Supplement to “Optimal Network Membership Estimation under Severe Degree Heterogeneity”

Zheng Tracy Ke
Department of Statistics, Harvard University

and

Jingming Wang
Department of Statistics, Harvard University

August 3, 2024

Contents

A Supplementary lemmas for the main text

In the main paper, owing to the space constraints, we stated some arguments without giving detailed proofs. In this section, we revisit and prove these arguments.

A.1 The simplex geometry and the oracle procedure

In Section 3, we considered the oracle case where $A = \Omega$ and claimed that there is a simplex geometry associated with the rows of $\hat{R}$. The next lemma makes this argument rigorous:

Lemma A.1 (The simplex geometry)**.** *Consider a DCMM model, where each community k has at least one pure node. Let $H_0 = \mathbb{E}[H]$ and $L_0 = H_0^{-\frac{1}{2}}\Omega H_0^{-\frac{1}{2}}$. Let λ_k be the kth largest eigenvalue (in magnitude) of L_0, and let ξ_k be the corresponding eigenvector. If we pick the sign of ξ_1 such that $\sum_{i=1}^n \xi_1(i) > 0$, then ξ_1 is a strictly positive vector. Furthermore, consider the matrix $R \in \mathbb{R}^{n\times(K-1)}$, where $R(i,k) = \xi_{k+1}(i)/\xi_1(i)$, $1 \le i \le n, 1 \le k \le K-1$. Write $R = [r_1, r_2, \dots, r_n]'$.*

- *There exists a simplex $\mathcal{S} \subset \mathbb{R}^{K-1}$ with K vertices $v_1, v_2, \ldots, v_K$, such that $r_1, r_2, \ldots, r_n$ are contained in $\mathcal{S}$. If node i is a pure node, then r_i falls on one vertex of this simplex; if node i is a mixed node, then r_i is in the interior of the simplex (it can be on an edge or a face, but cannot be 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 $w_i = \|\pi_i \circ b_1\|_1^{-1}(\pi_i \circ b_1)$, where $\circ$ is the Hardarmart product and b_1 is a K-dimensional vector with $b_1(k) = 1/\sqrt{\lambda_1 + v_k' \mathrm{diag}(\lambda_2, \ldots, \lambda_K) v_k}$, $1 \le k \le K$.*

Proof of Lemma A.1. The proof largely follows the one in [6], except that they considered a special case of $H = I_n$ while we allow for a general diagonal matrix H here.

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 $1 \le k \le K$, $1 \le \ell \le K-1$. Write $V = [v_1, v_2, \ldots, v_K]' \in \mathbb{R}^{K \times (K-1)}$. It follows that

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

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

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

Define $W = [\mathrm{diag}(\xi_1)]^{-1} H_0^{-\frac{1}{2}} \Theta \Pi \mathrm{diag}(b_1)$. The above equation implies that $\mathbf{1}_n = W \mathbf{1}_K$ and $R = WV$. Denote by w_i' the ith row of W. It follows that $w_i' \mathbf{1}_K = 1$ and $r_i = \sum_{k=1}^K w_i(k) v_k$.

Furthermore, under Condition 2.1(c), we can show that both ξ_1 and b_1 are strictly positive vectors; the proof is similar to the proof of Lemma B.4 of [6], 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 π_i. Write $\alpha_i = \xi_1^{-1}(i) H_0^{-\frac{1}{2}}(i,i)\theta_i$. Then, $w_i' = \alpha_i \cdot \pi_i' \mathrm{diag}(b_1) = \alpha_i \cdot (\pi_i \circ b_1)$. 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

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

where D_θ and G are as defined in Section 2 and we note that D_θ 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

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

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

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

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 of DCMM model in Section 2.1 says that $P(k,k) = 1$. Therefore, $b_1(k) = 1/\sqrt{\lambda_1 + v_k'\Lambda_1 v_k}$. □

A.2 Broadness of the θ-class $\mathcal{G}(\varrho, a_0)$

In Section 2, we introduced a technical condition on $F_n(\cdot)$ (see Definition 2.2) and defined $\mathcal{G}(\varrho, a_0)$, a class of θ. We claimed that this class is broad enough to include most interesting cases of degree heterogeneity. This is justified by the following lemma:

Lemma A.2. *The requirements in Definition 2.2 are satisfied if θ_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 which has its support in $(0, \infty)$ and satisfies one of the following conditions:*

- *$F(\cdot)$ is a discrete distribution;*
- *$F(\cdot)$ is a continuous distribution with support in $[c, \infty)$, for some $c > 0$;*
- *$F(\cdot)$ 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 $C > 0$.*

Proof of Lemma A.2. Recall that we assume θ_i's i.i.d. generated from $\kappa_n F(\cdot)$, where $\kappa_n > 0$, and $F(\cdot)$ is fixed distribution that is either continuous or discrete with finite mean m.

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

$$
\begin{aligned}
&F_n(c_n) = F(x_{L-1}) = 1 - \epsilon_L \leq 1 - a_0, \\
&\sum_{\ell=1}^{L-1} \frac{\epsilon_\ell}{\sqrt{m^{-1}x_\ell \wedge 1}} \geq (1-\epsilon_L) \sum_{\ell=1}^{L} \frac{\epsilon_\ell}{\sqrt{m^{-1}x_\ell \wedge 1}} \geq a_0 \sum_{\ell=1}^{L} \frac{\epsilon_\ell}{\sqrt{m^{-1}x_\ell \wedge 1}}
\end{aligned}
$$

which indeed verify the condition in Definition 2.2 for the chosen (ρ, a_0). This proves the first bullet point of Lemma A.2.

Next, we consider the case that $F(\cdot)$ is a continuous distribution with density $f(\cdot)$ and supp(f) $\subset [0, +\infty)$. Since $\int t dF_n(t) = 1$, it is not hard to see that $dF_n(t) = mf(mt)dt$. We can rewrite

$$
\int_{err_n^2}^{\infty} \frac{1}{\sqrt{t \wedge 1}} dF_n(t) = \int_{err_n^2 m}^{\infty} \frac{f(t)}{\sqrt{t/m \wedge 1}} dt \tag{A.1}
$$

The singularity of the integral on the RHS of (A.1) lies in the neighborhood of 0, or $err_n^2 m$.

If $f(t)t^{\frac{1}{2}-\epsilon_0} \to C$ as $t \to 0$ for some $\epsilon_0 > 0$, $C > 0$, then the integral on the RHS of (A.1) 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 $\tilde{c} > 0$, $\tilde{a} \in (0, 1)$ and $\rho \in (0, 1)$ such that $F(\tilde{c}) - F(\rho\tilde{c}) > C_2$ and $F(\tilde{c}) \le 1 - \tilde{a}$ for some constant $0 < C_2 < C_1$. We then set $c_n = \tilde{c}/m$ and $a_0 = \min\{\tilde{a}, C_2/C_1\}$. As a result,

$$F_n(c_n) = F(\tilde{c}) \le 1 - \tilde{a} \le 1 - a_0,$$
$$\int_{\rho c_n}^{c_n} \frac{1}{\sqrt{t \wedge 1}} dF_n(t) \ge \frac{F_n(\tilde{c}_n) - F_n(\rho\tilde{c}_n)}{\sqrt{c_n \wedge 1}} \ge C_2 \ge a_0 \int_{err_n^2}^{\infty} \frac{1}{\sqrt{t \wedge 1}} dF_n(t).$$

Here err_n can be replaced by any other sequence $x_n \to 0$. We remark that the case $F(\cdot)$ has a support bounded below from zero is also included in the current discussion. This proves the second bullet point in Lemma A.2 and part of the third bullet point.

If $f(t)t^{\frac{1}{2}+\epsilon_0} \to C$ as $t \to 0$ for some $\epsilon_0 > 0$ and $C > 0$, then the RHS of (A.1) is of the order $err_n^{-2\epsilon_0}$ and its mass is located in the neighborhood of $err_n^2 m$. Therefore, we can simply set $c_n = C_3\, err_n^2$ for some large $C_3 > 1$ such that $F(C_3 err_n^2 m) \le 1 - \tilde{a}$ for some $\tilde{a} > 0$ (this can be always achieved since $F(\cdot)$ is a fixed distribution with mean m). Let $\varrho = C_3^{-1}$. Then,

$$F_n(c_n) = F(C_3\, err_n^2\, m) \le 1 - \tilde{a},$$
$$\int_{\varrho c_n}^{c_n} \frac{1}{\sqrt{t \wedge 1}} dF_n(t) = \int_{err_n^2\, m}^{C_3\, err_n^2\, m} \frac{f(t)}{\sqrt{t/m}} dt > C_4 \int_{err_n^2 m}^{\infty} \frac{f(t)}{\sqrt{t/m \wedge 1}} dt,$$

for some $C_4 > 0$. We thus take $a_0 = \min\{\tilde{a}, C_4\}$. The arguments also hold if we replace err_n by any other sequence $x_n \to 0$. The condition in Definition 2.2 is satisfied for the chosen (ρ, a_0). This proves the remaining part of the third bullet point. □

Algorithm B.1: Mixed-SCORE-Laplacian.

Input: K, A, tuning parameters $(\tau, c, \gamma) = (1, 0.5, 0.05)$ (default), and a given VH algorithm.

1. Let L be the normalized graph Laplacian in (9). Let $\hat{\lambda}_k$ be the kth largest eigenvalue (in magnitude) of L, and let $\hat{\xi}_k$ be the associated eigenvector, $1 \leq k \leq K$. Define an $n \times (K-1)$ matrix $\hat{R}$ by

$$\hat{R}(i,k) = \hat{\xi}_{k+1}(i)/\hat{\xi}_1(i), \qquad 1 \leq i \leq n,\ 1 \leq k \leq K-1.$$

   Denote by $\hat{r}_1', \hat{r}_2', \ldots, \hat{r}_n'$ the rows of $\hat{R}$.

2. Let $\hat{\delta}_n = K|\hat{\lambda}_K|$ and $\hat{S}_n(c) = \{1 \leq i \leq n : d_i\hat{\delta}_n^2 \geq cK^3\log(n)\}$. For any $i \notin \hat{S}_n(c)$, set $\hat{\pi}_i = K^{-1}\mathbf{1}_K$.

3. Let $\hat{S}_n^*(c,\gamma) = \hat{S}_n(c) \cap \{1 \leq i \leq n : d_i \geq \gamma\bar{d}\}$. Run the given VH algorithm on the point cloud $\{\hat{r}_i\}_{i \in \hat{S}_n^*(c,\gamma)}$. Denote the output by $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_K$.

4. Let $\hat{\Lambda}_1 = \mathrm{diag}(\hat{\lambda}_2, \ldots, \hat{\lambda}_K)$ and obtain $\hat{b}_1 \in \mathbb{R}^K$ from

$$\hat{b}_1(k) = [\hat{\lambda}_1 + \hat{v}_k'\hat{\Lambda}_1\hat{v}_k]^{-1/2}, \qquad 1 \leq k \leq K,$$

   For each $i \in \hat{S}_n(c)$, solve $\hat{w}_i \in \mathbb{R}^K$ from the linear equation set:

$$\sum_{k=1}^K \hat{w}_i(k)\hat{v}_k = \hat{r}_i, \qquad \text{and} \qquad \sum_{k=1}^K \hat{w}_i(k) = 1.$$

   Let $\hat{\pi}_i^* \in \mathbb{R}^K$ be such that $\hat{\pi}_i^*(k) = \max\{\hat{w}_i(k)/\hat{b}_1(k),\, 0\}$, for $1 \leq k \leq K$. Output $\hat{\pi}_i = \hat{\pi}_i^*/\|\hat{\pi}_i^*\|_1$, for each $i \in \hat{S}_n(c)$.

Output: $\hat{\Pi}$.

B The Mixed-SCORE-Laplacian (MSL) algorithm

In Section 3, we explained the membership estimation steps in Figure 2 and gave a high-level description of the MSL algorithm in Algorithm 1. We now present Algorithm B.1, a detailed version of Algorithm 1. In this algorithm, we assume there is a given vertex hunting (VH) algorithm. The choices of the VH algorithm are discussed in Section B.1.

Algorithm B.2: Successive projection (SP).

Input: K and $x_1, x_2, \dots, x_m \in \mathbb{R}^d$ $(d \geq K)$.

For each $1 \leq k \leq K$, run the following steps:

- If $k \geq 2$, compute $\mathcal{P} = Y(Y'Y)^{-1}Y'$, where $Y = [x_{i_1}, x_{i_2}, \dots, x_{i_{k-1}}] \in \mathbb{R}^{d \times (k-1)}$.
- If $k \geq 1$, find $i_1 = \text{argmax}_i \|x_i\|$; otherwise, find $i_k = \text{argmax}_i \|(I_d - \mathcal{P})x_i\|$.

Output: $\hat{v}_k = x_{i_k}$, for $1 \leq k \leq K$.

Algorithm B.3: Sketched vertex search (SVS).

Input: K, $x_1, x_2, \dots, x_m \in \mathbb{R}^d$, and a tuning integer $L \geq K$.

- Run k-means clustering on $x_1, x_2, \dots, x_m$, assuming there are L clusters. Denote by $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_L$ the estimated cluster centers.
- Input $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_L$ to Algorithm B.2 to obtain $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_K$.

Output: $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_K$.

B.1 Choices of the plug-in VH algorithm

In our theoretical analysis and most simulations, we use successive projection (SP) [2] as the plug-in VH algorithm. The details of SP are presented in Algorithm B.2. When plugging this algorithm into Algorithm B.1, we need to pay attention to the dimension: Algorithm B.2 requires that the input point cloud is in a dimension $d \geq K$. However, each $\hat{r}_i$ is in dimension $K - 1$. To resolve this issue, we follow [6] to let

$$x_i = (1,\ \hat{r}_i')', \qquad 1 \leq i \leq n.$$

Now, each x_i is in dimensional K. We input x_i's to Algorithm B.2. The output $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_K$ will also be in dimension K. Since the first entry of each x_i is 1, the first entry of each $\hat{v}_k$ is also 1. We then remove this first entry and output the $(K-1)$-dimensional sub-vectors of $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_K$. As argued in [6], this has no effect on the vertex estimation accuracy.

SP performs well when the noise level in $x_1, x_2, \dots, x_m$ is relatively low. When the noise level is relatively high, [6] recommended to 'de-noise' before running SP. They proposed the sketched vertex search (SVS) algorithm, which uses k-means to denoise. The details of SVS can be found in Algorithm B.3. We also refer the readers to [7, Section 3.4] and [5] for more options of VH algorithms. In our simulation studies, we use SVS only in Experiment 2. This experiment studies the node-wise errors. Compared to the ℓ^1-loss used in other experiments, node-wise errors are more sensitive to the noise level. This motivates us to replace SP by SVS. SVS has one tuning integer L, which is set as $L = 5$ in Experiment 2.

Remark: In Algorithm B.1, we apply the plug-in VH algorithm on the trimmed point cloud: $\{\hat{r}_i : d_i \geq \gamma \bar{d}\}$. Our rationale is that the noise level on low-degree nodes is too high, so these points should be trimmed to improve performance. This trimming can also be viewed as a 'denoising' step. Therefore, when we plug SVS into Algorithm B.1, we actually conduct two rounds of denoising (trimming and k-means) before running SP.

C Additional simulation results

In this section, we present the additional simulation results, which are omitted in the main paper due to the space constraint.

C.1 The weighted ℓ^1-loss

We recall that most results in the main text are for the unweighted ℓ^1-loss. In Section 4.3, we extend the theoretical results to a general loss function parametrized by p and q. A special case of $p = 1/2$ and $q = 1$ is called the weighted ℓ^1-loss:

$$\mathcal{L}^w(\hat{\Pi}, \Pi) = \min_T \left\{ \frac{1}{n} \sum_{i=1}^n \sqrt{\frac{\theta_i}{\bar{\theta}}} \, \|T\hat{\pi}_i - \pi_i\|_1 \right\}. \tag{C.1}$$

Compared to the unweighted ℓ^1-loss, this performance metric down-weights those errors in low-degree nodes. In Figure 3, we report the unweighted ℓ^1-loss of MSL and Mixed-SCORE in four different cases of degree heterogeneity and various levels of network sparsity. We now provide more results of these simulations by reporting the weighted ℓ^1-loss in Figure C.1.

The conclusions for the weighted ℓ^1-loss are similar to those for the unweighted ℓ^1-loss:

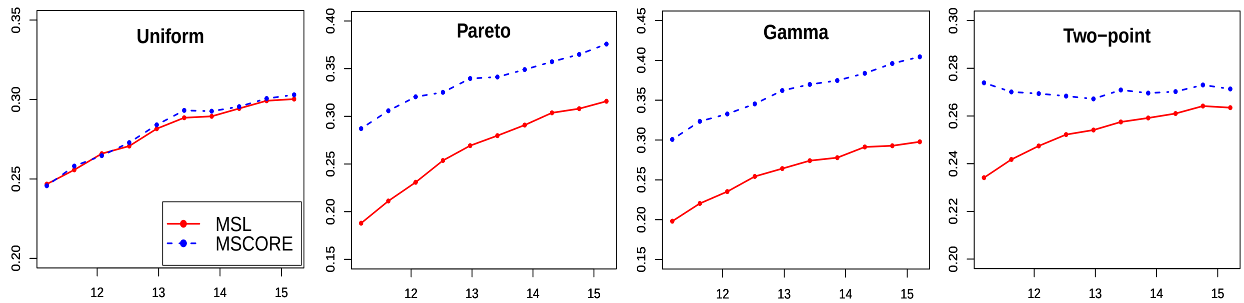

Figure C.1: MSL v.s. MSCORE ($n = 2000$, $K = 2$, x-axis is $\sqrt{n\bar{\theta}^2}$, and y-axis represents the unweighted ℓ^1-loss in (C.1)). This figure complements Figure 3 in the main text by reporting the unweighted ℓ^1-loss.

MSL greatly improves the conventional MSCORE in the last three cases, Pareto, Gamma, and Two-point mixture, which are the cases of severe degree heterogeneity. Furthermore, if we compare each panel in Figure C.1 with the corresponding panel in Figure 3, we find that the value of the weighted loss is significantly smaller than the value of the unweighted loss in the case of Pareto and Gamma. This is because the node-wise error is a decreasing function of θ_i; when low-degree nodes are down-weighted and high-degree nodes are up-weighted, the loss will decrease.

C.2 Other values of K

The simulation experiments in Section 6 focus on $K = 2$. We now consider other values of K and investigate how the performance of MSL changes with K.

Fix $n = 5000$ and let K range in $\{3, 4, \ldots, 10, 11\}$. Given any $\beta_n \in (0, 1)$ and $b_n > 0$, we let $P = \beta_n I_K + (1 - \beta_n)\mathbf{1}_K\mathbf{1}_K'$ and generate θ as follows: Draw $\theta_1^0, \theta_2^0, \ldots, \theta_n^0 \overset{iid}{\sim} \text{Uniform}(0, 1)$ and let $\theta_i = b_n \cdot n\theta_i^0/\|\theta^0\|_1$ for $1 \leq i \leq n$. We remark that this is a severe-degree-heterogeneity case: Since the support of Uniform$(0, 1)$ contains the neighborhood near zero, $\theta_{\max}/\theta_{\min}$ can be potentially large. We generate Π in the same way as in Experiments 1-2: Set $\pi_i = (1, 0)'$ and $\pi_i = (0, 1)'$ each for 15% of nodes, and let $\pi_i = (t_i, 1 - t_i)'$ for the remaining 70% of nodes, with $t_i \overset{iid}{\sim} \text{Uniform}(0, 1)$. Write SNR $:= \sqrt{n\bar{\theta}^2}(1 - P(1, 2)) = b_n\beta_n\sqrt{n}$. We set $b_n (= \bar{\theta})$ such that $nb_n^2 = 800$ and select β_n accordingly such that SNR $= \sqrt{500}$. Figure C.2 reports

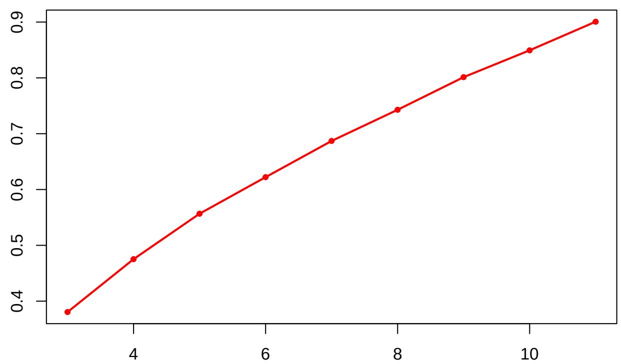

Figure C.2: The ℓ^1-loss of MSL under different values of K ($n = 5000$).

the ℓ^1-loss (averaged over 100 repetitions) of MSL for different K.

The results suggest that the ℓ^1-loss of MSL increases with K. We recall that the minimax rate is proportional to $K\sqrt{K}$ "asymptotically." However, for these "finite" K here, we observe that the error grows with K approximately linearly.

D Auxiliary lemmas on regularized graph Laplacian

Given $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$, recall that $\bar{\theta} = n^{-1}\sum_{i=1}^n \theta_i$. We introduce two disjoint index sets:

$$S_1 = \{1 \le i \le n : \theta_i \ge \bar{\theta}\}, \qquad S_2 = \{1 \le i \le n : \theta_i < \bar{\theta}\}. \tag{D.1}$$

D.1 Properties of L_0

Recall that $L_0 = H_0^{-\frac{1}{2}} \Omega H_0^{-\frac{1}{2}}$. We state the following three lemmas which give the spectrum properties of L_0 and also the estimates of the degree regularization matrix H_0.

Lemma D.1. *Under the conditions of Theorem 4.1,*

$$\lambda_1 > 0, \quad \lambda_1 \asymp 1, \quad |\lambda_K| \asymp K^{-1}\lambda_K(PG), \quad \lambda_1 - \max_{2\le k\le K} |\lambda_k| \ge c\lambda_1. \tag{D.2}$$

Lemma D.2. *Under the conditions of Theorem 4.1,*

$$\xi_1(i) \asymp \frac{1}{\sqrt{n}} \begin{cases} \sqrt{\theta_i/\bar{\theta}}, & i \in S_1, \\ \theta_i/\bar{\theta}, & i \in S_2, \end{cases} \qquad \|\Xi(i)\| \le \frac{C\sqrt{K}}{\sqrt{n}} \begin{cases} \sqrt{\theta_i/\bar{\theta}}, & i \in S_1, \\ \theta_i/\bar{\theta}, & i \in S_2, \end{cases} \tag{D.3}$$

and

$$H_0(i,i) \asymp \begin{cases} n\theta_i\bar{\theta}, & i \in S_1, \\ n\bar{\theta}^2, & i \in S_2. \end{cases} \tag{D.4}$$

Lemma D.3. *Under the conditions of Theorem 4.1, with probability* $1 - o(n^{-3})$,

$$\|I_n - H_0^{-1}H\| \le \frac{C\sqrt{\log(n)}}{\sqrt{n\bar{\theta}^2}}, \qquad \|H_0^{-1/2}(A-\Omega)H_0^{-1/2}\| \le \frac{C}{\sqrt{n\bar{\theta}^2}}. \tag{D.5}$$

The proof of Lemma D.1 is straightforward by noting that $D_\theta = H_0$ (see the definition of D_θ in Section 2) and therefore L_0 share the same eigenvalues as $K^{-1}PG$. Immediately, one can conclude (D.2) from Condition 2.1(b) and also the fact that $\lambda_1(PG) \asymp K$ which is derived from Condition 2.1(a) and (b). In the sequel, we show the proof of the Lemmas D.2 and D.3. Before that, we introduce the Bernstein inequality which we will use frequently to bound sum of independent Bernoulli entries.

Theorem D.1 (Bernstein inequality)**.** *Let* $X_1, \cdots, X_n$ *be independent zero-mean random variables. Suppose that* $|X_i| \le M$ *almost surely, for all* i*. Then for all* $t > 0$,

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n X_i\Big| \ge t\Big) \le 2\exp\Big(-\frac{t^2/2}{\sigma^2 + Mt/3}\Big),$$

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*

$$\Big|\sum_{i=1}^n X_i\Big| \le C(\sigma\sqrt{\log(n)} + M\log(n)) \quad \textit{with probability } 1 - o(n^{-5}).$$

Remark D.2. The exponent in the high probability $1 - o(n^{-5})$ can be replaced by $-\tilde{C}$ for any large positive constant integer $\tilde{C}$ by appropriately adjusting the constant C, which depends on $\tilde{C}$. Therefore, as long as we employ Bernstein inequality at most polynomial times in n to obtain the ultimate upper bound, the result still holds with high probability by adjusting the constant factor in the bound. Specifically, we can achieve a high probability of $1 - o(n^{-3})$ by choosing a sufficiently large constant C in the upper bound. Throughout the supplement, we sometimes omit specifying the high probability when applying Bernstein inequality for simplicity. It should be noted that in our analysis, Bernstein inequality is applied approximately $O(Kn)$ times.

Proof of Lemma D.2. First, we show (D.4). Uniformly for all $1 \le i \le n$,

$$\mathbb{E}d_i = \theta_i \sum_{j \neq i} \theta_j \pi_j' P \pi_i = \theta_i \sum_{j \neq i} \theta_j \sum_k \pi_j(k) e_k' P \pi_i \ge c_1 n \theta_i \bar{\theta}/K \cdot \mathbf{1}_K' P \pi_i \ge c_1 c_2 n \theta_i \bar{\theta} \tag{D.6}$$

by the last inequalities in Condition 2.1(a) and (b). On the other hand, $\pi_j' P \pi_i \le \max_{t,s} P(t,s)$ for all i, j, then $\mathbb{E}d_i = \theta_i \sum_{j \neq i} \theta_j \pi_j' P \pi_i \le cn\theta_i \bar{\theta}$ for all i. As a result, $\mathbb{E}d_i \asymp n\theta_i\bar{\theta}$, $\mathbb{E}\bar{d} \asymp n\bar{\theta}^2$; 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 (D.4). Next, we turn to prove (D.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

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

Using Condition 2.1(a) and $H_0 = D_\theta$, one gets $\|BB'\| \le Kc$, $\lambda_{\min}(BB') \ge Kc^{-1}$. Write $B = (b_1, \cdots, b_K)$. We have $Kc^{-1} \le \|b_i\|^2 \le Kc$ for $1 \le i \le K$. Taking the i-th row of Ξ,

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

For the leading eigenvector ξ_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 $H_0^{-\frac{1}{2}} \Theta \Pi P \Pi' \Theta H_0^{-1} \Theta \Pi B = H_0^{-\frac{1}{2}} \Theta \Pi B \Lambda$, which implies that $P \Pi' \Theta H_0^{-1} \Theta \Pi B = B \Lambda$. As a consequence, b_1 is the first right eigenvector of $P \Pi' \Theta H_0^{-1} \Theta \Pi$, and equivalently, the first right eigenvector of PG. Using Condition 2.1(c), we easily conclude that $b_1(k) > 0, b_1(k) \asymp 1$ for all $1 \le k \le K$. Then, $\pi_i' b_1 \asymp 1$ for all $1 \le i \le n$, and the entrywise estimate of ξ_1 simply follows from (D.4). □

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

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

By (D.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 achieved by employing Bernstein inequality. Applying the Bernstein inequality (Theorem D.1) to $d_i - \mathbb{E}d_i$, we see that

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

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

$$\sum_{j\neq i} \mathrm{var}A_{ij} \leq cn\theta_i\bar{\theta}.$$

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

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

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

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

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 $\sum_{j=1}^{n}\sqrt{\theta_j} \leq \sqrt{n}\sqrt{\sum_{j=1}^{n}\theta_j} = n\sqrt{\bar{\theta}}$. This finished the first estimate of (D.5). Now, we proceed to the second estimate. We crudely bound $\|H_0^{-1/2}(A - \Omega)H_0^{-1/2}\|$ by $\|H_0^{-1/2}WH_0^{-1/2}\| + \|H_0^{-1/2}\mathrm{diag}(\Omega)H_0^{-1/2}\|$. First, it is easy to get the bound

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

Next, we apply the non-asymptotic bounds for random matrices in [3] to bound the operator norm of $\widehat{W} := H_0^{-1/2} W H_0^{-1/2}$. Note that $\widehat{W}$ is a symmetric random matrix with independent upper triangular entries. Using Corollary 3.12 of [3] with Remark 3.13, we bound

$$\mathbb{P}(\|\widehat{W}\| \geq C\tilde{\sigma} + t) \leq n e^{-t^2/c\tilde{\sigma}_*^2}$$

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

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

Then, we take $t = c/\sqrt{n\bar{\theta}^2}$ for properly large $c > 0$ and use the assumption $n\bar{\theta}^2 \gg \log(n)$. It follows that $\|\widehat{W}\| \leq C/\sqrt{n\bar{\theta}^2}$ with probability $1 - o(n^{-3})$. We thus complete the proof of Lemma D.3. □

D.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 D.3 holds.

Lemma D.4. *Under the conditions in Theorem 4.1. Over the event E, for any fixed $1 \leq i \leq n$, the eigenvalues $\tilde{\lambda}_1^{(i)}, \cdots, \tilde{\lambda}_K^{(i)}$ of $\tilde{L}^{(i)}$ satisfy*

$$\tilde{\lambda}_1^{(i)} > 0, \quad \tilde{\lambda}_1^{(i)} \asymp 1, \quad |\tilde{\lambda}_K^{(i)}| \asymp K^{-1}|\lambda_K(PG)|, \quad \tilde{\lambda}_1^{(i)} - \max_{2 \leq k \leq K} |\tilde{\lambda}_k^{(i)}| \geq C^{-1}\tilde{\lambda}_1^{(i)}; \tag{D.7}$$

and for the associated eigenvectors,

$$\tilde{\xi}_1^{(i)}(j) \asymp \frac{1}{\sqrt{n}} \begin{cases} \sqrt{\theta_j/\bar{\theta}}, & j \in S_1, \\ \theta_j/\bar{\theta}, & j \in S_2, \end{cases} \qquad \|\tilde{\Xi}^{(i)}(j)\| \leq \frac{C\sqrt{K}}{\sqrt{n}} \begin{cases} \sqrt{\theta_j/\bar{\theta}}, & j \in S_1, \\ \theta_j/\bar{\theta}, & j \in S_2. \end{cases} \tag{D.8}$$

Lemma D.5. *Under the conditions of Theorem 4.1. Over the event E, for any fixed $1 \leq i \leq n$ and $\tilde{H}^{(i)}$,*

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

In addition to the above lemma, by Theorem D.1 and after elementary computations, we also have that for each $1 \le j \le n, j \neq i$, over the event E,

$$\begin{aligned} \big|\tilde{H}^{(i)}(j,j) - H(j,j)\big| &= \Big| - W(j,i) - \frac{2}{n}\Big(\sum_{s\neq i} W(i,s)\Big)\Big| \\ &\le \Big| - A(j,i) + \Omega(j,i) - \frac{2}{n}\Big(\sum_{s\neq i} W(i,s)\Big)\Big| \\ &\le A(j,i) + \theta_j\theta_i + C\frac{\sqrt{\theta_i\bar{\theta}\log n}}{\sqrt{n}} + \frac{C\log(n)}{n}; \end{aligned} \tag{D.10}$$

and

$$\big|\tilde{H}^{(i)}(i,i) - H(i,i)\big| \le (1+2/n)\Big|\sum_{s\neq i} W(i,s)\Big| \le C\sqrt{n\theta_i\bar{\theta}\log(n)} + C\log(n). \tag{D.11}$$

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

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

under the assumption of Theorem 4.1, then (D.7) can be derived simply by further applying Lemma D.1. Moreover, (D.8) follows from Lemmas 5.1, E.1 and D.2. Thereby, we omit the proofs of Lemmas D.4 and D.5. We comment here that the proof of Lemmas 5.1, E.1 only depends on the lemmas in Section D.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.

D.3 Variants of Davis-Kahan sinθ Theorem

In our analysis, we heavily rely on the use of Davis-Kahan sinθ theorem under different versions. For readers' convenience, we collect all the variants we employed in our theory below.

Theorem D.3 (Davis-Kahan sinΘ Theorem and its variants). *Let $\Sigma, \hat{\Sigma} \in \mathbb{R}^{p,p}$ be symmetric, with eigenvalues $\lambda_1 \ge \cdots \ge \lambda_p$ and $\hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_p$ respectively. Fix $1 \le r \le s \le p$,* *let* $U = (u_r, \cdots, u_s)$ *and* $U^\perp = (u_1, \cdots, u_{r-1}, u_{s+1}, \cdots, u_p)$ *be the orthonormal eigenvectors such that* $\Sigma u_j = \lambda_j u_j$*, similarly we define* $\hat{U}, \hat{U}^\perp$ *for* $\hat{\Sigma}$*. Denote* $\delta := \inf\{|\hat{\lambda} - \lambda| : \lambda \in [\lambda_r, \lambda_s], \hat{\lambda} \in (-\infty, \hat{\lambda}_{r-1}] \cup [\hat{\lambda}_{s+1}, \infty)\}$ *where we take the convention* $\hat{\lambda}_0 = -\infty$ *and* $\hat{\lambda}_{p+1} = \infty$*. Then,*

$$\|(\hat{U}^\perp)' U\| = \|\sin\Theta(\hat{U}, U)\| \leq \frac{\|\hat{\Sigma} - \Sigma\|}{\delta} \tag{D.12}$$

for some constant $C > 0$*. Moreover, there exists an orthogonal matrix* O *of dimension* $s - r + 1$ *such that*

$$\|\hat{U}'U - O'\| \leq C\left(\frac{\|\hat{\Sigma} - \Sigma\|}{\delta}\right)^2, \tag{D.13}$$

$$\|\hat{U} - UO\| \leq C\frac{\|\hat{\Sigma} - \Sigma\|}{\delta} \tag{D.14}$$

for some constant $C > 0$*.*

Note that (D.12) is the version of sinΘ theorem proved by Davis and Kahan's original paper [4]. The proof of (D.13) can be referred to Lemma B.2 in the Supplementary of [1]. More specifically, $O' = \text{sgn}(\hat{U}'U) = \bar{U}\bar{V}'$ where the SVD of $\hat{U}'U$ is given by $\bar{U}\bar{\Lambda}\bar{V}'$. By Chp I, Cor 5.4 of [8], the singular values in $\bar{\Lambda}$ are the cosines of canonical angles $0 \leq \bar{\theta}_1 \leq \cdots \leq \bar{\theta}_{s-r+1} \leq \pi/2$ between $\hat{U}$ and U. It follows that

$$\|\hat{U}'U - O'\| = 1 - \cos\bar{\theta}_{s-r+1} \leq 1 - \cos^2\bar{\theta}_{s-r+1} = \sin^2\bar{\theta}_{s-r+1} = \|\sin\Theta(\hat{U}, U)\|^2$$

Thus, (D.13) follows directly from (D.12). (D.14) is implied by the simple derivations

$$\|\hat{U} - UO\|^2 = \|2I_{s-r+1} - \hat{U}'UO - O'U'\hat{U}\| \leq 2\|\hat{U}'U - O'\|\,.$$

E Entrywise eigenvector analysis

Here we show the complete proof of Theorem 4.1 in our manuscript. In Sections E.1-E.3, we state the proofs of key lemmas for proving (J.3), while the claim of (J.3) is already presented in the manuscript. Section E.4 collects the proof of the second claim in Theorem 4.1 (i.e., (J.4)) which provides the entry-wise estimates for the 2- to K-th eigenvectors. Similarly to the proof of the first claim in Theorem 4.1 (i.e., (J.3)), we introduce three key lemmas, Lemmas E.1-E.3, counterpart to Lemmas 5.1-5.3. The proofs of Lemmas E.1-E.3 are provided correspondingly in Section E.5-E.7.

E.1 Proof of Lemma 5.1

In this subsection, we show the proof of Lemma 5.1 using the eigen-properties of L_0 in Section D.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,

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

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

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

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

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

As a result,

$$|\tilde{\xi}_1^{(i)}(j) - \xi_1(j)| \le \Big|\frac{1}{\tilde{Y}(j,j)} - 1\Big||\tilde{\xi}_1^{(i)}(j)| + \Big|\frac{\lambda_1(\xi_1'\tilde{Y}\tilde{\xi}_1^{(i)})}{\tilde{\lambda}_1^{(i)}} - 1\Big||\xi_1(j)| + \sum_{k=2}^{K}\Big|\frac{\lambda_k(\xi_k'\tilde{Y}\tilde{\xi}_1^{(i)})}{\tilde{\lambda}_1^{(i)}}\Big||\xi_k(j)|. \tag{E.2}$$

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

$$\|\tilde{Y} - I_n\| = \big\|I_n - (H_0^{-1}\tilde{H}^{(i)})^{-\frac{1}{2}}\big\| \le \frac{C\sqrt{\log(n)}}{\sqrt{n\bar{\theta}^2}}, \tag{E.3}$$

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

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

since $\lambda_1(PG) \leq CK$ under our model assumption. Furthermore, by Lemma D.1, the eigengap 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 λ_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 (D.13) that

$$\begin{aligned}|\xi_1'\tilde{Y}\tilde{\xi}_1^{(i)} - 1| &\leq |\xi_1'\tilde{\xi}_1^{(i)} - 1| + |\xi_1'(\tilde{Y} - I_n)\tilde{\xi}_1^{(i)}| \\ &\leq C(K\lambda_1^{-1}(PG)\|\tilde{L}^{(i)} - L_0\|)^2 + \|\tilde{Y} - I_n\| \leq C\|\tilde{Y} - I_n\|.\end{aligned}$$

Here $\mathrm{sgn}(\xi_1'\tilde{\xi}_1^{(i)}) = 1$ 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 $\sum_{k=2}^K |\xi_k'\tilde{Y}\tilde{\xi}_1^{(i)}||\xi_k(j)| \leq \|\Xi_1'\tilde{Y}\tilde{\xi}_1^{(i)}\|\|\Xi_1(j)\|$. And by sine-theta theorem (D.12),

$$\begin{aligned}\|\Xi_1'\tilde{Y}\tilde{\xi}_1^{(i)}\| &\leq \|(\tilde{\xi}_1^{(i)})'\Xi_1\| + \|\tilde{Y} - I_n\| \\ &\leq C\Big(K\lambda_1^{-1}(PG)\|\tilde{L}^{(i)} - L_0\| + \|\tilde{Y} - I_n\|\Big) \\ &\leq C\|\tilde{Y} - I_n\|.\end{aligned}$$

Plugging the above estimates, we have

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

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

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

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

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

Then, consider all i's together, we conclude (31) with probability $1 - o(n^{-3})$ simultaneously for all $1 \le i, j \le n$.

E.2 Proof of Lemma 5.2

In this subsection, we state the proof of Lemma 5.2 which heavily relies on the eigen-properties of $\tilde{L}^{(i)}$ in Section D.2.

Fix the index i, we first show (33) which is based on the decomposition

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

where $w = \mathrm{sgn}(\xi_1'\hat{\xi}_1)$ will be claimed later. It is not hard to derive

$$\begin{aligned}|e_i'\Delta(w\hat{\xi}_1 - \overline{\xi}_1^{(i)})| \le \|\Delta\| \, \|w\hat{\xi}_1 - \overline{\xi}_1^{(i)}\| &\le \|H_0^{-\frac{1}{2}} W H_0^{-\frac{1}{2}}\| \|H_0^{\frac{1}{2}} H^{-\frac{1}{2}}\| \|H_0^{\frac{1}{2}} (\tilde{H}^{(i)})^{-\frac{1}{2}}\| \, \|w\hat{\xi}_1 - \overline{\xi}_1^{(i)}\| \\ &\le \frac{C}{\sqrt{n\bar{\theta}^2}} \|w\hat{\xi}_1 - \overline{\xi}_1^{(i)}\|\end{aligned}$$

over the event E, in light of Lemmas D.3 and D.5. We thus end up with (33) . We now turn to prove (32). We study the perturbation from $\tilde{L}^{(i)} = (\tilde{H}^{(i)})^{-1/2}\Omega(\tilde{H}^{(i)})^{-1/2}$ to $L = H^{-1/2}AH^{-1/2}$. Write $X \equiv X(i) := (\tilde{H}^{(i)})^{1/2}H^{-1/2}$. We can rewrite

$$L = X(\tilde{H}^{(i)})^{-\frac{1}{2}} A (\tilde{H}^{(i)})^{-\frac{1}{2}} X = X\tilde{L}^{(i)}X - X(\tilde{H}^{(i)})^{-\frac{1}{2}} \mathrm{diag}(\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}} X + X\Delta$$

with $\Delta = (\tilde{H}^{(i)})^{-1/2} W H^{-1/2}$. By definition, $\tilde{L}^{(i)} = (\tilde{H}^{(i)})^{-1/2}\Omega(\tilde{H}^{(i)})^{-1/2} = \sum_{k=1}^K \tilde{\lambda}_k^{(i)} \tilde{\xi}_k^{(i)} (\tilde{\xi}_k^{(i)})'$ and $H^{-1/2}AH^{-1/2}\hat{\xi}_1 = \hat{\lambda}_1\hat{\xi}_1$. It follows that

$$\sum_{k=1}^K \tilde{\lambda}_k^{(i)} (\hat{\xi}_1' X \tilde{\xi}_k^{(i)}) X\tilde{\xi}_k^{(i)} - X(\tilde{H}^{(i)})^{-\frac{1}{2}} \mathrm{diag}(\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}} X\hat{\xi}_1 + X\Delta\hat{\xi}_1 = \hat{\lambda}_1\hat{\xi}_1.$$

As a result,

$$\begin{aligned}\hat{\xi}_1(i) =& \frac{\tilde{\lambda}_1^{(i)} (\hat{\xi}_1' X \tilde{\xi}_1^{(i)})}{\hat{\lambda}_1} X(i,i)\tilde{\xi}_1^{(i)}(i) + \sum_{k=2}^K \frac{\tilde{\lambda}_k^{(i)} (\hat{\xi}_1' X \tilde{\xi}_k^{(i)})}{\hat{\lambda}_1} X(i,i)\tilde{\xi}_k^{(i)}(i) \\ &- \frac{X^2(i,i)\Omega(i,i)}{\hat{\lambda}_1 \tilde{H}^{(i)}(i,i)} \hat{\xi}_1(i) + \frac{X(i,i)}{\hat{\lambda}_1} e_i'\Delta\hat{\xi}_1. \end{aligned} \tag{E.6}$$

By Lemma D.5, it is easy to deduce that

$$\|X - I_n\| \le C\frac{\sqrt{\log(n)}}{\sqrt{n\bar{\theta}^2}}, \tag{E.7}$$

Since (D.7), by Weyl's inequality,

$$\max_{1\le k\le K}\{|\hat{\lambda}_k - \tilde{\lambda}_k^{(i)}|\} \le C\|L - \tilde{L}^{(i)}\|; \tag{E.8}$$

and over the event E,

$$\begin{aligned}\|L - \tilde{L}^{(i)}\| &\le C\|X - I_n\|\|H^{-\frac{1}{2}}AH^{-\frac{1}{2}}\| + \|(\tilde{H}^{(i)})^{-1}H_0\|\|H_0^{-1/2}(A-\Omega)H_0^{-1/2}\| \\ &\le C\,\|X - I_n\|\|H_0^{-\frac{1}{2}}\Omega H_0^{-\frac{1}{2}}\| + \|(\tilde{H}^{(i)})^{-1}H_0\|\|H_0^{-1/2}(A-\Omega)H_0^{-1/2}\|, \\ &\le C\,\frac{K^{-1}\lambda_1(PG)\sqrt{\log(n)}+1}{\sqrt{n\bar{\theta}^2}} \le C\sqrt{\frac{\log(n)}{n\bar{\theta}^2}} \ll |\tilde{\lambda}_K^{(i)}| \end{aligned} \tag{E.9}$$

since Lemmas D.1, D.3 and (D.7) with the condition $K\beta_n^{-1}\sqrt{\log(n)}/\sqrt{n\bar{\theta}^2} \ll 1$. Therefore, $\hat{\lambda}_1, \cdots, \hat{\lambda}_K$ share the same asymptotics as $\tilde{\lambda}_1^{(i)}, \cdots, \tilde{\lambda}_K^{(i)}$. The eigengap between $\hat{\lambda}_1$ and $\tilde{\lambda}_2^{(i)}, \cdots, \tilde{\lambda}_K^{(i)}$ is $K^{-1}\lambda_1(PG)$. Let $w(i) = \mathrm{sgn}(\hat{\xi}_1'\tilde{\xi}_1^{(i)})$. It follows that

$$\begin{aligned}|\hat{\xi}_1'X\tilde{\xi}_1^{(i)} - w(i)| &\le \|X - I_n\| + |\hat{\xi}_1'\tilde{\xi}_1^{(i)} - w(i)| \\ &\le \|X - I_n\| + C\Big(K\lambda_1^{-1}(PG)\|L - \tilde{L}^{(i)}\|\Big)^2\end{aligned} \tag{E.10}$$

where the last step is due to (D.13) in sinΘ Theorem . In particular, further by (D.14)

$$\|\hat{\xi}_1'X(\tilde{\xi}_2^{(i)}, \cdots, \tilde{\xi}_K^{(i)})\| \le \|X - I_n\| + CK\lambda_1^{-1}(PG)\|L - \tilde{L}^{(i)}\|. \tag{E.11}$$

We can actually claim that $w(i) \equiv w := \mathrm{sgn}(\xi_1'\hat{\xi}_1)$ as follows. First notice $|\hat{\xi}_1'\xi_1 - \hat{\xi}_1'\tilde{\xi}_1^{(i)}| \le \|\tilde{\xi}_1^{(i)} - \xi_1\| = o(1)$. Next, $|\hat{\xi}_1'\tilde{\xi}_1^{(i)}| > c$ for some constant $c \in (0, 1)$. It follows immediately that $w(i) = \mathrm{sgn}(\hat{\xi}_1'\tilde{\xi}_1^{(i)}) = \mathrm{sgn}(\hat{\xi}_1'\xi_1) = w$. In the sequel, we directly write w instead of $w(i)$. We plug in (E.8), (E.10) and (E.11) into (E.6). By some elementary simplifications, it arrives at

$$\begin{aligned}|w\hat{\xi}_1(i) - \tilde{\xi}_1^{(i)}(i)| \le &\Big(\|X - I_n\| + K\lambda_1^{-1}(PG)\|L - \tilde{L}^{(i)}\|\Big)\Big(\big|\tilde{\xi}_1^{(i)}(i)\big| + \|\tilde{\Xi}_1^{(i)}(i)\|\Big) \\ &+ \Big|\frac{X^2(i,i)\Omega(i,i)}{\hat{\lambda}_1\tilde{H}^{(i)}(i,i)}\hat{\xi}_1(i)\Big| + \Big|\frac{X(i,i)}{\hat{\lambda}_1}e_i'\Delta\hat{\xi}_1\Big|.\end{aligned} \tag{E.12}$$

We can further derive

$$\left|\frac{X(i,i)^2\Omega(i,i)}{\hat{\lambda}_1\tilde{H}^{(i)}(i,i)}\hat{\xi}_1(i)\right| \le \frac{CK\theta_i^2}{n\bar{\theta}(\bar{\theta}\vee\theta_i)\lambda_1(PG)} \le CK\lambda_1^{-1}(PG)\kappa_i\Big(\sqrt{\frac{\theta_i}{\bar{\theta}}}\wedge 1\Big), \tag{E.13}$$

by the estimate $\tilde{H}^{(i)}(i,i) \asymp n\bar{\theta}(\bar{\theta}\vee\theta_i)$ following from (D.4) and the first estimate in (D.9), with $\kappa_i = \frac{\sqrt{\log(n)}}{n\bar{\theta}} \cdot \frac{\sqrt{\theta_i}}{\sqrt{\bar{\theta}}}$. Then, plugging (E.13), (E.7) and (E.9) into (E.12) gives

$$\begin{aligned}|w\hat{\xi}_1(i) - \tilde{\xi}_1^{(i)}(i)| &\le C\Big(\sqrt{\frac{\log n}{n\bar{\theta}^2}} + \frac{K}{\lambda_1(PG)\sqrt{n\bar{\theta}^2}}\Big)\Big(\big|\tilde{\xi}_1^{(i)}(i)\big| + \|\tilde{\Xi}_1^{(i)}(i)\|\Big) \\ &\quad + CK\lambda_1^{-1}(PG)\kappa_i\Big(\sqrt{\frac{\theta_i}{\bar{\theta}}}\wedge 1\Big) + CK\lambda_1^{-1}(PG)|e_i'\Delta\hat{\xi}_1| \\ &\le C\kappa_i\Big(\sqrt{\frac{\theta_i}{\bar{\theta}}}\wedge 1\Big)K^{\frac{3}{2}}\lambda_1^{-1}(PG) + CK\lambda_1^{-1}(PG)|e_i'\Delta\hat{\xi}_1|\end{aligned}$$

where in the last step, we plugged the bound of $\big|\tilde{\xi}_1^{(i)}(i)\big|$ and $\|\tilde{\Xi}^{(i)}(i)\|$ in (D.8). Since the assumption $\lambda_1(PG) \ge CK$, it gives that over the event E,

$$|w\hat{\xi}_1(i) - \tilde{\xi}_1^{(i)}(i)| \le C\sqrt{K}\kappa_i + C|e_i'\Delta\hat{\xi}_1|.$$

This concludes our proof by considering all i's together.

E.3 Proof of Lemma 5.3

In this section, we prove Lemma 5.3. We separate the proofs into three parts corresponding to the three estimates (34)-(36).

E.3.1 Proof of (34)

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

$$e_i'\Delta\tilde{\xi}_1^{(i)} = \frac{W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}X\tilde{\xi}_1^{(i)}}{\sqrt{\tilde{H}^{(i)}(i,i)}} = \frac{W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{\xi}_1^{(i)}}{\sqrt{\tilde{H}^{(i)}(i,i)}} + \frac{W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(X-I_n)\tilde{\xi}_1^{(i)}}{\sqrt{\tilde{H}^{(i)}(i,i)}} \tag{E.14}$$

First, we study the term $|W(i)(\tilde{H}^{(i)})^{-1/2}\tilde{\xi}_1^{(i)}|$. Write

$$W(i)(\tilde{H}^{(i)})^{-1/2}\tilde{\xi}_1^{(i)} = \sum_{1\le j\le n: j\ne i}\frac{W(i,j)}{\sqrt{\tilde{H}^{(i)}(j,j)}}\tilde{\xi}_1^{(i)}(j).$$

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

$$\sum_{j\neq i} \frac{\theta_i\theta_j}{n\bar{\theta}(\theta_j \vee \bar{\theta})} \big(\tilde{\xi}_1^{(i)}(j)\big)^2 \leq \sum_{j} \frac{\theta_i\theta_j}{n\theta_j\bar{\theta}} \big(\tilde{\xi}_1^{(i)}(j)\big)^2 \leq \frac{\theta_i}{n\bar{\theta}}.$$

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

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

following from (D.8) and the estimate $\tilde{H}^{(i)}(i,i) \asymp n\bar{\theta}(\bar{\theta} \vee \theta_i)$. Applying Theorem D.1, one see that over the event E,

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

Hence, over the event E,

$$\frac{W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{\xi}_1^{(i)}}{\sqrt{\tilde{H}^{(i)}(i,i)}} \leq C\frac{\sqrt{\log(n)}}{n\bar{\theta}}\Big(1 \wedge \frac{\sqrt{\theta_i}}{\sqrt{\bar{\theta}}} + \frac{\sqrt{\log(n)}}{\sqrt{n\bar{\theta}^2}}\Big) \leq C\,\widetilde{\kappa}_i \tag{E.15}$$

by using the estimate $\tilde{H}^{(i)}(i,i) \asymp n\bar{\theta}(\bar{\theta} \vee \theta_i)$ and the definition of $\widetilde{\kappa}_i$ in (34).

Next, we study the term $|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(X - I_n)\tilde{\xi}_1^{(i)}|$. Over the event E, by (D.10), we have

$$|X(j,j) - 1| \leq C\frac{|H(j,j) - \tilde{H}^{(i)}(j,j)|}{\tilde{H}^{(i)}(j,j)} \leq C\,\frac{A(i,j) + \theta_i\theta_j + \theta_i\bar{\theta} + \log(n)/n}{\tilde{H}^{(i)}(j,j)}, \qquad j \neq i; \tag{E.16}$$

It follows that

$$\begin{aligned}|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(X - I_n)\tilde{\xi}_1^{(i)}| &= \Big|\sum_{1\leq j\leq n: j\neq i} W(i,j)\frac{[X(j,j)-1]\tilde{\xi}_1^{(i)}(j)}{\sqrt{\tilde{H}^{(i)}(j,j)}}\Big| \\ &\leq \sum_{1\leq j\leq n: j\neq i} |W(i,j)|\frac{|X(j,j)-1||\tilde{\xi}_1^{(i)}(j)|}{\sqrt{\tilde{H}^{(i)}(j,j)}}\end{aligned}$$

$$\le C \sum_{1\le j\le n: j\ne i} \big(A(i,j) + \theta_i\theta_j + \theta_i\bar\theta + \frac{\log(n)}{n}\big)\frac{|\tilde\xi_1^{(i)}(j)|}{[\tilde H^{(i)}(j,j)]^{3/2}},$$

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

$$\sum_{j\ne i} \frac{\theta_i\theta_j + \theta_i\bar\theta + \log(n)/n}{(n\bar\theta(\theta_j\vee\bar\theta))^{3/2}}\cdot\frac{\sqrt{\theta_j}}{\sqrt{n\bar\theta}} \le \sum_j \frac{\theta_i + log(n)/(n\bar\theta)}{(n\bar\theta)^2} \le \frac{\theta_i}{n\bar\theta^2}\Big(1+\frac{\log(n)}{n\bar\theta\theta_i}\Big).$$

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

$$\sum_{j\ne i}\frac{\theta_i\theta_j}{(n\bar\theta(\theta_j\vee\bar\theta))^3}\big(\tilde\xi_1^{(i)}(j)\big)^2 \le \frac{1}{(n\bar\theta^2)^2}\sum_j \frac{\theta_i\theta_j}{n\theta_j\bar\theta}\big(\tilde\xi_1^{(i)}(j)\big)^2 \le \frac{1}{(n\bar\theta^2)^2}\cdot\frac{\theta_i}{n\bar\theta}.$$

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

$$\frac{|\tilde\xi_1^{(i)}(j)|}{[\tilde H^{(i)}(j,j)]^{3/2}} \le \frac{C}{n\bar\theta}\begin{Bmatrix} 1/(n\theta_j\bar\theta), & j\in S_1 \\ \theta_j/(n\bar\theta^2), & j\in S_2. \end{Bmatrix} \le \frac{C}{n\bar\theta}\cdot\frac{1}{n\bar\theta^2}.$$

We then have

$$|W(i)(\tilde H^{(i)})^{-\frac12}(X-I_n)\tilde\xi_1^{(i)}| \le C\frac{\theta_i}{n\bar\theta^2}\Big(1+\frac{\log(n)}{n\bar\theta\theta_i}\Big)$$

As a result, over the event E,

$$\frac{W(i)(\tilde H^{(i)})^{-\frac12}(X-I_n)\tilde\xi_1^{(i)}}{\sqrt{\tilde H^{(i)}(i,i)}} \le \frac{C}{n\bar\theta^2}\sqrt{\frac{\theta_i}{n\bar\theta}}\Big(1\wedge\frac{\sqrt{\theta_i}}{\sqrt{\bar\theta}} + \sqrt{\frac{\log(n)}{n\bar\theta\theta_i}}\sqrt{\frac{\log(n)}{n\bar\theta^2}}\Big) \le C\frac{\widetilde\kappa_i}{\sqrt{n\bar\theta^2\log(n)}} \tag{E.17}$$

We plug (E.15) and (E.17) into (E.14), and consider all i's over the event E, then we conclude the proof of (34).

E.3.2 Proof of (35)

Similarly to (E.14), we have

$$|e_i'\Delta(\bar\xi_1^{(i)} - \tilde\xi_1^{(i)})| \le \frac{|W(i)(\tilde H^{(i)})^{-1/2}(\bar\xi_1^{(i)}-\tilde\xi_1^{(i)})|}{\sqrt{\tilde H^{(i)}(i,i)}} + \frac{C|W(i)(\tilde H^{(i)})^{-1/2}(X-I_n)(\bar\xi_1^{(i)}-\tilde\xi_1^{(i)})|}{\sqrt{\tilde H^{(i)}(i,i)}} \tag{E.18}$$

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

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

We shall apply Bernstein's inequality since $(\tilde{H}^{(i)})^{-1/2}(\overline{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})$ is independent of $W(i)$. The variance is bounded by (up to some constant)

$$\sum_{j\neq i} \frac{\theta_i\theta_j}{n\bar{\theta}(\theta_j \vee \bar{\theta})}[\overline{\xi}_1^{(i)}(j) - \tilde{\xi}_1^{(i)}(j)]^2 \le \sum_j \frac{\theta_i\theta_j}{n\theta_j\bar{\theta}}[\overline{\xi}_1^{(i)}(j) - \tilde{\xi}_1^{(i)}(j)]^2 \le \frac{4\theta_i}{n\bar{\theta}}$$

Each individual term is bounded by $\|(\tilde{H}^{(i)})^{-1/2}(\overline{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})\|_\infty$. As a result,

$$\begin{aligned}
&|W(i)(\tilde{H}^{(i)})^{-1/2}(\overline{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})| \\
&\le C\frac{\sqrt{\theta_i \log(n)}}{\sqrt{n\bar{\theta}}} + C\log(n)\|(\tilde{H}^{(i)})^{-1/2}(\overline{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})\|_\infty \\
&\le C\frac{\sqrt{\theta_i \log(n)}}{\sqrt{n\bar{\theta}}} + C\log(n)\|(\tilde{H}^{(i)})^{-1/2}(w\hat{\xi}_1 - \tilde{\xi}_1^{(i)})\|_\infty + C\log(n)\|(\tilde{H}^{(i)})^{-1/2}(\overline{\xi}_1^{(i)} - w\hat{\xi}_1)\| \\
&\le C\frac{\sqrt{\theta_i \log(n)}}{\sqrt{n\bar{\theta}}} + C\log(n)\|(\tilde{H}^{(i)})^{-1/2}(w\hat{\xi}_1 - \tilde{\xi}_1^{(i)})\|_\infty + \frac{C\log(n)}{\sqrt{n\bar{\theta}^2}}\|\overline{\xi}_1^{(i)} - w\hat{\xi}_1\|.
\end{aligned}$$

Further,

$$\begin{aligned}
\frac{|W(i)(\tilde{H}^{(i)})^{-1/2}(\overline{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})|}{\sqrt{\tilde{H}^{(i)}(i,i)}} &\le C\kappa_i + C\kappa_i\frac{\sqrt{n\bar{\theta}\log(n)}}{\sqrt{\theta_i}}\|(\tilde{H}^{(i)})^{-1/2}(w\hat{\xi}_1 - \tilde{\xi}_1^{(i)})\|_\infty + \frac{C\log(n)}{n\bar{\theta}^2}\|\overline{\xi}_1^{(i)} - w\hat{\xi}_1\| \\
&\le C\widetilde{\kappa}_i + C\widetilde{\kappa}_i n\bar{\theta}\,\|(\tilde{H}^{(i)})^{-1/2}(w\hat{\xi}_1 - \tilde{\xi}_1^{(i)})\|_\infty + \frac{C\log(n)}{n\bar{\theta}^2}\|\overline{\xi}_1^{(i)} - w\hat{\xi}_1\|
\end{aligned} \tag{E.19}$$

by the definition of $\widetilde{\kappa}_i = \frac{1}{n\bar{\theta}}\sqrt{\frac{\log(n)}{n\bar{\theta}^2}}\sqrt{n\bar{\theta}\theta_i \vee \log(n)}$.

We then study the term $|W(i)(\tilde{H}^{(i)})^{-1/2}(X - I_n)(\overline{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})|$. On the event E, recall (E.16). It follows that

$$\begin{aligned}
&|W(i)(\tilde{H}^{(i)})^{-1/2}(X - I_n)(\overline{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})| \\
&= \Big|\sum_{1\le j\le n: j\neq i} W(i,j)\frac{[X(j,j)-1][\overline{\xi}_1^{(i)}(j) - \tilde{\xi}_1^{(i)}(j)]}{\sqrt{\tilde{H}^{(i)}(j,j)}}\Big|
\end{aligned}$$

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

$$\leq C \sum_{1\leq j\leq n: j\neq i} \frac{A(i,j)+\theta_i\theta_j+\theta_i\bar{\theta}+\log(n)/n}{\tilde{H}^{(i)}(j,j)} \frac{|\bar{\xi}_1^{(i)}(j) - \tilde{\xi}_1^{(i)}(j)|}{\sqrt{\tilde{H}^{(i)}(j,j)}}.$$

We now decompose the RHS above by $\mathcal{I}_1 + \mathcal{I}_2$, where

$$\mathcal{I}_1 := \sum_{1\leq j\leq n: j\neq i} \frac{A(i,j)+\theta_i\theta_j+\theta_i\bar{\theta}+\frac{\log(n)}{n}}{\tilde{H}^{(i)}(j,j)} \frac{|\bar{\xi}_1^{(i)}(j) - w\hat{\xi}_1(j)|}{\sqrt{\tilde{H}^{(i)}(j,j)}},$$

$$\mathcal{I}_2 := \sum_{1\leq j\leq n: j\neq i} \frac{A(i,j)+\theta_i\theta_j+\theta_i\bar{\theta}+\frac{\log(n)}{n}}{\tilde{H}^{(i)}(j,j)} \frac{|w\hat{\xi}_1(j) - \tilde{\xi}_1^{(i)}(j)|}{\sqrt{\tilde{H}^{(i)}(j,j)}}.$$

We bound the sub-terms separately as below. By Cauchy-Schwarz inequality,

$$\mathcal{I}_1 \leq \Big(\sum_{1\leq j\leq n: j\neq i} \frac{(A(i,j)+\theta_i\theta_j+\theta_i\bar{\theta}+\log(n)/n)^2}{\tilde{H}^{(i)}(j,j)^3}\Big)^{\frac{1}{2}} \|\bar{\xi}_1^{(i)} - w\hat{\xi}_1\|. \tag{E.20}$$

And we crudely bound

$$\mathcal{I}_2 \leq \sum_{1\leq j\leq n: j\neq i} \frac{A(i,j)+\theta_i\theta_j+\theta_i\bar{\theta}+\log(n)/n}{\tilde{H}^{(i)}(j,j)} \|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\tilde{\xi}_1^{(i)} - w\hat{\xi}_1)\|_\infty. \tag{E.21}$$

Applying Bernstein inequality (i.e., Theorem D.1), similarly to the analysis of the term $|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(X-I_n)\tilde{\xi}_1^{(i)}|$ in Section E.3.1, we can have the estimates

$$\begin{aligned}
\Big(\sum_{1\leq j\leq n: j\neq i} \frac{(A(i,j)+\theta_i\theta_j+\theta_i\bar{\theta}+\log(n)/n)^2}{\tilde{H}^{(i)}(j,j)^3}\Big)^{\frac{1}{2}} &\leq C\Big(\sum_{1\leq j\leq n: j\neq i} \frac{A(i,j)+\theta_i\theta_j+\theta_i\bar{\theta}+\log(n)/n}{\tilde{H}^{(i)}(j,j)^3}\Big)^{\frac{1}{2}} \\
&\leq C\Big(\frac{n\bar{\theta}\theta_i + \log(n)}{(n\bar{\theta}^2)^3}\Big)^{\frac{1}{2}} \\
&\leq C\Big(\frac{1}{n\bar{\theta}^2}\sqrt{\frac{\theta_i}{\bar{\theta}}} + \frac{1}{n\bar{\theta}^2}\sqrt{\frac{\log(n)}{n\bar{\theta}^2}}\Big) \qquad \text{(E.22)}
\end{aligned}$$

and

$$\sum_{1\leq j\leq n: j\neq i} \frac{A(i,j)+\theta_i\theta_j+\theta_i\bar{\theta}+\log(n)/n}{\tilde{H}^{(i)}(j,j)} \leq C\Big(\frac{\theta_i}{\bar{\theta}} + \frac{\log(n)}{n\bar{\theta}^2}\Big) \tag{E.23}$$

over the event E. We thus conclude that

$$|W(i)(\tilde{H}^{(i)})^{-1/2}(X-I_n)(\bar{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})|$$

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

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

$$\begin{aligned}&\frac{|W(i)(\tilde{H}^{(i)})^{-1/2}(X-I_n)(\bar{\xi}_1^{(i)}-\tilde{\xi}_1^{(i)})|}{\sqrt{\tilde{H}^{(i)}(i,i)}}\\ &\leq C\widetilde{\kappa}_i n\bar{\theta}\|(\tilde{H}^{(i)})^{-1/2}(w\hat{\xi}_1-\tilde{\xi}_1^{(i)})\|_\infty + C(n\bar{\theta}^2)^{-\frac{3}{2}}\|\bar{\xi}_1^{(i)}-w\hat{\xi}_1\| \end{aligned} \tag{E.24}$$

over the event E. Now, plugging (E.19) and (E.24) into (E.18) and combining all i's, we thus finish the proof of (35).

E.3.3 Proof of (36)

Note that $\bar{\xi}_1^{(i)}$ 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 $|\bar{\lambda}_2^{(i)}|$ is of order $K^{-1}\lambda_1(PG)\asymp 1$ in light of Weyl's inequality

$$\max_i|\bar{\lambda}_i^{(i)}-\tilde{\lambda}_i^{(i)}|\leq\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\tilde{A}^{(i)}-\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}}\|\leq C\sqrt{\frac{\log(n)}{n\bar{\theta}^2}} \tag{E.25}$$

and $K^{-1}\lambda_1(PG)\gg\sqrt{\log(n)/n\bar{\theta}^2}$. Similarly, the eigengap between $\hat{\lambda}_1$ and $|\bar{\lambda}_2^{(i)}|$ is of order $K^{-1}\lambda_1(PG)$. We claim that $\mathrm{sgn}(\hat{\xi}_1'\bar{\xi}_1^{(i)})=\mathrm{sgn}(\hat{\xi}_1'\tilde{\xi}_1^{(i)})\equiv w$. Notice that $|(\hat{\xi}_1')\bar{\xi}_1^{(i)}-(\hat{\xi}_1')\tilde{\xi}_1^{(i)}|\leq\|\bar{\xi}_1^{(i)}-\tilde{\xi}_1^{(i)}\|=o(1)$. This, together with the fact that $|(\hat{\xi}_1')\tilde{\xi}_1^{(i)}|>c$ for some constant $c\in(0,1)$, implies that $\mathrm{sgn}(\hat{\xi}_1'\bar{\xi}_1^{(i)})=\mathrm{sgn}(\hat{\xi}_1'\tilde{\xi}_1^{(i)})$. Moreover,

$$\|\bar{\xi}_1^{(i)}-w\hat{\xi}_1\|\leq K\lambda_1^{-1}(PG)\|\underbrace{\big((\tilde{H}^{(i)})^{-1/2}\tilde{A}^{(i)}(\tilde{H}^{(i)})^{-1/2}-H^{-1/2}AH^{-1/2}\big)}_{=:\tilde{\Delta}(i)\equiv\tilde{\Delta}}\hat{\xi}_1\|$$

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

$$\begin{aligned}\tilde{\Delta}&=(\tilde{H}^{(i)})^{-\frac{1}{2}}(\tilde{A}^{(i)}-A)(\tilde{H}^{(i)})^{-\frac{1}{2}}+\big((\tilde{H}^{(i)})^{-\frac{1}{2}}A(\tilde{H}^{(i)})^{-\frac{1}{2}}-H^{-\frac{1}{2}}AH^{-\frac{1}{2}}\big)\\ &=-(\tilde{H}^{(i)})^{-\frac{1}{2}}(e_iW(i)+W(i)'e_i')(\tilde{H}^{(i)})^{-\frac{1}{2}}+(\tilde{H}^{(i)})^{-\frac{1}{2}}A((\tilde{H}^{(i)})^{-\frac{1}{2}}-H^{-\frac{1}{2}})+((\tilde{H}^{(i)})^{-\frac{1}{2}}-H^{-1/2})AH^{-\frac{1}{2}}\\ &=-(\tilde{H}^{(i)})^{-\frac{1}{2}}(e_iW(i)+W(i)'e_i')(\tilde{H}^{(i)})^{-\frac{1}{2}}+(\tilde{H}^{(i)})^{-\frac{1}{2}}A(\tilde{H}^{(i)})^{-\frac{1}{2}}(I_n-X)+(X^{-1}-I_n)H^{-\frac{1}{2}}AH^{-\frac{1}{2}}.\end{aligned}$$

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

$$\tilde{\Delta}\hat{\xi}_1=-(\tilde{H}^{(i)})^{-\frac{1}{2}}(e_iW(i)+W(i)'e_i')(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\xi}_1+(\tilde{H}^{(i)})^{-\frac{1}{2}}A(\tilde{H}^{(i)})^{-\frac{1}{2}}(I_n-X)\hat{\xi}_1+\hat{\lambda}_1(X^{-1}-I_n)\hat{\xi}_1.$$

As a result,

$$
\begin{aligned}
\|\bar{\xi}_1^{(i)} - w\hat{\xi}_1\| &\le K\lambda_1^{-1}(PG)\frac{|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\xi}_1|}{\sqrt{\tilde{H}^{(i)}(i,i)}} + K\lambda_1^{-1}(PG)\frac{\|(\tilde{H}^{(i)})^{-\frac{1}{2}}W(i)'\|}{\sqrt{\tilde{H}^{(i)}(i,i)}}\cdot|\hat{\xi}_1(i)| + C\|(I_n - X)\hat{\xi}_1\| \\
&\le K\lambda_1^{-1}(PG)\frac{|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\xi}_1|}{\sqrt{\tilde{H}^{(i)}(i,i)}} + \frac{CK\lambda_1^{-1}(PG)}{\sqrt{n\bar{\theta}^2}}|\hat{\xi}_1(i)| + C\|(I_n - X)\hat{\xi}_1\|, \\
&\le CK\lambda_1^{-1}(PG)\Big(\frac{|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\xi}_1|}{\sqrt{\tilde{H}^{(i)}(i,i)}} + \frac{1}{\sqrt{n\bar{\theta}^2}}|\tilde{\xi}_1^{(i)}(i)| + \frac{1}{\sqrt{n\bar{\theta}^2}}|w\hat{\xi}_1(i) - \tilde{\xi}_1^{(i)}(i)|\Big) + C\|(I_n - X)\hat{\xi}_1\|, \\
&\le CK\lambda_1^{-1}(PG)\Big(\frac{|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\xi}_1|}{\sqrt{\tilde{H}^{(i)}(i,i)}} + \frac{\kappa_i}{\sqrt{\log(n)}} + \frac{1}{\sqrt{n\bar{\theta}^2}}|w\hat{\xi}_1(i) - \tilde{\xi}_1^{(i)}(i)|\Big) + C\|(I_n - X)\hat{\xi}_1\|,
\end{aligned}
\tag{E.26}
$$

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

$$
\begin{aligned}
(\tilde{H}^{(i)}(i,i))^{-\frac{1}{2}}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}W(i)'\| &\le \sqrt{W(i)(\tilde{H}^{(i)})^{-1}W(i)'}\Big/\sqrt{\tilde{H}^{(i)}(i,i)} \\
&\le \frac{\sqrt{(\theta_i/\bar{\theta}) \vee (\log(n)/n\bar{\theta}^2)}}{\sqrt{n\bar{\theta}(\bar{\theta} \vee \theta_i)}} \le C(n\bar{\theta}^2)^{-\frac{1}{2}}
\end{aligned}
\tag{E.27}
$$

by simply using Bernstein inequality to $W(i)(\tilde{H}^{(i)})^{-1}W(i)'$, and in the last line we have used (D.3).

We consider the first term in (E.26). Note that

$$
\begin{aligned}
\frac{|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\xi}_1|}{\sqrt{\tilde{H}^{(i)}(i,i)}} &\le \frac{|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{\xi}_1^{(i)}|}{\sqrt{\tilde{H}^{(i)}(i,i)}} + \frac{|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(\bar{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})|}{\sqrt{\tilde{H}^{(i)}(i,i)}} + \frac{\|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\|\|w\hat{\xi}_1 - \bar{\xi}_1^{(i)}\|}{\sqrt{\tilde{H}^{(i)}(i,i)}} \\
&\le \frac{|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{\xi}_1^{(i)}|}{\sqrt{\tilde{H}^{(i)}(i,i)}} + \frac{|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(\bar{\xi}_1^{(i)} - \tilde{\xi}_1^{(i)})|}{\sqrt{\tilde{H}^{(i)}(i,i)}} + \frac{C}{\sqrt{n\bar{\theta}^2}}\|w\hat{\xi}_1 - \bar{\xi}_1^{(i)}\|,
\end{aligned}
$$

In (E.15) and (E.19), we have seen that the first two terms are bounded by

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

up to some constant. Combining the above gives

$$
\frac{|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\xi}_1|}{\sqrt{\tilde{H}^{(i)}(i,i)}} \le C\widetilde{\kappa}_i\big(1 + n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(w\hat{\xi}_1 - \tilde{\xi}_1^{(i)})\|_\infty\big) + \Big(\frac{\log(n)}{n\bar{\theta}^2} + \frac{1}{\sqrt{n\bar{\theta}^2}}\Big)\|\bar{\xi}_1^{(i)} - w\hat{\xi}_1\|.
\tag{E.28}
$$

We plug it into (E.26) and move all terms of $\|\bar{\xi}_1^{(i)} - w\hat{\xi}_1\|$ to the left hand side. It follows that

$$\|\bar{\xi}_1^{(i)} - w\hat{\xi}_1\| \leq CK\lambda_1^{-1}(PG)\Big(\widetilde{\kappa}_i\big(1+n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(w\hat{\xi}_1-\tilde{\xi}_1^{(i)})\|_\infty\big)+\frac{|w\hat{\xi}_1(i)-\tilde{\xi}_1^{(i)}(i)|}{\sqrt{n\bar{\theta}^2}}\Big)+C\|(I_n-X)\hat{\xi}_1\| \tag{E.29}$$

Below, we bound $\|(I_n - X)\hat{\xi}_1\|$. Note that

$$\|(I_n - X)\hat{\xi}_1\| \leq \|(I_n - X)\tilde{\xi}_1^{(i)}\| + \|(I_n - X)(w\hat{\xi}_1 - \tilde{\xi}_1^{(i)})\|, \tag{E.30}$$

where

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

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

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

Recall the bound of $|X(j,j) - 1|$ in (E.16). It follows that over the event E,

$$(J_1) \leq C\sum_{j=1}^n \frac{|A(i,j) + \theta_i\theta_j + \theta_i\bar{\theta} + \frac{\log(n)}{n}|^2 (\tilde{\xi}_1^{(i)}(j))^2}{[\tilde{H}^{(i)}(j,j)]^2}$$

$$\leq C\sum_{j=1}^n \big(A(i,j) + \theta_i\theta_j + \theta_i\bar{\theta} + \frac{\log(n)}{n}\big)\frac{(\tilde{\xi}_1^{(i)}(j))^2}{[\tilde{H}^{(i)}(j,j)]^2},$$

$$(J_2) \leq C\sum_{j=1}^n \frac{|A(i,j) + \theta_i\theta_j + \theta_i\bar{\theta} + \frac{\log(n)}{n}|^2}{\tilde{H}^{(i)}(j,j)}$$

$$\leq C\sum_{j=1}^n \big(A(i,j) + \theta_i\theta_j + \theta_i\bar{\theta} + \frac{\log(n)}{n}\big)\frac{1}{\tilde{H}^{(i)}(j,j)},$$

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 D.1). For (J_1),

- The mean is bounded by (up to some constant)

$$\sum_{j=1}^n \Big(\theta_i\theta_j + \theta_i\bar{\theta} + \frac{\log(n)}{n}\Big)\frac{1}{(n\bar{\theta})^3(\bar{\theta} \vee \theta_j)} \leq \frac{1}{n\bar{\theta}^2} \cdot \frac{n\bar{\theta}\theta_i + \log(n)}{(n\bar{\theta})^2};$$

- The variance is bounded by (up to some constant)

$$\sum_{j=1}^{n} \frac{\theta_i \theta_j}{(n\bar{\theta})^6 (\bar{\theta} \vee \theta_j)^2} \leq \frac{1}{(n\bar{\theta}^2)^2} \cdot \frac{\theta_i}{(n\bar{\theta})^3};$$

- Each individual term is bounded by (up to some constant) $\frac{|(\tilde{\xi}_1^{(i)}(j))^2|}{[\tilde{H}^{(i)}(j,j)]^2} \leq \frac{1}{n\bar{\theta}^2} \cdot \frac{1}{(n\bar{\theta})^2}$.

We then have

$$(J_1) \leq \frac{C}{n\bar{\theta}^2} \frac{\theta_i}{n\bar{\theta}} \left(1 + \frac{\log(n)}{n\bar{\theta}\theta_i}\right). \tag{E.31}$$

For (J_2),

- The mean is bounded by (up to some constant)

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

- The variance is bounded by (up to some constant)

$$\sum_{j=1}^{n} \frac{\theta_i \theta_j}{(n\bar{\theta})^2 (\bar{\theta} \vee \theta_j)^2} \leq \frac{1}{n\bar{\theta}^2} \cdot \frac{\theta_i}{\bar{\theta}};$$

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

It follows that

$$(J_2) \leq \frac{C\theta_i}{\bar{\theta}} + \frac{C\log(n)}{n\bar{\theta}^2} \tag{E.32}$$

Plugging (E.31)-(E.32) into (E.30), we find out that

$$\|(I_n - X)\hat{\xi}_1\| \leq C \frac{\widetilde{\kappa}_i}{\sqrt{\log(n)}} \left(1 + n\bar{\theta} \|(\tilde{H}^{(i)})^{-\frac{1}{2}} (w\hat{\xi}_1 - \tilde{\xi}_1^{(i)})\|_\infty\right) \tag{E.33}$$

We plug (E.33) into (E.29), together with the assumption $K^{-1}\lambda_1(PG) \asymp 1$, to get

$$\|\overline{\xi}_1^{(i)} - w\hat{\xi}_1\| \leq C\widetilde{\kappa}_i\left(1 + n\bar{\theta}\|(\tilde{H}^{(i)})^{-1/2}(w\hat{\xi}_1 - \tilde{\xi}_1^{(i)})\|_\infty\right) + \frac{1}{\sqrt{n\bar{\theta}^2}} |w\hat{\xi}_1(i) - \tilde{\xi}_1^{(i)}(i)| \tag{E.34}$$

over the event E, which proved (36) by considering all i's altogether.

E.4 Proof of the second claim in Theorem 4.1

In this section, we show the proof of (J.4). Similarly to the proof of (J.3), 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, i.e., the maximum row-wise ℓ^2-norm of a matrix. Specifically, for any matrix $A \in \mathbb{R}^{n\times m}$ and vector $x \in \mathbb{R}^m$, $\|A\|_{2\to\infty} := \max_{\|x\|=1} \|Ax\|_\infty = \max_i \|A(i)\|$.

Lemma E.1. *Suppose the assumptions in Theorem 4.1 hold. Recall* $\kappa_t := \sqrt{\frac{\log(n)}{n\bar{\theta}^2}}\sqrt{\frac{\theta_t}{n\bar{\theta}}}$ *for* $1 \le t \le n$. *With probability* $1 - o(n^{-3})$, *simultaneously for* $1 \le i, t \le n$,

$$\|\tilde{\Xi}_1^{(i)}(t)O_2^{(i)} - \Xi_1(t)\| \le CK^{\frac{3}{2}}\beta_n^{-1}\kappa_t\Big(1 \wedge \sqrt{\frac{\theta_t}{\bar{\theta}}}\Big), \tag{E.35}$$

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

Lemma E.2. *Under the assumptions in Theorem 4.1. With probability* $1 - o(n^{-3})$, *simultaneously for* $1 \le i \le n$,

$$\|\hat{\Xi}_1(i) - \tilde{\Xi}_1^{(i)}(i)O_3^{(i)})\| \le CK^{\frac{3}{2}}\beta_n^{-1}\kappa_i + CK\beta_n^{-1}\|e_i'\Delta\hat{\Xi}_1\|, \tag{E.36}$$

$$\|e_i'\Delta\hat{\Xi}_1\| \le \|e_i'\Delta\tilde{\Xi}_1^{(i)}\| + \|e_i'\Delta(\overline{\Xi}_1^{(i)} - \tilde{\Xi}_1^{(i)}O_4^{(i)})\| + \frac{C}{\sqrt{n\bar{\theta}^2}}\|\hat{\Xi}_1 - \overline{\Xi}_1^{(i)}O_5^{(i)}\|, \tag{E.37}$$

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* $\Delta \equiv \Delta(i) := (\tilde{H}^{(i)})^{-1/2}WH^{-1/2}$ *for short.*

Lemma E.3. *Under the assumptions in Lemma E.2. Recall the notation of* $\widetilde{\kappa}_i$ *in (34) for* $1 \le i \le n$. *With probability* $1 - o(n^{-3})$, *simultaneously for* $1 \le i \le n$,

$$\|e_i'\Delta\tilde{\Xi}_1^{(i)}\| \le CK^{\frac{1}{2}}\widetilde{\kappa}_i \tag{E.38}$$

$$\|e_i'\Delta(\overline{\Xi}_1^{(i)} - \tilde{\Xi}_1^{(i)}O_4^{(i)})\| \le CK^{\frac{1}{2}}\Big(\widetilde{\kappa}_i\big(1 + n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\big) + \frac{\log(n)}{n\bar{\theta}^2}\|\hat{\Xi}_1 - \overline{\Xi}_1^{(i)}O_5^{(i)}\|\Big), \tag{E.39}$$

$$\|\hat{\Xi}_1 - \overline{\Xi}_1^{(i)}O_5^{(i)}\| \le CK^{\frac{3}{2}}\beta_n^{-1}\,\widetilde{\kappa}_i\Big(1 + n\bar{\theta}\|\tilde{H}^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\Big) + \frac{CK\beta_n^{-1}}{\sqrt{n\bar{\theta}^2}}\|\hat{\Xi}_1(i) - \tilde{\Xi}_1^{(i)}(i)O_3^{(i)}\|. \tag{E.40}$$

In the sequel, we will prove the second claim in Theorem 4.1 (i.e., (J.4)) based on the above lemmas. The proofs of the lemmas are postponed to the next three subsections.

Proof of (J.4) . Plugging Lemma E.3 into (E.37), we first have with probability $1-o(n^{-3})$, simultaneously for all $1 \le i \le n$,

$$\|e_i'\Delta\hat{\Xi}_1\| \le CK^{\frac{1}{2}}\Big(\tilde{\kappa}_i\big(1+n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\big)+\frac{CK\beta_n^{-1}\sqrt{\log(n)}}{n\bar{\theta}^2}\|\hat{\Xi}_1(i)-\tilde{\Xi}_1^{(i)}(i)O_3^{(i)}\|\Big)$$

which, further substituted to (E.36), implies that

$$\begin{aligned}\|\hat{\Xi}_1(i)-\tilde{\Xi}_1^{(i)}(i)O_3^{(i)})\| \le & CK^{\frac{3}{2}}\beta_n^{-1}\widetilde{\kappa}_i\big(1+n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\big)\\ &+\frac{CK^3\sqrt{\log(n)}}{\beta_n^2 n\bar{\theta}^2}\|\hat{\Xi}_1(i)-\tilde{\Xi}_1^{(i)}(i)O_3^{(i)})\|.\end{aligned}$$

Since $K^3\beta_n^{-2}\log(n)/n\bar{\theta}^2=o(1)$, we then arrive at

$$\|\hat{\Xi}_1(i)-\tilde{\Xi}_1^{(i)}(i)O_3^{(i)})\| \le CK^{\frac{3}{2}}\beta_n^{-1}\widetilde{\kappa}_i\big(1+n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\big).$$

Set $\tilde{O}_1 \equiv \tilde{O}_1^{(i)} := (O_3^{(i)})'O_2^{(i)}$. Using Lemma E.1 and let $t=i$, we will see that

$$\begin{aligned}\|\hat{\Xi}_1(i)-\Xi_1(i)\tilde{O}_1'\| &\le \|\hat{\Xi}_1(i)-\tilde{\Xi}_1^{(i)}(i)O_3^{(i)})\|+\|\tilde{\Xi}_1^{(i)}(i)O_2^{(i)}-\Xi_1(i)\|\\ &\le CK^{\frac{3}{2}}\beta_n^{-1}\widetilde{\kappa}_i\big(1+n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\big)\end{aligned}$$

over the event E. Suppose that $\hat{\Xi}_1'\Xi_1$ has the singular value decomposition (SVD) $\hat{\Xi}_1'\Xi_1 = U'\cos\Theta V$, we define $O_1=\mathrm{sgn}(\hat{\Xi}_1'\Xi_1) := U'V$. Using sine-theta theorem (i.e., (D.13) and (D.14)), we can derive

$$\begin{aligned}\|O_1-\tilde{O}_1\| &\le \|\hat{\Xi}_1'\Xi_1-O_1\|+\|\hat{\Xi}_1'\Xi_1-\tilde{O}_1\|\\ &\le \|\hat{\Xi}_1'\Xi_1-O_1\|+\|\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_3^{(i)}\|+\|(\tilde{\Xi}_1^{(i)})'\Xi_1-O_2^{(i)}\|\\ &\le CK\beta_n^{-1}\Big(\|L_0-L\|+\|\tilde{L}^{(i)}-L\|+\|\tilde{L}^{(i)}-L_0\|\Big)\\ &\le CK\beta_n^{-1}\sqrt{\frac{\log(n)}{n\bar{\theta}^2}},\end{aligned}$$

by which, we will obtain

$$\|\hat{\Xi}_1(i)-\Xi_1(i)O_1'\| \le \|\hat{\Xi}_1(i)-\Xi_1(i)\tilde{O}_1'\|+\|\Xi_1(i)\|\cdot\|O_1-\tilde{O}_1\|$$

$$\leq CK^{\frac{3}{2}}\beta_n^{-1}\widetilde{\kappa}_i\big(1+n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\big) \tag{E.41}$$

and

$$\|H_0^{-\frac{1}{2}}\Xi_1(O_1-\tilde{O}_1)')\|_{2\to\infty} \leq \|H_0^{-\frac{1}{2}}\Xi_1\|_{2\to\infty}\cdot\|O_1-\tilde{O}_1\| \leq CK^{\frac{3}{2}}\beta_n^{-1}\frac{1}{n\bar{\theta}}\sqrt{\frac{\log(n)}{n\bar{\theta}^2}} \tag{E.42}$$

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

Applying Lemma E.1 again together with (D.9), (E.42), it is easy to deduce that

$$\begin{aligned}
&\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\\
&\leq C\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1-\Xi_1\tilde{O}_1')\|_{2\to\infty}+C\|H_0^{-\frac{1}{2}}(\Xi_1(O_2^{(i)})'-\tilde{\Xi}_1^{(i)})O_3^{(i)}\|_{2\to\infty}\\
&\leq C\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1-\Xi_1O_1')\|_{2\to\infty}+C\|H_0^{-\frac{1}{2}}\Xi_1(O_1-\tilde{O}_1)')\|_{2\to\infty}+C\|H_0^{-\frac{1}{2}}(\Xi_1(O_2^{(i)})'-\tilde{\Xi}_1^{(i)})\|_{2\to\infty}\\
&\leq C\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1-\Xi_1O_1')\|_{2\to\infty}+\frac{CK^{\frac{3}{2}}\beta_n^{-1}}{n\bar{\theta}}\sqrt{\frac{\log(n)}{n\bar{\theta}^2}}
\end{aligned}$$

Thereby, according to the condition $K^3\beta_n^{-2}\log(n)/n\bar{\theta}^2 = o(1)$, (E.41) can be further improved to

$$\|\hat{\Xi}_1(i)-\Xi_1(i)O_1'\| \leq CK^{\frac{3}{2}}\beta_n^{-1}\widetilde{\kappa}_i\big(1+n\bar{\theta}\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1-\Xi_1O_1')\|_{2\to\infty}\big). \tag{E.43}$$

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,

$$\begin{aligned}
\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1-\Xi_1O_1')\|_{2\to\infty} &= \max_i \|e_i'H_0^{-\frac{1}{2}}(\hat{\Xi}_1O_1-\Xi_1)\|\\
&\leq CK^{\frac{3}{2}}\beta_n^{-1}(n\bar{\theta})^{-1}\sqrt{\frac{\log(n)}{n\bar{\theta}^2}}+o(\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1-\Xi_1O_1'\|_{2\to\infty})
\end{aligned} \tag{E.44}$$

Rearranging both sides of (E.44), we can conclude that

$$\|H_0^{-\frac{1}{2}}(\hat{\Xi}_1-\Xi_1O_1')\|_{2\to\infty} \leq CK^{\frac{3}{2}}\beta_n^{-1}(n\bar{\theta})^{-1}\sqrt{\frac{\log(n)}{n\bar{\theta}^2}}.$$

which, further substituted into (E.43), yields (J.4) due to the condition $K^3\beta_n^{-2}\log(n)/n\bar{\theta}^2 = o(1)$. □

E.5 Proof of Lemma E.1

We state the proof of Lemma E.1 which is quite similar to Lemma 5.1 with additional attention to the non-commutative multiplication of matrices. Fix the index i, we start with the perturbation from L_0 to $\tilde{L}^{(i)}$.

$$\tilde{\Xi}_1^{(i)}\tilde{\Lambda}_1^{(i)} = \tilde{L}^{(i)}\tilde{\Xi}_1^{(i)} = \big(H_0^{\frac{1}{2}}(\tilde{H}^{(i)})^{-\frac{1}{2}}\big)L_0\big(H_0^{\frac{1}{2}}(\tilde{H}^{(i)})^{-\frac{1}{2}}\big)\tilde{\Xi}_1^{(i)} = \tilde{Y}\lambda_1\xi_1\xi_1'\tilde{Y}\tilde{\Xi}_1^{(i)} + \tilde{Y}\Xi_1\Lambda_1\Xi_1'\tilde{Y}\tilde{\Xi}_1^{(i)}$$

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

$$\tilde{\Xi}_1^{(i)}(t) = \tilde{Y}(t,t)\lambda_1\xi_1(t)\xi_1'\tilde{Y}\tilde{\Xi}_1^{(i)}(\tilde{\Lambda}_1^{(i)})^{-1} + \tilde{Y}(t,t)\Xi_1(t)\Lambda_1\Xi_1'\tilde{Y}\tilde{\Xi}_1^{(i)}(\tilde{\Lambda}_1^{(i)})^{-1}. \tag{E.45}$$

Recall (D.7), over the event E, we first crudely bound $\|\lambda_1(\tilde{\Lambda}_1^{(i)})^{-1}\|$ by $\beta_n^{-1}\lambda_1(PG)$. Then, using the estimate (E.3), we can crudely bound the first term on the RHS of (E.45) by

$$\begin{aligned}\|\tilde{Y}(t,t)\lambda_1\xi_1(t)\xi_1'\tilde{Y}\tilde{\Xi}_1^{(i)}(\tilde{\Lambda}_1^{(i)})^{-1}\| &\le C\beta_n^{-1}\lambda_1(PG)\Big(\|\tilde{Y} - I_n\||\xi_1(t)| + \|\xi_1'\tilde{\Xi}_1^{(i)}\||\xi_1(t)|\Big)\\ &\le C\beta_n^{-1}\lambda_1(PG)\kappa_i\Big(1 \wedge \sqrt{\frac{\theta_i}{\bar{\theta}}}\Big)\end{aligned} \tag{E.46}$$

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

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

For the second term on the RHS of (E.45), we have

$$\|\tilde{Y}(t,t)\Xi_1(t)\Lambda_1\Xi_1'\tilde{Y}\tilde{\Xi}_1^{(i)}(\tilde{\Lambda}_1^{(i)})^{-1} - \Xi_1(t)\Lambda_1\Xi_1'\tilde{\Xi}_1^{(i)}(\tilde{\Lambda}_1^{(i)})^{-1}\| \le C\beta_n^{-1}\lambda_1(PG)\|\tilde{Y} - I_n\|\|\Xi_1(t)\|$$

and

$$\Xi_1(t)\Lambda_1\Xi_1'\tilde{\Xi}_1^{(i)}(\tilde{\Lambda}_1^{(i)})^{-1} = \Xi_1(t)\Xi_1'L_0\,\tilde{\Xi}_1^{(i)}(\tilde{\Lambda}_1^{(i)})^{-1} = \Xi_1(t)\Xi_1'\tilde{\Xi}_1^{(i)} + \Xi_1(t)\Xi_1'(L_0 - \tilde{L}^{(i)})\tilde{\Xi}_1^{(i)}(\tilde{\Lambda}_1^{(i)})^{-1}.$$

By singular value decomposition (SVD), we write $\Xi_1'\tilde{\Xi}_1^{(i)} = U\cos\Theta\, V'$ for some orthogonal matrices U, V and diagonal matrix $\cos\Theta$ all of which are i-dependent. Setting $O_2^{(i)} = \big(\mathrm{sgn}(\Xi_1'\tilde{\Xi}_1^{(i)})\big)' := VU'$ which is an orthogonal matrix, then we obtain that

$$\|\Xi_1'\tilde{\Xi}_1^{(i)} - (O_2^{(i)})'\| \le C(K\beta_n^{-1}\|\tilde{L}^{(i)} - L_0\|)^2 \le CK\beta_n^{-1}\|\tilde{L}^{(i)} - L_0\|. \tag{E.47}$$

Here we used the fact that $K\beta_n^{-1}\|\tilde{L}^{(i)} - L_0\| \le C\beta_n^{-1}\lambda_1(PG)\|\tilde{Y} - I_n\| = o(1)$. Further we crudely bound

$$\begin{aligned}\|\Xi_1(t)\Xi_1'(L_0 - \tilde{L}^{(i)})\tilde{\Xi}_1^{(i)}(\tilde{\Lambda}_1^{(i)})^{-1}\| &\le CK\beta_n^{-1}\|\tilde{L}^{(i)} - L_0\|\|\Xi_1(t)\| \\ &\le C\beta_n^{-1}\lambda_1(PG)\|\tilde{Y} - I_n\|\|\Xi_1(t)\|\end{aligned}$$

Hence,

$$\begin{aligned}\|\tilde{Y}(t,t)\Xi_1(t)\Lambda_1\Xi_1'\tilde{Y}\tilde{\Xi}_1^{(i)}(\tilde{\Lambda}_1^{(i)})^{-1} - \Xi_1(t)(O_2^{(i)})'\| &\le C\beta_n^{-1}\lambda_1(PG)\|\tilde{Y} - I_n\|\|\Xi_1(t)\| \\ &\le C\sqrt{K}\beta_n^{-1}\lambda_1(PG)\kappa_i\Big(1 \wedge \sqrt{\frac{\theta_i}{\bar{\theta}}}\Big) \qquad \text{(E.48)}\end{aligned}$$

over the event E.

Plugging in (E.46) and (E.48) back to (E.45), we simply conclude that

$$\|\tilde{\Xi}_1^{(i)}(t)O_2^{(i)} - \Xi_1(t)\| \le C\sqrt{K}\beta_n^{-1}\lambda_1(PG)\kappa_i\Big(1 \wedge \sqrt{\frac{\theta_i}{\bar{\theta}}}\Big)$$

over the event E. Combining all i's and t's together and noting $\mathbb{P}(E) = 1 - o(n^{-3})$. This finished the proof of Lemma E.1 by further noticing that $\lambda_1(PG) \asymp K$.

E.6 Proof of Lemma E.2

In this section, we prove Lemma E.2.

Let us fix the index i. The proof of (E.37) is straightforward by the decomposition

$$\hat{\Xi}_1 = \tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)} + (\overline{\Xi}_1^{(i)} - \tilde{\Xi}_1^{(i)}O_4^{(i)})O_5^{(i)} + \hat{\Xi}_1 - \overline{\Xi}_1^{(i)}O_5^{(i)}$$

where the two orthogonal matrices $O_4^{(i)}, O_5^{(i)}$ will be specified later. We further bound

$$\begin{aligned}\|e_i'\Delta(\hat{\Xi}_1 - \overline{\Xi}_1^{(i)}O_5^{(i)})\| &\le \frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}}\|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\|\,\|X\|\,\|\hat{\Xi}_1 - \overline{\Xi}_1^{(i)}O_5^{(i)}\| \\ &\le \frac{C}{\sqrt{n\bar{\theta}^2}}\|\hat{\Xi}_1 - \overline{\Xi}_1 O_5^{(i)}\|\end{aligned}$$

over the event E, by writing $\Delta = (\tilde{H}^{(i)})^{-1/2}W(\tilde{H}^{(i)})^{-1/2}X$ and using the fact $\|X\| \le C$ and $\|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\| \le \sqrt{\theta_i/\bar{\theta}} \vee \sqrt{\log(n)/n\bar{\theta}^2}$ (see (E.27)) over the event E. This, together with the trivial identities $\|e_i'\Delta\tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)}\| = \|e_i'\Delta\tilde{\Xi}_1^{(i)}\|$ and $\|e_i'\Delta(\overline{\Xi}_1^{(i)} - \tilde{\Xi}_1^{(i)}O_4^{(i)})O_5^{(i)}\| = \|e_i'\Delta(\overline{\Xi}_1^{(i)} - \tilde{\Xi}_1^{(i)}O_4^{(i)})\|$ implies (E.37).

We then turn to show (E.36). Note that

$$\hat{\Xi}_1\hat{\Lambda}_1 = L\,\hat{\Xi}_1 = X(\tilde{H}^{(i)})^{-\frac{1}{2}}A(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1 = X(\tilde{H}^{(i)})^{-\frac{1}{2}}\Omega(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1 + X(\tilde{H}^{(i)})^{-\frac{1}{2}}(A-\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1$$

by the notation $X = (\tilde{H}^{(i)})^{\frac{1}{2}}H^{-\frac{1}{2}}$. Then,

$$\begin{aligned}\hat{\Xi}_1(i) =& X(i,i)\tilde{\lambda}_1^{(i)}\tilde{\xi}_1^{(i)}(i)\big(\tilde{\xi}_1^{(i)}\big)'X\hat{\Xi}_1\hat{\Lambda}_1^{-1} + X(i,i)\tilde{\Xi}_1^{(i)}(i)\tilde{\Lambda}^{(i)}(\tilde{\Xi}_1^{(i)})'X\hat{\Xi}_1\hat{\Lambda}_1^{-1} \\ &+ X(i,i)e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}(A-\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1\hat{\Lambda}_1^{-1}.\end{aligned} \tag{E.49}$$

Recall the estimate $\|X - I_n\| \le C\sqrt{\log(n)}/\sqrt{n\bar{\theta}^2}$ following from Lemma D.5 and the properties of eigenvalues and eigenvectors of $\tilde{L}^{(i)}$ in Lemma D.4. Then, for the first term on the RHS of (E.49), we have

$$\|X(i,i)\tilde{\lambda}_1^{(i)}\tilde{\xi}_1^{(i)}(i)\big(\tilde{\xi}_1^{(i)}\big)'X\hat{\Xi}_1\hat{\Lambda}_1^{-1}\| \le CK^{-1}\lambda_1(PG)\|\hat{\Lambda}_1^{-1}\|\big|\tilde{\xi}_1^{(i)}(i)\big|\Big(\|X - I_n\| + \|\big(\tilde{\xi}_1^{(i)}\big)'\hat{\Xi}_1\|\Big). \tag{E.50}$$

Recall that $\hat{\lambda}_j$'s for $1 \le j \le K$ share the same asymptotic as λ_j's in (D.2) over the event E. By sin-theta theorem and (E.9), we have the bound

$$\|\big(\tilde{\xi}_1^{(i)}\big)'\hat{\Xi}_1\| \le CK\lambda_1^{-1}(PG)\|L - \tilde{L}^{(i)}\| \le C\Big(\sqrt{\frac{\log(n)}{n\bar{\theta}^2}} + \frac{K\lambda_1^{-1}(PG)}{\sqrt{n\bar{\theta}^2}}\Big). \tag{E.51}$$

Thus, plugging (E.51), (D.8) together with $\|X - I_n\| \le C\sqrt{\log(n)}/\sqrt{n\bar{\theta}^2}$ into (E.50), we arrive at

$$\|X(i,i)\tilde{\lambda}_1^{(i)}\tilde{\xi}_1^{(i)}(i)\big(\tilde{\xi}_1^{(i)}\big)'X\hat{\Xi}_1\hat{\Lambda}_1^{-1}\| \le CK\beta_n^{-1}\kappa_i\Big(1 \wedge \sqrt{\frac{\theta_i}{\bar{\theta}}}\,\Big), \tag{E.52}$$

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

To estimate the other two term in (E.49), we need the assistance of $\overline{\Xi}$, the eigenspace of $(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{A}^{(i)}(\tilde{H}^{(i)})^{-\frac{1}{2}}$, which is counterpart to $\tilde{\Xi}_1^{(i)}$ and $\hat{\Xi}_1$. Recall that $\tilde{A}^{(i)} = \Omega + \tilde{W}^{(i)} - \mathrm{diag}(\Omega)$ where $\tilde{W}^{(i)}$ is obtained by zeroing-out i-th row and column of W. Similarly to

(E.47), we can then claim that there exists an orthogonal matrix $O_4^{(i)}$ by sin-theta theorem such that

$$\|\overline{\Xi}_1^{(i)} - \tilde{\Xi}_1^{(i)} O_4^{(i)}\| \leq CK\beta_n^{-1}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{A}^{(i)}(\tilde{H}^{(i)})^{-\frac{1}{2}} - (\tilde{H}^{(i)})^{-\frac{1}{2}}\Omega(\tilde{H}^{(i)})^{-\frac{1}{2}}\| \leq \frac{CK\beta_n^{-1}}{\sqrt{n\bar{\theta}^2}} \tag{E.53}$$

over the event E, where $O_4^{(i)} = \mathrm{sgn}((\tilde{\Xi}_1^{(i)})'\overline{\Xi}_1^{(i)})$. We will also need an orthogonal matrix $O_5 \equiv O_5(i) := \mathrm{sgn}((\overline{\Xi}_1^{(i)})'\hat{\Xi}_1)$. Again by sin-theta theorem,

$$\begin{aligned}\|(\overline{\Xi}_1^{(i)})'\hat{\Xi}_1 - O_5\|^{\frac{1}{2}} &\leq CK\beta_n^{-1}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{A}^{(i)}(\tilde{H}^{(i)})^{-\frac{1}{2}} - H^{-\frac{1}{2}}AH^{-\frac{1}{2}}\| \\ &\leq C\beta_n^{-1}\lambda_1(PG)\|X - I_n\| + CK\beta_n^{-1}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\tilde{A}^{(i)} - A)(\tilde{H}^{(i)})^{-\frac{1}{2}}\| \\ &\leq CK\beta_n^{-1}\sqrt{\frac{\log(n)}{n\bar{\theta}^2}}. \end{aligned} \tag{E.54}$$

Here we used $\|(\tilde{H}^{(i)})^{-\frac{1}{2}}A(\tilde{H}^{(i)})^{-\frac{1}{2}}\| \asymp \|(\tilde{H}^{(i)})^{-\frac{1}{2}}\Omega(\tilde{H}^{(i)})^{-\frac{1}{2}}\| \asymp K^{-1}\lambda_1(PG)$ 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 (E.49), similarly to (E.48), we get

$$\begin{aligned}&\|X(i,i)\tilde{\Xi}_1^{(i)}(i)\tilde{\Lambda}^{(i)}(\tilde{\Xi}_1^{(i)})'X\hat{\Xi}_1\hat{\Lambda}_1^{-1} - \tilde{\Xi}_1^{(i)}(i)O_3^{(i)}\| \\ &\leq C\big(\beta_n^{-1}\lambda_1(PG)\|X - I_n\| + K\beta_n^{-1}\|L - \tilde{L}^{(i)}\| + \|(\tilde{\Xi}^{(i)})_1'\hat{\Xi}_1 - O_3^{(i)}\|\big)\|\tilde{\Xi}_1^{(i)}(i)\| \\ &\leq CK^{\frac{3}{2}}\beta_n^{-1}\kappa_i + C\|(\tilde{\Xi}_1^{(i)})'\hat{\Xi}_1 - O_3^{(i)}\|\|\tilde{\Xi}_1^{(i)}(i)\| \end{aligned} \tag{E.55}$$

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

$$\|(\tilde{\Xi}_1^{(i)})'\hat{\Xi}_1 - O_3^{(i)}\| \leq \|\overline{\Xi}_1^{(i)} - \tilde{\Xi}_1^{(i)}O_4^{(i)}\| + \|(\overline{\Xi}_1^{(i)})'\hat{\Xi}_1 - O_5^{(i)}\|$$

which with (E.53), (E.54) and (D.8) leads to

$$\|X(i,i)\tilde{\Xi}_1^{(i)}(i)\tilde{\Lambda}^{(i)}(\tilde{\Xi}^{(i)})_1'Y\hat{\Xi}_1\hat{\Lambda}_1^{-1} - \tilde{\Xi}_1^{(i)}(i)O_3^{(i)}\| \leq CK^{\frac{3}{2}}\beta_n^{-1}\kappa_i \tag{E.56}$$

Combining (E.52) and (E.56) back into (E.49), we get

$$\|\hat{\Xi}_1(i) - \tilde{\Xi}_1^{(i)}(i)O_3^{(i)}\| \leq CK^{\frac{3}{2}}\beta_n^{-1}\kappa_i + \|X(i,i)e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}(A - \Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1\hat{\Lambda}_1^{-1}\| \tag{E.57}$$

In the sequel, we proceed to the second term on the RHS above. First, using the trivial bound $|X(i,i)| \leq 2$ and $\|\hat{\Lambda}_1\|^{-1} \leq K\beta_n^{-1}$, we have

$$\begin{aligned}
&\|X(i,i)e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}(A-\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1\hat{\Lambda}_1^{-1}\| \\
&\leq CK\beta_n^{-1}\|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}(A-\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1\| \\
&\leq CK\beta_n^{-1}\Big(\|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}W(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1\| + \|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}\mathrm{diag}(\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1\|\Big)
\end{aligned}$$

We can simply get the bound

$$\begin{aligned}
\|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}\mathrm{diag}(\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1\| &= \|(\tilde{H}^{(i)})^{-1}(i,i)\Omega(i,i)X(i,i)\hat{\Xi}_1(i)\| \\
&\leq C\,\frac{\theta_i^2}{n\bar{\theta}(\bar{\theta}\vee\theta_i)}\,\|\hat{\Xi}_1(i)\| \\
&\leq \frac{\sqrt{K}}{\sqrt{\log(n)}}\kappa_i\Big(1\wedge\sqrt{\frac{\theta_i}{\bar{\theta}}}\Big)
\end{aligned}$$

This leads to

$$\|X(i,i)e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}(A-\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}}X\hat{\Xi}_1\hat{\Lambda}_1^{-1}\| \leq CK^{\frac{3}{2}}\beta_n^{-1}\kappa_i + CK\beta_n^{-1}\|e_i'\Delta\hat{\Xi}_1\| \qquad \text{(E.58)}$$

over the event E satisfying $\mathbb{P}(E) = 1-o(n^{-3})$. Combining (E.58) and (E.57) and considering all i's, we then conclude the proof of (E.36).

E.7 Proof of Lemma E.3

The proof of Lemma E.3 is rather complicated. We will show the three claims (i.e., (E.38)-(E.40)) separately in the following three parts.

E.7.1 Proof of (E.38)

Write $\Delta = (\tilde{H}^{(i)})^{-\frac{1}{2}}W(\tilde{H}^{(i)})^{-\frac{1}{2}}X$, we first crudely have

$$\|e_i'\Delta\tilde{\Xi}_1^{(i)}\| \leq \|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}W(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{\Xi}_1^{(i)}\| + \|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}W(\tilde{H}^{(i)})^{-\frac{1}{2}}(X-I_n)\tilde{\Xi}_1^{(i)}\| \qquad \text{(E.59)}$$

We start with the first term on the RHS of (E.59).

$$\|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}W(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{\Xi}_1^{(i)}\| = \Big\|\frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}}W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{\Xi}_1^{(i)}\Big\| = \Big\|\frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}}\sum_{t=1}^{n}\frac{W(i,t)}{\sqrt{\tilde{H}^{(i)}(t,t)}}\tilde{\Xi}_1^{(i)}(t)\Big\|.$$

Thanks to the independence between $\tilde{\Xi}_1^{(i)}$ and $W(i)$, we can estimate $\sum_{t=1}^{n} \frac{W(i,t)}{\sqrt{\tilde{H}^{(i)}(t,t)}} \tilde{\Xi}_1^{(i)}(t)$ componentwisely by Bernstein inequality with respect to the randomness of $W(i)$. For each $2 \le p \le K$, we can bound the variance of $\sum_{t=1}^{n} \frac{W(i,t)}{\sqrt{\tilde{H}^{(i)}(t,t)}} \tilde{\xi}_p^{(i)}(t)$ by

$$\mathrm{var}\Big(\sum_{t=1}^{n} \frac{W(i,t)}{\sqrt{\tilde{H}^{(i)}(t,t)}} \tilde{\xi}_p^{(i)}(t)\Big) \asymp \sum_{t=1}^{n} \frac{\theta_i \theta_t}{\tilde{H}^{(i)}(t,t)} \big(\tilde{\xi}_p^{(i)}(t)\big)^2 \le C \sum_{t=1}^{n} \frac{\theta_i \theta_t}{n\bar{\theta}(\theta_t \vee \bar{\theta})} \big(\tilde{\xi}_p^{(i)}(t)\big)^2 \le C \frac{\theta_i}{n\bar{\theta}}.$$

Each individual summand can be bounded by $C/n\bar{\theta}$ over the event E. As a result,

$$\Big|\sum_{t=1}^{n} \frac{W(i,t)}{\sqrt{\tilde{H}^{(i)}(t,t)}} \tilde{\xi}_p^{(i)}(t)\Big| \le C\Big(\sqrt{\frac{\theta_i \log(n)}{n\bar{\theta}}} + \frac{\log(n)}{n\bar{\theta}}\Big) \le C \frac{\sqrt{\log(n)}}{n\bar{\theta}} \sqrt{n\bar{\theta}\theta_i \vee \log(n)}$$

Further with $\tilde{H}^{(i)}(i,i) \asymp n\bar{\theta}(\theta_i \vee \bar{\theta})$, we finally conclude that

$$\|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}} W (\tilde{H}^{(i)})^{-\frac{1}{2}} \tilde{\Xi}_1^{(i)}\| = \Big\| \frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}} \sum_{t=1}^{n} \frac{W(i,t)}{\sqrt{\tilde{H}^{(i)}(t,t)}} \tilde{\Xi}_1^{(i)}(t) \Big\| \le C K^{\frac{1}{2}} \widetilde{\kappa}_i. \tag{E.60}$$

Next, regarding the term $\|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}} W (\tilde{H}^{(i)})^{-\frac{1}{2}} (X - I_n) \tilde{\Xi}_1^{(i)}\|$, using the estimate (E.16), we can derive

$$\begin{aligned}
\frac{\|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(X - I_n)\tilde{\Xi}_1^{(i)}\|}{\sqrt{\tilde{H}^{(i)}(i,i)}} &\le \frac{C}{\sqrt{\tilde{H}^{(i)}(i,i)}} \sum_{t=1, t\neq i}^{n} |W(i,t)| \frac{A(i,t) + \theta_i\theta_t + \theta_i\bar{\theta} + \frac{\log(n)}{n}}{|\tilde{H}^{(i)}(t,t)|^{\frac{3}{2}}} \|\tilde{\Xi}_1^{(i)}(t)\| \\
&\le \frac{C}{\sqrt{\tilde{H}^{(i)}(i,i)}} \sum_{t=1, t\neq i}^{n} \frac{A(i,t) + \theta_i\theta_t + \theta_i\bar{\theta} + \frac{\log(n)}{n}}{|\tilde{H}^{(i)}(t,t)|^{\frac{3}{2}}} \|\tilde{\Xi}_1^{(i)}(t)\| \\
&\le C\sqrt{K} \frac{\widetilde{\kappa}_i}{\sqrt{\log(n)}}
\end{aligned} \tag{E.61}$$

where the last step is analogous to how we get (E.17) by Bernstein's inequality and one can refer to the details in Section E.3.1. Combining (E.60) and (E.61) into (E.59), and considering all i's, we thus conclude (E.38).

E.7.2 Proof of (E.39)

The proof is similar to the proof of (35) in Section E.3.2. First, by definition, we bound

$$\|e_i'\Delta(\overline{\Xi}_1^{(i)} - \tilde{\Xi}_1^{(i)} O_4^{(i)})\| \le \|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}} W (\tilde{H}^{(i)})^{-\frac{1}{2}} (\overline{\Xi}_1^{(i)} - \tilde{\Xi}_1^{(i)} O_4^{(i)})\|$$

$$+\,\|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}W(\tilde{H}^{(i)})^{-\frac{1}{2}}(X-I_n)(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})\|. \tag{E.62}$$

We rewrite the first term on the RHS by

$$\|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}W(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})\| = \Big\|\frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}}W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})\Big\|.$$

According to the definition of $\overline{\Xi}_1^{(i)}$, $\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)}$ 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 $W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})$ by

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

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

$$\sum_t(\tilde{\Xi}_1^{(i)}(t)O_4^{(i)}e_p)^2 = e_p'(O_4^{(i)})'(\tilde{\Xi}_1^{(i)})'\sum_t e_te_t'\tilde{\Xi}_1^{(i)}O_4^{(i)}e_p = e_p'(O_4^{(i)})'(\tilde{\Xi}_1^{(i)})'\tilde{\Xi}_1^{(i)}O_4^{(i)}e_p = 1.$$

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 $W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})$ is bounded by $C\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})e_p\|_\infty$. Thus, for each $1\le p\le K-1$,

$$\big|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})e_p\big| \le C\sqrt{\frac{\theta_i\log(n)}{n\bar{\theta}}}+C\log(n)\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})e_p\|_\infty;$$

and therefore,

$$\begin{aligned}\|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})\| &\le C\Big(\sum_{p=1}^{K-1}\Big(\sqrt{\frac{\theta_i\log(n)}{n\bar{\theta}}}+\log(n)\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})e_p\|_\infty\Big)^2\Big)^{\frac{1}{2}}\\ &\le C\sqrt{K\log(n)}\sqrt{\frac{\theta_i}{n\bar{\theta}}}+C\log(n)\Big(\sum_{p=1}^{K-1}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})e_p\|_\infty^2\Big)^{\frac{1}{2}}.\end{aligned}$$

We further have

$$\begin{aligned}\Big(\sum_{p=1}^{K-1}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})e_p\|_\infty^2\Big)^{\frac{1}{2}} &\le \Big(\sum_{p=1}^{K-1}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)}\|_{2\to\infty}^2\Big)^{\frac{1}{2}}\\ &\le \sqrt{K}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)})\|_{2\to\infty}+\frac{C\sqrt{K}}{\sqrt{n\bar{\theta}^2}}\|\hat{\Xi}_1-\overline{\Xi}_1^{(i)}O_5^{(i)}\|.\end{aligned}$$

Thus, over the event E,

$$
\begin{aligned}
&\|e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}W(\tilde{H}^{(i)})^{-\frac{1}{2}}(\Xi_1^{(i)} - \tilde{\Xi}_1^{(i)}O_4^{(i)})\| \\
&\leq C\sqrt{K}\,\kappa_i + C\sqrt{K}\frac{\log(n)}{\sqrt{n\bar{\theta}(\bar{\theta}\vee\theta_i)}}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)})\|_{2\to\infty} + \frac{C\sqrt{K}\,\log(n)}{n\bar{\theta}^2}\|\hat{\Xi}_1 - \Xi_1^{(i)}O_5^{(i)}\|
\end{aligned} \tag{E.63}
$$

Next, for the second term of (E.62), using the estimate (E.16), we have

$$
\begin{aligned}
&\frac{\|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(X - I_n)(\Xi_1 - \tilde{\Xi}_1^{(i)}O_4^{(i)})\|}{\sqrt{\tilde{H}^{(i)}(i,i)}} \\
&\leq \frac{C}{\sqrt{\tilde{H}^{(i)}(i,i)}}\sum_{t=1,t\neq i}^{n}|W(i,t)|\,\frac{A(i,t)+\theta_i\theta_t+\theta_i\bar{\theta}+\log(n)/n}{\tilde{H}^{(i)}(t,t)}\,\frac{\|\Xi_1(t) - \tilde{\Xi}_1^{(i)}(t)O_4^{(i)}\|}{\sqrt{\tilde{H}^{(i)}(t,t)}} \\
&\leq \frac{C}{\sqrt{\tilde{H}^{(i)}(i,i)}}\sum_{t=1,t\neq i}^{n}\frac{A(i,t)+\theta_i\theta_t+\theta_i\bar{\theta}+\frac{\log(n)}{n}}{\tilde{H}^{(i)}(t,t)}\Big(\big\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)})\big\|_{2\to\infty} + \frac{\|\hat{\Xi}_1(t) - \Xi_1^{(i)}(t)O_5^{(i)}\|}{\sqrt{\tilde{H}^{(i)}(t,t)}}\Big)
\end{aligned} \tag{E.64}
$$

Similarly to the derivations of upper bounds of (E.20) and (E.21), we bound the two sums on the RHS of (E.64) corresponding to the two terms in the parenthesis separately as follows:

$$
\begin{aligned}
&\frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}}\sum_{t=1,t\neq i}^{n}\frac{A(i,t)+\theta_i\bar{\theta}+\log(n)/n}{\tilde{H}^{(i)}(t,t)}\big\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)})\big\|_{2\to\infty} \\
&\leq \frac{\theta_i/\bar{\theta}+\log(n)/n\bar{\theta}^2}{\sqrt{n\bar{\theta}(\bar{\theta}\vee\theta_i)}}\,\big\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)})\big\|_{2\to\infty} \\
&\leq \Big(\frac{1}{\bar{\theta}}\sqrt{\frac{\theta_i}{n\bar{\theta}}} + \frac{\log(n)}{(n\bar{\theta}^2)^{\frac{3}{2}}}\Big)\big\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)})\big\|_{2\to\infty}
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}}\sum_{t=1,t\neq i}^{n}\frac{A(i,t)+\theta_i\bar{\theta}+\log(n)/n}{(\tilde{H}^{(i)}(t,t))^{3/2}}\|\hat{\Xi}_1(t) - \Xi_1^{(i)}(t)O_5^{(i)}\| \\
&\leq \frac{1}{\sqrt{\tilde{H}^{(i)}(i,i)}}\Big(\sum_{t=1,t\neq i}^{n}\frac{\big(A(i,t)+\theta_i\bar{\theta}+\log(n)/n\big)^2}{(\tilde{H}^{(i)}(t,t))^3}\Big)^{\frac{1}{2}}\Big(\sum_{t=1,t\neq i}^{n}\|\hat{\Xi}_1(t) - \Xi_1^{(i)}(t)O_5^{(i)}\|^2\Big)^{\frac{1}{2}} \\
&\leq C(n\bar{\theta}^2)^{-\frac{3}{2}}\Big(\mathrm{tr}(\hat{\Xi}_1 - \Xi_1^{(i)}O_5^{(i)})'(\hat{\Xi}_1 - \Xi_1^{(i)}O_5^{(i)})\Big)^{\frac{1}{2}} \\
&\leq C\sqrt{K}(n\bar{\theta}^2)^{-\frac{3}{2}}\|\hat{\Xi}_1 - \Xi_1^{(i)}O_5^{(i)}\|,
\end{aligned}
$$

over the event E, in which, we applied (E.23) and (E.22). We plug the above two estimates into (E.64) and conclude that over the event E,

$$\frac{\|W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(X-I_n)(\overline{\Xi}_1^{(i)}-\tilde{\Xi}_1^{(i)}O_4^{(i)})\|}{\sqrt{\tilde{H}^{(i)}(i,i)}}$$
$$\leq C\widetilde{\kappa}_i n\bar{\theta}\,\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_4^{(i)}O_5^{(i)})\|_{2\to\infty}+\sqrt{K}(n\bar{\theta}^2)^{-\frac{3}{2}}\|\hat{\Xi}_1-\overline{\Xi}_1^{(i)}O_5^{(i)}\|$$

This, together with (E.63), concludes the proof of (E.39) for fixed i, by the fact that $\log(n)/\sqrt{n\bar{\theta}^2}\leq\widetilde{\kappa}_i\cdot n\bar{\theta}$. Combining all i's and the fact $\mathbb{P}(E)=1-o(n^{-3})$, we finish the proof.

E.7.3 Proof of (E.40)

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 (E.25)), analogously to (E.26), we first have

$$\begin{aligned}
&\|\hat{\Xi}_1-\overline{\Xi}_1^{(i)}O_5^{(i)}\|\\
&\leq K\beta_n^{-1}\|\big(H^{-\frac{1}{2}}AH^{-\frac{1}{2}}-(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{A}^{(i)}(\tilde{H}^{(i)})^{-\frac{1}{2}}\big)\hat{\Xi}_1\|\\
&\leq K\beta_n^{-1}\Big(\|(I_n-X^{-1})H^{-\frac{1}{2}}AH^{-\frac{1}{2}}\hat{\Xi}_1\|+\|(\tilde{H}^{(i)})^{-\frac{1}{2}}A(\tilde{H}^{(i)})^{-\frac{1}{2}}(X-I_n)\hat{\Xi}_1\|+\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(A-\tilde{A}^{(i)})(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\Xi}_1\|\Big)\\
&\leq CK\beta_n^{-1}\Big(\|(X-I_n)\hat{\Xi}_1\hat{\Lambda}_1\|+\|(\tilde{H}^{(i)})^{-\frac{1}{2}}A(\tilde{H}^{(i)})^{-\frac{1}{2}}(X-I_n)\hat{\Xi}_1\|+\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(e_iW(i)+W(i)'e_i')(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\Xi}_1\|\Big).
\end{aligned}\tag{E.65}$$

We start with a simple derivation,

$$\begin{aligned}
&\|(\tilde{H}^{(i)})^{-\frac{1}{2}}A(\tilde{H}^{(i)})^{-\frac{1}{2}}(X-I_n)\hat{\Xi}_1\|\\
&\leq\|(\tilde{H}^{(i)})^{-\frac{1}{2}}\Omega(\tilde{H}^{(i)})^{-\frac{1}{2}}\|\|(X-I_n)\hat{\Xi}_1\|+\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(A-\Omega)(\tilde{H}^{(i)})^{-\frac{1}{2}}\|\|(X-I_n)\hat{\Xi}_1\|\\
&\leq CK^{-1}\lambda_1(PG)\|(X-I_n)\hat{\Xi}_1\|;
\end{aligned}$$

Second, we have

$$\begin{aligned}
\|(\tilde{H}^{(i)})^{-\frac{1}{2}}W(i)'e_i'(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\Xi}_1\|&=\tilde{H}^{(i)}(i,i)^{-\frac{1}{2}}\|\hat{\Xi}_1(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}W(i)'\|\\
&\leq\frac{C}{\sqrt{n\bar{\theta}^2}}\|\hat{\Xi}_1(i)\|
\end{aligned}$$

$$\leq \frac{C\sqrt{K}\,\kappa_i}{\sqrt{\log(n)}} + \frac{C}{\sqrt{n\bar{\theta}^2}}\|\hat{\Xi}_1(i) - \tilde{\Xi}_1^{(i)}(i)O_3^{(i)}\|.$$

where in the second step we used (E.27) and we decomposed $\hat{\Xi}_1(i)$ as $\tilde{\Xi}^{(i)}O_3^{(i)} + \hat{\Xi}_1(i) - \tilde{\Xi}_1^{(i)}(i)O_3^{(i)}$ and employed (D.8) in the last step. Thus, we further bound the RHS of (E.65) as

$$\begin{aligned}\|\hat{\Xi}_1 - \overline{\Xi}_1^{(i)}O_5^{(i)}\| \leq & C\beta_n^{-1}\lambda_1(PG)\|(X-I_n)\hat{\Xi}_1\| + CK\beta_n^{-1}\|\tilde{H}^{(i)}(i,i)^{-\frac{1}{2}}W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\Xi}_1\| \\ & + \frac{CK^{\frac{3}{2}}\beta_n^{-1}\kappa_i}{\sqrt{\log(n)}} + \frac{CK\beta_n^{-1}}{\sqrt{n\bar{\theta}^2}}\|\hat{\Xi}_1(i) - \tilde{\Xi}_1^{(i)}(i)O_3^{(i)}\|. \end{aligned} \tag{E.66}$$

In the sequel, we analyze the first two terms on the RHS above. For $\|(X-I_n)\hat{\Xi}_1\|$, similarly to (E.30), we decompose $\hat{\Xi}_1$ and get that

$$\|(X-I_n)\hat{\Xi}_1\| \leq \|(X-I_n)\tilde{\Xi}_1^{(i)}\| + \|(X-I_n)(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_3^{(i)})\|$$

Then, we replicate the derivations for the two terms in (E.30) with $\tilde{\xi}_1^{(i)}$, $\hat{\xi}_1$ replaced by $\tilde{\Xi}_1^{(i)}$, $\hat{\Xi}_1$ and w replaced by O_3' to get

$$\begin{aligned}\|(X-I_n)\tilde{\Xi}_1^{(i)}\|^2 &\leq C\sum_{j=1}^{n}\Big(A(i,j) + \theta_i\theta_j + \theta_i\bar{\theta} + \frac{\log(n)}{n}\Big)\frac{\|\tilde{\Xi}_1^{(i)}(j)\|^2}{[\tilde{H}^{(i)}(j,j)]^2} \leq \frac{CK\widetilde{\kappa}_i^2}{\log(n)} \\ \|(X-I_n)(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_3^{(i)})\|^2 &\leq C\|(\tilde{H}^{(i)})^{-1/2}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}^2\sum_{j=1}^{n}\frac{A(i,j) + \theta_i\theta_j + \theta_i\bar{\theta} + \frac{\log(n)}{n}}{\tilde{H}^{(i)}(j,j)} \\ &\leq C\|(\tilde{H}^{(i)})^{-1/2}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}^2 \cdot \frac{n\bar{\theta}\theta_i + \log(n)}{n\bar{\theta}^2}\end{aligned}$$

over the event E. More detailed steps can be referred to derivations from (E.30)-(E.32). We thereby arrive at

$$\|(X-I_n)\hat{\Xi}_1\| \leq C\frac{\sqrt{K}}{\sqrt{\log(n)}}\widetilde{\kappa}_i\Big(1 + n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\Big) \tag{E.67}$$

Now we turn to study the term $\|\tilde{H}^{(i)}(i,i)^{-\frac{1}{2}}W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\Xi}_1\|$. Using (E.60), (E.63) and (E.27), we can deduce that

$$\begin{aligned}\|\tilde{H}^{(i)}(i,i)^{-\frac{1}{2}}W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\hat{\Xi}_1\| \leq & \|\tilde{H}^{(i)}(i,i)^{-\frac{1}{2}}W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}\tilde{\Xi}_1^{(i)}\| + \|\tilde{H}^{(i)}(i,i)^{-\frac{1}{2}}W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(\overline{\Xi}_1^{(i)} - \tilde{\Xi}_1^{(i)}O_4^{(i)})\| \\ & + \|\tilde{H}^{(i)}(i,i)^{-\frac{1}{2}}W(i)(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1 - \overline{\Xi}_1^{(i)}O_5^{(i)})\|\end{aligned}$$

$$\le C\sqrt{K}\,\widetilde{\kappa}_i\Big(1+n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\Big)$$
$$+C\Big(\sqrt{K}\,\frac{\log(n)}{n\bar{\theta}^2}+\frac{1}{\sqrt{n\bar{\theta}^2}}\Big)\|\hat{\Xi}_1-\overline{\Xi}_1^{(i)}O_5^{(i)}\| \qquad \text{(E.68)}$$

over the event E. Combining (E.67) and (E.68) into (E.66) and putting all terms equipped with factor $\|\hat{\Xi}_1-\overline{\Xi}_1^{(i)}O_5^{(i)}\|$ to the LHS, under the condition $K^3\beta_n^{-2}\log(n)/n\bar{\theta}^2=o(1)$ and $\lambda_1(PG)\le CK$, we finally see that

$$\|\hat{\Xi}_1-\overline{\Xi}_1^{(i)}O_5^{(i)}\|\le CK^{\frac{3}{2}}\beta_n^{-1}\,\widetilde{\kappa}_i\Big(1+n\bar{\theta}\|(\tilde{H}^{(i)})^{-\frac{1}{2}}(\hat{\Xi}_1-\tilde{\Xi}_1^{(i)}O_3^{(i)})\|_{2\to\infty}\Big)+\frac{CK\beta_n^{-1}}{\sqrt{n\bar{\theta}^2}}\|\hat{\Xi}_1(i)-\tilde{\Xi}_1^{(i)}(i)O_3^{(i)}\|$$

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

F Rate of Mixed-SCORE-Laplacian

We prove the error rate of Mixed-SCORE-Laplacian in this Section. In detail, in Section F.1 we prove Lemma 4.1; in Section F.2, we prove the first claim of Theorem 4.2; in Section F.3, we briefly state the proofs of Corollary 4.1 and the second claim of Theorem 4.2, as these arguments directly stem from Theorem 4.2.

F.1 Proof of Lemma 4.1

Fix the choice of $\hat{\xi}_1$ such that $w=1$ in (J.3). Choose the orthogonal matrix O_1 appeared in Theorem 4.1. By definition,

$$\|O_1'\hat{r}_i-r_i\|=\|e_i'\big(\hat{\Xi}_1O_1/\hat{\xi}_1(i)-\Xi_1/\xi_1(i)\big)\|\le\|e_i'(\hat{\Xi}_1O_1-\Xi_1)/\hat{\xi}_1(i)\|+\|\Xi_1(i)\|\Big|\frac{1}{\hat{\xi}_1(i)}-\frac{1}{\xi_1(i)}\Big|$$

Employing Theorem 4.1 with Lemma D.2, for $i\in S_n(c_0)$, we have

$$\|e_i'(\hat{\Xi}_1O_1-\Xi_1)/\hat{\xi}_1(i)\|\le C\frac{\|e_i'(\hat{\Xi}_1O_1-\Xi_1)\|}{\xi_1(i)}\le C\sqrt{\frac{K^3\log(n)}{n\bar{\theta}(\bar{\theta}\wedge\theta_i)\beta_n^2}}$$

and

$$\|\Xi_1(i)\|\Big|\frac{1}{\hat{\xi}_1(i)}-\frac{1}{\xi_1(i)}\Big|\le C\frac{\|\Xi_1(i)\|}{\xi_1(i)}\cdot\frac{|\hat{\xi}_1(i)-\xi_1(i)|}{\xi_1(i)}\le C\sqrt{\frac{K^3\log(n)}{n\bar{\theta}(\bar{\theta}\wedge\theta_i)\beta_n^2}}$$

with probability $1-o(n^{-3})$ simultaneously for $i\in S_n(c_0)$. Combining the above inequalities, we immediately get (20) simultaneously for $i\in S_n(c_0)$, with probability $1-o(n^{-3})$.

F.2 Proof of the first claim in Theorem 4.2

This theorem has two claims, one is about the node-wise error, and the other is about the bound for the ℓ^1-loss. The second claim follows easily from the first claim, and its proof is relegated to Section F.3. We now prove the first claim about the node-wise error.

We only focus on $i \in \hat{S}_n(c)$ (see (15)). For $i \notin \hat{S}_n(c)$, since we take trivial estimator $K^{-1}\mathbf{1}_K$, the estimation error is then trivially bounded by some constant. Recall the definition, for $i \in \hat{S}_n(c)$,

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

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 the errors of $\hat{w}_i$'s and $\hat{b}_1$ compared to w_i's and b_1 separately.

Suppose we are under the high probability $1 - o(n^{-3})$ event in which Lemma 4.1 holds. We refrain ourselves from stating the high probability in the following derivations. We first study $\hat{w}_i$'s. Thanks to the choice of a variant of successive projection in our vertex hunting algorithm, referring to Lemma 3.1 of [6], it is easy to deduce that

$$\|\mathsf{P}\widehat{V}O_1 - V\|_{2\to\infty} \leq C \max_{i\in\hat{S}_n^*(c,\gamma)} \|O_1'\hat{r}_i - r_i\| \leq C\sqrt{\frac{K^3\log(n)}{n\bar{\theta}^2\beta_n^2}}. \tag{F.1}$$

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

$$\hat{Q}\hat{w}_i = \begin{pmatrix} 1 \\ O_1'\hat{r}_i \end{pmatrix}, \qquad \hat{Q} := \begin{pmatrix} 1 & 1 & \cdots & 1 \\ O_1'\hat{v}_1 & O_1'\hat{v}_2 & \cdots & O_1'\hat{v}_K \end{pmatrix}$$

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

$$Qw_i = \begin{pmatrix} 1 \\ r_i \end{pmatrix}, \qquad Q := \begin{pmatrix} 1 & 1 & \cdots & 1 \\ v_1 & v_2 & \cdots & v_K \end{pmatrix}.$$

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,

$$\begin{aligned}\|\mathsf{P}\hat{w}_i - w_i\| &= \left\| (\hat{Q}\mathsf{P}')^{-1} \begin{pmatrix} 1 \\ O_1'\hat{r}_i \end{pmatrix} - Q^{-1} \begin{pmatrix} 1 \\ r_i \end{pmatrix} \right\| \\ &\leq \left\| \left((\hat{Q}\mathsf{P}')^{-1} - Q^{-1}\right) \begin{pmatrix} 1 \\ r_i \end{pmatrix} \right\| + \left\| \hat{Q}^{-1} \left[\begin{pmatrix} 1 \\ O_1\hat{r}_i \end{pmatrix} - \begin{pmatrix} 1 \\ r_i \end{pmatrix} \right] \right\|. \end{aligned} \tag{F.2}$$

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

$$\left\| \left((\hat{Q}\mathsf{P}')^{-1} - Q^{-1}\right) \begin{pmatrix} 1 \\ r_i \end{pmatrix} \right\| = \left\| \hat{Q}^{-1}(\hat{Q}\mathsf{P}' - Q)Q^{-1} \begin{pmatrix} 1 \\ r_i \end{pmatrix} \right\| = \|\hat{Q}^{-1}\| \, \|(\hat{Q}\mathsf{P}' - Q)w_i\|,$$

and

$$\|(\hat{Q}\mathsf{P}' - Q)w_i\| = \|(O_1'\widehat{V}'\mathsf{P}' - V')w_i\| \leq \|\mathsf{P}\widehat{V}O_1 - V\|_{2\to\infty}$$

If we can claim that $\|\hat{Q}^{-1}\| \leq C$, then we are done with the bound of the first term. Notice that one easily check

$$\|\hat{Q}\mathsf{P}' - Q\| \leq \sqrt{K}\|\mathsf{P}\widehat{V}O_1 - V\|_{2\to\infty} = o(\sqrt{K})$$

since $\frac{K^3\log(n)}{n\bar{\theta}^2\beta_n^2} \to 0$ 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

$$v_k(t) = \frac{b_t(k)}{b_1(k)}, \quad 2 \leq t \leq K, \tag{F.3}$$

which can be easily verified with some elementary derivations from the definition of R and the fact $\Xi = H_0^{-\frac{1}{2}}\Theta\Pi B$ with $B = (b_1, \cdots, b_K)$ (see the proof of Lemma A.1 in Section A.1). We will see that

$$Q = B'\mathrm{diag}(1/b_1(1), \cdots, 1/b_1(K));$$

And due to $b_1(k) \asymp 1$ (claimed in the Proof of Lemma D.2), we then obtain that

$$\|Q^{-1}\| = \|\mathrm{diag}(b_1(1), \cdots, b_1(K))B^{-1}\| \asymp \|B^{-1}\|.$$

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,

$$\left\| \Big((\hat{Q}\mathsf{P}')^{-1} - Q^{-1} \Big) \begin{pmatrix} 1 \\ r_i \end{pmatrix} \right\| \leq CK^{-\frac{1}{2}} \|\mathsf{P}\widehat{V}O_1 - V\|_{2\to\infty}.$$

Next, for the second term on the RHS of (F.2), one simply bounds it by $\|\hat{Q}^{-1}\|\|O_1\hat{r}_i - r_i\| \leq CK^{-\frac{1}{2}}\|O_1\hat{r}_i - r_i\|$. Combining these two estimates into (F.2), with the aids of (F.1) and Lemma 4.1, we conclude that

$$\|\mathsf{P}\hat{w}_i - w_i\| \leq CK^{-\frac{1}{2}}\|O_1\hat{r}_i - r_i\| \leq C\sqrt{\frac{K^2\log(n)}{n\bar{\theta}(\bar{\theta}\wedge\theta_i)\beta_n^2}}. \tag{F.4}$$

Next, we study the error between $1/e_k'\mathsf{P}\hat{b}_1$ 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,

$$\left| \frac{1}{\big(e_k'\mathsf{P}\hat{b}_1\big)^2} - \frac{1}{\big(b_1(k)\big)^2} \right| \leq \big|\hat{\lambda}_1 - \lambda_1\big| + \big|e_k'\mathsf{P}\widehat{V}\hat{\Lambda}_1\widehat{V}'\mathsf{P}'e_k - v_k'\Lambda_1 v_k\big|.$$

The eigenvalue difference is simply bounded by $\sqrt{\log(n)/n\bar{\theta}^2}$ 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 (F.3) and $b_1(k) \asymp 1$,

$$\|v_k\| \leq C\|e_k'B\| \leq C\|BB'\|^{\frac{1}{2}} \leq C\sqrt{K}.$$

We can then derive that

$$\begin{aligned}
&\big|e_k'\mathsf{P}\hat{V}\hat{\Lambda}_1\hat{V}'\mathsf{P}'e_k - v_k'\Lambda_1 v_k\big| \\
&\leq \big|e_k'(\mathsf{P}\widehat{V}O_1 - V)O_1'\hat{\Lambda}_1 O_1 O_1'\widehat{V}'\mathsf{P}'e_k\big| + \big|v_k O_1'\hat{\Lambda}_1 O_1(O_1'\widehat{V}'\mathsf{P}' - V)e_k\big| + \big|v_k'(O_1'\hat{\Lambda}_1 O_1 - \Lambda_1)v_k\big| \\
&\leq CK^{-\frac{1}{2}}|\lambda_2(PG)|\|\mathsf{P}\widehat{V}O_1 - V\|_{2\to\infty} + \big|v_k'(O_1'\hat{\Lambda}_1 O_1 - \Lambda_1)v_k\big|
\end{aligned}$$

Here in the last step, we used the trivial bound $\|\hat{\Lambda}_1\| \leq K^{-1}|\lambda_2(PG)|$. We further estimate the second term above. Notice that $O_1 = \mathrm{sgn}(\hat{\Xi}_1'\Xi_1)$ shown up in Theorem 4.1. By $L_0\Xi_1 = \Xi_1\Lambda_1$, $L\hat{\Xi}_1 = \hat{\Xi}_1\hat{\Lambda}_1$, and sine-theta theorem (D.13),

$$\big|v_k'(O_1'\hat{\Lambda}_1 O_1 - \Lambda_1)v_k\big| \leq \big|v_k'(O_1 - \hat{\Xi}_1'\Xi_1)'\hat{\Lambda}_1 O_1 v_k\big| + \big|v_k'\Xi_1'(L - L_0)\hat{\Xi}_1 O_1 v_k\big|$$

$$
\begin{aligned}
&+ \left| v_k' \Lambda_1 (O_1 - \hat{\Xi}_1' \Xi_1)' O_1 v_k \right| \\
&\le C |\lambda_2(PG)| \| O_1 - \hat{\Xi}_1' \Xi_1 \| + K \| L - L_0 \| \\
&\le C |\lambda_2(PG)| (K \beta_n^{-1} \| L - L_0 \|)^2 + K \| L - L_0 \| \\
&\le C \sqrt{\frac{K^3 \log(n)}{n \bar{\theta}^2 \beta_n^2}}
\end{aligned}
$$

where we also used $\|L - L_0\| \le C\sqrt{\log(n)/n\bar{\theta}^2}$ and $K^{\frac{3}{2}}\beta_n^{-1}\sqrt{\log(n)/n\bar{\theta}^2} = o(1)$. Next, using (F.1) and the last inequality in Condition 2.1(b), we bound

$$
K^{-\frac{1}{2}} |\lambda_2(PG)| \| \mathsf{P}\widehat{V} O_1 - V \|_{2\to\infty} \le C \sqrt{\frac{K^2 \log(n)}{n \bar{\theta}^2 \beta_n^2}} .
$$

It follows then

$$
\left| e_k' \mathsf{P} \hat{V} \hat{\Lambda}_1 \hat{V}' \mathsf{P}' e_k - v_k' \Lambda_1 v_k \right| \le C \sqrt{\frac{K^3 \log(n)}{n \bar{\theta}^2 \beta_n^2}} .
$$

As a consequence,

$$
\left| \frac{1}{(\mathsf{P}\hat{b}_1)(k)} - \frac{1}{b_1(k)} \right| \le C \sqrt{\frac{K^3 \log(n)}{n \bar{\theta}^2 \beta_n^2}} \tag{F.5}
$$

since $b_1(k) \asymp 1$.

Now, we are able to study $\mathsf{P}\hat{\pi}_i^*$ and further $\mathsf{P}\hat{\pi}_i$ by (F.4) and (F.5). If $(\mathsf{P}\hat{w}_i)(k) \le 0$, trivially we have

$$
\left| (\mathsf{P}\hat{\pi}_i^*)(k) - \pi_i^*(k) \right| = \pi_i^*(k) \asymp w_i(k) \le |(\mathsf{P}\hat{w}_i)(k) - w_i(k)|
$$

For the case that $(\mathsf{P}\hat{w}_i)(k) > 0$, we get the bound

$$
\left| (\mathsf{P}\hat{\pi}_i^*)(k) - \pi_i^*(k) \right| = \left| \frac{(\mathsf{P}\hat{w}_i)(k)}{(\mathsf{P}\hat{b}_1)(k)} - \frac{w_i(k)}{b_1(k)} \right| \le \left| (\mathsf{P}\hat{w}_i)(k) \right| \left| \frac{1}{(\mathsf{P}\hat{b}_1)(k)} - \frac{1}{b_1(k)} \right| + \frac{|(\mathsf{P}\hat{w}_i)(k) - w_i(k)|}{|b_1(k)|}
$$

Moreover, taking sum over k for both sides above,

$$
\| \mathsf{P}\hat{\pi}_i^* - \pi_i^* \|_1 \le C \max_k \left| \frac{1}{(\mathsf{P}\hat{b}_1)(k)} - \frac{1}{b_1(k)} \right| + \frac{\| \mathsf{P}\hat{w}_i - w_i \|_1}{\min_k |b_1(k)|} \le C \sqrt{\frac{K^3 \log(n)}{n \bar{\theta}(\bar{\theta} \wedge \theta_i) \beta_n^2}}
$$

Here we used the Cauchy-Schwarz inequality $\|\mathsf{P}\hat{w}_i - w_i\|_1 \le \sqrt{K}\,\|\mathsf{P}\hat{w}_i - w_i\|$ and further applied (F.4) and (F.5). As a result,

$$\big|(\mathsf{P}\hat{\pi}_i)(k) - \pi_i(k)\big| = \left|\frac{(\mathsf{P}\hat{\pi}_i^*)(k)}{\|\mathsf{P}\hat{\pi}_i^*\|_1} - \frac{\pi_i^*(k)}{\|\pi_i^*\|_1}\right| \le \big|(\mathsf{P}\hat{\pi}_i^*)(k)\big| \frac{\|\mathsf{P}\hat{\pi}_i^* - \pi_i^*\|_1}{\|\mathsf{P}\hat{\pi}_i^*\|_1\|\pi_i^*\|_1} + \frac{|(\mathsf{P}\hat{\pi}_i^*)(k) - \pi_i^*(k)|}{\|\pi_i^*\|_1}.$$

And summing up over k for both sides, we can further have

$$\|\mathsf{P}\hat{\pi}_i - \pi_i\|_1 \le \frac{\|\mathsf{P}\hat{\pi}_i^* - \pi_i^*\|_1}{\|\pi_i^*\|_1} \le C\sqrt{\frac{K^3\log(n)}{n\bar{\theta}(\bar{\theta}\wedge\theta_i)\beta_n^2}}$$

since $\|\pi_i^*\|_1 = \sum_k w_i(k)/b_1(k) \asymp 1$ by $b_1(k) \asymp 1$ for all $1 \le k \le K$. Therefore, we finished the proof.

F.3 Proofs of Corollary 4.1 and the second claim of Theorem 4.2

The proofs oare straightforward by employing Theorem 4.2. We shortly claim it below.

Proof of the second claim of Theorem 4.2. Recall the definition of the ℓ^1-loss $\mathcal{L}(\hat{\Pi}, \Pi)$ in (5). Employing the node-wise errors in Theorem 4.2 and taking average, we see that

$$\mathcal{L}(\hat{\Pi}, \Pi) \le C\sqrt{\log(n)}\int \min\Big\{\frac{err_n}{\sqrt{t\wedge 1}}, 1\Big\}dF_n(t),$$

with probability $1 - o(n^{-3})$. Further by the trivial bound $\mathcal{L}(\hat{\Pi}, \Pi) \le 2$, the high probability error rate implies the expected ℓ^1-loss rate, i.e., the $err_n(\theta)$ in (8). This finishes the proof. □

Proof of Corollary 4.1. Recall the loss metric $\mathcal{L}(\hat{\Pi}, \Pi; p, q)$ in (22). We crudely bound

$$\|T\hat{\pi}_i - \pi_i\|_q^q \le C_q\|\hat{\pi}_i - \pi_i\|_1^q$$

where C_q is some constant depending on q. Combining it with the node-wise error rate in Theorem 4.2 gives Corollary 4.1.

For the special case $p = 1/2$ and $q = 1$, we further bound

$$\mathcal{L}^w(\hat{\Pi}, \Pi) = \min_T\Big\{\frac{1}{n}\sum_{i=1}^n(\theta_i/\bar{\theta})^{1/2}\|T\hat{\pi}_i - \pi_i\|_1\Big\}$$

$$\leq \min_T \Big\{ \frac{1}{n} \sum_{i \in S_1} (\theta_i/\bar{\theta})^{1/2} \|T\hat{\pi}_i - \pi_i\|_1 \Big\} + \min_T \Big\{ \frac{1}{n} \sum_{i \in S_2} (\theta_i/\bar{\theta})^{1/2} \|T\hat{\pi}_i - \pi_i\|_1 \Big\} \quad \text{(F.6)}$$

where we recall the definition of S_1, S_2 in (D.1). For the first term, we use Cauchy-Schwarz inequality and get

$$\begin{aligned} \frac{1}{n} \sum_{i \in S_1} (\theta_i/\bar{\theta})^{1/2} \|T\hat{\pi}_i - \pi_i\|_1 &\leq \Big(\frac{1}{n} \sum_{i \in S_1} \theta_i/\bar{\theta} \Big)^{\frac{1}{2}} \Big(\frac{1}{n} \sum_{i \in S_1} \|T\hat{\pi}_i - \pi_i\|_1^2 \Big)^{\frac{1}{2}} \\ &\leq C \sqrt{\log(n)}\, err_n \end{aligned}$$

with probability $1 - o(n^{-3})$. Plugging in the above inequality into (F.6), and applying the error rate in Theorem 4.2 separately for $i \in S_2$, especially noticing that $\theta_i/\bar{\theta} \leq c\, err_n^2 \log(n)$ for $i \notin \hat{S}_n(c)$, one can easily obtain

$$\mathcal{L}^w(\hat{\Pi}, \Pi) \leq C \sqrt{\log(n)}\, err_n$$

with probability $1 - o(n^{-3})$. Further with trivial bound $\mathcal{L}^w(\hat{\Pi}, \Pi) \leq C$, we then conclude the proof. □

G Least-favorable configurations and proof of the lower bound

The key of proving the lower bound arguments in Theorem 2.1 and Theorem 4.3 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 ℓ^1-loss $\mathcal{L}^w(\hat{\Pi}, \Pi)$, and then modify it to construct the LFC for the standard ℓ^1-loss $\mathcal{L}(\hat{\Pi}, \Pi)$. The following notation is useful.

Definition G.1. *Given (n, K, β_n), $\theta \in \mathbb{R}^n$ and $P \in \mathbb{R}^{K \times K}$, let $\mathcal{Q}_n(K, \theta, P, \beta_n)$ denote the collection of eligible membership matrices Π such that Condition 2.1 is satisfied.*

First, we construct the LFC for proving the lower bound in Theorem 4.3. We take a special form of P,

$$P^* = \beta_n I_K + (1-\beta_n)\mathbf{1}_K\mathbf{1}_K', \quad \text{where } 0 < \beta_n < c < 1, \tag{G.1}$$

and construct a collection of Π. We need a well-known result.

Lemma G.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 $\omega^{(0)} = \mathbf{0}_s$ and $\|\omega^{(j)} - \omega^{(k)}\|_1 \geq s/8$, for all $0 \leq j < k \leq J$.*

In Theorem 4.3, we assume $F_n(err_n) \leq \check{c}$, for a constant $\check{c} \in (0,1)$. Let $c = \frac{1+\check{c}}{2} \in (0,1)$. Let $n_1 = \lfloor K^{-1}cn \rfloor$and $n_0 = n - Kn_1$. We set

$$\Pi^* = \Big(\underbrace{\frac{1}{K}\mathbf{1}_K, \cdots, \frac{1}{K}\mathbf{1}_K}_{n_0}, \underbrace{\mathbf{e}_1, \cdots, \mathbf{e}_1}_{n_1}, \cdots, \underbrace{\mathbf{e}_K, \cdots, \mathbf{e}_K}_{n_1} \Big)'. \tag{G.2}$$

Without loss of generality, we can assume that those θ_i's corresponding to the pure nodes in Π^* contains the top $\lfloor (c-\check{c})n \rfloor$ 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 θ_i's satisfy $\theta_i/\bar{\theta} \geq err_n^2$, by the assumption of $F_n(err_n^2) \leq \check{c}$. Note that we can always find a permutation to achieve such θ and re-construct Π^* correspondingly. Let $m = \lfloor n_0/2 \rfloor$ and $r = \lfloor K/2 \rfloor$. We apply Lemma G.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 row-wisely, 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:

$$\Gamma^{(j)} = \begin{bmatrix} H^{(j)} & -H^{(j)} & \mathbf{0}_{m\times 1} \\ -H^{(j)} & H^{(j)} & \mathbf{0}_{m\times 1} \\ \mathbf{0}_{1\times r} & \mathbf{0}_{1\times r} & 0 \\ \hdashline \mathbf{0}_{n_1\times r} & \mathbf{0}_{n_1\times r} & \mathbf{0}_{n_1\times 1} \end{bmatrix}, \qquad 0 \leq j \leq J. \tag{G.3}$$

In (G.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 $\Theta = \mathrm{diag}(\theta_1, \theta_2, \ldots, \theta_n)$. We construct $\Pi^{(0)}, \Pi^{(1)}, \ldots, \Pi^{(J)}$ by

$$\Pi^{(j)} = \Pi^* + \gamma_n \Theta^{-\frac{1}{2}} \Gamma^{(j)}, \qquad \text{where} \quad \gamma_n = c_0 K^{\frac{1}{2}} (n\bar{\theta}\beta_n^2)^{-\frac{1}{2}}, \quad \text{for } 0 \le j \le J, \tag{G.4}$$

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

Theorem G.1. *Fix c_1-c_4 in Condition 2.1 and $\check{c}$ in Theorem 4.3. Given any $(n, K, \alpha_n, \beta_n)$ and $\theta \in \mathbb{R}^n$ such that $F_n(err_n^2) \le \check{c}$, let P^* be as in (G.1), and construct $\Pi^{(0)}, \Pi^{(1)}, \ldots, \Pi^{(J)}$ as in (G.2)-(G.4). When c_0 in (G.4) is properly small, the following statements are true.*

- *For any constant $c_5 > 0$, let $\mathcal{Q}_n(\theta, P^*) = \mathcal{Q}_n(K, \theta, P^*, c_5\beta_n)$ (see Definition G.1). There exists a properly small c_5 such that $\Pi^{(j)}$ is contained in $\mathcal{Q}_n(\theta, P^*)$, for $0 \le j \le J$.*
- *There exists a constant $C_1 > 0$ such that $\mathcal{L}^w(\Pi^{(j)}, \Pi^{(k)}) \ge C_1 err_n$, for all $0 \le j < k \le J$.*
- *Let $\mathcal{P}_j$ be the probability measure of a DCMM model with $(\theta, P^*, \Pi^{(j)})$ and let $\mathrm{KL}(\cdot, \cdot)$ denote the Kullback-Leibler divergence. There exists a constant $\epsilon_1 \in (0, 1/8)$ such that $\sum_{1 \le j \le J} \mathrm{KL}(\mathcal{P}_j, \mathcal{P}_0) \le (1/8 - \epsilon_1) J \log(J)$.*

Furthermore, $\inf_{\hat{\Pi}} \sup_{\Pi \in \mathcal{Q}_n(\theta, P^*)} \mathbb{E}\mathcal{L}^w(\hat{\Pi}, \Pi) \ge C err_n$.

Theorem 4.3 follows immediately from Theorem G.1.

Next, we construct the LFC for proving the lower bound in Theorem 2.1. We still take P^* in (G.1) and construct a collection of Π. 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:

$$\int_0^{err_n^2} dF_n(t) \le C \int_{err_n^2}^{\infty} \frac{err_n}{\sqrt{t_n \wedge 1}} dF_n(t). \tag{G.5}$$

To understand (G.5), note that $\theta_i/\bar{\theta} \leq err_n^2$ is equivalent to $n\bar{\theta}\theta_i\beta_n^2 \leq K^3\log(n)$. For such a node i, the best estimator is the naive estimator $\hat{\pi}_i^{naive} = \frac{1}{K}\mathbf{1}_K$. In (G.5), the left hand side is the total contribution of these nodes in $\mathcal{L}(\hat{\Pi}, \Pi)$, and the right hand side is the contribution of remaining nodes. Therefore, (G.5) guarantees that the rate of convergence of the unweighted ℓ^1-loss is driven by those nodes for which we can indeed construct non-trivial estimators of π_i from data. When (G.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 next section and below we will focus on the case that (G.5) holds. Note that all examples in Section A.2 satisfy (G.5).

We need a technical lemma about the property of $F_n(\cdot)$ that satisfies the requirement in Theorem 2.1. It is proved in the next section.

Lemma G.2. *Fix $\rho > 0$ and $a_0 \in (0,1)$. Given any $\theta \in \mathcal{G}_n(\varrho, a_0)$ (see Definition 2.2), recall that $F_n(\cdot)$ is the empirical distribution associated with $\eta_i = \theta_i/\bar{\theta}$, $1 \leq i \leq n$. Let $\tilde{F}_n$ be the empirical distribution associated with $\tilde{\eta}_i = \eta_i \wedge 1$. For any $c > 0$ and $\epsilon \in (0,1)$, define*

$$\tau_n(c, \epsilon) = \inf\Big\{t > 0 : \int_{err_n^2}^{t} d\tilde{F}_n(x) \geq (1-\epsilon)\int_{err_n^2}^{c} d\tilde{F}_n(x)\Big\}.$$

If $F_n(\cdot)$ satisfies (G.5), then there exists a number $c_n > err_n^2$ and a constant $\tilde{a}_0 \in (0,1)$ such that $F_n(c_n) \leq 1 - \tilde{a}_0$ and

$$\int_{\tau_n(c_n,1/8)}^{c_n} \frac{1}{\sqrt{t \wedge 1}} dF_n(t) + \frac{\lceil n \cdot \varpi_n \rceil}{n\sqrt{\tau_n(c_n, 1/8) \wedge 1}} \;\geq\; \tilde{a}_0 \int_{err_n^2}^{\infty} \frac{1}{\sqrt{t \wedge 1}} dF_n(t), \tag{G.6}$$

where $\varpi_n = [F_n(c_n) - F_n(err_n^2-)]/8 - [F_n(c_n) - F_n(\tau_n(c_n, 1/8))]$ and for any distribution function $F(\cdot)$, we define $F(x-) = \lim_{\omega\to 0} F(x-\omega)$.

We now construct a collection of Π using c_n in Lemma G.2. We re-order θ_i's such that

$$\theta_{(1)} \leq \theta_{(2)} \leq \ldots \leq \theta_{(n)}. \tag{G.7}$$

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 G.2. Define

$$s_n = \max\{1 \leq i \leq n : \tilde{\eta}_{(i)} \leq err_n^2\}, \qquad n_0 = \max\{1 \leq i \leq n : \tilde{\eta}_{(i)} \leq c_n\} - s_n.$$

It follows from the definition of $\tilde{F}_n$ that n_0 is approximately the total number of $\tilde{\eta}_i$'s such that $err_n^2 < \tilde{\eta}_i \leq c_n$. It can be derived from the definition of $\tau_n(c_n, 1/8)$ and ϖ_n that $n[F_n(c_n) - F_n(\tau_n(c_n, 1/8))] + \lceil n\varpi_n \rceil = n_0 - \lfloor 7n_0/8 \rfloor$. Combining these claims with (G.6) gives

$$\frac{1}{n}\sum_{\lfloor 7n_0/8\rfloor < i - s_n \leq n_0} \frac{1}{\sqrt{\eta_{(i)} \wedge 1}} \gtrsim \int_{err_n^2}^{\infty} \frac{1}{\sqrt{t\wedge 1}} dF_n(t).$$

We multiply err_n on both hand sides. By the condition (G.5), we have $err_n \int_{err_n^2}^{\infty} \frac{1}{\sqrt{t\wedge 1}} dF_n(t) \geq C^{-1} \int \min\{\frac{err_n}{\sqrt{t\wedge 1}}, 1\} dF_n(t)$, which yields a lower bound for the right hand side. For the left hand side, we plug in $\eta_i = \theta_i/\bar{\theta}$. It follows that

$$\sqrt{\frac{K^3}{n\bar{\theta}\beta_n^2}} \cdot \frac{1}{n} \sum_{\lfloor 7n_0/8\rfloor < i - s_n \leq n_0} \frac{1}{\sqrt{\theta_{(i)} \wedge \bar{\theta}}} \gtrsim \int \min\Big\{\frac{err_n}{\sqrt{t\wedge 1}}, 1\Big\} dF_n(t).$$

Notice that for each individual i such that $\lfloor 7n_0/8\rfloor < i - s_n \leq n_0$, its contribution to the left hand side sum above is negligible since $n^{-1}/\sqrt{\theta_{(i)}} \leq n^{-1}/\sqrt{\bar{\theta} err_n^2} = n^{-1/2} K^{-3/2} \bar{\theta}^{1/2} \beta_n = o(1)$. Therefore, we can remove finitely many i from the left hand side sum without changing the inequality above. This further implies

$$\sqrt{\frac{K^3}{n\bar{\theta}\beta_n^2}} \cdot \frac{1}{n} \sum_{\lfloor 7n_0/8\rfloor + 2 < i - s_n \leq n_0} \frac{1}{\sqrt{\theta_{(i)} \wedge \bar{\theta}}} \gtrsim \int \min\Big\{\frac{err_n}{\sqrt{t\wedge 1}}, 1\Big\} dF_n(t).$$

Let $\mathcal{M}_0$ be the index set of the nodes ordered between s_n and $s_n + n_0$ in (G.7). Let $\gamma_n = c_0 K^{\frac{1}{2}} (n\bar{\theta}\beta_n^2)^{-\frac{1}{2}}$. The above implies that

$$\gamma_n \inf_{\substack{\mathcal{M}\subset\mathcal{M}_0,\\ |\mathcal{M}|\geq n_0/8-2}} \Big\{\frac{1}{n}\sum_{i\in\mathcal{M}} \frac{1}{\sqrt{\theta_i \wedge \bar{\theta}}}\Big\} \gtrsim K^{-1} \int \min\Big\{\frac{err_n}{\sqrt{t\wedge 1}}, 1\Big\} dF_n(t). \tag{G.8}$$

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 Π^* are of the same order (such an ordering always exists). After node re-arrangement, we construct Π^* and $\Gamma^{(0)}, \Gamma^{(1)}, \ldots, \Gamma^{(J)}$ in the same way as in (G.2)-(G.3). Let $\tilde{\theta}_i = \theta_i \wedge \bar{\theta}$ and $\widetilde{\Theta} = \mathrm{diag}(\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_n)$. Let

$$\Pi^{(j)} = \Pi^* + \gamma_n \widetilde{\Theta}^{-\frac{1}{2}} \Gamma^{(j)}, \qquad \text{for} \quad 0 \leq j \leq J, \tag{G.9}$$

where $\gamma_n = c_0 K^{\frac{1}{2}}(n\bar{\theta}\beta_n^2)^{-\frac{1}{2}}$ is the same as in (G.8). The following theorem is an analog of Theorem G.1 for the unweighted ℓ^1-loss and is proved in next section.

Theorem G.2. *Fix c_1-c_4 in Condition 2.1 and (ϱ, a_0) in Theorem 2.1. Given (n, K, β_n) and $\theta \in \mathcal{G}_n(\varrho, a_0)$, let P^* be as in (G.1), and construct $\Pi^{(0)}, \Pi^{(1)}, \ldots, \Pi^{(J)}$ as in (G.9). When c_0 in (G.4) is properly small, the following statements are true.*

- *For any constant $c_5 > 0$, let $\mathcal{Q}_n(\theta, P^*) = \mathcal{Q}_n(K, \theta, P^*, c_5\beta_n)$. There exists a properly small c_5 such that $\Pi^{(j)}$ is contained in $\mathcal{Q}_n(\theta, P^*)$, for $0 \leq j \leq J$.*
- *There exists a constant $C_2 > 0$ such that $\mathcal{L}(\Pi^{(j)}, \Pi^{(k)}) \geq C_2 \int \min\{\frac{err_n}{\sqrt{t \wedge 1}}, 1\} dF_n(t)$, 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 $\mathrm{KL}(\cdot,\cdot)$ denote the Kullback-Leibler divergence. There exists a constant $\epsilon_2 \in (0, 1/8)$ such that $\sum_{1 \leq j \leq J} \mathrm{KL}(\mathcal{P}_j, \mathcal{P}_0) \leq (1/8 - \epsilon_2) J \log(J)$.*

Furthermore, $\inf_{\hat{\Pi}} \sup_{\Pi \in \mathcal{Q}_n(\theta, P^*)} \mathbb{E}\mathcal{L}(\hat{\Pi}, \Pi) \geq C \int \min\{\frac{err_n}{\sqrt{t \wedge 1}}, 1\} dF_n(t)$.

Theorem 2.1 follows immediately from Theorem G.2.

Remark: In Theorems G.1-G.2, we fix $P = P^*$ and prove the lower bounds by taking supreme over a class of Π. Such lower bounds are not only θ-specific but also P-specific, and they are stronger than the θ-specific lower bounds in Theorems 2.1 and 4.3. In Section H.4 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 $P\mathbf{1}_K \propto \mathbf{1}_K$; (c) (K, P) can depend on n, and K can be unbounded, but $P\mathbf{1}_K \propto \mathbf{1}_K$ and $|\lambda_2(P)| \leq C\beta_n = o(1)$.

H Proofs in lower bound analysis

In this section, we complete the proofs of lower bounds, i.e., Theorems 2.1 and 4.3. To this end, we will show the proofs of Theorems G.1-G.2 and Lemma G.2 stated in Section G.

We organize this section as follows: In Section H.1, we provide the proof of Theorem G.1 regarding weighted loss metric $\mathcal{L}^w(\hat{\Pi}, \Pi)$. In Section H.2, we claim Lemma G.2 and prove Theorem G.2 under the condition (G.5). The proof of Theorem G.2 with (G.5) violated is relatively simpler and we state it in Section H.3 for completeness. In Section H.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 G.

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

H.1 Proof of Theorem G.1

We begin with the proof of the first claim. We first verify $\Pi^{(j)} \in \mathcal{Q}_n(\theta, P^*)$, for every $0 \leq j \leq J$ which are constructed in (G.2)-(G.4). By the definition of perturbation matrix $\Gamma^{(j)}$'s in (G.3), and the fact that $\gamma_n/\sqrt{\theta_i} \leq c_0/K$ for all $1 \leq i \leq n_0$ due to $\theta_i/\bar{\theta} \geq err_n^2$ 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 Condition 2.1. Note that Condition 2.1(d) and the last inequality in Condition 2.1(a) immediately hold because of the construction of Π^*. By definition, $G^{(j)} = K(\Pi^{(j)})'\Theta H_0^{-1}\Theta\Pi^{(j)}$ and

$$\|G^{(j)} - G^*\| \leq 2\|G^*\|^{\frac{1}{2}}\|K\gamma_n^2(\Gamma^{(j)})'\Theta^{\frac{1}{2}}H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)}\|^{\frac{1}{2}} + \|K\gamma_n^2(\Gamma^{(j)})'\Theta^{\frac{1}{2}}H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)}\|. \tag{H.1}$$

Elementary computations lead to

$$G^* = \Big(\sum_{i=1}^{n_0}\frac{\theta_i^2}{H_0(i,i)}\Big)\frac{1}{K}\mathbf{1}_K\mathbf{1}_K' + K\mathrm{diag}\Big(\sum_{i\in\mathcal{C}_{p,1}}\frac{\theta_i^2}{H_0(i,i)}, \cdots, \sum_{i\in\mathcal{C}_{p,K}}\frac{\theta_i^2}{H_0(i,i)}\Big).$$

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

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

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

$$\|K\gamma_n^2(\Gamma^{(j)})'\Theta^{\frac{1}{2}}H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)}\| \le c_0^2 err_n^2 = o(1) \tag{H.2}$$

following from $\sum_{i=1}^n \theta_i/H_0(i,i) \le 1/\bar{\theta}$ and $|e_i\Gamma^{(j)}x| \le \sqrt{K}$ for all $1 \le i \le n$ and any unit vector $x \in \mathbb{R}^K$. Therefore, by Weyl's inequality and (H.1), we can conclude that the first two inequalities in Condition 2.1(a) hold for all $G^{(j)}$'s. Further with our choice of special P^* which satisfies that $\mathbf{1}_K'P = \big(K-(K-1)\beta_n\big)\mathbf{1}_K' > (1-c)K\mathbf{1}_K$, $\lambda_1(P^*) \asymp K$, and $\lambda_k(P^*) = \beta_n$ for all $2 \le k \le K$, the requirements that $|\lambda_K(PG)| \ge c_5\beta_n$ for some $c_5 > 0$ hold for P^*G^* and $P^*G^{(j)}$'s. The eigengap condition, Condition 2.1(b), holds for P^*G^*. Condition 2.1(b) also holds for $P^*G^{(j)}$'s, as a result of the Weyl's inequality, where

$$|\lambda_1(P^*G^{(j)})\lambda_2(P^*G^{(j)})| \le \sigma_1(P^*G^{(j)})\sigma_2(P^*G^{(j)}) \le CK\beta_n$$

and $\lambda_1(P^*G^{(j)}) \asymp K$. Here we use $\sigma_1(P^*G^{(j)}), \sigma_2(P^*G^{(j)})$ to denote the first and second largest singular values of $P^*G^{(j)}$; and the right hand side upper bound is due to the expression $P^*G^{(j)} = \beta_n G^{(j)} + K(1-\beta_n)\frac{1}{K}\mathbf{1}_K\mathbf{1}_K'G^{(j)}$ with the fact that $\|G^{(j)}\| \le c$ and $\|(G^{(j)})^{-1}\| \le c$ for some constant c.

Lastly, we claim that Condition 2.1(c) holds for all $G^{(j)}$'s. Using *Perron's theorem*, we obtain that the first right eigenvector of $P^*G^{(j)}$ is positive for all $1 \le j \le J$. In particular, for P^*G^*, all of its entries are positive and $\asymp 1$. We then claim that $P^*G^{(j)}(i,k) \asymp 1$ for all $1 \le i,k \le K$. To see this, for each $1 \le i,k \le K$, we first write

$$\begin{aligned} P^*G^{(j)}(i,k) - P^*G^*(i,k) &= K\gamma_n e_i'P^*(\Gamma^{(j)})'\Theta^{\frac{1}{2}}H_0^{-1}\Theta\Pi^*e_k + K\gamma_n e_i'P^*(\Pi^*)'\Theta H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)}e_k \\ &\quad + K\gamma_n^2 e_i'P^*(\Gamma^{(j)})'\Theta^{\frac{1}{2}}H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)}e_k. \end{aligned}$$

Note that $P^*(\Gamma^{(j)})' = \beta_n(\Gamma^{(j)})'$ by the definition of $\Gamma^{(j)}$ such that $\mathbf{1}_K'(\Gamma^{(j)})' = 0$. We thus easily bound the first and third terms on the RHS above by

$$|K\gamma_n e_i'P^*(\Gamma^{(j)})'\Theta^{\frac{1}{2}}H_0^{-1}\Theta\Pi^*e_k| \le \beta_n\|G^*\|^{\frac{1}{2}}\|K\gamma_n^2(\Gamma^{(j)})'\Theta^{\frac{1}{2}}H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)}\|^{\frac{1}{2}} \le c\beta_n err_n$$

$$|K\gamma_n^2 e_i' P^*(\Gamma^{(j)})'\Theta^{\frac{1}{2}} H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)} e_k| \le \beta_n \|K\gamma_n^2(\Gamma^{(j)})'\Theta^{\frac{1}{2}} H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)}\| \le c\beta_n err_n^2$$

which are both of order $o(1)$. For the second term, by the definition of Π^*, we have

$$\begin{aligned}
|K\gamma_n e_i' P^*(\Pi^*)'\Theta H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)} e_k| &\le |K\gamma_n\beta_n e_i'(\Pi^*)'\Theta H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)} e_k| + |K\gamma_n(1-\beta_n)\mathbf{1}_n'\Theta H_0^{-1}\Theta^{\frac{1}{2}}\Gamma^{(j)} e_k| \\
&\le c\beta_n err_n + cK\gamma_n \sum_{i=1}^n \frac{\sqrt{\theta_i}}{n\bar{\theta}} \\
&\le c\, err_n
\end{aligned}$$

where we used Cauchy-Schwarz inequality $\sum_{i=1}^n \sqrt{\theta_i} \le \sqrt{n}(\sum_{i=1}^n \theta_i)^{1/2} = n\sqrt{\bar{\theta}}$ in the last step. Therefore, it follows from the above discussions that

$$P^*G^{(j)}(i,k) = P^*G^*(i,k) + \big(P^*G^{(j)}(i,k) - P^*G^*(i,k)\big) = P^*G^*(i,k) + o(1) \asymp 1$$

for all $1 \le i, k \le K$. As a result, $\min_{i,k} P^*G^{(j)}(i,k)/\max_{i,k} P^*G^{(j)}(i,k) > c$ for some constant $c > 0$. Then, $\eta_1^{(j)}$, the first right eigenvector of $P^*G^{(j)}$, satisfies

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

This proves Condition 2.1(c). 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,

$$\begin{aligned}
\mathcal{L}^w(\Pi^{(j)}, \Pi^{(k)}) &= \frac{1}{n}\sum_{i=1}^n (\theta_i/\bar{\theta})^{\frac{1}{2}} \|\pi_i^{(j)} - \pi_i^{(k)}\|_1 = \frac{c_0}{n}\sum_{i=1}^{n_0} \frac{\sqrt{K}}{\beta_n\sqrt{n\bar{\theta}^2}} \|e_i\big(\Gamma^{(j)} - \Gamma^{(k)}\big)\|_1 \\
&= \frac{c_0\sqrt{K}}{\beta_n\sqrt{n\bar{\theta}^2}} \cdot \frac{4}{n}\|\omega^{(j)} - \omega^{(k)}\|_1 \ge C_1 err_n,
\end{aligned}$$

for some constant $C_1 > 0$.

In the last part, we prove the third claim in Theorem G.1 regarding the KL divergence statement. Note that

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

Notice that $\Omega_{ij}^{(\ell)} = \Omega_{ij}^{(0)}$ for $n_0 < i < j \le n$. Only the pairs satisfying $0 < i < j \le n_0$ and $0 < i \le n_0 < j \le n$ have the contributions. We then write $KL(\mathcal{P}_\ell, \mathcal{P}_0) = (I) + (II)$ where

$$(I) := \sum_{0<i<j\le n_0} \Omega_{ij}^{(\ell)} \log(\Omega_{ij}^{(\ell)}/\Omega_{ij}^{(0)}) + \big(1-\Omega_{ij}^{(\ell)}\big) \log \frac{1-\Omega_{ij}^{(\ell)}}{1-\Omega_{ij}^{(0)}} \tag{H.4}$$

$$(II) := \sum_{0<i\le n_0<j\le n} \Omega_{ij}^{(\ell)} \log(\Omega_{ij}^{(\ell)}/\Omega_{ij}^{(0)}) + \big(1-\Omega_{ij}^{(\ell)}\big) \log \frac{1-\Omega_{ij}^{(\ell)}}{1-\Omega_{ij}^{(0)}} \tag{H.5}$$

We begin with the estimate of (I). For simplicity, we write $\Gamma^{(\ell)} = \big(\Gamma_1^{(\ell)}, \cdots, \Gamma_n^{(\ell)}\big)'$ for $\ell = 0, \cdots, J$. By definition, for $0 < i < j \le n_0$,

$$\Omega_{ij}^{(0)} = \theta_i\theta_j\Big(\frac{1}{K^2}\mathbf{1}_K' P \mathbf{1}_K\Big) = \theta_i\theta_j\big(1-(1-1/K)\beta_n\big) < 1;$$

and

$$\begin{aligned}\Omega_{ij}^{(\ell)} &= \theta_i\theta_j\big(\frac{1}{K}\mathbf{1}_K + \frac{\gamma_n}{\sqrt{\theta_i}}\Gamma_i^{(\ell)}\big)' P \big(\frac{1}{K}\mathbf{1}_K + \frac{\gamma_n}{\sqrt{\theta_j}}\Gamma_j^{(\ell)}\big) \\ &= \theta_i\theta_j\big(1-(1-1/K)\beta_n\big) + \theta_i\theta_j \frac{\gamma_n^2\beta_n}{\sqrt{\theta_i\theta_j}}\big(\Gamma_i^{(\ell)}\big)'\Gamma_j^{(\ell)} \\ &= \Omega_{ij}^{(0)}\Big(1+\Delta_{ij}^{(\ell)}\Big), \qquad\qquad j \ne 0,\end{aligned}$$

in which, $\Delta_{ij}^{(\ell)} := (\gamma_n^2/\sqrt{\theta_i\theta_j}) \cdot \beta_n\big(\Gamma_i^{(\ell)}\big)'\Gamma_j^{(\ell)}/[1-(1-1/K)\beta_n]$. Here we used the identity $\mathbf{1}_K'\Gamma_j^{(\ell)} = 0$ for all $1 \le j \le n$, $1 \le \ell \le J$. Further by $1-(1-1/K)\beta_n > c$ for some constant $c > 0$and the assumption that the first n_0 θ_i's satisfy $\theta_i/\bar{\theta} \ge err_n^2$, we notice that

$$\max_{0<i,j\le n_0} |\Delta_{ij}^{(\ell)}| \le C\frac{c_0^2K^2}{\beta_n\sqrt{(n\bar{\theta}\theta_i)(n\bar{\theta}\theta_j)}} \le Cc_0^2K^{-1}\beta_n$$

and $\Omega_{ij}^{(0)}\Delta_{ij}^{(\ell)} < c(1-\Omega_{ij}^{(0)})$ for some constant $c \in (0,1)$ when c_0 is small enough. Choosing sufficiently small c_0, we have the Taylor expansions

$$\Omega_{ij}^{(\ell)} \log(\Omega_{ij}^{(\ell)}/\Omega_{ij}^{(0)}) = \Omega_{ij}^{(0)}\big(1+\Delta_{ij}^{(\ell)}\big)\log\big(1+\Delta_{ij}^{(\ell)}\big) = \Omega_{ij}^{(0)}\big(\Delta_{ij}^{(\ell)} + \frac{1}{2}(\Delta_{ij}^{(\ell)})^2 + O((\Delta_{ij}^{(\ell)})^3)\big)$$

and

$$\big(1-\Omega_{ij}^{(\ell)}\big)\log\frac{1-\Omega_{ij}^{(\ell)}}{1-\Omega_{ij}^{(0)}} = \big(1-\Omega_{ij}^{(0)} - \Omega_{ij}^{(0)}\Delta_{ij}^{(\ell)}\big)\log\Big(1 - \frac{\Omega_{ij}^{(0)}}{1-\Omega_{ij}^{(0)}}\Delta_{ij}^{(\ell)}\Big)$$

$$= -\Omega_{ij}^{(0)}\Delta_{ij}^{(\ell)} + \frac{(\Omega_{ij}^{(0)})^2}{2(1-\Omega_{ij}^{(0)})}(\Delta_{ij}^{(\ell)})^2 + O\Big(\frac{(\Omega_{ij}^{(0)}\Delta_{ij}^{(\ell)})^3}{(1-\Omega_{ij}^{(0)})^2}\Big)$$

Combining the above two equations together into (H.4), we arrive at

$$(I) \le \sum_{0<i<j\le n_0} \frac{\Omega_{ij}^{(0)}}{(1-\Omega_{ij}^{(0)})}(\Delta_{ij}^{(\ell)})^2 = \sum_{0<i<j\le n_0} \frac{\gamma_n^4\beta_n^2[\big(\Gamma_i^{(\ell)}\big)'\Gamma_j^{(\ell)}]^2}{[1-(1-1/K)\beta_n](1-\Omega_{ij}^{(0)})}$$
$$\le C\gamma_n^4\beta_n^2 n_0^2 K^2 \le Cc_0^4 n_0 K \cdot \frac{K^3}{\beta_n^2 n\bar{\theta}^2} \le Cc_0^4 n_0 K \tag{H.6}$$

Here we used $K^3/\beta_n^2(n\bar{\theta}^2) \le c$ for some sufficiently small $c > 0$ and the crude bound $|\big(\Gamma_i^{(\ell)}\big)'\Gamma_j^{(\ell)}| \le K$, which follows from the definition of $\Gamma^{(\ell)}$ in (G.3).

In the sequel, we turn to study (II). Since $0 < i \le n_0 < j \le n$, we suppose that $j \in \mathcal{C}_{p,\hat{j}}$ for some $1 \le \hat{j} \le K$. Then,

$$\Omega_{ij}^{(0)} = \theta_i\theta_j\Big(\frac{1}{K}\mathbf{1}_K' P\mathbf{e}_{\hat{j}}\Big) = \theta_i\theta_j\big(1-(1-1/K)\beta_n\big);$$

and

$$\Omega_{ij}^{(\ell)} = \theta_i\theta_j\big(\frac{1}{K}\mathbf{1}_K + \frac{\gamma_n}{\sqrt{\theta_i}}\Gamma_i^{(\ell)}\big)'P\mathbf{e}_{\hat{j}}$$
$$= \theta_i\theta_j\big(1-(1-1/K)\beta_n\big) + \theta_i\theta_j\frac{\gamma_n\beta_n}{\sqrt{\theta_i}}\big(\Gamma_i^{(\ell)}\big)'\mathbf{e}_{\hat{j}}$$
$$= \Omega_{ij}^{(0)}\Big(1+\widetilde{\Delta}_{ij}^{(\ell)}\Big)$$

with $\widetilde{\Delta}_{ij}^{(\ell)} := (\gamma_n/\sqrt{\theta_i})\cdot\beta_n\big(\Gamma_i^{(\ell)}\big)'\mathbf{e}_{\hat{j}}/\big(1-(1-1/K)\beta_n\big)$. Similarly, one can easily check that

$$\max_{i,j}|\widetilde{\Delta}_{ij}^{(\ell)}| \le \frac{Cc_0\sqrt{K}}{\sqrt{n\bar{\theta}\theta_i}} \le Cc_0K^{-1}\beta_n$$

by our assumption that the first n_0 θ_i's satisfy $\theta_i/\bar{\theta} \ge err_n^2$. Therefore, in the same way as (H.6), we can derive

$$(II) \le \sum_{0<i\le n_0<j\le n} \frac{\Omega_{ij}^{(0)}}{(1-\Omega_{ij}^{(0)})}(\widetilde{\Delta}_{ij}^{(\ell)})^2 = \sum_{\substack{0<i\le n_0,\\ j\in\mathcal{C}_{p,\hat{j}},1\le\hat{j}\le K}} \frac{\theta_j\gamma_n^2\beta_n^2[\big(\Gamma_i^{(\ell)}\big)'\mathbf{e}_{\hat{j}}]^2}{[1-(1-1/K)\beta_n](1-\Omega_{ij}^{(0)})}$$
$$\le C\Big(\sum_{j=n_0+1}^{n}\theta_j\Big)\gamma_n^2\beta_n^2 n_0 \le Cc_0^2 n_0 K. \tag{H.7}$$

We now combine (H.6) and (H.7). They imply that

$$\sum_{\ell=1}^{J} KL(\mathcal{P}_\ell, \mathcal{P}_0) \le Cc_0^2 JnK.$$

Here C is a constant independent of choice of c_0 and n. At the same time, since $J \ge 2^{\lfloor n_0/2 \rfloor \times \lfloor K/2 \rfloor / 8}$, we obtain that $\log J \ge cnK$ for some constant $c > 0$ not relying on the other parameters. By properly choosing c_0, we can always find a constant $\epsilon_1 \in (0, 1/8)$ such that $Cc_0^2 JnK \le Cc_0^2/c \cdot J\log(J) \le (1/8 - \epsilon_1)J\log(J)$. This finishes the proof of the last claim. Furthermore, with standard techniques of lower bound analysis (e.g., [9, Theorem 2.5]), we ultimately obtain the lower bound stated in Theorem G.1.

H.2 Proof of Lemma G.2 and Theorem G.2

Proof of Lemma G.2. Recall Definition 2.2. For such c_n, ϱ and a_0, we see that if $\tau_n(c_n, 1/8) \ge \varrho c_n$, then by definition of $\tau_n(c_n, 1/8)$ and ϖ_n,

$$\begin{aligned}\int_{\tau_n(c_n,1/8)}^{c_n} \frac{1}{\sqrt{t \wedge 1}} \mathrm{d}F_n(t) + \frac{\lceil n \cdot \varpi_n \rceil}{n\sqrt{\tau_n(c_n, 1/8) \wedge 1}} &\ge \frac{1}{8\sqrt{c_n \wedge 1}} \tilde{F}_n(c_n) \\ &\ge \frac{\sqrt{\varrho}}{8\sqrt{\varrho c_n \wedge 1}} \tilde{F}_n(c_n) \ge \frac{\sqrt{\varrho}}{8} \int_{\varrho c_n}^{c_n} \frac{1}{\sqrt{t \wedge 1}} \mathrm{d}F_n(t) .\end{aligned}$$

By Definition 2.2, we conclude

$$\int_{\tau_n(c_n,1/8)}^{c_n} \frac{1}{\sqrt{t \wedge 1}} \mathrm{d}F_n(t) \ge \tilde{a}_0 \int_{err_n^2}^{\infty} \frac{1}{\sqrt{t \wedge 1}} \mathrm{d}F_n(t), \quad \tilde{a}_0 := \frac{\sqrt{\varrho}}{8} a_0.$$

In the case that $\tau_n(c_n, 1/8) < \rho c_n$, trivially, (G.6) holds with $\tilde{a}_0 = a_0$. □

With the help of Lemma G.2, we are able to prove Theorem G.2 below.

Proof of Theorem G.2. Since the least-favorable configurations Π^* and $\Pi^{(j)}$'s are quite similar to those for the weighted loss metric, only with slightly different perturbation scales. Such differences 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 G.1. We thus conclude the first statement without details.

Next, for the pairwise difference under unweight loss metric, by definition,

$$\mathcal{L}(\Pi^{(j)},\Pi^{(k)}) = \frac{1}{n}\sum_{i=1}^{n}\|\pi_i^{(j)}-\pi_i^{(k)}\|_1 = \frac{1}{n}\sum_{i=1}^{n_0}\frac{\gamma_n}{\sqrt{\theta_i\wedge\bar{\bar{\theta}}}}\|\mathbf{e}_i'\big(\Gamma^{(j)}-\Gamma^{(k)}\big)\|_1$$

Note that $\|H^{(j)}-H^{(k)}\|_1 \geq \lfloor n_0/2\rfloor \times \lfloor K/2\rfloor/8$. At least $\lfloor n_0/2\rfloor/8$ 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

$$\begin{aligned}
\frac{\gamma_n}{n}\sum_{i=1}^{n_0}\frac{1}{\sqrt{\theta_i\wedge\bar{\bar{\theta}}}}\|\mathbf{e}_i'\big(\Gamma^{(j)}-\Gamma^{(k)}\big)\|_1 &\geq \frac{\gamma_n}{n}\min_{\substack{\mathcal{M}\subset\mathcal{M}_0,\\ |\mathcal{M}|=2\lfloor\lfloor n_0/2\rfloor/8\rfloor}}\sum_{i\in\mathcal{M}}\frac{K-1}{\sqrt{\theta_i\wedge\bar{\bar{\theta}}}}\\
&\geq \frac{\gamma_n}{n}\min_{\substack{\mathcal{M}\subset\mathcal{M}_0,\\ |\mathcal{M}|=\lfloor n_0/8\rfloor-1}}\sum_{i\in\mathcal{M}}\frac{K-1}{\sqrt{\theta_i\wedge\bar{\bar{\theta}}}}\\
&\gtrsim \int\min\Big\{\frac{err_n}{\sqrt{t\wedge 1}},1\Big\}\mathrm{d}F_n(t)
\end{aligned}$$

where the last step is due to (G.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 G.1. We again define (I) and (II) as (H.4)-(H.5), and bound them separately. Thanks to the slight difference on the perturbation scale, one can simply mimic the proof of Theorem G.1 and obtain the following bounds under current settings.

$$\begin{aligned}
(I) &\leq \sum_{1\leq i<j\leq n_0}\frac{\Omega_{ij}^{(0)}}{(1-\Omega_{ij}^{(0)})}(\Delta_{ij}^{(\ell)})^2 = \sum_{1\leq i<j\leq n_0}\frac{\frac{\theta_i}{\theta_i\wedge\bar{\bar{\theta}}}\frac{\theta_j}{\theta_j\wedge\bar{\bar{\theta}}}\gamma_n^4\beta_n^2[\big(\Gamma_i^{(\ell)}\big)'\Gamma_j^{(\ell)}]^2}{[1-(1-1/K)\beta_n](1-\Omega_{ij}^{(0)})}\\
&\leq C\gamma_n^4\beta_n^2 n_0^2K^2 \leq Cc_0^4 n_0K\cdot\frac{K^3}{\beta_n^2 n\bar{\theta}^2}\leq Cc_0^4n_0K
\end{aligned}$$

and

$$\begin{aligned}
(II) &\leq \sum_{0<i\leq n_0<j\leq n}\frac{\Omega_{ij}^{(0)}}{(1-\Omega_{ij}^{(0)})}(\widetilde{\Delta}_{ij}^{(\ell)})^2 = \sum_{\substack{0<i\leq n_0,\\ j\in\mathcal{C}_{p,\hat{j}},1\leq\hat{j}\leq K}}\frac{\frac{\theta_i}{\theta_i\wedge\bar{\bar{\theta}}}\theta_j\gamma_n^2\beta_n^2[\big(\Gamma_i^{(\ell)}\big)'\mathbf{e}_{\hat{j}}]^2}{[1-(1-1/K)\beta_n](1-\Omega_{ij}^{(0)})}\\
&\leq C\Big(\sum_{j=n_0+1}^{n}\theta_j\Big)\gamma_n^2\beta_n^2 n_0\leq Cc_0^2n_0K.
\end{aligned}$$

Here to obtain the two upper bounds above, we used an estimate

$$\sum_{1\leq i\leq n_0}\frac{\theta_i}{\theta_i\wedge\bar{\bar{\theta}}}\leq\sum_{1\leq i\leq n_0}\Big(1+\frac{\theta_i}{\bar{\bar{\theta}}}\Big)\leq 2cn_0$$

for some constant $c > 0$, where the last step is due to our ordering of θ_i's and $\theta_i/\bar{\theta} \leq c_n \leq C$ for some constant $C > 0$, $1 \leq i \leq n_0$, which follows from Definition 2.2 and self-normalization of $F_n(\cdot)$ (i.e., $\int t dF_n(t) = 1$). As a result,

$$\frac{1}{J}\sum_{\ell=1}^{J} KL(\mathcal{P}_\ell, \mathcal{P}_0) \leq Cc_0^2 n_0 K \leq \tilde{C}c_0^2 \log J.$$

Properly choosing sufficiently small c_0, we thus complete the third statement. Furthermore, by standard techniques of lower bound analysis (e.g., [9, Theorem 2.5]), we ultimately obtain the lower bound stated in Theorem G.2.

□

H.3 Proof of Theorem G.2 without (G.5)

In this section, we show the proof of Theorem G.2 in the case that (G.5) violates. We will need a distinct sequence of least-favorable configurations. We still order θ_i's as (G.7). But we define

$$n_0 = \max\{1 \leq i \leq n : \theta_i/\bar{\theta} \leq err_n^2\}$$

which means n_0 is the total number of η_i's such that $0 < \eta_i \leq err_n^2$. 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 Π^* are of the same order as before. Π^* and $\Gamma^{(0)}, \Gamma^{(1)}, \ldots, \Gamma^{(J)}$are constructed in the same way as in (G.2)-(G.3). Different from (G.9), let

$$\Pi^{(j)} = \Pi^* + c_0 K^{-1}\Gamma^{(j)}, \qquad \text{for} \quad 0 \leq j \leq J. \tag{H.8}$$

First, following the first part of proof of Theorem G.1, we will see that G^*, P^*G^* satisfy the regularity conditions in Condition 2.1. Especially, in this case, we still have

$$K\sum_{i\in\mathcal{C}_{p,k}} \frac{\theta_i^2}{H_0(i,i)} \asymp \int_{err_n^2}^{\infty} tdF_n(t) \geq \int_0^{\infty} tdF_n(t) - err_n^2 \asymp 1$$

Furthermore, one can derive

$$\|c_0^2 K^{-2}(\Gamma^{(j)})'\Theta H_0^{-1}\Theta\Gamma^{(j)}\| \leq c_0^2.$$

Similarly to the analog in the proof of Theorem G.1 , by choosing properly small c_0, we have the regularity conditions hold for $P^*G^{(j)}$'s as well. Since the proofs are quite similar, we hence omit the details.

Second, under the construction (H.8),

$$\begin{aligned}\mathcal{L}(\Pi^{(j)}, \Pi^{(k)}) = \frac{1}{n}\sum_{i=1}^{n} \|\pi_i^{(j)} - \pi_i^{(k)}\|_1 &= \frac{c_0}{nK}\sum_{i=1}^{n_0} \|\mathbf{e}_i'\big(\Gamma^{(j)} - \Gamma^{(k)}\big)\|_1 \\ &= \frac{4c_0}{nK}\|H^{(j)} - H^{(k)}\|_1 \\ &\geq C\frac{n_0}{n}\end{aligned}$$

for some constant C not relying on the other parameters. Notice that $n_0/n = \int_0^{err_n^2} \mathrm{d}F_n(t)$. Since (G.5) violates, we thus conclude that

$$\mathcal{L}(\Pi^{(j)}, \Pi^{(k)}) > C\int_0^{err_n^2} \mathrm{d}F_n(t) \geq C_3 \int \min\{\frac{err_n}{\sqrt{t \wedge 1}}, 1\}\mathrm{d}F_n(t). \tag{H.9}$$

for some constant $C_3 > 0$. Third, we claim the KL divergence in the same way as previously, $KL(\mathcal{P}_\ell, \mathcal{P}_0) = (I) + (II)$ and (I), (II) are defined in (H.4)-(H.5). By our least-favorable configurations (H.8), we bound

$$\begin{aligned}(I) &\leq \sum_{1\leq i<j\leq n_0} \frac{\Omega_{ij}^{(0)}}{(1-\Omega_{ij}^{(0)})}(\Delta_{ij}^{(\ell)})^2 = \sum_{1\leq i<j\leq n_0} \frac{\theta_i\theta_j c_0^4 K^{-4}\beta_n^2[\big(\Gamma_i^{(\ell)}\big)'\Gamma_j^{(\ell)}]^2}{[1-(1-1/K)\beta_n](1-\Omega_{ij}^{(0)})} \\ &\leq Cc_0^4\Big(\sum_{i=1}^{n_0}\theta_i\Big)^2\beta_n^2K^{-2} \leq Cc_0^4 n_0 K\cdot\frac{K^3}{\beta_n^2 n\bar{\theta}^2} \leq Cc_0^4 n_0 K\end{aligned}$$

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

$$\begin{aligned}(II) &\leq \sum_{0<i\leq n_0<j\leq n} \frac{\Omega_{ij}^{(0)}}{(1-\Omega_{ij}^{(0)})}(\widetilde{\Delta}_{ij}^{(\ell)})^2 = \sum_{\substack{0<i\leq n_0,\\ j\in\mathcal{C}_{p,\hat{j}}, 1\leq\hat{j}\leq K}} \frac{\theta_i\theta_j c_0^2K^{-2}\beta_n^2[\big(\Gamma_i^{(\ell)}\big)'\mathbf{e}_{\hat{j}}]^2}{[1-(1-1/K)\beta_n](1-\Omega_{ij}^{(0)})} \\ &\leq Cc_0^2\Big(\sum_{i=1}^{n_0}\theta_i\Big)\Big(\sum_{j=n_0+1}^{n}\theta_j\Big)K^{-2}\beta_n^2 \leq Cc_0^2 n_0 K\end{aligned}$$

where $\widetilde{\Delta}_{ij}^{(\ell)} := c_0K^{-1}\cdot\beta_n\big(\Gamma_i^{(\ell)}\big)'\mathbf{e}_{\hat{j}}/\big(1-(1-1/K)\beta_n\big)$ for this case. Combining the upper bounds for (I) and (II), we finally get

$$\sum_{1\leq\ell\leq J} \mathrm{KL}(\mathcal{P}_\ell, \mathcal{P}_0) \leq Cc_0^2 J n_0 K \leq (1/8 - \epsilon_3)J\log(J) \tag{H.10}$$

for a constant $\epsilon_3 \in (0, 1/8)$, 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 G.2 when (G.5) violates. Further by standard techniques of lower bound analysis (e.g., [9, Theorem 2.5]), we ultimately obtain the lower bound.

H.4 Extension to P-specific lower bounds

In this subsection, we study the P-specific lower bounds of $\mathcal{L}^w(\hat{\Pi}, \Pi)$ and $\mathcal{L}(\hat{\Pi}, \Pi)$ 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 $P\mathbf{1}_K \propto \mathbf{1}_K$;

(c) (K, P) can depend on n, and K can be unbounded, but $P\mathbf{1}_K \propto \mathbf{1}_K$ 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 proofs for Theorems G.1-G.2, 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 G.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

$$(\gamma^{(j)})' = \big((\omega^{(j)})', \mathbf{0}_{1\times(n-n_0)} \big). \tag{H.11}$$

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

$$\eta'\mathbf{1}_{K_0} = 0, \qquad \eta' P_0 \mathbf{1}_{K_0} = 0, \qquad \|\eta\|_1 \asymp K_0 \tag{H.12}$$

Such η 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 (G.4) for weighted loss, (G.9) or (H.8) for unweighted loss. The verifications of regularity conditions and pairwise difference between the configurations can be claimed in the same way as in the proofs of Theorems G.1-G.2. The most distinguishing part appears in the KL divergence. Especially, for $0 < i < j \leq n_0$,

$$\Omega_{ij}^{(\ell)} = \Omega_{ij}^{(0)}\Big(1 + \Delta_{ij}^{(\ell)}\Big), \quad \Delta_{ij}^{(\ell)} \propto \gamma^{(\ell)}(i)\gamma^{(\ell)}(j)\eta' P\eta \tag{H.13}$$

where the coefficients we did not specify for Δ_{ij}'s rely on the perturbation scale we take from (G.4), or (G.9), or (H.8). Similarly for $0 < i \leq n_0 < j \leq n$, if $j \in \mathcal{C}_{p,\hat{j}}$,

$$\Omega_{ij}^{(\ell)} = \Omega_{ij}^{(0)}\Big(1 + \widetilde{\Delta}_{ij}^{(\ell)}\Big), \quad \widetilde{\Delta}_{ij}^{(\ell)} \propto \gamma^{(\ell)}(i)\eta' P e_{\hat{j}}. \tag{H.14}$$

Nevertheless, in this case, $\eta' P\eta \asymp 1$ and $|\eta' P e_{\hat{j}}| \leq C$. In particular, $\beta_n \asymp 1$. All of these facts lead to similar derivations on upper bounds of (I) and (II) (see definitions in (H.4)-(H.5)). 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 G.1-G.2.

(b) If both (K, P) may depend on n, but they satisfy that $K \leq C$ and $P\mathbf{1}_K \propto \mathbf{1}_K$. We take the same simplified least-favorable configurations as in Case (a). The regularity conditions and pairwise difference can be claimed likewisely. $\mathbf{1}_K$ is an eigenvector of P. We can take special η, the eigenvector associated to the smallest eigenvalue (in magnitude) of P. In (H.13) and (H.14), we have $\eta' P\eta \asymp \beta_n$ and $|\eta' P e_{\hat{j}}| \leq C\beta_n$, which fit the arguments in the proofs of KL divergence for Theorems G.1-G.2. Thereby, we can prove the KL divergence in the same way as the proofs of Theorems G.1-G.2.

(c) If both (K, P) may depend on n and K can be unbounded, but $P\mathbf{1}_K \propto \mathbf{1}_K$ and $|\lambda_2(P)| \leq C\beta_n = o(1)$. We adopt the same least-favorable configurations in Section G correspondingly to Theorems G.1-G.2. The different parts only appear in the quantities involving P. Notice that in this case, $\mathbf{1}_K$ is the eigenvector associated to the largest eigenvalue of P,

$$(\Gamma_i^{(\ell)})' P \mathbf{1}_K \propto (\Gamma_i^{(\ell)})' \mathbf{1}_K = 0, \qquad (\Gamma_i^{(\ell)})' P \Gamma_j^{(\ell)} \asymp \beta_n \tag{H.15}$$

since the other eigenvalues of P are asymptotically of order β_n. These two estimates exactly coincide with the ones in the proofs of Theorems G.1-G.2. Then, all the arguments in the proofs of Theorems G.1-G.2 can be directly applied in this setting. We thereby conclude our proof.

I Analysis for a general b

In Section 3.2, we explained the rationale of choosing $b = 1/2$. An important claim there is that for a general b, subject to a column permutation of $\hat{\Pi}$, it holds simultaneously that

$$\|\hat{\pi}_i - \pi_i\|_1 \lesssim \frac{C\sqrt{\log(n)}}{\sqrt{n\theta_i\bar{\theta}}} \cdot \delta(b, F_n), \qquad \text{where } \delta(b, F_n) = \frac{\sqrt{\int t^{3-4b} dF_n(t)}}{\int t^{2-2b} dF_n(t)}. \tag{I.1}$$

In this section, we discuss why (I.1) is true.

Let H be the same as in (9), and let $H_0 = \mathbb{E}H$. Given a fixed $b \geq 0$, let

$$L = H^{-b} A H^{-b}, \qquad L_0 = H_0^{-b} \Omega H_0^{-b}.$$

For $1 \leq k \leq K$, let $\hat{\lambda}_k$ be the kth largest eigenvalue (in magnitude) of L, and let $\hat{\xi}_k \in \mathbb{R}^n$ be the corresponding eigenvector. Similarly, let λ_k be the kth largest eigenvalue (in magnitude) of L_0, and let $\xi_k \in \mathbb{R}^n$ be the corresponding eigenvector. From the proof of Theorem 4.2, we can see that the node-wise error $\|\hat{\pi}_i - \pi_i\|_1$ is closely related to the following quantity:

$$\mathrm{ER}_i := \max_{1 \leq k \leq K} \left\{ \frac{|\hat{\xi}_k(i) - \xi_k(i)|}{\xi_1(i)} \right\}, \tag{I.2}$$

subject to a rotation matrix $O_1 \in \mathbb{R}^{(K-1)\times(K-1)}$ applied to the columns of $\hat{\Xi}_1 = [\hat{\xi}_2, \hat{\xi}_3, \ldots, \hat{\xi}_K]$. To study ER_i, we introduce a proxy to $\hat{\xi}_k$. Note that $\hat{\lambda}_k \hat{\xi}_k = H^{-b} A H^{-b} \hat{\xi}_k$, from the definition of eigenvalues and eigenvectors. It implies $\hat{\xi}_k = \hat{\lambda}_k^{-1} H^{-b} A H^{-b} \hat{\xi}_k$. We replace $(H, \hat{\xi}_k)$ on the right hand side by (H_0, ξ_k) to obtain

$$\hat{\xi}_k^* := \lambda_k^{-1} H_0^{-b} A H_0^{-b} \xi_k, \qquad 1 \le k \le K. \tag{I.3}$$

We then similarly define

$$\mathrm{ER}_i^* := \max_{1 \le k \le K} \left\{ \frac{|\hat{\xi}_k^*(i) - \xi_k(i)|}{\xi_1(i)} \right\}. \tag{I.4}$$

The difference between ER_i^* and ER_i is governed by $\zeta := \hat{\xi}_1^* - \hat{\xi}_1 \in \mathbb{R}^n$. How to control the effect of ζ is the central topic of entry-wise eigenvector analysis. In Section 5 and Section E, we give rigorous analysis in the case of $b = 1/2$. The analysis of a general b follows a similar vein and is omitted here (the regularity conditions may be slightly different). From now on, we assume that $|\mathrm{ER}_i - \mathrm{ER}_i^*|$ is negligible and focus on studying ER_i^* in (I.4).

Lemma I.1. *Consider the DCMM model in (1)-(2). Write $G_0 = [\mathrm{tr}(\Theta M_0^2 \Theta)]^{-1} \Pi' \Theta M_0^2 \Theta \Pi$ where $M_0 = H_0^{-b}$. As $n \to \infty$, suppose K is fixed and PG_0 converges to a fixed irreducible non-singular matrix that has distinct eigenvalues. Denote by $\theta_{\max}$ and $\theta_{\min}$ the maximum and minimum of θ_i's. We assume $n\bar{\theta}\theta_{\min}/\log(n) \to \infty$ and set $\tau = 0$ in the definition of H (see (9)). Let $\hat{\xi}_1^*, \ldots, \hat{\xi}_K^*$ be defined as in (I.3). With probability $1 - o(n^{-3})$, simultaneously for all $1 \le i \le n$,*

$$\mathrm{ER}_i^* \le C\left(\sqrt{\frac{\log(n)}{\theta_i}} \times \frac{\sqrt{\sum_{j=1}^n \theta_j^{3-4b}}}{\sum_{j=1}^n \theta_j^{2-2b}} + \frac{\log(n)}{\theta_i} \times \frac{\theta_{\max}^{1-2b}}{\sum_{j=1}^n \theta_j^{2-2b}} \right). \tag{I.5}$$

We have made some strong assumptions in Lemma I.1, e.g., K is fixed, PG_0 converges to a fixed matrix with distinct eigenvalues, and $n\bar{\theta}\theta_{\min}/\log(n) \to \infty$. These assumptions are only for convenience, as we want to avoid re-defining those regularity conditions in Section 2.2 for a general b. However, there is no technical hurdle of extending Lemma I.1 to allow for weaker conditions similar to those in Section 2.2.

We now look into the right hand side of (I.5) and write it as $\mathrm{ER}_i^* \le C(\omega_i^* + \widetilde{\omega}_i)$, where ω_i^* and $\widetilde{\omega}_i$ represent the two terms in the brackets, respectively. By Definition 2.1, $\sum_{i=1}^n \theta_i^\gamma =$

$\bar{\theta}^{\gamma}\int t^{\gamma}dF_n(t)$, for any $\gamma\in\mathbb{R}$. Consequently,

$$\omega_i^* = \sqrt{\frac{\log(n)}{\theta_i}} \times \frac{\sqrt{\sum_{j=1}^n \theta_j^{3-4b}}}{\sum_{j=1}^n \theta_j^{2-2b}} = \frac{C\sqrt{\log(n)}}{\sqrt{n\theta_i\bar{\theta}}} \times \underbrace{\frac{\sqrt{\int t^{3-4b}dF_n(t)}}{\int t^{2-2b}dF_n(t)}}_{\delta(b,F_n)}. \tag{I.6}$$

In addition, note that $\sqrt{\sum_j \theta_j^{3-4b}} \le \sqrt{\theta_{\max}^{2-4b}\sum_j \theta_j} \le \theta_{\max}^{1-2b}\sqrt{n\bar{\theta}}$. We then obtain:

$$\frac{\widetilde{\omega}_i}{\omega_i^*} = \frac{\sqrt{\theta_i}}{\sqrt{\log(n)}} \times \frac{\sqrt{\sum_{j=1}^n \theta_j^{3-4b}}}{\theta_{\max}^{1-2b}} \le \frac{\sqrt{n\bar{\theta}\theta_i}}{\sqrt{\log(n)}}. \tag{I.7}$$

As long as $n\bar{\theta}\theta_i/\log(n)\ge C$, $\widetilde{\omega}_i$ is dominated by ω_i^*, simultaneously for all fixed b. Furthermore, ω_i^* is minimized at $b=1/2$. It suggests that $b=1/2$ is the universally best choice (as claimed in Section 3.2 of the main text).

I.1 Proof of Lemma I.1

We recall that $\hat{\xi}_k^*$ is a proxy to ξ_k. It is tedious to bound the difference between $\hat{\xi}_k^*$ and ξ_k and handle the rotation matrix $O_1\in\mathbb{R}^{(K-1)\times(K-1)}$. In the special case of $b=1/2$, such analysis is detailed in Section 5 and Section E. For a general b, we skip this step but only study $\hat{\xi}_k^*$. Since the statement of Lemma I.1 is only about $\hat{\xi}_k^*$, the proof is much shorter.

Recall the proxy $\hat{\xi}_k^* := \lambda_k^{-1}H_0^{-b}AH_0^{-b}\xi_k$, where $H_0(i,i)=\mathbb{E}d_i\asymp n\bar{\theta}\theta_i$. Under the assumptions on G_0 and PG_0, it can be observed that the eigenvalues $\lambda_1,\cdots,\lambda_K$ are well separated by a order of $\sum_{i=1}^n \theta_i^{2-2b}/(n\bar{\theta})^{2b}$ and

$$|\lambda_k| \asymp \sum_{j=1}^n \theta_j^{2-2b}/(n\bar{\theta})^{2b} \quad \text{for all } 1\le k\le K.$$

In the same manner to prove Lemma D.2, under the assumptions in Lemma I.1, we can also claim that

$$\xi_1(i)\asymp \frac{\theta_i^{1-b}}{\sqrt{\sum_{j=1}^n \theta_i^{2-2b}}}, \quad \max_{1\le k\le K}|\xi_k(i)| \le C\frac{\theta_i^{1-b}}{\sqrt{\sum_{j=1}^n \theta_i^{2-2b}}}$$

by replacing H_0 by M_0^2. We skip the details for simplicity. Based on the above estimates, we can derive

$$\begin{aligned}\big|\hat{\xi}_k^*(i) - \xi_k(i)\big| &\lesssim \frac{(n\bar{\theta})^{2b} M_0(i,i)}{\sum_{j=1}^n \theta_j^{2-2b}} \sum_{j=1}^n \big(A(i,j) - \Omega(i,j)\big) M_0(j,j)\xi_k(j) \\ &\lesssim \frac{(n\bar{\theta})^b}{\big(\sum_{j=1}^n \theta_j^{2-2b}\big)\theta_i^b} \sum_{j\neq i}^n W(i,j) M_0(j,j)\xi_k(j) + \frac{\theta_i^{3-3b}}{\big(\sum_{j=1}^n \theta_j^{2-2b}\big)^{\frac{3}{2}}} \,. \end{aligned} \tag{I.8}$$

for any fixed $1 \le k \le K$. To proceed, we apply Bernstein inequality (i.e., Theorem D.1) to the summation on the RHS above and get

$$\begin{aligned}&\sum_{j\neq i}^n W(i,j) M_0(j,j)\xi_k(j) \\ &\le C\sqrt{\sum_{j=1}^n \theta_i\theta_j (n\bar{\theta}\theta_j)^{-2b} \frac{\theta_j^{2-2b}}{\sum_{j=1}^n \theta_j^{2-2b}} \cdot \log(n)} + C\max_j \left|(n\bar{\theta}\theta_j)^{-b} \frac{\theta_j^{1-b}}{\sqrt{\sum_{j=1}^n \theta_j^{2-2b}}}\right| \cdot \log(n) \\ &\le C\sqrt{\theta_i \log(n)}\, \frac{(\sum_{j=1}^n \theta_j^{3-4b})^{1/2}}{(n\bar{\theta})^b (\sum_{j=1}^n \theta_j^{2-2b})^{1/2}} + \frac{C\max_j \theta_j^{1-2b}}{(n\bar{\theta})^b (\sum_{j=1}^n \theta_j^{2-2b})^{1/2}} \cdot \log(n)\end{aligned}$$

with probability at least $1 - o(n^{-3})$. Substituting the above inequality back into (I.8), together with the estimate of $\xi_1(i)$, we obtain that

$$\frac{\max_{1\le k\le K} \big|\hat{\xi}_k^*(i) - \xi_k(i)\big|}{\xi_1(i)} \lesssim \frac{\sqrt{\log(n)}}{\sqrt{\theta_i}} \frac{\big(\sum_{j=1}^n \theta_j^{3-4b}\big)^{1/2}}{\sum_{j=1}^n \theta_j^{2-2b}} + \frac{\log n}{\theta_i} \frac{\max_j \theta_j^{1-2b}}{\sum_{j=1}^n \theta_j^{2-2b}}$$

with probability at least $1 - o(n^{-4})$. By combining this inequality for all i, we conclude our proof.

J Relaxed condition on P

In Section 2.2, we introduced one regularity condition on P in Condition 2.1 (b). Specifically, we assume that

$$\min_{1\le k\le K} \Big\{ \sum_{1\le \ell\le K} P(k,\ell) \Big\} \ge c_2 K.$$

As we explained in the main context of Section 2.2, this condition can be relaxed to

$$\min_{1\leq k\leq K}\Big\{\sum_{1\leq \ell\leq K} P(k,\ell)\Big\} \geq c_2\alpha_n, \quad \text{where } \alpha_n \in [1, K]. \tag{J.1}$$

Our theory still holds under certain modification of the regularity conditions in this case. We discuss more details in this section.

Let $\alpha_n \in [1, K]$. We re-define

$$G = \alpha_n \cdot \Pi'\Theta D_\theta^{-1}\Theta\Pi \tag{J.2}$$

The following theorem holds.

Theorem J.1. *Consider the DCMM model (1)-(2), where Condition 2.1(a), the first inequality of (b), and (c) are satisfied for G defined in (J.2). In addition, suppose (J.1) holds and $K^2\alpha_n\log(n)/(n\bar{\theta}^2\beta_n^2) \to 0$ as $n \to \infty$. With probability $1 - o(n^{-3})$, there exists $\omega \in \{1, -1\}$ and an orthogonal matrix $O_1 \in \mathbb{R}^{(K-1)\times(K-1)}$ such that simultaneously for all $1 \leq i \leq n$,*

$$|\omega\hat{\xi}_1(i) - \xi_1(i)| \leq C\sqrt{\frac{K^2\theta_i\log(n)}{n^2\bar{\theta}^3\alpha_n}}\left(1 + \sqrt{\frac{\log(n)}{n\bar{\theta}\theta_i(\alpha_n/K)}}\right), \tag{J.3}$$

$$\|e_i'(\hat{\Xi}_1 O_1 - \Xi_1)\| \leq C\sqrt{\frac{K^2\alpha_n\theta_i\log(n)}{n^2\bar{\theta}^3\beta_n^2}}\left(1 + \sqrt{\frac{\log(n)}{n\bar{\theta}\theta_i(\alpha_n/K)}}\right). \tag{J.4}$$

Based on the above theorem, we have the node-wise error and rate of MSL below.

Theorem J.2. *Suppose the assumptions in Theorem J.1 hold, and additionally, Condition 2.1(d) is satisfied. Let $\hat{\Pi}$ be the output of Algorithm 1. With probability $1 - o(n^{-3})$, there exists a permutation T on $\{1, 2, \ldots, K\}$, such that simultaneously for all $1 \leq i \leq n$,*

$$\|T\hat{\pi}_i - \pi_i\|_1 \leq C\min\left\{\sqrt{\frac{K^2\alpha_n\log(n)}{n\bar{\theta}(\bar{\theta}\wedge\theta_i)\beta_n^2}},\ 1\right\}, \tag{J.5}$$

In addition, let $\mathcal{L}(\hat{\Pi}, \Pi)$ be the ℓ^1-loss in (5). Then,

$$\mathbb{E}\mathcal{L}(\hat{\Pi}, \Pi) \leq C\sqrt{\log(n)} \cdot \frac{1}{n}\sum_{i=1}^n \min\left\{\frac{K\sqrt{\alpha_n}}{\beta_n\sqrt{n\bar{\theta}(\bar{\theta}\wedge\theta_i)}},\ 1\right\}.$$

References

[1] 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.

[2] 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.

[3] 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.

[4] Davis, C. and W. M. Kahan (1970). The rotation of eigenvectors by a perturbation. iii. *SIAM J. Numer. Anal. 7*(1), 1–46.

[5] Jin, J., T. Ke, G. Moryoussef, J. Tang, and J. Wang. Improved algorithm and bounds for successive projection. In *The Twelfth International Conference on Learning Representations*.

[6] Jin, J., Z. T. Ke, and S. Luo (2024). Mixed membership estimation for social networks. *J. Econometrics 239*(2), 105369.

[7] Ke, Z. T. and J. Jin (2023). Special invited paper: The SCORE normalization, especially for highly heterogeneous network and text data. *Stat 12*(1).

[8] Stewart, G. W. and J.-g. Sun (1990). Matrix perturbation theory. *Computer Science and Scientific Computing*.

[9] Tsybakov, A. B. (2009). Introduction to nonparametric estimation. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats.