n\theta_2^2\delta_n^2}}.
\] (D.1)

for some \( K \times K \) permutation matrix \( P \), where we denote by \( V = (v_1, v_2, \ldots, v_K)' \) and \( \hat{V} = (\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_K)' \). In our Mixed-SCORE-Laplacian algorithm, \( \hat{w}_i \)’s are solved from

\[
\hat{Q}\hat{\pi}_i = \left( \begin{array}{c} 1 \\ O_1'\hat{r}_i \end{array} \right), \quad \hat{Q} := \left( \begin{array}{cccc} 1 & 1 & \cdots & 1 \\ O_1'\hat{\pi}_1 & O_1'\hat{\pi}_2 & \cdots & O_1'\hat{\pi}_K \end{array} \right)
\]

Here, a little different from original linear system, we multiply \( \hat{r}_i \) and \( \hat{v}_1, \hat{v}_2, \ldots, \hat{v}_K \) by \( O_1' \) on the left. Analogously, for the oracle case,

\[
Qw_i = \left( \begin{array}{c} 1 \\ r_i \end{array} \right), \quad Q := \left( \begin{array}{cccc} 1 & 1 & \cdots & 1 \\ v_1 & v_2 & \cdots & v_K \end{array} \right).
\]

Note that since \( v_j \)’s, \( \hat{v}_j \)’s for \( 1 \leq j \leq K \) are the vertices, we easily get that both \( \hat{Q} \) and \( Q \) are of full-rank. Then,

\[
\|P\hat{w}_i - w_i\| = \left\| (\hat{Q}P')^{-1} \left( \begin{array}{c} 1 \\ O_1'\hat{\pi}_i \end{array} \right) - Q^{-1} \left( \begin{array}{c} 1 \\ r_i \end{array} \right) \right\|
\]

\[
\leq \left\| (\hat{Q}P')^{-1} - Q^{-1} \right\| \left| \left( \begin{array}{c} 1 \\ r_i \end{array} \right) \right\| + \left\| Q^{-1} \left[ \left( \begin{array}{c} 1 \\ O_1'\hat{\pi}_i \end{array} \right) - \left( \begin{array}{c} 1 \\ r_i \end{array} \right) \right] \right\|. \quad (D.2)
\]

For the first term on the RHS of (D.2), we have

\[
\left\| (\hat{Q}P')^{-1} - Q^{-1} \right\| \left| \left( \begin{array}{c} 1 \\ r_i \end{array} \right) \right| = \|\hat{Q}^{-1}(\hat{Q}P' - Q)Q^{-1} \left( \begin{array}{c} 1 \\ r_i \end{array} \right)\| = \|\hat{Q}^{-1}\| \| (\hat{Q}P' - Q)w_i \|,
\]
and
\[ \| (\hat{Q}P' - Q)w_i \| = \| (O'_i \hat{V}'P' - V')w_i \| \leq \| P\hat{V}O_1 - V \|_{2\to\infty} \]

If we can claim that \( \| \hat{Q}^{-1} \| \asymp 1 \), then we are done with the bound of the first term. Notice that one easily check
\[ \| \hat{Q}P' - Q \| \leq \sqrt{K}\| P\hat{V}O_1 - V \|_{2\to\infty} = o(\sqrt{K}) \]

since \( \frac{K^3 \log(n)}{n\theta^2\delta_n^2} \to 0 \) as \( n \to \infty \). Suppose that \( \| Q^{-1} \| \asymp K^{-\frac{1}{2}} \), then immediately \( \| \hat{Q}^{-1} \| \asymp K^{-\frac{1}{2}} \). To claim that \( \| Q^{-1} \| \asymp K^{-\frac{1}{2}} \), we use the identity
\[ v_k(t) = \frac{b_t(k)}{b_1(k)}, \quad 2 \leq t \leq K, \quad (D.3) \]

which is easy to be verified with some elementary derivations from the definition of \( R \) and the fact \( \Xi = H_0^{-\frac{1}{2}}\Theta\Pi B \) with \( B = (b_1, \cdots, b_K) \). We will see that
\[ Q = B'\text{diag}(1/b_1(1), \cdots, 1/b_1(K)) \]

And due to \( b_1(k) \asymp 1 \) (claimed in the Proof of Lemma \[ B.2 \]), we then obtain that
\[ \| Q^{-1} \| = \| \text{diag}(b_1(1), \cdots, b_1(K))B^{-1} \| \asymp \| B^{-1} \|. \]

Further recall that \( BB' = (\Pi'\Theta H_0^{-1}\Theta\Pi)^{-1} \). Hence, \( \lambda_{\min}(BB') = 1/\lambda_{\max}(\Pi'\Theta H_0^{-1}\Theta\Pi) \asymp K \), which leads to \( \| Q^{-1} \| \asymp \| B^{-1} \| \asymp K^{-\frac{1}{2}} \). As a consequence,
\[ \left\| (\hat{Q}P')^{-1} - Q^{-1} \right\| \left( \begin{array}{c} 1 \\ r_i \end{array} \right) \right\| \leq CK^{-\frac{1}{2}}\| P\hat{V}O_1 - V \|_{2\to\infty}. \]

Next, for the second term on the RHS of \( (D.2) \), one simply bounds it by \( \| \hat{Q}^{-1} \| \| O_1r_i - r_i \| \leq CK^{-\frac{1}{2}}\| O_1r_i - r_i \| \). Combining these two estimates into \( (D.2) \), with the aids of \( (D.1) \) and Lemma \[ 3.1 \], we conclude that
\[ \| P\hat{w}_i - w_i \| \leq CK^{-\frac{1}{2}}\| O_1r_i - r_i \| \leq C \sqrt{\frac{K\log n}{n\sigma(\hat{\theta} \wedge \theta)\delta_n^2}}. \quad (D.4) \]

Next, we study the error between \( 1/e'_k P\hat{b}_1 \) and \( b_1^{-1}(k) \). Here to the end of this section, with a little ambiguity of notation, we denote \( \{ e_k \}_{k=1}^K \) for the standard basis of \( \mathbb{R}^K \). By definition, since \( P \) is a permutation matrix,
\[ \left| \frac{1}{(e'_k P\hat{b}_1)^2} - \frac{1}{(b_1(k))^2} \right| \leq |\hat{\lambda}_1 - \lambda_1| + |e'_k \hat{P}V\hat{A}_1 \hat{V}'P' e_k - v_k'\Lambda_1 v_k|. \]

The eigenvalue difference is simply bounded by \( \sqrt{\log(n)/n\theta' B} \) by Weyl' inequality, which has been previously shown in entry-wise eigenvector analysis. To bound the second term above, we first claim \( \| v_k \| \leq C\sqrt{K} \). To see this, using \( (D.3) \) and \( b_1(k) \asymp 1 \),
\[ \| v_k \| \leq C\| e'_k B \| \leq C\| BB' \|^\frac{1}{2} \leq C\sqrt{K}. \]

56
We can then derive that
\[
|\varepsilon_k^* P\hat{V}\hat{A}_1\hat{V}'P'e_k - v_k'\Lambda_1 v_k| \\
\leq |\varepsilon_k^* (P\hat{V}O_1 - V)O_1'\hat{A}_1O_1'O_1'\hat{V}'P'e_k| + |v_k' O_1'\hat{A}_1 O_1 (O_1'\hat{V}'P' - V)e_k| + |v_k' (O_1'\hat{A}_1 O_1 - \Lambda_1) v_k| \\
\leq CK^{-\frac{1}{2}}|\lambda_2(PG)||P\hat{V}O_1 - V|_2 \to \infty + |v_k' (O_1'\hat{A}_1 O_1 - \Lambda_1) v_k|
\]
Here in the last step, we used the trivial bound $||\hat{A}_1|| \leq K^{-1}|\lambda_2(PG)|$. We further estimate the second above. Notice that $O_1 = \text{sgn}(\Xi_1'\Xi_1)$ shown up in Theorem 3.1. By $L_0\Xi_1 = \Xi_1\Lambda_1$, $L\Xi_1 = \Xi_1\hat{\Lambda}_1$, and sine-theta theorem,
\[
|v_k' (O_1'\Lambda_1 O_1 - \Lambda_1) v_k| \leq |v_k' (O_1 - \hat{\Xi}_1'\Xi_1)'\hat{\Lambda}_1 O_1 v_k| + |v_k' \Xi_1' (L - L_0)\hat{\Xi}_1 O_1 v_k| \\
+ |v_k' \Lambda_1 (O_1 - \hat{\Xi}_1'\Xi_1)'O_1 v_k| \\
\leq C|\lambda_2(PG)||O_1 - \hat{\Xi}_1'\Xi_1|| + K\|L - L_0\| \\
\leq C|\lambda_2(PG)|(K\beta_n^{-1}\|L - L_0\|^2 + K\|L - L_0\|) \\
\leq C\sqrt{K^3\log n \over n\theta^2\beta_n^2}
\]
where we also used $\|L - L_0\| \leq C\sqrt{\log(n)/n\theta^2}$ and $K^{\frac{1}{2}}\beta_n^{-1}\sqrt{\log(n)/n\theta^2} = o(1)$. As a consequence,
\[
{\frac{1}{(Pb_1)(k)}} - {\frac{1}{b_1(k)}} \leq C\sqrt{K^3\log n \over n\theta^2\beta_n^2}
\]
\[
(D.5)
\]
since $\lambda_2(PG) \leq C\sqrt{K}$.

Now, we are able to study $P\hat{\pi}_i^*$ and further $P\hat{\pi}_i$ combining (D.4) and (D.5). If $(P\hat{w}_i)(k) \leq 0$, trivially we have
\[
|(P\hat{\pi}_i^*)(k) - \pi_i^*(k)| = \pi_i^*(k) - w_i(k) \leq |\hat{w}_i(k) - w_i(k)|
\]
For the case that $(P\hat{w}_i)(k) > 0$, we get the bound
\[
|(P\hat{\pi}_i^*)(k) - \pi_i^*(k)| = \left|\frac{(P\hat{w}_i)(k) - w_i(k)}{(Pb_1)(k)} - \frac{1}{b_1(k)}\right| \leq \left|\frac{1}{(Pb_1)(k)}\right| \left|\frac{(P\hat{w}_i)(k) - w_i(k)}{b_1(k)}\right| \\
Moreover, taking sum over $k$ for both sides above,
\[
\|P\hat{\pi}_i^* - \pi_i^*\|_1 \leq C \max_k \left|\frac{1}{(Pb_1)(k)}\right| + \left|\frac{P\hat{w}_i - w_i}{b_1(k)}\right| \leq C\sqrt{K^3\log n \over n\theta(\theta + 1)\delta_n^2}
\]
Here we used the Cauchy-Schwarz inequality $\|P\hat{w}_i - w_i\|_1 \leq \sqrt{K}\|P\hat{w}_i - w_i\|$ and further applied (D.4) and (D.5). As a result,
\[
|(P\hat{\pi}_i)(k) - \pi_i(k)| = \left|\frac{(P\hat{\pi}_i)(k) - \pi_i^*(k)}{||P\hat{\pi}_i^*||_1} - \frac{\pi_i^*(k)}{||\pi_i^*||_1}\right| \leq \left|\frac{(P\hat{\pi}_i)(k) - \pi_i^*(k)}{||P\hat{\pi}_i^*||_1 ||\pi_i^*||_1}\right| + \left|\frac{(P\hat{\pi}_i)(k) - \pi_i^*(k)}{||\pi_i^*||_1}\right|
\]
And summing up over $k$ for both sides, we can further have
\[
\|P\hat{\pi}_i - \pi_i\|_1 \leq \frac{\|P\hat{\pi}_i^* - \pi_i^*\|_1}{\|\pi_i^*\|_1} \leq C \sqrt{\frac{K^3 \log n}{n\theta(\bar{\theta} \wedge \theta_i)\delta^2_n}}
\]
since $\|\pi_i^*\|_1 = \sum_k w_i(k)/b_1(k) \approx 1$ by $b_1(k) \approx 1$ for all $1 \leq k \leq K$. Therefore, we finished the proof.

D.2 Proofs of Corollary 3.2 and 3.3

The proofs of Corollary 3.2 and 3.3 are straightforward by employing Theorem 3.2. We shortly claim it below.

Proof of Corollary 3.2. Recall the definition of the $\ell^1$-loss $L(\hat{\Pi}, \Pi)$ in (1.5). Employing the node-wise errors in Theorem 3.2 and taking average, we see that
\[
L(\hat{\Pi}, \Pi) \leq C \sqrt{\log(n)} \int \min \left\{ \frac{\text{err}_n}{\sqrt{\ell \wedge 1}}, 1 \right\} dF_n(t),
\]
with probability $1 - o(n^{-3})$. Further by the trivial bound $L(\hat{\Pi}, \Pi) \leq 2$, we translate the high probability error rate into expected $\ell^1$-loss rate, i.e., (3.15) and conclude the proof.

Proof of Corollary 3.3. Recall the loss metric $L(\hat{\Pi}, \Pi; p, q)$ in (3.17). We crudely bound
\[
\|T\hat{\pi}_i - \pi_i\|_q^q \leq C_q \|\hat{\pi}_i - \pi_i\|_1^q
\]
where $C_q$ is some constant depending on $q$. We then use the error rate in Theorem 3.2 and conclude (3.3).

For the special case $p = 1/2$ and $q = 1$, we further bound
\[
L^w(\hat{\Pi}, \Pi) = \min_T \left\{ \frac{1}{n} \sum_{i=1}^n (\theta_i/\bar{\theta})^{1/2} \|T\hat{\pi}_i - \pi_i\|_1 \right\}
\]
\[
\leq \min_T \left\{ \frac{1}{n} \sum_{i \in S_1} (\theta_i/\bar{\theta})^{1/2} \|T\hat{\pi}_i - \pi_i\|_1 \right\} + \min_T \left\{ \frac{1}{n} \sum_{i \in S_2} (\theta_i/\bar{\theta})^{1/2} \|T\hat{\pi}_i - \pi_i\|_1 \right\}
\]
where we recall the definition of $S_1, S_2$ in (B.1). For the first term, we use Cauchy-Schwarz inequality and get
\[
\frac{1}{n} \sum_{i \in S_1} (\theta_i/\bar{\theta})^{1/2} \|T\hat{\pi}_i - \pi_i\|_1 \leq \left( \frac{1}{n} \sum_{i \in S_1} \theta_i/\bar{\theta} \right)^{1/2} \left( \frac{1}{n} \sum_{i \in S_1} \|T\hat{\pi}_i - \pi_i\|_1^2 \right)^{1/2} \leq C \sqrt{\log(n) \text{err}_n}
\]
with probability $1 - o(n^{-3})$. Plugging in the above inequality into (D.6), and applying the error rate in Theorem 3.2 separately for $i \in S_2$, one can easily obtain
\[
L^w(\hat{\Pi}, \Pi) \leq C \sqrt{\log(n) \text{err}_n}
\]
with probability $1 - o(n^{-3})$. Further with trivial bound $L^w(\hat{\Pi}, \Pi) \leq C$, we then conclude (3.19).
E Proofs of lower bounds

In this section, we complete the proofs of lower bounds, i.e., Theorems 3.3-3.4. To this end, we will show the proofs of Theorems 5.1-5.2 and Lemma 5.2 stated in Section 5. We organize this section as follows: In Section E.1, we provide the proof of Theorem 5.1 regarding weighted loss metric $L^w (\hat{\Pi}, \Pi)$. In Section E.2, we claim Lemma 5.2 and prove Theorem 5.2 under the condition (5.5). The proof of Theorem 5.2 with (5.5) violated is relatively simpler and we state it in Section E.3 for completeness. In the last subsection, Section E.4, we shortly show how to extend the lower bounds to $P$-specific case under some certain additional assumptions. This supports our arguments in the Remark in the end of Section 5.

Throughout this section, we will use $C_{p,k}$ to denote the index set collecting indices of the pure nodes in $k$-th community for $1 \leq k \leq K$.

E.1 Proof of Theorem 5.1

We begin with the proof of the first claim. We first verify $\Pi^*(j) \in Q^*_n$, for every $0 \leq j \leq J$ which are constructed in (5.2)-(5.4). By the definition of perturbation matrix $\Gamma(j)'s$ in (5.3), and the fact that $\gamma_n/\sqrt{\theta_i} \leq c_0/K$ for all $1 \leq i \leq n_0$ since $\theta_i/\bar{\theta} < \text{err}_n^2$ for all $1 \leq i \leq n_0$, it is easy to see that $\Pi^*(j)$’s are indeed membership matrices when choosing small $c_0$. Next, we check the regularity conditions (3.1)-(3.4). Note that (3.4) and the last inequality in (3.1) immediately hold because of the construction of $\Pi^*$. By definition, $G^*(j) = K(\Pi^*(j))'\Theta H_0^{-1}\Theta \Pi^*(j)$ and

$$
\|G^*(j) - G^*\|^2 \leq 2\|G^*\|_2^2 \gamma_n^2(\Pi^*)'\Theta^2 H_0^{-1}\Theta \Pi^*(j)\|_2^2 + \|\gamma_n^2(\Pi^*)'\Theta^2 H_0^{-1}\Theta \Pi^*(j)\|_2^2.
$$

(E.1)

Elementary computations lead to

$$
G^* = \left(\sum_{i=1}^{n_0} \frac{\theta_i^2}{H_0(i,i)}\right)^{1/2} K1_K + K\text{diag}\left(\sum_{i \in C_{p,1}} \frac{\theta_i^2}{H_0(i,i)}, \cdots, \sum_{i \in C_{p,K}} \frac{\theta_i^2}{H_0(i,i)}\right)
$$

According to our construction and assumptions on $\Pi^*$ and $\theta$, it can be derived from $\sum_{i=1}^{n} \theta_i^2/H_0(i,i) \asymp 1$ that

$$
K \sum_{i \in C_{p,k}} \frac{\theta_i^2}{H_0(i,i)} \asymp 1
$$

for all $1 \leq k \leq K$. It follows that $\|G^*\| \leq c_1$ and $\|(G^*)^{-1}\| \leq c_1$ for some constant $c_1$. Furthermore, one can also derive

$$
\|\gamma_n^2(\Pi^*)'\Theta^2 H_0^{-1}\Theta \Pi^*(j)\| \leq c_0^2 K^{-1}\text{err}_n^2 = o(1)
$$

(E.2)

following from $\sum_{i=1}^{n} \theta_i/H_0(i,i) \leq 1/\bar{\theta}$ and $|e_i \Pi^*(j)x| \leq \sqrt{K}$ for all $1 \leq i \leq n$ and any unit vector $x \in \mathbb{R}^K$. Therefore, by Weyl’s inequality and (E.1), we can conclude that the first
two inequalities in (3.1) hold for all $G^{(j)}$'s. Further with our choice of special $P^*$ which satisfies that $\lambda_1(P^*) = K$ and $\lambda_k(P^*) = \beta_n$ for all $2 \leq k \leq K$, (3.2) holds automatically for $P^*G^*$ and the first two inequalities in (3.1) hold for $P^*G^{(j)}$'s. The eigengap condition in (3.2) for $P^*G^{(j)}$'s follows from Weyl’s inequality and $\|P^*G^{(j)} - P^*G^*\| \leq c_{err} n$ for some constant $c$, which can be derived from (E.1). More precisely,

$$\lambda_1(P^*G^{(j)}) - |\lambda_2(P^*G^{(j)})| > (\lambda_1(P^*G^*) - c_{err} n) - (\lambda_2(P^*G^*) + c_{err} n) \geq CK - 2c_{err} n > C'K.$$ 

Lastly, we claim (3.3) holds for all $G^{(j)}$'s. Using Perron’s theorem, we obtain that the first right singular vector of $P^*G^{(j)}$ is positive for all $1 \leq j \leq J$. In particular, for $P^*G^*$, all of its entries are positive and of order 1, i.e., $\min_{i,j} P^*G^*(i,j)/\max_{i,j} P^*G^*(i,j) > c$ for some constant $c$. Then, $\eta_1^*$, the first right singular vector of $P^*G^*$, satisfies

$$\frac{\min_k \eta_1^j(k)}{\max_k \eta_1^j(k)} \geq \frac{\min_{i,j} P^*G^*(i,j) \sum_k \eta_1^j(k)}{\max_{i,j} P^*G^*(i,j) \sum_k \eta_1^j(k)} > c. \quad (E.3)$$

Thus $P^*G^*$ satisfies (3.3). For $P^*G^{(j)}$ with its the first right singular vector $\eta_1^{(j)}$, by sine-theta theorem, we have

$$\|\eta_1^{(j)} - \eta_1^*\| \leq \sqrt{2|\eta_1^{(j)}\eta_1^j - 1|} \leq CK^{-1}\|P^*G^* - P^*G^{(j)}\| \leq CK^{-1}err_n.$$ 

Then,

$$\frac{\min_k \eta_1^{(j)}(k)}{\max_k \eta_1^{(j)}(k)} > \frac{\min_k \eta_1^*(k) - CK^{-1}err_n}{\max_k \eta_1^*(k) + CK^{-1}err_n} > c \quad (E.4)$$

for some $c$, since $\eta_1^*(k) \asymp K^{-\frac{1}{2}}$ by (E.3) and we can choose sufficiently small $c_0$. We then conclude the proof of first statement.

Next, we proceed to prove the second statement, the pairwise difference between $\Pi^{(j)}$’s, under the weighted loss metric. By definition,

$$L^w(\Pi^{(j)}, \Pi^{(k)}) = \frac{1}{n} \sum_{i=1}^{n} (\theta_i/\bar{\theta})^{\frac{1}{2}} \|\pi_i^{(j)} - \pi_i^{(k)}\|_1 = \frac{c_0}{n} \sum_{i=1}^{n} \frac{\sqrt{K}}{\beta_n \sqrt{n\theta^2}} \|e_i(\Gamma^{(j)} - \Gamma^{(k)})\|_1$$

$$= \frac{c_0 \sqrt{K}}{\beta_n \sqrt{n\theta^2}} \cdot \frac{1}{n} \|\Gamma^{(j)} - \Gamma^{(k)}\|_1 \geq Cerr_n.$$ 

In the last part, we prove the third claim in Theorem 5.1 regarding the KL divergence statement. Note that

$$KL(\mathcal{P}_\ell, \mathcal{P}_0) = \sum_{1 \leq i < j \leq n} \Omega_{ij}^{(\ell)} \log(\Omega_{ij}^{(\ell)}/\Omega_{ij}^{(0)}) + (1 - \Omega_{ij}^{(\ell)}) \log \frac{1 - \Omega_{ij}^{(\ell)}}{1 - \Omega_{ij}^{(0)}}.$$ 

60
Notice that $\Omega^{(\ell)}_{ij} = \Omega^{(0)}_{ij}$ for $n_0 < i < j \leq n$. Only the pairs $0 < i < j \leq n_0$ and $0 < i \leq n_0 < j \leq n$ have the contributions. We then write $KL(P_i, P_0) = (I) + (II)$ where

\[(I) := \sum_{0<i<j\leq n_0} \Omega^{(\ell)}_{ij} \log(\Omega^{(\ell)}_{ij}/\Omega^{(0)}_{ij}) + (1 - \Omega^{(\ell)}_{ij}) \log \frac{1 - \Omega^{(\ell)}_{ij}}{1 - \Omega^{(0)}_{ij}} \quad (E.5)\]

\[(II) := \sum_{0<i\leq n_0<j\leq n} \Omega^{(\ell)}_{ij} \log(\Omega^{(\ell)}_{ij}/\Omega^{(0)}_{ij}) + (1 - \Omega^{(\ell)}_{ij}) \log \frac{1 - \Omega^{(\ell)}_{ij}}{1 - \Omega^{(0)}_{ij}} \quad (E.6)\]

We begin with the bound of $(I)$. For simplicity, we write $\Gamma^{(\ell)} = (\Gamma^{(\ell)}_1, \cdots, \Gamma^{(\ell)}_n)'$ for $\ell = 0, \cdots, J$. By definition, for $0 < i < j \leq n_0$,

$$\Omega^{(0)}_{ij} = \theta_i \theta_j \left( \frac{1}{K^2} 1_K^t P 1_K \right) = \theta_i \theta_j (1 - (1 - 1/K) \beta_n);$$

and

$$\Omega^{(\ell)}_{ij} = \theta_i \theta_j \left( \frac{1}{K} 1_K + \frac{\gamma_n}{\sqrt{\theta_i}} \Gamma^{(\ell)}_i \right)' P \left( \frac{1}{K} 1_K + \frac{\gamma_n}{\sqrt{\theta_j}} \Gamma^{(\ell)}_j \right) = \theta_i \theta_j (1 - (1 - 1/K) \beta_n) + \theta_i \theta_j \frac{\gamma_n^2 \beta_n}{\sqrt{\theta_i \theta_j}} (\Gamma^{(\ell)}_i)' \Gamma^{(\ell)}_j = \Omega^{(0)}_{ij} \left( 1 + \Delta^{(\ell)}_{ij} \right)$$

in which, $\Delta^{(\ell)}_{ij} := \gamma_n^2 / \sqrt{\theta_i \theta_j} \cdot \beta_n (\Gamma^{(\ell)}_i)' \Gamma^{(\ell)}_j / [1 - (1 - 1/K) \beta_n]$. Further notice that

$$\max_{0<i,j\leq n_0} |\Delta^{(\ell)}_{ij}| \leq C \frac{c_0^2 K^2}{\beta_n \sqrt{(n \theta_i)(n \theta_j)}} \leq C c_0^2 K^{-1} \beta_n$$

Choosing sufficiently small $c_0$, we have the Taylor expansions

$$\Omega^{(\ell)}_{ij} \log(\Omega^{(\ell)}_{ij}/\Omega^{(0)}_{ij}) = \Omega^{(0)}_{ij} (1 + \Delta^{(\ell)}_{ij}) \log (1 + \Delta^{(\ell)}_{ij}) = \Omega^{(0)}_{ij} \left( \Delta^{(\ell)}_{ij} + \frac{1}{2} (\Delta^{(\ell)}_{ij})^2 + O((\Delta^{(\ell)}_{ij})^3) \right)$$

and

$$(1 - \Omega^{(\ell)}_{ij}) \log \frac{1 - \Omega^{(\ell)}_{ij}}{1 - \Omega^{(0)}_{ij}} = (1 - \Omega^{(0)}_{ij} - \Omega^{(0)}_{ij} \Delta^{(\ell)}_{ij}) \log \left( 1 - \frac{\Omega^{(0)}_{ij}}{1 - \Omega^{(0)}_{ij}} \Delta^{(\ell)}_{ij} \right) = -\Omega^{(0)}_{ij} \Delta^{(\ell)}_{ij} + \frac{(\Omega^{(0)}_{ij})^2}{2(1 - \Omega^{(0)}_{ij})} (\Delta^{(\ell)}_{ij})^2 + O((\Omega^{(0)}_{ij} \Delta^{(\ell)}_{ij})^3)$$

Combining the above two equations together into (E.5), we arrive at

$$(I) \leq \sum_{0<i,j\leq n_0} \frac{\Omega^{(0)}_{ij}}{(1 - \Omega^{(0)}_{ij})} (\Delta^{(\ell)}_{ij})^2 = \sum_{0<i,j\leq n_0} \frac{\gamma_n^4 \beta_n^2 [(\Gamma^{(\ell)}_i)' \Gamma^{(\ell)}_j]^2}{1 - (1 - 1/K) \beta_n (1 - \Omega^{(0)}_{ij})}$$
and with parameters. By properly choosing $c$, we can derive
\[ K^3 / \beta_n^2 (n \theta^2) \leq c. \]

In the sequel, we turn to study (II). Since $0 < i \leq n_0 < j \leq n$, we suppose that $j \in C_{p,j}$ for some $1 \leq j \leq K$. Then,
\[
\Omega_{ij}^{(0)} = \theta_i \theta_j \left( \frac{1}{K} \mathbf{1}_K P e_j \right) = \theta_i \theta_j (1 - (1 / K) \beta_n);
\]
and
\[
\Omega_{ij}^{(\ell)} = \theta_i \theta_j \left( \frac{1}{K} \mathbf{1}_K + \frac{\gamma_n}{\sqrt{\theta_i}} \Gamma_i^{(\ell)} \right)^P e_j \nonumber \\
= \theta_i \theta_j (1 - (1 / K) \beta_n) + \theta_i \theta_j \frac{\gamma_n \beta_n}{\sqrt{\theta_i}} (\Gamma_i^{(\ell)})^P e_j \\
= \Omega_{ij}^{(0)} \left( 1 + \tilde{\Delta}_{ij}^{(\ell)} \right)
\]
with $\tilde{\Delta}_{ij}^{(\ell)} := \frac{\gamma_n}{\sqrt{\theta_i}} \cdot \beta_n (\Gamma_i^{(\ell)})^P e_j / (1 - (1 / K) \beta_n)$. Similarly, one can easily check that
\[
\max_{i,j} |\tilde{\Delta}_{ij}^{(\ell)}| \leq \frac{C \sqrt{K}}{\sqrt{n \theta_i}} \leq C \theta_n K^{-1} \beta_n.
\]

Therefore, in the same way as (E.7), we can derive
\[
(II) \leq \sum_{0 < i \leq n_0 < j \leq n} \frac{\Omega_{ij}^{(0)}}{(1 - \Omega_{ij}^{(0)})^2} (\tilde{\Delta}_{ij}^{(\ell)})^2 = \sum_{0 < i < j \leq n_0} \frac{\theta_i \theta_j \gamma_n^2 \beta_n^2 ((\Gamma_i^{(\ell)})^P e_j)^2}{[1 - (1 - 1 / K) \beta_n] (1 - \Omega_{ij}^{(0)})} \\
\leq C \left( \sum_{j = n_0 + 1}^n \theta_j \right) \gamma_n^2 \beta_n^2 n_0 \leq C \theta_n^2 n_0 K.
\]

We now combine (E.7) and (E.8). They imply that
\[
\sum_{\ell = 1}^J KL(\mathcal{P}_\ell, \mathcal{P}_0) \leq C \theta_n^2 J n K.
\]

Here $C$ is a constant independent of choice of $c_0$ and $n$. At the same time, since $J \geq 2^{[n_0/2] \times [K/2] / 8}$, we obtain that $\log J \geq c n K$ for some constant $c$ not relying on the other parameters. By properly choosing $c_0$, we finish the proof of the last claim. Further with standard techniques of lower bound analysis (e.g., [Tsybakov, 2009, Theorem 2.5]), we ultimately obtain the lower bound stated in Theorem 5.1.

### E.2 Proof of Lemma 5.2 and Theorem 5.2

**Proof of Lemma 5.2**. Recall Definition 3.1. For such $c_n$, $\gamma$ and $a_0$, we see that if $\tau_n(c_n, 1/8) \geq \gamma c_n$, then
\[
\int_{\tau_n(c_n, 1/8)}^{c_n} \frac{1}{\sqrt{t} \land 1} dF_n(t) \geq \frac{1}{8 \sqrt{c_n} \land 1} \tilde{F}_n(c_n) \geq \frac{\sqrt{\gamma}}{8 \sqrt{\gamma c_n} \land 1} \tilde{F}_n(c_n) \geq \frac{\sqrt{\gamma}}{8} \int_{\gamma c_n}^{c_n} \frac{1}{\sqrt{t} \land 1} dF_n(t)
\]

62
By Definition 3.1, we conclude
\[
\int_{\tau_n(c_n,1/8)}^{c_n} \frac{1}{\sqrt{t+1}} \, dF_n(t) \geq \tilde{a}_0 \int_{err_n}^{\infty} \frac{1}{\sqrt{t+1}} \, dF_n(t), \quad \tilde{a}_0 := \sqrt{\frac{r^*}{8}} a_0.
\]
In the case that \( \tau_n(c_n,1/8) < \gamma c_n \), trivially, (5.6) holds with \( \tilde{a}_0 = a_0 \).

With the help of Lemma 5.2, we are able to prove Theorem 5.2 below.

**Proof of Theorem 5.2.** Since the least-favorable configuration \( \Pi^* \) and \( \Pi^{(j)} \)'s are quite similar to those for the weighted loss metric, only with slightly different perturbation scales. This will not affect the regularity conditions. In fact, one can simply verify the regularity conditions in the same manner as the first part of the proof of Theorem 5.1. We thus conclude the first statement without details.

Next, for the pairwise difference under unweight loss metric, by definition,
\[
\mathcal{L}(\Pi^{(j)}, \Pi^{(k)}) = \frac{1}{n} \sum_{i=1}^{n} ||\pi_i^{(j)} - \pi_i^{(k)}||_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{\gamma_n}{\sqrt{\theta_i} \wedge \theta} ||\epsilon_i'(\Gamma^{(j)} - \Gamma^{(k)})||_1
\]
Note that \( ||H^{(j)} - H^{(k)}||_1 \geq [n_0/2] \times [K/2]/8 \). At least \( [n_0/2]/8 \) rows of \( H^{(j)} - H^{(k)} \) will contribute to the RHS term above. Since the construction of \( \Gamma^{(j)} \) based on \( H^{(j)} \), it is not hard to see that
\[
\frac{\gamma_n}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{\theta_i} \wedge \theta} ||\epsilon_i'(\Gamma^{(j)} - \Gamma^{(k)})||_1 \geq \frac{\gamma_n}{2n} \min_{M \subseteq M_0, |M| = n_0/8} \sum_{i \in M} \frac{K}{\sqrt{\theta_i} \wedge \theta}
\]
\[
\geq \frac{\gamma_n}{2n} \min_{M \subseteq M_0, |M| = n_0/8} \sum_{i \in M} \frac{K}{\sqrt{\theta_i} \wedge \theta}
\]
\[
\geq \int \min_{\{err_n, 1\}} \frac{1}{\sqrt{t+1}} \, dF_n(t)
\]
where the last step is due to (5.8). This concludes the second statement.

In the end, we briefly state the proof of the KL divergence bound since it is quite analogous to the counterpart proof of Theorem 5.1. We again define (I) and (II) as (E.5)–(E.6), and bound them separately. Thanks to the slight difference on the perturbation scale, one can simply mimic the proof of Theorem 5.1 and obtain the following bounds under current settings.

\[
(I) \leq \sum_{1 \leq i < j \leq n_0} \frac{\Omega_{ij}^{(0)}}{(1 - \Omega_{ij}^{(0)})}(\Delta^{(l)})^2 = \sum_{1 \leq i < j \leq n_0} \frac{\theta_i \wedge \theta_j \wedge \theta_j}{\theta_i \wedge \theta_j \wedge \theta_j} \frac{K^2}{\beta_n^2 n_0} \left[ (\Gamma_i^{(l)})'(\Gamma_j^{(l)}) \right]^2 \leq C_{n_0}^2 \beta_n^2 n_0 K^2 \leq C_{n_0}^2 n_0 K
\]
and
\[
(II) \leq \sum_{0 < i \leq n_0, j \leq n} \frac{\Omega_{ij}^{(0)}}{(1 - \Omega_{ij}^{(0)})}(\Delta^{(l)})^2 = \sum_{0 < i \leq n_0} \sum_{j \in C_{\rho, j}, 1 \leq j \leq K} \frac{\theta_i \wedge \theta_j \wedge \theta_j}{\theta_i \wedge \theta_j \wedge \theta_j} \frac{\gamma_n^2 \beta_n^2}{\beta_n^2 n_0} \left[ (\Gamma_i^{(l)})'(\epsilon_j) \right]^2 \leq C_{n_0}^2 \beta_n^2 n_0 K
\]
\[
\leq C \left( \sum_{j=n_0+1}^{n} \theta_j \right) \gamma^2 / \beta^2 n_0 \leq C c_0^2 n_0 K.
\]

Here to obtain the two upper bounds above, we used an estimate
\[
\sum_{1 \leq i \leq n_0} \frac{\theta_i}{\theta_i \wedge \bar{\theta}} \leq \sum_{1 \leq i \leq n_0} \left( 1 + \frac{\theta_i}{\bar{\theta}} \right) \leq 2n_0
\]
where the last step is owing to our ordering of \( \theta_i \)'s so that \( \sum_{1 \leq i \leq n_0} \theta_i / n_0 \leq \bar{\theta} \).

As a result,
\[
\frac{1}{J} \sum_{\ell=1}^{J} KL(P_\ell, P_0) \leq C c_0^2 n_0 K \leq \tilde{C} c_0^2 \log J.
\]

Properly choosing sufficiently small \( c_0 \), we thus complete the third statement. Furthermore, by standard techniques of lower bound analysis (e.g., (Tsybakov, 2009, Theorem 2.5)), we ultimately obtain the lower bound stated in Theorem 5.2.

\[\square\]

E.3 Proof of Theorem 5.2 without (5.5)

In this section, we show the proof of Theorem 5.2 in the case that (5.5) violates. We will need a distinct sequence of least-favorable configurations. We still order \( \theta_i \)'s as (5.7). But we define
\[
\frac{\theta_i}{n_0} = \max \{1 \leq i \leq n : \theta_i / \bar{\theta} \leq \text{err}^2_n \}
\]
which means \( n_0 \) is the total number of \( \eta_i \)'s such that \( 0 < \eta_i \leq \text{err}^2_n \). For the remaining \( n - n_0 \) nodes, we order them in the way that the average degrees of the pure nodes in different communities of \( \Pi^* \) are of the same order as before. And \( \Pi^* \) and \( \Gamma^{(0)}, \Gamma^{(1)}, \ldots, \Gamma^{(J)} \) are constructed in the same way as in (5.2)-(5.3). Different from (5.9), let
\[
\Pi^{(j)} = \Pi^* + c_0 K^{-1} \Gamma^{(j)}, \quad \text{for } 0 \leq j \leq J.
\]

First, following the first part of proof of Theorem 5.1, we see that \( G^*, P^* G^* \) satisfy the regularity conditions (3.1)- (3.4). Furthermore, one can derive
\[
\| c_0^2 K^{-1} (\Gamma^{(j)})' \Theta H_0^{-1} \Theta (\Gamma^{(j)}) \| \leq c_0^2 K^{-1}.
\]

Similarly to the analog in the proof of Theorem 5.1 (see (E.2)-(E.4)), by choosing properly small \( c_0 \), we have the regularity conditions hold for \( P^* G^{(j)} \)'s as well.

Second, under the construction (E.9),
\[
\mathcal{L}(\Pi^{(j)}, \Pi^{(k)}) = \frac{1}{n} \sum_{i=1}^{n} || \pi^{(j)}_i - \pi^{(k)}_i ||_1 = \frac{c_0}{nK} \sum_{i=1}^{n_0} || e'_i (\Gamma^{(j)} - \Gamma^{(k)}) ||_1
\]
\[
= \frac{4c_0}{nK} || H^{(j)} - H^{(k)} ||_1
\]

64
for some constant $C$ not relying on the other parameters. Notice that $n_0/n = \int_0^{err_2} dF_n(t)$. Since (5.5) violates, we thus conclude that

$$\mathcal{L}(\Pi^{(j)}, \Pi^{(k)}) \geq C \int_0^{err_2} dF_n(t) \geq C_2 \int \min\{\frac{err_n}{\sqrt{t + 1}}, 1\} dF_n(t). \quad (E.10)$$

for some constant $C_2 > 0$. Third, we claim the KL divergence in the same way as previously, $KL(\mathcal{P}_t, \mathcal{P}_0) = (I) + (II)$ and (I), (II) are defined in (E.5)-(E.6). By our least-favorable configurations (E.9), we bound

$$(I) \leq \sum_{1 \leq i < j \leq n} \frac{\Omega_{ij}^{(0)}}{(1 - \Omega_{ij}^{(0)})} (\Delta_{ij}^{(0)})^2 = \sum_{1 \leq i < j \leq n_0} \frac{\theta_i \theta_j c_0^4 K^{-4} \beta_n^2 [([\Gamma_i^{(0)}]' \Gamma_j^{(0)})^2]}{(1 - (1 - 1/K) \beta_n)(1 - \Omega_{ij}^{(0)})}$$

$$\leq C_0^4 \left( \sum_{i=1}^{n_0} \theta_i \right)^2 K^{-2} \beta_n^2 \leq C_0^4 n_0 K \frac{K^3}{\beta_n^2} \leq C_0^4 n_0 K$$

where in this case $\Delta_{ij}^{(0)} = c_0^2 K-2 \beta_n (\Gamma_i^{(0)})' \Gamma_j^{(0)}/(1 - (1 - 1/K) \beta_n)$, and we used the fact that $\theta_i \leq K^3 \beta_n^2/(n \theta)$ for all $1 \leq i \leq n_0$ to obtain the second inequality on the second row; and

$$(II) \leq \sum_{0 < i \leq n_0, j \leq n} \frac{\Omega_{ij}^{(0)}}{(1 - \Omega_{ij}^{(0)})} (\Delta_{ij}^{(0)})^2 = \sum_{0 < i \leq n_0, 1 \leq j \leq K} \frac{\theta_i \theta_j c_0^2 K^{-2} \beta_n^2 [([\Gamma_i^{(0)}]' e_j)]^2}{(1 - (1 - 1/K) \beta_n)(1 - \Omega_{ij}^{(0)})}$$

$$\leq C_0^2 \left( \sum_{i=1}^{n_0} \theta_i \right) \left( \sum_{j=n_0+1}^{n} \theta_j \right) K^{-2} \beta_n^2 \leq C_0^2 n_0 K$$

where $\Delta_{ij}^{(0)} := c_0 K^{-1} \beta_n (\Gamma_i^{(0)})' e_j / (1 - (1 - 1/K) \beta_n)$ for this case. Combining the upper bounds for (I) and (II), we finally get

$$\sum_{1 \leq i \leq J} KL(\mathcal{P}_t, \mathcal{P}_0) \leq C_0^2 Jn_0 K \leq (1/8 - \epsilon_1) J \log(J) \quad (E.11)$$

for a constant $\epsilon_1 \in (0, 1/8)$, by choosing sufficiently small $c_0$ and noting $n_0 K \asymp \log(J)$.

In conclusion, we proved the analogs of the three claims in Theorem 5.2 when (5.5) violates. Further by standard techniques of lower bound analysis (e.g., Tsybakov 2009, Theorem 2.5), we ultimately obtain the lower bound.

### E.4 Extension to $P$-specific lower bounds

In this subsection, we claim the $P$-specific lower bounds of $\mathcal{L}^w(\tilde{\Pi}, \Pi)$ and $\mathcal{L}(\tilde{\Pi}, \Pi)$ for arbitrary $P$ if one of the following condition holds as $n \to \infty$:

(a) $(K, P)$ are fixed;
(b) \((K, P)\) can depend on \(n\), but \(K \leq C\) and \(P1_K \propto 1_K\);

(c) \((K, P)\) can depend on \(n\), and \(K\) can be unbounded, but \(P1_K \propto 1_K\) and \(|\lambda_2(P)| \leq C\beta_n = o(1)\).

Since the proofs are quite analogous to the case of the special \(P\) in the manuscript, in the sequel, we point out the key differences compared to the full arguments for Theorems 3.3-3.4, and shortly state how to adapt the proofs in the previous subsections to the current cases.

(a) If \(K = K_0\) and \(P = P_0\), for a fixed integer \(K_0 \geq 2\) and a fixed matrix \(P_0\), we can simplify the construction of \(\Gamma^{(j)}\)'s and hence the configurations \(\Pi^{(j)}\)'s. More specifically, we apply Lemma 5.1 to \(n_0\) to get \(\omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(J)}\), where \(J \geq 2^{n_0/8}\). We insert \(\omega^{(j)}\)'s into \(n\)-dim vectors \(\gamma^{(j)}\)'s such that

\[
(\gamma^{(j)})' = ((\omega^{(j)})', 0_{1 \times (n-n_0)}).
\]  

(E.12)

Let \(\eta \in \mathbb{R}^{K_0}\) be a nonzero vector such that

\[
\eta'1_{K_0} = 0, \quad \eta'P_01_{K_0} = 0, \quad \|\eta\|_1 \asymp K_0
\]  

(E.13)

Such \(\eta\) always exists by solving certain linear system. Based on these notations, we re-define \(\Gamma^{(j)} = \gamma^{(j)}\eta\)' and re-define \(\Pi^{(j)}\) correspondingly as \([5.4, 5.9, E.9]\) for unweighted loss, \(([5.4, 5.9, E.9])\) for weighted loss, \(5.1, 5.2\).

The verification of regularity conditions and pairwise difference between the configurations can be claimed in the same way as in the proofs of Theorems 5.1-5.2. The most distinguishing part appears in the KL divergence. Especially, for \(0 < i < j \leq n_0\),

\[
\Omega^{(\ell)}_{ij} = \Omega^{(0)}_{ij} \left(1 + \Delta^{(\ell)}_{ij}\right), \quad \Delta^{(\ell)}_{ij} \propto \Gamma^{(\ell)}(i)\Gamma^{(\ell)}(j)\eta'P\eta
\]  

(E.14)

where the coefficients we did not specify for \(\Delta^{(\ell)}_{ij}\)'s rely on the perturbation scale we take from \([5.4, 5.9, E.9]\). Similarly for \(0 < i \leq n_0 < j \leq n\), if \(j \in \mathcal{C}_{p,j}\),

\[
\Omega^{(\ell)}_{ij} = \Omega^{(0)}_{ij} \left(1 + \tilde{\Delta}^{(\ell)}_{ij}\right), \quad \tilde{\Delta}^{(\ell)}_{ij} \propto \Gamma^{(\ell)}(i)\eta'P\hat{e}_j.
\]  

(E.15)

Nevertheless, in this case, \(\eta'P\eta \asymp 1\) and \(\eta'P\hat{e}_j \asymp 1\). In particular, \(\beta_n \asymp 1\). All of these facts lead to similar derivations on upper bounds of (I) and (II) (see definitions in \([E.5, E.6]\)). One can claim the desired upper bounded for KL divergence for the least-favorable configurations we constructed here. One can conclude the proof by mimicking the proofs of Theorems 5.1-5.2.

(b) If both \((K, P)\) may depend on \(n\), but they satisfy that \(K \leq C\) and \(P1_K \propto 1_K\). We take the same simplified least-favorable configurations as in Case (a). The regularity conditions and pairwise difference can be claimed likewise. \(1_K\) is an eigenvector of \(P\). We can take special \(\eta\), the eigenvector associated to the smallest eigenvalue (in magnitude) of \(P\). In \((E.14, E.15)\), we have \(\eta'P\eta \asymp \beta_n\) and \(\eta'P\hat{e}_j \asymp \beta_n\), which fit the arguments in the proofs of KL divergence for Theorems 5.1-5.2. Thereby, we can prove the KL divergence in the same way as the proofs of Theorems 5.1-5.2.
(c) If both $(K, P)$ may depend on $n$ and $K$ can be unbounded, but $P1_K \propto 1_K$ and $|\lambda_2(P)| \leq C\beta_n = o(1)$. We adopt the same least-favorable configurations in Section 5 correspondingly to Theorems 5.1, 5.2. The different parts only show up in the quantities involving $P$. Notice that in this case, $1_K$ is the eigenvector associated to the largest eigenvalue of $P$,

\[(\Gamma_i^{(t)})'P1_K \propto (\Gamma_i^{(t)})'1_K = 0, \quad (\Gamma_i^{(t)})'P\Gamma_j^{(t)} \asymp \beta_n \quad \text{(E.16)}\]

since the other eigenvalues of $P$ are asymptotically of order $\beta_n$. These two estimates exactly coincide with the ones in the proofs of Theorems 5.1, 5.2. Then, all the arguments in the proofs of Theorems 5.1, 5.2 can be directly applied in this setting. We thereby conclude our proof.

### E.5 Further remark on Definition 3.1

In Section 3.3 of the manuscript, we introduce a technical condition on $F_n(\cdot)$ (see Definition 3.1). And we remark in the manuscript that this condition only excludes those $F_n(\cdot)$ that have extremely ill behavior in the neighborhood of 0, which rarely appears in reality. Especially for $\theta_i$’s i.i.d. generated from $\kappa_n F(\cdot)$, where $\kappa_n > 0$ is a scalar and $F(\cdot)$ is fixed distribution that is either continuous or discrete with finite mean $m$, we provide some arguments that Definition 3.1 holds below.

- If $F(\cdot)$ is a discrete distribution, i.e., $F = \sum_{\ell=1}^L \epsilon_\ell \delta_{x_\ell}$ where $L$ is a fixed constant and $0 < x_1 < x_2 < \ldots < x_L$, $\epsilon_\ell$’s are all fixed, $\delta_{x}$ is a point mass at $x$, and $\sum_{\ell=1}^L \epsilon_\ell x_\ell = 1$. In this case, we simply set $c_n = x_{L-1}/m$, $\gamma = x_{L-1}/x_1$ and $a_0 = \min_\ell \epsilon_\ell$. One can easily check

\[
F_n(c_n) = F(x_{L-1}) = 1 - \epsilon_L \leq 1 - a_0,
\]

\[
\sum_{\ell=1}^{L-1} \frac{\epsilon_\ell}{\sqrt{m^{-1}x_\ell \wedge 1}} \geq (1 - \epsilon_L) \sum_{\ell=1}^L \frac{\epsilon_\ell}{\sqrt{m^{-1}x_\ell \wedge 1}} \geq a_0 \sum_{\ell=1}^L \frac{\epsilon_\ell}{\sqrt{m^{-1}x_\ell \wedge 1}}
\]

which indeed verify the condition in Definition 3.1.

- If $F(\cdot)$ is a continuous distribution with density $f(\cdot)$ and $\text{supp}(F) \subset [0, +\infty)$. Since $\int t f_n(t) = 1$, it is not hard to see that $dF_n(t) = m f(mt) dt$. We can rewrite

\[
\int_{\text{err}_n^2}^\infty \frac{1}{\sqrt{t} \wedge 1} dF_n(t) = \int_{\text{err}_n^2 m}^\infty \frac{f(t)}{\sqrt{t/m} \wedge 1} dt \quad \text{(E.17)}
\]

The singularity of the above RHS integral lies in the neighborhood of 0, or $\text{err}_n^2/m$.

If $f(t)t^{\frac{1}{2} - \epsilon_0} \leq c$ as $t \to 0$ for some $\epsilon_0 > 0$, $c > 0$, then the integral on the RHS of (E.17) converges and can be bounded by some constant $C_1 > 0$. Since $F(\cdot)$ is a fixed continuous distribution with finite mean $m$, we can always find $\bar{c} > 0$, $\bar{a} \in (0, 1)$ and $\gamma \in (0, 1)$ such that $F(\bar{c}) - F(\gamma \bar{c}) > C_2$ and $F(\bar{c}_n) \leq 1 - \bar{a}$ for some constant $0 < C_2 < C_1$. We then set $c_n = \bar{c}/m$ and $a_0 = \min\{\bar{a}, C_2/C_1\}$. As a result,

\[
F_n(c_n) = F(\bar{c}) \leq 1 - \bar{a} \leq 1 - a_0,
\]

67
\[
\int_{c_n}^{\infty} \frac{1}{\sqrt{t \wedge 1}} dF_n(t) \geq \frac{F_n(\tilde{c}_n) - F_n(\gamma \tilde{c}_n)}{\sqrt{c_n \wedge 1}} \geq C_2 \geq a_0 \int_{err_2^m}^{\infty} \frac{1}{\sqrt{t \wedge 1}} dF_n(t).
\]

We remark that the case \( F(\cdot) \) has a support bounded below from zero is also included in the current discussion.

If \( f(t) t^{1/2 + \epsilon_0} \geq c \) as \( t \to 0 \) for some \( \epsilon_0 > 0 \) and \( c > 0 \), the mass of the integral on the RHS of (E.17) should concentrate in the neighborhood of \( err_2^m \). Therefore, we can simply set \( c_n = Cerr_2^2 \) for some large \( C > 1 \) such that \( F(Cerr_2^2 m) \leq 1 - \tilde{a} \) for some \( \tilde{a} > 0 \) (this can be always achieved since \( F(\cdot) \) is a fixed distribution with mean \( m \)). Let \( \gamma = C^{-1} \), then

\[
F_n(c_n) = F(Cerr_2^2 m) \leq 1 - \tilde{a}, \quad \int_{\gamma c_n}^{c_n} \frac{1}{\sqrt{t \wedge 1}} dF_n(t) = \int_{err_2^2 m}^{Cerr_2^2 m} \frac{f(t)}{\sqrt{t/m}} dt > C_3 \int_{err_2^2 m}^{\infty} \frac{f(t)}{\sqrt{t/m \wedge 1}} dt
\]

for some \( C_3 > 0 \). We thus take \( a_0 = \min\{\tilde{a}, C_3\} \) and the condition in Definition 3.1 is satisfied.