7. Proofs

7.1. Proof of Theorem 1

Observe that $Y^* \in \mathcal{S}_{\text{psd}} \subset \mathcal{S}_{\text{nuclear}}$, so it suffices to prove the theorem assuming $\mathcal{S} = \mathcal{S}_{\text{nuclear}}$ in (3).

We need some additional notation. Suppose the size of the i-th cluster is K_i, and the rank$-r$ SVD of Y^* is $U\Sigma U^\top$. Note that $UU^\top$ is a block diagonal matrix with r blocks such that the i-th block has size $K_i \times K_i$ with all entries equal to $\frac{1}{K_i} \le \frac{1}{K}$. We define the projections $\mathcal{P}_T$ and $\mathcal{P}_{T^\perp}$ by

$$\mathcal{P}_T Z = UU^\top Z + ZUU^\top - UU^\top ZUU^\top$$

and

$$\mathcal{P}_{T^\perp} Z = Z - \mathcal{P}_T Z.$$

Define the matrix $W_{ij} := ((2A_{ij}-1)B_{ij})_{i,j=1}^n \in \mathbb{R}^{n\times n}$ and the quantities $\theta := \mathbb{E}\left[(1-2P_{ij})\, B_{ij}\right]$ and $\rho := \mathbb{E}\left[B_{ij}^2\right] - \theta^2$. Note that

$$\mathbb{E}W_{ij} = \mathbb{E}\left[\mathbb{E}\left[(2A_{ij}-1)\, B_{ij} | P\right]\right] = \left(2Y_{ij}^* - 1\right)\mathbb{E}\left[(1-2P_{ij})\, B_{ij}\right] = \left(2Y_{ij}^*-1\right)\theta$$

and

$$\text{Var}\left[W_{ij}\right] = \mathbb{E}\left[W_{ij}^2\right] - \left(\mathbb{E}W_{ij}\right)^2 = \mathbb{E}\left[(2A_{ij}-1)^2 B_{ij}^2\right] - \left(2Y_{ij}^*-1\right)^2\theta^2 = \rho.$$

Our proof requires two standard concentration results for the random matrix W.

Lemma 1. *If $0 \le W_{ij} \le b_0$ almost surely for all i,j and the condition (6) holds, then with high probability, we have*

$$\|W - \mathbb{E}\left[W\right]\| \le c_2\left(b\log n + \sqrt{\rho n \log n}\right) \tag{12}$$

and

$$\left\|UU^\top\left(W - \mathbb{E}\left[W\right]\right)\right\|_\infty \le c_3 \frac{\sqrt{b^2\log^2 n + \rho K \log n}}{K} \tag{13}$$

for some universal constants c_2, c_3.

We prove the lemma in Section 7.1.1 to follow. We now prove Theorem 1 assuming the two inequalities (12) and (13) in the lemma hold.

For any matrix Y, we define $\Delta(Y) := \langle Y^* - Y, W\rangle$. To prove the theorem, it suffices to show that $\Delta(Y) > 0$ for all feasible Y of the program 2–4 with $Y \neq Y^*$. We rewrite $\Delta(Y)$ as

$$\Delta(Y) = \langle \mathbb{E}W, Y^* - Y\rangle + \langle W - \mathbb{E}W, Y^* - Y\rangle. \tag{14}$$

We bound the two terms above. For any feasible Y obeying the constraint 4, the first term in (14) can be written as

$$\begin{aligned}\langle \mathbb{E}W, Y^* - Y\rangle &= \sum_{i,j}\left(2Y_{ij}^*-1\right)\theta\cdot\left(Y_{ij}^* - Y_{ij}\right)\\ &= \theta\|Y^* - Y\|_1,\end{aligned} \tag{15}$$

where the last equality follows from $0 \le Y_{ij} \le 1, \forall i,j$.

On the other hand, if we let $\lambda := c_2\left(\log n + \sqrt{\rho n \log n}\right)$, then by (12) we have

$$\left\|\frac{1}{\lambda}\mathcal{P}_{T^\perp}\left(W - \mathbb{E}W\right)\right\| \le \left\|\frac{1}{\lambda}\left(W - \mathbb{E}W\right)\right\| \le 1.$$

This means $UU^\top + \frac{1}{\lambda}\mathcal{P}_{T^\perp}\left(W - \mathbb{E}W\right)$ is a subgradient of the function $f(X) = ||X||_*$ at $X = Y^*$. Therefore, for any feasible Y we have

$$0 \ge \|Y\|_* - \|Y^*\|_* \ge \langle UU^\top + \frac{1}{\lambda}\mathcal{P}_{T^\perp}\left(W - \mathbb{E}W\right), Y - Y^*\rangle,$$

which means

$$\langle W - \mathbb{E}W, Y^* - Y\rangle \ge \left\langle \mathcal{P}_T\left(W - \mathbb{E}W\right) - \lambda UU^\top, Y^* - Y\right\rangle \tag{16}$$

We substitute (15) and (16) into (14) to obtain that for all feasible Y,

$$\begin{aligned}\Delta(Y) &\geq \theta\|Y^* - Y\|_1 + \langle \mathcal{P}_T\left(W - \mathbb{E}W\right) - \lambda UU^\top, Y^* - Y\rangle \\ &\overset{(a)}{\geq} \left(\theta - \lambda\|UU^\top\|_\infty - \|\mathcal{P}_T\left(W - \mathbb{E}W\right)\|_\infty\right)\|Y^* - Y\|_1 \\ &\overset{(b)}{\geq} \left(\theta - \frac{\lambda}{K} - \|\mathcal{P}_T\left(W - \mathbb{E}W\right)\|_\infty\right)\|Y^* - Y\|_1,\end{aligned}$$

where (a) follows from the Holder's inequality and (b) follows from the structure of U. But by definition of $\mathcal{P}_T$, we have

$$\begin{aligned}\||\mathcal{P}_T(W - \mathbb{E}W)\|_\infty &\leq \left\|UU^\top\left(W - \mathbb{E}w\right)\right\|_\infty + \left\|\left(W - \mathbb{E}W\right)UU^\top\right\|_\infty + \left\|UU^\top\left(W - \mathbb{E}W\right)UU^\top\right\|_\infty \\ &\leq 3\left\|UU^\top\left(W - \mathbb{E}W\right)\right\|_\infty \leq 3c_3\frac{\sqrt{b^2\log^2 n + K\rho\log n}}{K},\end{aligned}$$

where the last inequality follows from (13). It follows that

$$\Delta(Y) \geq \left(\theta - \frac{c_2\left(b\log n + \sqrt{\rho n\log n}\right)}{K} - 3c_3\frac{\sqrt{b^2\log^2 n + K\rho\log n}}{K}\right)\|Y^* - Y\|_1.$$

If the condition 6 in the theorem holds, then the quantity inside the parenthesis is positive (note that $\rho \leq \mathbb{E}\left[B_{ij}^2\right]$). This means $\Delta(Y) > 0$ for all $Y \neq Y^*$, which proves the theorem.

7.1.1. Proof of Lemma 1

Let e_i be the i-th standard basis vector in $\mathbb{R}^n$. For the first inequality in the lemma, note that

$$W - \mathbb{E}W = \sum_{i,j}\left(W_{ij} - \mathbb{E}W_{ij}\right)e_i e_j^\top,$$

which is the sum of n^2 i.i.d. zero-mean matrix. We compute

$$\left\|\left(W_{ij} - \mathbb{E}W_{ij}\right)e_i e_j^\top\right\| = |W_{ij} - \mathbb{E}W_{ij}| \leq b$$

for all (i, j) and

$$\begin{aligned}\left\|\mathbb{E}\sum_{i,j}\left(W_{ij} - \mathbb{E}W_{ij}\right)^2 e_j e_i^\top e_i e_j^\top\right\| &= \left\|\mathbb{E}\sum_{i,j}\left(W_{ij} - \mathbb{E}W_{ij}\right)^2 e_i e_j^\top e_j e_i^\top\right\| \\ &= \rho\left\|\sum_{i,j} e_i e_i^\top\right\| = \rho n.\end{aligned}$$

Applying the matrix Bernstein inequality (Tropp, 2012) gives that w.h.p.

$$\|W - \mathbb{E}W\| \leq c_2\left(b\log n + \sqrt{\rho n\log n}\right)$$

for some constant c_2.

We prove the second inequality. Fix (i, j). Assume node i belongs to the cluster k. Then

$$\left(UU^\top(W - \mathbb{E}W)\right)_{ij} = \frac{1}{K_k}\sum_{i' \in C_k^*}(W - \mathbb{E}W)_{i'j},$$

which is the average of K_k independent zero-mean random variables taking values in $[-b, b]$ with variance bounded by ρ. Therefore, by standard Bernstein inequality, we know that for some constant c_3 that

$$\left|\left(UU^\top(W - \mathbb{E}W)\right)_{ij}\right| \leq \frac{1}{K_k}c_3\left(b\log n + \sqrt{\rho K_k\log n}\right) \leq c_3\frac{\sqrt{b^2\log^2 n + K\rho\log n}}{K}, \text{ w.h.p.}$$

where the last inequality follows from $K_k \geq K$. The lemma follows from a union bound over all (i, j).

7.2. Proof of Corollary 1

For the first part of the corollary, we only need to show that the condition (6) in Theorem 1 is satisfied. Take $b_0 := 10 \log \frac{1}{\epsilon}$. Note that under the assumption of the theorem, we have almost surely

$$B_{ij} = B_{ij}^{\text{MLE}} = \log \frac{1 - \bar{P}_{ij}}{\bar{P}_{ij}} \leq \min \left\{ \frac{1 - 2\bar{P}_{ij}}{\bar{P}_{ij}}, \log \frac{1}{\epsilon} \right\} \leq 10 \left(1 - 2\bar{P}_{ij}\right) \log \frac{1}{\epsilon}, \tag{17}$$

so $B_{ij} \leq b_0$. The condition (8) in the corollary statement implies that

$$\mathbb{E}\left[\left(\frac{1}{2} - P_{ij}\right) B_{ij}\right] \geq c_1 \cdot \frac{n \log n}{K^2} \cdot \frac{b_0}{10} \geq c_1 \frac{b_0}{10} \frac{\log n}{K}$$

since $\bar{P}_{ij} \geq P_{ij}$ and $K \leq n$. On the other hand, the second term in the RHS of (6) can be upper bounded as follows:

$$\begin{aligned}
c_0 \sqrt{\mathbb{E}\left[B_{ij}^2\right]} \frac{\sqrt{n \log n}}{K} &= c_0 \sqrt{\mathbb{E}\left[\log \frac{1 - \bar{P}_{ij}}{\bar{P}_{ij}} \cdot \log \frac{1 - \bar{P}_{ij}}{\bar{P}_{ij}}\right]} \frac{\sqrt{n \log n}}{K} \\
&\overset{(a)}{\leq} 10 c_0 \sqrt{\mathbb{E}\left[\left(1 - 2\bar{P}_{ij}\right) \log \frac{1 - \bar{P}_{ij}}{\bar{P}_{ij}}\right]} \frac{\sqrt{n \log(1/\epsilon) \log n}}{K}, \\
&\overset{(b)}{\leq} \frac{c_0}{2} \mathbb{E}\left[(1 - 2P_{ij}) \log \frac{1 - \bar{P}_{ij}}{\bar{P}_{ij}}\right] \\
&= \frac{c_0}{2} \mathbb{E}\left[(1 - 2P_{ij}) B_{ij}\right],
\end{aligned}$$

where the inequality (a) follows from the previous bound (17), and (b) follows from the condition (8) in the corollary statement and $\bar{P}_{ij} \geq P_{ij}$. Combining the last two display equations proves that (6) is satisfied.

For the second part of the corollary, we note that $\bar{P}_{ij} := \max\left\{\frac{1}{16}, P_{ij}\right\} \geq \epsilon := \frac{1}{16}$. The RHS of (8) is upper bounded by $\log 16 \cdot c_1 \frac{n}{K^2} \log n$. Because $\log \frac{1-x}{x} \geq \frac{1}{10}(1 - 2x) = \frac{1}{5}\left(\frac{1}{2} - x\right)$ for all $x \leq \frac{1}{2}$, we have $\left(\frac{1}{2} - \bar{P}_{ij}\right) \log \frac{1 - \bar{P}_{ij}}{\bar{P}_{ij}} \geq \frac{1}{5}\left(\frac{1}{2} - \bar{P}_{ij}\right)^2 \geq \frac{1}{10}\left(\frac{1}{2} - P_{ij}\right)^2$ almost surely. It follows that the LHS of (8) is lower bounded by $\frac{1}{10}\mathbb{E}\left[\left(\frac{1}{2} - P_{ij}\right)^2\right]$. Under the condition 9, we conclude that (8) is satisfied.

7.3. Proof of Theorem 2

We prove the lemma using Fano's inequality. By Stirling's formula we have

$$|\mathcal{Y}| = \binom{n}{n/2} \geq 2^{n/2}.$$

Now suppose Y^* is sampled uniformly at random from $\mathcal{Y}$, and then P and A are generated according to our model. We have

$$\begin{aligned}
I\left(A, P; Y^*\right) &= H\left(A, P\right) - H\left(A, P | Y^*\right) \\
&\leq \binom{n}{2} \left[H\left(A_{11}, P_{11}\right) - H\left(A_{11}, P_{11} | Y_{11}^*\right)\right] \\
&\leq n^2 I\left(A_{11}, P_{11}; Y_{11}^*\right),
\end{aligned}$$

where the first inequality follows from symmetry. Let $a = A_{11}, p = P_{11}$ and $y = Y_{11}^*$. We now compute

$$\begin{aligned}
I\left(a, p; y\right) =& \mathbb{E}_{a,p,y}\left[\log \frac{\mathbb{P}\left(a, p | y\right)}{\mathbb{P}\left(a, p\right)}\right] = \mathbb{E}_{a,p,y}\left[\log \frac{\mathbb{P}\left(a | y, p\right) \mathbb{Q}(p)}{\mathbb{P}\left(a | p\right) \mathbb{Q}(p)}\right] \\
=& \mathbb{E}_{a,p,y}\left[\log \mathbb{P}\left(a | y, p\right)\right] - \mathbb{E}_{a,p,y}\left[\log \mathbb{P}\left(a | p\right)\right] \\
=& \mathbb{E}_{a,p,y}\left[\log \mathbb{P}\left(a | y, p\right)\right] - \mathbb{E}_{p}\left[p \log \frac{1}{2}\right] - \mathbb{E}_{p}\left[(1 - p) \log \frac{1}{2}\right].
\end{aligned}$$

One verifies that

$$\mathbb{E}_{a,p,y}\left[\log \mathbb{P}\left(a|y,p\right)\right] = \mathbb{E}_{a,p,y}\left[\mathbb{E}\left[\log \mathbb{P}\left(a|y,p\right)|a,y\right]\right] = \mathbb{E}_{p}\left[p\log p\right] + \mathbb{E}_{p}\left[(1-p)\log(1-p)\right].$$

Combining the last three display equations gives

$$\begin{aligned} I\left(A,P;Y^{*}\right) &\leq n^{2}\mathbb{E}_{p}\left[p\log 2p + (1-p)\log 2(1-p)\right] \\ &\leq n^{2}\mathbb{E}_{p}\left[p\left(2p-1\right) + \left(1-p\right)\left(1-2p\right)\right] \\ &= n^{2}\mathbb{E}_{p}\left[(1-2p)^{2}\right]. \end{aligned}$$

It follows that under the condition (10) of the theorem, we have

$$I\left(A,P;Y^{*}\right) \leq n^{2}\cdot\frac{c'}{n} = c'n \leq \frac{1}{4}\log|\mathcal{Y}|$$

provided c' is sufficiently small. Applying Fano's inequality, we obtain that for any $\hat{Y}$,

$$\mathbb{P}\left[\hat{Y}(A,P) \neq Y^{*}\right] \geq 1 - \frac{I\left(A,P;Y^{*}\right) + \log 2}{\log|\mathcal{Y}|} \geq \frac{1}{2},$$

where the probability is w.r.t. the randomness in Y^{*}, P and A. Because the supremum is lower bounded by the average, we conclude that

$$\sup_{Y^{*}\in\mathcal{Y}}\mathbb{P}\left[\hat{Y}(A,P) \neq Y^{*}\right] \geq \frac{1}{2}.$$

Taking the infimum over all $\hat{Y}$ proves the theorem.