SUPPLEMENTARY MATERIAL FOR "FAST COMMUNITY DETECTION BY SCORE"

By Jiashun Jin

Carnegie Mellon University

In this supplement we propose some variants of SCORE and present the technical proofs for the main theorems in [3] (in Appendix B) and for all the secondary lemmas in [3] (Appendix C). Equation and theorem references made to the main document do not contain letters.

APPENDIX A: VARIANTS OF SCORE

The key idea underlying the SCORE is that, in a broad context, the leading eigenvectors of the adjacency matrix A approximate those of the non-stochastic matrix Ω, where the latter are

$$\Theta\Big(\sum_{\ell=1}^{k}[a_k(\ell)/\|\theta^{(\ell)}\|]\mathbf{1}_\ell\Big), \qquad k = 1, 2, \ldots, K.$$

It is seen that

- The information of the community labels is contained in the term within the bracket, which depends on $\{\theta(i)\}_{i=1}^n$ only through the overall degree intensities $\|\theta^{(k)}\|/\|\theta\|$.
- The diagonal matrix Θ does not contain any information of the community labels.
- Therefore, $\{\theta(i)\}_{i=1}^n$ are almost nuisance parameters, the effect of which can be removed by many *scaling invariant* mappings, to be introduced below.

Definition A.1. *Let $W \subset R^K$ be a subset such that when $x \in W$, then $ax \in W$ for any $a > 0$. We call a mapping $\mathbb{M}$ from W to R^K scaling invariant if $\mathbb{M}(ax) = \mathbb{M}(x)$ for any $a > 0$ and $x \in W$.*

The following are some examples of scaling invariant mappings.

- (a). $W = \{x \in R^K, x(1) \neq 0\}$, and $\mathbb{M}x = x/x(1)$; $x(1)$ is the first coordinate of x.
- (b). $W = R^K \setminus \{0\}$, $\mathbb{M}x = x/\|x\|_q$, where $q > 0$ is a constant.

	SCORE	SCOREq	
		$q=1$	$q=2$
web blogs ($n=1222$)	58	61	64
karate ($n=34$)	1	1	1

TABLE 1
Comparison of number of errors for the SCORE and the SCOREq.

Given a scaling invariant $\mathbb{M}$, we have the following extension of SCORE.

- Obtain the K leading (unit-norm) eigenvectors of A. Arrange them in an $n \times K$ matrix $\hat{R}$ as follows so that ξ_i' is the i-th row of $\hat{R}$, $1 \leq i \leq n$:
$$\hat{R} = [\hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_K] = (\xi_1, \xi_2, \ldots, \xi_n)'.$$
- Obtain an $n \times K$ matrix $\hat{R}^*$ where the i-th row of $\hat{R}^*$ is $(\mathbb{M}\xi_i)'$.
- Apply k-means method to $\hat{R}^*$ for clustering with $\leq K$ classes.

For example, if we view each row of $\hat{R}$ as a point in R^k and apply $\mathbb{M}$ in (a), then we have the $\hat{R}$ matrix in (2.4) associated with the original SCORE (except for that in (2.4), the first column is removed for it is the vector of 1 and is thus non-informative for clustering). For another example, we take $\mathbb{M}$ as the mapping in (b), and call the resultant procedure SCOREq, where q is the parameter in the mapping $x \to x/\|x\|_q$.

We have investigated the performances of SCORE and SCOREq with the karate club data and the web blogs data, where we pick $q = 1$ and $q = 2$. The performances are largely similar, but SCORE is slightly better for the web blogs data. See Table 1 for details.

APPENDIX B: PROOF OF THE MAIN THEOREMS

In this section, we prove Theorems 2.1-2.2. The key for the proof is Lemmas 2.7-2.8, which contain bounds on $\|\hat{\eta}_k - \eta_k\|$ and $\|\Theta^{-1}(\hat{\eta}_k - \eta_k)\|$, respectively. To show Lemmas 2.7-2.8, we need tight moderate deviation bounds on matrices and vectors involving the noise matrix W. Such bounds are ensured by the assumption (2.17). In fact, by (2.17) and basic algebra, for any constant $C \geq 1$,

(B.1)
$$\sum_{i=1}^n \max\{\theta(i), \frac{\log(n)\theta_{max}^2}{\|\theta\|_3^3}\} \leq C\|\theta\|_1, \ \sum_{i=1}^n \max\{\frac{1}{\theta(i)}, \frac{\log(n)\theta_{max}^2}{\theta^2(i)\|\theta\|_3^3}\} \leq \sum_{i=1}^n \frac{C}{\theta(i)},$$

and

$$\frac{\log(n)}{\theta_{min}^2} \leq \max\{\frac{1}{\theta_{min}}\|\theta\|_1, \theta_{max}\sum_{i=1}^n \frac{1}{\theta(i)}\}. \quad \text{(B.2)}$$

We call (B.1)-(B.2) the Moderate Deviation conditions on Vectors (MDV) and the Moderate Deviation conditions on Matrices (MDM), respectively. These inequalities are used to control the moderate deviations of norms of vectors, say, $W\theta$, and norms of matrices, say, $\Theta^{-1}W$, respectively. The main results of the paper continue to hold if we replace (2.17) by (B.1)-(B.2).

Below, we first describe such moderate deviation bounds, and then give the proofs for Theorems 2.1-2.2. The proofs of Lemmas 2.7-2.8 are given in Section C.

B.1. Moderate deviation inequalities on vectors and matrices. The following theorem is proved in [5], which is the extension of the well-known Bernstein's inequality from the case of random variables to the case of random matrices. Recall that $\|\cdot\|$ denotes the spectral norm.

THEOREM B.1. *Consider a finite sequence $\{Z_k\}$ of independent (real-valued) $n \times p$ random matrices. Assume that each random matrix satisfies $E[Z_k] = 0$ and $\|Z_k\| \leq h_0$ almost surely. Then for all $t \geq 0$,*

$$P(\|\sum_k Z_i\| \geq t) \leq (n+p)\exp\left(-\frac{t^2/2}{\sigma^2 + h_0 t/3}\right),$$

where $\sigma^2 = \max\{\|\sum_k E[Z_k Z_k']\|, \|\sum_k E[Z_k' Z_k]\|\}$.

The following lemma provides moderate deviation bounds on $\|\Theta^{-1}W\|$.

LEMMA B.1. *If the Moderate Deviation condition on Matrices (B.2) holds, then $\|\Theta^{-1}W\|^2 \leq C\log(n)\max\{\theta_{min}^{-1}\|\theta\|_1, \theta_{max}\sum_{i=1}^n \frac{1}{\theta(i)}\}$, with probability at least $1 + o(n^{-3})$.*

The following lemma provides moderate deviation bounds on the norms of various vectors. The lemma is proved in Section C, where the classical Bennett's inequality is the key [4] (recall that $\theta^{(k)}$ is defined in (2.10)).

LEMMA B.2. *If the Moderate Deviation condition on Vectors (B.1) holds, then with probability at least $1 + o(n^{-3})$, for all $1 \leq k, \ell \leq K$,*

- $\|W\theta^{(k)}\|^2 \leq C\log(n)\|\theta\|_3^3\|\theta\|_1$.
- $\|\Theta^{-1}W\theta\|^2 \leq C\log(n)\|\theta\|_3^3\sum_{i=1}^n \frac{1}{\theta(i)}$.
- $|(\theta^{(k)})'W\theta^{(\ell)}|^2 \leq C\log(n)[\|\theta\|_3^6 + \log(n)\theta_{max}^4]$.

We are now ready to show the two main theorems.

B.2. Proof of Theorem 2.1. Let $\{\eta_i\}_{i=1}^K$ and $\{\hat{\eta}_i\}_{i=1}^K$ be as in Lemma 2.1 and Lemma 2.4, respectively. Let $\hat{S}$ be the set of "well-behaved" nodes as in (2.20), where $c_0 = 1/2$ for simplicity. By Lemmas 2.7-2.9, there is an event E_n such that $P(E_n^c) = o(n^{-3})$ and that over E_n, for all $1 \leq k \leq K$,

(B.3)
$$\|\hat{\eta}_k - \eta_k\|^2 \leq C\log(n)\|\theta\|_1\|\theta\|_3^3/\|\theta\|^6, \quad \|\Theta^{-1}(\hat{\eta}_k - \eta_k)\|^2 \leq C\log(n)err_n,$$

and

$$|V \setminus \hat{S}| \leq C\log(n)err_n. \tag{B.4}$$

Especially, combining (B.3) and (2.12), $\|\hat{\eta}_k\| \sim \|\eta_k\| = 1$. To show the claim, it is sufficient to show that over the event E_n, $\|\hat{R}^* - R\|_F^2 \leq C\log^3(n)err_n$.

To this end, we write

$$\|\hat{R}^* - R\|_F^2 = U_1 + U_2,$$

where U_1 is the sum of squares of the ℓ^2-norms of all "ill-behaved" rows of $\hat{R}^* - R$, and U_2 is that of all "well-behaved" rows.

Consider U_1. For any $i \notin \hat{S}$ and $1 \leq k \leq K-1$, it is seen that $|\hat{R}^*(i,k)| \leq T_n$ and $|R(i,k)| \leq |\eta_{k+1}(i)/\eta_1(i)| \leq C$, where $T_n = \log(n)$ and we have used Lemma 2.1 and (2.15). Combining these with Lemma 2.9 and (2.14),

$$U_1 \leq C\log^2(n)|V \setminus \hat{S}| \leq C\log^3(n)err_n. \tag{B.5}$$

Consider U_2. Recall that for any $i \in \hat{S}$,

$$|\hat{\eta}_1(i)/\eta_1(i) - 1| \leq 1/2. \tag{B.6}$$

Since $|R(i,k)| \leq C$, $|\hat{R}^*(i,k) - R(i,k)| \leq |\hat{R}(i,k) - R(i,k)|$. Write

$$\hat{R}(i,k) - R(i,k) = \frac{\|\hat{\eta}_1\|}{\|\hat{\eta}_{k+1}\|}\frac{\hat{\eta}_{k+1}(i)}{\hat{\eta}_1(i)} - \frac{\eta_{k+1}(i)}{\eta_1(i)} = \frac{\|\hat{\eta}_1\|}{\|\hat{\eta}_{k+1}\|}(I + II + III), \tag{B.7}$$

where

$$I = (\hat{\eta}_{k+1}(i) - \eta_{k+1}(i))/\eta_1(i), \qquad II = \hat{\eta}_{k+1}(i)(\eta_1(i) - \hat{\eta}_1(i))/(\hat{\eta}_1(i)\eta_1(i)),$$

and

$$III = (1 - \|\hat{\eta}_{k+1}\|/\|\hat{\eta}_1\|)\eta_{k+1}(i)/\eta_1(i).$$

Recall that $\|\hat{\eta}_k\| \sim 1$ for all $1 \leq k \leq K$.

Now, first, by Lemma 2.6,

$$|I| \leq C|\hat{\eta}_{k+1}(i) - \eta_{k+1}(i)|/\theta(i). \tag{B.8}$$

Second, write $II = IIa + IIb$, where

$$IIa = \frac{\eta_{k+1}(i)(\eta_1(i) - \hat{\eta}_1(i))}{\hat{\eta}_1(i)\eta_1(i)}, \qquad IIb = \frac{[\hat{\eta}_{k+1}(i) - \eta_{k+1}(i)][\eta_1(i) - \hat{\eta}_1(i)]}{\hat{\eta}_1(i)\eta_1(i)}.$$

By Lemma 2.1, Lemma 2.6, and (B.6), $|\eta_{k+1}(i)/[\hat{\eta}_1(i)\eta_1(i)]| \leq C/\theta(i)$ and $|[\eta_1(i)-\hat{\eta}_1(i)]/[\hat{\eta}_1(i)\eta_1(i)]| \leq C/\theta(i)$. Therefore, $|IIa| \leq C|\hat{\eta}_1(i)-\eta_1(i)|/\theta(i)$ and $|IIb| \leq C|\hat{\eta}_{k+1}(i) - \eta_{k+1}(i)|/\theta(i)$. Combining these gives

$$|II| \leq C\big[|\hat{\eta}_1(i) - \eta_1(i)| + |\hat{\eta}_{k+1}(i) - \eta_{k+1}(i)|\big]/\theta(i). \tag{B.9}$$

Third, recalling $\|\eta_k\| = 1$ and $\|\hat{\eta}_k\| \sim 1$ for all $1 \leq k \leq K$ and using triangle inequality,

$$\big|1 - \|\hat{\eta}_{k+1}\|/\|\hat{\eta}_1\|\big| \lesssim \big|\|\hat{\eta}_1\| - \|\hat{\eta}_{k+1}\|\big| \leq \big|\|\hat{\eta}_1\| - \|\eta_1\|\big| + \big|\|\hat{\eta}_{k+1}\| - \|\eta_{k+1}\|\big|,$$

where the right hand side does not exceed $\|\hat{\eta}_1 - \eta_1\| + \|\hat{\eta}_k - \eta_k\|$. At the same time, recall that $\eta_{k+1}(i)/\eta_1(i) \leq C$. Combining these gives

$$|III| \leq C[\|\hat{\eta}_1 - \hat{\eta}_{k+1}\|] \leq C[\|\hat{\eta}_1 - \eta_1\| + \|\hat{\eta}_{k+1} - \eta_{k+1}\|]. \tag{B.10}$$

Inserting (B.8)-(B.10) into (B.7), $|\hat{R}(i,k) - R(i,k)|$ does not exceed

$$C\big(\frac{1}{\theta(i)}\big[|\hat{\eta}_1(i) - \eta_1(i)| + |\hat{\eta}_{k+1}(i) - \eta_{k+1}(i)|\big] + \|\hat{\eta}_1 - \eta_1\| + \|\hat{\eta}_{k+1} - \eta_{k+1}\|\big).$$

Therefore, over the event E_n,

$$U_2 \leq C \sum_{k=1}^{K} \big(\|\Theta^{-1}(\hat{\eta}_k - \eta_k)\|^2 + n\|\hat{\eta}_k - \eta_k\|^2\big).$$

Combining this with (B.3) gives

$$U_2 \leq C\log(n)[err_n + n\|\theta\|_3^3\|\theta\|_1/\|\theta\|^6]. \tag{B.11}$$

Note that $n\|\theta\|_1/\|\theta\|^2 \leq \sum_{i=1}^n (1/\theta(i))$. Therefore, $n\|\theta\|_3^3\|\theta\|_1/\|\theta\|^6 \leq err_n$ by definitions. Combining this with (B.5) and (B.11) gives the claim. □

B.3. Proof of Theorem 2.2. Without loss of generality, assume $K > 2$. The proof for the case $K = 2$ is the same, except for that $\hat{R}^*$, R, and M^* are vectors rather than matrices, so that we have to change the terminology slightly. The following lemma is proved in Section C.

LEMMA B.3. *The $n \times (K-1)$ matrix R has exactly K distinct rows, and the ℓ^2-distance between any two distinct rows is no smaller than $\sqrt{2}$.*

We now show Theorem 2.2. For $1 \le i \le n$, let $\hat{r}_i$, r_i, and c_i denote the i-th row of $\hat{R}^*$, R, and M^* correspondingly. Fixing $\delta = \sqrt{2}/3$, we introduce a subset of V by $W = \{1 \le i \le n : \|\hat{r}_i - m_i\| \le \delta, \|r_i - m_i\| \le \delta\}$. Recalling that V partitions to K communities $V^{(1)}, V^{(2)}, \ldots, V^{(K)}$, we note that W has a similar partition $W = W^{(1)} \cup W^{(2)} \cup \ldots W^{(K)}$, where $W^{(k)} = V^{(k)} \cap W$, $1 \le k \le K$.

By Theorem 2.1, there is an event B such that $P(B^c) = o(n^{-2})$, and over the event B,

$$\|\hat{R}^* - R\|_F^2 \le C \log^3(n) err_n. \tag{B.12}$$

Note that the $n \times (K-1)$ matrix R has exactly K unique rows so $R \in \mathcal{M}_{n,K-1,K}$. By how the k-means procedure is constructed, $\|\hat{R}^* - M^*\|_F \le \|\hat{R}^* - R\|_F$, and so $\|R - M^*\|_F \le \|\hat{R}^* - R\|_F + \|\hat{R}^* - M^*\|_F \le 2\|\hat{R}^* - R\|_F$. Combining this with (B.12),

$$\|M^* - R\|_F^2 \le 4\|\hat{R}^* - R\|_F^2 \le C \log^3(n) err_n. \tag{B.13}$$

Combining this with (B.12), it follows from the definition of W that $|V \setminus W| \le C \log^3(n) err_n$. Comparing this with the desired claim, it is sufficient to show that all nodes in W are correctly labeled; equivalently, this is to show that for any $i, j \in W$ such that $i \in W^{(k)}$ and $j \in W^{(\ell)}$,

$$m_i = m_j \text{ if and only if } k = \ell. \tag{B.14}$$

We now show (B.14). First, by definitions and (B.12)-(B.13), the cardinality of $(V^{(k)} \setminus W^{(k)})$ does not exceed $\delta^{-2} \sum_{i \in V^{(k)}} (\|\hat{r}_i - r_i\|^2 + \|m_i - r_i\|^2) \le C \log^3(n) err_n$. Recall that we assume $\log^3(n) err_n \ll \min\{n_1, n_2, \ldots, n_k\}$. Combining these, $W^{(k)}$ is non-empty. Second, by Lemma B.3 and definitions, for any $i, j \in W$ such that $i \in W^{(k)}$, $j \in W^{(\ell)}$, where $1 \le k, \ell \le K$ and $k \ne \ell$,

$$\|m_i - m_j\| \ge \|r_i - r_j\| - (\|m_i - r_i\| + \|m_j - r_j\|) \ge \delta. \tag{B.15}$$

Therefore, if $m_i = m_j$ for some $i, j \in W$, then there is a $1 \le k \le K$ such that $i, j \in W^{(k)}$. Suppose we pick one node j_k from each $W^{(k)}$, $1 \le k \le K$. By (B.15), the K row vectors $\{m_{j_1}, ..., m_{j_K}\}$ are distinct. Note that the matrix M^* has at most K distinct rows, so if $i, j \in W^{(k)}$ for some $1 \le k \le K$, then $m_i = m_j$. Combining these gives (B.14). □

APPENDIX C: PROOF OF SECONDARY LEMMAS

In this section, we prove all the lemmas in the preceding sections.

C.1. Proof of Lemmas 2.1. Fix $1 \le k \le K$. Let λ_k be the nonzero eigenvalue of Ω with the k-th largest magnitude, let η_k be one of the (unit-norm) eigenvector associated with λ_k, and let a_k be the $K \times 1$ vector such that $a_k(i) = (\theta^{(i)}/\|\theta^{(i)}\|, \eta_k)$. In our model, we can rewrite Ω as

$$\Omega = \|\theta\|^2 \sum_{i,j=1}^{K} (DPD)(i,j)\big(\frac{\theta^{(i)}}{\|\theta^{(i)}\|}\big)\big(\frac{\theta^{(j)}}{|\theta^{(j)}\|}\big)', \tag{C.16}$$

and so by basic algebra and notations,

$$\Omega\eta_k = \|\theta\|^2 \sum_{i,j=1}^{K} (DPD)(i,j)(\frac{\theta^{(j)}}{\|\theta^{(j)}\|}, \eta_k)\frac{\theta^{(i)}}{\|\theta^{(i)}\|} = \|\theta\|^2 \sum_{i,j=1}^{K} (DPD)(i,j) a_k(j) \frac{\theta^{(i)}}{\|\theta^{(i)}\|}. \tag{C.17}$$

At the same time, since $\Omega\eta_k = \lambda_k \eta_k$,

$$a_k(i) = (\theta^{(i)}/\|\theta^{(i)}\|, \eta_k) = \frac{1}{\lambda_k}(\theta^{(i)}/\|\theta^{(i)}\|, \Omega\eta_k). \tag{C.18}$$

Note that $\{\theta^{(i)}/\|\theta^{(i)}\|\}_{i=1}^K$ is an orthonormal base. Inserting (C.17) to the right hand side of (C.18) gives $a_k(i) = (\|\theta\|^2/\lambda_k)\sum_{j=1}^K (DPD)(i,j)a_k(j)$, or in matrix form,

$$DPDa_k = (\lambda_k/\|\theta\|^2)a_k. \tag{C.19}$$

This says that $\lambda_k/\|\theta\|^2$ is an eigenvalue of DPD and a_k is one of the associated eigenvector. Moreover, inserting (C.19) into to the right hand side of (C.17) and recalling $\Omega\eta_k = \lambda_k\eta_k$,

$$\eta_k = \frac{1}{\lambda_k}\Omega\eta_k = \sum_{i=1}^{K} a_k(i)\theta^{(i)}/\|\theta^{(i)}\|,$$

and so $\|a_k\|^2 = \|\eta_k\|^2 = 1$. By our assumptions, all eigenvalues of DPD are simple. It follows that a_k is unique determined up to a factor of ± 1. $\square$

C.2. Proof of Lemma 2.2. By the definition of DCBM and (2.8), we have that $\|\mathrm{diag}(\Omega)\| \le \theta_{max}^2$, where $\theta_{max} \le g_0 < 1$. Note that by (2.11)-(2.12), $\theta_{max}\|\theta\|_1 \to \infty$. It follows that

$$\|\mathrm{diag}(\Omega)\| = o(\sqrt{\log(n)\theta_{max}\|\theta\|_1}).$$

Therefore, to show the claim, it is sufficient to show that with probability at least $1 + o(n^{-3})$,

$$\|W\| \le 3\sqrt{\log(n)\theta_{max}\|\theta\|_1}.$$

Let e_i be the $n \times 1$ vector such that $e_i(j) = 1$ if $i = j$ and 0 otherwise. Write $W = \sum_{1 \leq i < j \leq n} Z^{(i,j)}$, where $Z^{(i,j)} = W(i,j)[e_i e_j' + e_j e_i']$. Let $\sigma^2 = \| \sum_{1 \leq i < j \leq n} E[(Z^{(i,j)})^2] \|$. By elementary statistics and (2.8), $E[W^2(i,j)] \leq \theta(i)\theta(j)$. At the same time,

$$E[(Z^{(i,j)})^2] = E[W(i,j)^2] \cdot [e_i e_j' + e_j e_i']^2 = E[W(i,j)^2] \cdot [e_i e_i' + e_j e_j'].$$

Combining these gives that $\sigma^2 \leq \theta_{max} \|\theta\|_1$. Fix $q > 0$. Applying Theorem B.1 with $Z^{(i,j)} = W(i,j)[e_i e_j' + e_j e_i']$, $h_0 = 1$, $\sigma^2 = \| \sum_{i<j} E[(Z^{(i,j)})^2] \|$, and $t = \sqrt{2q \log(n) \theta_{max} \|\theta\|_1}$,

$$P(\|W\| \geq \sqrt{2q \log(n) \theta_{max} \|\theta\|_1}) \leq 2n\text{exp}\Big[-\frac{q \log(n)}{1 + (1/3)\sqrt{2q \log(n)/(\theta_{max}\|\theta\|_1)}} \Big].$$

Note that $\theta_{max}\|\theta\|_1 \geq \|\theta\|^2$, and that $\|\theta\|^2/\log(n) \to \infty$ as in the assumption (2.11). It follows that $q\log(n)/(\theta_{max}\|\theta\|_1) \to 0$, and the claim follows by taking $q = 9/2$.

C.3. Lemma 2.4. Let $\lambda_1 > \lambda_2 > \ldots > \lambda_K$ be the K nonzero eigenvalues of Ω. By Lemma 2.3 and (2.14)-(2.15), for all $1 \leq k \leq K-1$, $|\lambda_{k+1} - \lambda_k| \geq C\|\theta\|^2$, At the same time, by Lemma 2.2 and (2.11), it follows from basic algebra (e.g., [1, Page 473]) that with probability at least $1 + o(n^{-3})$,

$$|\hat{\lambda}_k - \lambda_k| / \|\theta\|^2 = o(1), \tag{C.20}$$

and so all the K eigenvalues are simple.

Now, fixing $1 \leq k \leq K$, let $\hat{\eta}_k$ be an eigenvector (the norms of which are not necessarily 1) associated with $\hat{\lambda}_k$. Writing for short $\hat{\theta}^{(i,k)} = [I_n - (W - \text{diag}(\Omega))/\hat{\lambda}_k]^{-1}\theta^{(i)}$, we let $\hat{b}_k$ be the $K \times 1$ vector such that

$$\hat{b}_k(i) = (\theta^{(i)}/\|\theta^{(i)}\|, \hat{\eta}_k), \qquad 1 \leq i \leq K,$$

and let

$$\hat{a}_k = (B^{(k)})^{-1}\hat{b}_k.$$

Since $A\hat{\eta}_k = \hat{\lambda}_k \eta_k$ and $A = \Omega + (W - \text{diag}(\Omega))$, it follows that

$$\hat{\eta}_k = [\hat{\lambda}_k I_n - (W - \text{diag}(\Omega))]^{-1}\Omega\hat{\eta}_k.$$

Recall that (e.g., (C.16)) $\Omega = \|\theta\|^2 \sum_{i,j=1}^K (DPD)(i,j)(\theta^{(i)}/\|\theta^{(i)}\|)(\theta^{(j)}/\|\theta^{(j)}\|)'$. Combining these and rearranging,

$$\hat{\eta}_k = (\frac{\|\theta\|^2}{\hat{\lambda}_k}) \sum_{i=1}^K \Big[\Big(\sum_{j=1}^K (DPD)(i,j)\hat{b}_k(j) \Big) \cdot \frac{\hat{\theta}^{(i,k)}}{\|\theta^{(i)}\|} \Big]. \tag{C.21}$$

Recall that $B^{(k)}(\ell, i) = (\theta^{(\ell)})'[I_n - (W - \mathrm{diag}(\Omega))/\hat{\lambda}_k]^{-1}\theta^{(i)}/[\|\theta^{(\ell)}\| \cdot \|\theta^{(i)}\|] \equiv (\theta^{(\ell)})'\hat{\theta}^{(i,k)}/[\|\theta^{(\ell)}\| \cdot \|\theta^{(i)}\|]$. Taking the inner product of two sides in (C.21) with $\theta^{(\ell)}/\|\theta^{(\ell)}\|$, it follows from the definitions of $\hat{b}_k$ that for any $1 \leq \ell \leq K$,

$$\hat{b}_k(\ell) = (\frac{\|\theta\|^2}{\hat{\lambda}_k}) \sum_{i,j=1}^{K} B^{(k)}(\ell, i)(DPD)(i, j)\hat{b}_k(j) = (\frac{\|\theta\|^2}{\hat{\lambda}_k}) \sum_{j=1}^{n} (B^{(k)}DPD)(\ell, j)\hat{b}_k(j),$$

or in matrix form,

$$\hat{b}_k = (\frac{\|\theta\|^2}{\hat{\lambda}_k}) B^{(k)} DPD\hat{b}_k.$$

This means that $\hat{\lambda}_k/\|\theta\|^2$ is an eigenvalue of $B^{(k)}DPD$ and $\hat{b}_k$ is one of the associated eigenvector. Recall that $\hat{a}_k = (B^{(k)})^{-1}\hat{b}_k$. By basic algebra, $\hat{\lambda}_k/\|\theta\|^2$ is an eigenvalue of $DPDB^{(k)}$, and $\hat{a}_k$ is one of the associated eigenvectors. Especially,

$$DPD\hat{b}_k = DPDB^{(k)}\hat{a}_k = [\hat{\lambda}_k/\|\theta\|^2]\hat{a}_k. \tag{C.22}$$

Inserting (C.22) into (C.21) and rearranging,

$$\hat{\eta}_k = \sum_{i=1}^{K} \hat{a}_k(i) \cdot \hat{\theta}^{(i,k)}/\|\theta^{(i)}\|.$$

We now check the uniqueness of $\hat{\lambda}_k$. By Lemma C.1 to be introduced below and (2.12), $\|B^{(k)} - I_K\|_F \leq C \log(n)\|\theta\|_1\|\theta\|_3^3/\|\theta\|^6 = o(1)$. By similar argument as in (C.20), $\mathrm{eigsp}(DPD) \geq C$. Combining these with basic algebra (e.g., [1, Page 473]), all eigenvalues of $DPDB^{(k)}$ are simple, and $\hat{b}_k$ (and so $\hat{\lambda}_k$, $\hat{a}_k$, and $\hat{\eta}_k$) are uniquely determined up to some scaling factors. If we further require $\|\hat{a}_k\| = 1$, then $\hat{a}_k$, $\hat{b}_k$, and $\hat{\eta}_k$ are all uniquely determined, up to a common factor that takes values from $\{-1, 1\}$. This gives the claim. □

C.4. Proof of Lemma 2.6. Recall that D is a diagonal matrix where the k-th diagonal is the k-th coordinate of the $K \times 1$ vector $d^{(n)}$, equalling $\|\theta^{(k)}\|/\|\theta\|$, $1 \leq k \leq K$; the superscript "(n)" emphasizes the dependence on n (same below). By Lemma 2.1, $\Theta^{-1}\eta_1 = \sum_{k=1}^K [a_1^{(n)}(k)/\|\theta^{(k)}\|]\mathbf{1}_k$, where $a_1^{(n)}$ is the eigenvector associated with the largest eigenvalue of DPD. By (2.14), to show the lemma, it suffices to show that for sufficiently large n,

$$OSC(a_1^{(n)}) \leq C. \tag{C.23}$$

Note that in the special case where $d^{(n)}$ does not depend on n, the claim follows directly by Perron's theorem [2, Page 508], since DPD is non-negative and irreducible. Consider the general case where $d^{(n)}$ may depend on n. If (C.23) does not hold, then we can find a subsequence of $n \in \{1, 2, \ldots, \}$ such that along this sequence, there are two $K \times 1$ vectors d_0 and a such that (a) $OSC(a_1^{(n)}) \to \infty$, (b) $d^{(n)} \to d_0$, and (c) $a_1^{(n)} \to a$. By the condition (2.15), $OSC(d_0) \leq C$, and a direct use of of Perron's theorem [2, Page 508] implies that $OSC(a) \leq C$. This contradicts with (a). The contradiction proves (C.23) and the claim follows. □

C.5. Proof of Lemmas 2.7-2.8. Write $\hat{\theta}^{(i,k)} = [I_n - (W - \text{diag}(\Omega))/\hat{\lambda}_k]^{-1}\theta^{(i)}$ for short. In our notations,

$$\eta_k = \sum_{i=1}^{K} [a_k(i)/\|\theta^{(i)}\|]\theta^{(i)}, \qquad \hat{\eta}_k = \sum_{i=1}^{K} [\hat{a}_k(i)/\|\theta^{(i)}\|]\hat{\theta}^{(i,k)}, \tag{C.24}$$

where a_k are the eigenvectors of DPD and $\hat{a}_k$ are the eigenvectors of $DPD(B^{(k)})$. To show the claim, we first characterize $\|\hat{a}_k - a_k\|$, and then characterize $\|\hat{\theta}^{(i,k)} - \theta^{(i)}\|$.

Consider $\|\hat{a}_k - a_k\|$ first. Let I_K be the $K \times K$ identity matrix. The following lemma is proved below (implicitly, we assume that in Lemma C.1, the conditions of Lemmas 2.7-2.8 hold; same for Lemmas C.3-C.4).

LEMMA C.1. *With probability at least $1 + o(n^{-3})$,*

$$\|B^{(k)} - I_K\|_F \leq C \log(n)(\|\theta\|_1 \cdot \|\theta\|_3^3)/\|\theta\|^6.$$

Note that by (2.11), the right hand side tends to 0 as $n \to \infty$.

We also need a lemma on *eigenvector sensitivity*. Suppose U and Err are both symmetric $K \times K$ matrix where $\|Err\| < (1/2)\text{eigsp}(U)$, so that all the eigenvalues of U and $U + Err$ are simple. Let $\lambda_1^{(1)} > \lambda_2^{(1)} > \ldots > \lambda_K^{(1)}$ and $\lambda_1^{(2)} > \lambda_2^{(2)} > \ldots > \lambda_K^{(2)}$ be the eigenvalues of U and $U + Err$, respectively, and let $\xi_2^{(1)}, \xi_2^{(1)}, \ldots, \xi_K^{(1)}$ and $\xi_1^{(1)}, \xi_2^{(2)}, \ldots, \xi_K^{(2)}$ be the corresponding (unit-norm) eigenvectors, of U and $U + Err$, respectively. The following lemma is proved below.

LEMMA C.2. *If $\|Err\| < \text{eigsp}(U)/2$, then for any $1 \leq k \leq K$, $\|\xi_k^{(1)} - \xi_k^{(2)}\| \leq 2\sqrt{2}\frac{\|Err\|}{\text{eigsp}(U)}$.*

Note that $\|DPD\| \leq C$. Using Lemma C.1 and basic algebra, with probability at least $1 + o(n^{-3})$,

$$\|DPD(B^{(k)}) - DPD\| \leq C\|(B^{(k)}) - I_K\| \leq C \log(n)(\|\theta\|_1 \cdot \|\theta\|_3^3)/\|\theta\|^6.$$

Applying Lemma C.2 with $U = DPD$ and $Err = DPD[(B^{(k)}) - I_K]$, it follows from the eigen-space condition (2.14) that with probability at least $1 + o(n^{-3})$, for $1 \leq k \leq K$,

$$\|\hat{a}_k - a_k\| \leq C\|Err\| \leq C\log(n)(\|\theta\|_1 \cdot \|\theta\|_3^3)/\|\theta\|^6.$$

By (2.11), the right hand side tends to 0, so

$$\|\hat{a}_k - a_k\|^2 \leq \|\hat{a}_k - a_k\| \leq C\log(n)(\|\theta\|_1 \cdot \|\theta\|_3^3)/\|\theta\|^6. \tag{C.25}$$

Next, we consider $\|\hat{\theta}^{(i,k)} - \theta^{(k)}\|$. The following lemmas are proved below.

Lemma C.3. *With probability at least $1 + o(n^{-3})$, for all $1 \leq k, i \leq K$,*

$$\|\hat{\theta}^{(i,k)} - \theta^{(i)}\|^2 \leq C\log(n)\|\theta\|_1\|\theta\|_3^3/\|\theta\|^4.$$

Lemma C.4. *With probability at least $1 + o(n^{-3})$, for any $1 \leq k, i \leq K$,*

$$\|\Theta^{-1}(\hat{\theta}^{(i,k)} - \theta^{(i)})\|^2 \leq C\log(n)\frac{\|\theta\|_3^3}{\|\theta\|^4}[\sum_{i=1}^n \frac{1}{\theta(i)} + \frac{1}{\theta_{min}}\frac{\log(n)\|\theta\|_1^2}{\|\theta\|^4}].$$

We now show Lemmas 2.7-2.8. Consider Lemma 2.7 first. By (C.24) and basic algebra,

$$\|\hat{\eta}_k - \eta_k\|^2 \leq C(I + II),$$

where $I = \sum_{i=1}^K \hat{a}_k^2(i)(\|\hat{\theta}^{(i,k)} - \theta^{(i)}\|^2/\|\theta^{(i)}\|^2)$, and $II = \sum_{i=1}^K(\hat{a}_k(i) - a_k(i))^2 = \|\hat{a}_k - a_k\|^2$. Since $\hat{a}_k$ has unit norm and $\|\theta^{(i)}\|^2 \geq C\|\theta\|^2$, combining these with (C.25) and Lemma gives

$$\|\hat{\eta}_k - \eta_k\|^2 \leq C\log(n)[\|\theta\|_1 \cdot \|\theta\|_3^3]/\|\theta\|^6,$$

and the claim follows.

Next, consider Lemma 2.8. Similarly, $\|\Theta^{-1}[\hat{\eta}_k - \eta_k]\|^2 \leq C(I + II)$, where $I = \sum_{i=1}^K \hat{a}_k^2(i)(\|\Theta^{-1}[\hat{\theta}^{(i,k)} - \theta^{(i)}]\|^2/\|\theta^{(i)}\|^2)$, and $II = \sum_{i=1}^K |\hat{a}_k(i) - a_k(i)|^2\|\mathbf{1}_i\|^2/\|\theta^{(i)}\|^2$. By Lemma C.4 and similar argument,

$$\|\Theta^{-1}[\hat{\eta}_k - \eta_k]\|^2 \leq C\log(n)\frac{\|\theta\|_3^3}{\|\theta\|^6}[\sum_{i=1}^n \frac{1}{\theta(i)} + \frac{\log(n)}{\theta_{min}}\frac{\|\theta\|_1^2}{\|\theta\|^4} + \frac{n\|\theta\|_1}{\|\theta\|^2}].$$

Note that $n\|\theta\|_1 \leq \|\theta\|^2 \sum_{i=1}^n \frac{1}{\theta(i)}$, it follows that

$$\|\Theta^{-1}[\hat{\eta}_k - \eta_k]\|^2 \leq C\log(n)\frac{\|\theta\|_3^3}{\|\theta\|^6}[\sum_{i=1}^n \frac{1}{\theta(i)} + \frac{\log(n)}{\theta_{min}}\frac{\|\theta\|_1^2}{\|\theta\|^4}],$$

and the claim follows by the definition of err_n; see (2.18). ∎

C.6. Proof of Lemma B.1. Similar to that in the proof of Lemma 2.2, let e_i be the $n \times 1$ vector such that $e_i(j) = 1$ if $i = j$ and 0 otherwise. Write $\Theta^{-1}W = \sum_{i<j} Z^{(i,j)}$, where $Z^{(i,j)} = W(i,j)\Theta^{-1}[e_i e_j' + e_j e_i']$. Let

$$\sigma^2 = \max\{\| \sum_{1 \leq i < j \leq n} E[Z^{(i,j)}(Z^{(i,j)})']\|, \; \| \sum_{1 \leq i < j \leq n} E[(Z^{(i,j)})' Z^{(i,j)}]\|\}.$$

First, by (2.8) and basic statistics, $E[W^2(i,j)] \leq \theta(i)\theta(j)$. It is seen

$$E[Z^{(i,j)}(Z^{(i,j)})'] = E[W^2(i,j)]\Theta^{-1}[e_i e_j' + e_j e_i']^2 \Theta^{-1} = E[W^2(i,j)]\Theta^{-1}[e_i e_i' + e_j e_j']\Theta^{-1},$$

which is a diagonal matrix, where the i-th diagonal $\leq \theta(j)/\theta(i)$, the j-th diagonal $\leq \theta(i)/\theta(j)$, and all other diagonals are 0. Therefore, $\sum_{1 \leq i < j \leq n} E[Z^{(i,j)}(Z^{(i,j)})']$ is a diagonal matrix where the i-th coordinate does not exceed $\|\theta\|_1/\theta(i)$, and the matrix norm of which $\leq \theta_{min}^{-1}\|\theta\|_1$.

Second, we similarly have

$$E[(Z^{(i,j)})' Z^{(i,j)}] = E[W^2(i,j)][e_i e_j' + e_j e_i']\Theta^{-2}[e_i e_j' + e_j e_i'],$$

which is a diagonal matrix where the i-th coordinate does not exceed $\theta(i)/\theta(j)$, the j-th coordinate $\leq \theta(j)/\theta(i)$. As a result, $\sum_{1 \leq i < j \leq n} E[(Z^{(i,j)})' Z^{(i,j)}]$ is a diagonal matrix where the i-th coordinate does not exceed $\theta(i) \sum_{j=1}^n (1/\theta(j))$, and the matrix norm $\leq \theta_{max} \sum_{i=1}^n (1/\theta(i))$. Combining these gives

$$\sigma^2 \leq \max\{\frac{1}{\theta_{min}}\|\theta\|_1, \; \theta_{max} \sum_{i=1}^n \frac{1}{\theta(i)}\} \equiv \sigma_0^2.$$

Fix $q > 0$. Applying Theorem B.1 with $h_0 = 1/\theta_{min}$ and $t = \sigma_0\sqrt{2q\log(n)}$ gives

$$P(\|\Theta^{-1}W\| \geq \sigma_0\sqrt{2q\log(n)}) \leq 2n\exp\Big[-\frac{q\log(n)}{1 + (1/3)\sqrt{2q\log(n)}\theta_{min}^{-1}/\sigma_0}\Big].$$

By (B.2), $\log(n)\theta_{min}^{-2} \leq \sigma_0^2$, and the claim follows by picking q to be a sufficiently large constant. □

C.7. Proof of Lemma B.2. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with $|Y_k| \leq b$, $E[Y_k] = 0$, and $\mathrm{var}(Y_k) \leq \sigma_k^2$ for $1 \leq k \leq n$. Write for short $\sigma^2 = \sigma_1^2 + \ldots \sigma_n^2$. We claim that with probability at least $1 + o(1/n^3)$,

$$|\sum_{i=1}^n Y_i|^2 \leq C\log(n)\max\{\sigma^2, \log(n)b^2\}. \tag{C.26}$$

In detail, using Bennett's Lemma [4, Page 851], for all $\lambda > 0$,

$$P(|\sum_{i=1}^n Y_i| \geq \lambda) \leq \begin{cases} 2\exp(-\frac{c_0}{2\sigma^2}\lambda^2), & \lambda b \leq \sigma^2, \\ 2\exp(-\frac{c_0}{2}\frac{\lambda}{b}), & \lambda b \geq \sigma^2. \end{cases}$$

where $c_0 = \psi(1)$, with ψ as in [4, Page 851]; note that $c_0 \approx 0.773$. Now, when $\sigma/b \geq 2\sqrt{2\log(n)}$, we take $\lambda = 2\sqrt{2\log(n)}\sigma$. It is seen $\lambda b \leq \sigma^2$, and so

$$P(\big|\sum_{i=1}^n Y_i\big| \geq \lambda) \leq 2\exp(-4c_0\log(n)) = o(n^{-3}).$$

When $\sigma/b < 2\sqrt{2\log(n)}$, we take $\lambda = 8b\log(n)$. It is seen $\lambda b \geq \sigma^2$. It follows that

$$P(\big|\sum_{i=1}^n Y_i\big| \geq \lambda) \leq 2\exp(-4c_0\log(n)) = o(n^{-3}).$$

Combining these, except for a probability of $o(n^{-3})$,

$$\big|\sum_{i=1}^n Y_i\big| \leq 2\sqrt{2\log(n)}\sigma 1\{\sigma/b \geq 2\sqrt{2\log(n)}\} + 8b\log(n)1\{\sigma/b < 2\sqrt{2\log(n)}\},$$

and (C.26) follows.

We now show Lemma B.2. The last item follows directly from (C.26), and the proofs for first two items are similar, so we only show the second item. Let e_i be the $n \times 1$ vector such that $e_i(j) = 0$ if $i = j$ and 0 otherwise. Write $\|W\theta^{(k)}\|^2 = \sum_{i=1}^n (e_i' W\theta^{(k)})^2$. For each fixed i, applying (C.26) to $Y_j = W(i,j)\theta^{(k)}(j)$, $b = \theta_{max}$, and $\sigma^2 = E[(\sum_{j\in V^{(k)}} \theta(j)W(i,j))^2]$, we have that with probability at least $1 + o(1/n^3)$,

$$|e_i' W\theta^{(k)}|^2 \leq C\log(n)\max\{\sigma^2, \log(n)\theta_{max}^2\}.$$

Now, direct calculation shows that $\sigma^2 \leq \theta(i)\|\theta\|_3^3$. It follows that with probability at least $1 + o(n^{-2})$ that

$$\|W\theta^{(k)}\|^2 \leq C\log(n)\sum_{i=1}^n \max\{\theta(i)\|\theta\|_3^3, \log(n)\theta_{max}^2\},$$

and the claim follows by the first MDV assumption in (B.1). □

C.8. Proof of Lemma B.3. We expand R to be an $n \times K$ matrix by adding a column of ones to the left. For notational simplicity, we still call the matrix by R. It is sufficient to show that R has exactly K distinct rows, and the ℓ^2-distance for each pair of such distinct rows is no smaller than $\sqrt{2}$.

With the new notations, since Θ is a diagonal matrix, for any $1 \leq i \leq n$ and $1 \leq k \leq K$,

$$R(i,k) = \frac{\eta_k(i)}{\eta_1(i)} = \frac{(\Theta^{-1}\eta_k)(i)}{(\Theta^{-1}\eta_1)(i)},$$

where $\eta_1, \eta_2, \ldots, \eta_K$ are the K leading eigenvectors of Ω. Combining this with Lemma 2.1 and recalling that $d_j = \|\theta^{(j)}\|/\|\theta\|$,

$$R(i,k) = \frac{\sum_{j=1}^K a_k(j)\mathbf{1}_j(i)/d_j}{\sum_{j=1}^K a_1(j)\mathbf{1}_j(i)/d_j},$$

which equals to $a_k(\ell)/a_1(\ell)$ if and only if node i belongs to the ℓ-th community $V^{(\ell)}$, $\ell = 1, 2, \ldots, K$. It is now evident that R has K distinct rows, each is one of the following row-vectors:

$$\frac{1}{a_1(\ell)}\big(a_1(\ell), a_2(\ell), \ldots, a_K(\ell)\big), \qquad \ell = 1, 2, \ldots, K.$$

Fix $k \neq \ell$. The square of the ℓ^2-distance between the vector $\frac{1}{a_1(k)}\big(a_1(k), \ldots, a_K(k)\big)$ and the vector $\frac{1}{a_1(\ell)}\big(a_1(\ell), \ldots, a_K(\ell)\big)$ is

$$\frac{1}{a_1^2(k)}\sum_{j=1}^K a_j^2(k) + \frac{1}{a_1^2(\ell)}\sum_{j=1}^K a_j^2(\ell) - \frac{2}{a_1(k)a_1(\ell)}\sum_{j=1}^K a_j(k)a_j(\ell).$$

Since $a_1, a_2, \ldots, a_K$ form an orthonormal base, $\sum_{j=1}^K a_j^2(k) = 1$, $\sum_{j=1}^K a_j^2(\ell) = 1$, and $\sum_{j=1}^K a_j(k)a_j(\ell) = 0$. Therefore, the square of the ℓ^2-distance between these two vectors is $a_1^{-2}(k) + a_1^{-2}(\ell)$ and the claim follows since $|a_1(k)| \leq 1$ and $|a_1(\ell)| \leq 1$. □

C.9. Proof of Lemma C.1. Write for short $U = \mathrm{diag}(\Omega)$ and $H = (W - \mathrm{diag}(\Omega))/\hat{\lambda}_k$. For $1 \leq i, j \leq K$, $B^{(k)}(i,j) = (\theta^{(i)})'[I_n - H]^{-1}\theta^{(j)}/(\|\theta^{(i)}\| \cdot \|\theta^{(j)}\|)$. By (2.15), for all $1 \leq i \leq K$, $\|\theta^{(i)}\| \asymp \|\theta\|$. All we need to show is

(C.27)
$$\big|(\theta^{(i)})'[(I_n - H)^{-1} - I_n]\theta^{(j)}\big| \leq C\log(n)\|\theta\|_1\|\theta\|_3^3/\|\theta\|^4, \qquad 1 \leq i, j \leq K;$$

note that $(\theta^{(i)})'\theta^{(j)} = \|\theta^{(i)}\|^2$ if $i = j$ and 0 otherwise.

Write

$$(\theta^{(i)})'[(I_n - H)^{-1} - I_n]\theta^{(j)} = I + II, \tag{C.28}$$

where

$$I = (\theta^{(i)})' H \theta^{(j)}, \qquad II = (\theta^{(i)})' H [I_n - H]^{-1} H \theta^{(j)}.$$

Consider I first. By $H = (W - U)/\hat{\lambda}_k$, we have

$$I = \frac{1}{\hat{\lambda}_k}(Ia - Ib), \tag{C.29}$$

where $Ia = (\theta^{(i)})' W \theta^{(j)}$, $Ib = (\theta^{(i)})' U \theta^{(j)}$. First, by (2.8) and that all $\theta(i) \leq 1$, $|Ib| \leq \|\theta\|_4^4 \leq \|\theta\|_3^3$ with probability at least $1 + o(n^{-3})$. Second, by Lemma B.2, $Ia \leq C\sqrt{\log(n)} \max\{\|\theta\|_3^3, \sqrt{\log(n)}\theta_{max}^2\}$. Last, by (2.16), with probability at least $1+o(n^{-3})$, $|\hat{\lambda}_k| \asymp \|\theta\|^2$. Inserting these into (C.29) gives

$$|I| \leq C\sqrt{\log(n)} \max\{\|\theta\|_3^3, \sqrt{\log(n)}\theta_{max}^2\}/\|\theta\|^2. \tag{C.30}$$

Consider II next. First, by Schwartz inequality, $|II| \leq \|(I_n - H)^{-1/2} H \theta^{(i)}\| \cdot \|(I_n - H)^{-1/2} H \theta^{(j)}\|$. Second, by (2.11) and Lemma 2.3, with probability at least $1 + o(n^{-3})$, $\|I_n - H\|^{-1/2} \lesssim 1$ and $\hat{\lambda}_k \asymp \|\theta\|^2$. Therefore, for any $1 \leq i \leq K$, $\|(I_n - H)^{-1/2} H \theta^{(i)}\|^2$ does not exceed

$$\frac{1}{\hat{\lambda}_k^2}\|(I_n - H)^{-1/2}(W-U)\theta^{(i)}\|^2 \leq C\|(W-U)\theta^{(i)}\|^2/\|\theta\|^4 \leq (IIa + IIb)/\|\theta\|^4,$$

where $IIa = \|W\theta^{(i)}\|^2$ and $IIb = \|U\theta^{(i)}\|^2$. Now, on one hand, by Lemma B.2, with probability at least $1 + o(n^{-3})$,

$$\|W\theta^{(i)}\|^2 \leq C\log(n)\|\theta\|_1\|\theta\|_3^3.$$

On the other hand, by basic algebra and that $\|\theta^{(i)}\|_\infty < 1$,

$$\|U\theta^{(i)}\|^2 \leq \|\theta\|_6^6 \leq \|\theta\|_3^3.$$

Note that by (2.11)-(2.12), $\log(n)\|\theta\|_1 \geq \|\theta\|^4 \gg 1$. Combining these gives that with probability at least $1 + o(1/n^2)$,

$$|II| \leq C\log(n)\|\theta\|_1\|\theta\|_3^3/\|\theta\|^4. \tag{C.31}$$

Inserting (C.30) and (C.31) into (C.28) gives that with probability at least $1 + o(1/n^2)$,

$$\left|(\theta^{(i)})'[(I_n - H)^{-1} - I_n]\theta^{(j)}\right| \leq \frac{C\log(n)}{\|\theta\|^4}[\max\{\|\theta\|_3^3, \sqrt{\log(n)}\theta_{max}^2\}\|\theta\|^2 + \|\theta\|_1\|\theta\|_3^3]. \tag{C.32}$$

Write $\max\{\|\theta\|_3^3, \sqrt{\log(n)}\theta_{max}^2\}\|\theta\|^2 \leq \|\theta\|_3^3\|\theta\|^2 + \sqrt{\log(n)}\theta_{max}^2\|\theta\|^2$. First, since $\|\theta\|_\infty < 1$, $\|\theta\|^2\|\theta\|_3^3 \leq \|\theta\|_1\|\theta\|_3^3$. Second, by (2.11), $\sqrt{\log(n)}\theta_{max}^2\|\theta\|^2 \leq \|\theta\|^4 \leq \|\theta\|_3^3\|\theta\|_1$. Inserting these into (C.32) gives (C.27) and the claim follows. □

C.10. Proof of Lemma C.2. By the assumptions and elementary algebra, it is seem that $(\lambda_k^{(1)}, \xi_k^{(1)})$ and $(\lambda_k^{(2)}, \xi_k^{(2)})$ take real values. By definitions, $(U + Err)\xi_k^{(2)} = \lambda_k^{(2)}\xi_k^{(2)}$, and so

$$(\xi_k^{(2)})'[\lambda_k^{(2)}I_K - U]^2\xi_k^{(2)} = (\xi_k^{(2)})'(Err)^2\xi_k^{(2)}.$$

Since $\{\xi_1^{(1)}, \xi_2^{(1)}, \ldots, \xi_K^{(1)}\}$ constitute an orthonormal base, we have

$$(\xi_k^{(2)})'[\lambda_k^{(2)}I_K - U]^2\xi_k^{(2)} = \sum_{i=1}^K (\lambda_k^{(2)} - \lambda_i^{(1)})^2(\xi_k^{(2)}, \xi_i^{(1)})^2 \geq \sum_{i\neq k}(\lambda_k^{(2)} - \lambda_i^{(1)})^2(\xi_k^{(2)}, \xi_i^{(1)})^2.$$

Combining these gives

$$\sum_{i\neq k}(\lambda_k^{(2)} - \lambda_i^{(1)})^2(\xi_k^{(2)}, \xi_i^{(1)})^2 \leq (\xi_k^{(2)})'(Err)^2\xi_k^{(2)} \leq \|Err\|^2. \tag{C.33}$$

By the assumption of $\|Err\| \leq (1/2)\text{eigsp}(U)$, for all $1 \leq i \leq K$ and $i \neq k$, $|\lambda_k^{(2)} - \lambda_i^{(1)}| \geq (1/2)\text{eigsp}(U)$. Inserting this into (C.33) gives

$$\sum_{i\neq k}(\xi_k^{(2)}, \xi_i^{(1)})^2 \leq 4\|Err\|^2/\text{eigsp}(U)^2.$$

Since $(\xi_k^{(2)}, \xi_k^{(1)})^2 = 1 - \sum_{i\neq k}(\xi_k^{(2)}, \xi_i^{(1)})^2$, it follows that $(\xi_k^{(2)}, \xi_k^{(1)})^2 \geq 1 - 4\|Err\|^2/\text{eigsp}(U)^2$, and the claim follows by basic algebra. □

C.11. Proof of Lemma C.3. Write for short $U = \text{diag}(\Omega)$ and $H = (W - \text{diag}(\Omega))/\hat{\lambda}_k$. Similarly, write

$$\hat{\theta}^{(i,k)} - \theta^{(i)} = [(I_n - H)^{-1} - I_n]\theta^{(i)} = (I_n - H)^{-1}H\theta^{(i)}.$$

By Lemma 2.3, with probability at least $1 + o(n^{-3})$, $\|H\| = o(1)$,

$$\|\hat{\theta}^{(i,k)} - \theta^{(i)}\| \leq \|(I_n - H)^{-1}\|\|H\theta^{(i)}\| \lesssim \|H\theta^{(i)}\|. \tag{C.34}$$

Next, by (2.16), with probability at least $1 + o(n^{-3})$, $\hat{\lambda}_k \asymp \|\theta\|^2$. It follows from basic algebra that

$$\|H\theta^{(i)}\|^2 \leq \frac{1}{\hat{\lambda}_k^2}(I + II) \leq C(I + II)/\|\theta\|^4, \tag{C.35}$$

where $I = \|W\theta^{(i)}\|^2$ and $II = \|U\theta^{(i)}\|^2$. Note that, first, since $\|\theta\|_\infty < 1$,

$$II \leq \|\theta\|_6^6 \leq \|\theta\|_3^3. \tag{C.36}$$

Second, by the assumption (B.1) and Lemma B.2, with probability at least $1 + o(n^{-3})$,

$$I \leq C\log(n)\|\theta\|_1\|\theta\|_3^3. \tag{C.37}$$

Combining (C.36)-(C.37) with (C.34)-(C.35) gives

$$\|\hat{\theta}^{(i,k)} - \theta^{(i)}\|^2 \leq C\log(n)(\|\theta\|_1 + 1)\|\theta\|_3^3/\|\theta\|^4. \tag{C.38}$$

By (2.11) and basic algebra, $\|\theta\|_1 \geq \|\theta\|^2 \geq \log(n)$, and so $\|\theta\|_1 + 1 \leq 2\|\theta\|_1$. Inserting this into (C.38) gives the claim. □

C.12. Proof of Lemma C.4. Write for short $U = \mathrm{diag}(\Omega)$ and $H = (W - \mathrm{diag}(\Omega))/\hat{\lambda}_k$. Since $(I_n - H)^{-1} - I_n = H + H(I_n - H)^{-1}H$, it follows from definitions and basic algebra that

$$\|\Theta^{-1}[\hat{\theta}^{(i,k)} - \theta^{(i)}]\|^2 = \|\Theta^{-1}[(I_n - H)^{-1} - I_n]\theta^{(i)}\|^2 \leq 2(I + II), \tag{C.39}$$

where

$$I = \|\Theta^{-1}H\theta^{(i)}\|^2, \qquad II = \|\Theta^{-1}H(I_n - H)^{-1}H\theta^{(i)}\|^2.$$

Consider I first. Similarly, by Lemma 2.3, with probability at least $1 + o(n^{-3})$, $\hat{\lambda}_k \asymp \|\theta\|^2$, and so

$$I = \frac{1}{\hat{\lambda}_k^2}\|\Theta^{-1}W\theta^{(i)} - \Theta^{-1}U\theta^{(i)}\| \leq \frac{C}{\|\theta\|^4}[Ia + Ib], \tag{C.40}$$

where

$$Ia = \|\Theta^{-1}W\theta^{(i)}\|^2, \qquad Ib = \|\Theta^{-1}U\theta^{(i)}\|^2.$$

Now, first, since $\|\theta\|_\infty < 1$,

$$Ib \leq \|\theta\|_4^4 \leq \|\theta\|_3^3. \tag{C.41}$$

Second, by (B.1) and Lemma B.2, with probability at least $1 + o(n^{-3})$,

$$Ia \leq C\log(n)\|\theta\|_3^3\sum_{i=1}^{n}(1/\theta(i)). \tag{C.42}$$

Inserting (C.41)-(C.42) into (C.40) gives that with probability at least $1 + o(n^{-3})$,

$$I \le C\log(n)\frac{\|\theta\|_3^3}{\|\theta\|^4}\cdot\Big[1+\sum_{i=1}^n \frac{1}{\theta(i)}\Big] \le C\log(n)\frac{\|\theta\|_3^3}{\|\theta\|^4}\cdot\Big[\sum_{i=1}^n \frac{1}{\theta(i)}\Big]. \tag{C.43}$$

Next, we analyze II. By definitions, $II = \frac{1}{\hat{\lambda}_k^4}\|\Theta^{-1}(W-U)(I_n-H)^{-1}(W-H)\theta^{(i)}\|^2$. Recalling that $\hat{\lambda}_k \asymp \|\theta\|^2$ with probability at least $1+o(n^{-3})$, and so by basic algebra,

$$II \le \frac{1}{\|\theta\|^8}\|\Theta^{-1}(W-U)(I_n-H)^{-1}(W-U)\theta^{(i)}\|^2 \le \frac{1}{\|\theta\|^8} IIa\cdot IIb, \tag{C.44}$$

where $IIa = \|\Theta^{-1}(W-U)(I-H)^{-1}\|^2$ and $IIb = \|(W-U)\theta^{(i)}\|^2$.

Consider IIa first. By Lemma 2.2, with probability $1+o(n^{-3})$, $\|H\| = o(1)$. Therefore,

$$IIa \lesssim \|\Theta^{-1}(W-U)\|^2 \le C\big[\|\Theta^{-1}W\|^2 + \|\Theta^{-1}U\|^2\big].$$

Next, by (B.2) and Lemma B.1, we have with probability at least $1+o(n^{-3})$,

$$\|\Theta^{-1}W\|^2 \le C\log(n)\max\{\theta_{max}\sum_{i=1}^n \frac{1}{\theta(i)},\ \frac{1}{\theta_{min}}\|\theta\|_1\}.$$

At the same time, it is seen $\|\Theta^{-1}U\|^2 \le 1$, which is much smaller than the right hand side of the equation above. Combining these gives

$$IIa \le C\log(n)\max\{\theta_{max}\sum_{i=1}^n \frac{1}{\theta(i)},\ \frac{1}{\theta_{min}}\|\theta\|_1\}. \tag{C.45}$$

Next, we consider IIb. Write

$$IIb = \|(W-U)\theta^{(i)}\|^2 \le C\big[\|W\theta^{(i)}\|^2 + \|U\theta^{(i)}\|^2\big].$$

On one hand, by Lemma B.2, with probability at least $1+o(n^{-3})$,

$$\|W\theta^{(i)}\|^2 \le C\log(n)\|\theta\|_1\cdot\|\theta\|_3^3.$$

On the other hand, since $\|\theta\|_\infty < 1$, by definitions and direct calculations,

$$\|U\theta^{(i)}\|^2 \le \|\theta\|_4^4 \le \|\theta\|_3^3.$$

Recall that that (2.11) implies $\|\theta\|_1 \geq \|\theta\|^2 \geq 1$. Combining these gives

$$IIb \leq C\log(n)[1 + \|\theta\|_1]\|\theta\|_3^2 \leq C\log(n)\|\theta\|_1 \cdot \|\theta\|_3^3. \tag{C.46}$$

Inserting (C.45)-(C.46) into (C.44) gives

$$II \leq C\log^2(n)\frac{\|\theta_1\| \cdot \|\theta\|_3^3}{\|\theta\|^8}\max\{\theta_{max}\sum_{i=1}^n \frac{1}{\theta(i)},\ \frac{1}{\theta_{min}}\|\theta\|_1\}. \tag{C.47}$$

Inserting (C.43) and (C.47) into (C.39), $\|\Theta^{-1}[\hat{\theta}^{(i,k)} - \theta^{(i)}]\|^2$ does not exceed

$$\frac{C\log(n)\|\theta\|_3^3}{\|\theta\|^4}\Bigg[\sum_{i=1}^n \frac{1}{\theta(i)} + \frac{\log(n)\|\theta\|_1}{\|\theta\|^4}\max\{\theta_{max}\sum_{i=1}^n \frac{1}{\theta(i)},\ \frac{1}{\theta_{min}}\|\theta\|_1\}\Bigg].$$

By (2.11), $\log(n)\theta_{max}\|\theta\|_1/\|\theta\|^4 \to 0$, and so

$$\|\Theta^{-1}[\hat{\theta}^{(i,k)} - \theta^{(i)}]\|^2 \leq \frac{C\log(n)\|\theta\|_3^3}{\|\theta\|^4}\Bigg[\sum_{i=1}^n \frac{1}{\theta(i)} + \frac{1}{\theta_{min}}\frac{\log(n)\|\theta\|_1^2}{\|\theta\|^4}\Bigg],$$

and the claim follows. □

REFERENCES

[1] Bai Z, Silverstein J (2009) *Spectral analysis of large dimensional random matrices.* Springer, NY.

[2] Horn R, Johnson C (1985) *Matrix analysis.* Cambridge Press.

[3] Jin, J. (2013). Fast community detection by SCORE. *arXiv:1211.5803.*

[4] Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

[5] Tropp J (2012) User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12**, 389-434.

J. Jin
Department of Statistics
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
USA
E-mail: jiashun@stat.cmu.edu