Supplementary Material for "ORTHOGONAL SYMMETRIC NON-NEGATIVE MATRIX FACTORIZATION UNDER THE STOCHASTIC BLOCK MODEL"

Subhadeep Paul and Yuguo Chen

The Ohio State University and University of Illinois at Urbana-Champaign

1. Proofs of Theoretical Results

Proof of Propositions 1 and 2

Proof. We first prove Proposition 1. If there are two non-zero elements in a row of H, say $H_{ik}, H_{il} > 0$, then their product would be a positive quantity. However since the columns of H are orthonormal, $\sum_i H_{ik}H_{il} = 0$. This would require the product $H_{i'k}H_{i'l}$ to be negative for some other i'. However, this is not possible since all the elements of H are non-negative as well.

Now we prove Proposition 2. Suppose the OSNTF as defined in Equation (2.2) is not unique and there is another order K factorization of A as $A = H'S'H'^T$. Then $H' = HQ$ and $S' = R^T SR$ where $QR^T = I$. Moreover, if $rank(A)$ is K, then both H and H' span the same subspace and must be related through an orthogonal change of basis matrix. Consequently, this is the only source of non-uniqueness. However for OSNTF even this ambiguity of an orthogonal matrix is not possible due to the orthogonality and non-negativity constraints except for permutation matrices. If HQ is a solution, then HQ must have orthonormal columns, i.e., $(HQ)^T HQ = I$ which implies $Q^T Q = I$. However, except $Q = I$ or a permutation matrix, at least one element of Q must be negative in order for it to be an orthogonal matrix (Ding et al., 2006). However, if an element of Q, say Q_{kl}, is negative, then $(HQ)_{il} = \sum_k H_{ik} Q_{kl} < 0$ for all rows i of H such that the only non-zero element in the row is in the kth place (note that such a row always exists, since no column of the rank K matrix H can be all 0's). This will make HQ contain at least one negative element, which violates the non-negativity constraint. Hence the factorization is unique up to permutations.

□

Upper bound on mis-clustering rate

Recovery in noiseless case

The next lemma shows that the procedure OSNTF can recover the class assignments perfectly from the population adjacency matrix or the Laplacian matrix generated by the stochastic block model. Hence even though for any given matrix both proving the existence and evaluation of *exact* OSNTF is NP hard, if we know that the matrix is formed according to the stochastic block model, the factorization can recover true class assignments.

Lemma 1. *Define the matrix $Q = (Z^T Z) \in \mathcal{R}^{K\times K}$, which is a diagonal matrix with strictly positive entries. The solutions to the OSNTF of $\mathcal{A}$ and $\mathcal{L}_\tau$ are $[\bar{H} = ZQ^{-1/2}, \bar{S} = Q^{1/2} B Q^{1/2}]$ and $[\bar{H}_L = ZQ^{-1/2}, \bar{S}_L = Q^{1/2} B_{L,\tau} Q^{1/2}]$, unique up to a permutation matrix P respectively. Moreover,*

$$\bar{H}_i = \bar{H}_j \iff Z_i = Z_j,$$

where $\bar{H}_i$ and Z_i are ith rows of $\bar{H}$ and Z respectively. Equivalently,

$$\arg\max_k \bar{H}_{ik} = \arg\max_k \bar{H}_{jk} \iff \arg\max_k Z_{ik} = \arg\max_k Z_{jk}.$$

The previous lemma shows that OSNTF of rank K applied to the population adjacency or the Laplacian matrix of an SBM obtains factors $[\bar{H}, \bar{S}]$ such that any two rows of $\bar{H}$ are equal if and only if the corresponding rows are equal in Z. Now assigning rows to communities on the basis of the largest entry in $\bar{H}$ as in Equation (2.4) effectively means doing the same on rows of Z, which by definition will result into correct community assignments. However due to the ambiguity in terms of a permutation matrix P, the community labels can be identified only up to a permutation.

We now prove a parallel result on recovery of class assignments from the population adjacency and Laplacian matrices of DCSBM.

Lemma 2. *Define the matrices* $Q = (Z^T\Theta^2 Z)$, $Q_L = (Z^T\Theta Z) \in \mathcal{R}^{K\times K}$, *which are diagonal matrices with strictly positive entries. The solutions of OSNTF of* $\mathcal{A}$ *and* $\mathcal{L}_\tau$ *are* $[\bar{H} = \Theta Z Q^{-1/2},\ \bar{S} = Q^{1/2}B'Q^{1/2}]$ *and* $[\bar{H}_L = \Theta^{1/2} Z Q_L^{-1/2},\ \bar{S}_L = Q_L^{1/2} B_{L,\tau}^{'} Q_L^{1/2}]$, *unique up to a permutation matrix* P*, respectively. Moreover,*

$$\arg\max_k \bar{H}_{ik} = \arg\max_k \bar{H}_{jk} \iff \arg\max_k Z_{ik} = \arg\max_k Z_{jk}.$$

Note for the DCSBM, the $\bar{H}$ OSNTF extracts does not have the same row for all nodes in the same community. However, the community detected from $\bar{H}$ by finding the location of the maximum value in the row is the same for all nodes in the same community. Therefore, the OSNTF still accurately estimates the community labels.

Uniform convergence of objective function

Although OSNTF can perfectly recover Z from the population adjacency matrix $\mathcal{A}$ and the population Laplacian matrix $\mathcal{L}$, in practice we do not observe $\mathcal{A}$ or $\mathcal{L}$. Instead we observe the sample version (or perturbed version) of $\mathcal{A}$, the sample adjacency matrix A.

The regularized sample adjacency matrix A_τ and sample Laplacian matrix L_τ may not have exact OSNTFs. In that case, let the optimization problem in (2.3) or equivalently in (2.6), obtain a solution $[\hat{H}, \hat{S}]$ as OSNTF of A_τ. The matrix approximating A_τ is then $\hat{A} = \hat{H}\hat{S}\hat{H}^T$ and we assign the nodes to the communities using the matrix $\hat{H}$.

We denote the objective function in the optimization problem of (2.6) as $F(A_\tau, H) = \|H^T A_\tau H\|_F$. This is a function of the regularized adjacency matrix A_τ and the factor matrix H. We can define a corresponding "population" version of this objective function with the population adjacency matrix as $F(\mathcal{A}, H) = \|H^T \mathcal{A} H\|_F$. The corresponding observed and population versions for the Laplacian matrix are defined by $F(L_\tau, H) = \|H^T L_\tau H\|_F$ and $F(\mathcal{L}, H) = \|H^T \mathcal{L}_\tau H\|_F$, respectively. The next lemma, which is an intermediate result, shows two uniform convergences. We show that for any $H \in \mathcal{H}_+^{N\times K}$, the difference between $F(A_\tau, H)$ and $F(\mathcal{A}, H)$ and that between $F(L_\tau, H)$ and $F(\mathcal{L}_\tau, H)$ are bounded in high probability.

Lemma 3. *For any $H \in \mathcal{H}_+^{N\times K}$, there exists a constant $c_1(r_1) > 0$, such that we have with probability at least $1 - n^{-r_1}$*

$$|F(A_\tau, H) - F(\mathcal{A}, H)| \le c_1 K \Delta^{3/2}, \tag{11}$$

and there exists a constant $c_2(r_2) > 0$, such that we have with probability at least $1 - o(1)$,

$$|F(L_\tau, H) - F(\mathcal{L}_\tau, H)| \le \frac{c_2 K}{\sqrt{\Delta}}. \tag{12}$$

Characterizing mis-clustering lemmas

Although OSNTF can perfectly recover Z from $\mathcal{A}$, in practice we obtain the matrix $\hat{H} \in \mathcal{H}_+^{N\times K}$ from the observed adjacency matrix A instead of obtaining $\bar{H}$. Consequently, community assignment using the largest entry in each row of $\hat{H}$ as in Equation (2.4) will lead to some error. We quantify the error through a measure called mis-clustering rate which, given a ground truth community assignment and a candidate community assignment, computes the proportion of nodes for which the assignments do not agree. Let $\bar{e}$ denote the ground truth and $\hat{e}$ denote a candidate assignment. Then we define the mis-clustering rate $r = \frac{1}{N} \inf_{\Pi} d_H(\bar{e}, \Pi(\hat{e}))$, where $\Pi(\cdot)$ is a permutation of the labels and $d_H(\cdot, \cdot)$ is the Hamming distance between two vectors.

The next result relates the error with the difference of the matrices $\hat{H}$ and $\bar{H}$ for SBM.

Lemma 4. *Let Z be the true community assignment matrix for a network generated from the stochastic block model and $Q = Z^T Z$. Let $(\hat{H}, \hat{S})$ be the factorization of the adjacency matrix as in (2.3). Then any mis-clustered node i must satisfy*

$$\|\hat{H}_i - \bar{H}_i P\| > \frac{1}{\sqrt{N_{max}}}, \tag{13}$$

where $\hat{H}_i$ and $\bar{H}_i$ denote the ith row of the matrices $\hat{H}$ and $\bar{H}$, respectively, P is a permutation matrix, and $N_{max} = \max_{k \in \{1,\dots,K\}} Q_{kk}$, i.e., the population of the largest block. This is also a necessary condition for mis-clustering node i in OSNTF of the Laplacian matrix.

For DCSBM we again prove a lemma connecting the event of mis-clustering with the difference between matrices $\hat{H}$ and $\bar{H}$, and matrices $\hat{H}_L$ and $\bar{H}_L$ for A and L respectively .

Lemma 5. *For a network generated from the DCSBM with parameter (Θ, Z, B) as in Equation (3.8), let $(\hat{H}, \hat{S})$ be the factorization of the adjacency matrix as in (2.2). Then a necessary condition for any node i to be mis-clustered is*

$$\|\hat{H}_i - \bar{H}_i P\| \geq m, \tag{14}$$

where $m = \min_{i \in \{1,\dots,N\}} \theta_i / \sqrt{(Z^T \Theta^2 Z)_{kk}}$ with k being the community to which the node i truly belongs. The corresponding necessary condition for the OSNTF in Laplacian matrix is

$$\|\hat{H}_{L,i} - \bar{H}_{L,i} P\| \geq m', \tag{15}$$

with $m' = \min_{i\in\{1,\ldots,N\}} \sqrt{\theta_i/(Z^T\Theta Z)_{kk}}$.

Proofs of the lemmas

Proof of Lemma 1

Proof. We have by definition of the stochastic block model,

$$\mathcal{A} = ZBZ^T, \quad Z^TZ = Q^{K\times K}, \quad det(B) \neq 0,$$

where Q is a diagonal matrix whose diagonal elements $\{Q_{11},\ldots,Q_{KK}\}$ are the population of the different blocks. Clearly an OSNTF of order K applied to the matrix $\mathcal{A}$ will not yield the matrices Z and B, since $Z^TZ \neq I$. However, notice that

$$\mathcal{A} = ZBZ^T = Z(Z^TZ)^{-1/2}(Z^TZ)^{1/2}B(Z^TZ)^{1/2}(Z^TZ)^{-1/2}Z^T = \bar{H}\bar{S}\bar{H}^T, \tag{16}$$

where $\bar{H} = Z(Z^TZ)^{-1/2} = ZQ^{-1/2}$ and $\bar{S} = (Z^TZ)^{1/2}B(Z^TZ)^{1/2} = Q^{1/2}BQ^{1/2}$. Since we assume all the communities in the stochastic block model have at least one member, all the elements of the diagonal matrix Q are strictly positive quantities. Hence both the square root matrix $Q^{1/2}$ and its inverse exist and are well defined. Clearly, $\bar{H}^T\bar{H} = I$ and $\bar{H},\bar{S} \geq 0$. Hence, $[\bar{H},\bar{S}]$ is an OSNTF of rank K for $\mathcal{A}$. Any other OSNTF of rank K for the matrix $\mathcal{A}$ is unique up to a permutation matrix P by Proposition 2. Therefore $\bar{H}_i = \bar{H}_j \implies Z_i = Z_j$.

For the result on $\mathcal{L}$, we have,

$$\mathcal{L} = ZB_LZ^T = ZQ^{-1/2}Q^{1/2}B_LQ^{1/2}Q^{-1/2}Z^T. \tag{17}$$

Hence, following the preceding argument, an OSNTF of rank K applied to the matrix $\mathcal{L}$ will recover the factor matrices as $\bar{H}_L = ZQ^{-1/2}$ and $\bar{S}_L = Q^{1/2}B_LQ^{1/2}$ unique up to a permutation matrix P. Since $Q^{1/2}$ and $Q^{-1/2}$ exist, $Z_iQ^{-1/2} = Z_jQ^{-1/2} \iff Z_i = Z_j$ in both cases.

For $\mathcal{L}_\tau$ we can present the same arguments by noting that $\mathcal{L}_\tau$ under SBM can be written as $\mathcal{L}_\tau = Z(B_L + \frac{\Delta}{n})Z^T$.

□

Proof of Lemma 2

Proof. The population adjacency matrix of the DCSBM, as in Equation (3.8), is

$$\begin{aligned}\mathcal{A} &= \Theta Z B Z^T \Theta \\ &= \Theta Z (Z^T\Theta^2 Z)^{-1/2}(Z^T\Theta^2 Z)^{1/2} B (Z^T\Theta^2 Z)^{1/2}(Z^T\Theta^2 Z)^{-1/2} Z^T \Theta \\ &= \bar{H}\bar{S}\bar{H}^T, \end{aligned} \tag{18}$$

where $\bar{H} = \Theta Z(Z^T\Theta^2 Z)^{-1/2} = \Theta Z Q^{-1/2}$ and $\bar{S} = (Z^T\Theta^2 Z)^{1/2} B (Z^T\Theta^2 Z)^{1/2} = Q^{1/2} B Q^{1/2}$. Note that the matrix $Q = (Z^T\Theta^2 Z) = (\Theta Z)^T(\Theta Z) \in R^{K\times K}$, is a diagonal matrix. Clearly all the elements are strictly positive and hence the matrix admits both a square root and an inverse. We compute

$$\bar{H}^T\bar{H} = (Z^T\Theta^2 Z)^{-1/2}(Z^T\Theta^2 Z)(Z^T\Theta^2 Z)^{-1/2} = I,$$

and $\bar{H}, \bar{S} \geq 0$. Hence, $[\bar{H}, \bar{S}]$ is an OSNTF of rank K for $\mathcal{A}$ under DCSBM. Any other OSNTF of rank K for the matrix $\mathcal{A}$ is unique up to a permutation matrix P by Proposition 2.

Since both $Q^{1/2}$ and $Q^{-1/2}$ exist, we have $Z_i Q^{-1/2} = Z_j Q^{-1/2}$ if and only if $Z_i = Z_j$. Moreover, since Z_i contains only one non-zero element, say at position k, and Q is a diagonal matrix, $(ZQ^{-1/2})_i$ also has only one non-zero element, whose position within the row is also k. Now,

$$\arg\max_k \bar{H}_{ik} = \arg\max_k \theta_i (ZQ^{-1/2})_{ik} = \arg\max_k (ZQ^{-1/2})_{ik}$$

Hence, nodes i and j will be assigned to the same community if and only if $Z_i = Z_j$.

Similarly for $\mathcal{L}$, we have

$$\begin{aligned}
\mathcal{L} &= \Theta^{1/2} Z B_L Z^T \Theta^{1/2} \\
&= \Theta^{1/2} Z (Z^T \Theta Z)^{-1/2} (Z^T \Theta Z)^{1/2} B_L (Z^T \Theta Z)^{1/2} (Z^T \Theta Z)^{-1/2} Z^T \Theta^{1/2} \\
&= \bar{H}_L \bar{S}_L \bar{H}_L^T, \qquad (19)
\end{aligned}$$

where $\bar{H}_L = \Theta^{1/2} Z (Z^T \Theta Z)^{-1/2} = \Theta^{1/2} Z Q_L^{-1/2}$ and $\bar{S}_L = (Z^T \Theta Z)^{1/2} B_L (Z^T \Theta Z)^{1/2} = Q_L^{1/2} B_L Q_L^{1/2}$. We note that the matrix $Q_L = Z^T \Theta Z \in R^{K \times K}$ is also a diagonal matrix with strictly positive diagonal entries and hence both square root and inverse are well defined. Since $\bar{H}_L^T \bar{H}_L = I$ and $\bar{H}_L, \bar{S}_L \geq 0$, $[\bar{H}_L, \bar{S}_L]$ is an OSNTF of rank K for the matrix $\mathcal{L}$. As before, this is unique up to a permutation matrix P.

The proof for the second part is identical to the previous case with $\mathcal{A}$. □

Proof of Lemma 3

Proof. We have for any $H \in \mathcal{H}_+^{N \times K}$,

$$\begin{aligned}
&|F(A_\tau, H) - F(\mathcal{A}, H)| \\
&= |\|H^T A_\tau H\|_F^2 - \|H^T \mathcal{A} H\|_F^2| \\
&= |(\|H^T A_\tau H\|_F - \|H^T \mathcal{A} H\|_F)^2 + 2(\|H^T A_\tau H\|_F - \|H^T \mathcal{A} H\|_F)(\|H^T \mathcal{A} H\|_F)| \\
&\leq (\|H^T A_\tau H\|_F - \|H^T \mathcal{A} H\|_F)^2 + 2\|H^T \mathcal{A} H\|_F |(\|H^T A_\tau H\|_F - \|H^T \mathcal{A} H\|_F)|,
\end{aligned}$$

where the equality in the third line follows because $(a^2 - b^2) = (a - b)^2 + 2b(a - b)$. Next, we bound the two terms in the last line above separately. First, we use the following result from Theorem 2.1 in Le et al. (2017). For any $C' > 0$, there exists a C dependent on C' such that

$$\|A_\tau - \mathcal{A}\|_2 \leq C\sqrt{\Delta},$$

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

For the first term we have

$$\begin{aligned}(\|H^T A_\tau H\|_F - \|H^T \mathcal{A} H\|_F)^2 &\leq \|H^T(A_\tau - \mathcal{A})H\|_F^2 \\ &\leq K\|H^T(A_\tau - \mathcal{A})H\|_2^2 \\ &\leq K\|H^T\|_2^2\|H\|_2^2\|(A_\tau - \mathcal{A})\|_2^2 \\ &\leq K\|(A_\tau - \mathcal{A})\|_2^2 \leq KC_1\Delta.\end{aligned}$$

with probability at least $1 - n^{-C'}$. The second line follows due to the fact that $(H^T A_\tau H - H^T \mathcal{A} H)$ is a $K \times K$ matrix and the equivalence of norm relation, $\|X\|_F \leq \sqrt{\text{rank}(X)}\|X\|_2$. The third line is due to the property of spectral norm that $\|ABC\|_2 \leq \|A\|_2\|B\|_2\|C\|_2$, while the fourth line follows from Theorem 2.1 of Le et al. (2017) as mentioned above, and the fact that $\|H\|_2^2 = \lambda_{\max}(H^T H) = \lambda_{\max}(I_K) = 1$.

Now the second term can be bounded with probability at least $1 - n^{-C_2'}$ as follows:

$$\begin{aligned}&2\|H^T \mathcal{A} H\|_F |(\|H^T A_\tau H\|_F - \|H^T \mathcal{A} H\|_F)| \\ &\leq 2\sqrt{K}\|H^T \mathcal{A} H\|_2 |(\|H^T A_\tau H - H^T \mathcal{A} H\|_F)| \\ &\leq 2C_2\sqrt{K}\Delta\sqrt{K}\sqrt{\Delta} = C_3 K \Delta^{3/2},\end{aligned}$$

since $\lambda_{\max}(\mathcal{A}) \leq \Delta$. Therefore combing the two terms we have

$$|F(A_\tau, H) - F(\mathcal{A}, H)| \leq C_4 K \Delta^{3/2}$$

with probability at least $1 - n^{-C_3'}$ for some constant C_4 which depends on C_3'.

Similarly, for the objective function on the regularized Laplacian matrix, we have the following result. From Theorem 1.2 in Le et al. (2017) with $\tau = 2\Delta$, we have the result

$$\|L_\tau - \mathcal{L}_\tau\|_2 \leq \frac{C}{\sqrt{\Delta}}$$

with probability at least $1 - o(1)$ for some constant C. Then we have for any $H \in \mathcal{H}_+^{N\times K}$,

$$|F(L_\tau, H) - F(\mathcal{L}_\tau, H)| \le KC_1\frac{1}{\Delta} + KC_2\sqrt{\frac{1}{\Delta}} \le \frac{KC_3}{\sqrt{\Delta}}$$

with probability $1 - o(1)$, since $\lambda_{\max}(\mathcal{L}) \le 1$. □

Proof of Lemma 4

Proof. Since $\hat{H} \in \mathcal{H}_+^{N\times K}$, by Proposition 1, each row of $\hat{H}$ has at most one non-zero element. If $\bar{H}_{ik} = (ZQ^{-1/2}P)_{ik} > 0$, then a correct assignment for row i would require $\hat{H}_{ik} > 0$. This implies if node i is incorrectly assigned, then

$$\begin{aligned}\|\hat{H}_i - \bar{H}_i\|^2 &= \|\hat{H}_i - Z_iQ^{-1/2}P\|^2 = \|\hat{H}_i\|^2 + \|Z_iQ^{-1/2}P\|^2 \\ &\ge \|Z_iQ^{-1/2}P\|^2 = \frac{1}{Q_{kk}} \ge \frac{1}{N_{max}}.\end{aligned}$$

Hence, every mis-clustered node i must have $\|\hat{H}_i - Z_iQ^{-1/2}P\|$ at least as large as $\frac{1}{\sqrt{N_{max}}}$. The matrix $\bar{H}_L = ZQ^{-1/2}$ is the same for OSNTF in Laplacian matrix as it is for OSNTF in adjacency matrix, and hence the necessary condition for mis-clustering is also $\|\hat{H}_{L,i} - \bar{H}_{L,i}P\| \ge \frac{1}{N_{max}}$. □

Proof of Theorem 1

Proof. Using Lemma 4, the mis-clustering rate r_A has the following relationship:

$$\|\hat{H} - \bar{H}P\|_F^2 = \sum_i \|\hat{H}_i - \bar{H}_iP\|^2 \ge \sum_{i:\ i \text{ is mis-clustered}} \|\hat{H}_i - \bar{H}_iP\|^2 \ge \frac{Nr_A}{N_{\max}}.$$

Now let $S_1 = \hat{H}^T\mathcal{A}\hat{H}$ and $\mathcal{A}_1 = \hat{H}S_1\hat{H}^T$. Then $F(\mathcal{A}, \bar{H}) = \|\bar{H}^T\mathcal{A}\bar{H}\|_F^2 = \|\bar{S}\|_F^2$ and $F(\mathcal{A}, \hat{H}) = \|\hat{H}^T\mathcal{A}\hat{H}\|_F^2 = \|S_1\|_F^2$. Moreover, $[\hat{H}, S_1]$ is an exact OSNTF of the matrix $\mathcal{A}_1$.

From the discussion in Section 2, the columns of $\bar{H}$ and $\hat{H}$ span reducing subspaces of $\mathcal{A}$ and $\mathcal{A}_1$ respectively. We can then look at the matrix $\mathcal{A}$ as a perturbed version of the matrix

$\mathcal{A}_1$ and use the Davis-Kahan Perturbation Theorem (Davis and Kahan, 1970) to relate the difference between the subspaces $\mathcal{R}(\hat{H})$ and $\mathcal{R}(\bar{H})$ with the difference between $\mathcal{A}_1$ and $\mathcal{A}$. In the next proposition we first reproduce the perturbation theorem mentioned in Theorem 3.4, Chapter 5 of Stewart and Sun (1990) in terms of canonical angles between subspaces. Note that for any matrix A, $\Lambda(A)$ denotes the set of its eigenvalues. For two subspaces $\mathcal{E}$ and $\mathcal{F}$, the matrix $\Theta(\mathcal{E}, \mathcal{F})$ is a diagonal matrix that contains the canonical angles between the subspaces in the diagonal. See Stewart and Sun (1990) and Vu and Lei (2013) for more details on canonical angles. We use $\sin\Theta(\mathcal{E}, \mathcal{F})$ to denote the matrix that applies sine on every element of $\Theta(\mathcal{E}, \mathcal{F})$.

Proposition 1. *(Stewart and Sun, 1990) Let the columns of $H_1^{N\times K}$ span a reducing subspace of the matrix $\mathcal{B}$, and let the spectral resolution of $\mathcal{B}$ be*

$$\begin{pmatrix} H_1^T \\ H_2^T \end{pmatrix} \mathcal{B}(H_1, H_2) = \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix}, \tag{110}$$

where (H_1, H_2) is an orthogonal matrix with $H_1 \in \mathcal{R}^{N\times K}$, and $K_1 \in \mathcal{R}^{K\times K}$ and $K_2 \in \mathcal{R}^{(N-K)\times(N-K)}$ are real symmetric matrices. Let $X \in \mathcal{R}^{N\times K}$ be the analogous quantity of H_1 in the perturbed matrix B, i.e., X has orthonormal columns and there exists a real symmetric matrix $M \in \mathcal{R}^{K\times K}$ such that $BX = XM$. Define $R = \mathcal{B}X - XM = (\mathcal{B} - B)X$. If $\delta = \min_{\lambda_1\in\Lambda(K_2),\lambda_2\in\Lambda(M)} |\lambda_1 - \lambda_2| > 0$, then

$$\|\sin\Theta(\mathcal{R}(H_1), \mathcal{R}(X))\|_F \leq \frac{\|R\|_F}{\delta} \leq \frac{\|\mathcal{B} - B\|_F}{\delta}.$$

To use the proposition in our context, let $\mathcal{B} = \mathcal{A}_1$, $B = \mathcal{A}$, $H_1 = \hat{H}$, $X = \bar{H}$. Then we have $K_1 = S_1$ and $M = \bar{S}$. Since S_1 contains all the non-zero eigenvalues of $\mathcal{A}_1$ (Section 3.1), in this case $\Lambda(K_2)$ contains only 0's. On the other hand $\Lambda(M)$ contains all the non-zero eigenvalues of $\mathcal{A}$. Consequently, $\delta = \min_{\lambda_1\in\Lambda(K_2),\lambda_2\in\Lambda(M)} |\lambda_1 - \lambda_2| = \lambda^{\mathcal{A}}$.

By Proposition 2.2 of Vu and Lei (2013) there exists a K dimensional orthogonal matrix O such that

$$\frac{1}{2}\|\hat{H}-\bar{H}O\|_F^2 \leq \|\sin\Theta(\mathcal{R}(\hat{H}),\mathcal{R}(\bar{H}))\|_F^2 \leq \frac{\|\mathcal{A}-\mathcal{A}_1\|_F^2}{(\lambda^{\mathcal{A}})^2}. \tag{111}$$

Next note that,

$$\begin{aligned}
\|\mathcal{A}-\mathcal{A}_1\|_F^2 &= \|\mathcal{A}\|_F^2 + \|\mathcal{A}_1\|_F^2 - 2tr(\mathcal{A}\mathcal{A}_1) \\
&= \|\bar{S}\|_F^2 + \|S_1\|_F^2 - 2tr(\mathcal{A}\hat{H}\hat{H}^T\mathcal{A}\hat{H}\hat{H}^T) \\
&= \|\bar{S}\|_F^2 + \|S_1\|_F^2 - 2tr(\hat{H}^T\mathcal{A}\hat{H}\hat{H}^T\mathcal{A}\hat{H}) \\
&= \|\bar{S}\|_F^2 + \|S_1\|_F^2 - 2tr(S_1S_1) \\
&= \|\bar{S}\|_F^2 - \|S_1\|_F^2 \\
&= \|\bar{H}^T\mathcal{A}\bar{\mathcal{H}}\|_F^2 - \|\hat{H}^T\mathcal{A}\hat{\mathcal{H}}\|_F^2 \\
&= F(\mathcal{A},\bar{H}) - F(\mathcal{A},\hat{H}).
\end{aligned}$$

Also we have

$$\begin{aligned}
F(\mathcal{A},\bar{H}) - F(\mathcal{A},\hat{H}) &\leq F(\mathcal{A},\bar{H}) - F(\mathcal{A},\hat{H}) + F(A,\hat{H}) - F(A,\bar{H}) \\
&\leq |F(\mathcal{A},\bar{H}) - F(A,\bar{H})| + |F(\mathcal{A},\hat{H}) - F(A,\hat{H})| \\
&\leq 2CK\Delta^{3/2},
\end{aligned}$$

where the first inequality follows from the fact that $F(A,\hat{H}) \geq F(A,\bar{H})$, since $\hat{H}$ maximizes the function $F(A,H)$. Hence from Equation (111) we have with probability at least $1-n^{-r_1}$,

$$\frac{1}{2}\|\hat{H}-\bar{H}P\|_F^2 \leq \frac{2CK\Delta^{3/2}}{(\lambda^{\mathcal{A}})^2}.$$

Consequently,

$$r_A \leq \frac{N_{\max}c_1K\Delta^{3/2}}{N(\lambda^{\mathcal{A}})^2},$$

for some constant c_1 with probability at least $1 - n^{-r_1}$.

The result for r_L follows by repeating the same arguments. We have

$$r_L \leq \frac{c_2 N_{\max} K}{N(\lambda^{\mathcal{L}_\Delta})^2 \sqrt{\Delta}},$$

for some constant c_2 with probability at least $1 - o(1)$.

□

Proof of Lemma 5

Proof. Following the previous arguments for the case of SBM in Lemma 4, if node i is incorrectly assigned, then

$$\|\hat{H}_i - \bar{H}_i P\|^2 = \|\hat{H}_i - \theta_i Z_i Q^{-1/2} P\|^2 = \|\hat{H}_i\|^2 + \|\theta_i Z_i Q^{-1/2} P\|^2$$
$$\geq \|\theta_i Z_i Q^{-1/2} P\|^2 = \frac{\theta_i^2}{(Z^T \Theta^2 Z)_{kk}} \geq m^2.$$

For OSNTF of the Laplacian matrix, this necessary condition for mis-clustering becomes

$$\|\hat{H}_{L,i} - \bar{H}_{L,i} P\|^2 = \|\hat{H}_i - \theta_i^{1/2} Z_i Q_L^{-1/2} P\|^2 \geq \frac{\theta_i}{(Z^T \Theta Z)_{kk}} \geq (m')^2.$$

□

Proof of Theorem 2

Proof. The proof follows similar arguments as in the proof of Theorem 1. From Lemma 5, we have

$$\|\hat{H} - \bar{H} P\|_F^2 \geq N r_A m^2.$$

Further, we have from the proof of Theorem 1,

$$\|\hat{H} - \bar{H} P\|_F^2 \leq \frac{4 c_1 K \Delta^{3/2}}{(\lambda^{\mathcal{A}})^2}.$$

Then combining the two results, we have

$$r_A \leq \frac{4c_1 K \Delta^{3/2}}{Nm^2(\lambda^{\mathcal{A}})^2},$$

with probability at least $1 - n^{-r_1}$, for some constants $c_1, r_1 > 0$.

Similarly, for the case of r_L, from Lemma 5 we have $\|\hat{H}_L - \bar{H}_L P\|_F^2 \geq N r_L (m')^2$. Therefore the result follows: $r_L \leq \frac{4c_2 K}{Nm^2(\lambda^{\mathcal{L}}\Delta)^2\sqrt{\Delta}}$, with probability at least $1 - o(1)$ for some constant $c_2 > 0$.

□

2. Additional Tables and Figures

Number of simulations different methods perform the best

In Tables S1 and S2, we report the number of times out of 90 repetitions the different algorithms returned the best correct clustering rate for various scenarios.

Average density	Reg. OSNTF	OSNTF	Spectral	Reg. Spectral	SCOREplus	SBM refine	DCBM refine
0.025	52	13	0	0	11	6	8
0.027	44	29	0	0	10	5	2
0.029	48	29	0	0	5	6	2
0.031	47	35	0	0	2	4	2
0.033	56	26	0	0	1	6	1
0.036	45	34	0	0	1	6	4

Table S1: Number of cases a method returned the best correct clustering rate out of 90 simulations for different (low) average densities

Degree Heterogeneity	Reg. OSNTF	OSNTF	Spectral	Reg. Spectral	SCOREplus	DCBM refine
1.8	8	67	1	6	2	6
1.9	2	77	1	5	2	3
2.0	0	87	0	0	0	3
2.1	2	84	1	0	0	3
2.2	1	84	2	0	0	3
2.3	0	87	2	1	0	0
2.4	0	89	1	0	0	0
2.5	0	84	4	1	1	0

Table S2: Number of cases a method returned the best correct clustering rate out of 90 simulations for different extents of degree heterogeneity

Simulation for computing time

We plot the computing time taken by OSNTF and Regularized OSNTF along with other methods with increasing number of nodes N in Figure S1. We find that OSNTF has a higher computational cost compared to non-iterative methods like SCOREplus and regularized spectral clustering, and the computational cost is comparable to the iterative methods like SBM refine and DCBM refine. As N increases, the computational cost for the iterative methods, including that of OSNTF, increases quite a bit.

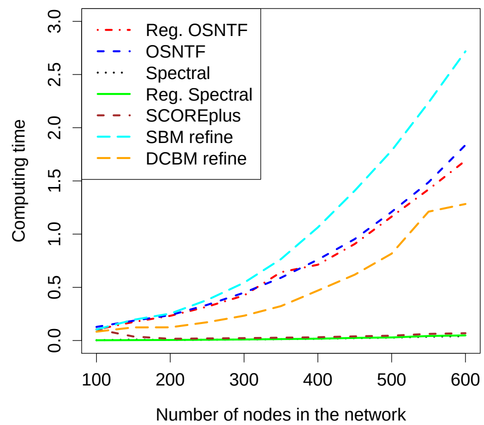

Figure S1: Comparison of computing time for various methods

References

Davis, C. and W. M. Kahan (1970). The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis 7*(1), 1–46.

Ding, C., T. Li, W. Peng, and H. Park (2006). Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 126–135.

Le, C. M., E. Levina, and R. Vershynin (2017). Concentration and regularization of random graphs. *Random Structures & Algorithms 51*(3), 538–561.

Stewart, G. W. and J.-g. Sun (1990). *Matrix Perturbation Theory*. Academic Press, Boston, MA.

Vu, V. Q. and J. Lei (2013). Minimax sparse principal subspace estimation in high dimensions. *The Annals of Statistics 41*(6), 2905–2947.