SUPPLEMENT TO "COMMUNITY DETECTION IN DEGREE-CORRECTED BLOCK MODELS"

By Chao Gao, Zongming Ma, Anderson Y. Zhang, and Harrison H. Zhou

University of Chicago, University of Pennsylvania, Yale University and Yale University

In this supplement, we first present additional graphical comparisons of numeric studies. Then we present proofs of the main results, including the proofs of Theorem 1 for the case of $k = 2$, of Theorem 2, and of all main results in Section 3. Moreover, we provide a careful investigation of key properties of the quantity $J_t(p, q)$ defined in (13), followed by proofs of auxiliary results.

APPENDIX A: ADDITIONAL GRAPHICS OF NUMERIC RESULTS

A.1. Comparison of Running Times. We compare running times of all six algorithms mentioned in Section 4 under both Scenarios 1 and 2. The comparison was conducted in Matlab on a laptop computer with 1.3 GHz Intel Core i5 CPU and 8GB 1600 MHz DDR3 RAM. The implementation of CMM was kindly provided by the original authors.

From both Figure 4 and Figure 5, we can clearly see that CMM is much slower than all the other methods. Among all the remaining five methods, SCORE and RSC outperform in speed. However, their advantages in speed come at a cost of statistical accuracy.

A.2. Comparison of Performance of Algorithm 1 with Different τ's. We consider Algorithm 1 under Scenarios 1 and 2 with different choices of parameter τ. We let $\tau = c(\sum_{i,j} A_{i,j})/n$, and vary the value of c among $\{2, 5, 10, 20, 30, 50\}$. Shown in Figure 6, the errors of Algorithm 1 are nearly identical once c is greater than 5.

APPENDIX B: ADDITIONAL PROOFS OF MAIN RESULTS

In order for Lemma 2 to be applied to lower bounding the performance of community detection in DCBM, we need a version of Lemma 2 that can handle approximately equal sizes. To be specific, suppose $X = (X_1, \ldots, X_m, X_{m+1}, \ldots, X_{m+m_1})$ have independent Bernoulli entries. Given $1 \geq p > q \geq 0$ and $\theta_0, \theta_1, \ldots, \theta_{m+m_1} > 0$ such that

$$\frac{1}{m}\sum_{i=1}^{m}\theta_i, \frac{1}{m_1}\sum_{i=m+1}^{m+m_1}\theta_i \in [1-\delta, 1+\delta]. \tag{34}$$

When m and m_1 are approximately equal, we are interested in understanding the minimum possible Type I+II error of testing

$$\begin{aligned} H_0 : X \sim \bigotimes_{i=1}^{m}\text{Bern}\left(\theta_0\theta_i p\right) \otimes \bigotimes_{i=m+1}^{m+m_1}\text{Bern}\left(\theta_0\theta_i q\right) \\ \text{vs.} \quad H_1 : X \sim \bigotimes_{i=1}^{m}\text{Bern}\left(\theta_0\theta_i q\right) \otimes \bigotimes_{i=m+1}^{m+m_1}\text{Bern}\left(\theta_0\theta_i p\right). \end{aligned} \tag{35}$$

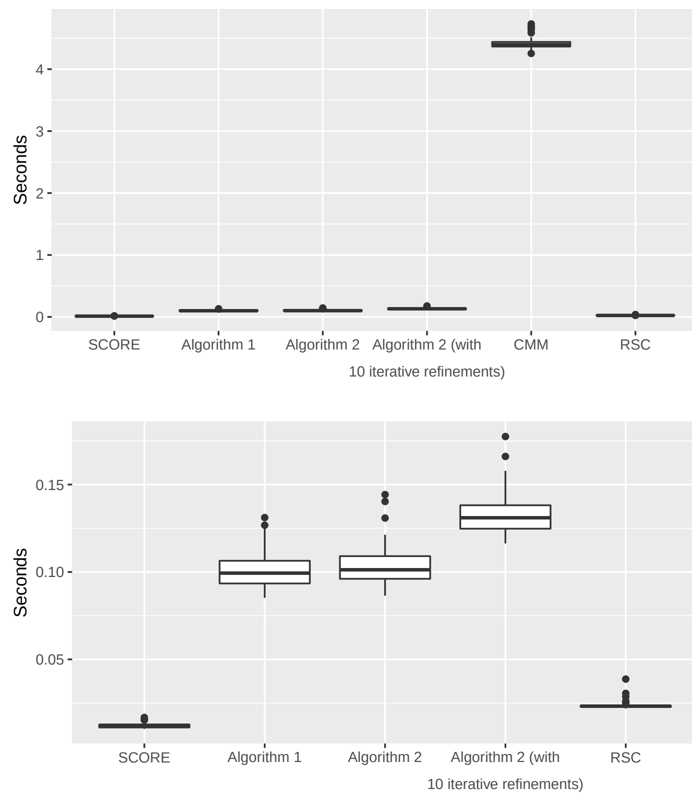

FIG 4. *Comparison of speed for all six methods under Scenario 1 (top), and the same comparison without CMM displayed (bottom).*

The setting of Lemma 2 is a special case where $m = m_1$ and $\delta = 0$.

LEMMA 5. *Suppose that as* $m \to \infty$, $1 < p/q = O(1)$, $p \max_{0 \le i \le 2m} \theta_i^2 = o(1)$, $\delta = o(1)$ *and* $\left|\frac{m}{m_1} - 1\right| = o(1)$. *If* $\theta_0 m(\sqrt{p} - \sqrt{q})^2 \to \infty$,

$$\inf_{\phi} \left(P_{H_0}\phi + P_{H_1}(1-\phi)\right) \ge \exp\left(-(1+o(1))\theta_0 m(\sqrt{p}-\sqrt{q})^2\right).$$

Otherwise, there exists a constant $c \in (0,1)$ *such that* $\inf_\phi \left(P_{H_0}\phi + P_{H_1}(1-\phi)\right) \ge c$.

B.1. Proof of Theorem 1 for $k = 2$. By the definition of the loss function, $n\ell(\tilde{z}, z) \le n/2$ for any $\tilde{z} \in [2]^n$. Therefore, we only need to calculate $\mathbb{P}(n\ell(\hat{z}, z) = m)$ for $1 \le m \le n/2$. We will keep using the definitions $\Gamma_{u,v} = \{i : z(i) = u, \tilde{z}(i) = v\}$, $\mathcal{C}_u = \Gamma_{u,1} \cup \Gamma_{u,2}$ and $\Gamma = \Gamma_{1,2} \cup \Gamma_{2,1}$ for all $u, v \in [2]$. Recall in (24) we have shown

$$\mathbb{P}\left(L(\tilde{z}) > L(z)\right) \le \prod_{\substack{i<j \\ \tilde{Y}_{ij} \ne Y_{ij}}} \exp\left(-\frac{1}{2}\theta_i\theta_j(\sqrt{p}-\sqrt{q})^2\right).$$

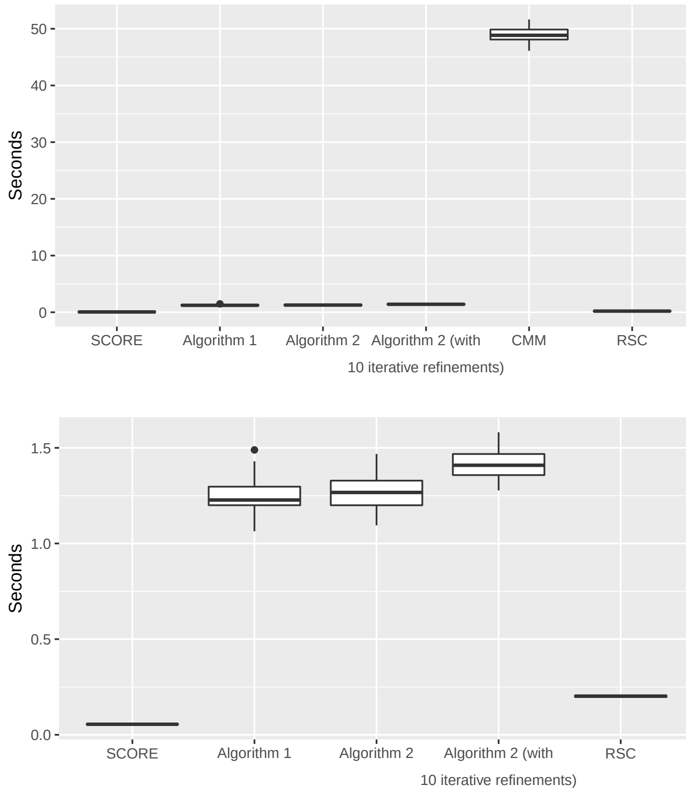

FIG 5. *Comparison of speed for all six methods under Scenario 2 (top), and the same comparison without CMM displayed (bottom).*

Since

$$\begin{aligned}\sum_{\substack{i<j\\ \tilde{Y}_{ij}\neq Y_{ij}}}\theta_i\theta_j &= \sum_{i\in\Gamma_{1,1}}\theta_i\sum_{i\in\Gamma_{1,2}}\theta_i+\sum_{i\in\Gamma_{1,1}}\theta_i\sum_{i\in\Gamma_{2,1}}\theta_i+\sum_{i\in\Gamma_{1,2}}\theta_i\sum_{i\in\Gamma_{2,1}}\theta_i+\sum_{i\in\Gamma_{2,1}}\theta_i\sum_{i\in\Gamma_{2,2}}\theta_i\\ &\geq \sum_{i\in\Gamma}\theta_i\left((1-\delta)n-\sum_{i\in\Gamma}\theta_i\right),\end{aligned}$$

we have

$$\mathbb{P}\left(L(\tilde{z})>L(z)\right)\leq\exp\left(-\frac{1}{2}\sum_{i\in\Gamma}\theta_i\left((1-\delta)n-\sum_{i\in\Gamma}\theta_i\right)(\sqrt{p}-\sqrt{q})^2\right). \tag{36}$$

Denote $m'=\eta n$ for some $\eta=o(1)$ satisfying $\eta^{-1}=o(I)$. We define τ exactly the same way as in Section 5.2. We use the notation

$$R_t=\frac{1}{n}\sum_{i=1}^{n}\exp\left(-(1-t)\theta_i\frac{n}{2}(\sqrt{p}-\sqrt{q})^2\right).$$

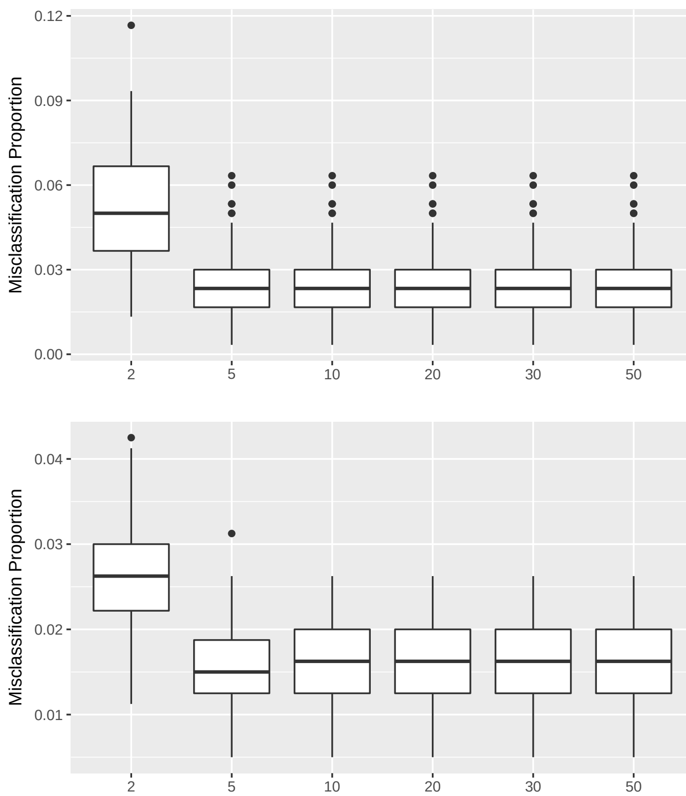

FIG 6. *The effect of different choices of τ for Scenario 1 (left) and Scenario 2 (right). The x-axis is the value of c with $\tau = c(\sum_{i,j} A_{i,j})/n$.*

Recall the constant M used in Section 5.2.
Case 1: $1 \leq m \leq M$. By (36), we have

$$\mathbb{P}(n\ell(\hat{z}, z) = m) \leq \sum_{|\Gamma|=m} \exp\left(-\frac{1}{2}\sum_{i\in\Gamma}\theta_i\left((1-\delta)n - \sum_{i\in\Gamma}\theta_i\right)(\sqrt{p}-\sqrt{q})^2\right).$$

Using the argument in Section 5.2, we have $\mathbb{P}(n\ell(\tilde{z}, z) = m) \leq (2nMR_{\delta+2\eta})^m$.
Case 2: $M \leq m \leq m'$. We have $\sum_{i\in\tau(\Gamma)}\theta_i \leq 2|\tau(\Gamma)| \leq 2\eta n$ due to (26). Note that $n - \sum_{i\in\Gamma}\theta_i = \sum_{i\in\Gamma^c}\theta_i \geq |\Gamma^c|\theta_{\min} \geq n\theta_{\min}/2$. For any $m \leq m'$, using the monotone property of $x(1-x)$ for $x \in [0,1]$, we have

$$\sum_{i\in\Gamma}\theta_i\left((1-\delta)n - \sum_{i\in\Gamma}\theta_i\right) \geq \sum_{i\in\tau(\Gamma)}\theta_i\left((1-\delta)n - \sum_{i\in\tau(\Gamma)}\theta_i\right) \geq \sum_{i\in\tau(\Gamma)}\theta_i(1-\delta-2\eta)n.$$

Thus, by (36), we have

$$\mathbb{P}\left(L(\tilde{z}) > L(z)\right) \leq \prod_{i\in\tau(\Gamma)}\exp\left(-\theta_i(1-\delta-2\eta)\frac{n}{2}(\sqrt{p}-\sqrt{q})^2\right).$$

Using the argument in Section 5.2, we have

$$\mathbb{P}(n\ell(\tilde{z}, z) = m) \leq (2eM)^m \left(\frac{e^2 M n R_{\delta+2\eta}}{m/M}\right)^{m/M}.$$

Case 3: $m > m'$. Under this scenario, we can take an arbitrary subset $\Gamma' \subset \tau(\Gamma)$ such that $|\Gamma'| = \eta m/M$, which leads to $\sum_{i\in\Gamma'} \theta_i \leq 2\eta n$. Recall $n - \sum_{i\in\Gamma} \theta_i = \sum_{i\in\Gamma^c} \theta_i \geq |\Gamma^c|\theta_{\min} \geq n\theta_{\min}/2$. Using (36), together with the property of $x(1-x)$ for $x \in [0,1]$, we have

$$\begin{aligned}\mathbb{P}\left(L(\tilde{z}) > L(z)\right) &\leq \exp\left(-\frac{1}{2}\sum_{i\in\Gamma'} \theta_i \left((1-\delta)n - \sum_{i\in\Gamma'} \theta_i\right)(\sqrt{p}-\sqrt{q})^2\right)\\ &\leq \prod_{i\in\Gamma'} \exp\left(-\theta_i(1-\delta-2\eta)\frac{n}{2}(\sqrt{p}-\sqrt{q})^2\right).\end{aligned}$$

By the argument used in Section 5.2, we have

$$\mathbb{P}(n\ell(\tilde{z}, z) = m) \leq (2eM)^m \left(\frac{e^2 M n R_{\delta+2\eta}}{\eta^2 m/M}\right)^{\eta m/M}.$$

Note that the above rate involves $R_{\delta+2\eta}$ instead of $R^{1/2}_{\delta+2\eta}$ for the case $k \geq 3$ in Section 5.2. This results in a tighter bound for $\mathbb{P}\left(L(\tilde{z}) > L(z)\right)$.

Finally by applying the same techniques used in Section 5.2, we obtain the desired bound for $\mathbb{E}\ell(\tilde{z}, z)$.

B.2. Proof of Theorem 2. We only state the proof for the case $k \geq 3$. The proof for the case $k = 2$ can be derived using essentially the same argument. For a label vector, recall the notation $n_u(z) = |\{i \in [n] : z(i) = u\}|$. Under Condition N, there exists a $z^* \in [k]^n$ such that $n_1(z^*) \leq n_2(z^*) \leq n_3(z^*) \leq \cdots \leq n_k(z^*)$ with $n_1(z^*) = n_2(z^*) = \lfloor n/(\beta k)\rfloor$, and that $(n_u(z^*))^{-1}\sum_{i:z^*(i)=u} \theta_i \in (1 - \frac{\delta}{4}, 1 + \frac{\delta}{4})$ for all $u \in [k]$.

1° As a first step, we define a community detection problem on a subset of the parameter space such that we can avoid the complication of label permutation. To this end, given z^*, for each $u \in [k]$, let $T_u \subset \{i : z^*(i) = u\}$ with cardinality $\left\lceil n_u(z^*) - \frac{\delta n}{4k^2\beta}\right\rceil$ collect the indices of the largest θ_i's in $\{\theta_i : z(i) = u\}$. Let $T = \cup_{u=1}^k T_u$. Define

$$Z^* = \left\{z \in [k]^n : \; z(i) = z^*(i) \text{ for all } i \in T, \frac{n}{\beta k} - 1 \leq n_u(z) \leq \frac{\beta n}{k} + 1 \text{ for all } u \neq v \in [k]\right\}.$$

Since $z^* \in Z^*$, the latter is not empty. By the definition of T and Condition N, $\max_{i\in T^c} \theta_i$ is bounded by a constant. Thus, for any z such that $z(i) = z^*(i)$ for all $i \in T$, we have

$$\frac{1}{n_u(z)} \sum_{\{i:z(i)=u\}} \theta_i \in (1-\delta, 1+\delta), \quad \text{for all } u \in [k].$$

Therefore, we can define a smaller parameter space $\mathcal{P}_n^0 = \mathcal{P}_n^0(\theta, p, q, k, \beta; \delta) \subset \mathcal{P}_n(\theta, p, q, k, \beta; \delta)$ where

(37)
$$\mathcal{P}_n^0(\theta, p, q, k, \beta; \delta) = \left\{P : P_{ij} = \theta_i\theta_j B_{z(i)z(j)}, z \in Z^*, B_{uu} = p, \forall u \in [k],\; B_{uv} = q, \forall u \neq v\right\}.$$

So we have

$$\inf_{\hat z}\sup_{\mathcal{P}_n(\theta,p,q,k,\beta;\delta)}\mathbb{E}n\ell(\hat z,z)\geq\inf_{\hat z}\sup_{\mathcal{P}_n^0}\mathbb{E}n\ell(\hat z,z)=\inf_{\hat z}\sup_{\mathcal{P}_n^0}\mathbb{E}H(\hat z,z), \tag{38}$$

where $H(\cdot,\cdot)$ is the Hamming distance. Here, the equality is due to the fact that for any two $z_1,z_2\in Z^*$ they share the same labels for all indices in T. Thus, we have $H(z_1,z_2)\leq\frac{1}{2}\delta\frac{n}{\beta k}$, and so when δ is small we have $n\ell(z_1,z_2)=\inf_{\pi\in\Pi_k}H(\pi(z_1),z_2)=H(z_1,z_2)$.

2° We now turn to lower bounding the rightmost side of (38), which relies crucially on our previous discussion in Section 5.1. To this end, observe that

$$\begin{aligned}\inf_{\hat z}\sup_{\mathcal{P}_0}\mathbb{E}H(\hat z,z)&\geq\inf_{\hat z}\operatorname*{ave}_{Z^*}\mathbb{E}H(\hat z,z)\geq\sum_{i\in T^c}\inf_{\hat z(i)}\operatorname*{ave}_{Z^*}\mathbb{P}(\hat z(i)\neq z(i))\\&\geq c\frac{\delta}{k}n\frac{1}{|T^c|}\sum_{i\in T^c}\inf_{\hat z(i)}\operatorname*{ave}_{Z^*}\mathbb{P}(\hat z(i)\neq z(i)),\end{aligned}\tag{39}$$

for some constant $c>0$. Here, ave stands for arithmetic average. The first inequality holds since minimax risk is lower bounded by Bayes risk. The second inequality is due to the fact that for any $z\in Z^*$, $z(i)=z^*(i)$ for all $i\in T$, and so infimum can be taken over all $\hat z$ with $\hat z(i)=z^*(i)$ for $i\in T$. The last inequality holds because $|T^c|\geq c\frac{\delta n}{k}$ for some constant c by its definition.

We now focus on lower bounding $\inf_{\hat z(i)}\operatorname{ave}_{Z^*}\mathbb{P}(\hat z(i)\neq z(i))$ for each $i\in T^c$. Without loss of generality, suppose $1\in T^c$. Then we partition Z^* into disjoint subsets $Z^*=\cup_{u=1}^k Z_u^*$ where

$$Z_u^*=\{z\in Z^*:z(1)=u\},\quad u\in[k].$$

Note that for any $u\neq v$, there is a 1-to-1 correspondence between the elements in Z_u^* and Z_v^*. In particular, for each $z\in Z_u^*$, there exists a unique $z'\in Z_v^*$ such that $z(i)=z'(i)$ for all $i\neq 1$. Thus, we can simultaneously index all $\{Z_u^*\}_{u=1}^k$ by the second to the last coordinates of the z vectors contained in them. We use z_{-1} to indicate the subvector in $[k]^{n-1}$ excluding the first coordinate and collect all the different z_{-1}'s into a set Z_{-1}. Then we have

$$\begin{aligned}&\inf_{\hat z(1)}\operatorname*{ave}_{Z^*}\mathbb{P}(\hat z(1)\neq z(1))\\&\geq\frac{1}{k(k-1)}\sum_{u<v}\inf_{\hat z(1)}\left(\operatorname*{ave}_{Z_u^*}\mathbb{P}(\hat z(1)\neq u)+\operatorname*{ave}_{Z_v^*}\mathbb{P}(\hat z(1)\neq v)\right)\\&\geq\frac{1}{k(k-1)}\inf_{\hat z(1)}\left(\operatorname*{ave}_{Z_1^*}\mathbb{P}(\hat z(1)\neq 1)+\operatorname*{ave}_{Z_2^*}\mathbb{P}(\hat z(1)\neq 2)\right)\\&\geq\frac{1}{k(k-1)}\frac{1}{|Z_{-1}|}\sum_{z_{-1}\in Z_{-1}}\inf_{\hat z(1)}\left(\mathbb{P}_{z=(1,z_{-1})}(\hat z(1)\neq 1)+\mathbb{P}_{z=(2,z_{-1})}(\hat z(1)\neq 2)\right).\end{aligned}\tag{40}$$

Note that by the definition of z^* and Z^*, it is guaranteed that for either $(1,z_{-1})$ or $(2,z_{-1})$, $|\frac{n_1}{n_2}-1|=o(1)$. Therefore, we can apply Lemma 5 to bound from below each term in the summation of the rightmost side of the last display by $\exp\left(-(1+\eta)\theta_1\frac{n}{\beta k}(\sqrt{p}-\sqrt{q})^2\right)$ for some $\eta = o(1)$. Together with (38) – (40), this implies that

$$\begin{aligned}\inf_{\hat{z}} \sup_{\mathcal{P}} \mathbb{E}\ell(\hat{z}, z) &\geq c\frac{\delta}{k^3}\frac{1}{|T^c|}\sum_{i\in T^c}\exp\left(-(1+\eta)\theta_i\frac{n}{\beta k}(\sqrt{p}-\sqrt{q})^2\right)\\ &\geq c\frac{\delta}{k^3}\frac{1}{n}\sum_{i=1}^{n}\exp\left(-(1+\eta)\theta_i\frac{n}{\beta k}(\sqrt{p}-\sqrt{q})^2\right)\\ &\geq c\frac{\delta}{k^3}\exp\left(-(1+\eta)I\right) = \exp\left(-(1+o(1))I\right).\end{aligned}$$

Here, the first inequality is simple algebra. The second inequality holds since T^c only contains within each community defined by z^* the nodes with the smallest θ_i's. The third inequality is a direct application of Jensen's inequality, and the last equality holds since $\log k = o(I)$ and $\log \frac{1}{\delta} = o(I)$. This completes the proof.

B.3. Proofs of Lemma 1 and Corollary 2. We now prove Lemma 1 and Corollary 2, which characterize the performance of Algorithm 1. To prove Lemma 1, we need two auxiliary lemmas, whose proofs will be given in Appendix D. In the rest of this part, we let $P = (P_{ij}) = (\theta_i\theta_j B_{z(i)z(j)})$ for notational convenience.

The following lemma characterizes the connection between measure on misclassification and geometry of the point cloud. The result is not tied to any specific clustering algorithm or choice of norm.

LEMMA 6. *Let $z \in [k]^n$ be the true label for a DCBM in $\mathcal{P}_n'(\theta, p, q, k, \beta; \delta, \alpha)$. Given any $\tilde{z} \in (\{0\} \cup [k])^n$, any $\{\tilde{v}_u\}_{u\in[k]}, \{V_i\}_{i\in[n]} \subset \mathbb{R}^n$ and any $b > 0$, define*

$$\tilde{V}_i = \tilde{v}_{\tilde{z}(i)} \quad \text{for all } i \in S_0^c,$$

where $S_0 = \{i \in [n] : \tilde{z}(i) = 0\}$. Then, for any norm $\|\cdot\|$ satisfying triangle inequality, as long as

$$\min_{z(i)\neq z(j)} \|V_i - V_j\| \geq 2b, \tag{41}$$

we have

$$\min_{\pi\in\Pi_k} \sum_{\{i:\tilde{z}(i)\neq\pi(z(i))\}} \theta_i \leq \sum_{i\in S_0}\theta_i + (2\beta^2+1)\sum_{i\in S}\theta_i,$$

where $S = \left\{i \in S_0^c : \|\tilde{V}_i - V_i\| \geq b\right\}$.

LEMMA 7. *Under the settings of Lemma 1, let $\tau = C_1\left(np\|\theta\|_\infty^2 + 1\right)$ for some sufficiently large $C_1 > 0$ in Algorithm 1. Then, for any constant $C' > 0$, there exists some $C > 0$ only depending on C_1, C' and α such that*

$$\|\hat{P} - P\|_{\mathrm{F}} \leq C\sqrt{k(np\|\theta\|_\infty^2 + 1)},$$

with probability at least $1 - n^{-(1+C')}$ uniformly over $\mathcal{P}_n'(\theta, p, q, k, \beta; \delta, \alpha)$.

PROOF OF LEMMA 1. Let P_i denote the i^{th} row of P and $\bar{P}_i = \|P_i\|_1^{-1} P_i$ the ℓ_1 normalized row. By definition, for sufficiently large values of n,

$$\frac{pn}{2\beta k} \leq \frac{\|P_i\|_1}{\theta_i} = \sum_{j:j\neq i} \theta_j B_{z(i)z(j)} \leq 2\alpha np, \text{ for any } i \in [n], \tag{42}$$

under the conditions $\delta = o(1)$ and $\|\theta\|_\infty = o(n/k)$.

Note that $\bar{P}_i = \bar{P}_j$ when $z(i) = z(j)$. Our first task is to lower bound $\|\bar{P}_i - \bar{P}_j\|_1$ when $z(i) \neq z(j)$, which serves as the separation condition among different clusters. For any i and j such that $z(i) = u \neq v = z(j)$, we assume $\|P_i\|_1/\theta_i \leq \|P_j\|_1/\theta_j$ without loss of generality. Then,

$$\begin{aligned}
\|\bar{P}_i - \bar{P}_j\|_1 &\geq \sum_{l:z(l)=u} |P_{il} - P_{jl}| = \sum_{l:z(l)=u} \left| \frac{\theta_l B_{uu}}{\|P_i\|_1/\theta_i} - \frac{\theta_l B_{uv}}{\|P_j\|_1/\theta_j} \right| \\
&= \frac{1}{\|P_j\|_1/\theta_j} \sum_{l:z(l)=u} \theta_l \left| \frac{\|P_j\|_1/\theta_j}{\|P_i\|_1/\theta_i} B_{uu} - B_{uv} \right| \\
&\geq \frac{p-q}{\|P_j\|_1/\theta_j} \frac{n}{2\beta k} \qquad (43)\\
&\geq \frac{p-q}{4\alpha\beta kp}. \qquad (44)
\end{aligned}$$

Here, (43) holds since $\frac{\|P_j\|_1/\theta_j}{\|P_i\|_1/\theta_i} B_{uu} \geq p$ and $B_{uv} \leq q$, and (44) is due to (42). By switching i and j, the foregoing argument also works for the case where $\|P_i\|_1/\theta_i > \|P_j\|_1/\theta_j$. Hence,

$$\min_{z(i)\neq z(j)} \|\bar{P}_i - \bar{P}_j\|_1 \geq \frac{p-q}{4\alpha\beta kp}. \tag{45}$$

Let $\hat{z} \in (\{0\} \cup [k])^n$ and $\hat{v}_1, ..., \hat{v}_k \in \mathbb{R}^n$ denote a solution to the optimization problem (8) (with all nodes in S_0 assigned to the 0^{th} community). Define matrix $\hat{V} \in \mathbb{R}^{n\times n}$ with the i^{th} row $\hat{V}_i = \hat{v}_{\hat{z}(i)}$. If $\hat{z}(i) = 0$, set $\hat{V}_i$ as the zero vector. Define $S = \{i \in [n] : \|\hat{V}_i - \bar{P}_i\|_1 \geq \frac{p-q}{8\alpha\beta kp}\}$ and recall $S_0 = \{i \in [n] : \|\hat{P}_i\|_1 = 0\}$. Then, by the separation condition (45) and Lemma 6, we have

$$\min_{\pi\in\Pi_k} \sum_{i:\hat{z}(i)\neq\pi(z(i))} \theta_i \leq (2\beta^2 + 1) \sum_{i\in S} \theta_i + \sum_{i\in S_0} \theta_i. \tag{46}$$

In what follows, we derive bounds for $\sum_{i\in S} \theta_i$ and $\sum_{i\in S_0} \theta_i$, respectively. Recall that $\tilde{P}_i = \|\hat{P}_i\|_1^{-1} \hat{P}_i$. By the definition of $\hat{z}$ and $\hat{V}$, we have

$$\sum_{i=1}^n \|\hat{P}_i\|_1 \|\hat{V}_i - \tilde{P}_i\|_1 \leq (1+\epsilon) \sum_{i=1}^n \|\hat{P}_i\|_1 \|\bar{P}_i - \tilde{P}_i\|_1. \tag{47}$$

In order to bound $\sum_{i\in S}\theta_i$, we first derive a bound for $\sum_{i\in S}\|\hat{P}_i\|_1$. That is,

$$
\begin{aligned}
\sum_{i\in S}\|\hat{P}_i\|_1 &\leq \frac{8\alpha\beta kp}{p-q}\sum_{i\in S}\|\hat{P}_i\|_1\|\hat{V}_i-\bar{P}_i\|_1 && (48)\\
&\leq \frac{8\alpha\beta kp}{p-q}\sum_{i\in S}\Big(\|\hat{P}_i\|_1\|\hat{V}_i-\tilde{P}_i\|_1+\|\hat{P}_i\|_1\|\bar{P}_i-\tilde{P}_i\|_1\Big)\\
&\leq \frac{8(2+\epsilon)\alpha\beta kp}{p-q}\sum_{i=1}^{n}\|\hat{P}_i\|_1\|\bar{P}_i-\tilde{P}_i\|_1 && (49)\\
&\leq \frac{16(2+\epsilon)\alpha\beta kp}{p-q}\sum_{i=1}^{n}\|\hat{P}_i-P_i\|_1 && (50)\\
&\leq \frac{16(2+\epsilon)\alpha\beta nkp}{p-q}\|\hat{P}-P\|_{\mathrm{F}}, && (51)
\end{aligned}
$$

where (48) uses the definition of S, (49) is by the inequality (47), and (50) is by the inequality $\|\|x\|_1^{-1}x-\|y\|_1^{-1}y\|_1\leq\frac{2\|x-y\|_1}{\|x\|_1\vee\|y\|_1}$ which in turn is due to the triangle inequality.

Now we are ready to bound $\sum_{i\in S}\theta_i$ as

$$
\begin{aligned}
\sum_{i\in S}\theta_i &\leq \frac{2\beta k}{pn}\sum_{i\in S}\|P_i\|_1 && (52)\\
&\leq \frac{2\beta k}{pn}\sum_{i\in S}\Big(\|\hat{P}_i\|_1+\|\hat{P}_i-P_i\|_1\Big) && (53)\\
&\leq \frac{2\beta k}{pn}\left(\frac{16(2+\epsilon)\alpha\beta nkp}{p-q}\|\hat{P}-P\|_{\mathrm{F}}+n\|\hat{P}-P\|_{\mathrm{F}}\right) && (54)\\
&\leq \frac{(66+32\epsilon)\alpha\beta^2k^2}{p-q}\|\hat{P}-P\|_{\mathrm{F}}, && (55)
\end{aligned}
$$

where (52) is by the inequality (42), (53) is due to the triangle inequality, (54) uses (51) and Cauchy-Schwarz, and (55) holds since $\alpha,\beta,k\geq 1$.

We now turn to bounding $\sum_{i\in S_0}\theta_i$. To this end, simple algebra leads to

$$
\begin{aligned}
\sum_{i\in S_0}\theta_i &\leq \frac{2\beta k}{pn}\sum_{i\in S_0}\|P_i\|_1 && (56)\\
&\leq \frac{2\beta k}{pn}\sum_{i=1}^{n}\|\hat{P}_i-P_i\|_1 && (57)\\
&\leq \frac{2\beta k}{p}\|\hat{P}-P\|_{\mathrm{F}}\leq\frac{\alpha\beta^2k^2}{p-q}\|\hat{P}-P\|_{\mathrm{F}}, && (58)
\end{aligned}
$$

where (56) is by the inequality (42), (57) uses the definition of S_0 and (58) is due to the Cauchy-Schwarz inequality.

Combining the bounds in (55), (58) and (46), we have

$$
\min_{\pi\in\Pi_k}\sum_{\{i:\hat{z}(i)\neq\pi(z(i))\}}\theta_i\leq\frac{C(1+\epsilon)k^2}{p-q}\|\hat{P}-P\|_{\mathrm{F}}, \qquad (59)
$$

where we have absorbed α and β into the constant C. By Lemma 7, we have

$$\min_{\pi\in\Pi_k}\sum_{\{i:\hat{z}(i)\neq\pi(z(i))\}}\theta_i\leq C\frac{(1+\epsilon)k^{5/2}\sqrt{n\|\theta\|_\infty^2 p+1}}{p-q},$$

with probability at least $1-n^{-(1+C')}$. This completes the proof. □

PROOF OF COROLLARY 2. Under the condition $\min_i\theta_i=\Omega(1)$, the loss $\min_{\pi\in\Pi_k}\sum_{\{i:\hat{z}(i)\neq\pi(z(i))\}}\theta_i$ can be lower bounded by $n\ell(\hat{z},z)$ multiplied by a constant. Moreover, since $p\geq n^{-1}$, $k=O(1)$ and $\|\theta\|_\infty=O(1)$, the rate $\frac{k^{5/2}\sqrt{n\|\theta\|_\infty^2 p+1}}{p-q}$ is bounded by $O\left(\frac{\sqrt{np}}{p-q}\right)=O\left(\sqrt{n}|\sqrt{p}-\sqrt{q}|^{-1}\right)$. Thus, it is sufficient to show $n^{-1/2}|\sqrt{p}-\sqrt{q}|^{-1}=O(I^{-1/2})$. This is true by observing that

$$e^{-I}\geq\frac{1}{n}\sum_{i=1}^n\exp\left(-\theta_i\frac{n}{k}(\sqrt{p}-\sqrt{q})^2\right)\geq\exp\left(-\|\theta\|_\infty\frac{n}{k}(\sqrt{p}-\sqrt{q})^2\right).$$

Thus, the proof is complete. □

B.4. Proofs of Theorem 3, Theorem 4 and Corollary 3. Now we are going to give proofs of Theorem 3, Theorem 4 and Corollary 3. Note that both Theorem 3 and Corollary 3 are direct consequences of Theorem 4. The main argument in the proof of Theorem 4 is the following lemma.

LEMMA 8. *Suppose $1<p/q=O(1)$ and $\delta=o\left(\frac{p-q}{p}\right)$. If there exist two sequences $\gamma_1=o\left(\frac{p-q}{kp}\right)$ and $\gamma_2=o\left(\frac{p-q}{k^2p}\right)$, a constant $C_1>0$ and permutations $\{\pi_i\}_{i\in[n]}\subset\Pi_k$ such that*

$$\min_{i\in[n]}\mathbb{P}\left(\frac{1}{n}\sum_{j=1}^n\theta_j\mathbf{1}_{\{\hat{z}^0_{-i}(j)\neq\pi_i(z(j))\}}\leq\gamma_1,\frac{1}{n}\sum_{j=1}^n\mathbf{1}_{\{\hat{z}^0_{-i}(j)\neq\pi_i(z(j))\}}\leq\gamma_2\right)\geq 1-n^{-(1+C_1)},\tag{60}$$

uniformly for all probability distributions in $\mathcal{P}_n'(\theta,p,q,k,\beta;\delta,\alpha)$. Then, we have for all $i\in[n]$,

$$\mathbb{P}\left(\hat{z}^0_{-i}(i)\neq\pi_i(z(i))\right)\leq(k-1)\exp\left(-(1-\eta)\theta_i(n_{(1)}+n_{(2)})J_{t^*}(p,q)/2\right)+n^{-(1+C_1)}$$

uniformly for all probability distributions in $\mathcal{P}_n'(\theta,p,q,k,\beta;\delta,\alpha)$, where $\eta=o(1)$.

PROOF. In what follows, let E_i denote the event in (60). We are going to derive a bound for $\mathbb{P}\left(\hat{z}^0_{-1}(1)\neq\pi_1(z(1))\text{ and }E_1\right)$. For the sake of brevity, we are going to use $\hat{z}$ and z to denote $\hat{z}^0_{-1}$ and $\pi_1(z)$ in the proof with slight abuse of notaion. Define $n_u=|\{i\in[n]:z(i)=u\}|$, $m_u=|\{i\in[n]:\hat{z}(i)=u\}|$ and $\hat{\Theta}_u=\sum_{\{i:\hat{z}(i)=u,z(i)=u\}}\theta_i$. Without loss of generality, consider the case $z(i)=1$. Then,

$$\mathbb{P}(\hat{z}(1)\neq 1\text{ and }E_1)\leq\sum_{l=2}^k\mathbb{P}(\hat{z}(1)=l\text{ and }E_1).$$

The arguments for bounding $\mathbb{P}(\hat{z}(1) = l \text{ and } E_1)$ are the same for $l = 2, ..., k$. Thus, we only give the bound for $l = 2$ in details. By the definition, we have

$$\mathbb{P}(\hat{z}(1) = 2 \text{ and } E_1) \leq \mathbb{P}\left(\frac{1}{m_2}\sum_{\{i:\hat{z}(i)=2\}} A_{1i} \geq \frac{1}{m_1}\sum_{\{i:\hat{z}(i)=1\}} A_{1i} \text{ and } E_1\right). \tag{61}$$

Define independent random variables $X_i \sim \text{Bernoulli}(\theta_1\theta_i q)$, $Y_i \sim \text{Bernoulli}(\theta_1\theta_i p)$, and $Z_i \sim \text{Bernoulli}(\theta_1\theta_i p)$ for all $i \in [n]$. Then, a stochastic order argument bounds the right hand side of (61) by

$$\mathbb{P}\left(\frac{1}{m_2}\sum_{\{i:\hat{z}(i)=2,z(i)=2\}} X_i + \frac{1}{m_2}\sum_{\{i:\hat{z}(i)=2,z(i)=1\}} Z_i \geq \frac{1}{m_1}\sum_{\{i:\hat{z}(i)=1,z(i)=1\}} Y_i \text{ and } E_1\right). \tag{62}$$

Using Chernoff bound, for any $\lambda > 0$, we upper bound (62) by

$$\begin{aligned}
& \mathbb{E}\Bigg\{\prod_{\{i:\hat{z}(i)=2,z(i)=2\}}(\theta_1\theta_i q e^{\lambda/m_2} + 1 - \theta_1\theta_i q)\prod_{\{i:\hat{z}(i)=2,z(i)=1\}}(\theta_1\theta_i\alpha p e^{\lambda/m_2} + 1 - \theta_1\theta_i\alpha p) \\
& \quad \prod_{\{i:\hat{z}(i)=1,z(i)=1\}}(\theta_1\theta_i p e^{-\lambda/m_1} + 1 - \theta_1\theta_i p)\mathbf{1}_{\{E_1\}}\Bigg\} \\
\leq\ & \mathbb{E}\Bigg\{\exp\Bigg(\sum_{\{i:\hat{z}(i)=2,z(i)=2\}}(\theta_1\theta_i q e^{\lambda/m_2} - \theta_1\theta_i q) + \sum_{\{i:\hat{z}(i)=2,z(i)=1\}}(\theta_1\theta_i\alpha p e^{\lambda/m_2} - \theta_1\theta_i\alpha p)\Bigg) \\
& \quad \exp\Bigg(\sum_{\{i:\hat{z}(i)=1,z(i)=1\}}(\theta_1\theta_i p e^{-\lambda/m_1} - \theta_1\theta_i p)\Bigg)\mathbf{1}_{\{E_1\}}\Bigg\} \\
=\ & \mathbb{E}\left\{\exp\left(\theta_1 m_2 q(e^{\lambda/m_2} - 1) + \theta_1 m_1 p(e^{-\lambda/m_1} - 1)\right)\mathbf{1}_{\{E_1\}}\right\} && (63) \\
& \times\mathbb{E}\left\{\exp\left((\hat{\Theta}_2 - m_2)\theta_1 q(e^{\lambda/m_2} - 1\right)\mathbf{1}_{\{E_1\}}\right\} && (64) \\
& \times\mathbb{E}\left\{\exp\left((\hat{\Theta}_1 - m_1)\theta_1 p(e^{-\lambda/m_1} - 1)\right)\mathbf{1}_{\{E_1\}}\right\} && (65) \\
& \times\mathbb{E}\left\{\exp\left(\sum_{\{i:\hat{z}(i)=2,z(i)=1\}}\theta_1\theta_i\alpha p(e^{\lambda/m_2} - 1)\right)\mathbf{1}_{\{E_1\}}\right\}. && (66)
\end{aligned}$$

In what follows, we set

$$\lambda = \frac{m_1 m_2}{m_1 + m_2}\log\frac{p}{q},$$

We are going to give bounds for the four terms (63), (64), (65) and (66), respectively. On the event E_1,

$$|\hat{\Theta}_2 - m_2| \leq \left|\sum_{\{i:z(i)=2\}}\theta_i - n_2\right| + |n_2 - m_2| + \sum_{\{i:z(i)=2,\hat{z}(i)=1\}}\theta_i \leq \left(\gamma_1 + \gamma_2 + \frac{\delta\beta}{k}\right)n,$$

and

$$q|e^{\lambda/m_2}-1| = p^{\frac{m_1}{m_1+m_2}}q^{\frac{m_2}{m_1+m_2}} - q \le p-q,$$

we have

$$\mathbb{E}\left\{\exp\left((\hat{\Theta}_2-m_2)\theta_1 q(e^{\lambda/m_2}-1\right)\mathbf{1}_{\{E_1\}}\right\} \le \exp\left(n\left(\gamma_1+\gamma_2+\frac{\delta\beta}{k}\right)\theta_1(p-q)\right),$$

which is a bound for (64). A similar argument leads to a bound (65), which is

$$\mathbb{E}\left\{\exp\left((\hat{\Theta}_1-m_1)\theta_1 p(e^{-\lambda/m_1}-1)\right)\mathbf{1}_{\{E_1\}}\right\} \le \exp\left(n\left(\gamma_1+\gamma_2+\frac{\delta\beta}{k}\right)\theta_1(p-q)\right).$$

The last term (66) has a bound

$$\mathbb{E}\left\{\exp\left(\sum_{\{i:\hat{z}(i)=2,z(i)=1\}}\theta_1\theta_i\alpha p(e^{\lambda/m_2}-1)\right)\mathbf{1}_{\{E_1\}}\right\} \le \exp\left(n\gamma_1\alpha\theta_1(p-q)\right).$$

Finally, we need a bound for (63). With the current choice of λ,

$$-m_2q(e^{\lambda/m_2}-1)-m_1p(e^{-\lambda/m_1}-1) = \frac{1}{2}(m_1+m_2)J_{\frac{m_1}{m_1+m_2}}(p,q).$$

Note that

$$\begin{aligned}
& \left|(m_1+m_2)J_{\frac{m_1}{m_1+m_2}}(p,q)-(n_1+n_2)J_{\frac{n_1}{n_1+n_2}}(p,q)\right| \\
\le\ & |n_1+n_2-m_1-m_2|J_{\frac{n_1}{n_1+n_2}}(p,q)+(m_1+m_2)\left|J_{\frac{m_1}{m_1+m_2}}(p,q)-J_{\frac{n_1}{n_1+n_2}}(p,q)\right| \\
\le\ & n\gamma_2 J_\tau(p,q)+n(1+\gamma_2)\left|J_\tau(p,q)-J_{\hat{\tau}}(p,q)\right|,
\end{aligned}$$

where $\tau=\frac{n_1}{n_1+n_2}$ and $\hat{\tau}=\frac{m_1}{m_1+m_2}$. We will give a bound for $|J_\tau(p,q)-J_{\hat{\tau}}(p,q)|$. Since $\left|\frac{\partial}{\partial\tau}J_\tau(p,q)\right| = \frac{1}{2}\left|(p-q)-p^tq^{1-t}\log\frac{p}{q}\right| \le \frac{1}{2}|p-q|+\frac{1}{2}p|\log p-\log q| \le |p-q|$, we have $|J_\tau(p,q)-J_{\hat{\tau}}(p,q)| \le |p-q||\tau-\hat{\tau}| \le \beta k\gamma_2(p-q)$. Hence, we have a bound for (63), which is

$$\begin{aligned}
& \mathbb{E}\left\{\exp\left(\theta_1 m_2 q(e^{\lambda/m_2}-1)+\theta_1 m_1 p(e^{-\lambda/m_1}-1)\right)\mathbf{1}_{\{E_1\}}\right\} \\
\le\ & \exp\left(-\frac{1}{2}\theta_1(n_1+n_2)J_{\frac{n_1}{n_1+n_2}}(p,q)+\frac{1}{2}\theta_1 n\gamma_2 J_{\frac{n_1}{n_1+n_2}}(p,q)+\frac{1}{2}\theta_1 n(1+\gamma_2)k\beta\gamma_2(p-q)\right).
\end{aligned}$$

Combining the above bounds for (63), (64), (65) and (66), we have

$$\begin{aligned}
& \mathbb{P}\left(\hat{z}(1)=2 \text{ and } E_1\right) \\
\le\ & \exp\left(-\frac{1}{2}\theta_1(n_1+n_2)J_{\frac{n_1}{n_1+n_2}}(p,q)\right) \\
& \times\exp\left(-\frac{1}{2}\theta_1 n\gamma_2 J_{\frac{n_1}{n_1+n_2}}(p,q)+\left[(2+\alpha)\gamma_1+(2+k\beta)\gamma_2+\frac{2\delta\beta}{k}\right]\theta_1 n(p-q)\right).
\end{aligned}$$

By the property of $J_t(p,q)$ stated in Lemma 12, $J_{\frac{n_1}{n_1+n_2}}(p,q) \geq (4\beta^2)^{-1}\frac{(p-q)^2}{p}$. Then, under the assumptions $\gamma_1 = o\left(\frac{p-q}{p}\right)$, $\gamma_2 = o\left(\frac{p-q}{pk}\right)$ and $\delta = o\left(\frac{k(p-q)}{p}\right)$, we have

$$\mathbb{P}\left(\hat{z}(1) = 2 \text{ and } E_1\right) \leq \exp\left(-\frac{1}{2}(1-\eta)\theta_1(n_1+n_2)J_{\frac{n_1}{n_1+n_2}}(p,q)\right),$$

for some $\eta = o(1)$. The same bound also holds for $\mathbb{P}\left(\hat{z}(1) = l \text{ and } E_1\right)$ for $l = 2, ..., k$. Thus, a union bound argument gives

$$\mathbb{P}(\hat{z}(1) \neq 1 \text{ and } E_1) \leq (k-1)\exp\left(-\frac{1}{2}(1-\eta)\theta_1(n_1+n_2)J_{\frac{n_1}{n_1+n_2}}(p,q)\right).$$

Hence,

$$\mathbb{P}(\hat{z}(1) \neq 1) \leq (k-1)\exp\left(-\frac{1}{2}(1-\eta)\theta_1(n_1+n_2)J_{\frac{n_1}{n_1+n_2}}(p,q)\right) + n^{-(1+C_1)}.$$

Now let us use the original notation and apply the above argument for each node, which leads to the bound

$$\mathbb{P}\left(\hat{z}^0_{-i}(i) \neq \pi_i(z(i))\right) \leq (k-1)\exp\left(-\frac{1}{2}(1-\eta)\theta_i \min_{u\neq v}\left[(n_u+n_v)J_{\frac{n_u}{n_u+n_v}}(p,q)\right]\right) + n^{-C_1},$$

for all $\in [n]$. By the property of $J_t(p,q)$ stated in Lemma 9, $\min_{u\neq v}\left[(n_u+n_v)J_{\frac{n_u}{n_u+n_v}}(p,q)\right] = (n_{(1)}+n_{(2)})J_{t^*}(p,q)$, with t^* specified by (15). Thus, the proof is complete. □

PROOF OF THEOREM 4. It is sufficient to check that the initial clustering step satisfies (60) with $\gamma_1 = o\left(\frac{p-q}{kp}\right)$ and $\gamma_2 = o\left(\frac{p-q}{k^2p}\right)$. This can be done using the bound in Lemma 1 under the assumptions (16) and (17). Note that the n initial clustering results $\{\hat{z}^0_{-i}\}$ may not correspond to the same permutation. This problem can be taken care of by the consensus step (12). Details of the argument are referred to the proof of Theorem 2 in [2]. □

PROOFS OF THEOREM 3 AND COROLLARY 3. Theorem 3 is a direct implication of Theorem 4 by observing $I = J$ when $\beta = 1$. The fact that $(n_{(1)}+n_{(2)})J_{t^*}(p,q) \geq 2n_{(1)}J_{1/2}(p,q) \geq \frac{2n}{\beta k}(\sqrt{p}-\sqrt{q})^2$ by Lemma 11 implies the result for the case $k \geq 3$ in Corollary 3. For $k = 2$, observe that

$$\frac{1}{n}\sum_{i=1}^{n}\exp\left(-\theta_i\frac{n}{2\beta}(\sqrt{p}-\sqrt{q})^2\right) \leq \left[\frac{1}{n}\sum_{i=1}^{n}\exp\left(-\theta_i\frac{n}{2}(\sqrt{p}-\sqrt{q})^2\right)\right]^{\beta}.$$

This implies the result for $k = 2$ in Corollary 3. □

APPENDIX C: PROPERTIES OF $J_T(P,Q)$

In this section, we study the quantity $J_t(p,q)$ defined in (13). We will state some lemmas about some useful properties of $J_t(p,q)$ that we have used in the paper. Recall that for $p, q, t \in (0,1)$,

$$J_t(p,q) = 2(tp + (1-t)q - p^tq^{1-t}).$$

LEMMA 9. *Given $p, q \in (0,1)$, let $f(x_1, x_2) = x_1 p + x_2 q - (x_1 + x_2) p^{\frac{x_1}{x_1+x_2}} q^{\frac{x_2}{x_1+x_2}}$ where $x_1, x_2 > 0$. Then the function f is increasing in terms of x_1 and x_2, respectively.*

PROOF. By differentiating f against x_1 we get

$$\frac{\partial f(x_1,x_2)}{\partial x_1} = p - q\left(\frac{p}{q}\right)^{\frac{x_1}{x_1+x_2}} - q\left(\frac{p}{q}\right)^{\frac{x_1}{x_1+x_2}} \log\left(\frac{p}{q}\right)\frac{x_2}{x_1+x_2}.$$

Thus $\lim_{x_1\to\infty} \frac{\partial f(x_1,x_2)}{\partial x_1} = 0$. Moreover,

$$\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2} = -q\left(\frac{p}{q}\right)^{\frac{x_1}{x_1+x_2}} \log^2\left(\frac{p}{q}\right)\frac{x_2^2}{(x_1+x_2)^3} \leq 0,$$

Therefore, $\frac{\partial f(x_1,x_2)}{\partial x_1} \geq 0$ for all $x_1, x_2 > 0$. This shows $f(x_1, x_2)$ is increasing with respect to x_1. Similarly we can prove that $f(x_1, x_2)$ is also an increasing function in terms of x_2. □

LEMMA 10. *For any $0 < q < p < 1$ and $0 < t \leq \frac{1}{2}$, we have*

$$J_t(p,q) \leq J_{1-t}(p,q).$$

PROOF. Define $S(t) = \frac{1}{2}\left(J_t(p,q) - J_{1-t}(p,q)\right) = (2t-1)(p-q) - \left(q(\frac{p}{q})^t - p(\frac{q}{qp})^t\right)$. Then, we have

$$S''(t) = -\log^2\left(\frac{q}{p}\right)\left(p^t q^{1-t} - p^{1-t}q^t\right) \geq 0.$$

Since $S(0) = S(1/2) = 0$, we have $S(t) \leq 0$ for all $t \in (0, 1/2]$. □

LEMMA 11. *For any $0 < q < p < 1$ and $0 < x_1 \leq x_2$, we have*

$$2x_1 J_{1/2}(p,q) \leq (x_1+x_2) J_{\frac{x_1}{x_1+x_2}}(p,q) \leq (x_1+x_2) J_{1/2}(p,q).$$

PROOF. The first inequality $2x_1 J_{1/2}(p,q) \leq (x_1+x_2) J_{\frac{x_1}{x_1+x_2}}(p,q)$ is a consequence of Lemma 9 and $x_1 \leq x_2$. Since $\left(\frac{\partial}{\partial t}\right)^2 J_t(p,q) = -2p^t q^{1-t}\log^2\left(\frac{p}{q}\right) \leq 0$, $J_t(p,q)$ is concave in t. Thus,

$$\frac{1}{2}\left(J_t(p,q) + J_{1-t}(p,q)\right) \leq J_{1/2}(p,q). \tag{67}$$

When $t \in (0, 1/2]$, $J_t(p,q) \leq \frac{1}{2}\left(J_t(p,q) + J_{1-t}(p,q)\right)$ according to Lemma 10. Thus, $J_t(p,q) \leq J_{1/2}(p,q)$, which leads to the second inequality $(x_1+x_2) J_{\frac{x_1}{x_1+x_2}}(p,q) \leq (x_1+x_2) J_{1/2}(p,q)$ by the assumption $x_1 \leq x_2$. □

LEMMA 12. *For any $0 < p, q, t < 1$, we have*

$$2\min(t, 1-t)(\sqrt{p} - \sqrt{q})^2 \leq J_t(p,q) \leq 2(\sqrt{p} - \sqrt{q})^2. \tag{68}$$

Moreover, if $\max(p/q, q/p) \leq M$, then we have

$$J_t(p,q) \leq \left(2 + \frac{4M^4}{3}\right)\min(t, 1-t)\frac{(p-q)^2}{\min(p,q)}. \tag{69}$$

PROOF. Without loss of generality, let $p > q$. We first consider the case $0 < t \leq 1/2$. By Lemma 11, we have

$$2\frac{x_1}{x_1+x_2}J_{1/2}(p,q) \leq J_{\frac{x_1}{x_1+x_2}}(p,q) \leq J_{1/2}(p,q).$$

Let $\frac{x_1}{x_1+x_2} = t$, and we have

$$2t(\sqrt{p}-\sqrt{q})^2 \leq J_t(p,q) \leq (\sqrt{p}-\sqrt{q})^2. \tag{70}$$

Now we consider the case $1/2 \leq t < 1$. Let $s = 1-t$. By Lemma 10 and (70), we have

$$J_t(p,q) \geq J_s(p,q) \geq 2s(\sqrt{p}-\sqrt{q})^2 = 2(1-t)(\sqrt{p}-\sqrt{q})^2.$$

Using (67) and (70), we have

$$J_t(p,q) \leq 2J_{1/2}(p,q) - J_s(p,q) \leq 2J_{1/2}(p,q) - 2s(\sqrt{p}-\sqrt{q})^2 = 2t(\sqrt{p}-\sqrt{q})^2 \leq 2(\sqrt{p}-\sqrt{q})^2.$$

Hence,

$$2(1-t)(\sqrt{p}-\sqrt{q})^2 \leq J_t(p,q) \leq 2(\sqrt{p}-\sqrt{q})^2. \tag{71}$$

Combine (70) and (71), and we can derive (68) for $p > q$. A symmetric argument leads to the same result for $p < q$. When, $p = q$, the result trivially holds. Thus, the proof for (68) is complete.

To prove (69), we use the identity

$$\frac{1}{2t(1-t)}J_t(p,q) = p\frac{1}{1-t}\left(1-\left(\frac{q}{p}\right)^{1-t}\right) + q\frac{1}{t}\left(1-\left(\frac{p}{q}\right)^{t}\right).$$

By Taylor's theorem, we have

$$\frac{1}{\alpha}(1-x^\alpha) = 1-x+\frac{1}{2}(1-\alpha)(x-1)^2 - \frac{1}{6}(\alpha-1)(\alpha-2)\xi^{\alpha-3}(x-1)^3,$$

for some ξ between x and 1. Thus, using the condition that $\max(p/q, q/p) \leq M$, we have

$$\frac{1}{2t(1-t)}J_t(p,q) \leq \left(1+\frac{2M^4}{3}\right)\frac{(p-q)^2}{\min(p,q)}.$$

Then, we can derive (69) by the fact that $t(1-t) \leq \min(t, 1-t)$. □

APPENDIX D: PROOFS OF AUXILIARY RESULTS

PROOF OF LEMMA 2. Note that (19) is a simple vs. simple hypothesis testing problem. By the Neyman–Pearson lemma, the optimal test is the likelihood ratio test ϕ which rejects H_0 if

$$\prod_{i=1}^{m}(\theta_0\theta_i p)^{X_i}(1-\theta_0\theta_i p)^{1-X_i}\prod_{i=m+1}^{2m}(\theta_0\theta_i q)^{X_i}(1-\theta_0\theta_i q)^{1-X_i}$$
$$< \prod_{i=1}^{m}(\theta_0\theta_i q)^{X_i}(1-\theta_0\theta_i q)^{1-X_i}\prod_{i=m+1}^{2m}(\theta_0\theta_i p)^{X_i}(1-\theta_0\theta_i p)^{1-X_i}.$$

Therefore,

$$\mathbb{P}_{H_0}\phi = \mathbb{P}\left(\sum_{i=1}^{m}\left(X_i \log\frac{q(1-\theta_0\theta_i p)}{p(1-\theta_0\theta_i q)} - \log\frac{1-\theta_0\theta_i p}{1-\theta_0\theta_i q}\right) + \sum_{i=m+1}^{2m}\left(X_i \log\frac{p(1-\theta_0\theta_i q)}{q(1-\theta_0\theta_i p)} - \log\frac{1-\theta_0\theta_i q}{1-\theta_0\theta_i p}\right) > 0\right).$$

To establish the desired bound for this quantity, we employ below a refined version of the Cramer–Chernoff argument [3, Proposition 14.23]. To this end, for any fixed $t > 0$, define independent random variables $\{W_i\}_{i=1}^{2m}$ by

$$\mathbb{P}\left(W_i = t\log\frac{q}{p}\right) = \theta_0\theta_i p, \quad \mathbb{P}\left(W_i = t\log\frac{1-\theta_0\theta_i q}{1-\theta_0\theta_i p}\right) = 1-\theta_0\theta_i p, \quad \text{for } i = 1, ..., m,$$

and

$$\mathbb{P}\left(W_i = t\log\frac{p}{q}\right) = \theta_0\theta_i q, \quad \mathbb{P}\left(W_i = t\log\frac{1-\theta_0\theta_i p}{1-\theta_0\theta_i q}\right) = 1-\theta_0\theta_i q, \quad \text{for } i = m+1, ..., 2m.$$

In addition, let

$$B_i = \begin{cases} (\theta_0\theta_i p)^{1-t}(\theta_0\theta_i q)^t + (1-\theta_0\theta_i p)^{1-t}(1-\theta_0\theta_i q)^t, & i = 1, \dots, m; \\ (\theta_0\theta_i q)^{1-t}(\theta_0\theta_i p)^t + (1-\theta_0\theta_i q)^{1-t}(1-\theta_0\theta_i p)^t, & i = m+1, \dots, 2m. \end{cases}$$

We lower bound $P_{H_0}\phi$ by

$$\begin{aligned} P_{H_0}\phi &= \mathbb{P}\left(\sum_{i=1}^{m} W_i + \sum_{i=m+1}^{2m} W_i > 0\right) \\ &\geq \sum_{0<\sum_i w_i<L}\prod_{i=1}^{2m}\mathbb{P}(W_i = w_i) \\ &\geq \left(\prod_{i=1}^{2m} B_i\right)e^{-L}\sum_{0<\sum_i w_i<L}\prod_{i=1}^{2m}\frac{P_i(w_i)e^{w_i}}{B_i} \\ &= \left(\prod_{i=1}^{2m} B_i\right)e^{-L}\sum_{0<\sum_i w_i<L}\prod_{i=1}^{2m} Q_i(w_i) \\ &= \left(\prod_{i=1}^{2m} B_i\right)e^{-L}\mathbb{Q}\left(0 < \sum_{i=1}^{2m} W_i < L\right), \end{aligned}$$

where

$$Q_i\left(W_i = t\log\frac{q}{p}\right) = \frac{(\theta_0\theta_i p)^{1-t}(\theta_0\theta_i q)^t}{B_i}, \quad Q_i\left(W_i = t\log\frac{1-\theta_0\theta_i q}{1-\theta_0\theta_i p}\right) = \frac{(1-\theta_0\theta_i p)^{1-t}(1-\theta_0\theta_i q)^t}{B_i}$$

for $i = 1, ..., m$ and

$$Q_i\left(W_i = t\log\frac{p}{q}\right) = \frac{(\theta_0\theta_i q)^{1-t}(\theta_0\theta_i p)^t}{B_i}, \quad Q_i\left(W_i = t\log\frac{1-\theta_0\theta_i p}{1-\theta_0\theta_i q}\right) = \frac{(1-\theta_0\theta_i q)^{1-t}(1-\theta_0\theta_i p)^t}{B_i}$$

for $i = m+1, ..., 2m$. We have also used the abbreviations $P_i(w_i) = \mathbb{P}(W_i = w_i)$ and $Q_i(w_i) = \mathbb{Q}(W_i = w_i)$.

To obtain the desired lower bound, we set t to be the minimizer of $\prod_{i=1}^{2m} B_i$. Since the minimizer is a stationary point, it satisfies

$$\sum_{i=1}^{2m} \mathbb{E}_{\mathbb{Q}} W_i = 0. \tag{72}$$

For any $t, a, b \in (0,1)$, recall the definition of $J_t(a,b)$ in (13). By Lemma 12, we have

$$J_t(1-a, 1-b) \leq CaJ_t(a,b), \tag{73}$$

where C only depends on the ratio a/b. Therefore, under the condition $a \asymp b = o(1)$, (73) implies

$$\log\left(1 - \frac{1}{2}J_t(a,b) - \frac{1}{2}J_t(1-a,1-b)\right) \geq -\frac{1}{2}(1+\eta)J_t(a,b), \tag{74}$$

for some $\eta = o(1)$ independent of t. Using (74), under the assumption that $1 < p/q = O(1)$, we have

$$\begin{aligned} \prod_{i=1}^{2m} B_i &\geq \exp\left(-\frac{1+\eta}{2}\sum_{i=1}^{m} J_t(\theta_0\theta_i q, \theta_0\theta_i p) - \frac{1+\eta}{2}\sum_{i=m+1}^{2m} J_t(\theta_0\theta_i p, \theta_0\theta_i q)\right) \\ &= \exp\left(-(1+\eta)\theta_0 m \left(p + q - p^{1-t}q^t - q^{1-t}p^t\right)\right). \end{aligned} \tag{75}$$

Hence,

$$\begin{aligned} \min_{0\leq t\leq 1}\prod_{i=1}^{2m} B_i &\geq \exp\left(-(1+\eta)\theta_0 m \max_{0\leq t\leq 1}\left(p + q - p^{1-t}q^t - q^{1-t}p^t\right)\right) \\ &= \exp\left(-(1+\eta)\theta_0 m(\sqrt{p} - \sqrt{q})^2\right). \end{aligned}$$

We now turn to lower bounding $e^{-L}\mathbb{Q}\left(0 < \sum_{i=1}^{2m} W_i < L\right)$ with t satisfying (72). To this end, we first calculate the variances of the W_i's. For $i = 1, ..., m$, there exists some constant $C > 0$ such that

$$\begin{aligned} \mathrm{Var}_{\mathbb{Q}}(W_i) &\leq \mathbb{E}_{\mathbb{Q}}(W_i^2) \\ &\leq \left(t\log\frac{p}{q}\right)^2 Q_i\left(t\log\frac{q}{p}\right) + \left(t\log\frac{1-\theta_0\theta_i q}{1-\theta_0\theta_i p}\right)^2 Q_i\left(t\log\frac{1-\theta_0\theta_i q}{1-\theta_0\theta_i p}\right) \\ &\leq C\theta_0\theta_i p\left(\log(\theta_0\theta_i p) - \log(\theta_0\theta_i q)\right)^2 + \left(\log(1-\theta_0\theta_i p) - \log(1-\theta_0\theta_i q)\right)^2 \\ &\leq C\theta_0\theta_i p\frac{(\theta_0\theta_i p - \theta_0\theta_i q)^2}{(\theta_0\theta_i q)^2} + C(\theta_0\theta_i p - \theta_0\theta_i q)^2 \\ &\leq C\frac{\theta_0\theta_i(p-q)^2}{p}. \end{aligned}$$

In addition, we have

$$\mathbb{E}_{\mathbb{Q}}(W_i^2) \geq \left(t\log\frac{p}{q}\right)^2 Q_i\left(t\log\frac{q}{p}\right) \gtrsim \frac{\theta_0\theta_i(p-q)^2}{p}, \quad \text{and} \quad (\mathbb{E}W_i)^2 = o\left(\frac{\theta_0\theta_i(p-q)^2}{p}\right).$$

Similar bounds hold for W_i, $i = m+1, ..., 2m$. Thus, we obtain that

$$\sum_{i=1}^{2m} \mathrm{Var}_{\mathbb{Q}}(W_i) \asymp \frac{\theta_0 m(p-q)^2}{p}.$$

Note that with $t \in [\gamma, 1-\gamma]$ and $p/q = O(1)$, the value of W_i is bounded by constant, for any $i \in [2m]$. Under the assumption that $\theta_0 m(\sqrt{p}-\sqrt{q})^2 \to \infty$, we have $\sum_{i=1}^{2m} \mathrm{Var}_{\mathbb{Q}}(W_i) \to \infty$, implying the indicator function $\mathbf{1}_{\left\{|W_i - \mathbb{E}W_i| > \epsilon\sqrt{\sum_{i=1}^{2m} \mathrm{Var}_{\mathbb{Q}}(W_i)}\right\}}$ goes to 0 for every i.

Thus

$$\lim_{m\to\infty} \sum_{i=1}^{2m} \mathbb{E}(W_i - \mathbb{E}W_i)^2 \mathbf{1}_{\left\{|W_i - \mathbb{E}W_i| > \epsilon\sqrt{\sum_{i=1}^{2m} \mathrm{Var}_{\mathbb{Q}}(W_i)}\right\}} = 0,$$

for any constant $\epsilon > 0$. Together with (72), the Lindeberg condition implies that under $\mathbb{Q}$, $\frac{\sum_{i=1}^{2m} W_i}{\sqrt{\sum_{i=1}^{2m} \mathrm{Var}_{\mathbb{Q}}(W_i)}}$ converges to $N(0,1)$. Taking $L = \sqrt{\sum_{i=1}^{2m} \mathrm{Var}_{\mathbb{Q}}(W_i)}$, we have that for any $\eta = o(1)$,

$$e^{-L}\mathbb{Q}\left(0 < \sum_{i=1}^{2m} W_i < L\right) \geq \exp\left(-\eta\theta_0 m(\sqrt{p}-\sqrt{q})^2\right)$$

for sufficiently large values of m. This completes the proof when $\theta_0 m(\sqrt{p}-\sqrt{q})^2 \to \infty$.

When $\theta_0 m(\sqrt{p}-\sqrt{q})^2 = O(1)$, then we have

$$\begin{aligned}
\inf_{\phi}(P_{H_0}\phi + P_{H_1}(1-\phi)) &= \int dP_{H_0} \wedge dP_{H_1} \\
&\geq \frac{1}{2}\left(\int \sqrt{dP_{H_0} dP_{H_1}}\right)^2 \\
&= \frac{1}{2}\left(\prod_{i=1}^{2m}\left(\theta_0\theta_i\sqrt{pq} + \sqrt{(1-\theta_0\theta_i p)(1-\theta_0\theta_i q)}\right)\right)^2 \\
&\geq \frac{1}{2}\exp\left(-(2+\eta)\theta_0 m(\sqrt{p}-\sqrt{q})^2\right) \\
&\geq c.
\end{aligned}$$

This completes the proof. □

PROOF OF (21). We bound $P_{H_0}\phi$ by

$$
\begin{aligned}
P_{H_0}\phi &\leq \left(\prod_{i=n/2+1}^{n}\mathbb{E}e^{tX_i}\right)\left(\prod_{i=1}^{n/2}\mathbb{E}e^{-tX_i}\right) \\
&= \exp\left(\sum_{i=n/2+1}^{n}\log\left(1-\theta_0\theta_i q+\theta_0\theta_i q e^{t}\right)+\sum_{i=1}^{n/2}\log\left(1-\theta_0\theta_i p+\theta_0\theta_i p e^{-t}\right)\right) \\
&\leq \exp\left(\sum_{i=n/2+1}^{n}\left(-\theta_0\theta_i q+\theta_0\theta_i q e^{t}\right)+\sum_{i=1}^{n/2}\left(-\theta_0\theta_i p+\theta_0\theta_i p e^{-t}\right)\right) \\
&= \exp\left(-\frac{\theta_0 n}{2}\left(p+q-pe^{-t}-qe^{t}\right)\right) \\
&= \exp\left(-\frac{\theta_0 n}{2}(\sqrt{p}-\sqrt{q})^2\right),
\end{aligned}
$$

where we have set $e^t=\sqrt{p/q}$. The same bound can be established for $P_{H_1}(1-\phi)$. □

PROOF OF LEMMA 5. The proof is very similar to that of Lemma 2. Therefore, we only sketch the difference. Without loss of generality, let $\theta_1\geq\theta_2\geq\ldots\geq\theta_m$, $\theta_{m+1}\geq\theta_{m+2}\geq\ldots\geq\theta_{m+m_1}$, and $m\leq m_1$. Then, we have

$$
\inf_{\phi}\left(P_{H_0}\phi+P_{H_1}(1-\phi)\right)\geq\inf_{\phi}\left(P_{\bar{H}_0}\phi+P_{\bar{H}_1}(1-\phi)\right),
$$

where $\bar{H}_0$ and $\bar{H}_1$ correspond to the following two hypotheses.

$$
\bar{H}_0: X\sim\bigotimes_{i=1}^{m}\mathrm{Bern}\left(\theta_0\theta_i p\right)\otimes\bigotimes_{i=m+1}^{2m}\mathrm{Bern}\left(\theta_0\theta_i q\right)
$$

$$
\text{vs.}\quad \bar{H}_1: X\sim\bigotimes_{i=1}^{m}\mathrm{Bern}\left(\theta_0\theta_i q\right)\otimes\bigotimes_{i=m+1}^{2m}\mathrm{Bern}\left(\theta_0\theta_i p\right).
$$

Bounding $\inf_\phi\left(P_{\bar{H}_0}\phi+P_{\bar{H}_1}(1-\phi)\right)$ is handled by the proof of Lemma 2 except that we do not have the relation (18) exactly. This slightly change the derivation of (75), as we will illustrate below. By the definition of $J_t(\cdot,\cdot)$, we have

$$
\begin{aligned}
&\frac{1}{2}\left(\sum_{i=1}^{n}J_t(\theta_0\theta_i q,\theta_0\theta_i p)+\sum_{i=m+1}^{2m}J_t(\theta_0\theta_i p,\theta_0\theta_i q)\right) \\
&= \left(\theta_0\sum_{i=1}^{m}\theta_i\right)\left(tq+(1-t)p-q^tp^{1-t}\right)+\left(\theta_0\sum_{i=m+1}^{2m}\theta_i\right)\left(tp+(1-t)p-p^tq^{1-t}\right) \\
&= \theta_0 m(p+q-p^{1-t}q^t-q^{1-t}p^t)+\theta_0 m\left|\frac{1}{m}\sum_{i=1}^{m}\theta_i-1\right|\left(tq+(1-t)p-q^tp^{1-t}\right) \\
&\quad +\theta_0 m\left|\frac{1}{m}\sum_{i=m+1}^{2m}\theta_i-1\right|\left(tp+(1-t)p-p^tq^{1-t}\right) \\
&\leq \theta_0 m(p+q-p^{1-t}q^t-q^{1-t}p^t)+C\eta\theta_0 m(\sqrt{p}-\sqrt{q})^2,
\end{aligned}
$$

for some $\eta = o(1)$. The last inequality uses Lemma 12 and the fact that $\delta = o(1)$. Since the term $C\eta\theta_0 m(\sqrt{p}-\sqrt{q})^2$ is of smaller order compared with the targeted exponent, the desired result can be derived following the remaining proof of Lemma 2. $\square$

PROOF OF LEMMA 6. For each $u \in [k]$, we define

$$\mathcal{C}_u = \left\{ i \in z^{-1}(u) \cap S_0^c : \|\tilde{V}_i - V_i\| < b \right\}.$$

Following [1], we divide the sets $\{\mathcal{C}_u\}_{u\in[k]}$ into three groups. Define

$$\begin{aligned} R_1 &= \{u \in [k] : \mathcal{C}_u = \varnothing\}, \\ R_2 &= \{u \in [k] : \mathcal{C}_u \neq \varnothing, \forall i, j \in \mathcal{C}_u, \tilde{z}(i) = \tilde{z}(j)\}, \\ R_3 &= \{u \in [k] : \mathcal{C}_u \neq \varnothing, \exists i, j \in \mathcal{C}_u, \text{s.t. } i \neq j, \tilde{z}(i) \neq \tilde{z}(j)\}. \end{aligned}$$

Then, it is easy to see that $\cup_{u\in[k]}\mathcal{C}_u = S_0^c \backslash S^c$ and $\mathcal{C}_u \cap \mathcal{C}_v = \varnothing$ for any $u \neq v$. Suppose there exists some $i \in \mathcal{C}_u$ and $j \in \mathcal{C}_v$ such that $u \neq v$ but $\tilde{z}(i) = \tilde{z}(j)$. Then, by the fact $\tilde{V}_i = \tilde{V}_j$, we have

$$\|V_i - V_j\| \le \|V_i - \tilde{V}_i\| + \|V_j - \tilde{V}_j\| < 2b,$$

contradicting (41). This means $\tilde{z}(i)$ and $\tilde{z}(j)$ take different values if i and j are not in the same $\mathcal{C}_u$'s. By the definition of R_2, the nodes in $\cup_{u\in R_2}\mathcal{C}_u$ have the same partition induced by z and $\tilde{z}$. Therefore,

$$\min_{\pi\in\Pi_k} \sum_{\{i:\hat{z}(i)\neq\pi(z(i))\}} \theta_i \le \sum_{i\in S_0}\theta_i + \sum_{i\in S}\theta_i + \sum_{i\in\cup_{u\in R_3}\mathcal{C}_u}\theta_i.$$

It is sufficient to bound $\sum_{i\in\cup_{u\in R_3}\mathcal{C}_u}\theta_i$. By the definition of R_3, we observe that each $\mathcal{C}_u$ for some $u \in R_3$ contains at least two different labels given by $\tilde{z}$. Thus we have $|R_2| + 2|R_3| \le k$. Moreover, since $k = |R_1| + |R_2| + |R_3|$, we have $|R_3| \le |R_1|$. This leads to

$$\begin{aligned} \sum_{i\in\cup_{u\in R_3}\mathcal{C}_u}\theta_i &\le |R_3|(1+\delta)\frac{\beta n}{k} \\ &\le |R_1|(1+\delta)\frac{\beta n}{k} \\ &\le \frac{1+\delta}{1-\delta}\beta^2 \sum_{i\in\cup_{u\in R_1}\{i\in[n]:z(i)=u\}}\theta_i \\ &\le \frac{1+\delta}{1-\delta}\beta^2\sum_{i\in S}\theta_i \\ &\le 2\beta^2\sum_{i\in S}\theta_i. \end{aligned}$$

This completes the proof. $\square$

PROOF OF LEMMA 7. Define the matrix $P' \in \mathbb{R}^{n\times n}$ by $P'_{ij} = \theta_i\theta_j B_{z(i)z(j)}$ for each $i, j \in [n]$. Then, P' has rank at most k and differs from P only by the diagonal entries. By the definition of $\hat{P}$, we have $\|\hat{P}-T_\tau(A)\|_{\rm F}^2 \leq \|P'-T_\tau(A)\|_{\rm F}^2$. After rearrangement, we have

$$\begin{aligned}
\|\hat{P}-P\|_{\rm F}^2 &\leq 2\left|\left\langle \hat{P}-P', T_\tau(A)-P\right\rangle\right| + \|P'-P\|_{\rm F}^2 \\
&\leq 2\|\hat{P}-P'\|_{\rm F} \sup_{\{K:\|K\|_{\rm F}=1,\text{rank}(K)\leq 2k\}} |\langle K, T_\tau(A)-P\rangle| + \|P'-P\|_{\rm F}^2 \\
&\leq \frac{1}{4}\|\hat{P}-P'\|_{\rm F}^2 + 4 \sup_{\{K:\|K\|_{\rm F}=1,\text{rank}(K)\leq 2k\}} |\langle K, T_\tau(A)-P\rangle|^2 + \|P'-P\|_{\rm F}^2 \\
&\leq \frac{1}{2}\|\hat{P}-P\|_{\rm F}^2 + \frac{3}{2}\|P'-P\|_{\rm F}^2 + 4 \sup_{\{K:\|K\|_{\rm F}=1,\text{rank}(K)\leq 2k\}} |\langle K, T_\tau(A)-P\rangle|^2 .
\end{aligned}$$

Therefore,

$$\|\hat{P}-P\|_{\rm F}^2 \leq 3\|P'-P\|_{\rm F}^2 + 8 \sup_{\{K:\|K\|_{\rm F}=1,\text{rank}(K)\leq 2k\}} |\langle K, T_\tau(A)-P\rangle|^2 . \tag{76}$$

Apply singular value decomposition to K and we get $K=\sum_{l=1}^{2k}\lambda_l u_l u_l^T$. Then,

$$|\langle K, T_\tau(A)-P\rangle| \leq \sum_{l=1}^{2k}|\lambda_l||u_l^T(T_\tau(A)-P)u_l| \leq \|T_\tau(A)-P\|_{\rm op}\sum_{l=1}^{2k}|\lambda_l| \leq \sqrt{2k}\|T_\tau(A)-P\|_{\rm op}.$$

By Lemma 5 of [2], $\|T_\tau(A)-P\|_{\rm op} \leq C\sqrt{n\alpha p\|\theta\|_\infty^2+1}$ with probability at least $1-n^{-C'}$, where the constant C' can be made arbitrarily large. Hence,

$$8 \sup_{\{K:\|K\|_{\rm F}=1,\text{rank}(K)\leq 2k\}} |\langle K, T_\tau(A)-P\rangle|^2 \leq C_1 k(n\alpha p\|\theta\|_\infty^2+1),$$

with probability at least $1-n^{-C'}$ Moreover,

$$3\|P'-P\|_{\rm F}^2 = 3\sum_{i=1}^n \theta_i^2 B_{z(i)z(i)}^2 \leq 3\alpha^2p^2\|\theta\|_\infty n(1+\delta) \leq C_2\alpha^2 p\|\theta\|_\infty^2 n.$$

Using (76), the proof is complete by absorbing α into the constant. □

REFERENCES

[1] Y. Chen, X. Li, and J. Xu. Convexified modularity maximization for degree-corrected stochastic block models. *arXiv preprint arXiv:1512.08425*, 2015.

[2] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving optimal misclassification proportion in stochastic block model. *arXiv preprint arXiv:1505.03772*, 2015.

[3] A. W. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 2000.