Biometrika (2017), **xx**, x, *pp*. 1–13

Printed in Great Britain

Supplementary material for 'Covariate-assisted spectral clustering'

BY N. BINKIEWICZ
Department of Statistics, University of Wisconsin, 1300 University Ave., Madison 53706, U.S.A.
norbertbin@gmail.com

J. T. VOGELSTEIN
Department of Biomedical Engineering, Johns Hopkins University, 720 Rutland Ave., Baltimore 21205, U.S.A.
jovo@jhu.edu

AND K. ROHE
Department of Statistics, University of Wisconsin, 1300 University Ave., Madison 53706, U.S.A.
karlrohe@stat.wisc.edu

1. DISCONTINUOUS TRANSITIONS IN THE LEADING EIGENSPACE OF $\tilde{L}$

Discontinuous changes in the leading eigenspace of $\tilde{L}(\alpha)$ are a major concern when determining an optimal α value since they have a large effect on the clustering results. They can be studied algebraically by expressing $\tilde{L}(\alpha)$ in terms of the eigenvectors of $L_\tau L_\tau$ and XX^T. This approach is motivated by Brand (2006).

Let $L_\tau L_\tau = V\Lambda V^T$ and P be the orthogonal basis of the column space of $(I - VV^T)XX^T$, the component of XX^T orthogonal to V. Let $XX^T = \tilde{V}\tilde{\Lambda}\tilde{V}^T$ and $\tilde{X}_i = \tilde{\lambda}_i^{1/2}\tilde{V}_i$, so $XX^T = \tilde{X}\tilde{X}^T$. Then, $\tilde{L}$ can be written as follows.

$$
\begin{aligned}
\tilde{L} &= L_\tau L_\tau + \alpha XX^T \\
&= L_\tau L_\tau + \alpha \tilde{X}\tilde{X}^T \\
&= \begin{bmatrix} V & \tilde{X} \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & \alpha I \end{bmatrix} \begin{bmatrix} V & \tilde{X} \end{bmatrix}^T \\
&= \begin{bmatrix} V & P \end{bmatrix} \begin{bmatrix} I & V^T\tilde{X} \\ 0 & P^T(I - VV^T)\tilde{X} \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & \alpha I \end{bmatrix} \begin{bmatrix} I & 0 \\ \tilde{X}^T V & \tilde{X}^T(I - VV^T)P \end{bmatrix} \begin{bmatrix} V & P \end{bmatrix}^T \\
&= \begin{bmatrix} V & P \end{bmatrix} \begin{bmatrix} \Lambda + \alpha V^T\tilde{X}\tilde{X}^T V & \alpha V^T\tilde{X}\tilde{X}^T(I - VV^T)P \\ \alpha P^T(I - VV^T)\tilde{X}\tilde{X}^T V & \alpha P^T(I - VV^T)\tilde{X}\tilde{X}^T(I - VV^T)P \end{bmatrix} \begin{bmatrix} V & P \end{bmatrix}^T \\
&= \begin{bmatrix} V & P \end{bmatrix} S \begin{bmatrix} V & P \end{bmatrix}^T \\
&= \Big(\begin{bmatrix} V & P \end{bmatrix} V' \Big) \Lambda' \Big(V'^T \begin{bmatrix} V & P \end{bmatrix}^T \Big).
\end{aligned}
$$

Note that

$$(\Lambda + \alpha V^T\tilde{X}\tilde{X}^T V)_{ij} = \lambda_i\delta_{ij} + \alpha \sum_k (V_i^T\tilde{X}_k)(\tilde{X}_k^T V_j)$$

and

$$P^T(I - VV^T)\tilde{X}\tilde{X}^T V = \{P^T(I - VV^T)\tilde{X}\}[(\tilde{X}_i^T V_j)_{ij}].$$

Hence, for any j such that $\tilde{X}_i^T V_j = 0$, for all i, the jth row and column of S will be zero except for the diagonal element. This means that U_j will not be rotated by V' and will be an eigenvector of $\tilde{L}$ for all values of α. The eigenvalue λ_j will not change either, but its position relative to the other eigenvalues will change with α. The change in the relative position of λ_j will result in a discontinuous transition in the leading eigenspace of $\tilde{L}$ if $j \geq K$.

For any i such that $\tilde{X}_i^T V_j = 0$ for all j, $\tilde{V}_i$ is a column in P by construction. Row i in the lower left block of S is

$$\begin{aligned}\tilde{V}_i^T(I - VV^T)\tilde{X}[(\tilde{X}_i^T V_j)_{ij}] &= [0, ..., \tilde{\lambda}_i^{1/2}, 0, ...][(\tilde{X}_i^T V_j)_{ij}] \\ &\quad - [0, ..., 1, 0, ...]\mathrm{diag}\left(\tilde{\lambda}_1^{1/2}, ..., \tilde{\lambda}_R^{1/2}\right)[(\tilde{X}_i^T V_j)_{ij}] \\ &= [0, ..., 0],\end{aligned}$$

and, since S is symmetric, this is also column i in the upper right block of S. The lower right block of S has row i, and by symmetry column i, given by

$$\begin{aligned}\tilde{V}_i^T(I - VV^T)\tilde{X}\tilde{X}^T(I - VV^T)P &= \tilde{v}_i^T(\tilde{X}\tilde{X}^T - \tilde{X}\tilde{X}^T VV^T)P \\ &= \tilde{\lambda}_i \tilde{V}_i^T P \\ &= [0, ..., \tilde{\lambda}_i, 0, ...].\end{aligned}$$

Thus, for any i such that $\tilde{X}_i^T V_j = 0$, for all j the ith row and column of S will be zero except for the diagonal element. This means that $\tilde{V}_i$ and $\tilde{\lambda}_i$ will be an eigenvector and eigenvalue of $\tilde{L}$ for all values of α, but will occupy different relative positions in the eigendecomposition based on the value of α. The change in the relative position of $\tilde{\lambda}_i$ will result in a discontinuous transition in the leading eigenspace of $\tilde{L}$ if $i \geq K$.

Knowing the interval on which such discontinuous transitions are possible can reduce the computational burden of choosing an optimal α. The values of α for which transitions occur can be identified as points at which the eigengap equals zero, $\lambda_K(\tilde{L}) - \lambda_{K+1}(\tilde{L}) = 0$. First, consider the lowest possible value of α for which such a transition can occur, $\alpha = \mathrm{argmin}_\alpha\{\alpha : \lambda_K(\tilde{L}) - \lambda_{K+1}(\tilde{L}) = 0\}$. Note that $\lambda_K(\tilde{L}) \geq \lambda_K(L_\tau L_\tau)$, where the equality holds when V_K is orthogonal to X and α is sufficiently small, and $\lambda_{K+1}(\tilde{L}) \leq \lambda_{K+1}(L_\tau L_\tau) + \alpha\lambda_1(XX^T)$, where the equality holds when V_{K+1} is identical to $\tilde{V}_1$. Hence, the earliest possible transition occurs when

$$\lambda_K(L_\tau L_\tau) - \{\lambda_{K+1}(L_\tau L_\tau) + \alpha_{min}\lambda_1(XX^T)\} = 0,$$
$$\alpha_{min} = \frac{\lambda_K(L_\tau L_\tau) - \lambda_{K+1}(L_\tau L_\tau)}{\lambda_1(XX^T)}.$$

For the highest value of α for which such a transition is possible, consider $\alpha^{-1}\tilde{L}$. Following the above argument for α^{-1} with XX^T and $L_\tau L_\tau$ interchanged, a symmetric result is obtained with the additional dependence on the number of covariates, R. This result yields,

$$\alpha_{max} = \frac{\lambda_1(L_\tau L_\tau)}{\lambda_R(XX^T)\mathbf{1}_{(R\leq K)} + \{\lambda_K(XX^T) - \lambda_{K+1}(XX^T)\}\mathbf{1}_{(R>K)}}.$$

Therefore, discontinuous transitions in the leading eigenspace of $\tilde{L}(\alpha)$ can only occur in the interval $[\alpha_{\min}, \alpha_{\max}]$.

2. EMPIRICAL RESULTS FOR CHOOSING α

Figure 1 presents some empirical details to demonstrate how the within cluster sum of squares and the mis-clustering rate vary with the tuning parameter α. The simulations shown in the figure use the same model structure described in §4 of the paper. The results show the minimum of the within cluster sum of squares falls within the prescribed range of α, $[\alpha_{\min}, \alpha_{\max}]$. Furthermore, the minimum of the within cluster sum of squares tends to align with the minimum of the mis-clustering rate. Similar results were observed for other parameter settings.

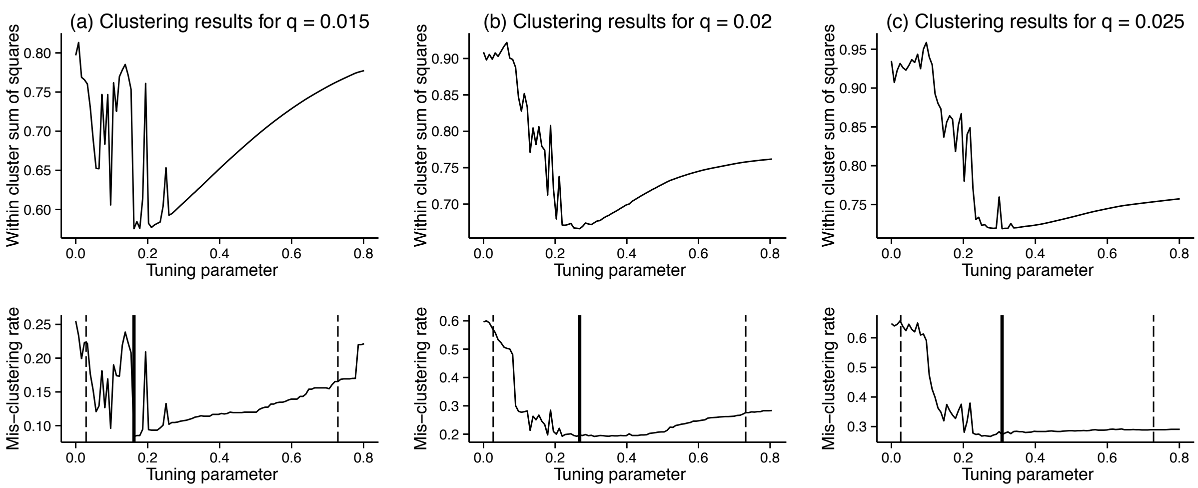

Fig. 1. The results of assortative covariate-assisted spectral clustering for a range of α values. The solid line in bottom graphs indicates the α value chosen by the optimization procedure and the dased lines indicate the interval $[\alpha_{\min}, \alpha_{\max}]$. The fixed parameters are $N = 1500$, $p = 0.03$, $m_1 = 0.8$, and $m_2 = 0.2$.

3. PROOF OF LEMMA 1

This proof follows the approach used in Rohe et al. (2011) to establish the equivalence between block membership and a subset of the population eigenvectors. Note that $\tilde{\mathcal{L}} = (\mathcal{D} + \tau I)^{-1/2} ZBZ^T(\mathcal{D}+\tau I)^{-1}ZBZ^T(\mathcal{D}+\tau I)^{-1/2} + \alpha E(XX^T)$. Define $c_l = \sum_i \mathrm{var}(X_{il} \mid Z_i = l)$, a diagonal matrix $\tilde{C}$ such that $\tilde{C}_{ll} = c_l$, and a diagonal matrix C such that $CZ = Z\tilde{C}$.

If we let $\mathcal{D}_B = \mathrm{diag}(BZ^T\mathbf{1}_n + \tau)$, then $\tilde{\mathcal{L}} = Z\{\mathcal{D}_B^{-1/2}BZ^T(\mathcal{D}+\tau I)^{-1}ZB\mathcal{D}_B^{-1/2} + \alpha MM^T\}Z^T + \alpha C$. Recall that B is symmetric and full rank by assumption. Let $\tilde{B} = \mathcal{D}_B^{-1/2}BZ^T(\mathcal{D}+\tau I)^{-1}ZB\mathcal{D}_B^{-1/2} + \alpha MM^T$, which is positive definite for all $\alpha \geq 0$. Assume α is chosen such that $\tilde{B}$ is full rank, which is true for all α with the possible exception of a set of values of measure zero. Let $\tilde{P} = Z^TZ$ and note that $\det(\tilde{B}\tilde{P}) = \det(\tilde{B})\det(\tilde{P}) > 0$. Hence, $\tilde{B}\tilde{P} + \alpha\tilde{C}$ is symmetric and has real eigenvalues. By spectral decomposition, let

$$\tilde{B}\tilde{P} + \alpha\tilde{C} = \mu\Lambda\mu^T.$$

Then,

$$\begin{aligned}\tilde{\mathcal{L}}Z\mu &= (Z\tilde{B}Z^T + \alpha C)Z\mu \\ &= (Z\tilde{B}Z^T Z + \alpha CZ)\mu \\ &= (Z(\tilde{B}\tilde{P}) + \alpha Z\tilde{C})\mu \\ &= Z\mu\Lambda.\end{aligned}$$

Therefore, $Z\mu$ is the matrix of K eigenvectors of $\tilde{\mathcal{L}}$, but not necessarily the top K. Also, $\det(\mu) > 0$ so μ^{-1} exists and $Z_i\mu = Z_j\mu \iff Z_i = Z_j$. This establishes the equivalence between block membership and a subset of the population eigenvectors. A condition will now be derived for which this equivalence holds for the top K population eigenvectors. Let x be a normalized eigenvector orthogonal to the span of $Z\mu$. Because μ has orthogonal columns, it is full rank. As such, $x^T Z = 0$.

Define $\bar{c} = \sum^K c_l/K$, $\bar{C} = \bar{c}I$, and $\varkappa = \max_l |c_l - \bar{c}|$, then

$$\begin{aligned}x^T\tilde{\mathcal{L}}x &= x^T(Z\tilde{B}Z^T + \alpha C)x \\ &= \alpha x^T C x \\ &= \alpha x^T(\bar{C} + (C - \bar{C}))x \\ &= \alpha x^T \bar{c} I x + \alpha x^T (C - \bar{C})x \\ &= \alpha\bar{c} + \alpha x^T(C - \bar{C})x \\ &\le \alpha\bar{c} + \alpha||C - \bar{C}|| \\ &= \alpha(\bar{c} + \varkappa).\end{aligned}$$

The kth eigenvalue of $\tilde{B}\tilde{P} + \alpha\tilde{C}$ is given by

$$\begin{aligned}\lambda_K(\tilde{B}\tilde{P} + \alpha\tilde{C}) &= \min_{||x||=1} x^T(\tilde{B}\tilde{P} + \alpha\tilde{C})x \\ &= \min x^T[(\tilde{B}\tilde{P} + \alpha\bar{c}I) + (\alpha\tilde{C} - \alpha\bar{c}I)]x \\ &\ge \min x^T(\tilde{B}\tilde{P} + \alpha\bar{c}I)x + \alpha\min x^T(\tilde{C} - \bar{c}I)x \\ &= \min x^T\tilde{B}\tilde{P}x + \alpha\bar{c} - \alpha\max x^T(\bar{c}I - \tilde{C})x \\ &\ge \lambda_K(\tilde{B}\tilde{P}) + \alpha\bar{c} - \alpha\max_u |c_u - \bar{c}| \\ &= \lambda_K(\tilde{B}\tilde{P}) + \alpha(\bar{c} - \varkappa).\end{aligned}$$

Hence, a positive eigengap exists between the eigenvectors in $Z\mu$ and x if

$$\begin{aligned}0 &< \lambda_K(\tilde{B}\tilde{P} + \alpha\tilde{C}) - \max_x x^T(Z\tilde{B}Z^T + \alpha C)x \\ &< \lambda_K(\tilde{B}\tilde{P}) + \alpha(\bar{c} - \varkappa) - \alpha(\bar{c} + \varkappa) \\ &= \lambda_K(\tilde{B}\tilde{P}) - 2\alpha\varkappa.\end{aligned}$$

Assume (i) $\lambda_K(\tilde{B}\tilde{P}) > 2\alpha\varkappa$, then the top K eigenvectors of $\tilde{\mathcal{L}}$ are given by $Z\mu$, where $Z_i\mu = Z_j\mu \iff Z_i = Z_j$. Hence, there is an equivalence between block membership and the top K population eigenvectors.

4. Proof of Theorem 1

4·1. *Triangle inequality bound*

The spectral norm of the difference between the sample and population covariate-assisted Laplacians is bounded by first applying the triangle inequality and bounding the resulting terms individually.

$$
\begin{aligned}
\|\tilde{L} - \tilde{\mathcal{L}}\| \leq{}& \|\alpha XX^T - E(\alpha XX^T)\| && (1)\\
&+ \|\mathcal{D}_\tau^{-1/2} A \mathcal{D}_\tau^{-1} A \mathcal{D}_\tau^{-1/2} - \mathcal{D}_\tau^{-1/2} \mathcal{A} \mathcal{D}_\tau^{-1} \mathcal{A} \mathcal{D}_\tau^{-1/2}\| && (2)\\
&+ \|D_\tau^{-1/2} A D_\tau^{-1} A D_\tau^{-1/2} - \mathcal{D}_\tau^{-1/2} A \mathcal{D}_\tau^{-1} A \mathcal{D}_\tau^{-1/2}\|. && (3)
\end{aligned}
$$

4·2. *Bound for Equation (1)*

For equation (1), use the matrix Bernstein inequality (Tropp, 2012). Note that $\alpha XX^T = \sum_k \alpha X_k X_k^T$, where X_k is the kth column of X. Now bound the spectral norm of $\alpha X_k X_k^T - E(\alpha X_k X_k^T)$.

$$
\begin{aligned}
\|\alpha X_k X_k^T - E(\alpha X_k X_k^T)\| &= \alpha \|X_k X_k^T - \mathcal{X}_k \mathcal{X}_k^T - \mathrm{diag}(\mathcal{X}_k^{(2)} - \mathcal{X}_k^2)\|\\
&\leq \alpha(\|X_k X_k^T\| + \|\mathcal{X}_k \mathcal{X}_k^T\| + \max|\mathcal{X}_k^{(2)} - \mathcal{X}_k^2|)\\
&\leq \alpha(NJ^2 + NJ^2 + J^2)\\
&\leq 3\alpha N J^2\\
&\equiv S.
\end{aligned}
$$

Next, find a bound on the spectral norm of the variance of αXX^T. Let $\mathcal{X}_k^{(i)}$ be the ith moment of X_k. Note that vector products are element-wise where dictated by vector dimensions.

$$
\begin{aligned}
E(X_k X_k^T) ={}& \mathcal{X}_k \mathcal{X}_k^T - \mathrm{diag}(\mathcal{X}_k^2 - \mathcal{X}_k^{(2)}).\\
E(X_k X_k^T) E(X_k X_k^T) ={}& \{\mathcal{X}_k \mathcal{X}_k^T - \mathrm{diag}(\mathcal{X}_k^2 - \mathcal{X}_k^{(2)})\}\{\mathcal{X}_k \mathcal{X}_k^T - \mathrm{diag}(\mathcal{X}_k^2 - \mathcal{X}_k^{(2)})\}\\
={}& \mathcal{X}_k \mathcal{X}_k^T \mathcal{X}_k \mathcal{X}_k^T - \mathcal{X}_k \mathcal{X}_k^T \mathrm{diag}(\mathcal{X}_k^2 - \mathcal{X}_k^{(2)})\\
&- \mathrm{diag}(\mathcal{X}_k^2 - \mathcal{X}_k^{(2)}) \mathcal{X}_k \mathcal{X}_k^T + \mathrm{diag}\{(\mathcal{X}_k^2 - \mathcal{X}_k^{(2)})^2\}\\
={}& \Big(\sum_i \mathcal{X}_{ik}^2\Big) \mathcal{X}_k \mathcal{X}_k^T - \mathcal{X}_k \{\mathcal{X}_k (\mathcal{X}_k^2 - \mathcal{X}_k^{(2)})\}^T\\
&- \{\mathcal{X}_k (\mathcal{X}_k^2 - \mathcal{X}_k^{(2)})\} \mathcal{X}_k^T + \mathrm{diag}\{(\mathcal{X}_k^2 - \mathcal{X}_k^{(2)})^2\}.
\end{aligned}
$$

$$
\begin{aligned}
E(X_k X_k^T X_k X_k^T) ={}& E\Big\{\Big(\sum_i X_{ik}^2\Big) X_k X_k^T\Big\}\\
={}& \begin{cases} \mathcal{X}_{ik}\mathcal{X}_{jk} \sum_{l \neq i,j} \mathcal{X}_{lk}^{(2)} + \mathcal{X}_{ik}\mathcal{X}_{jk}^{(3)} + \mathcal{X}_{jk}\mathcal{X}_{ik}^{(3)} & i \neq j\\ \mathcal{X}_{ik}^{(2)} \sum_{l \neq i} \mathcal{X}_{lk}^{(2)} + \mathcal{X}_{ik}^{(4)} & i = j \end{cases}\\
={}& \Big(\sum \mathcal{X}_{ik}^{(2)}\Big) \mathcal{X}_k \mathcal{X}_k^T - \mathcal{X}_k (\mathcal{X}_k \mathcal{X}_k^{(2)})^T - (\mathcal{X}_k \mathcal{X}_k^{(2)}) \mathcal{X}_k^T\\
&+ \mathcal{X}_k \mathcal{X}_k^{(3)^T} + \mathcal{X}_k^{(3)} \mathcal{X}_k^T\\
&+ \mathrm{diag}\Big\{(\mathcal{X}_k^{(2)} - \mathcal{X}_k^2)\Big(\sum_i \mathcal{X}_{ik}^{(2)}\Big) - \mathcal{X}_k^{(2)^2} + 2\mathcal{X}_k^2 \mathcal{X}_k^{(2)} - 2\mathcal{X}_k \mathcal{X}_k^{(3)} + \mathcal{X}_k^{(4)}\Big\}.
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(X_k X_k^T) =& \mathcal{X}_k \mathcal{X}_k^T \sum_i (\mathcal{X}_{ik}^{(2)} - \mathcal{X}_{ik}^2) + \mathcal{X}_k \{\mathcal{X}_k(\mathcal{X}_k^2 - 2\mathcal{X}_k^{(2)}) + \mathcal{X}_k^{(3)}\}^T + \{\mathcal{X}_k(\mathcal{X}_k^2 - 2\mathcal{X}_k^{(2)}) + \mathcal{X}_k^{(3)}\}\mathcal{X}_k^T \\
&+ \mathrm{diag}\left\{(\mathcal{X}_k^{(2)} - \mathcal{X}_k^2)\Big(\sum_i \mathcal{X}_{ik}^{(2)}\Big) - \mathcal{X}_k^{(2)^2} + 2\mathcal{X}_k^2\mathcal{X}_k^{(2)} - 2\mathcal{X}_k\mathcal{X}_k^{(3)} + \mathcal{X}_k^{(4)} - (\mathcal{X}_k^{(2)} - \mathcal{X}_k^2)^2\right\}.
\end{aligned}
$$

$$
\begin{aligned}
\Big|\Big|\sum_k \mathrm{var}(X_k X_k^T)\Big|\Big| \le& \sum_k\sum_i \Big|\mathcal{X}_{ik}^2 \sum_l (\mathcal{X}_{lk}^{(2)} - \mathcal{X}_{lk}^2)| + 2|\mathcal{X}_{ik}^4 - 2\mathcal{X}_{ik}^2\mathcal{X}_{ik}^{(2)} + \mathcal{X}_{ik}\mathcal{X}_{ik}^{(3)}\Big| \\
&+ \max_i \Big|(\mathcal{X}_{ik}^{(2)} - \mathcal{X}_{ik}^2)\Big(\sum_l \mathcal{X}_{lk}^{(2)}\Big) - \mathcal{X}_{ik}^{(2)^2} + 2\mathcal{X}_{ik}^2\mathcal{X}_{ik}^{(2)} - 2\mathcal{X}_{ik}\mathcal{X}_{ik}^{(3)} + \mathcal{X}_{ik}^{(4)} - (\mathcal{X}_{ik}^{(2)} - \mathcal{X}_{ik}^2)^2\Big| \\
\le& \sum_k\sum_i \mathcal{X}_{ik}^2 \sum_l (\mathcal{X}_{lk}^{(2)} - \mathcal{X}_{lk}^2) + 2(\mathcal{X}_{ik}^2\mathcal{X}_{ik}^{(2)} - \mathcal{X}_{ik}^4) + 2|\mathcal{X}_{ik}^2\mathcal{X}_{ik}^{(2)} - \mathcal{X}_{ik}\mathcal{X}_{ik}^{(3)}| \\
&+ \max_i \Big\{(\mathcal{X}_{ik}^{(2)} - \mathcal{X}_{ik}^2)\Big(\sum_l \mathcal{X}_{lk}^{(2)}\Big) + |2\mathcal{X}_{ik}\mathcal{X}_{ik}^{(3)} - \mathcal{X}_{ik}^{(4)} - \mathcal{X}_{ik}^4| + 2(\mathcal{X}_{ik}^{(2)} - \mathcal{X}_{ik}^2)^2\Big\} \\
\le& \sum_k\sum_i 3\mathcal{X}_{ik}^2 \sum_l (\mathcal{X}_{lk}^{(2)} - \mathcal{X}_{lk}^2) + 2|\mathcal{X}_{ik}^2\mathcal{X}_{ik}^{(2)} - \mathcal{X}_{ik}\mathcal{X}_{ik}^{(3)}| \\
&+ \max_i \Big\{3(\mathcal{X}_{ik}^{(2)} - \mathcal{X}_{ik}^2)\Big(\sum_l \mathcal{X}_{lk}^{(2)}\Big) + |2\mathcal{X}_{ik}\mathcal{X}_{ik}^{(3)} - \mathcal{X}_{ik}^{(4)} - \mathcal{X}_{ik}^4|\Big\} \\
\le& 8\sum_k \left\{\sum_i \mathcal{X}_{ik}^{(2)} \sum_l (\mathcal{X}_{lk}^{(2)} - \mathcal{X}_{lk}^2) + \mathcal{X}_{ik}^{(4)}\right\}.
\end{aligned}
$$

Thus,

$$
\Big|\Big|\sum_k \mathrm{var}(\alpha X_k X_k^T)\Big|\Big| \le 8\alpha^2 \sum_k \left\{\sum_i \mathcal{X}_{ik}^{(2)} \sum_l (\mathcal{X}_{lk}^{(2)} - \mathcal{X}_{lk}^2) + \mathcal{X}_{ik}^{(4)}\right\} \equiv \varpi.
$$

Let $b = \{3\varpi \log(8N/\epsilon)\}^{1/2}$ and assume (iii) $\varpi/S^2 > 3\log(8N/\epsilon)$, then $b < \varpi/S$. Note that assumption (iii) requires that $R \ge \Theta(\log N)$. Applying the matrix Bernstein inequality gives,

$$
\begin{aligned}
P(||\alpha XX^T - E(\alpha XX^T)|| > b) &\le 2N\exp\left(-\frac{b^2}{2\sigma^2 + 2Sb/3}\right) \\
&\le 2N\exp\left\{-\frac{3S\varpi\log(8N/\epsilon)}{2\varpi + 2Sb/3}\right\} \\
&\le 2N\exp\left\{-\frac{3\varpi\log(8N/\epsilon)}{3\varpi}\right\} \\
&= \epsilon/4.
\end{aligned}
$$

Hence, with with probability $1 - \epsilon/4$,

$$
\|\alpha XX^T - E(\alpha XX^T)\| \le b.
$$

4·3. *Bound for Equation* (2)

Equation (2) can be decomposed into three terms using properties of the spectral norm.

$$\begin{aligned}
&\|\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1}A\mathcal{D}_\tau^{-1/2} - \mathcal{D}_\tau^{-1/2}\mathcal{A}\mathcal{D}_\tau^{-1}\mathcal{A}\mathcal{D}_\tau^{-1/2}\| \\
&\le \|\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1}A\mathcal{D}_\tau^{-1/2} - E(\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1/2})E(\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1/2})\| \\
&\le \|\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1/2} - E(\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1/2})\|\|\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1/2} + E(\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1/2})\|.
\end{aligned}$$

The first term above can be bounded following the proof in the Supplement of Qin & Rohe (2013). Under the assumption that (ii) $d+\tau > 3\log(8N/\epsilon)$, where $d = \min \mathcal{D}_{ii}$, let $a = [\{3\log(8N/\epsilon)\}/(d+\tau)]^{1/2}$, so $a<1$. Then, with probability at least $1-\epsilon/4$,

$$\|\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1/2} - E(\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1/2})\| \le a.$$

Using the fact that $\|\mathcal{L}_\tau\| \le 1$, $\|L_\tau\| \le 1$, and $\|\mathcal{D}_\tau^{-1/2}D_\tau^{-1/2}\| \le a+1$, with probability $1-\epsilon/4$, as shown in the Supplement of Qin & Rohe (2013), the second term can be bounded with probability $1-\epsilon/4$ as follows.

$$\begin{aligned}
&\|\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1/2} + E(\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1/2})\| \\
&\le \|\mathcal{D}_\tau^{-1/2}D_\tau^{-1/2}L_\tau D_\tau^{-1/2}\mathcal{D}_\tau^{-1/2}\| + \|\mathcal{L}_\tau\| \\
&\le \|\mathcal{D}_\tau^{-1/2}D_\tau^{-1/2}\|\|L_\tau\|\|D_\tau^{-1/2}\mathcal{D}_\tau^{-1/2}\| + 1 \\
&\le (a+1)^2 + 1.
\end{aligned}$$

Hence, with with probability $1-\epsilon/4$,

$$\|\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1}A\mathcal{D}_\tau^{-1/2} - E(\mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1}A\mathcal{D}_\tau^{-1/2})\| \le a(a+1)^2 + a.$$

4·4. *Bound for Equation* (3)

Note that $\|\mathcal{D}_\tau^{-1/2}D_\tau^{-1/2} - I\| \le a$, with probability $1-\epsilon/4$, as shown in the Supplement of Qin & Rohe (2013), and $\|D_\tau^{-1/2}\mathcal{D}_\tau^{-1}D_\tau^{-1/2} - I\| \le a$, which can be derived by the same approach. Using these results, equation (3) can be bounded with probability $1-\epsilon/2$ as follows.

$$\begin{aligned}
&\|D_\tau^{-1/2}AD_\tau^{-1}AD_\tau^{-1/2} - \mathcal{D}_\tau^{-1/2}A\mathcal{D}_\tau^{-1}A\mathcal{D}_\tau^{-1/2}\| \\
&= \|L_\tau L_\tau - \mathcal{D}_\tau^{-1/2}D_\tau^{1/2}L_\tau D_\tau^{1/2}\mathcal{D}_\tau^{-1}D_\tau^{1/2}L_\tau D_\tau^{1/2}\mathcal{D}_\tau^{-1/2}\| \\
&= \|L_\tau L_\tau - L_\tau D_\tau^{1/2}\mathcal{D}_\tau^{-1}D_\tau^{1/2}L_\tau D_\tau^{1/2}\mathcal{D}_\tau^{-1/2} + (I - \mathcal{D}_\tau^{-1/2}D_\tau^{1/2})L_\tau D_\tau^{1/2}\mathcal{D}_\tau^{-1}D_\tau^{1/2}L_\tau D_\tau^{1/2}\mathcal{D}_\tau^{-1/2}\| \\
&\le \|L_\tau(L_\tau - D_\tau^{1/2}\mathcal{D}_\tau^{-1}D_\tau^{1/2}L_\tau D_\tau^{1/2}\mathcal{D}_\tau^{-1/2})\| + a(a+1)^2 \\
&\le \|D_\tau^{1/2}\mathcal{D}_\tau^{-1}D_\tau^{1/2}L_\tau(D_\tau^{1/2}\mathcal{D}_\tau^{-1/2} - I) - (D_\tau^{1/2}\mathcal{D}_\tau^{-1}D_\tau^{1/2} - I)L_\tau\| + a(a+1)^2 \\
&\le a(a+1) + a + a(a+1)^2.
\end{aligned}$$

Consequently, joining the results for the five terms, gives the desired bound. With probability at least $1-\epsilon$,

$$\begin{aligned}
\|\tilde{L} - \tilde{\mathcal{L}}\| &\le 2a^3 + 5a^2 + 5a + b \\
&\le 12a + b \\
&= \{\varpi^{1/2} + 12(d+\tau)^{-1/2}\}\{3\log(8N/\epsilon)\}^{1/2}.
\end{aligned}$$

Let $\delta \equiv \varpi^{1/2} + 12(d+\tau)^{-1/2}$, then the bound becomes

$$\|\tilde{L} - \tilde{\mathcal{L}}\| \le \delta\{3\log(8N/\epsilon)\}^{1/2}.$$

5. PROOF OF THEOREM 2

Using Lemma 9 from McSherry (2001), let $P_{\tilde{L}}$ be the projection onto the span of the first K left singular eigenvectors of $\tilde{L}$. Then, $P_{\tilde{L}}$ is the optimal rank K approximation to $\tilde{L}$ and

$$\|P_{\tilde{L}} - \tilde{\mathcal{L}}\|_F^2 \leq 8K\|\tilde{L} - \tilde{\mathcal{L}}\|^2.$$

Next, apply the Davis–Kahan Theorem to $\tilde{\mathcal{L}}$ (Davis & Kahan, 1970). Let $W \subset \mathbb{R}$ be an interval and define the distance between W and the spectrum of $\tilde{\mathcal{L}}$ outside of W as

$$\Lambda = \min\{|\lambda - r|; \lambda \text{ eigenvalue of } \tilde{\mathcal{L}}, \lambda \notin W, r \in W\}.$$

Choose $W = (\lambda_K/2, \infty)$, where λ_K is the Kth eigenvalue of $\tilde{L}$. Then, $\Lambda = \lambda_K/2$. Let ω_K be the Kth largest eigenvalue of $\tilde{\mathcal{L}}$, then under the assumption that $\delta\{3\log(8N/\epsilon)\}^{1/2} \leq \lambda_K/2$,

$$|\lambda_K - \omega_K| \leq \delta\{3\log(8N/\epsilon)\}^{1/2} \leq \lambda_K/2.$$

Hence, $\omega_K \in W$, and U has the same dimension as $\mathcal{U}$. The Davis-Kahan Theorem implies,

$$\begin{aligned}
\|U - \mathcal{U}O\|_F &\leq \frac{2^{1/2}\|P_{\tilde{L}}\tilde{L} - \tilde{\mathcal{L}}\|_F}{\Lambda} \\
&\leq \frac{8^{1/2}\|P_{\tilde{L}}\tilde{L} - \tilde{\mathcal{L}}\|_F}{\lambda_K} \\
&\leq \frac{8K^{1/2}\|\tilde{L} - \tilde{\mathcal{L}}\|}{\lambda_K} \\
&\leq \frac{8\delta\{3K\log(8N/\epsilon)\}^{1/2}}{\lambda_K}
\end{aligned}$$

with probability at least $1 - \epsilon$.

6. PROOF OF THEOREM 3

This proof follows the arguments given in Qin & Rohe (2013). Let $P = \max_i(Z^TZ)_{ii}$ and

$$\begin{aligned}
||\mathcal{C}_i - \mathcal{C}_j||_2 &\geq ||Z_i(Z^TZ)^{-1/2}V - Z_j(Z^TZ)^{-1/2}V||_2 \\
&\geq 2^{1/2}||Z^TZ||_2 \\
&\geq \left(\frac{2}{P}\right)^{1/2}.
\end{aligned}$$

For all $Z_j \neq Z_i$, a sufficient condition for one observed centroid to be closest to the population centroid is

$$||C_i\mathcal{O}^T - \mathcal{C}_i||_2 < \frac{1}{(2P)^{1/2}} \Rightarrow ||C_i\mathcal{O}^T - \mathcal{C}_i||_2 < ||C_i\mathcal{O}^T - \mathcal{C}_j||_2,$$

since

$$\begin{aligned}
||C_i\mathcal{O}^T - \mathcal{C}_i||_2 < \frac{1}{(2P)^{1/2}} \Rightarrow ||C_i\mathcal{O}^T - \mathcal{C}_j||_2 &\geq ||\mathcal{C}_i - \mathcal{C}_j||_2 - ||C_i\mathcal{O}^T - \mathcal{C}_i||_2 \\
&\geq \left(\frac{2}{P}\right)^{1/2} - \left(\frac{1}{2P}\right)^{1/2} \geq \frac{1}{(2P)^{1/2}}.
\end{aligned}$$

Let $\mathcal{G} = \{i : ||C_i\mathcal{O}^T - \mathcal{C}_i||_2 \geq \frac{1}{(2P)^{1/2}}\}$, so $\mathcal{M} \subset \mathcal{G}$. Define $Q \in \mathbb{R}^{N\times K}$ where the ith row is C_i. By the definition of k-means, $||U - Q||_2 \leq ||U - \mathcal{U}\mathcal{O}||_2$. Applying the triangle inequality gives

$$||Q - Z\mu\mathcal{O}||_2 = ||Q - \mathcal{U}\mathcal{O}||_2 \leq ||U - Q||_2 + ||U - \mathcal{U}\mathcal{O}||_2 \leq 2||U - \mathcal{U}\mathcal{O}||_2.$$

So,

$$\begin{aligned}\frac{||\mathcal{M}||}{N} \leq \frac{||\mathcal{G}||}{N} &= \frac{1}{N}\sum_{i\in\mathcal{G}} 1\\ &\leq \frac{2P}{N}\sum_{i\in\mathcal{G}} ||C_i\mathcal{O}^T - \mathcal{C}_i||_2^2\\ &= \frac{2P}{N}\sum_{i\in\mathcal{G}} ||C_i - Z_i\mu\mathcal{O}||_2^2\\ &\leq \frac{2P}{N}||Q - Z\mu\mathcal{O}||_F^2\\ &\leq \frac{8P}{N}||U - \mathcal{U}\mathcal{O}||_F^2.\end{aligned}$$

Thus, using the result from Theorem 2, with probability at least $1 - \epsilon$,

$$\frac{||\mathcal{M}||}{N} \leq \frac{c_0 KP\delta^2 \log(8N/\epsilon)}{N\lambda_k^2},$$

where $c_0 = 3 \times 2^6$.

7. Proof of Corollary 1

In order to investigate the mis-clustering bound and the accompanying conditions, we make some simplifying assumptions. Assume $B_{i,i} = p$, for all i and $B_{i,j} = q$, for all $i \neq j$; in addition, $M_{i,i} = m_1$, for all i; $M_{i,j} = m_2$, for all $i \neq j$; and $R > 1$. For computational convenience, assume that each block has the same number of nodes N/K and R is a multiple of K. Recall, $\tilde{\mathcal{L}} = Z(\mathcal{D}_B^{-1/2}BZ^T\mathcal{D}_\tau^{-1}ZB\mathcal{D}_B^{-1/2} + \alpha MM^T)Z^T = Z\tilde{B}Z^T$. Therefore,

$$\begin{aligned}\tilde{B} = &\left[\frac{1}{N\{p + (K-1)q\}/K + \tau}\right]^2 \left(\frac{N}{K}\right)[(p-q)^2 I + \{2pq + (K-1)q\}\mathbf{1}_K\mathbf{1}_K^T]\\ &+ \alpha\{(m_p - m_q)I + m_q\mathbf{1}_K\mathbf{1}_K^T\},\end{aligned}$$

where $m_p = R\{m_1^2 + (K-1)m_2^2\}/K$ and $m_q = R\{2m_1m_2 + (K-2)m_2^2\}/K$. For matrices of the form $aI + b\mathbf{1}_K\mathbf{1}_K^T$, $\lambda_K = a$. Note that $m_p - m_q = R(m_1 - m_2)^2/K$. Thus,

$$\lambda_K(\tilde{B}) = \left[\frac{p-q}{N\{p + (K-1)q\}/K + \tau}\right]^2 \left(\frac{N}{K}\right) + \alpha R(m_1 - m_2)^2/K.$$

Recall that $\tilde{\mathcal{L}}$ has the same eigenvalues as $(Z^TZ)^{1/2}\tilde{B}(Z^TZ)^{1/2} = (N/K)^{1/2}I\tilde{B}(N/K)^{1/2}I = (N/K)\tilde{B}$. Hence, the population eigengap is

$$\lambda_K(\tilde{\mathcal{L}}) = \left\{\frac{p-q}{p + (K-1)q + K\tau/N}\right\}^2 + \frac{\alpha NR(m_1 - m_2)^2}{K^2}.$$

Hence, the mis-clustering bound for a growing number of covariates is given by

$$\frac{|\mathcal{M}|}{N} \leq \frac{\{(d+\tau)^{-1} + \alpha(d+\tau)^{-1/2}\Theta(NR^{1/2}) + \alpha^2\Theta(N^2R)\}\Theta(\log N)}{\alpha\Theta(N^2R^2) + \alpha\Theta(NR) + \Theta(1)}.$$

Two conditions in Theorem 3 depend on R. Condition (iii) becomes $R > \Theta(\log N)$ and condition (iv) becomes $\{\alpha NR^{1/2} + (d+\tau)^{-1/2}\}(\log N)^{1/2} \leq \alpha NR + c_0$, which is satisfied for $R \geq \Theta(\log N)$.

Let $R = \Theta\{(\log N)^{a+1}\}$, $d+\tau = \Theta\{(\log N)^{b+1}\}$, and $\alpha = \Theta\{N^{-1}(\log N)^{-1-c}\}$, where $a, b, c \geq 0$, then the mis-clustering rate becomes

$$\frac{|\mathcal{M}|}{N} \leq c_2 \frac{(\log N)^{a-2c} + (\log N)^{(a-b)/2-c} + (\log N)^{-b}}{(\log N)^{2(a-c)} + (\log N)^{a-c} + \Theta(1)}.$$

If c is chosen such that $a > c$, then $(\log N)^{2(a-c)}$ is the dominant term in the denominator and $|\mathcal{M}|/N = O\{(\log N)^{-a}\} + O\{(\log N)^{-(3a+b)/2-3c}\} + O\{(\log N)^{2(c-a)-b}\}$. The mis-clustering rate is minimized when $c = 0$, so the rate becomes $|\mathcal{M}|/N = O\{(\log N)^{-a}\}$.

If c is chosen such that $a \leq c$, $|\mathcal{M}|/N = O\{(\log N)^{a-2c}\} + O\{(\log N)^{(a-b)/2-c}\} + O\{(\log N)^{-b}\}$. The mis-clustering rate is minimized when $c = \frac{a+b}{2}$, so the rate becomes $\frac{|\mathcal{M}|}{N} = O\{(\log N)^{-b}\}$.

Hence, to minimize the mis-clustering rate when $a \leq b$ choose $c = \frac{a+b}{2}$, which yields a mis-clustering rate of $O\{(\log N)^{-b}\}$, and when $a > b$ choose $c = 0$, which gives a mis-clustering rate of $O\{(\log N)^{-a}\}$. If we consider the special case where $a = 0$ or $R = \Theta(\log N)$ and $b = 0$ or $d + \tau = \Theta(\log N)$. The theoretical results above suggest $\alpha = \Theta\{(N\log N)^{-1}\}$. This result agrees with the value suggested by the empirical procedure in §2·3, which yields $\alpha_{\min} = \alpha_{\max} = \Theta\{(N\log N)^{-1}\}$ when $R = \Theta(\log N)$ based on the population eigenvalues.

8. PROOF OF COROLLARY 2

Perfect clustering requires that $|\mathcal{M}| < 1$. Based on the bound in Theorem 3, this corresponds to $\delta\{c_0 KP\log(8N/\epsilon)\}^{1/2} < \lambda_K$. Under the same simplifying assumptions as above, this becomes

$$\begin{aligned} c'\{\alpha NR^{1/2} + (d+\tau)^{-1/2}\}(N\log N)^{1/2} &< \alpha NR + \Theta(1), \\ c''\alpha NR^{1/2}(N\log N)^{1/2} &< \alpha NR, \\ R &\geq \Theta(N\log N). \end{aligned}$$

9. PROOF OF THEOREM 4

This proof uses Fano's inequality to derive the lower bound following an approach similar to Chaudhuri et al. (2012). Let G_S be a partition given by a specific S, the set of all nodes in the first block, and let F be the family of all such partitions. Fano's inequality states

$$\sup_{G_S \in F} P_{G_S}(\Psi \neq G_S) \geq 1 - \frac{\beta + \log 2}{\log r},$$

where $KL(G_S, G_{S'}) \leq \beta$, $r = |F| - 1$, and Ψ is the estimated node partition based on the observed edges and node covariates.

First, by independence the KL-divergence can be written as follows,

$$KL(G_S, G_{S'}) = \sum_{e \in E} KL(\rho_e, \rho'_e) + \sum_{v \in V} KL(\gamma_v, \gamma'_v).$$

Let ρ_e and ρ'_e be the distribution for edge e and γ_v and γ'_v be the covariate distribution for node v in G_S and $G_{S'}$, respectively. Recall $B_{1,1} \geq B_{2,2} \geq B_{1,2}$ and let $b_i \in \{B_{1,1}, B_{2,2}, B_{1,2}\}$. For a single edge when $\rho_e \neq \rho'_e$,

$$\begin{aligned}
KL(\rho_e, \rho'_e) &\in \{b_i \log \frac{b_i}{b_j} + (1 - b_i) \log \frac{1 - b_i}{1 - b_j}\} \\
&\leq B_{1,1} \log \frac{B_{1,1}}{B_{1,2}} + (1 - B_{1,1}) \log \frac{1 - B_{1,1}}{1 - B_{1,2}} + B_{1,2} \log \frac{B_{1,2}}{B_{1,1}} + (1 - B_{1,2}) \log \frac{1 - B_{1,2}}{1 - B_{1,1}} \\
&= (B_{1,1} - B_{1,2}) \log \left\{ 1 + \frac{B_{1,1} - B_{1,2}}{B_{1,2}(1 - B_{1,1})} \right\} \\
&\leq \frac{(B_{1,1} - B_{1,2})^2}{B_{1,2}(1 - B_{1,1})}.
\end{aligned}$$

Now find the KL-divergence of the covariates on a single node. For $\gamma_v \neq \gamma'_v$,

$$KL(\gamma_v, \gamma'_v) = \sum_{j}^{R} KL(\gamma_{v_j}, \gamma'_{v_j}) \equiv \Gamma.$$

For the case of Bernoulli random variables where the jth covariate has probability $M_{1,j}$ in block 1 and $M_{2,j}$ in block 2, this is

$$\begin{aligned}
KL(\gamma_{v_j}, \gamma'_{v_j}) &= \begin{cases} M_{1,j} \log \frac{M_{1,j}}{M_{2,j}} + (1 - M_{1,j}) \log \frac{1 - M_{1,j}}{1 - M_{2,j}}, & v \in \text{block } 1 \\ M_{2,j} \log \frac{M_{2,j}}{M_{1,j}} + (1 - M_{2,j}) \log \frac{1 - M_{2,j}}{1 - M_{1,j}}, & v' \in \text{block } 1 \end{cases} \\
&\leq (M_{1,j} - M_{2,j}) \log \frac{M_{1,j}(1 - M_{2,j})}{M_{2,j}(1 - M_{1,j})}.
\end{aligned}$$

Therefore, the KL-divergence is bounded by

$$KL(G_S, G_{S'}) \leq \binom{N}{2} \frac{(B_{1,1} - B_{1,2})^2}{B_{1,2}(1 - B_{1,1})} + N\Gamma \leq \frac{N^2}{2} \frac{(B_{1,1} - B_{1,2})^2}{B_{1,2}(1 - B_{1,1})} + N\Gamma.$$

The number of partitions can be bounded as follows,

$$\begin{aligned}
|F| = \frac{1}{2} \binom{N}{N/2} &= \frac{N!}{2\{(N/2)!\}^2} \\
&\geq \frac{(2\pi N)^{1/2} (N/e)^N}{[e(N/2)^{1/2} \{N/(2e)\}^{N/2}]^2} \\
&\geq \frac{2^{N-2.1}}{(N/2)^{1/2}},
\end{aligned}$$

where the first inequality uses $(2\pi N)^{1/2}(N/e)^N \leq N! \leq eN^{1/2}(N/e)^N$. Now the $\log$ term is bounded by

$$\begin{aligned}\log(|F|-1) &\geq \log\left\{\frac{2^{N-2.1}}{(N/2)^{1/2}}-1\right\}\\ &\geq (N-3)\log 2 - \frac{1}{2}\log(N/2)\\ &\geq \frac{\log 2}{2}N \text{ for } N \geq 8.\end{aligned}$$

Thus, by Fano's inequality, in order to correctly determine the block assignments with probability at least $1-\epsilon$ requires

$$\epsilon \geq 1 - \frac{N^2(B_{1,1}-B_{1,2})^2/\{2B_{1,2}^2(1-B_{1,1})^2\} + N\Gamma + \log 2}{(N\log 2)/2},$$

$$B_{1,1}-B_{1,2} \geq B_{1,2}(1-B_{1,1})[\frac{2}{N}\left\{\frac{\log 2}{2}(1-\epsilon) - \Gamma - \frac{\log 2}{N}\right\}]^{1/2}.$$

Fix $B_{1,1}$ and let $\Delta = B_{1,1}-B_{1,2}$, then rewrite this bound as

$$\Delta \geq \frac{B_{1,1}(1-B_{1,1})}{\left[\frac{2}{N}\left\{\frac{\log 2}{2}(1-\epsilon)-\Gamma-\frac{\log 2}{N}\right\}\right]^{-1/2} + (1-B_{1,1})}.$$

10. COMPARISON OF THE GENERAL LOWER BOUND TO THEOREM 3

First, simplify the general lower bound given in Theorem 4 to make the comparison with Theorem 3 easier.

$$\begin{aligned}\Delta &\geq \frac{B_{1,1}(1-B_{1,1})}{\left[\frac{2}{N}\left\{\frac{\log 2}{2}(1-\epsilon)-\mathcal{K}-\frac{\log 2}{N}\right\}\right]^{-1/2} + (1-B_{1,1})}\\ &\geq \frac{B_{1,1}(1-B_{1,1})}{3/2\left[\frac{2}{N}\left\{\frac{\log 2}{2}(1-\epsilon)-\mathcal{K}-\frac{\log 2}{N}\right\}\right]^{-1/2}}\\ &\geq B_{1,1}(1-B_{1,1})\left(\frac{2}{3}\right)\left[\frac{2}{N}\left\{\frac{\log 2}{2}(1-\epsilon)-\mathcal{K}-\frac{\log 2}{8}\right\}\right]^{1/2}\\ &\geq \frac{c_4}{N^{1/2}}.\end{aligned}$$

According to Theorem 3 to achieve perfect clustering with probability $1-\epsilon$, requires $\delta\{c_0 KP\log(8N/\epsilon)\}^{1/2} < \lambda_K$. As shown in §8, this requires $R \geq \Theta(N\log N)$.

REFERENCES

BRAND, M. (2006). Fast low-rank modifications of the thin singular value decomposition. *Linear algebra and its applications* **415**, 20–30.

CHAUDHURI, K., CHUNG, F. & TSIATAS, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. *Journal of Machine Learning Research* **2012**, 1–23.

DAVIS, C. & KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. *SIAM Journal on Numerical Analysis* **7**, 1–46.

MCSHERRY, F. (2001). Spectral partitioning of random graphs. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science, 2001.*, M. Naor, ed. Los Alamitos, CA: IEEE, pp. 529–537.

QIN, T. & ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems 26*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani & K. Weinberger, eds. Red Hook, NY: Curran Associates, Inc., pp. 3120–3128.

ROHE, K., CHATTERJEE, S. & YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics* **39**, 1878–1915.

TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics* **12**, 389–434.