Supplement for Estimating Mixed Memberships with Sharp Eigenvector Deviations

Abstract

In this document we present technical details and accompanying lemmas which are necessary for the main results in the main manuscript. When we make references to equations or theorems etc. in the main document, we follow the numbering scheme of the main document, and the references do not have any alphabets in them.

I Identifiability

Our proof links the MMSB parameters $\mathbf{\Theta}$ and $\mathbf{B}$ to the eigen-decomposition of the probability matrix $\mathbf{P}$, and then exploits its geometric structure. Specifically, we show that the eigenvector row corresponding to any node lies inside a polytope whose vertices correspond to pure nodes. When $\mathbf{B}$ is full rank, the polytope has K linearly independent vertices, and the community memberships $\boldsymbol{\theta}_i$ of each node i are fixed by the position of its eigenvector row with respect to these vertices. This proves part (a) of Theorem 2.1. When $\mathbf{B}$ is rank-deficient, the points corresponding to the pure nodes are linearly dependent. However, under the conditions of part (b), the constraints on $\mathbf{\Theta}$ and $\mathbf{B}$ are shown to make the model identifiable. In other cases, we construct a new $\mathbf{\Theta}'$ that still yields the same probability matrix $\mathbf{P}$. This proves part (c).

Proof of Theorem 2.1. Without loss of generality, we absorb ρ in $\mathbf{B}$, and reorder nodes so that the first K nodes contain one pure node from each community. Thus, $\mathbf{\Theta}(1:K,:) = \mathbf{I}_K$.

Let $\mathbf{P} = \mathbf{V}\mathbf{E}\mathbf{V}^T$ be the eigen-decomposition of $\mathbf{P}$, with $\mathbf{V} \in \mathbb{R}^{n\times \mathrm{rank}(\mathbf{B})}$. Let $\mathbf{V}_P = \mathbf{V}(1:K,:)$. Lemma 2.3 shows that $\mathbf{V} = \mathbf{\Theta}\mathbf{V}_P$. Thus, for any node i, $\mathbf{V}(i,:)$ lies in the convex hull of the K rows of $\mathbf{V}_P$, that is, $\mathbf{V}(i,:) \in \mathrm{Conv}(\mathbf{V}_P)$. We will slightly abuse the classical notation to denote by $\mathrm{Conv}(\mathbf{M})$ the convex hull of the rows of matrix $\mathbf{M}$.

Now, suppose $\mathbf{P}$ can be generated by another set of parameters $(\mathbf{\Theta}', \mathbf{B}')$, where $\mathbf{\Theta}'$ has a different set of pure nodes, with indices $\mathcal{I} \neq 1:K$. By the previous argument, we must have $\mathbf{V}(\mathcal{I},:) \subseteq \mathrm{Conv}(\mathbf{V}_P)$. Since $(\mathbf{\Theta}', \mathbf{B}')$ and $(\mathbf{\Theta}, \mathbf{B})$ have the same probability matrix $\mathbf{P}$, they have the same eigen-decomposition up to a permutation of the communities. Thus, swapping the roles of $\mathbf{\Theta}$ and $\mathbf{\Theta}'$ and reapplying the above argument, we find that $\mathbf{V}_P \subseteq \mathrm{Conv}(\mathbf{V}(\mathcal{I},:))$. Then $\mathrm{Conv}(\mathbf{V}_P) \subseteq \mathrm{Conv}(\mathbf{V}(\mathcal{I},:)) \subseteq \mathrm{Conv}(\mathbf{V}_P)$, so we must have $\mathrm{Conv}(\mathbf{V}_P) = \mathrm{Conv}(\mathbf{V}(\mathcal{I},:))$. This means the pure nodes in $\mathbf{\Theta}$ and $\mathbf{\Theta}'$ are aligned up to a permutation, that is, $\mathbf{V}(\mathcal{I},:) = \mathbf{M}\mathbf{V}_P$, where $\mathbf{M} \in \mathbb{R}^{K\times K}$ is a permutation matrix.

Now, $\mathbf{V} = \mathbf{\Theta}\mathbf{V}_P = \mathbf{\Theta}'\mathbf{V}(\mathcal{I}, :) = \mathbf{\Theta}'\mathbf{M}\mathbf{V}_P$, which implies

$$(\mathbf{\Theta} - \mathbf{\Theta}'\mathbf{M})\mathbf{V}_P = 0 \tag{26}$$

Since $\mathbf{V} = \mathbf{\Theta}\mathbf{V}_P$ and $\text{rank}(\mathbf{\Theta}) = K$, we have $\text{rank}(\mathbf{V}_P) = \text{rank}(\mathbf{V}) = \text{rank}(\mathbf{B})$. Hence, if $\text{rank}(\mathbf{B}) = K$, $\mathbf{V}_P$ is full rank, so $\mathbf{\Theta} = \mathbf{\Theta}'\mathbf{M}$. Thus, $\mathbf{\Theta}$ and $\mathbf{\Theta}'$ are identical up to a permutation. To have the same $\mathbf{P}$, $\mathbf{B}$ and $\mathbf{B}'$ must also be identical up to the same permutation. Hence, the MMSB model is identifiable. This proves part (a).

Now, suppose $\text{rank}(\mathbf{B}) = K - \ell < K$. We first permute the columns of $\mathbf{\Theta}$, and the rows and columns of $\mathbf{B}$, so that

$$\mathbf{B} = \left[\begin{array}{c|c} \mathbf{C} & \mathbf{C}\mathbf{W} \\ \hline \mathbf{W}^T\mathbf{C} & \mathbf{W}^T\mathbf{C}\mathbf{W} \end{array}\right], \tag{27}$$

where $\mathbf{C} \in \mathbb{R}^{(K-\ell)\times(K-\ell)}$ is full rank, and $\mathbf{W} \in \mathbb{R}^{(K-\ell)\times\ell}$. We see that

$$\begin{aligned} \mathbf{C}\left[\begin{array}{c|c} \mathbf{I}_{K-\ell} & \mathbf{W} \end{array}\right] &= \mathbf{V}\left(1 : (K-\ell), :\right)\mathbf{E}\mathbf{V}_P^T, \\ \mathbf{W}^T\mathbf{C}\left[\begin{array}{c|c} \mathbf{I}_{K-\ell} & \mathbf{W} \end{array}\right] &= \mathbf{V}\left((K-\ell+1) : K, :\right)\mathbf{E}\mathbf{V}_P^T. \end{aligned}$$

The first equation shows that $\text{rank}(\mathbf{V}\left(1 : (K-\ell), :\right)) = \text{rank}(\mathbf{C}) = K-\ell$, so $\mathbf{V}\left(1 : (K-\ell), :\right)$ is full rank. Hence,

$$\mathbf{V}\left((K-\ell+1) : K, :\right) = \mathbf{W}^T\mathbf{V}\left(1 : (K-\ell), :\right) \Rightarrow \mathbf{V}_P = \left[\begin{array}{c} \mathbf{I}_{K-\ell} \\ \hline \mathbf{W}^T \end{array}\right]\mathbf{V}\left(1 : (K-\ell), :\right) \tag{28}$$

Case 1: $\text{rank}(\mathbf{B}) = K - 1$ *(so* $\mathbf{W}$ *is a vector) and* $\mathbf{W}^T\mathbf{1}_{K-\ell} \neq 1$*.*

Now using Eqs (26) and (28), we have

$$(\mathbf{\Theta} - \mathbf{\Theta}'\mathbf{M})\left[\begin{array}{c} \mathbf{I}_{K-\ell} \\ \hline \mathbf{W}^T \end{array}\right]\mathbf{V}\left(1 : (K-\ell), :\right) = \mathbf{0} \Rightarrow \mathbf{\Theta} = \mathbf{\Theta}'\mathbf{M}. \tag{29}$$

The above equation is derived using $\mathbf{\Theta}\mathbf{1}_K = \mathbf{\Theta}'\mathbf{1}_K = \mathbf{1}_n$, and $\mathbf{W}^T\mathbf{1}_{K-\ell} \neq 1$.

Clearly $\mathbf{B}' = \mathbf{M}\mathbf{B}\mathbf{M}'$ as well, so the MMSB model is identifiable. From Eq (27), we have $\mathbf{B}\left((K-\ell+1):K,:\right) = \mathbf{W}^T\mathbf{B}\left(1:(K-\ell),:\right)$, so $\mathbf{W}^T\mathbf{1}_{K-\ell} \neq 1$ if and only if the last row of $\mathbf{B}$ is not a affine combination of the remaining rows. It is easy to see that the same holds for any row of $\mathbf{B}$. This proves part (b).

Case 2: rank$(\mathbf{B}) = K-1$ *and* $\mathbf{W}^T\mathbf{1}_{K-\ell} = 1$*, or* rank$(\mathbf{B}) < K-1$.

We will construct a $\mathbf{\Theta}' \neq \mathbf{\Theta}$ that yields the same probability matrix $\mathbf{P}$. Let the completely mixed node be m, so $\theta_{mj} > 0$ for all communities j. We use

$$\boldsymbol{\theta}'_j = \begin{cases} \boldsymbol{\theta}_j & \text{if } j \neq m \\ \boldsymbol{\theta}_m + \epsilon\boldsymbol{\beta}^T\left[-\mathbf{W}^T \mid \mathbf{I}_l\right] & \text{if } j = m, \end{cases},$$

where ϵ is small enough that $\theta'_{mj} \in (0,1)$ for all communities j, and $\boldsymbol{\beta} \in \mathbb{R}^\ell \neq \mathbf{0}$ is such that $\boldsymbol{\beta}^T\left[-\mathbf{W}^T\mathbf{1}_{K-\ell} + \mathbf{1}_\ell\right] = 0$. Note that such a $\boldsymbol{\beta}$ always exists when $\ell > 1$ and can be arbitrary vector when $\mathbf{W}^T\mathbf{1}_{K-\ell} = \mathbf{1}_\ell$. Hence, each row of $\mathbf{\Theta}'$ sums to 1, and $\mathbf{\Theta}'$ is a valid community-membership matrix. Additionally, $\mathbf{\Theta}'\mathbf{V}_P = \mathbf{\Theta}\mathbf{V}_P$.

Finally, we will show that $(\mathbf{\Theta}', \mathbf{B})$ and $(\mathbf{\Theta}, \mathbf{B})$ generate the same probability matrix. Note that $\mathbf{B} = \mathbf{P}_{1:K,1:K} = \mathbf{V}_P\mathbf{E}\mathbf{V}_P^T$. Hence,

$$\mathbf{\Theta}\mathbf{B}\mathbf{\Theta}^T = \mathbf{\Theta}\mathbf{V}_P\mathbf{E}\mathbf{V}_P^T\mathbf{\Theta}^T = \mathbf{V}\mathbf{E}\mathbf{V}^T = \mathbf{P} = \mathbf{\Theta}'\mathbf{V}_P\mathbf{E}\mathbf{V}_P^T\mathbf{\Theta}'^T = \mathbf{\Theta}'\mathbf{B}\mathbf{\Theta}'^T.$$

This proves part (c). □

Proof of Theorem 2.2. Consider an MMSB model parameterized by $(\mathbf{\Theta}^{(1)}, \mathbf{B}^{(1)})$, with $\mathbf{P} = \mathbf{\Theta}^{(1)}\mathbf{B}^{(1)}{\mathbf{\Theta}^{(1)}}^T$ (we absorb ρ in $\mathbf{B}$ without loss of generality). We want to construct a $(\mathbf{\Theta}^{(2)}, \mathbf{B}^{(2)})$ that gives the same probability matrix $\mathbf{P}$. The idea is to construct a matrix $\mathbf{M}$ such that $\mathbf{\Theta}^{(2)} = \mathbf{\Theta}^{(1)}\mathbf{M}$ and $\mathbf{B}^{(2)} = \mathbf{M}^{-1}\mathbf{B}^{(1)}(\mathbf{M}^T)^{-1}$. The difficulty is in ensuring that all constraints are satisfied: $\mathbf{\Theta}^{(2)}\mathbf{1}_K = \mathbf{1}_n$, $\mathbf{\Theta}^{(2)} \geq 0$, and $0 \leq \mathbf{B}^{(2)}_{ij} \leq 1$ for all i, j.

Without loss of generality, suppose that the first community does not have any pure nodes. In other words, for all nodes $i \in [n]$, $\theta^{(1)}_{i1} \leq 1 - \delta$ for some $\delta > 0$. Consider the following $\mathbf{M}$:

$$\mathbf{M} = \left[\begin{array}{c|c} 1+(K-1)\epsilon^2 & -\epsilon^2 \mathbf{1}_{K-1}^T \\ \hline \mathbf{0} & \epsilon \mathbf{1}_{K-1}\mathbf{1}_{K-1}^T + (1-(K-1)\epsilon)\mathbf{I}_{K-1} \end{array}\right],$$

where ϵ is a small positive number ($0 < \epsilon < \delta$). It is easy to check that $\mathbf{M}$ is full rank (for small enough ϵ) and $\mathbf{M} \cdot \mathbf{1}_K = \mathbf{1}_K$. Hence, $\mathbf{\Theta}^{(2)}\mathbf{1}_K = \mathbf{\Theta}^{(1)}\mathbf{M}\mathbf{1}_K = \mathbf{1}_n$

For any node i and for $j > 1$,

$$\begin{aligned} \theta_{i1}^{(2)} &= \theta_{i1}^{(1)}(1+(K-1)\epsilon^2) \geq 0, \\ \theta_{ij}^{(2)} &= -\theta_{i1}^{(1)}\epsilon^2 + \sum_{\ell=2}^{K} \theta_{i\ell}^{(1)}\mathbf{M}_{\ell j} \geq -\theta_{i1}^{(1)}\epsilon^2 + \epsilon \sum_{\ell=2}^{K} \theta_{i\ell}^{(1)} \geq \epsilon\delta^2 > 0, \end{aligned}$$

where we used $\epsilon < \delta$ and $\boldsymbol{\theta}_i^{(1)}\mathbf{1}_K = 1$. Hence, $\mathbf{\Theta}^{(2)} \geq 0$.

Finally, we must show that $\mathbf{B}^{(2)} = \mathbf{M}^{-1}\mathbf{B}^{(1)}(\mathbf{M}^T)^{-1}$ has all elements between 0 and 1. Note that

$$\mathbf{M} - \mathbf{I}_K = \left[\begin{array}{c|c} (K-1)\epsilon^2 & -\epsilon^2 \mathbf{1}_{K-1}^T \\ \hline \mathbf{0} & \epsilon \mathbf{1}_{K-1}\mathbf{1}_{K-1}^T - (K-1)\epsilon\mathbf{I}_{K-1} \end{array}\right],$$

so $\|\mathbf{M} - \mathbf{I}_K\|_F \to 0$ as $\epsilon \to 0$. Since $\mathbf{M}^{-1}$ is continuous in $\mathbf{M}$, we have $\|\mathbf{M}^{-1} - \mathbf{I}_K\|_F \to 0$. Thus,

$$\begin{aligned} \left\|\mathbf{B}^{(2)} - \mathbf{B}^{(1)}\right\|_F &= \left\|\mathbf{M}^{-1}\mathbf{B}^{(1)}(\mathbf{M}^T)^{-1} - \mathbf{B}^{(1)}\right\|_F \\ &\leq \left\|\mathbf{M}^{-1} - \mathbf{I}_K\right\|_F^2 \left\|\mathbf{B}^{(1)}\right\|_F + 2\left\|\mathbf{M}^{-1} - \mathbf{I}_K\right\|_F \left\|\mathbf{B}^{(1)}\right\|_F \\ &\to 0 \text{ as } \epsilon \to 0. \end{aligned}$$

Since $\mathbf{B}_{ij}^{(1)} \in (0,1)$, we have $\mathbf{B}_{ij}^{(2)} \in (0,1)$ for ϵ small enough, completing the proof. □

II Some Auxiliary Results, Proof of Lemmas 3.2

Definition II.1. (A construction of rotation matrix) Consider the discretization defined in Definition 5.1. The Davis-Kahan Theorem states that there exists a rotation matrix $\hat{\mathbf{O}}$ such that $\|\hat{\mathbf{V}} - \mathbf{V}\hat{\mathbf{O}}\|_F$ is small. In this definition we will carefully construct this matrix. Consider the intervals resulting from the discretization of population eigenvalues in Definition 5.1. Now, from Theorem 2 of [1], $\exists \hat{\mathbf{O}}_k$ such that

$$\|\mathbf{R}_k\|_F = \|\hat{\mathbf{V}}_{S_k} - \mathbf{V}_{S_k}\hat{\mathbf{O}}_k\|_F \leq \frac{\sqrt{8n_k}\|\mathbf{A} - \mathbf{P}\|}{g_k} \tag{30}$$

Typically the denominator is $f_k := \min(\lambda_{s_k} - \lambda_{s_k-1}, \min(\lambda_{e_k} - \lambda_{e_k+1}, \lambda_{e_k}))$. We now construct our $\hat{\mathbf{O}}$ by stacking the $\hat{\mathbf{O}}_k$ matrices on the diagonal of a $K \times K$ matrix. This is also a valid rotation matrix. Now, let $\mathbf{E}_k$ by the submatrix of $\mathbf{E}$ corresponding to eigenvalues in S_k. Similarly define $\hat{\mathbf{E}}_k$. Furthermore, let $\mathbf{R} := [\mathbf{R}_1| \dots |\mathbf{R}_I]$.

Lemma II.1. *If Assumption 3.1 holds, then there exists an orthogonal matrix* $\hat{\mathbf{O}} \in \mathbb{R}^{K \times K}$ *constructed using Definition II.1 that satisfies*

$$\|\mathbf{R}\|_F \leq \frac{\sqrt{8K}\|\mathbf{A} - \mathbf{P}\|}{\lambda^*(\mathbf{P})} \tag{31}$$

$$\left\|\hat{\mathbf{E}} - \hat{\mathbf{O}}^T\mathbf{E}\hat{\mathbf{O}}\right\|_F = O_P\left(K^2\sqrt{n\rho}\right) \tag{32}$$

with probability larger than $1 - n^{-3}$.

Proof. Consider the rotation matrix $\hat{\mathbf{O}}$, the residual matrix $\mathbf{R}$ constructed as in Definition II.1. This gives us the Frobenius norm of $\mathbf{R}$ as follows, since by construction $g_k \geq \lambda^*(\mathbf{P})$.

$$\|\mathbf{R}\|_F \leq \sqrt{\sum_k \|\mathbf{R}_k\|_F^2} \leq \frac{\sqrt{8K}\|\mathbf{A} - \mathbf{P}\|}{\lambda^*(\mathbf{P})}$$

Finally note that, using Lemma III.1, since $\lambda_{s_k} \leq \sum_{i=1}^k n_i g_k \leq K g_k$,

$$\|\mathbf{R}_k\hat{\mathbf{O}}_k^T\mathbf{E}_k\|_F \leq \|\mathbf{R}_k\|_F\|\mathbf{E}_k\| \leq \frac{\sqrt{8n_k}\|\mathbf{A} - \mathbf{P}\|\lambda_{s_k}}{g_k} = O_P(K\sqrt{n_k}\sqrt{n\rho}) \tag{33}$$

Now we use these intervals as follows.

$$\begin{aligned}\left\|\hat{\mathbf{E}} - \hat{\mathbf{O}}^T\mathbf{E}\hat{\mathbf{O}}\right\|_F &= \left\|\hat{\mathbf{V}}\hat{\mathbf{E}}\hat{\mathbf{V}}^T - \hat{\mathbf{V}}\hat{\mathbf{O}}^T\mathbf{E}\hat{\mathbf{O}}\hat{\mathbf{V}}^T\right\|_F \\ &= \|\mathbf{A}_K - (\mathbf{V}+\mathbf{R}\hat{\mathbf{O}}^T)\mathbf{E}(\mathbf{V}+\mathbf{R}\hat{\mathbf{O}}^T)^T\|_F \\ &\le \|\mathbf{A}_K - \mathbf{P}\|_F + 2\underbrace{\|\mathbf{R}\hat{\mathbf{O}}^T\mathbf{E}\mathbf{V}^T\|_F}_{P_1} + \underbrace{\|\mathbf{R}\hat{\mathbf{O}}^T\mathbf{E}\hat{\mathbf{O}}\mathbf{R}^T\|_F}_{P_2} \\ &= O_P(\sqrt{Kn\rho}) + P_1 + P_2\end{aligned}$$

The last step is true because $\|\mathbf{A}_K - \mathbf{P}\|_F \le \sqrt{K}\|\mathbf{A}_K - \mathbf{P}\| \le \sqrt{K}(\|\mathbf{A}-\mathbf{A}_K\| + \|\mathbf{A}-\mathbf{P}\|) \le 2\sqrt{Kn\rho}$ with probability at least $1 - n^{-r}$ using Weyl's inequality and Theorem 5.2 of [2]. As for P_1, note that: $P_1 \le \|\mathbf{R}\hat{\mathbf{O}}^T\mathbf{E}\|_F \le \sqrt{\sum_k \|\mathbf{R}_k\hat{\mathbf{O}}_k^T\mathbf{E}_k\|_F^2} =: O_P(K^{3/2}\sqrt{n\rho})$.

As for P_2, we have:

$$P_2 \le \|\mathbf{R}\hat{\mathbf{O}}^T\mathbf{E}\|_F\|\mathbf{R}\|_F = O_P\left(\frac{K^2 n\rho}{\lambda^*(\mathbf{P})}\right)$$

Thus the final bound is $O_P(K^2\sqrt{n\rho}(\max(1/K^{3/2}, 1/\sqrt{K}, \sqrt{n\rho}/\lambda^*(\mathbf{P}))) = O_P(K^2\sqrt{n\rho})$ using Assumption 3.1. The failure probability comes from the failure of event $\|\mathbf{A}-\mathbf{P}\| = O_P(\sqrt{n\rho})$. Taking $r = 3$ we get the required bound. □

Lemma II.2. *For $\mathbf{\Theta} \in \mathbb{R}^{n\times K}$, with $\|\boldsymbol{\theta}_i\|_1 = 1$ and $\theta_{ij} \ge 0$, $\forall i,j \in [n]$, $\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta}) \le \max_j \mathbf{1}_n^T\mathbf{\Theta}\mathbf{e}_j \le n$ and $\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}) \le \min_j \mathbf{1}_n^T\mathbf{\Theta}\mathbf{e}_j$.*

Proof. By Proposition 2.4 of [3], as $\mathbf{\Theta}^T\mathbf{\Theta}$ is a nonnegative matrix, $\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})$ is upper bounded by its largest row sum and $\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})$ is lower bounded by its smallest row sum. For the i-th row of $\mathbf{\Theta}^T\mathbf{\Theta}$, its row sum is:

$$\mathbf{e}_i^T\mathbf{\Theta}^T\mathbf{\Theta}\mathbf{1}_K = \mathbf{e}_i^T\mathbf{\Theta}^T\mathbf{1}_n = \mathbf{1}_n^T\mathbf{\Theta}\mathbf{e}_i \le n.$$

Thus the result follows. □

Lemma II.3. *Under Assumption 2.1, we have $\mathbf{\Theta}^T\mathbf{\Theta} = \left(\mathbf{V}_P\mathbf{V}_P^T\right)^{-1}$, which implies $\lambda_1(\mathbf{V}_P\mathbf{V}_P^T) = 1/\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})$, $\lambda_K(\mathbf{V}_P\mathbf{V}_P^T) = 1/\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})$ and $\kappa(\mathbf{V}_P\mathbf{V}_P^T) = \kappa\left(\mathbf{\Theta}^T\mathbf{\Theta}\right)$.*

Proof. From Lemma 2.3, $\mathbf{V} = \boldsymbol{\Theta}\mathbf{V}_P$, so

$$\mathbf{I} = \mathbf{V}^T\mathbf{V} = \mathbf{V}_P^T\boldsymbol{\Theta}^T\boldsymbol{\Theta}\mathbf{V}_P.$$

As $\mathbf{V}_P$ is full rank, we have $\boldsymbol{\Theta}^T\boldsymbol{\Theta} = \left(\mathbf{V}_P\mathbf{V}_P^T\right)^{-1}$, which gives

$$\lambda_1(\mathbf{V}_P\mathbf{V}_P^T) = \frac{1}{\lambda_K\left(\boldsymbol{\Theta}^T\boldsymbol{\Theta}\right)} \quad \text{and} \quad \lambda_K(\mathbf{V}_P\mathbf{V}_P^T) = \frac{1}{\lambda_1\left(\boldsymbol{\Theta}^T\boldsymbol{\Theta}\right)},$$

so $\kappa(\mathbf{V}_P\mathbf{V}_P^T) = \kappa\left(\boldsymbol{\Theta}^T\boldsymbol{\Theta}\right)$. □

Proof of Lemma 3.6. If $\boldsymbol{\theta}_i \sim \text{Dirichlet}(\boldsymbol{\alpha})$, let us consider $\boldsymbol{\theta}_i$ as a random variable. Denote

$$\hat{\mathbf{M}} = \boldsymbol{\Theta}^T\boldsymbol{\Theta} = \sum_{i=1}^{n} \boldsymbol{\theta}_i\boldsymbol{\theta}_i^T.$$

Note that $\hat{\mathbf{M}} - \mathrm{E}[\hat{\mathbf{M}}] = \sum_i \mathbf{X}_i$ where $\mathbf{X}_i$ are independent mean zero symmetric $K \times K$ random matrices. We have

$$\mathrm{E}[\boldsymbol{\theta}_i\boldsymbol{\theta}_i^T] = \frac{\mathrm{diag}(\boldsymbol{\alpha}) + \boldsymbol{\alpha}\boldsymbol{\alpha}^T}{\alpha_0(1+\alpha_0)} \qquad \mathrm{Cov}(\boldsymbol{\theta}_i) = \frac{\alpha_0\mathrm{diag}(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\boldsymbol{\alpha}^T}{\alpha_0^2(1+\alpha_0)}.$$

Furthermore, since $\|\boldsymbol{\theta}_i\|_1 = 1$, and $\alpha_0 = \sum_i \alpha_i$, we have $\|\mathbf{X}_i\| \le \boldsymbol{\theta}_i^T\boldsymbol{\theta}_i + \|\mathrm{E}[\boldsymbol{\theta}_i\boldsymbol{\theta}_i^T]\| \le 1 + \frac{\alpha_{\max} + \|\boldsymbol{\alpha}\|^2}{\alpha_0(1+\alpha_0)} \le 2$. Finally, since the operator norm is convex, Jensen's inequality gives: $\|\mathrm{E}[\mathbf{X}_i^2]\| \le \mathrm{E}[\|\mathbf{X}_i^2\|] \le \mathrm{E}[\|\mathbf{X}_i\|^2] \le 4$. Using standard Matrix Bernstein type concentration results (Theorem 1.4 of [4]), for large n we get:

$$\mathrm{P}(\|\hat{\mathbf{M}} - \mathrm{E}[\hat{\mathbf{M}}]\| \ge t) \le K \exp\left(-\frac{t^2}{8n + 4t/3}\right) =: \delta_t$$

Now Weyl's inequality gives, with probability at least $1 - \delta_t$,

$$|\lambda_1(\hat{\mathbf{M}}) - \lambda_1(\mathrm{E}[\hat{\mathbf{M}}])| \le t \qquad |\lambda_K(\hat{\mathbf{M}}) - \lambda_K(\mathrm{E}[\hat{\mathbf{M}}])| \le t$$

For the population quantities,

$$\lambda_1(\mathrm{E}[\hat{\mathbf{M}}]) \le \frac{\alpha_{\max} + \|\boldsymbol{\alpha}\|^2}{\alpha_0(1+\alpha_0)} n, \qquad \lambda_K(\mathrm{E}[\hat{\mathbf{M}}]) \ge \frac{\alpha_{\min}}{\alpha_0(1+\alpha_0)} n$$

For $\hat{\lambda}_1(\hat{\mathbf{M}})$ we take $t = \frac{n}{2}\frac{\alpha_{\max}+\|\boldsymbol{\alpha}\|^2}{\alpha_0(1+\alpha_0)} \in [\frac{n}{2\nu(1+\alpha_0)}, \frac{n}{2}]$ and hence $\delta_t \le K \exp\left(-\frac{n}{36\nu^2(1+\alpha_0)^2}\right)$. For $\hat{\lambda}_K(\hat{\mathbf{M}})$, we take $t = \frac{n}{2}\frac{\alpha_{\min}}{\alpha_0(1+\alpha_0)} \in [\frac{n}{2\nu(1+\alpha_0)}, \frac{n}{2}]$. Hence $\delta_t \le K \exp(-\frac{n}{36\nu^2(1+\alpha_0)^2})$.

Hence the condition number of $\hat{\mathbf{M}}$ can also be bounded as:

$$\kappa(\boldsymbol{\Theta}^T\boldsymbol{\Theta}) = \frac{\lambda_1(\boldsymbol{\Theta}^T\boldsymbol{\Theta})}{\lambda_K(\boldsymbol{\Theta}^T\boldsymbol{\Theta})} \le \frac{\frac{3n}{2}\frac{\alpha_{\max}+\|\boldsymbol{\alpha}\|^2}{\alpha_0(\alpha_0+1)}}{\frac{n}{2}\frac{\alpha_{\min}}{\alpha_0(\alpha_0+1)}} = 3\frac{\alpha_{\max}+\|\boldsymbol{\alpha}\|^2}{\alpha_{\min}}$$

□

Lemma II.4. *Let $\lambda^*(\mathbf{P})$ denote the K^{th} largest singular value of $\mathbf{P}$. We have $\lambda^*(\mathbf{P}) \ge \rho\lambda^*(\mathbf{B})\lambda_K(\boldsymbol{\Theta}^T\boldsymbol{\Theta})$.*

Proof. First note that, by Theorem 1.3.22 of [5], we have $(\mathbf{B}\boldsymbol{\Theta}^T)\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}(\mathbf{B}\boldsymbol{\Theta}^T)$ have the same K largest eigenvalues in magnitude, then

$$\lambda^*(\mathbf{P}) = \lambda^*(\rho\boldsymbol{\Theta}\mathbf{B}\boldsymbol{\Theta}^T) = \lambda^*(\rho\mathbf{B}\boldsymbol{\Theta}^T\boldsymbol{\Theta}) \ge \rho\lambda^*(\mathbf{B})\lambda_K(\boldsymbol{\Theta}^T\boldsymbol{\Theta}). \tag{34}$$

The inequality holds because for all full rank positive definite matrix $\mathbf{M}_1$, $\mathbf{M}_2 \in \mathbb{R}^{K\times K}$, $\|(\mathbf{M}_1\mathbf{M}_2)^{-1}\| \le \|\mathbf{M}_1^{-1}\|\|\mathbf{M}_2^{-1}\|$, and as $\sigma_K(\mathbf{M}_1) = 1/\|\mathbf{M}_1^{-1}\|$ (same for $\mathbf{M}_1$ and $\mathbf{M}_1\mathbf{M}_2$), where $\sigma_K(.)$ denotes the K^{th} largest singular value of a matrix. Then we have:

$$\sigma_K(\mathbf{M}_1\mathbf{M}_2) \ge \sigma_K\left(\mathbf{M}_1\right)\sigma_K\left(\mathbf{M}_2\right).$$

□

Proof of Lemma 3.2. Note that $\boldsymbol{\Theta}^T\boldsymbol{\Theta} = \left(\mathbf{V}_P\mathbf{V}_P^T\right)^{-1}$ by Lemma II.3, thus for pure nodes,

$$\max_i \left\|\mathbf{e}_i^T\mathbf{V}_P\right\|^2 = \max_i \mathbf{e}_i^T\mathbf{V}_P\mathbf{V}_P^T\mathbf{e}_i \le \max_{\|\mathbf{x}\|=1} \mathbf{x}^T\mathbf{V}_P\mathbf{V}_P^T\mathbf{x} = \lambda_1\left(\mathbf{V}_P\mathbf{V}_P^T\right) = \frac{1}{\lambda_K\left(\boldsymbol{\Theta}^T\boldsymbol{\Theta}\right)}$$

As for other nodes, their rows are convex combinations of the rows of pure nodes and would be smaller than or equal to the norm of the pure nodes. Thus the result follows.

Note that by Lemma 2.3, $\mathbf{e}_i^T\mathbf{V} = \boldsymbol{\theta}_i^T\mathbf{V}_P$, then

$$\begin{aligned}\min_i \left\|\mathbf{e}_i^T\mathbf{V}\right\|^2 &= \min_i \boldsymbol{\theta}_i^T\mathbf{V}_P\mathbf{V}_P^T\boldsymbol{\theta}_i^T = \min_i \|\boldsymbol{\theta}_i\|^2\frac{\boldsymbol{\theta}_i^T}{\|\boldsymbol{\theta}_i\|}\mathbf{V}_P\mathbf{V}_P^T\frac{\boldsymbol{\theta}_i}{\|\boldsymbol{\theta}_i\|} \ge \min_i \|\boldsymbol{\theta}_i\|^2 \min_{\|\mathbf{x}\|=1} \mathbf{x}^T\mathbf{V}_P\mathbf{V}_P^T\mathbf{x} \\ &= \min_i \|\boldsymbol{\theta}_i\|^2\lambda_K\left(\mathbf{V}_P\mathbf{V}_P^T\right) = \frac{\min_i \|\boldsymbol{\theta}_i\|^2}{\lambda_1\left(\boldsymbol{\Theta}^T\boldsymbol{\Theta}\right)} \ge \frac{1}{K\lambda_1\left(\boldsymbol{\Theta}^T\boldsymbol{\Theta}\right)}\end{aligned}$$

where for the last inequality we use for any $i \in [n]$, $\|\boldsymbol{\theta}_i\| \geq \|\boldsymbol{\theta}_i\|_1/\sqrt{K} = 1/\sqrt{K}$. Thus the result follows. □

III Proofs for Section 5

III.1 Proofs of Lemma 5.1

Lemma III.1. *Consider the intervals defined in Definition 5.1. We have for positive eigenvalues:* $\lambda_{s_k} \leq \sum_{i=1}^{k} n_i g_k$.

Proof. We prove this by induction. First, note that the smallest positive eigenvalue is larger than $\lambda^*(\mathbf{P})$ by definition. For $k = 1$, $\lambda_{s_1} - \lambda_{e_1} \leq (n_1 - 1)\lambda_{e_1}$, and hence $\lambda_{s_1} \leq n_1 \lambda_{e_1} = n_1 g_1$. Now assume that $\lambda_{s_k} \leq \sum_{i=1}^{k} n_i g_k$. Hence,

$$\lambda_{s_{k+1}} \leq (n_{k+1} - 1) g_{k+1} + (\lambda_{e_{k+1}} - \lambda_{s_k}) + \lambda_{s_k} = n_{k+1} g_{k+1} + \sum_{i=1}^{k} n_i g_k \leq \sum_{i=1}^{k+1} n_i g_{k+1}.$$

The last step holds since $g_k < g_{k+1}$. □

Proofs of Lemma 5.1. First consider positive eigenvalues. By Lemma III.1, $\lambda_{s_k} \leq g_k \sum_{i=1}^{k} n_i \leq g_k \sum_{i=1}^{I^+} n_i \leq g_k K$ and by definition $\lambda_{s_k}/g_k \leq \kappa(\mathbf{P})$, so $\lambda_{s_k}/g_k \leq \min\{K, \kappa(\mathbf{P})\}$. From construction of the eigenvalue intervals, we have: $\lambda_{s_k} - \lambda_{s_{k-1}} \leq n_k g_k$. Also note that $\sum_{k=1}^{I^+} \left(\lambda_{s_k} - \lambda_{s_{k-1}}\right)/g_k = \sum_{k=1}^{I^+} \left(\lambda_{s_k} - \lambda_{s_{k-1}}\right)/\lambda^*(\mathbf{P}) = \lambda_{s_{I^+}}/\lambda^*(\mathbf{P}) \leq \sigma_1(\mathbf{P})/\lambda^*(\mathbf{P}) = \kappa(\mathbf{P})$, where $\sigma_1(\mathbf{P})$ is the largest singular value of $\mathbf{P}$, we have $\sum_{k=1}^{I^+} \left(\left(\lambda_{s_k} - \lambda_{s_{k-1}}\right)/g_k\right) \leq \min\{K, \kappa(\mathbf{P})\}$. A similar argument can be made for negative eigenvalues. So $\psi(\mathbf{P}) \leq 2\min\{K, \kappa(\mathbf{P})\}^2$.

If eigenvalues of $\mathbf{P}$ can be divided by a constant number of bins where in each bin the eigenvalues are of the same order, for each bin, there will be at most a constant of intervals defined in Definition 5.1, or the eigenvalues can not be of the same order. In that case, λ_{s_k}, λ_{e_k} and g_k are of the same order and $I^+ + I^-$ is a constant, so $\psi(\mathbf{P}) = O(1)$. □

III.2 Proof of Lemma 5.2

Proof. Since $\mathbf{G}_{\mathbf{A}}(z) - \mathbf{G}_{\mathbf{P}}(z) = \mathbf{G}_{\mathbf{P}}(z)(\mathbf{P}-\mathbf{A})\mathbf{G}_{\mathbf{A}}(z)$,

$$\mathbf{G}_{\mathbf{A}}(z) - \mathbf{G}_{\mathbf{P}}(z) = \left(\mathbf{M}_z - \frac{\mathbf{I}}{z}\right)(\mathbf{P}-\mathbf{A})(\mathbf{G}_{\mathbf{A}}(z) - \mathbf{G}_{\mathbf{P}}(z)) + \mathbf{G}_{\mathbf{P}}(z)(\mathbf{P}-\mathbf{A})\mathbf{G}_{\mathbf{P}}(z)$$

Bringing $\frac{\mathbf{A}-\mathbf{P}}{z}(\mathbf{G}_{\mathbf{A}}(z) - \mathbf{G}_{\mathbf{P}}(z))$ to the LHS, and using the definition of the resolvent of $\mathbf{A}-\mathbf{P}$ we get:

$$\mathbf{G}_{\mathbf{A}}(z) - \mathbf{G}_{\mathbf{P}}(z) = z\mathbf{G}_{\mathbf{A}-\mathbf{P}}(z)\left(\underbrace{\mathbf{M}_z(\mathbf{A}-\mathbf{P})(\mathbf{G}_{\mathbf{A}}(z) - \mathbf{G}_{\mathbf{P}}(z))}_{R_0} + \underbrace{\mathbf{G}_{\mathbf{P}}(z)(\mathbf{A}-\mathbf{P})\mathbf{G}_{\mathbf{P}}(z)}_{R}\right) \tag{35}$$

As it turns out, each of the rows of zR either have small Frobenius norm or they disappear when combined with $z\mathbf{G}_{\mathbf{A}-\mathbf{P}}(z)$ post integration. We will show this step by step. Note that, using Eq (9), R in the above equation can be decomposed as:

$$\mathbf{G}_{\mathbf{P}}(z)(\mathbf{A}-\mathbf{P})\mathbf{G}_{\mathbf{P}}(z) = \underbrace{\frac{\mathbf{A}-\mathbf{P}}{z^2}}_{R_1} + \underbrace{\mathbf{M}_z(\mathbf{A}-\mathbf{P})\mathbf{G}_{\mathbf{P}}(z)}_{R_2} - \underbrace{\frac{\mathbf{A}-\mathbf{P}}{z}\mathbf{M}_z}_{R_3}$$

Next, we show that $z\mathbf{G}_{\mathbf{A}-\mathbf{P}}(z)R_1$ disappears upon integration. Since by construction $\forall z \in \mathcal{C}_k, \forall k$, $|z| \geq a_k > \|\mathbf{A}-\mathbf{P}\|$, none of the contours contain zero, Eq (21) immediately gives:

$$\frac{1}{2\pi\sqrt{-1}}\oint_{\mathcal{C}_k} z\mathbf{G}_{\mathbf{A}-\mathbf{P}}(z)R_1 dz = -\sum_{t\geq 1}\oint_{\mathcal{C}_k}\frac{1}{z}\left(\frac{\mathbf{A}-\mathbf{P}}{z}\right)^t dz = 0 \tag{36}$$

Thus Eqs (12), (35) and (36) give us,

$$\mathbf{e}_x^T(\mathbf{V}_k\mathbf{V}_k^T - \hat{\mathbf{V}}_k\hat{\mathbf{V}}_k^T) = -\frac{1}{2\pi\sqrt{-1}}\oint_{\mathcal{C}_k}\mathbf{e}_x^T\mathbf{G}_{\mathbf{A}-\mathbf{P}}(z)z(R_0 + R_2 - R_3)dz$$
$$\|\mathbf{e}_x^T(\mathbf{V}_k\mathbf{V}_k^T - \hat{\mathbf{V}}_k\hat{\mathbf{V}}_k^T)\| \leq \frac{b_k - a_k + 2\gamma_k}{\pi}\max_{z\in\mathcal{C}_k}\|\mathbf{e}_x^T\mathbf{G}_{\mathbf{A}-\mathbf{P}}(z)z(R_0 + R_2 - R_3)\|$$

Now we bound each part individually.

$$\begin{aligned}\|\mathbf{e}_x^T \mathbf{G}_{\mathbf{A}-\mathbf{P}}(z) z(R_0+R_2)\| &= \|\mathbf{e}_x^T \mathbf{G}_{\mathbf{A}-\mathbf{P}}(z) z\mathbf{M}_z(\mathbf{A}-\mathbf{P})\mathbf{G}_{\mathbf{A}}(z)\| \\ &\leq \|\mathbf{e}_x^T \mathbf{G}_{\mathbf{A}-\mathbf{P}}(z) z\mathbf{M}_z\|_F \|\mathbf{A}-\mathbf{P}\| \|\mathbf{G}_{\mathbf{A}}(z)\| \\ &\overset{(i)}{\leq} |z| \|\mathbf{G}_{\mathbf{A}}(z)\| \|\mathbf{A}-\mathbf{P}\| \|\mathbf{E}_z\| \|\mathbf{e}_x^T \mathbf{G}_{\mathbf{A}-\mathbf{P}}(z)\mathbf{V}\| =: P_1(z)\end{aligned}$$

Step (i) uses Eq (8). Finally we also have:

$$\|\mathbf{e}_x^T z\mathbf{G}_{\mathbf{A}-\mathbf{P}}(z) R_3\| \leq \|\mathbf{e}_x^T \mathbf{G}_{\mathbf{A}-\mathbf{P}}(z)(\mathbf{A}-\mathbf{P})\mathbf{V}\| \|\mathbf{E}_z\| =: P_2(z)$$

Thus, the statement of the lemma follows. □

III.3 Proof of Lemma 5.3

Proof. For ease of notation we will first prove this for one population eigenvector $\mathbf{v}$. Recall that $\mathbf{H} := (\mathbf{A}-\mathbf{P})/\sqrt{n\rho}$. Let $X_j = (\mathbf{A}_{ij} - \mathbf{P}_{ij})v_j$, where v_j is the j^{th} component of $\mathbf{v}$. We have $|X_j| \leq \|\mathbf{v}\|_\infty =: M$. Since $\mathbf{\Theta}$ is assumed to be fixed for this lemma, $\mathbf{P}$ is fixed and X_j are mean zero independent random variables. Also note that, since $\|\mathbf{v}\| = 1$ and $\mathbf{P}_{ij} \leq \rho$,

$$\sum_j \mathrm{var}(X_j) = \sum_j \mathrm{E}\left[(\mathbf{A}_{ij} - \mathbf{P}_{ij})^2 v_j^2\right] = \sum_j \mathbf{P}_{ij}(1-\mathbf{P}_{ij})v_j^2 \leq \rho$$

An application of Bernstein's inequality gives us:

$$\mathrm{P}\left(|\sum_j X_j| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2(\sum_j \mathrm{var}(X_j) + tM/3)}\right) =: 2\exp(-A)$$

First note that the RHS of the above equation is a decreasing function of t. We set $t = 4\max(M, \sqrt{\rho})\log n$. Consider the following two cases:

Case 1: $M > \sqrt{\rho}$: We have $t = 4M\log n$. Hence,

$$\exp(-A) \leq \exp\left(-\frac{16M^2\log^2 n}{2\rho + 8/3M^2\log n}\right) \leq \exp\left(-\frac{16M^2\log^2 n}{2M^2 + 8/3M^2\log n}\right) \leq \frac{1}{n^3}$$

Case 2: $M \leq \sqrt{\rho}$: We have $t = 4\sqrt{\rho}\log n$. Hence,

$$\exp(-A) \leq \exp\left(-\frac{16\rho\log^2 n}{2\rho + 8/3M\sqrt{\rho}\log n}\right) \leq \exp\left(-\frac{16\rho\log^2 n}{2\rho + 8/3\rho\log n}\right) \leq \frac{1}{n^3}$$

Applying this to all K population eigenvectors we have:

$$\mathrm{P}\left(\exists k \in [K], |\mathbf{e}_i^T(\mathbf{A}-\mathbf{P})\mathbf{v}_k| \geq 4\max(\|\mathbf{v}_k\|_\infty, \sqrt{\rho})\log n\right) \leq \frac{2K}{n^3} \tag{37}$$

Recall from Lemma 3.2 that, $\forall k \in [K], \|\mathbf{v}_k\|_\infty \leq \max_i \|\mathbf{e}_i^T\mathbf{V}\| \leq \sqrt{\frac{1}{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}$. Then if Assumption 3.1 is satisfied, we have, $\|\mathbf{v}_k\|_\infty \leq \sqrt{\rho}$, $\forall k \in [K]$. So from Eq. (37),

$$\mathrm{P}\left(\exists k \in [K], |\mathbf{e}_i^T(\mathbf{A}-\mathbf{P})\mathbf{v}_k| \geq 4\sqrt{\rho\log^2 n}\right) \leq \frac{2K}{n^3}.$$

Note that $\forall k \in [K]$, $\|\mathbf{v}_k\|_\infty \geq 1/\sqrt{n}$, then,

$$\mathrm{P}\left(\exists k \in [K], |\mathbf{e}_i^T\mathbf{H}\mathbf{v}_k| \geq 4\log n\|\mathbf{v}_k\|_\infty\right) \leq \mathrm{P}\left(\exists k \in [K], |\mathbf{e}_i^T\mathbf{H}\mathbf{v}_k| \geq 4\sqrt{\frac{\log^2 n}{n}}\right) = O\left(\frac{K}{n^3}\right).$$

□

III.4 Proof of Lemma 5.4

Proof of Lemma 5.4. For $t \geq 2$, we claim that this result follows via straightforward modifications of the proof of Lemma 7.10 in [6], where the main two elements are:

1. $\mathrm{E}[|\mathbf{H}_{ij}|^m] \leq \frac{1}{n}$ for $m \geq 2$. Note that for our setting, Assumption 3.1 implies that $n\rho \geq 1$. Hence $|\mathbf{H}_{ij}| \leq 1$, and hence

$$\mathrm{E}[|\mathbf{H}_{ij}|^m] \leq \mathrm{E}[|\mathbf{H}_{ij}|^2] \leq \frac{\mathbf{P}_{ij}}{n\rho} \leq \frac{1}{n}$$

2. The authors use a higher order Markov inequality. This inequality upper bounds the number of terms that are non-zero in the summand via a multigraph construction for path counting. Then these non-zero elements are bounded by their absolute value and hence, even though $\mathbf{v}$ does not equal $\mathbf{e}$, just the fact that it is fixed and hence independent of $\mathbf{H}_{ij}$, is enough to apply the proof directly to get the required result.

Using an almost identical argument as [6], we have:

$$\mathrm{E}[|\mathbf{e}_i^T\mathbf{H}^t\mathbf{v}|^p] \leq (tp)^{tp}\|\mathbf{v}\|_\infty^p$$

Now a higher order Markov inequality, with $p = (\log n)^\xi/2t$ gives:

$$\begin{aligned}\mathrm{P}\left(|\mathbf{e}_i^T\mathbf{H}^t\mathbf{v}| \geq (\log n)^{t\xi}\|\mathbf{v}\|_\infty\right) &\leq \frac{(tp)^{tp}\|\mathbf{v}\|_\infty^p}{(\log n)^{tp\xi}\|\mathbf{v}\|_\infty^p} = \frac{1}{\sqrt{2}^{(\log n)^\xi}}\\ &= \exp(-(\log n)^\xi \log\sqrt{2}) \leq \exp(-(\log n)^\xi/3)\end{aligned}$$

□

IV Comparison with [7] on row-wise deviation of eigenspace

Here we give a row-wise error bound for eigenspace using Abbe et al. [7]'s result.

Lemma IV.1. *For $\mathbf{P} = \mathbf{V}\mathbf{E}\mathbf{V}^T$ and $\mathbf{A} = \hat{\mathbf{V}}\hat{\mathbf{E}}\hat{\mathbf{V}}^T$ as $\mathbf{P}$ and $\mathbf{A}$'s top-K eigen-decomposition respectively, we have*

$$\begin{aligned}\|\hat{\mathbf{V}}sgn(\hat{\mathbf{V}}^T\mathbf{V}) - \mathbf{A}\mathbf{V}\mathbf{E}^{-1}\|_{2\to\infty} &= O_P\left(\frac{(\kappa(\mathbf{\Theta}^T\mathbf{\Theta}))^2K\sqrt{n}}{\sqrt{\rho}(\lambda^*(\mathbf{B}))^3(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right)\\ \|\hat{\mathbf{V}}sgn(\hat{\mathbf{V}}^T\mathbf{V}) - \mathbf{V}\|_{2\to\infty} &= O_P\left(\max\left(\frac{(\kappa(\mathbf{\Theta}^T\mathbf{\Theta}))^2K\sqrt{n}}{\sqrt{\rho}(\lambda^*(\mathbf{B}))^3(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}, \frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}\right)\right),\end{aligned}$$

where $\|\mathbf{U}\|_{2\to\infty} = \max_i \|\mathbf{e}_i^T\mathbf{U}\|$ denotes the maximum row norm of a matrix $\mathbf{U}$, and $sgn(\hat{\mathbf{V}}^T\mathbf{V})$ is the matrix sign function

$$sgn(\hat{\mathbf{V}}^T\mathbf{V}) = \mathbf{U}_1\mathbf{U}_2^T, \quad \text{SVD of } \hat{\mathbf{V}}^T\mathbf{V} \text{ is } \hat{\mathbf{V}}^T\mathbf{V} = \mathbf{U}_1\mathbf{\Sigma}\mathbf{U}_{\mathbf{2}}^{\mathbf{T}}.$$

Proof. First from Assumption A3 of [7] we have $c\sqrt{\rho n} \leq \gamma\Delta^*$ for some constant c. Δ^* is the eigen-gap, which is $\lambda^*(\mathbf{P}) \geq \rho\lambda^*(\mathbf{B})\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})$ from Lemma II.4. This requires

$$\gamma \geq \frac{c\sqrt{n}}{\sqrt{\rho}\lambda^*(\mathbf{B})\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}. \tag{38}$$

Using Eqs (13) of Corollary 2.1 in [7], we have,

$$\begin{aligned}\|\hat{\mathbf{V}}sgn(\hat{\mathbf{V}}^T\mathbf{V}) - \mathbf{AVE}^{-1}\|_{2\to\infty} &\le \kappa(\kappa+\varphi(1))(\gamma+\varphi(\gamma))\|\mathbf{V}\|_{2\to\infty} \\ \text{(Lemma 3.2)} \quad &\le \kappa(\kappa+\varphi(1))(\gamma+\varphi(\gamma))\frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}.\end{aligned}$$

κ is the condition number of $\mathbf{P}$ which is upper bounded by:

$$\kappa \le \kappa(\mathbf{\Theta}^T\mathbf{\Theta})\kappa(\mathbf{B}) \le \kappa(\mathbf{\Theta}^T\mathbf{\Theta})\frac{\|\mathbf{B}\|}{\lambda^*(\mathbf{B})} \le \kappa(\mathbf{\Theta}^T\mathbf{\Theta})\frac{\sqrt{K}}{\lambda^*(\mathbf{B})}.$$

Since $\varphi(x)$ is typically bounded by a constant and $\varphi(x)/x$ non-increasing, we have,

$$\begin{aligned}\|\hat{\mathbf{V}}sgn(\hat{\mathbf{V}}^T\mathbf{V}) - \mathbf{AVE}^{-1}\|_{2\to\infty} &= O_P\left(\kappa^2\gamma\frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}\right) \\ &= O_P\left(\frac{(\kappa(\mathbf{P}))^2\sqrt{n}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right) \\ &= O_P\left(\frac{(\kappa(\mathbf{\Theta}^T\mathbf{\Theta}))^2K}{(\lambda^*(\mathbf{B}))^2}\frac{\sqrt{n}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right) \\ &= O_P\left(\frac{(\kappa(\mathbf{\Theta}^T\mathbf{\Theta}))^2K\sqrt{n}}{\sqrt{\rho}(\lambda^*(\mathbf{B}))^3(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right). \quad (39)\end{aligned}$$

Furthermore, using Eqs (14) of Corollary 2.1 in [7], we have,

$$\begin{aligned}\|\hat{\mathbf{V}}sgn(\hat{\mathbf{V}}^T\mathbf{V}) - \mathbf{V}\|_{2\to\infty} &\le \|\hat{\mathbf{V}}sgn(\hat{\mathbf{V}}^T\mathbf{V}) - \mathbf{AVE}^{-1}\|_{2\to\infty} + \varphi(1)\|\mathbf{V}\|_{2\to\infty} \\ &= O_P\left(\max\left(\frac{(\kappa(\mathbf{P}))^2\sqrt{n}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}, \frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}\right)\right) \\ &= O_P\left(\max\left(\frac{(\kappa(\mathbf{\Theta}^T\mathbf{\Theta}))^2K\sqrt{n}}{\sqrt{\rho}(\lambda^*(\mathbf{B}))^3(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}, \frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}\right)\right) \quad (40)\end{aligned}$$

□

Remark IV.1. *Our bound in Theorem 3.1 has better dependency on* $\lambda^*(\mathbf{B})$ *comparing to Eq* (39) *when* $\lambda^*(\mathbf{B})$ *goes to 0 (*$\kappa(\mathbf{P})$ *goes to infinity). If* $\kappa(\mathbf{P}) = \Theta(1)$ *or* $\lambda^*(\mathbf{B}) = \Theta(1)$ *or* $K = \Theta(1)$*, the bound in Theorem 3.1 is comparable or better than that in Eq* (39)*.*

However, in comparison to their eigenvector deviation bound from Eq (40)*, we have a tighter bound by an order of* $1/\sqrt{n\rho}$ *when* $K = \Theta(1)$*,* $\lambda^*(\mathbf{B}) = \Theta(1)$*,* $\kappa(\mathbf{P}) = \Theta(1)$ *and* $\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}) = \Omega(n)$*. As a matter of fact, when* $\boldsymbol{\theta}_i \sim \text{Dirichlet}(\boldsymbol{\alpha})$ *and* $\alpha_{\max} \leq C\alpha_{\min}$ *for some constant* $C \geq 1$*, we have* $\nu = \alpha_0/\alpha_{\min} = \Theta(K)$ *and by Lemma 3.6, with high probability* $\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}) = \Omega(n/\nu) = \Omega(n)$ *when* $K = \Theta(1)$*.*

V Comparison with [8] on row-wise deviation of eigenspace

Here we give a row-wise error bound for eigenspace applying Cape et al. [8]'s result.

Lemma V.1. *For* $\mathbf{P} = \mathbf{V}\mathbf{E}\mathbf{V}^T$ *and* $\mathbf{A} = \hat{\mathbf{V}}\hat{\mathbf{E}}\hat{\mathbf{V}}^T$ *as* $\mathbf{P}$ *and* $\mathbf{A}$*'s top-*K *eigen-decomposition respectively, if* $\lambda^*(\mathbf{P}) \geq \|\mathbf{A}-\mathbf{P}\|_\infty$*, then there exists an orthognal matrix* $\mathbf{W}_\mathbf{V} \in \mathbb{R}^{K\times K}$ *such that*

$$\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{W}_\mathbf{V}\|_{2\to\infty} = O_P\left(\frac{n}{\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right). \tag{41}$$

Proof. From Lemma 3.2, $\|\mathbf{V}\|_{2\to\infty} \leq \frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}$. By applying Theorem 4.2 of [8], if $\lambda^*(\mathbf{P}) \geq \|\mathbf{A}-\mathbf{P}\|_\infty$,

$$\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{W}_\mathbf{V}\|_{2\to\infty} \leq 14\left(\frac{\|\mathbf{A}-\mathbf{P}\|_\infty}{\lambda^*(\mathbf{P})}\right)\|\mathbf{V}\|_{2\to\infty},$$

where $\|\mathbf{A}-\mathbf{P}\|_\infty = \max_i \sum_j |\mathbf{A}_{ij}-\mathbf{P}_{ij}|$. Note that as $\mathrm{E}[\sum_j |\mathbf{A}_{ij}-\mathbf{P}_{ij}|] = \sum_j \mathrm{E}[|\mathbf{A}_{ij}-\mathbf{P}_{ij}|] = \sum_j [\mathbf{P}_{ij}(1-\mathbf{P}_{ij}) + (1-\mathbf{P}_{ij})\mathbf{P}_{ij}] = O(\rho n)$. By applying Chernoff bound, it can be shown that for all $i \in [n]$, $\sum_j |\mathbf{A}_{ij}-\mathbf{P}_{ij}| = O_P(\rho n)$. Then $\|\mathbf{A}-\mathbf{P}\|_\infty = \max_i \sum_j |\mathbf{A}_{ij}-\mathbf{P}_{ij}| = O(\rho n)$ with high probability. From Lemma II.4, we have $\lambda^*(\mathbf{P}) \geq \rho\lambda^*(\mathbf{B})\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})$. Then we have,

$$\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{W}_\mathbf{V}\|_{2\to\infty} = O_P\left(\frac{n}{\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right).$$

□

Remark V.1. *Our bound in Theorem 3.1 is tighter by an order of* $1/\sqrt{n\rho}$ *comparing to Eq* (41) *when* $K = \Theta(1)$.

Note that Lemma V.1 is a direct application of the perturbation bound in [8] to the MMSB model. If we use the more careful analysis of the authors for the ρ-correlated SBM graphs, together with our Lemma 5.3, we can get a better bound as in the following Lemma.

Lemma V.2. *For* $\mathbf{P} = \mathbf{V}\mathbf{E}\mathbf{V}^T$ *and* $\mathbf{A} = \hat{\mathbf{V}}\hat{\mathbf{E}}\hat{\mathbf{V}}^T$ *as* $\mathbf{P}$ *and* $\mathbf{A}$*'s top-K eigen-decomposition respectively, and* $\mathbf{W}_{\mathbf{V}} = sgn(\mathbf{V}^T\hat{\mathbf{V}}) \in \mathbb{R}^{K\times K}$*, we have*

$$\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{W}_{\mathbf{V}}\|_{2\to\infty} = O_P\left(\frac{n}{\rho(\lambda^*(\mathbf{B}))^2(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^2}\right), \tag{42}$$

when $\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}) = \Omega(K)$ *and* $\rho(\lambda^*(\mathbf{B}))^2 \le \frac{n}{K\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}$

Proof. Using Corallary 3.3 and Proposition 6.5 of [8], noting that $(\mathbf{I} - \mathbf{V}\mathbf{V}^T)\mathbf{P} = \mathbf{0}$, we have the following decomposition and bound:

$$\begin{aligned}\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{W}_{\mathbf{V}}\|_{2\to\infty} \le& \|(\mathbf{I} - \mathbf{V}\mathbf{V}^T)(\mathbf{A} - \mathbf{P})\mathbf{V}\mathbf{W}_{\mathbf{V}}\hat{\mathbf{E}}^{-1}\|_{2\to\infty} \\ &+ \|(\mathbf{I} - \mathbf{V}\mathbf{V}^T)(\mathbf{A} - \mathbf{P})(\hat{\mathbf{V}} - \mathbf{V}\mathbf{W}_{\mathbf{V}})\hat{\mathbf{E}}^{-1}\|_{2\to\infty} \\ &+ \|\mathbf{V}(\mathbf{V}^T\hat{\mathbf{V}} - \mathbf{W}_{\mathbf{V}})\|_{2\to\infty} \\ \le& \|(\mathbf{A} - \mathbf{P})\mathbf{V}\|_{2\to\infty}\|\hat{\mathbf{E}}^{-1}\| + \|\mathbf{V}\|_{2\to\infty}\|\mathbf{V}^T\|\|(\mathbf{A} - \mathbf{P})\mathbf{V}\|\|\hat{\mathbf{E}}^{-1}\| \\ &+ \|\mathbf{I} - \mathbf{V}\mathbf{V}^T\|_{2\to\infty}\|\mathbf{A} - \mathbf{P}\|\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{W}_{\mathbf{V}}\|\|\hat{\mathbf{E}}^{-1}\| \\ &+ \|\mathbf{V}\|_{2\to\infty}\|\mathbf{V}^T\hat{\mathbf{V}} - \mathbf{W}_{\mathbf{V}}\|.\end{aligned}$$

By Lemma 5.3, we have $\|(\mathbf{A} - \mathbf{P})\mathbf{V}\|_{2\to\infty} = O_P(\sqrt{Kn\rho})\|\mathbf{V}\|_{2\to\infty}$, so

$$\|(\mathbf{A} - \mathbf{P})\mathbf{V}\| \le \sqrt{K}\|(\mathbf{A} - \mathbf{P})\mathbf{V}\|_{2\to\infty} = O_P(K\sqrt{n\rho})\|\mathbf{V}\|_{2\to\infty}.$$

We also have $\|\mathbf{A} - \mathbf{P}\| = O_P(\sqrt{n\rho})$, $\|\hat{\mathbf{E}}^{-1}\| = O_P\left(\frac{1}{\rho\lambda^*(\mathbf{B})\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\right)$ from results in Sec II. By Lemmas 6.7 and 6.8 of [8] and using $\lambda^*(\mathbf{P}) \ge \rho\lambda^*(\mathbf{B})\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})$ from Lemma II.4, we have $\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{W}_{\mathbf{V}}\| = O_P\left(\frac{\sqrt{n}}{\sqrt{\rho}\lambda^*(\mathbf{B})\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\right)$ and $\|\mathbf{V}^T\hat{\mathbf{V}} - \mathbf{W}_{\mathbf{V}}\| = O_P\left(\frac{n}{\rho(\lambda^*(\mathbf{B}))^2(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^2}\right)$. From

Lemma 3.2, $\|\mathbf{V}\|_{2\to\infty} \le \frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}$. It is also easy to get $\|\mathbf{I}-\mathbf{V}\mathbf{V}^T\|_{2\to\infty} \le 2$. Putting these bounds together, we have:

$$\begin{aligned}\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{W}_{\mathbf{V}}\|_{2\to\infty} &= O_P\left(\max\left(\frac{n}{\rho(\lambda^*(\mathbf{B}))^2(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^2}, \frac{\sqrt{Kn}\left(1+\sqrt{K/\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\right)}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right)\right) \\ &= O_P\left(\frac{n}{\rho(\lambda^*(\mathbf{B}))^2(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^2}\right),\end{aligned}$$

when $\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}) = \Omega(K)$ and $\rho(\lambda^*(\mathbf{B}))^2 \le \frac{n}{K\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}$. □

Remark V.2. *Our bound in Theorem 3.1 has better dependency on $\lambda^*(\mathbf{B})$ comparing to Eq (42) when $\lambda^*(\mathbf{B})$ goes to 0 ($\kappa(\mathbf{P})$ goes to infinity). When $\lambda^*(\mathbf{B}) = \Theta(1)$, $\kappa(\mathbf{P}) = \Theta(1)$, and $\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}) = \Omega(n/K)$, our bound in Theorem 3.1 is tighter by a factor of $\sqrt{\rho}$ comparing to Eq (42). As discussed in Sec IV, when $\boldsymbol{\theta}_i \sim \text{Dirichlet}(\boldsymbol{\alpha})$ and $\alpha_{\max} \le C\alpha_{\min}$ for some constant $C \ge 1$, with high probability $\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}) = \Omega(n/K) \gg K$, and $\rho(\lambda^*(\mathbf{B}))^2 \le \frac{n}{K\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})} = \Theta(1)$ is corresponding to the interesting regime when ρ or $\lambda^*(\mathbf{B})$ is small.*

VI Row-wise eigenspace concentration for general low rank matrix

Note that although our focus of this paper is on MMSB, Theorem 3.1 can be easily extended to handle any low rank matrix. The proof is almost identical to that of Theorem 3.1, just instead of assuming Assumption 3.1 is satisfied, we have some general conditions. The new events should be:

$$\begin{aligned}
&\mathcal{E}'_1 := \left\{\max_i \|\mathbf{v}_i\|_\infty \le \sqrt{\rho}\right\} && \mathrm{P}(\bar{\mathcal{E}'}_1) \le \delta_1 \\
&\mathcal{E}'_2 := \left\{\lambda^*(\mathbf{P}) \ge 4\sqrt{n\rho}(\log n)^\xi\right\} && \mathrm{P}(\bar{\mathcal{E}'}_2) \le \delta_2 \\
&\mathcal{E}' := \{\|\mathbf{A}-\mathbf{P}\| \le C\sqrt{n\rho}\} && \mathrm{P}(\bar{\mathcal{E}'}) \le n^{-3} \\
&\mathcal{E}_1 := \left\{\left|\mathbf{e}_i^T\mathbf{H}\mathbf{v}_k\right| \le 4\log n\|\mathbf{v}_k\|_\infty, \forall k \in [K]\right\} && \mathrm{P}(\bar{\mathcal{E}}_1) \le O\left(K/n^3\right) + \delta_1 \\
&\mathcal{E}_t := \left\{\left|\mathbf{e}_i^T\mathbf{H}^t\mathbf{v}_k\right| \le (\log n)^{t\xi}\|\mathbf{v}_k\|_\infty, \forall k \in [K]\right\} && \mathrm{P}(\bar{\mathcal{E}}_t) \le K\exp(-(\log n)^\xi/3), 1 < t \le \log n
\end{aligned} \tag{43}$$

If we use the new events in Eq (43) in the proof, we can get the following Theorem:

Theorem VI.1. *Suppose* $\mathbf{P}$ *has rank* K*,* $\max_{i,j}\mathbf{P}_{ij} \le \rho$*. Let* $\mathbf{A}_{ij} = \mathbf{A}_{ji} \sim \mathrm{Ber}(\mathbf{P}_{ij})$*,* $\mathbf{V}$ *and* $\hat{\mathbf{V}}$ *are* $\mathbf{P}$ *and* $\mathbf{A}$*'s* K *leading eigenvectors respectively. If* $\mathrm{P}(\max_i \|\mathbf{v}_i\|_\infty > \sqrt{\rho}) \le \delta_1$*, and for some constant* $\xi > 1$*,* $\rho n = \Omega((\log n)^{2\xi})$ *and* $\mathrm{P}(\lambda^*(\mathbf{P}) < 4\sqrt{n\rho}(\log n)^\xi) < \delta_2$*, then with probability at least* $1-\delta_1-\delta_2-O(Kn^{-2})$*,*

$$\max_{i\in[n]} \|\mathbf{e}_i^T(\hat{\mathbf{V}}\hat{\mathbf{V}}^T - \mathbf{V}\mathbf{V}^T)\| = O\left(\frac{\psi(\mathbf{P})\sqrt{Kn\rho}}{\lambda^*(\mathbf{P})}\right)\left((1+(\log n)^\xi)\max_i \|\mathbf{v}_i\|_\infty + 2n^{-2\xi}\right).$$

Remark VI.1. *For MMSB, it is easy to check that the condition* $\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}) \ge 1/\rho$ *in Assumption 3.1 is only used in the proof of Lemma 5.3 in Sec III.3 to show* $\max_i \|\mathbf{v}_i\|_\infty \le \sqrt{\rho}$*, so conditioned on* $\mathcal{E}'_1$ *and* $\mathcal{E}'_2$*, the proof goes through. If we plug in the upper bound of* $\max_i \|\mathbf{v}_i\|_\infty$ *from Lemma 3.2 and lower bound of* $\lambda^*(\mathbf{P})$ *in Lemma II.4, we can get the bound in Theorem 3.1 using Theorem VI.1.*

VII Consistency of estimated quantities

Proof of Lemma 3.4. To see that the pruning algorithm returns identical nodes (up-to ties) is straightforward. This is because the pruning algorithm proceeds by calculating Euclidean distances between pairs of nodes for nearest neighbor computation. We have

$$\|\hat{\mathbf{V}}\hat{\mathbf{V}}^T(\mathbf{e}_i - \mathbf{e}_j)\|^2 = (\mathbf{e}_i - \mathbf{e}_j)^T\hat{\mathbf{V}}\hat{\mathbf{V}}^T(\mathbf{e}_i - \mathbf{e}_j) = \|\hat{\mathbf{V}}^T(\mathbf{e}_i - \mathbf{e}_j)\|^2$$

Thus the pairwise distances between columns of $\hat{\mathbf{V}}\hat{\mathbf{V}}^T$ are the same as that between columns of $\hat{\mathbf{V}}^T$. As for the SPA algorithm, we prove the claim by induction.

Base case: For step $k = 1$, as

$$\|\hat{\mathbf{V}}\hat{\mathbf{V}}^T\mathbf{e}_i\|^2 = \mathbf{e}_i^T\hat{\mathbf{V}}\hat{\mathbf{V}}^T\hat{\mathbf{V}}\hat{\mathbf{V}}^T\mathbf{e}_i = \mathbf{e}_i^T\hat{\mathbf{V}}\hat{\mathbf{V}}^T\mathbf{e}_i = \|\hat{\mathbf{V}}^T\mathbf{e}_i\|^2,$$

picking max norm will give the same index, denoted as k_1.

Now for $\hat{\mathbf{V}}^T$, the vector whose projection is removed is $\hat{\mathbf{V}}^T\mathbf{e}_{k_1}$, and the normalized vector is $\mathbf{u} = \hat{\mathbf{V}}^T\mathbf{e}_{k_1}/\|\hat{\mathbf{V}}^T\mathbf{e}_{k_1}\|$, then for $\hat{\mathbf{V}}\hat{\mathbf{V}}^T$, the vector whose projection is removed is $\hat{\mathbf{V}}\hat{\mathbf{V}}^T\mathbf{e}_{k_1}$ and its normalized vector is $\mathbf{u}_1 = \hat{\mathbf{V}}\hat{\mathbf{V}}^T\mathbf{e}_{k_1}/\|\hat{\mathbf{V}}\hat{\mathbf{V}}^T\mathbf{e}_{k_1}\| = \hat{\mathbf{V}}\hat{\mathbf{V}}^T\mathbf{e}_{k_1}/\|\hat{\mathbf{V}}^T\mathbf{e}_{k_1}\| = \hat{\mathbf{V}}\mathbf{u}$.

Now

$$\begin{aligned}\|(\mathbf{I} - \mathbf{u}_1\mathbf{u}_1^T)\hat{\mathbf{V}}\hat{\mathbf{V}}^T\mathbf{e}_i\|^2 &= \|(\mathbf{I} - \hat{\mathbf{V}}\mathbf{u}\mathbf{u}^T\mathbf{V}^T)\hat{\mathbf{V}}\hat{\mathbf{V}}^T\mathbf{e}_i\|^2 \\ &= \|\hat{\mathbf{V}}(\mathbf{I} - \mathbf{u}\mathbf{u}^T)\hat{\mathbf{V}}^T\mathbf{e}_i\|^2 = \|(\mathbf{I} - \mathbf{u}\mathbf{u}^T)\hat{\mathbf{V}}^T\mathbf{e}_i\|^2,\end{aligned}$$

then for step $k = 2$, picking max norm will also give the same index.

Induction: Suppose for first $k - 1 \in [K - 1]$ steps SPA on $\hat{\mathbf{V}}^T$ and on $\hat{\mathbf{V}}\hat{\mathbf{V}}^T$ will give the same indices as S_{k-1}, then for the k-th step, we are removing the projections of the $k - 1$ columns in S_{k-1} selected before, now denote the singular value decomposition $(\hat{\mathbf{V}}_{S_{k-1}})^T = \mathbf{U}\mathbf{S}\mathbf{H}^T$, then the projection matrix on columns of $(\hat{\mathbf{V}}_{S_{k-1}})^T$ is $\mathbf{U}\mathbf{U}^T$. Also note that $\hat{\mathbf{V}}(\hat{\mathbf{V}}_{S_{k-1}})^T = (\hat{\mathbf{V}}\mathbf{U})\mathbf{S}\mathbf{H}^T$, it is easy to check that this is singular value decomposition of $\hat{\mathbf{V}}(\hat{\mathbf{V}}_{S_{k-1}})^T$, and the projection matrix on columns of $\hat{\mathbf{V}}(\hat{\mathbf{V}}_{S_{k-1}})^T$ is $\hat{\mathbf{V}}\mathbf{U}(\hat{\mathbf{V}}\mathbf{U})^T = \hat{\mathbf{V}}\mathbf{U}\mathbf{U}^T\hat{\mathbf{V}}^T$. Now the norm we need to pick from for SPA on $\hat{\mathbf{V}}\hat{\mathbf{V}}^T$ is

$$\|(\mathbf{I} - \hat{\mathbf{V}}\mathbf{U}\mathbf{U}^T\hat{\mathbf{V}}^T)\hat{\mathbf{V}}\hat{\mathbf{V}}^T\mathbf{e}_i\| = \|\hat{\mathbf{V}}(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\hat{\mathbf{V}}^T\mathbf{e}_i\| = \|(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\hat{\mathbf{V}}^T\mathbf{e}_i\|,$$

so the norms to pick for SPA on $\hat{\mathbf{V}}^T$ and on $\hat{\mathbf{V}}\hat{\mathbf{V}}^T$ will still be same and picking max norm will also give the same index. □

Lemma VII.1. *(Theorem 3 of Gillis et al. [9]). Let* $\mathbf{M}' = \mathbf{M} + \mathbf{N} = \mathbf{W}\mathbf{H} + \mathbf{N} \in \mathbb{R}^{m\times n}$, *where* $\mathbf{M} = \mathbf{W}\mathbf{H} = \mathbf{W}[\mathbf{I}_r|\mathbf{H}']$, $\mathbf{W} \in \mathbb{R}^{m\times r}$, $\mathbf{H} \in \mathbf{R}_+^{r\times n}$ *and* $\sum_{k=1}^r \mathbf{H}'_{kj} \leq 1$, $\forall j$ *and* $r \geq 2$.

Let $K(\mathbf{W}) = \max_i \|\mathbf{W}(:,i)\|_2$*, and* $\|\mathbf{N}(:,i)\|_2 \leq \epsilon$ *for all* i *with*

$$\epsilon < \sigma_r(\mathbf{W}) \min\left(\frac{1}{2\sqrt{r-1}}, \frac{1}{4}\right)\left(1 + 80\frac{K(\mathbf{W})^2}{\sigma_r^2(\mathbf{W})}\right)^{-1}$$

and J *be the index set of cardinality* r *extracted by SPA, where* $\sigma_r(\mathbf{W})$ *is the* r*-th singular value of* $\mathbf{W}$*. Then there exists a permutation* P *of* $\{1, 2, \cdots, r\}$ *such that*

$$\max_{1\leq j\leq r} \|\mathbf{M}'(:,J(j)) - \mathbf{W}(:,P(j))\| \leq \bar{\epsilon} = \epsilon\left(1 + 80\frac{K(\mathbf{W})^2}{\sigma_r^2(\mathbf{W})}\right).$$

Theorem VII.2. *Let* $\mathcal{S}_p$ *be the indices set returned by SPA in Algorithm 1,* $\hat{\mathbf{V}}_p = \hat{\mathbf{V}}(\mathcal{S}_p,:)$*. If Assumptions 2.1 and 3.1 are satisfied, then there exists a permutation matrix* $\mathbf{\Pi} \in \mathbb{R}^{K\times K}$ *such that*

$$\max_{1\leq j\leq K} \left\|\mathbf{e}_j^T\left(\hat{\mathbf{V}}_p - \mathbf{\Pi}^T\mathbf{V}_P(\mathbf{V}^T\hat{\mathbf{V}})\right)\right\| = O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right)$$

with probability larger than $1 - O(Kn^{-2})$*, where* $\epsilon = \tilde{O}\left(\frac{\psi(\mathbf{P})\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right)$ *is the row-wise error from Theorem 3.1, and the rows of* $\mathbf{V}_P \in \mathbb{R}^{K\times K}$ *correspond to pure nodes.*

Proof of Theorem VII.2. Note that from Lemma 2.3, $\mathbf{V} = \mathbf{\Theta}\mathbf{V}_P$. Let $\mathbf{M}' = \hat{\mathbf{V}}\hat{\mathbf{V}}^T$, $\mathbf{W} = \mathbf{V}\mathbf{V}_P^T$, $\mathbf{H} = \mathbf{\Theta}^T$, $r = K$, then for $\mathbf{M}' = \mathbf{W}\mathbf{H} + \mathbf{N}$, we have $\|\mathbf{N}(:,i)\|_2 \leq \epsilon$ uniformly with probability larger than $1 - O(Kn^{-2})$ by Theorem 3.1. W.L.O.G., let the first K rows of $\mathbf{\Theta}$ be K different pure nodes. Now use Lemma VII.1, there exists a permutation π of $\{1, 2, \cdots, K\}$ such that

$$\max_{1\leq j\leq K} \|\mathbf{M}'(:,\mathcal{S}_p(j)) - \mathbf{W}(,:\pi(j))\| = \epsilon\left(1 + 80\frac{K(\mathbf{W})^2}{\sigma_K^2(\mathbf{W})}\right) = O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right),$$

since $K(\mathbf{W}) = \max_i \|\mathbf{W}(:,i)\|_2 \leq \sigma_1(\mathbf{W})$ and $\frac{\sigma_1(\mathbf{W})}{\sigma_K(\mathbf{W})} = \kappa(\mathbf{W}) \leq \kappa(\mathbf{V}_P) = O\left(\sqrt{\kappa(\mathbf{\Theta}^T\mathbf{\Theta})}\right)$ by Lemma II.3.

So $\exists$ a permutation matrix $\mathbf{\Pi} \in \mathbb{R}^{K\times K}$ such that

$$\max_{1\leq j\leq K} \|\left(\hat{\mathbf{V}}\hat{\mathbf{V}}_p^T - \mathbf{W}\mathbf{\Pi}\right)\mathbf{e}_j\| = O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right),$$

taking transpose, it gives

$$\max_{1\le j\le K} \|\mathbf{e}_j^T \left(\hat{\mathbf{V}}_p \hat{\mathbf{V}}^T - \mathbf{\Pi}^T \mathbf{V}_P \mathbf{V}^T\right)\| = O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right),$$

and

$$\begin{aligned}\max_{1\le j\le K} \left\|\mathbf{e}_j^T \left(\hat{\mathbf{V}}_p - \mathbf{\Pi}^T \mathbf{V}_P (\mathbf{V}^T\hat{\mathbf{V}})\right)\right\| &= \max_{1\le j\le K} \left\|\mathbf{e}_j^T \left(\hat{\mathbf{V}}_p \hat{\mathbf{V}}^T - \mathbf{\Pi}^T \mathbf{V}_P \mathbf{V}^T\right)\hat{\mathbf{V}}\right\| \\ &\le \max_{1\le j\le K} \left\|\mathbf{e}_j^T \left(\hat{\mathbf{V}}_p \hat{\mathbf{V}}^T - \mathbf{\Pi}^T \mathbf{V}_P \mathbf{V}^T\right)\right\| \left\|\hat{\mathbf{V}}\right\| = O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right)\end{aligned}$$

with probability larger than $1 - O(Kn^{-2})$. The inequality follows from Proposition 5.6 of [8]. □

Lemma VII.3. *Let $\mathcal{S}_p$ be the set of of pure nodes extracted using SPACL. Let $\hat{\mathbf{V}}_p$ denote the rows of $\hat{\mathbf{V}}$ indexed by $\mathcal{S}_p$, and $\mathbf{V}_P$ denote the pure nodes of $\mathbf{V}$. Then, if Assumptions 2.1, 3.1, and 3.2 are satisfied,*

$$\max_{i\in[n]} \left\|\mathbf{e}_i^T \mathbf{V}(\mathbf{V}^T\hat{\mathbf{V}}) \left(\hat{\mathbf{V}}_p^{-1} - (\mathbf{\Pi}^T\mathbf{V}_P(\mathbf{V}^T\hat{\mathbf{V}}))^{-1}\right)\right\| = O\left(\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right)$$

with probability larger than $1 - O(Kn^{-2})$, where $\epsilon = \tilde{O}\left(\frac{\psi(\mathbf{P})\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right)$ is the row-wise error from Theorem 3.1, and rows of $\mathbf{V}_P \in \mathbb{R}^{K\times K}$ are corresponding to pure nodes.

Proof of Lemma VII.3. Define by $\mathbf{F} := \mathbf{V}^T\hat{\mathbf{V}}$, and $\tilde{\mathbf{V}}_P := \mathbf{\Pi}^T\mathbf{V}_P\mathbf{F}$, then,

$$\begin{aligned}&\left\|\mathbf{e}_i^T \mathbf{V}(\mathbf{V}^T\hat{\mathbf{V}}) \left(\hat{\mathbf{V}}_p^{-1} - (\mathbf{\Pi}^T\mathbf{V}_P(\mathbf{V}^T\hat{\mathbf{V}}))^{-1}\right)\right\| \\ &= \left\|\mathbf{e}_i^T\mathbf{V}\mathbf{F}(\hat{\mathbf{V}}_p^{-1} - \tilde{\mathbf{V}}_P^{-1})\right\| = \left\|\mathbf{e}_i^T\mathbf{V}\mathbf{F}\tilde{\mathbf{V}}_P^{-1}\left(\tilde{\mathbf{V}}_P - \hat{\mathbf{V}}_p\right)\hat{\mathbf{V}}_p^{-1}\right\| \\ &\le \left\|\mathbf{e}_i^T\mathbf{V}\mathbf{F}\mathbf{F}^{-1}\mathbf{V}_P^{-1}\mathbf{\Pi}\left(\tilde{\mathbf{V}}_P - \hat{\mathbf{V}}_p\right)\right\| \left\|\hat{\mathbf{V}}_p^{-1}\right\| = \left\|\mathbf{e}_i^T\mathbf{\Theta}\mathbf{\Pi}\left(\tilde{\mathbf{V}}_P - \hat{\mathbf{V}}_p\right)\right\| \left\|\hat{\mathbf{V}}_p^{-1}\right\| \\ &\le \max_{1\le i\le K} \left\|\mathbf{e}_i^T\left(\hat{\mathbf{V}}_p - \mathbf{\Pi}^T\mathbf{V}_P\mathbf{V}^T\hat{\mathbf{V}}\right)\right\| \left\|\hat{\mathbf{V}}_p^{-1}\right\| = O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right)\left\|\hat{\mathbf{V}}_p^{-1}\right\| \qquad (44)\end{aligned}$$

where the first inequality is true because rows of $\mathbf{\Theta\Pi}$ are still nonnegative and have unit ℓ_1 norm, and the last step follows from Theorem VII.2. Now we will prove a bound on $\|\hat{\mathbf{V}}_p^{-1}\|$. Let $\hat{\sigma}_i$ be the i^{th} singular value of $\hat{\mathbf{V}}_p$, then,

$$\|\hat{\mathbf{V}}_p^{-1}\| = \frac{1}{\hat{\sigma}_K}. \qquad (45)$$

From Lemma II.3, $\sigma_K(\mathbf{V}_P) = 1/\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}$ and $\sigma_1(\mathbf{V}_P) = 1/\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}$.

Now for using the orthogonal matrix $\hat{\mathbf{O}} \in \mathbb{R}^{K\times K}$ constructed using Definition II.1,

$$\left(\hat{\mathbf{V}}_p\hat{\mathbf{V}}^T - \mathbf{\Pi}^T\mathbf{V}_P\mathbf{V}^T\right)\hat{\mathbf{V}} = \left(\hat{\mathbf{V}}_p - \mathbf{\Pi}^T\mathbf{V}_P\hat{\mathbf{O}}\right) + \mathbf{\Pi}^T\mathbf{V}_P\left(\hat{\mathbf{O}}\hat{\mathbf{V}}^T - \mathbf{V}^T\right)\hat{\mathbf{V}},$$

then by Lemma II.3, Theorem VII.2 and Lemma II.4, we have,

$$\begin{aligned}
\|\hat{\mathbf{V}}_p - \mathbf{\Pi}^T\mathbf{V}_P\hat{\mathbf{O}}\|_F &\leq \|\hat{\mathbf{V}}_p\hat{\mathbf{V}}^T - \mathbf{\Pi}^T\mathbf{V}_P\mathbf{V}^T\|_F \cdot \|\hat{\mathbf{V}}\| + \|\mathbf{V}_P\| \cdot \|\hat{\mathbf{O}}\hat{\mathbf{V}}^T - \mathbf{V}^T\|_F \cdot \|\hat{\mathbf{V}}\| \\
&\leq \sqrt{K}\max_{1\leq j\leq K}\left\|\mathbf{e}_j^T\left(\hat{\mathbf{V}}_p - \mathbf{\Pi}^T\mathbf{V}_P(\mathbf{V}^T\hat{\mathbf{V}})\right)\right\| + \frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}\|\hat{\mathbf{O}}\hat{\mathbf{V}}^T - \mathbf{V}^T\|_F \\
&\leq O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K}\epsilon\right) + \frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}O\left(\frac{\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\right) \\
&= O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K}\epsilon\right) + O\left(\frac{\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right) \\
&= O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K}\epsilon\right) \quad \text{with probability larger than } 1 - O(Kn^{-2}).
\end{aligned} \tag{46}$$

Now, Weyl's inequality for singular values gives us:

$$\begin{aligned}
&|\hat{\sigma}_i - \sigma_i(\mathbf{V}_P)| \leq \|\hat{\mathbf{V}}_p - \mathbf{\Pi}^T\mathbf{V}_P\hat{\mathbf{O}}\| \leq \|\hat{\mathbf{V}}_p - \mathbf{\Pi}^T\mathbf{V}_P\hat{\mathbf{O}}\|_F = O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K}\epsilon\right) \\
&\hat{\sigma}_K \geq \frac{1}{\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}}\left(1 - O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\epsilon\right)\right) \\
&\hat{\sigma}_1 \leq \frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}\left(1 + O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\epsilon\right)\right).
\end{aligned} \tag{47}$$

Plugging this into Eq (45) we get:

$$\|\hat{\mathbf{V}}_p^{-1}\| = \sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\left(1 + O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\epsilon\right)\right) = O\left(\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\right).$$

The last step is true since Assumption 3.2 implies $\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\epsilon = O(1)$. Note that we also have

$$\|\mathbf{V}_P^{-1}\| = \frac{1}{\sigma_K(\mathbf{V}_P)} = O\left(\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\right).$$

Finally putting everything together with Eq (44) we get, with probability larger than $1 - O(Kn^{-2})$,

$$\max_{i\in[n]} \left\| \mathbf{e}_i^T \mathbf{V}(\mathbf{V}^T\hat{\mathbf{V}}) \left(\hat{\mathbf{V}}_p^{-1} - (\mathbf{\Pi}^T \mathbf{V}_P (\mathbf{V}^T\hat{\mathbf{V}}))^{-1} \right) \right\| = O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right) \left\| \hat{\mathbf{V}}_p^{-1} \right\|$$
$$= O\left(\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right).$$

The failure probability comes from the event that Theorem 3.1 fails, giving $O(Kn^{-2})$. □

Proof of Theorem 3.5. We break this up into proofs of Eqs (3) and (4). Recall that $\epsilon = \tilde{O}\left(\frac{\psi(\mathbf{P})\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right)$ is the row-wise error from Theorem 3.1.

Proof of Eq (3). Recall that $\hat{\mathbf{\Theta}} = \hat{\mathbf{V}}\hat{\mathbf{V}}_p^{-1}$. We have uniformly $\forall i \in [n]$,

$$\begin{aligned}
\left\| \mathbf{e}_i^T \left(\hat{\mathbf{\Theta}} - \mathbf{\Theta}\mathbf{\Pi} \right) \right\| &= \left\| \mathbf{e}_i^T \left(\hat{\mathbf{V}}\hat{\mathbf{V}}_p^{-1} - \mathbf{V}\mathbf{V}_P^{-1}\mathbf{\Pi} \right) \right\| \\
&\leq \left\| \mathbf{e}_i^T (\hat{\mathbf{V}} - \mathbf{V}(\mathbf{V}^T\hat{\mathbf{V}}))\hat{\mathbf{V}}_p^{-1} \right\| + \left\| \mathbf{e}_i^T \mathbf{V}(\mathbf{V}^T\hat{\mathbf{V}}) \left(\hat{\mathbf{V}}_p^{-1} - (\mathbf{\Pi}^T\mathbf{V}_P(\mathbf{V}^T\hat{\mathbf{V}}))^{-1} \right) \right\| \\
&\overset{(i)}{\leq} \left\| \mathbf{e}_i^T \left(\hat{\mathbf{V}} - \mathbf{V}(\mathbf{V}^T\hat{\mathbf{V}}) \right) \right\| \left\| \hat{\mathbf{V}}_p^{-1} \right\| + O\left(\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right) \\
&\overset{(ii)}{\leq} \epsilon \cdot O(\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}) + O\left(\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right) \\
&= O\left(\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\epsilon\right) \\
&= O\left(\sqrt{\lambda_1(\mathbf{\Theta}^T\mathbf{\Theta})}\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\right) \tilde{O}\left(\frac{\psi(\mathbf{P})\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right) \\
&= \tilde{O}\left(\frac{\psi(\mathbf{P})(\kappa(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\right)
\end{aligned}$$

with probability larger than $1 - O(Kn^{-2})$. Here (i) and (ii) follow from Lemma VII.3 and its proof, and the failure probability comes from the event that Theorem 3.1 does not hold.

Proof of Eq (4). Note $\hat{\rho}\hat{\mathbf{B}} = \hat{\mathbf{V}}_p\hat{\mathbf{E}}\hat{\mathbf{V}}_p^T$ and $\rho\mathbf{B} = \mathbf{V}_P\mathbf{E}\mathbf{V}_P^T$. Note that $\|\mathbf{E}\| \leq \max_i \|\mathbf{e}_i^T\mathbf{P}\|_1 = O(\rho n)$, and $\|\hat{\mathbf{E}}\| \leq \|\mathbf{E}\| + \|\mathbf{A} - \mathbf{P}\| = O(\rho n)$ using Weyl's inequality and Theorem 5.2 of

[2]. Then we have the following decomposition

$$
\begin{aligned}
&\left\|\hat{\rho}\hat{\mathbf{B}} - \rho\mathbf{\Pi}^T\mathbf{B}\mathbf{\Pi}\right\|_F = \left\|\hat{\mathbf{V}}_p\hat{\mathbf{E}}\hat{\mathbf{V}}_p^T - \mathbf{\Pi}^T\mathbf{V}_P\mathbf{E}\mathbf{V}_P^T\mathbf{\Pi}\right\|_F \\
=&\left\|\left(\hat{\mathbf{V}}_p - \mathbf{\Pi}^T\mathbf{V}_P\hat{\mathbf{O}}\right)\hat{\mathbf{E}}\hat{\mathbf{V}}_p^T + \mathbf{\Pi}^T\mathbf{V}_P\left(\hat{\mathbf{O}}\hat{\mathbf{E}} - \mathbf{E}\hat{\mathbf{O}}\right)\hat{\mathbf{V}}_p^T + \mathbf{\Pi}^T\mathbf{V}_P\mathbf{E}\hat{\mathbf{O}}\left(\hat{\mathbf{V}}_p^T - \hat{\mathbf{O}}^T\mathbf{V}_P^T\mathbf{\Pi}\right)\right\|_F \\
\leq&\left\|\hat{\mathbf{V}}_p - \mathbf{\Pi}^T\mathbf{V}_P\hat{\mathbf{O}}\right\|_F\left\|\hat{\mathbf{E}}\right\|\left\|\hat{\mathbf{V}}_p\right\| + \|\mathbf{V}_P\|\left\|\hat{\mathbf{O}}\hat{\mathbf{E}} - \mathbf{E}\hat{\mathbf{O}}\right\|_F\left\|\hat{\mathbf{V}}_p\right\| + \|\mathbf{V}_P\|\|\mathbf{E}\|\left\|\hat{\mathbf{V}}_p^T - \hat{\mathbf{O}}^T\mathbf{V}_P^T\mathbf{\Pi}\right\|_F \\
\leq&2\cdot O\left(\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K}\epsilon\right)\cdot O(\rho n)\cdot\frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}} + \frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}\left\|\hat{\mathbf{O}}\hat{\mathbf{E}} - \mathbf{E}\hat{\mathbf{O}}\right\|_F\frac{1}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}} \\
=&O\left(\frac{\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K}\rho n\epsilon}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}\right) + O\left(\frac{1}{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\right)\left\|\hat{\mathbf{O}}\hat{\mathbf{E}} - \mathbf{E}\hat{\mathbf{O}}\right\|_F
\end{aligned}
$$

using Eqs (46) and (47) and Lemma II.3.

Now by Lemma II.1,

$$
\begin{aligned}
\frac{1}{\rho}\|\hat{\rho}\hat{\mathbf{B}} - \rho\mathbf{\Pi}^T\mathbf{B}\mathbf{\Pi}\|_F &\leq O\left(\frac{\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K}n\epsilon}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}\right) + O\left(\frac{1}{\rho\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\right)\left\|\hat{\mathbf{O}}\hat{\mathbf{E}} - \mathbf{E}\hat{\mathbf{O}}\right\|_F \\
&= O\left(\frac{\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\sqrt{K}n}{\sqrt{\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}}\right)\cdot\tilde{O}\left(\frac{\psi(\mathbf{P})\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}}\right) + O\left(\frac{1}{\rho\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\right)\cdot O\left(K^2\sqrt{n\rho}\right) \\
&= \tilde{O}\left(\frac{\psi(\mathbf{P})\kappa(\mathbf{\Theta}^T\mathbf{\Theta})Kn^{1.5}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^2}\right)
\end{aligned}
$$

with probability larger than $1 - O(Kn^{-2})$. The failure probability comes from the event that Theorem 3.1 does not hold.

□

Proof Corollary 3.7. Define the event

$$\Omega := \{\mathbf{\Theta} : \lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}) \geq 1/\rho, \lambda^*(\mathbf{P}) \geq 4\sqrt{n\rho}(\log n)^\xi \text{ for some constant } \xi > 1\}.$$

If $\boldsymbol{\theta}_i \sim \text{Dirichlet}(\boldsymbol{\alpha})$ and Assumption 3.3 is satisfied, we have $\mathrm{P}(\mathbf{\Theta} \in \Omega) \geq 1 - Kn^{-3}$. If Assumption 2.1 holds, and $\lambda^*(\mathbf{B}) = \tilde{\Omega}(\frac{\min\{K,\kappa(\mathbf{B})\}^2K^2}{\sqrt{n\rho}})$, for $\mathbf{\Theta} \in \Omega$, by Theorem 3.1 and

Lemma 3.6,

$$\max_{i\in[n]}\left\|\mathbf{e}_i^T\left(\hat{\mathbf{\Theta}}-\mathbf{\Theta}\mathbf{\Pi}\right)\right\| = \tilde{O}\left(\frac{\psi(\mathbf{P})(\kappa(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\right) = \tilde{O}\left(\frac{\psi(\mathbf{P})\left(\frac{\alpha_{\max}+\|\boldsymbol{\alpha}\|^2}{\alpha_{\min}}\right)^{1.5}\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})\frac{n}{2\nu(1+\alpha_0)}}\right)$$
$$= \tilde{O}\left(\frac{\min\{K,\kappa(\mathbf{B})\}^2K^{1.5}}{\sqrt{\rho n}\lambda^*(\mathbf{B})}\right). \tag{48}$$

Since $\max_a \alpha_a \le C\min_a \alpha_a$ for some constant $C \ge 1$, and $\alpha_0 = O(1)$, the last step uses that

$$\frac{\alpha_{\max}+\|\boldsymbol{\alpha}\|^2}{\alpha_{\min}} \le \frac{\alpha_{\max}+\alpha_{\max}}{\alpha_{\min}} = (1+\alpha_0)\frac{\alpha_{\max}}{\alpha_{\min}} = O(1),$$

and by the worst case bound from Lemma 5.1, we have, $\psi(\mathbf{P}) \le \min\{K,\kappa(\mathbf{P})\}^2 \le \min\{K,\kappa(\mathbf{\Theta}^T\mathbf{\Theta})\kappa(\mathbf{B})\}^2 = O(\min\{K,\kappa(\mathbf{B})\}^2)$.

Now we are ready to obtain the failure probability of Eq (48). Consider the event $\mathcal{A}$ that $\hat{\mathbf{\Theta}}$ does not satisfy Eq (48). Then, by Theorem 3.1,

$$\mathrm{P}(\mathcal{A}) = \int_{\mathbf{\Theta}\in\Omega}\mathrm{P}\left(\mathcal{A}|\mathbf{\Theta}\right)\mathrm{P}(\mathbf{\Theta})d\mathbf{\Theta} + \int_{\mathbf{\Theta}\notin\Omega}\mathrm{P}\left(\mathcal{A}|\mathbf{\Theta}\right)\mathrm{P}(\mathbf{\Theta})d\mathbf{\Theta}$$
$$= O\left(\frac{K}{n^2}\right) + 1 - \mathrm{P}(\mathbf{\Theta}\in\Omega) = O\left(\frac{K}{n^2}\right). \tag{49}$$

Similarly, by Theorem 3.1 and Lemma 3.6,

$$\frac{1}{\rho}\|\hat{\rho}\hat{\mathbf{B}}-\rho\mathbf{\Pi}^T\mathbf{B}\mathbf{\Pi}\|_F = \tilde{O}\left(\frac{\psi(\mathbf{P})\kappa(\mathbf{\Theta}^T\mathbf{\Theta})Kn^{1.5}}{\sqrt{\rho}\lambda^*(\mathbf{B})(\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}))^2}\right) = \tilde{O}\left(\frac{\psi(\mathbf{P})\left(\frac{\alpha_{\max}+\|\boldsymbol{\alpha}\|^2}{\alpha_{\min}}\right)Kn^{1.5}}{\sqrt{\rho}\lambda^*(\mathbf{B})\left(\frac{n}{2\nu(1+\alpha_0)}\right)^2}\right)$$
$$= \tilde{O}\left(\frac{\min\{K,\kappa(\mathbf{B})\}^2K^3}{\sqrt{\rho n}\lambda^*(\mathbf{B})}\right). \tag{50}$$

By an argument analogous to that in Eq (49), we can show that the failure probability of Eq (50) is $O(Kn^{-2})$. □

VIII Comparison with [10]

We first translate some key assumptions in [10] (Eqs (2.14) and (2.15)) with our notation.

Assumption VIII.1. Assume for some constants $C > 0$ and $c_1 > 0$,

$$\frac{n}{CK} \leq \lambda_K(\mathbf{\Theta}^T\mathbf{\Theta}) \leq \lambda_1(\mathbf{\Theta}^T\mathbf{\Theta}) \leq \frac{Cn}{K}$$
$$\frac{c_1 n}{K}\lambda^*(\mathbf{B}) \leq |\lambda_K(\mathbf{B}\mathbf{\Theta}^T\mathbf{\Theta})| \leq |\lambda_2(\mathbf{B}\mathbf{\Theta}^T\mathbf{\Theta})| \leq \frac{n}{c_1 K}\lambda^*(\mathbf{B})$$
$$|\lambda_2(\mathbf{B}\mathbf{\Theta}^T\mathbf{\Theta})| \leq (1-c_1)\lambda_1(\mathbf{B}\mathbf{\Theta}^T\mathbf{\Theta})$$

Lemma VIII.1. *If Assumption VIII.1 is satisfied, for $\hat{\mathbf{\Theta}}$ estimated by SPACL, we have,*

$$\left\|\mathbf{e}_i^T\left(\hat{\mathbf{\Theta}} - \mathbf{\Theta}\mathbf{\Pi}\right)\right\| = \tilde{O}\left(\frac{K^{1.5}}{\sqrt{n\rho}\lambda^*(\mathbf{B})}\right).$$

Proof. By Theorem 1.3.22 of [5], $\rho\mathbf{B}\mathbf{\Theta}^T\mathbf{\Theta}$ and $\mathbf{P} = \rho\mathbf{\Theta}\mathbf{B}\mathbf{\Theta}^T$ have the same K largest eigenvalues in magnitude. So Assumption VIII.1 implies that:

$$\frac{c_1 n\rho}{K}\lambda^*(\mathbf{B}) \leq |\lambda_K(\mathbf{P})| \leq |\lambda_2(\mathbf{P})| \leq \frac{n\rho}{c_1 K}\lambda^*(\mathbf{B})$$
$$|\lambda_2(\mathbf{P})| \leq (1-c_1)\lambda_1(\mathbf{P}).$$

Then the eigenvalues of $\mathbf{P}$ can be divided into at most 2 groups where eigenvalues in each group are of the same order, by Lemma 5.1, we have $\psi(\mathbf{P}) = O(1)$.

On the other hand, if Assumption VIII.1 is satisfied, we have $\kappa(\mathbf{\Theta}^T\mathbf{\Theta}) = O(1)$, and by Theorem 3.5,

$$\left\|\mathbf{e}_i^T\left(\hat{\mathbf{\Theta}} - \mathbf{\Theta}\mathbf{\Pi}\right)\right\| = \tilde{O}\left(\frac{(\kappa(\mathbf{\Theta}^T\mathbf{\Theta}))^{1.5}\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})\lambda_K(\mathbf{\Theta}^T\mathbf{\Theta})}\right) = \tilde{O}\left(\frac{\sqrt{Kn}}{\sqrt{\rho}\lambda^*(\mathbf{B})n/K}\right) = \tilde{O}\left(\frac{K^{1.5}}{\sqrt{n\rho}\lambda^*(\mathbf{B})}\right)$$

□

Remark VIII.1. *Since [10] shows ℓ_1 norm error bound, our result in Lemma VIII.1 matches theirs with an extra $\sqrt{K}$ factor up-to logarithm factor, if we convert the bound in Lemma VIII.1 to ℓ_1 norm by multiplying $\sqrt{K}$.*

IX Comparison with [11]

Lemma IX.1. *Let $\boldsymbol{\theta}_i \sim \text{Dirichlet}(\alpha)$. If Assumptions 2.1 and 3.3 hold, and $\lambda^*(\mathbf{B}) = \tilde{\Omega}(\frac{\min\{K,(1+\alpha_0)\kappa(\mathbf{B})\alpha_{\max}/\alpha_{\min}\}^2 K^2}{\sqrt{n\rho}})$, there exists a permutation matrix $\mathbf{\Pi}$ such that with probability at least $1 - O(K/n^2)$, $\forall i \in [n]$,*

$$\left\|\hat{\mathbf{\Theta}} - \mathbf{\Theta\Pi}\right\|_1 = \tilde{O}\left(\left(\frac{\alpha_{\max}}{\alpha_{\min}}\right)^{1.5}\sqrt{\frac{n}{\rho}}\frac{\min\{K,(1+\alpha_0)\kappa(\mathbf{B})\alpha_{\max}/\alpha_{\min}\}^2 K\nu\,(1+\alpha_0)^{2.5}}{\lambda^*(\mathbf{B})}\right),$$

where $\|\mathbf{M}\|_1 = \sum_{i,j}|\mathbf{M}_{ij}|$ is the ℓ_1 norm for a matrix $\mathbf{M}$.

Proof. First note from the proof of Corollary 3.7, we have $(\alpha_{\max} + \|\boldsymbol{\alpha}\|^2)/\alpha_{\min} \leq (1+\alpha_0)\alpha_{\max}/\alpha_{\min}$, and by Lemma 3.6, we have, with high probability $\psi(\mathbf{P}) \leq \min\{K, \kappa(\mathbf{P})\}^2 \leq \min\{K, \kappa(\mathbf{\Theta}^T\mathbf{\Theta})\kappa(\mathbf{B})\}^2 = O(\min\{K,(1+\alpha_0)\kappa(\mathbf{B})\alpha_{\max}/\alpha_{\min}\}^2)$. Now by Theorem 3.5, if we sum up the squared error bound for each row, we can get a Frobenius bound:

$$\frac{1}{\sqrt{n}}\left\|\hat{\mathbf{\Theta}} - \mathbf{\Theta}\right\|_F = \tilde{O}\left(\left(\frac{\alpha_{\max}}{\alpha_{\min}}\right)^{1.5}\frac{\min\{K,(1+\alpha_0)\kappa(\mathbf{B})\alpha_{\max}/\alpha_{\min}\}^2 K^{0.5}\nu\,(1+\alpha_0)^{2.5}}{\sqrt{\rho n}\lambda^*(\mathbf{B})}\right)$$

and so

$$\begin{aligned}\left\|\hat{\mathbf{\Theta}} - \mathbf{\Theta}\right\|_1 &\leq \sqrt{Kn}\left\|\hat{\mathbf{\Theta}} - \mathbf{\Theta}\right\|_F \\ &= \tilde{O}\left(\left(\frac{\alpha_{\max}}{\alpha_{\min}}\right)^{1.5}\sqrt{\frac{n}{\rho}}\frac{\min\{K,(1+\alpha_0)\kappa(\mathbf{B})\alpha_{\max}/\alpha_{\min}\}^2 K\nu\,(1+\alpha_0)^{2.5}}{\lambda^*(\mathbf{B})}\right)\end{aligned}$$

□

Remark IX.1. *By Theorem 9 of [11], we have:*

$$\begin{aligned}\left\|\hat{\mathbf{\Theta}} - \mathbf{\Theta}\right\|_1 &= \tilde{O}\left(\frac{\alpha_{\max}}{\alpha_0}\left(\frac{\alpha_{\max}}{\alpha_{\min}}\right)^{0.5}\frac{\sqrt{n}K\nu^{1.5}(1+\alpha_0)^{1.5}\sqrt{\max_i(\rho\mathbf{e}_i^T\mathbf{B}\boldsymbol{\alpha})}}{\rho\sqrt{\alpha_0}\lambda^*(\mathbf{B})}\right) \\ &= \tilde{O}\left(\frac{\alpha_{\max}}{\alpha_0}\left(\frac{\alpha_{\max}}{\alpha_{\min}}\right)^{0.5}\sqrt{\frac{n}{\rho}}\frac{K\nu^{1.5}(1+\alpha_0)^{1.5}}{\lambda^*(\mathbf{B})}\right) \qquad (51)\end{aligned}$$

When $\max_a \alpha_a \leq C\min_a \alpha_a$ for some constant $C \geq 1$, $\alpha_0 = O(1)$ and $\kappa(\mathbf{B}) = \Theta(1)$, we have $\nu = O(K)$, $\alpha_{\max}/\alpha_0 = O(1/K)$, $\alpha_{\max}/\alpha_{\min} = O(1)$ and $\min\{K,(1+\alpha_0)\kappa(\mathbf{B})\alpha_{\max}/\alpha_{\min}\}^2 = O(\min\{K,\kappa(\mathbf{B})\}^2) = O(1)$, so our bound in Lemma IX.1 is worse by $\sqrt{K}$ than Eq. (51).

For worst case analysis, $\alpha_{\max}/\alpha_0 = O(1)$, $\alpha_{\max}/\alpha_0 = O(\nu)$, and $\min\{K, (1+\alpha_0)\kappa(\mathbf{B})\alpha_{\max}/\alpha_{\min}\}^2 = K^2$*, so our bound in Lemma IX.1 is worse by $K^2\sqrt{\nu}(1+\alpha_0)$ than Eq.* (51)*.*

Note that the proposed algorithm in [11] requires prior knowledge on α_0 while our algorithm does not need α_0 as input.

X Why Pruning Works

Proving the pruning algorithm requires strong distributional conditions on the residuals of the rows of eigenvectors. Here we present a heuristic argument of why pruning works. Note that in the pruning algorithm, essentially we are estimating the density of points in an ϵ-ball around every point i which has sufficiently large norm. This should work only if the points outside the population simplex have lower density in their ϵ-balls than the corners of the simplex. Otherwise, the pruning will remove the corners of the population simplex, diminishing the quality of the pure nodes. We consider $K \in \{2, \ldots, 10\}$ and

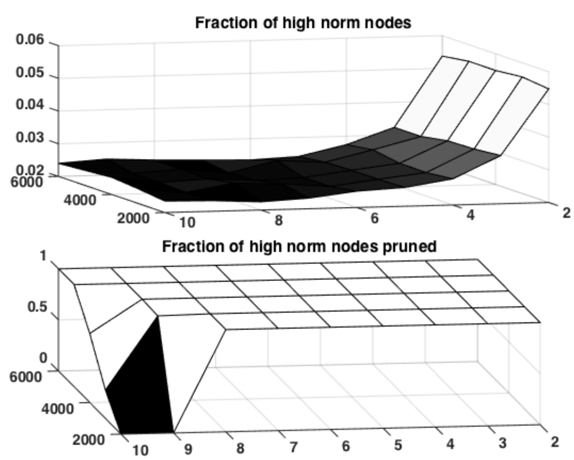


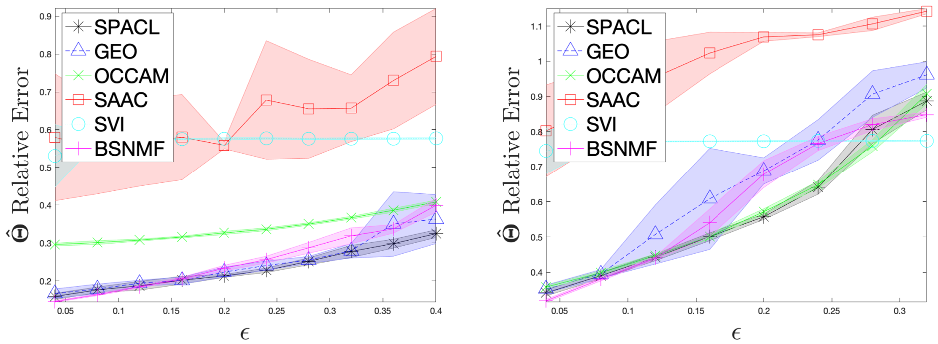
Figure I: Top panel: fraction of nodes with high norm. Bottom panel: fraction of nodes with high norm pruned. We vary $K \in \{2, \ldots, 10\}$ on the X axis and vary $n \in \{2000, 3000, \ldots, 6000\}$ on the Y axis.

$n \in \{2000, 3000, \ldots, 6000\}$, $\boldsymbol{\alpha} = \mathbf{1}_K/K$, $\mathbf{B}_{ii} = 1, \mathbf{B}_{ij} = 0.001$ and $\rho = \log n/n$. For each combination we use ϵ as the median of the row-wise difference of the empirical eigenvectors from their suitably rotated population counterpart. Let $y = \max_i \|\mathbf{V}_i\|$ denote the largest row-wise norm of the population eigenvectors; recall that this occurs at one of the corners of the simplex. Let S_0 denote the set of nodes with high empirical eigenvector row-norms (the "high-norm" nodes), defined as $S_0 := \{i : \|\hat{\mathbf{V}}_i\| \geq y + \epsilon\}$. SPA will choose at least one of these nodes (and possibly several of them) as its estimated corners. Let $B(x, \epsilon)$ denote the ℓ_2 ball of size ϵ centered at point x. For each of the K corners c_i of the population simplex (c_i equals some row of $\mathbf{V}_P$), we compute the number of neighborhood points $x_i := |\{j | \hat{\mathbf{V}}_j \in B(c_i, \epsilon)\}|$; let $\delta := \min_i x_i$ be the minimum neighborhood size among these corners. Similarly, for each $i \in S_0$, we compute $z_i = |\{j | \hat{\mathbf{V}}_j \in B(\hat{\mathbf{V}}_i, \epsilon)\}|$. Now we count the fraction of nodes in S_0 that could be pruned without pruning the corners c_i of the population simplex. This fraction is $m = \frac{\sum_{i \in S_0} \mathbf{1}\{z_i < \delta\}}{|S_0|}$. Fig I shows that for almost all combinations of K and n, we have $m = 1$, i.e., all the nodes in S_0 do get pruned, except for $K = 10, n = 2000$. This is expected, since for large K and small n the pure node density around the corners of the population simplex will be small. Fig I shows the fraction $|S_0|/n$ of high-norm nodes. For all (K, n) combinations pruning removes about a 2% to 6% of the nodes.

XI Extra simulation results

Changing B: In Fig II (i), we plot the relative error in estimating $\boldsymbol{\Theta}$ against increasing off diagonal noise ϵ of $\mathbf{B}$. We take $K = 3$, $\rho = 0.15$, $\alpha_i = 3/K = 1$, $\mathbf{B}_{ii} = 1$, $i \in [K]$. We see that SPACL outperforms SAAC, SVI, and OCCAM over the entire parameter range. For large ϵ, it is also better than GeoNMF and BSNMF.

We also include simulation results with $K = 7$. We take $\rho = 0.15$, $\alpha_i = 3/K = 3/7$, $\mathbf{B}_{ii} = 1$, $i \in [K]$. We see in Fig II (ii) that SAAC performs poorly, and OCCAM performs similarly with SPACL, which can also be implied from the simulation results on changing



(i) (ii)

Figure II: Error against ϵ: we use $\mathbf{B}_{ii} = 1$, $\mathbf{B}_{ij} = \epsilon$ for $i \neq j$. (i) $K = 3$. (ii) $K = 7$.

K. SPACL is more stable and outperforms GeoNMF and BSNMF.

References

[1] Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.

[2] J. Lei, A. Rinaldo, et al. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

[3] H. Schaefer. *Banach Lattices and Positive Operators*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1974. ISBN 9783540069362.

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

[5] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[6] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, et al. Spectral statistics of erdős–rényi graphs i: local semicircle law. *The Annals of Probability*, 41(3B):2279–2375, 2013.

[7] E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *arXiv preprint arXiv:1709.09565*, 2017.

[8] J. Cape, M. Tang, C. E. Priebe, et al. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47 (5):2405–2439, 2019.

[9] N. Gillis and S. A. Vavasis. Fast and robust recursive algorithmsfor separable nonnegative matrix factorization. *PAMI*, 36(4):698–714, 2014.

[10] J. Jin, Z. T. Ke, and S. Luo. Estimating network memberships by simplex vertex hunting. *arXiv preprint arXiv:1708.07852*, 2017.

[11] A. Anandkumar, R. Ge, D. Hsu, and S. M. Kakade. A tensor approach to learning mixed membership community models. *JMLR*, 15(1):2239–2312, 2014.