PERFECT SPECTRAL CLUSTERING WITH DISCRETE COVARIATES

Jonathan Hehir, Xiaoyue Niu, and Aleksandra Slavković

Penn State University

Supplementary Material

These materials contain proofs and derivations for the theorems provided in the main paper.

S1 Preliminaries

We begin by defining the matrix absolute value and discussing some of its properties.

Definition S1. For a matrix $A \in \mathbb{R}^{m\times n}$, we define the matrix absolute value $|A| = \sqrt{A^T A}$. In particular, when $D = \mathrm{diag}(d_1, \ldots, d_n)$, we have $|D| = \mathrm{diag}(|d_1|, \ldots, |d_n|)$. For symmetric matrices $A = A^T$ with eigendecomposition $A = U\Lambda U^T$, we have $|A| = U|\Lambda|U^T$.

Fact S2. *$|A|$ is the unique positive semi-definite square root of $A^T A$.*

Proof. See Horn and Johnson (2012, Theorem 7.3.1). □

Fact S3. *If $A = A^T$ and $A = U\Sigma V^T$ is a singular value decomposition of A, then $|A| = U\Sigma U^T$.*

Proof. We may write $A^T A = AA^T = U\Sigma V^T V\Sigma U^T = U\Sigma^2 U^T$. Note that

$$U\Sigma U^T \succeq 0 \quad \text{and} \quad (U\Sigma U^T)(U\Sigma U^T) = A^2 = A^T A.$$

So by Fact S2, $|A| = U\Sigma U^T$ is the unique positive semi-definite square root of $A^T A$. □

Fact S4. *Suppose $A = XDX^T$, where $X^T X$ is diagonal and D is a diagonal matrix with diagonal entries in $\{\pm 1\}$. Then $|A| = XX^T$.*

Proof. Write $A^T A$ as follows:

$$\begin{aligned} A^T A &= XDX^T XDX^T \\ &= XD^2(X^T X)X^T \quad \text{(diagonals commute)} \\ &= XX^T XX^T \quad (D^2 = I) \\ &= (XX^T)^2. \end{aligned}$$

Since $XX^T \succeq 0$, $|A| = XX^T$ is the unique positive semi-definite square root of $A^T A$. □

Fact S5. *If U is orthogonal, then $|UAU^T| = U|A|U^T$.*

Proof.

$$\begin{aligned} (U|A|U^T)^2 &= U|A||A|U^T \\ &= UA^T AU^T \quad (|A|^2 = A^T A) \\ &= UA^T U^T UAU^T \\ &= (UAU^T)^T (UAU^T). \end{aligned}$$

Since $U|A|U^T \succeq 0$, $U|A|U^T$ is the unique positive semi-definite square root of $(UAU^T)^T(UAU^T)$. □

Fact S6. *Suppose $A = c\mathbf{1}_n\mathbf{1}_n^T + dI_n$. Then $|A| = c'\mathbf{1}_n\mathbf{1}_n^T + d'I_n$, where:*

$$c' = \frac{|cn + d| - |d|}{n}, \quad d' = |d|.$$

Proof. Let $U\Lambda U^T$ be an eigendecomposition of $\mathbf{1}_n\mathbf{1}_n^T$. Then $\Lambda = \text{diag}(n, 0, \dots, 0)$. Now we write an eigendecomposition for A:

$$\begin{aligned} A &= c\mathbf{1}_n\mathbf{1}_n^T + dI_n \\ &= cU\Lambda U^T + dUU^T \\ &= U(c\Lambda + dI_n)U^T. \end{aligned} \tag{S1.1}$$

By definition, then:

$$|A| = U|c\Lambda + dI_n|U^T,$$

which is of the same form as eq. (S1.1), albeit with different constants. The result follows by solving the following for c' and d':

$$\text{diag}(|cn + d|, |d|, \dots, |d|) = |c\Lambda + dI_n| = c'\Lambda + d'I_n = \text{diag}(c'n + d', d', \dots, d').$$

□

Fact S7. *Suppose* $A = c\mathbf{1}_n\mathbf{1}_n^T + dI_n$*, and* $A_{ij} > 0$ *for all* $i, j \in [n]$*. Then* $|A|_{ij} > 0$ *for all* $i, j \in [n]$*.*

Proof. We begin with the trivial cases: If $d \geq 0$, then $A \succeq 0$ and $A = |A|$. Also if $n = 1$, then A is scalar, and $|A|$ is the usual scalar absolute value.

Assume then that $d < 0$ and $n \geq 2$. Let $|A| = c'\mathbf{1}_n\mathbf{1}_n^T + d'I_n$ as defined in Fact S6. Since all entries in A are positive, then $c > -d = |d|$. Consequently:

$$cn + d = cn - |d| > |d|n - |d| = |d|(n-1) \geq |d|$$

As a result, c' must be positive, since $|cn + d| = cn + d > |d|$. Since d' is also positive, every entry in $|A|$ is positive. □

Fact S8. *For any two square matrices of equal dimension,* $\|\, |A| - |B| \,\|_F \leq \sqrt{2}\|A - B\|_F$*.*

Proof. See Bhatia (2013), Theorem VII.5.7 and eq. (VII.39). □

We recall our definition of the binary matrix operator $\boxplus$.

Definition S9. Let $A \in \mathbb{R}^{m\times m}, B \in \mathbb{R}^{n\times n}$. Then:

$$A \boxplus B = (A \otimes \mathbf{1}_n\mathbf{1}_n^T) + (\mathbf{1}_m\mathbf{1}_m^T \otimes B).$$

The operation $\boxplus$ is similar to the more standard Kronecker sum $A \oplus B = (A \otimes I_n) + (I_m \otimes B)$, but with identity matrices replaced by $\mathbf{1}\mathbf{1}^T$. Fact S10 below also resembles a property that the Kronecker sum satisfies, but replacing the matrix exponential with an element-wise exponential.

Fact S10. *For two square matrices* A *and* B*,* $\exp(A \boxplus B) = \exp(A) \otimes \exp(B)$*, where* $\exp$ *is evaluated element-wise.*

Proof. Observe that the Kronecker product of two square matrices $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{n\times n}$ may be written $A \otimes B = (A \otimes \mathbf{1}_n\mathbf{1}_n^T) \odot (\mathbf{1}_m\mathbf{1}_m^T \otimes B)$, where $\odot$ denotes the Hadamard product (i.e., element-wise multiplication). From here it follows that:

$$\begin{aligned}
\exp(A \boxplus B) &= \exp(A \otimes \mathbf{1}_n\mathbf{1}_n^T + \mathbf{1}_m\mathbf{1}_m^T \otimes B) \\
&= \exp(A \otimes \mathbf{1}_n\mathbf{1}_n^T) \odot \exp(\mathbf{1}_m\mathbf{1}_m^T \otimes B) \\
&= \left(\exp(A) \otimes \mathbf{1}_n\mathbf{1}_n^T\right) \odot \left(\mathbf{1}_m\mathbf{1}_m^T \otimes \exp(B)\right) \\
&= \exp(A) \otimes \exp(B).
\end{aligned}$$

□

In light of the Kronecker representation of $\exp(A \boxplus B)$, we review some facts about Kronecker products and inspect their matrix absolute values.

Fact S11. *If $A = A^T$ and $B = B^T$, then $A \otimes B = (A \otimes B)^T$.*

Proof. By Horn and Johnson (1991, eq. 4.2.5), $(A \otimes B)^T = A^T \otimes B^T = A \otimes B$. □

Fact S12. *Let $A = A^T, B = B^T$ with eigendecompositions $A = U\Lambda U^T, B = V\Psi V^T$. If $C = A \otimes B$, then:*

$$|C| = (U \otimes V)|\Lambda \otimes \Psi|(U \otimes V)^T = |A| \otimes |B|.$$

Proof. We begin by writing SVDs for A and B, namely:

$$\begin{aligned} A &= U|\Lambda|(\mathrm{sign}(\Lambda)U^T) \\ B &= V|\Psi|(\mathrm{sign}(\Psi)V^T), \end{aligned}$$

where $\mathrm{sign}(\cdot)$ is taken element-wise. It is easy to verify that $\mathrm{sign}(\Lambda)U^T$ and $\mathrm{sign}(\Psi)V^T$ are indeed orthogonal.

Armed with these decompositions, we may apply Horn and Johnson (1991, Theorem 4.2.15) to find an SVD for C:

$$\begin{aligned} C &= (U \otimes V)(|\Lambda| \otimes |\Psi|)(\mathrm{sign}(\Lambda)U^T \otimes \mathrm{sign}(\Psi)V^T) \\ &= (U \otimes V)|\Lambda \otimes \Psi|(\mathrm{sign}(\Lambda)U^T \otimes \mathrm{sign}(\Psi)V^T) \end{aligned}$$

Since $A = A^T$ and $B = B^T$, we have that $C = C^T$ (Fact S11). Therefore:

$$\begin{aligned} |C| &= (U \otimes V)|\Lambda \otimes \Psi|(U \otimes V)^T \quad \text{(Fact S3)} \\ &= (U \otimes V)(|\Lambda| \otimes |\Psi|)(U \otimes V)^T \\ &= (U|\Lambda| \otimes V|\Psi|) \otimes (U^T \otimes V^T) \\ &= (U|\Lambda|U^T) \otimes (V|\Psi|V^T) \\ &= |A| \otimes |B|. \end{aligned}$$

□

Finally, we give two useful facts about sums and permutations.

Fact S13. *Let $x_1, \ldots, x_n \in \mathbb{R}$. Then for any $\sigma \in S_{[n]}$:*

$$\sum_{i=1}^{n} x_i x_{\sigma(i)} \leq \sum_{i=1}^{n} x_i^2.$$

Proof. This is an application of Cauchy–Schwarz in disguise:

$$\begin{aligned}\left(\sum_{i=1}^{n} x_i x_{\sigma(i)}\right)^2 &\le \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} x_{\sigma(i)}^2\right) \\ &= \left(\sum_{i=1}^{n} x_i^2\right)^2.\end{aligned}$$

The final statement comes by taking the square root of both sides. □

Fact S14. *Let $A \in \mathbb{R}^{n\times n}$ such that $A \succeq 0$. Then for any $\sigma \in S_{[n]}$:*

$$\sum_{i=1}^{n} A_{i\sigma(i)} \le \sum_{i=1}^{n} A_{ii}.$$

Moreover, if $\operatorname{rank}(A) = n$ *and* $\sigma \neq \mathrm{id}$*, the inequality is strict.*

Proof. Since $A \succeq 0$, let $A = XX^T$. Fix $\sigma \in S_{[n]}$. Then:

$$\begin{aligned}\sum_{i=1}^{n} A_{i\sigma(i)} &= \sum_{i=1}^{n} e_i^T A e_{\sigma(i)} \\ &= \sum_{i=1}^{n} \langle X^T e_i, X^T e_{\sigma(i)}\rangle \\ \text{ⓐ} &\le \sum_{i=1}^{n} \|X^T e_i\| \|X^T e_{\sigma(i)}\| \quad \text{(Cauchy–Schwarz)} \\ &\le \sum_{i=1}^{n} \|X^T e_i\|^2 \quad \text{(Fact S13)} \\ &= \sum_{i=1}^{n} \langle X^T e_i, X^T e_i\rangle \\ &= \sum_{i=1}^{n} e_i^T A e_i = \sum_{i=1}^{n} A_{ii}.\end{aligned}$$

If $\sigma \neq \mathrm{id}$, the inequality ⓐ is made strict when X has linearly independent rows, i.e., when A is full-rank. □

S2 Proofs of Results

Representation Results. We prove that ACSBM can be represented as an SBM by explicitly constructing such a representation.

Proof of Proposition 1. Consider first the case when $M = 1$, i.e., $Z = Z_{*1}$. Every edge is an independent Bernoulli random variable whose probability depends on (θ_i, Z_{i1}) and (θ_j, Z_{j1}). It will be convenient to map these tuples to scalars. Let $\tau(k, \ell) = L_1(k-1)+\ell$, a bijection from $[K] \times [L_1]$ to $[KL_1]$. Let $\tilde{\theta}^{(1)} \in [KL_1]^n = (\tau(\theta_i, Z_{1i}))_{i=1}^n$. We will now write the edge probabilities in terms of these new scalar quantities. It can be shown (if a bit tediously) that:

$$\begin{aligned}\mathbf{P}(Y_{ij} = 1 \mid \tilde{\theta}_i^{(1)} = t_1, \tilde{\theta}_j^{(1)} = t_2) &= g^{-1}\left([B \otimes \mathbf{1}_{L_1}\mathbf{1}_{L_1}^T + \mathbf{1}_K\mathbf{1}_K^T \otimes \beta_1 I_{L_1}]_{t_1t_2} \right) \\ &= \left[g^{-1}(B \boxplus \beta_1 I_{L_1}) \right]_{t_1t_2},\end{aligned}$$

where g^{-1} is taken element-wise in the final line. This is precisely the form of the SBM given in Definition 1. Thus when $M = 1$, we can say Y is equal to an SBM with $\tilde{L} = KL_1$ communities, $\tilde{\theta} = L_1(\theta - \mathbf{1}_n) + Z_{*1}$, and edge probabilities $\tilde{B} = g^{-1}(B \boxplus \beta_1 I_{L_1})$.

The case when $M \geq 2$ follows inductively. Let $Y_1 \sim \text{ACSBM}(\theta, B, Z_1, \beta_1, g) \overset{D}{=} \text{SBM}(\tilde{\theta}^{(1)}, \tilde{B}^{(1)})$. Define $Y_2 = \text{ACSBM}(\theta, B, [Z_1 \mid Z_2], (\beta_1, \beta_2)^T, g)$. This network is equal in distribution to $Y_2' \sim \text{ACSBM}(\tilde{\theta}^{(1)}, g(\tilde{B}^{(1)}), Z_2, \beta_2, g)$. By the $M = 1$ case above, these networks are equal in distribution to an SBM with KL_1L_2 communities:

$$\tilde{\theta}^{(2)} = L_2(\tilde{\theta}^{(1)} - \mathbf{1}_n) + Z_{*2} = L_2(L_1(\theta - \mathbf{1}_n) + Z_{*1} - \mathbf{1}_n) + Z_{*2}$$

and edge probabilities:

$$g^{-1}\left(g(\tilde{B}^{(1)}) \boxplus \beta_2 I_{L_2} \right) = g^{-1}(B \boxplus \beta_1 I_{L_1} \boxplus \beta_2 I_{L_2}),$$

where once again, g and g^{-1} are element-wise.

Proceed inductively to find the forms of $Y_3, \ldots, Y_M$, defined analogously to Y_2, so that $Y \overset{D}{=} Y_M$. □

The gRDPG representation now follows immediately as a corollary.

Proof of Proposition 2. By Proposition 1, we may represent Y as an SBM, i.e., $Y \overset{D}{=} \text{SBM}(\tilde{\theta}, \tilde{B})$. The ability to represent an SBM as a gRDPG using latent positions derived from spectral decomposition is a well established practice in the gRDPG literature, e.g., Rubin-Delanchy et al. (2017, Section 2.1). Thus Proposition 2 follows as a corollary to Proposition 1. □

Consistency of Part 1. Consistency of Part 1 of the algorithm was stated in Theorem 1, proven here.

Proof of Theorem 1. By Lemma 1, we know that:

$$\max_{i\in[n]} \|Q\hat{X}_i - X_{\tilde{B}}(\theta_i, Z_i)\|_2 = O_P\left(\frac{\log^c n}{\sqrt{n}}\right)$$

for some sequence of matrices $Q \in \mathbb{O}(p,q)$. We might prefer a statement in terms of $\hat{X}_i$, rather than $Q\hat{X}_i$, which we can make as follows:

$$\max_{i\in[n]} \|\hat{X}_i - QX_{\tilde{B}}(\theta_i, Z_i)\|_2 \leq \|Q^{-1}\|_2 \left(\max_{i\in[n]} \|Q\hat{X}_i - X_{\tilde{B}}(\theta_i, Z_i)\|_2\right).$$

We have seemingly done little here but move the troublesome Q and impose an additional nuisance term. However, Rubin-Delanchy et al. (2017, Lemma 5) states a key result: $\|Q\|_2$ and $\|Q^{-1}\|_2$ are bounded almost surely. This allows us to eliminate the nuisance term:

$$\max_{i\in[n]} \|\hat{X}_i - QX_{\tilde{B}}(\theta_i, Z_i)\|_2 = O_P\left(\frac{\log^c n}{\sqrt{n}}\right).$$

We still have to grapple with $QX_{\tilde{B}}$. Observe that for z fixed, the canonical latent positions $X_{\tilde{B}}(1,z),\ldots,X_{\tilde{B}}(K,z)$ are distinct by construction. Since Q is full-rank, this also applies to $QX_{\tilde{B}}(1,z),\ldots,QX_{\tilde{B}}(K,z)$. Moreover, in light of the bounded spectral norms of Q and Q^{-1}, which bound the singular values of Q in an interval away from zero, the asymptotic distortion of distances is limited. In particular, $\|Q(X_{\tilde{B}}(k_1,z) - X_{\tilde{B}}(k_2,z))\|_2 = \Theta(\sqrt{\alpha_n})$ almost surely. Combining these facts yields the result, as follows.

Let $\mathcal{B}(x,r)$ denote a ball centered at x with radius r. From our argument above, there exists a sequence of radii $r = O_P(\log^c n/\sqrt{n})$ such that $\hat{X}_i \in \mathcal{B}(QX_{\tilde{B}}(\theta_i,z),r)$ for all $i \in \mathcal{I}_z$. Since $\|Q(X_{\tilde{B}}(k_1,z) - X_{\tilde{B}}(k_2,z))\|_2$ scales with $\sqrt{\alpha_n} = \omega(\log^{2c} n/\sqrt{n})$, these balls shrink in size faster than they converge to the origin. More concretely, let $\mathcal{B}_{k,z} = \mathcal{B}(QX_{\tilde{B}}(k,z),r)$ for $k \in [K]$. Then for any $k_1,k_2 \in [K]$:

$$\mathbf{P}(\mathcal{B}_{k_1,z} \cap \mathcal{B}_{k_2,z} = \emptyset) = \mathbf{P}\left(r < \frac{1}{2}\|QX_{\tilde{B}}(k_1,z) - QX_{\tilde{B}}(k_2,z)\|_2\right) \to 1,$$

since $\|QX_{\tilde{B}}(k_1,z) - QX_{\tilde{B}}(k_2,z)\|_2 = \Theta(\sqrt{\alpha_n})$ almost surely, and $r = o_P(\sqrt{\alpha_n})$. □

Consistency of Part 2. Consistency of Part 2 of the algorithm was stated in Theorem 2.

Proof of Theorem 2. Suppose $Y_{gen} \sim \mathrm{SBM}(\tilde{\theta}, B_{gen})$ for some symmetric matrix $B_{gen} \in \mathbb{R}^{K\tilde{L}\times K\tilde{L}}$. This model is more general than $Y \sim \mathrm{SBM}(\tilde{\theta}, \tilde{B})$. Suppose we have a perfect estimate of $\tilde{\theta}$ (up to a permutation), and we wish to estimate B_{gen}. In this case, the natural approach to estimating B_{gen} via the empirical density of each block is precisely the maximum likelihood estimator, which has been well-studied (e.g., Bickel et al., 2013).

Under the theorem hypothesis, we have indeed recovered $\tilde{\theta}$ up to a permutation of labels. This is true since $\tilde{\theta}((\tau_{z_i} \circ \hat{\theta}_{z_i})(i)), z_i) = \tilde{\theta}_i$ for all i, and the function $\tilde{\theta}(\cdot, \cdot)$ is a bijection. Let $\tau \in S_{[K\tilde{L}]}$ denote this permutation, and let T denote the corresponding permutation matrix. Then $T^{-1}\hat{\tilde{B}}T$ is the maximum likelihood estimator for a model $Y_{gen} \sim \text{SBM}(\tilde{\theta}, B_{gen})$, and so we may apply the maximum likelihood results of Bickel et al. (2013, Lemma 1) or, more conveniently, Tang et al. (2022, Theorem 1). Per these results, we can say that for any $k_1, k_2 \in [K\tilde{L}]$:

$$n\alpha_n^{-1/2}\left((T^{-1}\hat{\tilde{B}}T)_{k_1k_2} - \tilde{B}_{k_1k_2}\right) \xrightarrow{D} \mathcal{N}(0, v_{k_1k_2}),$$

where $\xrightarrow{D} \mathcal{N}(\cdot, \cdot)$ denotes convergence in distribution to the normal distribution, and $v_{k_1k_2} > 0$ is a constant depending on k_1 and k_2. In other words:

$$(T^{-1}\hat{\tilde{B}}T)_{k_1k_2} - \tilde{B}_{k_1k_2} = O_P\left(\frac{\sqrt{\alpha_n}}{n}\right).$$

Since $\tilde{B}$ scales with α_n, we rewrite this to be in terms of the constant quantity $\alpha_n^{-1}\tilde{B}$:

$$\alpha_n^{-1}\left((T^{-1}\hat{\tilde{B}}T)_{k_1k_2} - \tilde{B}_{k_1k_2}\right) = O_P\left(\frac{1}{n\sqrt{\alpha_n}}\right) = o_P\left(\frac{1}{\sqrt{n\log^c n}}\right).$$

Since K and $\tilde{L}$ are kept constant in n, these entrywise bounds may be taken as a bound for the Frobenius norm, $\|T^{-1}\hat{\tilde{B}}T - \tilde{B}\|_F$. Moreover, since the Frobenius norm is unitarily invariant, we may write:

$$\|\hat{\tilde{B}} - T\tilde{B}T^{-1}\|_F = o_P\left(\frac{1}{\sqrt{n\log^c n}}\right).$$

□

Consistency of Part 3. We first show that the matching problem selects the appropriate permutations in the absence of estimation error, i.e., when applied to the true latent positions $X_{\tilde{B}}$. Note that the role of the permutation σ in Theorem S15 below differs slightly from its role in Algorithm 1. In the algorithm, there is an unknown permutation that we are looking to reverse for each choice of z; in the theorem below, there is no such permutation, so the correct choice of σ is the identity permutation.

Theorem S15. *Assume Y from the setting of Section 4. Let $X_{\tilde{B}}$ as in Proposition 1. For any fixed $z \in [L_1] \times \cdots \times [L_M]$:*

$$\arg\min_{\sigma \in S_{[K]}} \sum_{k=1}^{K} \|X_{\tilde{B}}(\sigma(k), z) - X_{\tilde{B}}(k, \mathbf{1}_M)\|_2^2 = \mathrm{id}. \tag{S2.2}$$

Moreover, if $\exp(B)$ *is full-rank,* $\sigma = \mathrm{id}$ *is the unique minimizer.*

Proof. To simplify notation for the proof, let $x_{kz} = X_{\tilde{B}}(k, z)$. We begin by unpacking the squared norm:

$$\begin{aligned}
\sum_{k=1}^{K} \|x_{\sigma(k)z} - x_{k\mathbf{1}}\|_2^2 &= \sum_{k=1}^{K} \langle x_{\sigma(k)z} - x_{k\mathbf{1}}, x_{\sigma(k)z} - x_{k\mathbf{1}} \rangle \\
&= \sum_{k=1}^{K} \big(\langle x_{\sigma(k)z}, x_{\sigma(k)z} \rangle + \langle x_{k\mathbf{1}}, x_{k\mathbf{1}} \rangle - 2 \langle x_{\sigma(k)z}, x_{k\mathbf{1}} \rangle \big) \\
&= \sum_{k=1}^{K} \langle x_{kz}, x_{kz} \rangle + \sum_{k=1}^{K} \langle x_{k\mathbf{1}}, x_{k\mathbf{1}} \rangle - 2 \sum_{k=1}^{K} \langle x_{\sigma(k)z}, x_{k\mathbf{1}} \rangle
\end{aligned}$$

Since only the final sum depends on σ, the optimization problem (S2.2) is equivalent to finding:

$$\arg\max_{\sigma \in S_{[K]}} \sum_{k=1}^{K} \langle x_{\sigma(k)z}, x_{k\mathbf{1}} \rangle.$$

Fix $z \in [L_1] \times \cdots \times [L_M]$, and let $\tilde{B}$ as in Proposition 1. Next, we will assemble yet another matrix. For any $k_1, k_2 \in [K]$, let $Q_{k_1 k_2} = \langle x_{k_1 z}, x_{k_2 \mathbf{1}} \rangle$. If we can show that $Q \succ 0$, the result will follow from Fact S14. This is our plan. Observe that:

$$\langle x_{k_1 z}, x_{k_2 \mathbf{1}} \rangle_{pq} = \tilde{B}_{\tilde{\theta}(k_1, z), \tilde{\theta}(k_2, \mathbf{1})},$$

where (p, q) is the signature of the gRDPG corresponding to Y. Following from Fact S4, the inner products that form the entries of Q can be found in $|\tilde{B}|$, i.e.:

$$Q_{k_1 k_2} = \langle x_{k_1 z}, x_{k_2 \mathbf{1}} \rangle = |\tilde{B}|_{\tilde{\theta}(k_1, z), \tilde{\theta}(k_2, \mathbf{1})}.$$

Since $g = \log$, by Fact S10, we can write $\tilde{B}$ like so:

$$\tilde{B} = \exp(B) \otimes \exp(\beta_1 I_{L_1}) \otimes \cdots \otimes \exp(\beta_M I_{L_M}).$$

Lemma S12 gives the convenient form of $|\tilde{B}|$:

$$|\tilde{B}| = |\exp(B)| \otimes |\exp(\beta_1 I_{L_1})| \otimes \cdots \otimes |\exp(\beta_M I_{L_M})|.$$

In particular, this means:

$$\begin{aligned} Q_{k_1 k_2} &= |\tilde{B}|_{\tilde{\theta}(k_1,z),\tilde{\theta}(k_2,\mathbf{1})} \\ &= |\exp(B)|_{k_1 k_2} \left[\, |\exp(\beta_1 I_{L_1})| \otimes \cdots \otimes |\exp(\beta_M I_{L_M})|\, \right]_{\tilde{\theta}(1,z),1} \\ &= c_z\, |\exp(B)|_{k_1 k_2}, \end{aligned}$$

where $c_z = \left[\, |\exp(\beta_1 I_{L_1})| \otimes \cdots \otimes |\exp(\beta_M I_{L_M})|\, \right]_{\tilde{\theta}(1,z),1}$ is a strictly positive constant. This follows from Fact S7, which says that each of the $|\exp(\beta_m I_{L_m})|$ matrices have positive entries. Since $|\exp(B)| \succeq 0$ by construction, we have then that $Q \succeq 0$. Moreover, when $\exp(B)$ is full-rank, $Q \succ 0$.

Applying Fact S14, we have that $\sigma = \mathrm{id}$ is a solution to our optimization problem; moreover, it is the unique solution when $\exp(B)$ is full-rank. □

The following corollary generalizes Theorem S15 to arbitrary link functions and differential homophily, as discussed in Section 4, "Generalizations."

Corollary S16. *Suppose M, L, θ, and Z are as defined for the ACSBM, but let $\beta_{m\ell} \in \mathbb{R}$ for $m \in [M], \ell \in [L_m]$ denote differential homophily coefficients for the model*

$$Y_{ij} \overset{ind}{\sim} \text{Bernoulli}\left(\alpha_n g^{-1}\left(B_{\theta_i\theta_j} + \sum_{m=1}^{M}\sum_{\ell=1}^{L_m} \beta_{m\ell}\mathbb{I}(Z_{im} = Z_{jm} = \ell)\right)\right), \quad i < j,$$

*where g is any link function. If $\tilde{B} \succeq 0, g^{-1}(B) \succ 0$ and **(A1)**, **(A3)** hold, then for any fixed $z \in [L_1] \times \cdots \times [L_M]$, $\sigma = \mathrm{id}$ is the unique minimizer of the quantity:*

$$\sum_{k=1}^{K} \|X_{\tilde{B}}(\sigma(k), z) - X_{\tilde{B}}(k, \mathbf{1}_M)\|_2^2.$$

Proof. The proof follows directly from the above, noting that in this case $|\tilde{B}| = \tilde{B}$, while the submatrix of interest is now $Q = \alpha_n g^{-1}(B)$. We needed to show that $Q \succ 0$, which follows now by assumption. □

Next, we show that the estimation error due to use of $\hat{X}_{\tilde{B}}$ in place of $X_{\tilde{B}}$ vanishes asymptotically. Note that relabeling permutations appear here.

Lemma S17. *Assume the conditions of Theorem 3 hold. Let $X_{\tilde{B}}$ as in Proposition 1 and $\hat{X}_{\tilde{B}}$ as in Algorithm 1. For any fixed $z \in [L_1] \times \cdots \times [L_M]$, let:*

$$\hat{L}_z(\sigma) = \sum_{k=1}^{K} \|\hat{X}_{\tilde{B}}(\sigma(k), z) - \hat{X}_{\tilde{B}}(k, \mathbf{1}_M)\|_2^2$$
$$L_z(\sigma) = \sum_{k=1}^{K} \|X_{\tilde{B}}((\sigma \circ \tau_z)(k), z) - \hat{X}_{\tilde{B}}(\tau_{\mathbf{1}_M}(k), \mathbf{1}_M)\|_2^2.$$

Then for any $\sigma_1, \sigma_2 \in S_{[K]}$:

$$\alpha_n^{-1}(\hat{L}_z(\sigma_1) - \hat{L}_z(\sigma_2)) = \alpha_n^{-1}(L_z(\sigma_1) - L_z(\sigma_2)) + o_P\left(\frac{1}{\sqrt{n \log^c n}}\right).$$

Proof. By an argument similar to the proof of Theorem S15, we observe that:

$$\hat{L}_z(\sigma) = \hat{c}_z - 2\sum_{k=1}^{K} \langle \hat{X}_{\tilde{B}}(\sigma(k), z), \hat{X}_{\tilde{B}}(k, \mathbf{1}_M)\rangle$$
$$L_z(\sigma) = c_z - 2\sum_{k=1}^{K} \langle X_{\tilde{B}}((\sigma \circ \tau_z)(k), z), \hat{X}_{\tilde{B}}(\tau_{\mathbf{1}_M}(k), \mathbf{1}_M)\rangle$$

for some constants $\hat{c}_z$ and c_z. Moreover, continuing to extend the arguments from the proof of Theorem S15, we have:

$$\begin{aligned}
\langle \hat{X}_{\tilde{B}}(\sigma(k), z), \hat{X}_{\tilde{B}}(k, \mathbf{1}_M)\rangle &= |\hat{\tilde{B}}|_{\tilde{\theta}(\sigma(k), z), \tilde{\theta}(k, \mathbf{1})} \\
\langle X_{\tilde{B}}((\sigma \circ \tau_z)(k), z), \hat{X}_{\tilde{B}}(\tau_{\mathbf{1}_M}(k), \mathbf{1}_M)\rangle &= |\tilde{B}|_{\tilde{\theta}((\sigma \circ \tau_z)(k), z), \tilde{\theta}(\tau_{\mathbf{1}_M}(k), \mathbf{1})} \\
&= (T|\tilde{B}|T^{-1})_{\tilde{\theta}(\sigma(k), z), \tilde{\theta}(k, \mathbf{1})} \\
&= |T\tilde{B}T^{-1}|_{\tilde{\theta}(\sigma(k), z), \tilde{\theta}(k, \mathbf{1})},
\end{aligned}$$

where T is the permutation matrix from Theorem 2. Note that the last line follows from Fact S5. Therefore:

$$\begin{aligned}
&\hat{L}_z(\sigma_1) - \hat{L}_z(\sigma_2) - (L_z(\sigma_1) - L_z(\sigma_2)) \\
&\quad = -2\sum_{k=1}^{K} |\hat{\tilde{B}}|_{\tilde{\theta}(\sigma_1(k),z),\tilde{\theta}(k,\mathbf{1})} + 2\sum_{k=1}^{K} |\hat{\tilde{B}}|_{\tilde{\theta}(\sigma_2(k),z),\tilde{\theta}(k,\mathbf{1})} \\
&\qquad + 2\sum_{k=1}^{K} |T\tilde{B}T^{-1}|_{\tilde{\theta}(\sigma_1(k),z),\tilde{\theta}(k,\mathbf{1})} - 2\sum_{k=1}^{K} |T\tilde{B}T^{-1}|_{\tilde{\theta}(\sigma_2(k),z),\tilde{\theta}(k,\mathbf{1})} \\
&\quad = 2\sum_{k=1}^{K} \left(|\hat{\tilde{B}}|_{\tilde{\theta}(\sigma_2(k),z),\tilde{\theta}(k,\mathbf{1})} - |T\tilde{B}T^{-1}|_{\tilde{\theta}(\sigma_2(k),z),\tilde{\theta}(k,\mathbf{1})}\right) \\
&\qquad - 2\sum_{k=1}^{K} \left(|\hat{\tilde{B}}|_{\tilde{\theta}(\sigma_1(k),z),\tilde{\theta}(k,\mathbf{1})} - |T\tilde{B}T^{-1}|_{\tilde{\theta}(\sigma_1(k),z),\tilde{\theta}(k,\mathbf{1})}\right).
\end{aligned}$$

Observe that the final expression consists of $2K$ terms of the form $2(|\hat{\tilde{B}}|_{ij} - |T\tilde{B}T^{-1}|_{ij})$. Combining Theorem 2 and Fact S8, we know that:

$$\alpha_n^{-1} \| \, |\hat{\tilde{B}}| - |T\tilde{B}T^{-1}| \, \|_F = o_P\left(\frac{1}{\sqrt{n \log^c n}}\right),$$

from which we claim a bound on the entrywise error for any $i, j \in [K\tilde{L}]$:

$$\alpha_n^{-1}(|\hat{\tilde{B}}|_{ij} - |T\tilde{B}T^{-1}|_{ij}) = o_P\left(\frac{1}{\sqrt{n \log^c n}}\right).$$

Summarizing, then, we have:

$$\alpha_n^{-1}\left(\hat{L}_z(\sigma_1) - \hat{L}_z(\sigma_2) - (L_z(\sigma_1) - L_z(\sigma_2))\right) = 4K \cdot o_P\left(\frac{1}{\sqrt{n \log^c n}}\right).$$

Since K is constant, the final result follows by simple rearrangement. □

For completeness, we end with a formal proof of Theorem 3.

Proof of Theorem 3. Let $\hat{L}_z : S_{[K]} \to \mathbb{R}$ and $L_z : S_{[K]} \to \mathbb{R}$ as in the statement of Lemma S17. We first rewrite the result of Theorem S15 in a permuted order. For any fixed z:

$$\begin{aligned}
&\arg\min_{\sigma \in S_{[K]}} L_z(\sigma) \\
&\quad = \arg\min_{\sigma \in S_{[K]}} \sum_{k=1}^{K} \| X_{\tilde{B}}\left((\sigma \circ \tau_z)(k), z\right) - X_{\tilde{B}}\left(\tau_{\mathbf{1}_M}(k), \mathbf{1}_M\right) \|_2^2 \\
&\quad = \tau_{\mathbf{1}_M} \circ \tau_z^{-1}.
\end{aligned}$$

This follows from the commutativity of the sum and the fact that $S_{[K]}$ is closed under composition. In other words, we may think of the sum as going in order of $\tau_{\mathbf{1}_M}(1), \ldots, \tau_{\mathbf{1}_M}(K)$ and minimizing over $\sigma \circ \tau_z \in S_{[K]}$ instead, if we prefer, in which case recovering the identity permutation is equivalent to recovering $\sigma \circ \tau_z = \tau_{\mathbf{1}_M}$.

For each z, let $\sigma_z^* = \tau_{\mathbf{1}_M} \circ \tau_z^{-1}$ denote the optimal permutation, and let:

$$\begin{aligned} a_z &= L_z(\sigma_z^*), \\ b_z &= \arg\min_{\sigma \neq \sigma_z^*} L_z(\sigma), \text{ and} \\ \Delta_z &= b_z - a_z, \end{aligned}$$

so that Δ_z denotes the gap between the optimal and second-best permutation. Let $\Delta_0 = \min_z \Delta_z$. Since $X_{\tilde{B}}$ scales with $\sqrt{\alpha_n}$, $L_z(\cdot)$ scales with α_n, and the quantity $\alpha_n^{-1}\Delta_0$ is constant. By assumption ***(A2)***, we may further assume $\Delta_0 > 0$.

By Lemma S17, we have that for any permutation $\sigma \in S_{[K]}$:

$$\alpha_n^{-1}(\hat{L}_z(\sigma) - \hat{L}_z(\sigma_z^*)) = \alpha_n^{-1}(L_z(\sigma) - L_z(\sigma_z^*)) + o_P\left(\frac{1}{\sqrt{n \log^c n}}\right).$$

We would like these error terms to be less than $\alpha_n^{-1}\Delta_0/2$ for all z. Since $\alpha_n^{-1}\Delta_0/2$ is constant, this happens with high probability for sufficiently large n. In this case, we have:

$$\hat{\sigma}_z = \arg\min_{\sigma \in S_{[K]}} \hat{L}_z(\sigma) = \arg\min_{\sigma \in S_{[K]}} L_z(\sigma) = \sigma_z^* = \tau_{\mathbf{1}_M} \circ \tau_z^{-1}.$$

Consequently, for all $i \in \mathcal{I}_z$, since $\hat{\theta}_z(i) = \tau_z(\theta_i)$, we have our desired result:

$$\hat{\sigma}_z(\hat{\theta}_z(i)) = \tau_{\mathbf{1}_M}(\tau_z^{-1}(\tau_z(\theta_i))) = \tau_{\mathbf{1}_M}(\theta_i).$$

□

References

Bhatia, R. (2013). *Matrix analysis*, Volume 169. Springer Science & Business Media.

Bickel, P., D. Choi, X. Chang, and H. Zhang (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics 41*(4), 1922–1943.

Horn, R. A. and C. R. Johnson (1991). *Topics in Matrix Analysis.* Cambridge University Press.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis*. Cambridge University Press.

Rubin-Delanchy, P., J. Cape, M. Tang, and C. E. Priebe (2017). A statistical interpretation of spectral embedding: the generalised random dot product graph. *arXiv preprint arXiv:1709.05506*.

Tang, M., J. Cape, and C. E. Priebe (2022). Asymptotically efficient estimators for stochastic blockmodels: The naive mle, the rank-constrained mle, and the spectral estimator. *Bernoulli 28*(2), 1049–1073.