Community Detection in Sparse Networks Using the Symmetrized Laplacian Inverse Matrix (SLIM)

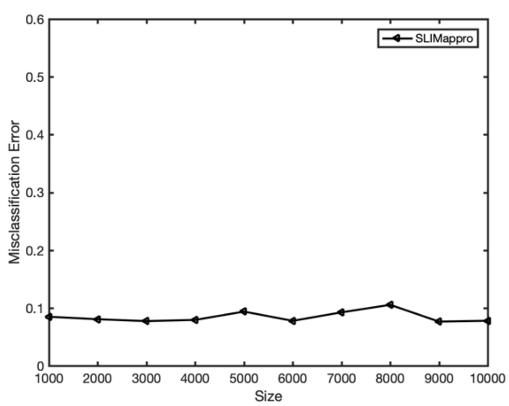
Bing-Yi Jing[1], Ting Li[2], Ningchen Ying[1] and Xianshi Yu[1]

[1]*Hong Kong University of Science and Technology and* [2]*Hong Kong Polytechnic University*

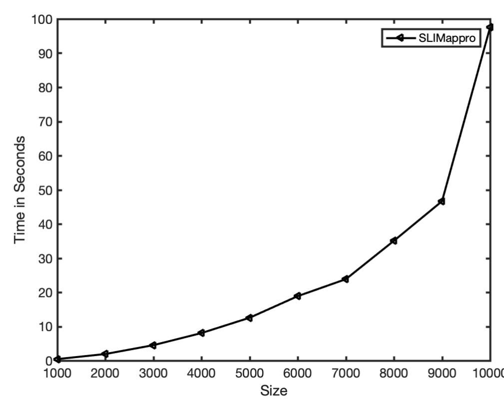
Supplementary Material

S1 Simulation with varying n.

We previously brought up an approximation approach SLIMappro to address the case where n is large and the calculation of $\hat{W}$ is time consuming. In section 4.2.3, we examined this method by networks with $n = 1200$. Here. we provide more simulation results for the SLIMappro, demonstrating its performance when n is large.



(a) ERROR RATE with varying n.



(b) Time consumption in seconds on the SLIMappro.

Figure 1: Performance of the SLIMappro with networks of varying n: Networks are simulated from the SBM with $K = 3$, $\rho = 0$, $\pi = (1/3, 1/3, 1/3)$, $\lambda = 4$ and $\beta = 0.08$ with 20 repetitions; (a) reports the average missclassification rate of SLIMappro; (b) reports the average time consumption of SLIMappro in seconds.

S2 Proof of Theorem 3.1

For the SLIM with regularization, misclassification comes from two sources: the difference between M_τ and $\hat{M}_\tau$ and the randomness of the clustering method, i.e., k-means. For convenience, we omit the subscript τ in M_τ and $\hat{M}_\tau$, and we use M_0 to specify the original one if needed.

S2.1 Misclassification Rate of K-means Algorithm

The following lemma describes the eigen-structure of M and is similar to Lemma 2.1 in Lei and Rinaldo (2015).

Lemma S2.1. Let the pair (Θ, B) parametrize the SBM with K communities, where B is of full rank. Let $\alpha < 1$, which makes $I - D^{-1}P\alpha$ invertible. Let UHU^T be the eigen-decomposition of $M - I$. Then $U = \Theta X$ where $X \in \mathbb{R}^{K\times K}$ and $\|X_{k*} - X_{l*}\| = \sqrt{{n_k}^{-1} + {n_l}^{-1}}$ for all $1 \le k < l \le K$.

Proof. Clearly $M-I$ is a block matrix of rank K. Let O be a $K \times K$ full rank matrix and

$$M - I = \Theta O \Theta^T = \Theta\Delta^{-1}\Delta O \Delta(\Theta\Delta^{-1})^T \ here\ \Delta = diag(\sqrt{n_1}, \cdots, \sqrt{n_K}).$$

Let $ZHZ^T = \Delta O \Delta$ be the eigen-decomposition of $\Delta O \Delta$. Because $M - I = UHU^T$, we have $U = \Theta\Delta^{-1}Z$ and $X = \Delta^{-1}Z$. The rows of X are perpendicular to each other and the kth row has length $\|(\Delta Z)_{k*}\| = \sqrt{1/n_k}$. In addition, the eigenvector of M is the same with $M - I$'s. □

Now, we bound the error of k-means by citing Lemma 5.3 in Lei and Rinaldo (2015).

Lemma S2.2. For $\varepsilon > 0$ and any two matrices $\hat{U}, U \in \mathbb{R}^{n\times K}$ such that $U = \Theta X$ with $\Theta \in \mathbb{F}_{n,K}, X \in \mathbb{R}^{K\times K}$, let $(\hat{\Theta}, \hat{X})$ be the $(1+\varepsilon)$-approximate solution to the k-means problem (see Kumar et al. (2004)), and $\bar{U} = \hat{\Theta}\hat{X}$. For any $\delta_k \le \min_{l\neq k} \|X_{k*} - X_{l*}\|$, define $S_k = \{i \in G_k(\Theta) : \|\bar{U}_{i*} - U_{i*}\| \ge \delta_k/2\}$, then

$$\sum_{k=1}^{K} |S_k|\delta_k^2 \le 4(4 + 2\varepsilon)\|\hat{U} - U\|_F^2. \tag{S2.1}$$

Moreover, if

$$4(4+2\varepsilon)\|\hat{U}-U\|_F^2/\delta_k^2 < n_k \quad for\ all\ k, \tag{S2.2}$$

then there exists a $K \times K$ permutation matrix J such that $\hat{\Theta}_{G*} = \Theta_{G*}J$, where $G = \bigcup_{k=1}^{K}(G_k \setminus S_k)$.

In the next lemma, similar to Lemma 5.1 in Lei and Rinaldo (2015), we bound $\|\hat{U}-U\|_F$ by $\|\hat{M}-M\|$. Here $\|F\|$ is the operator norm of matrix F.

Lemma S2.3. Assume that $M \in \mathbb{R}^{n\times n}$ is a symmetric matrix with singular value $\gamma_1 \geq \cdots \geq \gamma_n$. Let $\hat{M}$ be any symmetric matrix and $\hat{U}, U \in \mathbb{R}^{n\times K}$ be the K leading eigenvectors of $\hat{M}$ and M, respectively. Then there exists a $K\times K$ orthogonal matrix Q such that

$$\|\hat{U}-UQ\|_F \leq \frac{2\sqrt{2K}}{|\gamma_K-\gamma_{K+1}|}\|\hat{M}-M\|.$$

Proof. The proof follows the lines of Lemma 5.1 in Lei and Rinaldo (2015) using the Davis-Kahan $\sin\Theta$ theorem, and hence omitted. □

Remark S2.1. Under Condition 3.1, we can calculate the eigenvelues of M. It can be shown that the eigenvalues of $D^{-1}P$ are $\lambda_1 = 1, \lambda_2 = \cdots = \lambda_K = (a-b)(a+(K-1)b+K\tau)^{-1}, \lambda_{K+1} = \cdots = \lambda_n = 0$. And the eigenvalues of M are $\gamma_1 = \alpha(1-\alpha)^{-1}+1, \gamma_2 = \cdots = \gamma_K = \alpha\lambda_2(1-\alpha\lambda_2)^{-1}+1, \gamma_{K+1} = \cdots = \gamma_n = 1$. So we have

$$\|\hat{U}-UQ\|_F \leq \frac{2\sqrt{2K}}{\alpha}|1-\alpha+\frac{Kb}{a-b}+\frac{K\tau}{a-b}|\|\hat{M}-M\|$$

S2.2 Concentration of $\hat{M}$

We now bound $\|\hat{M}-M\|$. Following Le et al. (2017), we handle the sparsity issue by separating nodes into core points, whose degree is close to the mean, and extreme points, which have a vary large or a very small degree. The main differences from Le et al. (2017) are: the random walk Laplacian matrix is asymmetric instead of symmetric; and we control the low degree nodes by adding a constant τ and the high degree nodes by replacing their degree by cτ, here c is a sufficiently large constant.

We first bound $\|\hat{M}-M\|$ by the corresponding difference of their random walk Laplacian matrices.

Lemma S2.4. If $\alpha < 1/\sqrt{c+1}$ then

$$\|\hat{M} - M\| \leq \frac{\alpha}{(1-\sqrt{c+1}\alpha)^2}\|\hat{D}^{-1}A - D^{-1}P\|.$$

Proof. Using lemma S2.5 and the condition of α we have $\|\hat{D}^{-1}A\alpha\| \leq \sqrt{\hat{d}_{max}/\hat{d}_{min}}\alpha \leq \sqrt{(c+1)\tau/\tau}\alpha < 1$. Therefore $\|\hat{W}\| = \|(I - \hat{D}^{-1}A\alpha)^{-1}\| \leq (1 - \|\hat{D}^{-1}A\alpha\|)^{-1} \leq (1-\sqrt{c+1}\alpha)^{-1}$. So

$$\begin{aligned}\|\hat{M} - M\| &\leq \|\hat{W} - W\| \\ &= \|\hat{W}(W^{-1} - \hat{W}^{-1})W\| \\ &\leq \alpha\|\hat{W}\|\|W\|\|\hat{D}^{-1}A - D^{-1}P\| \\ &\leq \frac{\alpha}{(1-\sqrt{c+1}\alpha)^2}\|\hat{D}^{-1}A - D^{-1}P\|.\end{aligned}$$

□

Lemma S2.5. Let $L(A) = \hat{D}^{-1}A$ be the transition matrix of A, and $d_{\max} = \max[D_{ii}], d_{\min} = \min[D_{ii}]$. Then

$$\|L(A)\| \leq \sqrt{d_{\max}/d_{\min}}.$$

Proof. From the definition of $L(A)$ we have

$$\begin{aligned}\|L(A)\| &= \|D^{-\frac{1}{2}}D^{-\frac{1}{2}}AD^{-\frac{1}{2}}D^{\frac{1}{2}}\| \\ &\leq \|D^{-\frac{1}{2}}\|\|D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\|\|D^{\frac{1}{2}}\| \\ &\leq \sqrt{\frac{d_{max}}{d_{min}}}\|D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\|.\end{aligned}$$

It can be easily checked that $\|D^{-1/2}AD^{-1/2}\| = 1$. This completes the proof. □

Similar to Theorem 1.2 in Le et al. (2017), we can bound $\|\hat{D}^{-1}A - D^{-1}P\|$ as follows.

Lemma S2.6. Let A_0 be a random matrix generated from the SBM. For any $C' > 0$, there exists some $C > 0$ such that

$$\|\hat{D}^{-1}A - D^{-1}P\| \leq C\sqrt{\frac{\log d}{d}}$$

with probability at least $1 - n^{-C'}$ uniformly over $\tau \in [C_1 d, C_2 d]$ for some sufficiently large constants C_1, C_2, where $d = np_{max} + 1$ and $p_{max} = \max_{u \geq v} P_{uv}$.

Proof. First, there is a set of nodes with degrees close to their expected degree. From Lemma S2.7 we can find a set J containing all but at most n/d nodes from $[n]$ which satisfies:

$$\|(\hat{D}^{-1}A - D^{-1}P)_{JJ}\| \leq C_3(\frac{\sqrt{d\log d}(d+2\tau+\sqrt{d\log d})}{\tau^2}).$$

Now, let us deal with the residual. We consider nodes with a high degree in the original network first. By applying the SLIM with regularization, we have already changed the degree of these nodes to τ. From Lemma S2.9 we can find a set J_1 containing at most $n/(4\tau)$ nodes from $[n]$ which satisfies:

$$\|(\hat{D}^{-1}A - \hat{D}^{-1}P)_{J_1\times[n]\cup([n]-J_1)\times J_1}\| \leq \frac{C_4}{\sqrt{\tau}}.$$

In addition, it is easy to show that $J \cap J_1 = \emptyset$. Let $J_2 = [n] - J_1 - J$. We have $\max\{\hat{d}_u : u \in J_2\} < 4\tau$. Then we decompose the left nodes into two blocks $J_2 \times [n]$ and $([n] - J_2) \times J_2$. The first block has at most n/d rows, so Lemma S2.10 indicates that

$$\|L(A_\tau)_{J_2\times[n]}\| \leq \sqrt{\frac{\max\{\hat{d}_u : u \in J_2\}+\tau}{\min\{\hat{d}_u : u \in J_2\}+\tau}}\|\mathcal{L}(A_\tau)_{J_2\times[n]}\| \leq 3(\frac{2}{\sqrt{d}} + \frac{\sqrt{40r\log d}}{\sqrt{\tau}}).$$

Similarly, from Lemma S2.11, we have

$$\|L(P_\tau)_{J_2\times[n]}\| \leq \sqrt{\frac{\max\{d_u : u \in J_2\}+\tau}{\min\{d_u : u \in J_2\}+\tau}}\|\mathcal{L}(P_\tau)_{J_2\times[n]}\| \leq 3(\frac{2}{\sqrt{d}} + \frac{2}{\sqrt{\tau}}).$$

As for $([n]-J_2)\times J_2$, we can bound it in the same way. Finally we complete the proof using the triangle inequality and taking $\tau = C_5 d$ with a sufficiently large constant $C_5 > 0$. □

Lemma S2.7. For any $C' > 0$, there exists some $C > 0$ such that with probability at least $1 - n^{-C'}$, there exists a subset $J \subset [n]$ satisfying $n - |J| \leq n/d$, $\max_{v\in J} |d_v - \hat{d}_v| \leq C\sqrt{d\log d}$ and

$$\|(\hat{D}^{-1}A - D^{-1}P)_{JJ}\| \leq C(\frac{\sqrt{d\log d}(d+2\tau+\sqrt{d\log d})}{\tau^2}),$$

where $d = np_{max} + 1$.

Proof. The existance of such a subset J satisfying $n-|J| \leq n/d$ and $\max_{v\in J} |d_v - \hat{d}_v| \leq C\sqrt{d\log d}$ can be proved by Lemma 14 in Gao et al. (2017) from the beginning to inequation (85). Now let

$$A = (A_\tau)_{JJ}, \quad \hat{D} = (\hat{D}_\tau)_{JJ}, \quad P = (P_\tau)_{JJ}, \quad D = (D_\tau)_{JJ}.$$

We have $\|D^{-1}\| \leq \frac{1}{\tau}$, $\|P\| \leq d+\tau$. We also have

$$\|D^{-1} - \hat{D}^{-1}\| \leq \max_{v\in J} |\frac{1}{d_v+\tau} - \frac{1}{\hat{d}_v+\tau}| \leq \frac{C\sqrt{d\log d}}{\tau^2}.$$

Finally, we obtain

$$\begin{aligned}
\|\hat{D}^{-1}A - D^{-1}P\| &= \|\hat{D}^{-1}A - D^{-1}A + D^{-1}A - D^{-1}P\| \\
&\leq \|\hat{D}^{-1} - D^{-1}\|\|A\| + \|D^{-1}\|\|A-P\| \\
&\leq \|\hat{D}^{-1} - D^{-1}\|(\|P\| + \|A-P\|) + \|D^{-1}\|\|A-P\| \\
&\leq C(\frac{\sqrt{d\log d}(d+2\tau+\sqrt{d\log d})}{\tau^2})
\end{aligned}$$

for some constant $C > 0$. This completes the proof. □

The following result is Lemma 11 in Gao et al. (2017).

Lemma S2.8. For any $\tau > C(1+np_{max})$ with some sufficiently large $C > 0$, we have

$$|\{u \in [n] : d_u \geq \tau\}| \leq \frac{n}{\tau}$$

with probability at least $1 - e^{-C'n}$ for some constant $C' \geq 0$.

Lemma S2.9. For any $\tau > Cd$ with some sufficiently large $C > 0$, $J_1 = \{u \in [n] : \hat{d}_u > \tau\}$, there exists a positive constant C_1. We have

$$\|(\hat{D}^{-1}A - \hat{D}^{-1}P)_{J_1\times[n]\cup([n]-J_1)\times J_1}\| \leq \frac{C_1}{\sqrt{\tau}},$$

with probability at least $1 - e^{-C'n}$ for some constant $C' \geq 0$.

Proof. From Lemma S2.8, $|J_1| \leq n/\tau$ with probability at least $1 - e^{-C'n}$ for some constant $C' \geq 0$.

$$\begin{aligned}
\|(\hat{D}^{-1}A - \hat{D}^{-1}P)_{J_1\times[n]\cup([n]-J_1)\times J_1}\| &\leq \|(\hat{D}^{-1}A - \hat{D}^{-1}P)_{J_1\times[n]\cup([n]-J_1)\times J_1}\|_F \\
&\leq \sqrt{2n|J_1|max(\frac{1}{n} - \frac{1}{d_{i,j}+\tau}(P_{i,j}+\frac{\tau}{n}))} \\
&< \frac{\sqrt{2}n}{\sqrt{\tau}}(\frac{1}{n}+\frac{1}{\tau}(p_{max}+\frac{\tau}{n})) \\
&\leq \frac{C_1}{\sqrt{\tau}}
\end{aligned}$$

for some positive constant C_1. □

The following lemma is derived from Theorem 4.1 in Le et al. (2017).

Lemma S2.10. Let A_0 be a random matrix from the SBM. Then for any $r \geq 1$ the following holds with probability $1 - 2n^{-2r}$. Any sub-matrix $\mathcal{L}(A_\tau)_{\mathcal{I}\times\mathcal{J}}$ of the regularized Laplacian $\mathcal{L}(A_\tau)$ with at most n/d rows or columns satisfies

$$\|\mathcal{L}(A_\tau)_{\mathcal{I}\times\mathcal{J}}\| \leq \frac{2}{\sqrt{d}} + \frac{\sqrt{40r\log d}}{\sqrt{\tau}} \; for\ any\ \tau > 0.$$

Here $\mathcal{L}(A) = D^{-1/2}AD^{-1/2}$ is the symmetric normalized Laplacian of A.

Similarly, we can bound the Laplacian of the regularized P_τ.

Lemma S2.11. Let matrix P as assumption. Then any sub-matrix $\mathcal{L}(P_\tau)_{\mathcal{I}\times\mathcal{J}}$ of the regularized Laplacian $\mathcal{L}(P_\tau)$ with at most n/d rows or columns satisfies

$$\|\mathcal{L}(P_\tau)_{\mathcal{I}\times\mathcal{J}}\| \leq \frac{2}{\sqrt{d}} + \frac{2}{\sqrt{\tau}} \; for\ any\ \tau > 0.$$

Finally, we are ready to prove Theorem 3.1 now.

Proof. From Lemma S2.1 we have $UQ = \Theta XQ = \Theta X'$ where $\|X'_{k*} - X'_{l*}\| = \sqrt{1/n_k + 1/n_l}$. Here Q is a $K \times K$ orthogonal matrix. Then, we choose $\delta_k = \sqrt{n_k^{-1} + \max\{n_l : l \neq k\}^{-1}}$ in Lemma S2.2 so that $n_k\delta_k^2 \geq 1$ for all k. We have $L(\hat{\Theta}_\tau, \Theta) \leq \sum_{k=1}^K |S_k|(n_k^{-1} + \max\{n_l : l \neq k\}^{-1}) \leq 4(4+2\varepsilon)\|\hat{U}_\tau - UQ\|_F^2$. Then, using Lemma S2.3, we have $L(\hat{\Theta}_\tau, \Theta) \leq C_1\|\hat{M} - M\|^2(\gamma_{\tau,K} - \gamma_{\tau,K+1})^{-2}$ for some positive constant C_1. Following Lemma S2.4, we have $L(\hat{\Theta}_\tau, \Theta) \leq C\|\hat{D}^{-1}A - D^{-1}P\|(\gamma_{\tau,K} - \gamma_{\tau,K+1})^{-2}$. We obtain the final result by applying Lemma S2.6.

□

Bibliography

Gao, C., Z. Ma, A. Y. Zhang, and H. H. Zhou (2017). Achieving optimal misclassification proportion in stochastic block models. *The Journal of Machine Learning Research 18*(1), 1980–2024.

Kumar, A., Y. Sabharwal, and S. Sen (2004). A simple linear time (1+/spl epsiv/)-approximation algorithm for k-means clustering in any dimensions. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pp. 454–462. IEEE.

Le, C. M., E. Levina, and R. Vershynin (2017). Concentration and regularization of random graphs. *Random Structures and Algorithms*.

Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics 43*(1), 215–237.

List of Notation

A	Adjacency matrix A, which is an $n \times n$ 0–1 symmetric matrix.
A_τ	$A_\tau = A + \frac{\tau}{n}\mathbf{1}\mathbf{1}^{\mathbf{T}}$ and then set $A_{\tau J_1 *} = \frac{\tau}{n}$ *and* $A_{\tau * J_1} = \frac{\tau}{n}$.
B	$K \times K$ matrix with b_{ij} indicating the connecting probability between a pair of nodes from community i and j.
D	The expected degree diagonal matrix which is equal to $diag(P\mathbf{1}\mathbf{1})$.
D_τ	The expected regularized degree diagonal matrix which is equal to $diag(P_\tau \mathbf{1}\mathbf{1})$.
$F_{\mathcal{I}*}$	For a matrix F and index sets $\mathcal{I} \subseteq [n]$, $F_{\mathcal{I}*}$ is the sub-matrix of F consisting of the corresponding rows.
I_K	The $K \times K$ identity matrix.
J_1	$J_1 = \{u \in [n] : \hat{d}_u \geq \tau\}$.
K	Number of communities.
$L(F)$	For any matrix F, $L(F) = D_F^{-1}F$ which is the transition matrix of F.
$L(\hat{\Theta}, \Theta)$	The overall proportion of misclassification nodes, $L(\hat{\Theta}, \Theta) = n^{-1} \min_{J \in E_K} \|\hat{\Theta}J - \Theta\|_0$.
M	$M = \frac{1}{2}((I - D^{-1}P\alpha) + (I - D^{-1}P\alpha)^T)$.
M_τ	$M_\tau = \frac{1}{2}((I - D_\tau^{-1}P_\tau\alpha) + (I - D_\tau^{-1}P_\tau\alpha)^T)$.

P	Edge probability matrix P, with $P = \Theta B \Theta^T$.
P_τ	$P_\tau = P + \frac{\tau}{n}\mathbf{1}\mathbf{1}^{\mathbf{T}}$.
Θ	Membership matrix, $\Theta \in \mathbb{F}_{n,K}$, and $\Theta_{i,g_i} = 1$.
α	$\alpha = e^{-\gamma}$.
$\hat{D}$	The degree diagonal matrix which is equal to $diag(A\mathbf{1}\mathbf{1})$.
$\hat{D}_\tau$	The regularized degree diagonal matrix which equal to $diag(A_\tau \mathbf{1}\mathbf{1})$.
$\hat{M}_\tau$	$\hat{M}_\tau = \frac{1}{2}((I - \hat{D}_\tau^{-1} A_\tau \alpha) + (I - \hat{D}_\tau^{-1} A_\tau \alpha)^T)$.
$\hat{M}$	$\hat{M} = \frac{1}{2}((I - \hat{D}^{-1} A \alpha) + (I - \hat{D}^{-1} A \alpha)^T)$.
$\hat{\Theta}$	Estimated membership matrix, $\hat{\Theta} \in \mathbb{F}_{n,K}$.
$\hat{d}_i$	The degree of node i, which is $\hat{D}_{i,i}$.
$\hat{d}_{\tau,i}$	The regularized degree of node i, which is $\hat{D}_{\tau,i,i}$.
$\mathbb{F}_{n,K}$	The collection of all $n \times K$ matrices where each row has only one 1 and $(K-1)$ $0's$.
τ	The regularization number which is in $[C_1 d, C_2 d]$ for relatively large positive C_1 and C_2.
d	$d = np_{max} + 1$.
$diag(F)$	For any matrix F, $diag(F)$ denotes the matrix obtained by setting all off-diagonal entries of F to 0.
n	Number of nodes.
n_i	Number of nodes belonging to community i.
p_{max}	$p_{max} = \max_{u \geq v} P_{uv}$.
$\mathbf{1}_K$	The $K \times 1$ vector of 1's.