# Appendix for
# "On Mixed Memberships and Symmetric Nonnegative Matrix Factorizations"

**Xueyu Mao** [1] **Purnamrita Sarkar** [2] **Deepayan Chakrabarti** [3]

## A. Identifiability

**Lemma A.1.** (Lemma 1.1 of (Minc, 1988)) The inverse of a nonnegative matrix matrix $\mathbf{M}$ is nonnegative *if and only if* $\mathbf{M}$ is a generalized permutation matrix.

*Proof of Theorem 2.1.* Suppose there are two parameter settings ($\mathbf{\Theta}^{(1)}$, $\mathbf{B}^{(1)}$, $\rho^{(1)}$) and ($\mathbf{\Theta}^{(2)}$, $\mathbf{B}^{(2)}$, $\rho^{(2)}$) that yield the same probability matrix:

$$\mathbf{P} = \rho^{(1)}\mathbf{\Theta}^{(1)}\mathbf{B}^{(1)}{\mathbf{\Theta}^{(1)}}^T = \rho^{(2)}\mathbf{\Theta}^{(2)}\mathbf{B}^{(2)}{\mathbf{\Theta}^{(2)}}^T.$$

Pick up pure node indices set $\mathcal{I}_1$ of $\mathbf{\Theta}^{(1)}$ such that $\mathbf{\Theta}^{(1)}_{\mathcal{I}_1} = \mathbf{I}$, and denote $\mathbf{M} = \mathbf{\Theta}^{(2)}_{\mathcal{I}_1}$. Similarly, pick up pure node indices set $\mathcal{I}_2$ of $\mathbf{\Theta}^{(2)}$ such that $\mathbf{\Theta}^{(2)}_{\mathcal{I}_2} = \mathbf{I}$, and let $\mathbf{W} = \mathbf{\Theta}^{(1)}_{\mathcal{I}_2}$.

Then

$$\rho^{(1)}\mathbf{B}^{(1)} = \rho^{(2)}\mathbf{M}\mathbf{B}^{(2)}\mathbf{M}^T \text{ and } \rho^{(1)}\mathbf{W}\mathbf{B}^{(1)}\mathbf{W}^T = \rho^{(2)}\mathbf{B}^{(2)}.$$

Denote $\mathbf{T} = \mathbf{M}\mathbf{W}$, then

$$\mathbf{B}^{(1)} = \frac{1}{\rho^{(1)}}\mathbf{M}\rho^{(1)}\mathbf{W}\mathbf{B}^{(1)}\mathbf{W}^T\mathbf{M}^T = \mathbf{T}\mathbf{B}^{(1)}\mathbf{T}^T. \tag{1}$$

Note that $\mathbf{M} \cdot \mathbf{1} = \mathbf{\Theta}^{(2)}_{\mathcal{I}_1} \cdot \mathbf{1} = \mathbf{1}$ and $\mathbf{W} \cdot \mathbf{1} = \mathbf{\Theta}^{(1)}_{\mathcal{I}_2} \cdot \mathbf{1} = \mathbf{1}$, so $\mathbf{T} \cdot \mathbf{1} = \mathbf{M}\mathbf{W} \cdot \mathbf{1} = \mathbf{1}$. We can consider $\mathbf{T}$ as a transition matrix of a Markov chain, whose states are the nodes of the graph. Keep applying equation (1) to its RHS, we get

$$\mathbf{B}^{(1)} = \mathbf{T}^k\mathbf{B}^{(1)}{\mathbf{T}^k}^T,$$

which implies $\mathbf{B}^{(1)} = \mathbf{T}_\infty\mathbf{B}^{(1)}\mathbf{T}_\infty^T$, where $\mathbf{T}_\infty = \lim_{k\to\infty} \mathbf{T}^k$.

Given that $\mathbf{B}^{(1)}$ has full rank $K$, we must have $\mathbf{T}_\infty$ has full rank. Now we prove that stationary point of the Markov chain, $\mathbf{T}_\infty$, must be identity matrix.

The nodes of a finite-size Markov chain can be split into a finite number of communication classes, and possibly some transient nodes.

1. If a communication class has at least two nodes and is aperiodic, then the rows corresponding to those nodes in $\mathbf{T}_\infty$ are the stationary distribution for that class. Hence, $\mathbf{T}_\infty$ has identical rows, so it cannot be full rank.

2. The probability of a Markov chain ending in a transient node goes to zero as the number of iterations $k$ grows, so the column of $\mathbf{T}_\infty$ corresponding to any transient node is identically zero. Again, this means that $\mathbf{T}_\infty$ cannot be full rank.

[1]Department of Computer Science. [2]Department of Statistics and Data Sciences. [3]Department of Information, Risk, and Operations Management. The University of Texas at Austin, TX, USA. Correspondence to: Xueyu Mao <xmao@cs.utexas.edu>, Purnamrita Sarkar <purna.sarkar@austin.utexas.edu>, Deepayan Chakrabarti <deepay@utexas.edu>.

*Proceedings of the 34$^{th}$ International Conference on Machine Learning*, Sydney, Australia, PMLR 70, 2017. 

Hence, the only configuration in which $\mathbf{T}_\infty$ has full rank is when it contains $K$ communication classes, each with one node. This implies that $\mathbf{T}_\infty = \mathbf{I}$, and hence $\mathbf{T} = \mathbf{I}$. Note that if the communication classes are periodic, we can consider $\mathbf{T}^t$ where $t$ is the product of the periods of all the classes; the matrix $\mathbf{T}^t$ is now aperiodic for all the communication classes, and the above argument still applies to $\mathbf{T}_\infty = \lim_{k\to\infty} (\mathbf{T}^t)^k$.

As $\mathbf{I} = \mathbf{T} = \mathbf{MW}$, $\mathbf{M}$ and $\mathbf{W}$ have full rank, then $\mathbf{M}^{-1} = \mathbf{W}$, which is the case that a nonnegative matrix $\mathbf{M}$ has nonnegative inverse $\mathbf{W}$, using Lemma A.1, we know that $\mathbf{M}$ is a generalized permutation matrix, and note that each row of $\mathbf{M}$ sums to 1, the scale goes away and thus $\mathbf{M}$ is a permutation matrix, which implies $\mathbf{W}$ is also a permutation matrix. As largest element of $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ are equals as 1, we should have $\rho^{(1)} = \rho^{(2)}$ and thus $\mathbf{B}^{(1)} = \mathbf{MB}^{(2)}\mathbf{M}^T$.

Also since we have

$$\begin{aligned}\rho^{(1)}\mathbf{B}^{(1)}{\mathbf{\Theta}^{(1)}}^T &= \rho^{(1)}\mathbf{\Theta}^{(1)}_{\mathcal{I}_1}\mathbf{B}^{(1)}{\mathbf{\Theta}^{(1)}}^T = \rho^{(2)}\mathbf{\Theta}^{(2)}_{\mathcal{I}_1}\mathbf{B}^{(2)}{\mathbf{\Theta}^{(2)}}^T = \rho^{(2)}\mathbf{MB}^{(2)}{\mathbf{\Theta}^{(2)}}^T \\ &= \rho^{(2)}\mathbf{MB}^{(2)}\mathbf{M}^T\mathbf{M}{\mathbf{\Theta}^{(2)}}^T = \rho^{(1)}\mathbf{B}^{(1)}\mathbf{M}{\mathbf{\Theta}^{(2)}}^T,\end{aligned}$$

left multiply $\left(\rho^{(1)}\mathbf{B}^{(1)}\right)^{-1}$ on both sides, we have $\mathbf{\Theta}^{(1)} = \mathbf{\Theta}^{(2)}\mathbf{M}^T$.

Thus we have shown that MMSB is identifiable up to a permutation. □

## B. Uniqueness of SNMF for MMSB networks

**Lemma B.1** (Huang et al. (2014))**.** If rank$(\mathbf{P}) = K$, the Symmetric NMF $\mathbf{P} = \mathbf{WW}^T$ is unique if and only if the non-negative orthant is the only self-dual simplicial cone $\mathcal{A}$ with $K$ extreme rays that satisfies $\mathrm{cone}(\mathbf{W}^T) \subseteq \mathcal{A} = \mathcal{A}^*$, where $\mathcal{A}^*$ is the dual cone of $\mathcal{A}$, defined as $\mathcal{A}^* = \{\mathbf{y}|\mathbf{x}^T\mathbf{y} \geq 0, \forall \mathbf{x} \in \mathcal{A}\}$.

*Proof of Theorem 2.2.* When $\mathbf{B}$ is diagonal, it has a square root $\mathbf{C} = \mathbf{B}^{1/2}$, where $\mathbf{C}$ is also a positive diagonal matrix. It is easy to see that $\mathrm{cone}(\mathbf{C})$ is the non-negative orthant $\mathbb{R}^K_+$, so we have

$$\mathrm{cone}(\mathbf{W}^T) = \mathrm{cone}(\mathbf{C}^T\mathbf{\Theta}^T) = \mathrm{cone}(\mathbf{C}^T) = \mathrm{cone}(\mathbf{C}) = \mathbb{R}^K_+ = \mathbb{R}^{K*}_+.$$

The second equality follows from the fact that $\mathbf{\Theta}$ contains all pure nodes, and other nodes are convex combinations of these pure nodes. The fourth equality is due to the diagonal form of $\mathbf{C}$.

To see that this is unique, suppose there is another self-dual simplicial cone satisfying $\mathrm{cone}(\mathbf{W}^T) \subseteq \mathcal{A} = \mathcal{A}^*$. Then we have $\mathbb{R}^K_+ \subseteq \mathcal{A}$ and $\mathcal{A} = \mathcal{A}^* \subseteq \left(\mathbb{R}^K_+\right)^* = \mathbb{R}^K_+$, which implies $\mathcal{A} = \mathbb{R}^K_+$.

Hence, by Lemma B.1, an identifiable MMSB model with a diagonal $\mathbf{B}$ is sufficient for the Symmetric NMF solution to be unique and correct. □

## C. Concentration of the Laplacian

We will use $X = c(1 \pm \epsilon)$ to denote $X \in c[1-\epsilon, 1+\epsilon]$ for ease of notation from now onwards.

**Lemma C.1.** For $\mathbf{\Theta} \in \mathbb{R}^{n\times K}$, where each row $\boldsymbol{\theta}_i \sim \mathrm{Dirichlet}(\boldsymbol{\alpha})$, $\forall j \in [K]$,

$$\sum_{i=1}^n \theta_{ij} = n\frac{\alpha_j}{\alpha_0}\left(1 \pm O_P\left(\sqrt{\frac{\alpha_0}{\alpha_j}\frac{\log n}{n}}\right)\right)$$

with probability larger than $1 - 1/n^3$.

*Proof.* By using Chernoff bound

$$\mathrm{P}\left(\left|\sum_{i=1}^n \theta_{ij} - n\frac{\alpha_j}{\alpha_0}\right| > \epsilon n\frac{\alpha_j}{\alpha_0}\right) \leq \exp\left(-\frac{\epsilon^2 n\frac{\alpha_j}{\alpha_0}}{3}\right),$$

so by setting $\epsilon = O_P\left(3\sqrt{\frac{\log n}{n\alpha_j/\alpha_0}}\right)$, $\left|\sum_{i=1}^n \theta_{ij} - n\frac{\alpha_j}{\alpha_0}\right| \leq 3\sqrt{\frac{\alpha_j}{\alpha_0}n\log n}$, with probability larger than $1 - 1/n^3$.

That is

$$\sum_{i=1}^{n}\theta_{ij} = n\frac{\alpha_j}{\alpha_0} \pm O_P\left(\sqrt{\frac{\alpha_j}{\alpha_0}n\log n}\right) = n\frac{\alpha_j}{\alpha_0}\left(1 \pm O_P\left(\sqrt{\frac{\alpha_0}{\alpha_j}\frac{\log n}{n}}\right)\right).$$

□

**Lemma C.2.** (Theorem 5.2 of (Lei et al., 2015)) Let $\mathbf{A}$ be the adjacency matrix of a random graph on $n$ nodes in which edges occur independently. Set $\mathrm{E}[\mathbf{A}] = \mathbf{P}$ and assume that $n\max_{i,j}\mathbf{P}_{ij} \le d$ for $d \ge c_0 \log n$ and $c_0 > 0$. Then, for any $r > 0$ there exists a constant $C = C(r, c_0)$ such that:

$$\mathrm{P}(\|\mathbf{A} - \mathbf{P}\| \le C\sqrt{d}) \ge 1 - n^{-r}.$$

**Fact C.1.** If $\mathbf{M}$ is rank $k$, then $\|\mathbf{M}\|_F^2 \le k\|\mathbf{M}\|^2$.

**Lemma C.3.** (Variant of Davis-Kahan (Yu et al., 2015)). Let $\mathbf{P}$, $\hat{\mathbf{A}} \in \mathbb{R}^{n\times n}$ be symmetric, with eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n$ and $\hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_n$ respectively. Fix $1 \le r \le s \le n$, and assume that $\min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}) > 0$, where we define $\lambda_0 = \infty$ and $\lambda_{n+1} = -\infty$. Let $d = s - r + 1$, and let $\mathbf{V} = (\mathbf{v}_r, \mathbf{v}_{r+1}, \cdots, \mathbf{v}_s) \in \mathbb{R}^{n\times d}$ and $\hat{V} = (\hat{\mathbf{v}}_r, \hat{\mathbf{v}}_{r+1}, \cdots, \hat{\mathbf{v}}_s) \in \mathbb{R}^{n\times d}$ have orthonormal columns satisfying $\mathbf{P}\mathbf{v}_j = \lambda_j\mathbf{v}_j$ and $\hat{\mathbf{A}}\hat{\mathbf{v}}_j = \hat{\lambda}_j\hat{\mathbf{v}}_j$ for $j = r, r+1, \cdots, s$. Then there exists an orthogonal matrix $\hat{\mathbf{O}} \in \mathbb{R}^{d\times d}$ such that

$$\left\|\hat{\mathbf{V}} - \mathbf{V}\hat{\mathbf{O}}\right\|_F \le \frac{2^{3/2}\min\left(d^{1/2}\left\|\hat{\mathbf{A}} - \mathbf{P}\right\|, \left\|\hat{\mathbf{A}} - \mathbf{P}\right\|_F\right)}{\min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1})}$$

**Lemma C.4.** (Lemma A.1. of (Tang et al., 2013)). Let $\mathbf{H}_1$, $\mathbf{H}_2 \in \mathbb{R}^{n\times n}$ be positive semidefinite with $\mathrm{rank}(\mathbf{H}_1) = \mathrm{rank}(\mathbf{H}_2) = K$. Let $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n\times K}$ be of full column rank such that $\mathbf{X}\mathbf{X}^T = \mathbf{H}_1$ and $\mathbf{Y}\mathbf{Y}^T = \mathbf{H}_2$. Let $\lambda_K(\mathbf{H}_2))$ be the smallest nonzero eigenvalue of $\mathbf{H}_2$. Then there exists an orthogonal matrix $\mathbf{R} \in \mathbb{R}^{K\times K}$ such that:

$$\|\mathbf{X}\mathbf{R} - \mathbf{Y}\|_F \le \frac{\sqrt{K}\,\|\mathbf{H}_1 - \mathbf{H}_2\|\left(\sqrt{\|\mathbf{H}_1\|} + \sqrt{\|\mathbf{H}_2\|}\right)}{\lambda_K(\mathbf{H}_2)}.$$

**Lemma C.5.** Recall that $\hat{\mathbf{A}}_1 = \hat{\mathbf{V}}_1\hat{\mathbf{E}}_1\hat{\mathbf{V}}_1^T$ and $\mathbf{P}_1 = \mathbf{P}(\mathcal{S}, \mathcal{S})$ in Algorithm 1. If $\rho n = \Omega(\log n)$, then

$$\left\|\hat{\mathbf{A}}_1 - \mathbf{P}_1\right\| = O_P(\sqrt{\rho n}), \text{ and } \left\|\hat{\mathbf{A}}_1 - \mathbf{P}_1\right\|_F = O_P(\sqrt{K\rho n})$$

with probability larger than $1 - 1/n^3$.

*Proof.* Lemma C.2 gives the spectral bound of binary symmetric random matrices, in our model,

$$\frac{n}{2}\max_{i,j}\mathbf{P}_1(i,j) \le \frac{n}{2}\max_{i,j}\mathbf{P}(i,j) = \frac{n}{2}\max_{i,j}\rho\boldsymbol{\theta}_i\mathbf{B}\boldsymbol{\theta}_j^T \le \frac{n}{2}\max_{i,j}\rho\boldsymbol{\theta}_i\mathbf{I}\boldsymbol{\theta}_j^T \le \rho\frac{n}{2}.$$

Note that we need to use $\mathbf{B}$ is diagonal probability matrix and $\boldsymbol{\theta}_i$, $i \in [\frac{n}{2}]$ has $\ell_1$ norm 1 and all nonnegative elements for the last two inequality.

Since $\rho n = \Omega(\log n)$, $\exists c_0 \ge 0$ that $\rho\frac{n}{2} \ge c_0 \log\frac{n}{2}$.

Let $d = \rho\frac{n}{2}$, then $d \ge \frac{n}{2}\max_{i,j}\mathbf{P}_1(i,j)$ and $d \ge c_0\log\frac{n}{2}$, by Lemma C.2, $\forall r \ge 0$, $\exists\, C > 0$ that

$$\mathrm{P}\left(\|\mathbf{A}_1 - \mathbf{P}_1\| \le C\sqrt{\rho\frac{n}{2}}\right) \ge 1 - (\frac{n}{2})^{-r},$$

where $\mathbf{A}_1 = \mathbf{A}(\mathcal{S}, \mathcal{S})$. So $\|\mathbf{A}_1 - \mathbf{P}_1\| = O_P(\sqrt{\rho n})$, specially, taking $r = 3$ then it is with probability larger than $1 - 1/n^3$. Hence

$$\left\|\hat{\mathbf{A}}_1 - \mathbf{P}_1\right\| \le \left\|\hat{\mathbf{A}}_1 - \mathbf{A}_1 + \mathbf{A}_1 - \mathbf{P}_1\right\| \le \left\|\hat{\mathbf{A}}_1 - \mathbf{A}_1\right\| + \|\mathbf{A}_1 - \mathbf{P}_1\| = \hat{\sigma}_{K+1} + O_P(\sqrt{\rho n}) = O_P(\sqrt{\rho n}),$$

where $\hat{\sigma}_{K+1}$ is the $(K+1)$-th eigenvalue of $\mathbf{A}_1$ and is $O_P(\sqrt{\rho n})$ by Weyl's inequality.

Since $\hat{\mathbf{A}}_1$ and $\mathbf{P}_1$ have rank $K$, then by Fact C.1,

$$\left\|\hat{\mathbf{A}}_1 - \mathbf{P}_1\right\|_F \leq \sqrt{2K}\left\|\hat{\mathbf{A}}_1 - \mathbf{P}_1\right\| = O_P(\sqrt{K\rho n}).$$

□

**Lemma C.6** (Concentration of degrees)**.** Denote $\beta_{\min} = \min_a \mathbf{B}_{aa}$. Let $\mathbf{P} = \rho\mathbf{\Theta}^{(1)}\mathbf{B}{\mathbf{\Theta}^{(2)}}^T$, where $\rho$, $\mathbf{B}$, $\mathbf{\Theta}^{(1)} \in \mathbb{R}^{\frac{n}{2}\times K}$, and $\mathbf{\Theta}^{(2)} \in \mathbb{R}^{\frac{n}{2}\times K}$ follow the restrictions of MMSB model. Let $\mathbf{D}$ and $\boldsymbol{\mathcal{D}}$ be diagonal matrices representing the sample and population node degrees. Then

$$\boldsymbol{\mathcal{D}}_{ii} = O_P(\rho n/K),\ \boldsymbol{\mathcal{D}}_{ii} = \Omega(\beta_{\min}\rho n/K),\ \text{ and }\ |\mathbf{D}_{ii} - \boldsymbol{\mathcal{D}}_{ii}| = O_P(\sqrt{\rho n\log n/K})$$

with probability larger than $1 - O_P(1/n^3)$.

*Proof.* $\forall i \in [\frac{n}{2}]$, we have

$$\begin{aligned}
\boldsymbol{\mathcal{D}}_{ii} &= \sum_{j=1}^{n/2} P_{ij} = \sum_{j=1}^{n/2}\sum_{\ell=1}^{K}\rho\theta_{i\ell}^{(1)}\mathbf{B}_{\ell\ell}\theta_{j\ell}^{(2)} \leq \sum_{\ell=1}^{K}\rho\theta_{i\ell}^{(1)}\sum_{j=1}^{n/2}\theta_{j\ell}^{(2)} = \rho\sum_{\ell=1}^{K}\theta_{i\ell}^{(1)}\sum_{j=1}^{n/2}\theta_{j\ell}^{(2)} && (\max_a \mathbf{B}_{aa} = 1 \text{ by definition})\\
&= \rho n\frac{\sum_{\ell=1}^{K}\theta_{i\ell}^{(1)}\alpha_l}{\alpha_0}\left(\frac{1}{2} + O_P\left(\sqrt{\frac{\alpha_0}{\alpha_l}\frac{\log n}{n}}\right)\right) && (\text{from Lemma C.1})\\
&= \frac{\rho n}{2K}\left(1 + O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right), && (\text{when } \alpha_k = \tfrac{\alpha_0}{K}, \forall k \in [K])
\end{aligned}$$

so $\boldsymbol{\mathcal{D}}_{ii} = O_P(\rho n/K)$.

Similarly,

$$\begin{aligned}
\boldsymbol{\mathcal{D}}_{ii} &\geq \sum_{\ell=1}^{K}\beta_{\min}\rho\theta_{i\ell}^{(1)}\sum_{j=1}^{n/2}\theta_{j\ell}^{(2)} = \beta_{\min}\rho\sum_{\ell=1}^{K}\theta_{i\ell}^{(1)}\sum_{j=1}^{n/2}\theta_{j\ell}^{(2)}\\
&= \beta_{\min}\rho n\frac{\sum_{\ell=1}^{K}\theta_{i\ell}^{(1)}\alpha_l}{\alpha_0}\left(\frac{1}{2} + O_P\left(\sqrt{\frac{\alpha_0}{\alpha_l}\frac{\log n}{n}}\right)\right) && (\text{from Lemma C.1})\\
&= \frac{\beta_{\min}}{2K}\rho n\left(1 + O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right), && (\text{when } \alpha_l = \tfrac{\alpha_0}{K}, \forall l)
\end{aligned}$$

so $\boldsymbol{\mathcal{D}}_{ii} = \Omega(\beta_{\min}\rho n/K)$.

Then using Chernoff bound, we have

$$\mathrm{P}\left(|\mathbf{D}_{ii} - \boldsymbol{\mathcal{D}}_{ii}| > \epsilon\boldsymbol{\mathcal{D}}_{ii}]\right) \leq \exp\left(-\frac{\epsilon^2\boldsymbol{\mathcal{D}}_{ii}}{3}\right),$$

so when $\epsilon = O_P\left(3\sqrt{\frac{K\log n}{\rho n}}\right)$, $|\mathbf{D}_{ii} - \boldsymbol{\mathcal{D}}_{ii}| \leq \epsilon\boldsymbol{\mathcal{D}}_{ii} = O_P(\sqrt{\rho n\log n/K})$ with probability at least $1 - 1/n^3$. Note we have used Lemma C.1, so in total it is with probability larger than $1 - O_P(1/n^3)$.

□

**Lemma C.7.** Denote $\beta_{\min} = \min_a \mathbf{B}_{aa}$. If $\rho n = \Omega(\log n)$, then

$$\lambda_K(\mathbf{P}_1) = \Omega(\beta_{\min}\rho n/K), \quad \lambda_1(\mathbf{P}_1) = O_P\left(\rho n/K\right)$$

and

$$\lambda_K(\mathbf{A}_1) = \Omega(\beta_{\min}\rho n/K), \quad \lambda_1(\mathbf{A}_1) = O_P\left(\rho n/K\right)$$

with probability larger than $1 - O_P(K^2/n^3)$.

*Proof.* For conciseness, we will omit the subscript 1 (see Lemma C.5) in the following proof without loss of generality.

The $K$-th eigenvalue of $\mathbf{P}$ is

$$\lambda_K(\mathbf{P}) = \lambda_K(\rho\boldsymbol{\Theta}\mathbf{B}\boldsymbol{\Theta}^T) = \lambda_K(\rho\boldsymbol{\Theta}\mathbf{B}^{1/2}\mathbf{B}^{1/2}\boldsymbol{\Theta}^T) = \lambda_K(\rho\mathbf{B}^{1/2}\boldsymbol{\Theta}^T\boldsymbol{\Theta}\mathbf{B}^{1/2}).$$

Here we consider $\boldsymbol{\theta}_i$ as a random variable. Denote

$$\hat{\mathbf{M}} = \frac{1}{n/2}\rho\mathbf{B}^{1/2}\boldsymbol{\Theta}^T\boldsymbol{\Theta}\mathbf{B}^{1/2} = \frac{1}{n/2}\sum_{i=1}^{n/2}\rho\mathbf{B}^{1/2}\boldsymbol{\theta}_i^T\boldsymbol{\theta}_i\mathbf{B}^{1/2},$$

then $\hat{\mathbf{M}}_{ab} = \frac{1}{n/2}\sqrt{\beta_a\beta_b}\sum_{i=1}^{n/2}\rho\theta_{ia}\theta_{ib}$.

Consider $\boldsymbol{\theta}_i \sim \mathrm{Dirichlet}(\boldsymbol{\alpha})$, then

$$\mathrm{E}[\theta_{ia}\cdot\theta_{ib}] = \begin{cases}\mathrm{Cov}[\theta_{ia},\theta_{ib}] + \mathrm{E}[\theta_{ia}]\cdot\mathrm{E}[\theta_{ib}] = \frac{\alpha_a\alpha_b}{\alpha_0(\alpha_0+1)}, & \text{if } a \neq b\\ \mathrm{Var}[\theta_{ia}] + \mathrm{E}^2[\theta_{ia}] = \frac{\alpha_a(\alpha_a+1)}{\alpha_0(\alpha_0+1)}, & \text{if } a = b\end{cases}$$

so $\mathrm{E}[\hat{\mathbf{M}}_{ab}] = \sqrt{\beta_a\beta_b}\rho\mathrm{E}[\theta_{ia}\cdot\theta_{ib}] \leq \rho$. And

$$\mathrm{E}[\hat{\mathbf{M}}] = \rho(\mathrm{diag}(\mathbf{B}\boldsymbol{\alpha}) + \mathbf{B}^{1/2}\boldsymbol{\alpha}\boldsymbol{\alpha}^T\mathbf{B}^{1/2})/(\alpha_0(\alpha_0+1)).$$

Using Chernoff bound, we have

$$\mathrm{P}\left(\left|\hat{\mathbf{M}}_{ab} - \mathrm{E}[\hat{\mathbf{M}}_{ab}]\right| > \epsilon\mathrm{E}[\hat{\mathbf{M}}_{ab}]\right) \leq \exp\left(-\frac{\epsilon^2\frac{n}{2}\mathrm{E}[\hat{\mathbf{M}}_{ab}]}{3}\right) \leq \exp\left(-\frac{\epsilon^2\rho n}{6}\right),$$

so when $\epsilon = O_P\left(\sqrt{\frac{18\log n}{\rho n}}\right)$, $\left|\hat{\mathbf{M}}_{ab} - \mathrm{E}[\hat{\mathbf{M}}_{ab}]\right| \leq \epsilon\mathrm{E}[\hat{\mathbf{M}}_{ab}]$ with probability larger than $1 - 1/n^3$. Thus

$$\left\|\hat{\mathbf{M}} - \mathrm{E}[\hat{\mathbf{M}}]\right\| \leq \left\|\hat{\mathbf{M}} - \mathrm{E}[\hat{\mathbf{M}}]\right\|_F \leq \sqrt{K^2\epsilon^2\mathrm{E}^2[\hat{\mathbf{M}}_{ab}]} \leq K\epsilon\rho.$$

Note that

$$\begin{aligned}\lambda_K(\mathrm{E}[\hat{\mathbf{M}}]) &= \rho\lambda_K(\mathrm{diag}(\mathbf{B}\boldsymbol{\alpha}) + \mathbf{B}^{1/2}\boldsymbol{\alpha}\boldsymbol{\alpha}^T\mathbf{B}^{1/2})/(\alpha_0(\alpha_0+1))\\ &\geq \rho\left(\lambda_K(\mathrm{diag}(\mathbf{B}\boldsymbol{\alpha})) + \lambda_K(\mathbf{B}^{1/2}\boldsymbol{\alpha}\boldsymbol{\alpha}^T\mathbf{B}^{1/2})\right)/(\alpha_0(\alpha_0+1))\\ &= \rho\left(\min_a \beta_a\alpha_a + 0\right)/(\alpha_0(\alpha_0+1))\\ &= \rho\frac{\min_a\beta_a\alpha_a}{\alpha_0(\alpha_0+1)}\\ &= \frac{\beta_{\min}\rho}{K(\alpha_0+1)}, & \text{(when } \alpha_a = \tfrac{\alpha_0}{K}, \forall a)\end{aligned}$$

the first inequality is by definition of the smallest eigenvalue and property of $\min$ function; the second equality is by the smallest eigenvalue of a $K \times K$ rank-1 matrix ($K > 1$) is 0.

By Weyl's inequality,

$$\left|\lambda_K(\hat{\mathbf{M}}) - \lambda_K(\mathrm{E}[\hat{\mathbf{M}}])\right| \leq \left\|\hat{\mathbf{M}} - \mathrm{E}[\hat{\mathbf{M}}]\right\| = O_P\left(K\sqrt{\frac{\rho\log n}{n}}\right),$$

so

$$\lambda_K(\mathbf{P}) = \frac{n}{2}\lambda_K(\hat{\mathbf{M}}) \geq \frac{n}{2}\left(\frac{\beta_{\min}\rho}{K(\alpha_0+1)} - O_P\left(K\sqrt{\frac{\rho\log n}{n}}\right)\right)$$

with probability larger than $1 - K^2/n^3$, and thus $\lambda_K(\mathbf{P}) = \Omega(\beta_{\min}\rho n/K)$.

With similar argument we can get

$$\begin{aligned}\lambda_1(\mathrm{E}[\hat{\mathbf{M}}]) &\le \rho\left(1 + \frac{\alpha_0}{K}\|\boldsymbol{\beta}\|_1\right)/(K(\alpha_0+1)) \\ &\le \rho\left(1+\alpha_0\right)/(K(\alpha_0+1)) \\ &= \frac{\rho}{K},\end{aligned}$$

then $\lambda_1\left(\mathbf{P}\right) = O_P\left(\rho n/K\right)$.

From Weyl's inequality, we have

$$|\lambda_i(\mathbf{A}) - \lambda_i(\mathbf{P})| \le \|\mathbf{A} - \mathbf{P}\| = O_P\left(\sqrt{\rho n}\right),$$

so

$$\begin{aligned}\lambda_K(\mathbf{A}) \ge \lambda_K(\mathbf{P}) - O_P\left(\sqrt{\rho n}\right) &\Longrightarrow \lambda_K(\mathbf{A}) = \Omega(\beta_{\min}\rho n/K) \\ \lambda_1(\mathbf{A}) \le \lambda_K(\mathbf{P}) + O_P\left(\sqrt{\rho n}\right) &\Longrightarrow \lambda_1(\mathbf{A}) = O_P\left(\rho n/K\right).\end{aligned}$$

□

**Lemma C.8.** If $\rho n = \Omega(\log n)$, $\exists$ orthogonal matrix $\hat{\mathbf{O}}_1 \in \mathbb{R}^{K\times K}$,

$$\left\|\hat{\mathbf{V}}_1 - \mathbf{V}_1\hat{\mathbf{O}}_1\right\|_F = O_P\left(\frac{K^{3/2}}{\beta_{\min}\sqrt{\rho n}}\right)$$

with probability larger than $1 - O_P(K^2/n^3)$.

*Proof.* From Lemma C.7 we know that

$$\lambda_K(\mathbf{P}_1) = \Omega(\beta_{\min}\rho n/K)$$

with probability larger than $1 - O_P(K^2/n^3)$. Because $\mathbf{P}_1$ has rank $K$, its $K+1$ eigenvalue is 0, and the gap between the $K$-th and $(K+1)$-th eigenvalue of $\mathbf{P}_1$ is $\delta = \Omega(\beta_{\min}\rho n/K)$. Using variant of Davis-Kahan's theorem (Lemma C.3), setting $r = 1$, $s = K$, then $d = K$ is the interval corresponding to the first $K$ principle eigenvalues of $\mathbf{P}_1$, we have $\exists\, \hat{\mathbf{O}}_1 \in \mathbb{R}^{K\times K}$,

$$\left\|\hat{\mathbf{V}}_1 - \mathbf{V}_1\hat{\mathbf{O}}_1\right\|_F \le \frac{2^{3/2}\min\left(\sqrt{K}\left\|\hat{\mathbf{A}}_1 - \mathbf{P}_1\right\|, \left\|\hat{\mathbf{A}}_1 - \mathbf{P}_1\right\|_F\right)}{\delta},$$

using Lemma C.5,

$$\left\|\hat{\mathbf{V}}_1 - \mathbf{V}_1\hat{\mathbf{O}}_1\right\|_F = O_P\left(\frac{2^{3/2}\sqrt{K\rho n}}{\beta_{\min}\rho n/K}\right) = O_P\left(\frac{K^{3/2}}{\beta_{\min}\sqrt{\rho n}}\right)$$

with probability larger than $1 - O_P(K^2/n^3)$.

□

**Lemma C.9.** If $\rho n = \Omega(\log n)$, then the orthogonal matrix $\hat{\mathbf{O}}_1 \in \mathbb{R}^{K\times K}$ of Lemma C.8 satisfies

$$\left\|\hat{\mathbf{E}}_1 - \hat{\mathbf{O}}_1^T\mathbf{E}_1\hat{\mathbf{O}}_1\right\|_F = O_P(\sqrt{K\rho n}/\beta_{\min})$$

with probability larger than $1 - O_P(K^2/n^3)$.

*Proof.* We have

$$\left\|\hat{\mathbf{V}}_1\hat{\mathbf{E}}_1\hat{\mathbf{V}}_1^T - \mathbf{V}_1\mathbf{E}_1\mathbf{V}_1^T\right\|_F = \left\|\hat{\mathbf{A}}_1 - \mathbf{P}_1\right\|_F = O_P(\sqrt{K\rho n})$$

with probability larger than $1 - 1/n^3$ by Lemma C.5, and

$$\left\|\hat{\mathbf{V}}_1 - \mathbf{V}_1\hat{\mathbf{O}}_1\right\|_F = O_P\left(\frac{K^{3/2}}{\beta_{\min}\sqrt{\rho n}}\right)$$

with probability larger than $1 - O_P(K^2/n^3)$ by Lemma C.8. Also,

$$\hat{\mathbf{V}}_1\hat{\mathbf{E}}_1\hat{\mathbf{V}}_1^T - \mathbf{V}_1\mathbf{E}_1\mathbf{V}_1^T = \hat{\mathbf{V}}_1\left(\hat{\mathbf{E}}_1 - \hat{\mathbf{O}}_1^T\mathbf{E}_1\hat{\mathbf{O}}_1\right)\hat{\mathbf{V}}_1^T + \hat{\mathbf{V}}_1\hat{\mathbf{O}}_1^T\mathbf{E}_1\left(\hat{\mathbf{O}}_1\hat{\mathbf{V}}_1^T - \mathbf{V}_1^T\right) + \left(\hat{\mathbf{V}}_1\hat{\mathbf{O}}_1^T - \mathbf{V}_1\right)\mathbf{E}_1\mathbf{V}_1^T.$$

So

$$\begin{aligned}
&\left\|\hat{\mathbf{E}}_1 - \hat{\mathbf{O}}_1^T\mathbf{E}_1\hat{\mathbf{O}}_1\right\|_F = \left\|\hat{\mathbf{V}}_1\left(\hat{\mathbf{E}}_1 - \hat{\mathbf{O}}_1^T\mathbf{E}_1\hat{\mathbf{O}}_1\right)\hat{\mathbf{V}}_1^T\right\|_F \\
\le&\left\|\hat{\mathbf{V}}_1\hat{\mathbf{E}}_1\hat{\mathbf{V}}_1^T - \mathbf{V}_1\mathbf{E}_1\mathbf{V}_1^T\right\|_F + \left\|\hat{\mathbf{V}}_1\hat{\mathbf{O}}_1^T\mathbf{E}_1\left(\hat{\mathbf{O}}_1\hat{\mathbf{V}}_1^T - \mathbf{V}_1^T\right)\right\|_F + \left\|\left(\hat{\mathbf{V}}_1\hat{\mathbf{O}}_1^T - \mathbf{V}_1\right)\mathbf{E}_1\mathbf{V}_1^T\right\|_F \\
\le& O_P(\sqrt{K\rho n}) + 2\left\|\mathbf{E}_1\right\|\left\|\hat{\mathbf{V}}_1 - \mathbf{V}_1\hat{\mathbf{O}}_1^T\right\|_F \\
=& O_P(\sqrt{K\rho n}) + O_P\left(\frac{\rho n}{K}\cdot\frac{K^{3/2}}{\beta_{\min}\sqrt{\rho n}}\right) \\
=& O_P\left(\sqrt{K\rho n}/\beta_{\min}\right),
\end{aligned}$$

with probability larger than $1 - O_P(K^2/n^3)$. □

**Lemma C.10.** If $\rho n = \Omega(\log n)$, then $\exists$ an orthogonal matrix $\mathbf{R}_1 \in \mathbb{R}^{K\times K}$, together with the orthogonal matrix $\hat{\mathbf{O}} \in \mathbb{R}^{K\times K}$ of Lemma C.8 satisfies

$$\left\|\mathbf{R}_1\mathbf{E}_1^{1/2}\hat{\mathbf{O}}_1 - \hat{\mathbf{E}}_1^{1/2}\right\|_F = O_P\left(K^{3/2}/\beta_{\min}^2\right)$$

with probability larger than $1 - O_P(K^2/n^3)$.

*Proof.* From Lemma C.9 we have

$$\left\|\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1\hat{\mathbf{O}}_1^T - \mathbf{E}_1\right\|_F = \left\|\hat{\mathbf{E}}_1 - \hat{\mathbf{O}}_1^T\mathbf{E}_1\hat{\mathbf{O}}_1\right\|_F = O_P\left(\sqrt{K\rho n}/\beta_{\min}\right).$$

with probability larger than $1 - O_P(K^2/n^3)$.

By Lemma C.4, there exists an orthogonal matrix $\mathbf{R}_1 \in \mathbb{R}^{d\times d}$ such that:

$$\begin{aligned}
\left\|\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1^{1/2}\mathbf{R}_1 - \mathbf{E}_1^{1/2}\right\|_F &\le \frac{\sqrt{K}\left\|\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1\hat{\mathbf{O}}_1^T - \mathbf{E}_1\right\|\left(\sqrt{\left\|\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1\hat{\mathbf{O}}_1^T\right\|} + \sqrt{\|\mathbf{E}_1\|}\right)}{\lambda_K(\mathbf{E})} \\
&\le \frac{\sqrt{K}\cdot O_P\left(\sqrt{K\rho n}/\beta_{\min}\right)\left(O_P\left(\sqrt{\frac{\rho n}{K}}\right) + O_P\left(\sqrt{\frac{\rho n}{K}}\right)\right)}{\Omega(\beta_{\min}\rho n/K)} && \text{(from Lemma C.7)} \\
&\le O_P\left(K^{3/2}/\beta_{\min}^2\right).
\end{aligned}$$

Note that

$$\left\|\mathbf{R}_1\mathbf{E}_1^{1/2}\hat{\mathbf{O}}_1 - \hat{\mathbf{E}}_1^{1/2}\right\|_F = \left\|\mathbf{E}_1^{1/2} - \mathbf{R}_1^T\hat{\mathbf{E}}_1^{1/2}\hat{\mathbf{O}}_1^T\right\|_F = \left\|\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1^{1/2}\mathbf{R}_1 - \mathbf{E}_1^{1/2}\right\|_F,$$

so

$$\left\|\mathbf{R}_1\mathbf{E}_1^{1/2}\hat{\mathbf{O}}_1 - \hat{\mathbf{E}}_1^{1/2}\right\|_F = O_P\left(K^{3/2}/\beta_{\min}^2\right).$$

□

*Proof of Lemma 4.1.* Note that if $\boldsymbol{\alpha} = u\mathbf{1}$, $u > 0$, then

$$\boldsymbol{\mathcal{D}}_{21}(i,i) = \sum_{a\in[K],j\in\mathcal{S}} \rho\theta_{ia}\mathbf{B}_{aa}\theta_{ja} = \sum_{a\in[K]} \rho\theta_{ia}\mathbf{B}_{aa}\sum_{j\in\mathcal{S}}\theta_{ja} = \frac{\rho n}{2K}\sum_{a\in[K]}\theta_{ia}\mathbf{B}_{aa}\left(1 \pm O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right).$$

(by Lemma C.1)

Because

$$\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\boldsymbol{\mathcal{D}}_{21}^{-1/2}\boldsymbol{\Theta}_2\mathbf{B}^{1/2}\right\|_F^2 = \frac{\rho\left\|\mathbf{e}_i^T\boldsymbol{\Theta}_2\mathbf{B}^{1/2}\right\|_F^2}{\boldsymbol{\mathcal{D}}_{21}(i,i)} \in \frac{\rho\sum_{a\in[K]}\theta_{ia}^2\mathbf{B}_{aa}}{\frac{\rho n}{2K}\sum_{a\in[K]}\theta_{ia}\mathbf{B}_{aa}}\left[\frac{1}{1+O_P(\sqrt{K\log n/n})}, \frac{1}{1-O_P(\sqrt{K\log n/n})}\right]$$
$$= \frac{2K}{n}\cdot\frac{\sum_{a\in[K]}\theta_{ia}^2\mathbf{B}_{aa}}{\sum_{a\in[K]}\theta_{ia}\mathbf{B}_{aa}}\left[1-O_P(\sqrt{K\log n/n}), 1+O_P(\sqrt{K\log n/n})\right],$$

also note that

$$\frac{2K}{n}\cdot\frac{\sum_{a\in[K]}\theta_{ia}^2\mathbf{B}_{aa}}{\sum_{a\in[K]}\theta_{ia}\mathbf{B}_{aa}} \le \frac{2K}{n}\cdot\max_a\theta_{ia} \le \frac{2K}{n},$$

where the first inequality is an equality when $\forall k\in[K]$, $\theta_{ik} = \max_a\theta_{ia}$ or 0. The second inequality becomes an equality when $\max_a\theta_{ia} = 1$ (i.e. $i$ is a pure node). This implies that the LHS of the above equation equals $2K/n$ if and only if $i$ corresponds to a pure node. Then we have

$$\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\boldsymbol{\mathcal{D}}_{21}^{-1/2}\boldsymbol{\Theta}_2\mathbf{B}^{1/2}\right\|_F^2 \le \frac{2K}{n}\cdot\max_a\theta_{ia}\left(1+O_P(\sqrt{K\log n/n})\right),$$

and

$$\left|\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\boldsymbol{\mathcal{D}}_{21}^{-1/2}\boldsymbol{\Theta}_2\mathbf{B}^{1/2}\right\|_F^2 - \frac{2K}{n}\cdot\frac{\sum_{a\in[K]}\theta_{ia}^2\mathbf{B}_{aa}}{\sum_{a\in[K]}\theta_{ia}\mathbf{B}_{aa}}\right| = \frac{2K}{n}\cdot\frac{\sum_{a\in[K]}\theta_{ia}^2\mathbf{B}_{aa}}{\sum_{a\in[K]}\theta_{ia}\mathbf{B}_{aa}}\cdot O_P(\sqrt{K\log n/n})$$
$$= O_P\left(\frac{2K}{n}\cdot\sqrt{K\log n/n}\right)$$

with probability larger than $1 - O(1/n^3)$.

So $\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\boldsymbol{\mathcal{D}}_{21}^{-1/2}\boldsymbol{\Theta}_2\mathbf{B}^{1/2}\right\|_F^2$ concentrates around $\frac{2K}{n}$ for pure nodes. Note that we implicitly assume that the impure nodes have $\max_a\theta_{ia}$ bounded away from one, and hence have norm bounded away from $2K/n$.

□

*Proof of Theorem 4.2.* Denote $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}(\mathcal{S},:)$ and $\boldsymbol{\Theta}_2 = \boldsymbol{\Theta}(\bar{\mathcal{S}},:)$. Denote $\mathbf{A}_{12} = \mathbf{A}(\mathcal{S},\bar{\mathcal{S}})$ and $\mathbf{A}_{21} = \mathbf{A}(\bar{\mathcal{S}},\mathcal{S})$, $\mathbf{D}_{12}$ and $\mathbf{D}_{21}$ are the (row) degree matrix of $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$. GeoNMF projects $\mathbf{D}_{21}^{-1/2}\mathbf{A}_{21}$ onto $\hat{\mathbf{V}}_1\hat{\mathbf{E}}_1^{-1/2}$, and $\mathbf{D}_{12}^{-1/2}\mathbf{A}_{12}$ onto $\hat{\mathbf{V}}_2\hat{\mathbf{E}}_2^{-1/2}$.

Now, $\mathbf{V}_1\mathbf{E}_1\mathbf{V}_1^T = \mathbf{P}_1 = \rho\boldsymbol{\Theta}_1\mathbf{B}\boldsymbol{\Theta}_1^T$, with both $\mathbf{E}_1$ and $\mathbf{B}$ diagonal. This imples that there exists an orthogonal matrix $\mathbf{Q}_1$ such that $\mathbf{V}_1\mathbf{E}_1^{1/2}\mathbf{Q}_1 = \sqrt{\rho}\cdot\boldsymbol{\Theta}_1\mathbf{B}^{1/2}$ (by Lemma A.1 of (Tang et al., 2013)).

Also, as shown in Lemmas C.8 and C.10, there exists orthogonal matrices $\hat{\mathbf{O}}_1$ and and $\mathbf{R}_1$ such that

$$\left\|\hat{\mathbf{V}}_1 - \mathbf{V}_1\hat{\mathbf{O}}_1\right\|_F = O_P\left(\frac{K^{3/2}}{\beta_{\min}\sqrt{\rho n}}\right), \quad \text{and} \quad \left\|\mathbf{R}_1\mathbf{E}_1^{1/2}\hat{\mathbf{O}}_1 - \hat{\mathbf{E}}_1^{1/2}\right\|_F = O_P\left(K^{3/2}/\beta_{\min}^2\right)$$

with probability larger than $1 - O(K^2/n^3)$.

Then we have:

$$\begin{aligned}
&\left\|\mathbf{e}_i^T\mathbf{P}_{21}\mathbf{V}_1\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1^{-1/2}-\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F\\
=&\left\|\rho\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}\mathbf{\Theta}_1^T\mathbf{V}_1\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1^{-1/2}-\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F && \text{(by } \mathbf{P}_{21}=\rho\mathbf{\Theta}_2\mathbf{B}\mathbf{\Theta}_1^T\text{)}\\
=&\left\|\rho\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\left(\mathbf{B}^{1/2}\mathbf{\Theta}_1^T\right)\mathbf{V}_1\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1^{-1/2}-\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F\\
=&\sqrt{\rho}\cdot\left\|\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\left(\mathbf{V}_1\mathbf{E}_1^{1/2}\mathbf{Q}_1\right)^T\mathbf{V}_1\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1^{-1/2}-\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F && \text{(by Lemma A.1 of (Tang et al., 2013))}\\
=&\sqrt{\rho}\cdot\left\|\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\left(\mathbf{R}_1\mathbf{E}_1^{1/2}\hat{\mathbf{O}}_1\right)\hat{\mathbf{E}}_1^{-1/2}-\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F\\
=&\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\left(\mathbf{R}_1\mathbf{E}_1^{1/2}\hat{\mathbf{O}}_1-\hat{\mathbf{E}}_1^{1/2}\right)\hat{\mathbf{E}}_1^{-1/2}\right\|_F\\
\le&\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F\left\|\mathbf{R}_1\mathbf{E}_1^{1/2}\hat{\mathbf{O}}_1-\hat{\mathbf{E}}_1^{1/2}\right\|\left\|\hat{\mathbf{E}}_1^{-1/2}\right\|\\
\le&\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F\cdot O_P\left(K^{3/2}/\beta_{\min}^2\right)\cdot O_P\left(\sqrt{\frac{K}{\beta_{\min}\rho n}}\right) && \text{(by Lemmas 4.1, C.6, C.7 and C.10)}
\end{aligned}$$

$$\Longrightarrow\left\|\mathbf{e}_i^T\mathbf{P}_{21}\mathbf{V}_1\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1^{-1/2}-\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F=\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F\cdot O_P\left(\frac{K^2}{\beta_{\min}^{5/2}\sqrt{\rho n}}\right). \tag{2}$$

Now that

$$\begin{aligned}
&\left\|\mathbf{e}_i^T\mathbf{D}_{21}^{-1/2}\mathbf{A}_{21}\hat{\mathbf{V}}_1\hat{\mathbf{E}}_1^{-1/2}-\sqrt{\rho}\cdot\mathbf{e}_i^T\boldsymbol{\mathcal{D}}_{21}^{-1/2}\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F\\
\le&\left\|\mathbf{e}_i^T\mathbf{A}_{21}\hat{\mathbf{V}}_1\hat{\mathbf{E}}_1^{-1/2}\left(1+O_P\left(\sqrt{K\log n/n\rho}\right)\right)-\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F/\sqrt{\boldsymbol{\mathcal{D}}_{21}(i,i)} && \text{(by Lemma C.6)}\\
\le&\frac{\left(1+O_P\left(\sqrt{K\log n/n\rho}\right)\right)\cdot\left\|\mathbf{e}_i^T\left[(\mathbf{A}_{21}-\mathbf{P}_{21})\hat{\mathbf{V}}_1+\mathbf{P}_{21}\left(\hat{\mathbf{V}}_1-\mathbf{V}_1\hat{\mathbf{O}}_1\right)+\mathbf{P}_{21}\mathbf{V}_1\hat{\mathbf{O}}_1\right]\hat{\mathbf{E}}_1^{-1/2}-\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F}{\sqrt{\boldsymbol{\mathcal{D}}_{21}(i,i)}}\\
&+O_P\left(\sqrt{K\log n/n\rho}\right)\cdot\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F/\sqrt{\boldsymbol{\mathcal{D}}_{21}(i,i)}\\
\le&\left(1+O_P\left(\sqrt{K\log n/n\rho}\right)\right)\cdot\left\{\left\|\mathbf{e}_i^T(\mathbf{A}_{21}-\mathbf{P}_{21})\hat{\mathbf{V}}_1\hat{\mathbf{E}}_1^{-1/2}\right\|_F+\left\|\mathbf{e}_i^T\mathbf{P}_{21}\left(\hat{\mathbf{V}}_1-\mathbf{V}_1\hat{\mathbf{O}}_1\right)\hat{\mathbf{E}}_1^{-1/2}\right\|_F\right.\\
&\left.+\left\|\mathbf{e}_i^T\mathbf{P}_{21}\mathbf{V}_1\hat{\mathbf{O}}_1\hat{\mathbf{E}}_1^{-1/2}-\sqrt{\rho}\cdot\mathbf{e}_i^T\mathbf{\Theta}_2\mathbf{B}^{1/2}\mathbf{Q}_1^T\mathbf{R}_1^T\right\|_F\right\}/\sqrt{\boldsymbol{\mathcal{D}}_{21}(i,i)}+O_P\left(K\sqrt{\log n/n^2\rho}\right)\\
& && \text{(by Lemmas 4.1 and C.6)}\\
\le&\left(1+O_P\left(\frac{1}{\beta_{\min}}\sqrt{K\log n/n\rho}\right)\right)\cdot\left\{O_P\left(\sqrt{K\log n}\right)\cdot O_P\left(\sqrt{\frac{K}{\beta_{\min}\rho n}}\right)\right. && \text{(by Azuma's and Lemma C.7)}\\
&\left.+O_P\left(\sqrt{\frac{\rho n}{K}}\right)\cdot O_P\left(\frac{K^{3/2}}{\beta_{\min}\sqrt{\rho n}}\right)\cdot O_P\left(\sqrt{\frac{K}{\beta_{\min}\rho n}}\right)\right\}/\sqrt{\beta_{\min}\rho n/K} && \text{(by Lemmas C.6, C.8, and Eq. (2))}\\
&+\sqrt{\frac{2K}{n}}\cdot O_P\left(\frac{K^2}{\beta_{\min}^{5/2}\sqrt{\rho n}}\right)+O_P\left(K\sqrt{\log n/n^2\rho}\right)\\
=&O_P\left(\frac{K^{5/2}\sqrt{\log n}}{\beta_{\min}^{5/2}\rho n}\right).
\end{aligned}$$

In the last step we use the fact that $\left\|\mathbf{e}_i^T(\mathbf{A}_{21}-\mathbf{P}_{21})\hat{\mathbf{V}}_1\right\|_F^2$ is a sum of $K$ projections of $\mathbf{e}_i^T(\mathbf{A}_{21}-\mathbf{P}_{21})$ on a fixed unit vector (since the eigenvectors come from the different partition of the graph). Now Azuma's inequality gives $\left\|\mathbf{e}_i^T(\mathbf{A}_{21}-\mathbf{P}_{21})\hat{\mathbf{V}}_1\right\|_F=O_P(\sqrt{K\log n})$ with probability larger than $1-O(1/n^3)$.

Now as

$$\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\boldsymbol{\mathcal{D}}_{21}^{-1/2}\boldsymbol{\Theta}_2\mathbf{B}^{1/2}\right\|_F^2=\frac{\rho\left\|\mathbf{e}_i^T\boldsymbol{\Theta}_2\mathbf{B}^{1/2}\right\|_F^2}{\boldsymbol{\mathcal{D}}_{21}(i,i)}=\frac{\rho\cdot\mathbf{e}_i^T\boldsymbol{\Theta}_2\mathbf{B}\boldsymbol{\Theta}_2^T\mathbf{e}_i}{\boldsymbol{\mathcal{D}}_{21}(i,i)}$$
$$=\frac{\mathbf{P}_2(i,i)}{\boldsymbol{\mathcal{D}}_{21}(i,i)}=\Omega\left(\frac{\rho}{\rho n/K}\right)=O_P\left(\frac{K}{n}\right),$$

let $\boldsymbol{O}=\mathbf{Q}_1^T\mathbf{R}_1^T$, then $\forall i$,

$$\frac{\left\|\mathbf{e}_i^T\mathbf{D}_{21}^{-1/2}\mathbf{A}_{21}\hat{\mathbf{V}}_1\hat{\mathbf{E}}_1^{-1/2}-\sqrt{\rho}\cdot\mathbf{e}_i^T\boldsymbol{\mathcal{D}}_{21}^{-1/2}\boldsymbol{\Theta}_2\mathbf{B}^{1/2}\boldsymbol{O}\right\|_F}{\left\|\sqrt{\rho}\cdot\mathbf{e}_i^T\boldsymbol{\mathcal{D}}_{21}^{-1/2}\boldsymbol{\Theta}_2\mathbf{B}^{1/2}\right\|_F}=O_P\left(\frac{K^{5/2}\sqrt{\log n}}{\beta_{\min}^{5/2}\rho n}\cdot\sqrt{\frac{n}{K}}\right)$$
$$=O_P\left(\frac{K^2\sqrt{\log n}}{\beta_{\min}^{5/2}\rho\sqrt{n}}\right)$$

with probability larger than $1-n\cdot O(K^2/n^3)$=$1-O(K^2/n^2)$. □

## D. Correctness of Pure node clusters

*Proof of Lemma 4.3.* Recall that $\max_i\|\mathbf{X}_i\|$ concentrates around $\sqrt{2K/n}$ and this is achieved at the pure nodes. For ease of analysis let us introduce $\hat{\boldsymbol{Y}}:=\sqrt{n/2K}\hat{\mathbf{X}}$ and $\mathbf{Y}=\sqrt{n/2K}\mathbf{X}$. Recall that from Theorem 4.2 we have entry-wise consistency on $\|\hat{\mathbf{Y}}_i-\mathbf{Y}_i\boldsymbol{O}\|\le\epsilon'=O_P\left(\frac{K^2\sqrt{\log n}}{\beta_{\min}^{5/2}\rho\sqrt{n}}\right)$ with probability larger than $1-O_P(K^2/n^2)$.

Let $\epsilon_{\text{norm}}=O_P\left(\sqrt{\frac{K\log n}{n}}\right)=O_P(\epsilon')$ be the error of the norm of pure nodes in Lemma 4.1. Then $\forall i\in\mathcal{F}$,

$$\|\hat{\mathbf{X}}_i\|\ge(1-\epsilon_0)\max_j\|\hat{\mathbf{X}}_j\|\ge(1-\epsilon_0)(1-\epsilon')\max_j\|\mathbf{X}_j\|\ge(1-\epsilon_0)(1-\epsilon')(1-\epsilon_{\text{norm}})\sqrt{2K/n}.$$

Hence we have a series of inequalities,

$$(1-\epsilon_0)(1-\epsilon')(1-\epsilon_{\text{norm}})\le\|\hat{\mathbf{Y}}_i\|\le\|\hat{\mathbf{Y}}_i-\mathbf{Y}_i\boldsymbol{O}\|+\|\mathbf{Y}_i\|\le\epsilon'+\|\mathbf{Y}_i\|.$$

Hence

$$\|\mathbf{Y}_i\|^2\ge(1-\epsilon_0-2\epsilon'-\epsilon_{\text{norm}})^2\ge1-2(\epsilon_0+2\epsilon'+\epsilon_{\text{norm}})$$

And from the proof of Lemma 4.1,

$$1-2(\epsilon_0+2\epsilon'+\epsilon_{\text{norm}})\le\|\mathbf{Y}_i\|^2\le\frac{\sum_{a\in[K]}\theta_{ia}^2\mathbf{B}_{aa}}{\sum_{a\in[K]}\theta_{ia}\mathbf{B}_{aa}}\le\max_a\theta_{ia}\left(1+\epsilon_{\text{norm}}\right)$$
$$\Longrightarrow\max_a\theta_{ia}\ge1-2(\epsilon_0+2\epsilon'+1.5\epsilon_{\text{norm}})=1-O_P(\epsilon_0+\epsilon')$$

for $\epsilon=2(\epsilon_0+2\epsilon'+1.5\epsilon_{\text{norm}})=O_P(\epsilon_0+\epsilon')$, with probability larger than $1-O_P(K^2/n^2)$.

Note that $\|\mathbf{X}_i\|^2$ of those nearly pure nodes with $\max_a\theta_{ia}\ge1-\epsilon$ also concentrate around $\frac{2K}{n}$. These nearly pure nodes can also be used along with the pure nodes to recover the MMSB model asymptotically correctly. □

**Lemma D.1.** Let $\mathcal{F}$ be the set of nodes with $\|\hat{\mathbf{X}}_i\|\ge(1-\epsilon_0)\max_j\|\hat{\mathbf{X}}_j\|$. Then when $\epsilon_0=O_P(\epsilon')$ and $\epsilon=O_P(\epsilon_0+\epsilon')$ from Lemma 4.3,

$$\min_{i\in\mathcal{F}}\mathbf{D}_2(i,i)=\frac{\rho n}{2K}\left(\beta_{\min}\pm O_P(\epsilon)\right),\quad\text{and}\quad\max_{i\in\mathcal{F}}\mathbf{D}_2(i,i)=\frac{\rho n}{2K}\left(1\pm O_P(\epsilon)\right)$$

with probability larger than $1-O_P(K^2/n^2)$.

*Proof.* From Lemma 4.3 we know that $\forall i \in \mathcal{F}$, $\exists a_i$ that $\theta_{ia_i} \geq 1-\epsilon$, where $\epsilon = O_P(\epsilon_0 + \epsilon') = O_P(\epsilon')$. Then

$$\begin{aligned}
\boldsymbol{\mathcal{D}}_2(i,i) &= \sum_{j=1}^{n/2} \mathbf{P}_{ij} = \sum_{j=1}^{n/2}\sum_{\ell=1}^{K} \rho\theta_{i\ell}^{(2)}\mathbf{B}_{\ell\ell}\theta_{j\ell}^{(2)} = \rho\sum_{\ell=1}^{K}\theta_{i\ell}^{(2)}\mathbf{B}_{\ell\ell}\sum_{j=1}^{n/2}\theta_{j\ell}^{(2)} \\
&= \rho\sum_{\ell=1}^{K}\theta_{i\ell}^{(2)}\mathbf{B}_{\ell\ell}\frac{n}{2K}\left(1\pm O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right) \qquad \text{(from Lemma C.1)} \\
&= \frac{\rho n}{2K}\left((1-O_P(\epsilon))\mathbf{B}_{a_ia_i} + O_P(\epsilon)\right)\left(1\pm O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right) \\
&= \frac{\rho n}{2K}\left(\mathbf{B}_{a_ia_i} \pm O_P(\epsilon)\right).
\end{aligned}$$

Using the proof in Lemma C.6, we have $\mathbf{D}_2(i,i) \in \boldsymbol{\mathcal{D}}_2(i,i)\left[1-O_P\left(\sqrt{\frac{K\log n}{\rho n}}\right), 1+O_P\left(\sqrt{\frac{K\log n}{\rho n}}\right)\right]$, so

$$\mathbf{D}_2(i,i) = \frac{\rho n}{2K}\left(\mathbf{B}_{a_ia_i} \pm O_P(\epsilon)\right).$$

Then

$$\begin{aligned}
\min_{i\in\mathcal{F}} \mathbf{D}_2(i,i) &= \frac{\rho n}{2K}\left(\beta_{\min} \pm O_P(\epsilon)\right), \\
\max_{i\in\mathcal{F}} \mathbf{D}_2(i,i) &= \frac{\rho n}{2K}\left(1 \pm O_P(\epsilon)\right)
\end{aligned}$$

with probability larger than $1-O_P(K^2/n^2)$. □

*Proof of Theorem 4.4.* To prove this theorem, it is equivalent to prove that the upper bound of Euclidean distances within each community's (nearly) pure nodes is far more smaller than the lower bound of Euclidean distances between different communities' (nearly) pure nodes.

Recall that from Lemma 4.3, for $i \neq j \in \mathcal{F}$, $\exists\, a,b \in [K]$, such that $\theta_{ia} \geq 1-\epsilon$ and $\theta_{jb} \geq 1-\epsilon$. Note that $\epsilon = O_P(\epsilon_0+\epsilon')$ for $\epsilon' = O_P\left(\frac{K^2\sqrt{\log n}}{\beta_{\min}^{5/2}\rho\sqrt{n}}\right)$ and $\epsilon_0 = O_P(\epsilon')$.

Using a similar argument as in the proof of Lemma 4.1, we have:

1. if $a \neq b$,

$$\begin{aligned}
\|\mathbf{Y}_i - \mathbf{Y}_j\|_2^2 &\geq \left[\left(\frac{\theta_{ia}\mathbf{B}_{aa}^{1/2}}{\sqrt{\sum_k \theta_{ik}\mathbf{B}_{kk}}} - \frac{\theta_{ja}\mathbf{B}_{aa}^{1/2}}{\sqrt{\sum_k \theta_{jk}\mathbf{B}_{kk}}}\right)^2 + \left(\frac{\theta_{ib}\mathbf{B}_{bb}^{1/2}}{\sqrt{\sum_k \theta_{ik}\mathbf{B}_{kk}}} - \frac{\theta_{jb}\mathbf{B}_{bb}^{1/2}}{\sqrt{\sum_k \theta_{jk}\mathbf{B}_{kk}}}\right)^2\right]\cdot\left(1-O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right)^2 \\
&\geq \left[\mathbf{B}_{aa}\left(\frac{1-\epsilon}{\sqrt{\beta_{\max}}} - \frac{\epsilon}{\sqrt{\beta_{\min}}}\right)^2 + \mathbf{B}_{bb}\left(\frac{1-\epsilon}{\sqrt{\beta_{\max}}} - \frac{\epsilon}{\sqrt{\beta_{\min}}}\right)^2\right]\cdot\left(1-O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right) \\
&\geq \left[2\beta_{\min}\left(\frac{1-\epsilon}{\sqrt{\beta_{\max}}} - \frac{\epsilon}{\sqrt{\beta_{\min}}}\right)^2\right]\cdot\left(1-O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right) \\
&\geq \left\{2\frac{\beta_{\min}}{\beta_{\max}}\left[1-\left(1+\sqrt{\frac{\beta_{\max}}{\beta_{\min}}}\right)\epsilon\right]^2\right\}\cdot\left(1-O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right) \\
&= \left\{2\frac{\beta_{\min}}{\beta_{\max}}\left[1-2\left(1+\sqrt{\frac{\beta_{\max}}{\beta_{\min}}}\right)\epsilon + O_P(\epsilon^2)\right]\right\}\cdot\left(1-O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right) \\
&= 2\frac{\beta_{\min}}{\beta_{\max}}\left(1-O_P\left(\epsilon\sqrt{\frac{\beta_{\max}}{\beta_{\min}}}\right)\right).
\end{aligned}$$

So,

$$\|\mathbf{X}_i - \mathbf{X}_j\|_2 \geq \sqrt{2\frac{\beta_{\min}}{\beta_{\max}}}\sqrt{\frac{2K}{n}}\left(1 - O_P\left(\epsilon\sqrt{\frac{\beta_{\max}}{\beta_{\min}}}\right)\right),$$

and then,

$$\begin{aligned}
\left\|\hat{\mathbf{X}}_i - \hat{\mathbf{X}}_j\right\|_2 &\geq \|\mathbf{X}_i - \mathbf{X}_j\|_2 - \left\|\hat{\mathbf{X}}_i - \mathbf{X}_i\right\|_2 - \left\|\hat{\mathbf{X}}_j - \mathbf{X}_j\right\|_2 \\
&\geq \sqrt{2\frac{\beta_{\min}}{\beta_{\max}}}\sqrt{\frac{2K}{n}}\left(1 - O_P\left(\epsilon\sqrt{\frac{\beta_{\max}}{\beta_{\min}}}\right)\right) - 2\|\mathbf{X}_i\| \cdot \epsilon' \\
&\geq \sqrt{2\frac{\beta_{\min}}{\beta_{\max}}}\sqrt{\frac{2K}{n}}\left(1 - O_P\left(\epsilon\sqrt{\frac{\beta_{\max}}{\beta_{\min}}}\right)\right) - 2\sqrt{\frac{2K}{n}} \cdot (1 + O_P(\epsilon)) \cdot \epsilon' \\
&= 2\sqrt{\frac{K\beta_{\min}}{n}} - O_P\left(\epsilon\sqrt{\frac{K}{n}}\right). \qquad (\beta_{\max} = 1 \text{ by definition})
\end{aligned}$$

2. if $a = b$, first of all we have

$$(1-\epsilon)\beta_a \leq \sum_k \theta_{ik}\mathbf{B}_{kk} \leq \beta_a + \epsilon\sum_{k\neq a}\beta_k,$$

then

$$\begin{aligned}
\|\mathbf{Y}_i - \mathbf{Y}_j\|_2^2 &= \sum_l \left(\frac{\theta_{il}\mathbf{B}_{ll}^{1/2}}{\sqrt{\sum_k \theta_{ik}\mathbf{B}_{kk}}} - \frac{\theta_{jl}\mathbf{B}_{ll}^{1/2}}{\sqrt{\sum_k \theta_{jk}\mathbf{B}_{kk}}}\right)^2 \cdot \left(1 + O_P\left(\sqrt{\frac{K\log n}{n}}\right)\right)^2 \\
&\leq \left[\left(\frac{\mathbf{B}_{aa}^{1/2}}{\sqrt{(1-\epsilon)\beta_a}} - \frac{(1-\epsilon)\mathbf{B}_{aa}^{1/2}}{\sqrt{\beta_a + \epsilon\sum_{k\neq a}\beta_k}}\right)^2 + \sum_{k\neq a}\frac{\beta_{\max}}{\beta_{\min}}\epsilon^2\right] \cdot (1 + O_P(\epsilon)) \\
&= \left\{\left[1 + \frac{\epsilon}{2} + O_P(\epsilon^2) - (1-\epsilon)\left(1 - \frac{\epsilon}{2}\frac{\sum_{k\neq a}\beta_k}{\beta_a} + O_P(\epsilon^2)\right)\right]^2 + (K-1)\frac{\beta_{\max}}{\beta_{\min}}\epsilon^2\right\} \cdot (1 + O_P(\epsilon)) \\
&\leq \left\{\left[\left(\frac{3}{2} + \frac{\sum_{k\neq a}\beta_{\max}}{2\beta_{\min}}\right)\epsilon + O_P(\epsilon^2)\right]^2 + O_P\left(K\frac{\beta_{\max}}{\beta_{\min}}\epsilon^2\right)\right\} \cdot (1 + O_P(\epsilon)) \\
&= O_P\left(K\frac{\beta_{\max}}{\beta_{\min}}\epsilon^2\right).
\end{aligned}$$

So,

$$\|\mathbf{X}_i - \mathbf{X}_j\|_2 \leq \sqrt{\frac{2K}{n}}O_P\left(\epsilon\sqrt{K\frac{\beta_{\max}}{\beta_{\min}}}\right)$$

and then,

$$\begin{aligned}
\left\|\hat{\mathbf{X}}_i - \hat{\mathbf{X}}_j\right\|_2 &\leq \|\mathbf{X}_i - \mathbf{X}_j\|_2 + \left\|\hat{\mathbf{X}}_i - \mathbf{X}_i\right\|_2 + \left\|\hat{\mathbf{X}}_j - \mathbf{X}_j\right\|_2 \\
&\leq \sqrt{\frac{2K}{n}}O_P\left(\epsilon\sqrt{K\frac{\beta_{\max}}{\beta_{\min}}}\right) + 2\|\mathbf{X}_i\| \cdot \epsilon' \\
&\leq \sqrt{\frac{2K}{n}}O_P\left(\epsilon\sqrt{K\frac{\beta_{\max}}{\beta_{\min}}}\right) + 2\sqrt{\frac{2K}{n}} \cdot (1 + O_P(\epsilon)) \cdot \epsilon' \\
&= O_P\left(\epsilon\sqrt{\frac{K^2}{n\beta_{\min}}}\right). \qquad (\beta_{\max} = 1 \text{ by definition})
\end{aligned}$$

Now we can see $\frac{1}{2}\sqrt{\frac{K\beta_{\min}}{n}}$ can be used as a threshold to separate different clusters. However, in the algorithm we do not know $\beta_{\min}$ in advance, so we need to approximate it with some computable statistics. From Lemma D.1, we know that $\mathbf{D}_2(i,i) = \frac{\rho n}{2K}\left(\mathbf{B}_{a_i a_i} \pm O_P(\epsilon')\right)$ when $\theta_{ia_i} \geq 1-\epsilon$. So $\min_{i\in\mathcal{F}} \mathbf{D}_2(i,i)$ and $\max_{i\in\mathcal{F}} \mathbf{D}_2(i,i)$ can be used to estimate $\beta_{\min}$.

$$\tau = \sqrt{\frac{K}{4n}\frac{\min_{i\in\mathcal{F}} \mathbf{D}_2(i,i)}{\max_{i\in\mathcal{F}} \mathbf{D}_2(i,i)}} = \sqrt{\frac{K}{4n}\frac{\rho n\left(\beta_{\min} \pm O_P(\epsilon)\right)/(2K)}{\rho n\left(1 \pm O_P(\epsilon)\right)/(2K)}} = \frac{1}{2}\sqrt{\frac{K\beta_{\min}}{n}} \pm O_P\left(\epsilon\sqrt{\frac{K}{n\beta_{\min}}}\right).$$

Clearly,

$$2\sqrt{\frac{K\beta_{\min}}{n}} \pm O_P\left(\epsilon\sqrt{\frac{K}{n}}\right) > 2\tau \gg O_P\left(\epsilon\sqrt{\frac{K^2}{n\beta_{\min}}}\right),$$

which means PartitionPureNodes($\hat{\mathbf{X}}(\mathcal{F},:),\tau$) can exactly give us $K$ clusters of different (nearly) pure nodes and return one (nearly) pure node from each of the $K$ clusters with probability larger than $1 - O_P(K^2/n^2)$. □

## E. Consistency of inferred parameters

*Proof of Theorem 4.5.* Let $\hat{\boldsymbol{Y}} := \sqrt{n/2K}\hat{\mathbf{X}}$ and $\mathbf{Y} = \sqrt{n/2K}\mathbf{X}$. Let $\epsilon = O_P(\epsilon_0 + \epsilon') = O(\epsilon')$ from Lemma 4.3, where we show that $\|\mathbf{Y}_i\|^2 \geq 1-\epsilon$ for $i \in \mathcal{S}_P$. Furthermore for ease of exposition let us assume that the pure nodes are arranged so that $\hat{\mathbf{\Theta}}_{2p} = \hat{\mathbf{\Theta}}_2(\mathcal{S}_p,:)$ is close to an identity matrix, i.e., the columns are arranged with a particular permutation.

Thus $\|\mathbf{Y}_p\|_F^2 = \sum_i \|\mathbf{Y}_p(i,:)\|^2 \geq K(1-\epsilon)$ and so $\|\mathbf{Y}_p\|_F \geq \sqrt{K(1-\epsilon)}$.

We have also shown that $\|\mathbf{Y}_p(i,:)\|^2 \leq 1+\epsilon$, so $\|\mathbf{Y}_p\|_F \leq \sqrt{K}\,(1+\epsilon)$.

We will use

$$\|\hat{\mathbf{Y}}_p^{-1} - \left(\mathbf{Y}_p\boldsymbol{O}\right)^{-1}\|_F \leq \|\left(\mathbf{Y}_p\boldsymbol{O}\right)^{-1}\left(\mathbf{Y}_p\boldsymbol{O} - \hat{\mathbf{Y}}_p\right)\hat{\mathbf{Y}}_p^{-1}\|_F \leq \|\mathbf{Y}_p^{-1}\|_F\|\mathbf{Y}_p\boldsymbol{O} - \hat{\mathbf{Y}}_p\|_F\|\hat{\mathbf{Y}}_p^{-1}\|. \tag{3}$$

First we will prove a bound on $\|\hat{\mathbf{Y}}_p^{-1}\|$. Let $\hat{\sigma}_i$ be the $i^{th}$ singular value of $\hat{\mathbf{Y}}_p$,

$$\|\hat{\mathbf{Y}}_p^{-1}\| = \frac{1}{\hat{\sigma}_K}. \tag{4}$$

We can bound $\hat{\sigma}_K$ by bounding $\sigma_K$. In what follows we use $M_{1p}$ to denote the rows of $M_1$ indexed by $\mathcal{S}_p$ when $M_1$ is $n/2 \times K$ and by the square submatrix $M_1(\mathcal{S}_p,\mathcal{S}_p)$ is when $M_1$ is $n/2 \times n/2$. Note that $\|\mathbf{\Theta}_{2p} - I\|_F = \sqrt{K}\epsilon$, $\|\mathbf{B}^{1/2}\|_F = O_P(\sqrt{K})$, $\|\mathbf{\Theta}_{2p}\|_F = O_P(\sqrt{K})$ and $\|\boldsymbol{\mathcal{D}}_{21p}^{-1}\|_F = O_P(K^{3/2}/\beta_{\min}\rho n)$,

$$\begin{aligned}
\sigma_i^2 = \lambda_i(\mathbf{Y}_p\mathbf{Y}_p^T) &= \frac{\rho n}{2K}\lambda_i(\boldsymbol{\mathcal{D}}_{21p}^{-1/2}\mathbf{\Theta}_{2p}\mathbf{B}\mathbf{\Theta}_{2p}^T\boldsymbol{\mathcal{D}}_{21p}^{-1/2}) \\
&= \frac{\rho n}{2K}\lambda_i(\mathbf{B}^{1/2}\mathbf{\Theta}_{2p}^T\boldsymbol{\mathcal{D}}_{21p}^{-1}\mathbf{\Theta}_{2p}\mathbf{B}^{1/2}) \\
&= \frac{\rho n}{2K}\lambda_i\left(\mathbf{B}^{1/2}\left(\boldsymbol{\mathcal{D}}_{21p}^{-1} + (\mathbf{\Theta}_{2p} - I)^T\boldsymbol{\mathcal{D}}_{21p}^{-1}\mathbf{\Theta}_{2p} + \boldsymbol{\mathcal{D}}_{21p}^{-1}(\mathbf{\Theta}_{2p} - I)\right)\mathbf{B}^{1/2}\right).
\end{aligned}$$

Note that the matrix $\mathbf{B}^{1/2}\boldsymbol{\mathcal{D}}_{21p}^{-1}\mathbf{B}^{1/2}$ is a diagonal matrix with the $(i,i)^{th}$ diagonal element being $\beta_i/\boldsymbol{\mathcal{D}}_{21p}(i,i)$.

With similar arguments in the proof of Lemma D.1, we can get

$$\boldsymbol{\mathcal{D}}_{21p}(i,i) = \frac{n\rho}{2K}\left(\beta_i \pm O_P(\epsilon)\right),$$

so

$$\lambda_K(\mathbf{B}^{1/2}\boldsymbol{\mathcal{D}}_{21p}^{-1}\mathbf{B}^{1/2}) = \frac{2K}{\rho n}\left(1 \pm O_P(\epsilon/\beta_{\min})\right).$$

By Weyl's inequality and note that operator norm is less than or equal to Frobenius norm, it immediately gives us:

$$\begin{aligned}\left|\sigma_i^2 - \frac{\rho n}{2K}\lambda_i\left(\mathbf{B}^{1/2}\boldsymbol{\mathcal{D}}_{21p}^{-1}\mathbf{B}^{1/2}\right)\right| \leq& \frac{\rho n}{2K}\cdot\left\|\mathbf{B}^{1/2}(\mathbf{\Theta}_{2p}-I)^T\boldsymbol{\mathcal{D}}_{21p}^{-1}\mathbf{\Theta}_{2p}\mathbf{B}^{1/2}+\mathbf{B}^{1/2}\boldsymbol{\mathcal{D}}_{21p}^{-1}(\mathbf{\Theta}_{2p}-I)\mathbf{B}^{1/2}\right\| \\ \leq& \frac{\rho n}{2K}\cdot\left\|\mathbf{B}^{1/2}\right\|_F\cdot\|\mathbf{\Theta}_{2p}-I\|_F\cdot\left\|\boldsymbol{\mathcal{D}}_{21p}^{-1}\right\|_F\cdot 2\,\|\mathbf{\Theta}_{2p}\|_F\cdot\left\|\mathbf{B}^{1/2}\right\|_F \\ =& O_P\left(\frac{\rho n}{2K}\cdot\sqrt{K}\cdot\sqrt{K}\epsilon\cdot\frac{K^{3/2}}{\beta_{\min}\rho n}\cdot\sqrt{K}\cdot\sqrt{K}\right) \\ =& O_P\left(\frac{K^{9/2}\sqrt{\log n}}{\beta_{\min}^{7/2}\rho\sqrt{n}}\right) \\ \implies \quad \sigma_K^2 =& 1\pm O_P\left(\frac{K^{9/2}\sqrt{\log n}}{\beta_{\min}^{7/2}\rho\sqrt{n}}\right).\end{aligned}$$

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

$$\begin{aligned}|\hat{\sigma}_i-\sigma_i| &\leq \|\hat{\boldsymbol{Y}}_p-\mathbf{Y}_p\boldsymbol{O}\| \leq \|\hat{\boldsymbol{Y}}_p-\mathbf{Y}_p\boldsymbol{O}\|_F = O_P(\sqrt{K}\epsilon) \\ \hat{\sigma}_K &= \left(1\pm O_P\left(\frac{K^{9/2}\sqrt{\log n}}{\beta_{\min}^{7/2}\rho\sqrt{n}}\right)\right)^{1/2}\left(1\pm O_P\left(\sqrt{K}\cdot\frac{K^2\sqrt{\log n}}{\beta_{\min}^{5/2}\rho\sqrt{n}}\right)\right) = 1\pm O_P\left(\frac{K^{9/2}\sqrt{\log n}}{\beta_{\min}^{7/2}\rho\sqrt{n}}\right).\end{aligned}$$

Plugging this into Equation (4) we get:

$$\|\hat{\boldsymbol{Y}}_p^{-1}\| = 1\pm O_P\left(\frac{K^{9/2}\sqrt{\log n}}{\beta_{\min}^{7/2}\rho\sqrt{n}}\right).$$

Finally putting everything together with Equation (3) we get:

$$\frac{\|\hat{\mathbf{X}}_p^{-1}-(\mathbf{X}_p\boldsymbol{O})^{-1}\|_F}{\|\mathbf{X}_p^{-1}\|_F} = \frac{\|\hat{\boldsymbol{Y}}_p^{-1}-(\mathbf{Y}_p\boldsymbol{O})^{-1}\|_F}{\|\mathbf{Y}_p^{-1}\|_F} \leq \|\mathbf{Y}_p\boldsymbol{O}-\hat{\boldsymbol{Y}}_p\|_F\|\hat{\boldsymbol{Y}}_p^{-1}\| = O_P\left(\frac{K^{5/2}\sqrt{\log n}}{\beta_{\min}^{5/2}\rho\sqrt{n}}\right) \tag{5}$$

with probability larger than $1-O_P(K^2/n^2)$. □

*Proof of Theorem 4.6.* Recall that $\hat{\mathbf{\Theta}}_2 = \hat{\mathbf{\Theta}}(\bar{\mathcal{S}}) = \mathbf{D}_{12}^{1/2}\hat{\mathbf{X}}\hat{\mathbf{X}}_p^{-1}\mathbf{D}_{21}^{-1/2}(\mathcal{S}_p,\mathcal{S}_p)$. First note that if one plugs in the population counterparts of the the terms in $\hat{\mathbf{\Theta}}_2$, then for some permutation matrix $\mathbf{\Pi}$ that $\mathbf{\Theta}_{2p} := \mathbf{\Theta}_2(\mathcal{S}_p,:)\cdot\mathbf{\Pi}$ is close to an identity matrix, and

$$\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\mathbf{X}_p^{-1}\boldsymbol{\mathcal{D}}_{21p}^{-1/2} = \boldsymbol{\mathcal{D}}_{21}^{1/2}\left(\sqrt{\rho}\cdot\boldsymbol{\mathcal{D}}_{21}^{-1/2}\mathbf{\Theta}_2\mathbf{B}^{1/2}\right)\left(\frac{1}{\sqrt{\rho}}\cdot\mathbf{B}^{-1/2}\mathbf{\Pi}\mathbf{\Theta}_{2p}^{-1}\boldsymbol{\mathcal{D}}_{21p}^{1/2}\right)\boldsymbol{\mathcal{D}}_{21p}^{-1/2} = \mathbf{\Theta}_2\mathbf{\Pi}\mathbf{\Theta}_{2p}^{-1},$$

so

$$\mathbf{\Theta}_2\mathbf{\Pi} = \boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\mathbf{X}_p^{-1}\boldsymbol{\mathcal{D}}_{21p}^{-1/2}\mathbf{\Theta}_{2p}.$$

We have the following decomposition

$$\begin{aligned}\left\|\hat{\mathbf{\Theta}}_2-\mathbf{\Theta}_2\mathbf{\Pi}\right\|_F \leq& \left\|(\mathbf{D}_{21}^{1/2}-\boldsymbol{\mathcal{D}}_{21}^{1/2})\hat{\mathbf{X}}\hat{\mathbf{X}}_p^{-1}\mathbf{D}_{21p}^{-1/2}\right\|_F + \left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}(\hat{\mathbf{X}}-\mathbf{X}\boldsymbol{O})\hat{\mathbf{X}}_p^{-1}\mathbf{D}_{21p}^{-1/2}\right\|_F \\ &+\left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\boldsymbol{O}(\hat{\mathbf{X}}_p^{-1}-(\mathbf{X}_p\boldsymbol{O})^{-1})\mathbf{D}_{21p}^{-1/2}\right\|_F + \left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\mathbf{X}_p^{-1}(\mathbf{D}_{21p}^{-1/2}-\boldsymbol{\mathcal{D}}_{21p}^{-1/2})\right\|_F \\ &+\left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\mathbf{X}_p^{-1}\boldsymbol{\mathcal{D}}_{21p}^{-1/2}\left(\mathbf{I}-\mathbf{\Theta}_{2p}\right)\right\|_F.\end{aligned}$$

From the proof of Lemma C.6 we have $\sqrt{\mathbf{D}_{21}(i,i)} = \sqrt{\boldsymbol{\mathcal{D}}_{21}(i,i)}(1\pm O_P(\sqrt{K\log n/n\rho}))$ and hence

$$\left\|\mathbf{D}_{21}^{1/2}-\boldsymbol{\mathcal{D}}_{21}^{1/2}\right\| = \left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\right\|O_P(\sqrt{K\log n/n\rho}),$$

and

$$\left\|\mathbf{D}_{21p}^{-1/2} - \boldsymbol{\mathcal{D}}_{21p}^{-1/2}\right\| \leq \left\|\boldsymbol{\mathcal{D}}_{21p}^{-1/2}\right\| O_P(\sqrt{K\log n/n\rho}).$$

And $\|\hat{\mathbf{X}}_p^{-1}\| = \sqrt{n/(2K)}\|\hat{\mathbf{Y}}_p^{-1}\| = O_P(\sqrt{n/K})$, as we have shown in the proof of Theorem 4.5. Furthermore, by Fact C.1, $\|\hat{\mathbf{X}}_p^{-1}\|_F = O_P(\sqrt{n})$.

From the proof of Theorem 4.2 we can get $\|\hat{\mathbf{X}} - \mathbf{X}\boldsymbol{O}\|_F = O_P(\sqrt{K}\epsilon')$. Theorem 4.5 gives

$$\begin{aligned}\|\hat{\mathbf{X}}_p^{-1} - (\mathbf{X}_p\boldsymbol{O})^{-1}\|_F &= \|\mathbf{X}_p^{-1}\|_F \cdot O_P\left(\frac{K^{5/2}\sqrt{\log n}}{\beta_{\min}^{5/2}\rho\sqrt{n}}\right)\\ &= O_P\left(\sqrt{K}\cdot\sqrt{\frac{n}{K}}\right)\cdot O_P\left(\frac{K^{5/2}\sqrt{\log n}}{\beta_{\min}^{5/2}\rho\sqrt{n}}\right) = O_P\left(\frac{K^{5/2}\sqrt{\log n}}{\beta_{\min}^{5/2}\rho}\right).\end{aligned}$$

Also $\|\hat{\mathbf{X}}\|_F = O_P(\sqrt{K})$, since it concentrates around its population entry-wisely, and the max norm of any row of the population is $\sqrt{2K/n}$, so $\|\mathbf{X}\|_F = O_P(\sqrt{K})$. And

$$\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\| = \|\boldsymbol{\mathcal{D}}_{21}^{1/2}\sqrt{\rho}\cdot\boldsymbol{\mathcal{D}}_{21}^{-1/2}\mathbf{\Theta}_2\mathbf{B}^{1/2}\| = \|\sqrt{\rho}\cdot\mathbf{\Theta}_2\mathbf{B}^{1/2}\| = \sqrt{\|\mathbf{P}\|} = O_P(\sqrt{\rho n/K}).$$

Hence,

$$\begin{aligned}
&\left\|(\mathbf{D}_{21}^{1/2} - \boldsymbol{\mathcal{D}}_{21}^{1/2})\hat{\mathbf{X}}\hat{\mathbf{X}}_p^{-1}\mathbf{D}_{21p}^{-1/2}\right\|_F \leq \left\|\mathbf{D}_{21}^{1/2} - \boldsymbol{\mathcal{D}}_{21}^{1/2}\right\|\left\|\hat{\mathbf{X}}\right\|_F\left\|\hat{\mathbf{X}}_p^{-1}\right\|\left\|\mathbf{D}_{21p}^{-1/2}\right\|\\
&= O_P\left(\sqrt{\rho n/K}\right)O_P\left(\sqrt{K\log n/n\rho}\right)\cdot O_P\left(\sqrt{K}\right)\cdot O_P\left(\sqrt{n/K}\right)\cdot O_P(\sqrt{K/\beta_{\min}\rho n}) = O_P\left(\sqrt{K\log n/\beta_{\min}\rho}\right),\\
&\left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}(\hat{\mathbf{X}} - \mathbf{X}\boldsymbol{O})\hat{\mathbf{X}}_p^{-1}\mathbf{D}_{21p}^{-1/2}\right\|_F \leq \left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\right\|\left\|\hat{\mathbf{X}} - \mathbf{X}\boldsymbol{O}\right\|_F\left\|\hat{\mathbf{X}}_p^{-1}\right\|\left\|\mathbf{D}_{21p}^{-1/2}\right\|\\
&= O_P\left(\sqrt{\rho n/K}\right)\cdot O_P\left(\sqrt{K}\epsilon'\right)\cdot O_P\left(\sqrt{n/K}\right)\cdot O_P(\sqrt{K/\beta_{\min}\rho n}) = O_P\left(\sqrt{\frac{n}{\beta_{\min}}}\epsilon'\right) = O_P\left(\frac{K^2\sqrt{\log n}}{\beta_{\min}^3\rho}\right),\\
&\left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}(\hat{\mathbf{X}}_p^{-1} - (\mathbf{X}_p\boldsymbol{O})^{-1})\mathbf{D}_{21p}^{-1/2}\right\|_F \leq \left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\right\|\left\|\hat{\mathbf{X}}_p^{-1} - (\mathbf{X}_p\boldsymbol{O})^{-1}\right\|_F\left\|\mathbf{D}_{21p}^{-1/2}\right\|\\
&= O_P\left(\sqrt{\rho n/K}\right)\cdot O_P\left(\frac{K^{5/2}\sqrt{\log n}}{\beta_{\min}^{5/2}\rho}\right)\cdot O_P(\sqrt{K/\beta_{\min}\rho n}) = O_P\left(\frac{K^{5/2}\sqrt{\log n}}{\beta_{\min}^3\rho}\right),\\
&\left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\mathbf{X}_p^{-1}(\mathbf{D}_{21p}^{-1/2} - \boldsymbol{\mathcal{D}}_{21p}^{-1/2})\right\|_F \leq \left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\right\|\|\mathbf{X}_p^{-1}\|\left\|\mathbf{D}_{21p}^{-1/2} - \boldsymbol{\mathcal{D}}_{21p}^{-1/2}\right\|_F\\
&= O_P\left(\sqrt{\rho n/K}\right)\cdot O_P(\sqrt{K}\cdot\sqrt{n/K})\cdot O_P(\sqrt{K/\beta_{\min}\rho n})O_P\left(\sqrt{K\log n/n\rho}\right) = O_P\left(\sqrt{K\log n/\beta_{\min}\rho}\right),\\
&\left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\mathbf{X}_p^{-1}\boldsymbol{\mathcal{D}}_{21p}^{-1/2}(\mathbf{I} - \mathbf{\Theta}_{2p})\right\|_F \leq \left\|\boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\right\|\|\mathbf{X}_p^{-1}\|\left\|\boldsymbol{\mathcal{D}}_{21p}^{-1/2}\right\|_F\|\mathbf{I} - \mathbf{\Theta}_{2p}\|_F\\
&= O_P\left(\sqrt{\rho n/K}\right)\cdot O_P(\sqrt{K}\cdot\sqrt{n/K})\cdot O_P(\sqrt{K/\beta_{\min}\rho n})\cdot\sqrt{K}\epsilon' = O_P\left(\sqrt{\frac{Kn}{\beta_{\min}}}\epsilon'\right) = O_P\left(\frac{K^{5/2}\sqrt{\log n}}{\beta_{\min}^3\rho}\right).
\end{aligned}$$

So

$$\left\|\hat{\mathbf{\Theta}}_2 - \mathbf{\Theta}_2\mathbf{\Pi}\right\|_F = O_P\left(\frac{K^{5/2}\sqrt{\log n}}{\beta_{\min}^3\rho}\right).$$

Since $\|\mathbf{\Theta}_2\|_F^2 = \Omega(n/K)$, we finally have:

$$\frac{\left\|\hat{\mathbf{\Theta}}_2 - \mathbf{\Theta}_2\mathbf{\Pi}\right\|_F}{\|\mathbf{\Theta}_2\|_F} \leq O_P\left(\frac{K^3\sqrt{\log n}}{\beta_{\min}^3\rho\sqrt{n}}\right)$$

with probability larger than $1 - O(K^2/n^2)$. □

*Proof of Theorem 4.7.* Recall that $\hat{\rho}\hat{\beta}_a = \left\|\mathbf{e}_a^T \mathbf{D}_{21}^{1/2}(\mathcal{S}_p, \mathcal{S}_p)\hat{\mathbf{X}}_p\right\|_F^2$, and for some permutation matrix $\mathbf{\Pi}$ that $\mathbf{\Theta}_{2p} := \mathbf{\Theta}_2(\mathcal{S}_p, :) \cdot \mathbf{\Pi}$ is close to an identity matrix, if one plugs in the population counterparts of the the terms in $\hat{\mathbf{\Theta}}_2$,

$$\begin{aligned}\left\|\mathbf{e}_a^T \boldsymbol{\mathcal{D}}_{21}^{1/2}(\mathcal{S}_p, \mathcal{S}_p)\mathbf{X}_p\right\|_F &= \left\|\sqrt{\rho}\cdot \mathbf{e}_a^T \mathbf{\Theta}_{2p}\mathbf{\Pi}^T\mathbf{B}^{1/2}\right\|_F = \left\|\sqrt{\rho}\cdot \mathbf{e}_a^T\left(\mathbf{\Theta}_{2p} - \mathbf{I}\right)\mathbf{\Pi}^T\mathbf{B}^{1/2} + \sqrt{\rho}\cdot \mathbf{e}_a^T \mathbf{\Pi}^T\mathbf{B}^{1/2}\right\|_F \\ &\le \left\|\sqrt{\rho}\cdot \mathbf{e}_a^T\left(\mathbf{\Theta}_{2p} - \mathbf{I}\right)\mathbf{\Pi}^T\mathbf{B}^{1/2}\right\|_F + \left\|\sqrt{\rho}\cdot \mathbf{e}_a^T \mathbf{\Pi}^T\mathbf{B}^{1/2}\right\|_F \\ &= \sqrt{\rho}\epsilon' + \sqrt{\rho\beta_{a'}},\end{aligned}$$

where $a' \in [K]$ satisfies $\mathbf{\Pi}_{a'a} = 1$.

Using the bounds mentioned in the proof of Theorem 4.6, we have:

$$\begin{aligned}&\left\|\mathbf{e}_i^T(\mathbf{D}_{21}^{1/2}\hat{\mathbf{X}} - \boldsymbol{\mathcal{D}}_{21}^{1/2}\mathbf{X}\boldsymbol{O})\right\| = \left\|\mathbf{e}_i^T(\mathbf{D}_{21}^{1/2} - \boldsymbol{\mathcal{D}}_{21}^{1/2})\hat{\mathbf{X}} + \mathbf{e}_i^T\boldsymbol{\mathcal{D}}_{21}^{1/2}(\hat{\mathbf{X}} - \mathbf{X}\boldsymbol{O})\right\| \\ &\le \left\|\left(\sqrt{\mathbf{D}_{21}(i,i)} - \sqrt{\boldsymbol{\mathcal{D}}_{21}(i,i)}\right)\mathbf{e}_i^T\hat{\mathbf{X}}\right\| + \left\|\sqrt{\boldsymbol{\mathcal{D}}_{21}(i,i)}\mathbf{e}_i^T(\hat{\mathbf{X}} - \mathbf{X}\boldsymbol{O})\right\| \\ &\le \left(\sqrt{\mathbf{D}_{21}(i,i)} - \sqrt{\boldsymbol{\mathcal{D}}_{21}(i,i)}\right)\left(\left\|\mathbf{e}_i^T\left(\hat{\mathbf{X}} - \mathbf{X}\boldsymbol{O}\right)\right\| + \|\mathbf{e}_i^T\mathbf{X}\|\right) + \sqrt{\boldsymbol{\mathcal{D}}_{21}(i,i)}\left\|\mathbf{e}_i^T(\hat{\mathbf{X}} - \mathbf{X}\boldsymbol{O})\right\| \\ &= O_P\left(\sqrt{\frac{n\rho}{K}}\right)O_P\left(\sqrt{\frac{K\log n}{n\rho}}\right)\left[O_P\left(\frac{\sqrt{K^5\log n}}{\beta_{\min}^{5/2}\rho n}\right) + O_P\left(\sqrt{\frac{K}{n}}\right)\right] + O_P\left(\sqrt{\frac{n\rho}{K}}\right)\cdot O_P\left(\frac{\sqrt{K^5\log n}}{\beta_{\min}^{5/2}\rho n}\right) \\ &= O_P\left(\frac{K^{5/2}\log n}{\beta_{\min}^{5/2}\sqrt{\rho n}}\right).\end{aligned}$$

As a result,

$$\left|\sqrt{\hat{\rho}\hat{\beta}_a} - \sqrt{\rho\beta_{a'}}\right| \le O_P\left(\frac{K^{5/2}\log n}{\beta_{\min}^{5/2}\sqrt{\rho n}}\right) + \sqrt{\rho}\epsilon' = O_P\left(\frac{K^{5/2}\log n}{\beta_{\min}^{5/2}\sqrt{\rho n}}\right),$$

and note that $\rho\beta_{a'} = \Omega(\rho)$, we have

$$\hat{\rho}\hat{\beta}_a \in \rho\beta_{a'}\left[1 - O_P\left(\frac{K^{5/2}\log n}{\beta_{\min}^{5/2}\rho\sqrt{n}}\right), 1 + O_P\left(\frac{K^{5/2}\log n}{\beta_{\min}^{5/2}\rho\sqrt{n}}\right)\right]$$

with probability larger than $1 - O(K^2/n^2)$. $\square$

## References

Huang, Kejun, Sidiropoulos, Nicholas, and Swami, Ananthram. Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. *Signal Processing, IEEE Transactions on*, 62(1):211–224, 2014.

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

Minc, Henryk. Nonnegative matrices. 1988.

Tang, Minh, Sussman, Daniel L, Priebe, Carey E, et al. Universally consistent vertex classification for latent positions graphs. *The Annals of Statistics*, 41(3):1406–1430, 2013.

Yu, Yi, Wang, Tengyao, Samworth, Richard J, et al. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.