

Robust LLG Proof

Runze, Minh, Carey, Jovo, ...

November 7, 2016

1 Problem Description

1.1 Uncontaminated Model

Consider a positive semi-definite stochastic blockmodel parameterized by $B \in [0, R]^{K \times K}$, and $\rho \in (0, 1)^K$ with $\sum_{i=1}^K \rho_i = 1$. Note that K is the number of blocks, and $R > 0$ is an upper bound for the parameter. Assuming that $\text{rank}(B) = d$, then there exists $\nu \in \mathbb{R}^{K \times d}$ such that $B = \nu\nu^T$. Let τ represent the block membership of each vertex, i.e. $\tau(i) = k$ means vertex i belongs to block k . Then the probability matrix $P \in [0, R]^{n \times n}$ can be written as XX^T , where $X = [X_1, \dots, X_n]^T \in \mathbb{R}^{n \times d}$ and $X_i = \nu_{\tau_i}$.

Then we can sample m conditionally i.i.d. symmetric and hollow graphs $G^{(1)}, \dots, G^{(m)}$, such that conditioned on X , $G_{ij}^{(t)} \stackrel{\text{ind}}{\sim} \text{Exp}(P_{ij})$ for each $1 \leq t \leq m$, $1 \leq i < j \leq n$.

1.2 Contaminated Observations

Now we assume the observed edges are contaminated with probability ϵ .

That is to say, consider another positive semi-definite stochastic blockmodel parameterized by $B' \in [0, R]^{K' \times K'}$, and $\rho' \in (0, 1)^{K'}$ with $\sum_{i=1}^{K'} \rho'_i = 1$. Note that K' is the number of blocks which is not necessarily equal to K , but R is the same upper bound for the parameter. Assuming that $\text{rank}(B') = d'$, then there exists $\nu' \in \mathbb{R}^{K' \times d'}$ such that $B' = \nu'\nu'^T$. Let τ' represent the block membership of each vertex, i.e. $\tau'(i) = k$ means vertex i belongs to block k . Then the probability matrix $C \in [0, R]^{n \times n}$ can be written as YY^T , where $Y = [Y_1, \dots, Y_n]^T \in \mathbb{R}^{n \times d'}$ and $Y_i = \nu'_{\tau'_i}$.

Then we can sample m conditionally i.i.d. symmetric and hollow graphs $A^{(1)}, \dots, A^{(m)}$ such that conditioning on X and Y , $A_{ij}^{(t)} \stackrel{\text{ind}}{\sim} (1 - \epsilon)\text{Exp}(P_{ij}) + \epsilon\text{Exp}(C_{ij})$ for each $1 \leq t \leq m$, $1 \leq i < j \leq n$.

1.3 Goal

Given the contaminated observation of adjacency matrices of m graphs, i.e. $A^{(1)}, \dots, A^{(m)}$, we want to estimate the mean of the collection of uncontaminated graphs P .

2 Candidate Estimators

After observing contaminated adjacency matrices of m graphs $A^{(1)}, \dots, A^{(m)}$, we want to propose a good estimator for the mean of the collection of graphs P .

2.1 $\hat{P}^{(1)}$ based on entry-wise MLE

Under the independent edge setting, we can simplify the problem to finding an entry-wise estimate of P . And MLE is always our first choice, which exists and happen to be \bar{A} , the entry-wise mean in this case. For consistency, we define $\hat{P}^{(1)} = \bar{A}$.

2.2 $\hat{P}^{(q)}$ based on entry-wise MLqE

Since the observations are contaminated, robust estimators are preferred. A modified MLE estimator, the maximum likelihood L- q estimator [2, 6], is considered in this case. Note that there might be multiple solution to the ML q equation, we define the MLqE to be the largest solution (which is still less than MLE when the model is exponential distribution). Denote $\hat{P}^{(q)}$ as the entry-wise MLqE.

Remark: MLE is a special case of MLqE when $q = 1$. So we notate the entry-wise MLE to be $\hat{P}^{(1)}$ in consistent with entry-wise MLqE $\hat{P}^{(q)}$.

2.3 $\tilde{P}^{(1)}$ based on ASE of entry-wise MLE

By taking advantages of the graph structure, we expect a better performance after applying a rank-reduction procedure to the entry-wise MLE $\hat{P}^{(1)}$ under the SBM. So we first apply ASE to $\hat{P}^{(1)}$ to get the latent positions $\hat{X}^{(1)}$ in dimension $d^{(1)}$, and then define $\tilde{P}^{(1)} = (\hat{X}^{(1)} \hat{X}^{(1)T})_{\text{tr}}$, where each element is a projection of $\hat{X}_i^{(1)} \hat{X}_j^{(1)T}$ onto $[0, \min(\hat{P}_{ij}^{(1)}, R)]$.

2.4 $\tilde{P}^{(q)}$ based on ASE of entry-wise MLqE

Similarly, we also expect a better performance after applying a rank-reduction procedure to the entry-wise MLqE $\hat{P}^{(q)}$ under the SBM. So we first apply ASE to $\hat{P}^{(q)}$ to get the latent positions $\hat{X}^{(q)}$ in dimension $d^{(q)}$, and then define $\tilde{P}^{(q)} = (\hat{X}^{(q)} \hat{X}^{(q)T})_{\text{tr}}$, where each element is a projection of $\hat{X}_i^{(q)} \hat{X}_j^{(q)T}$ onto $[0, \min(\hat{P}_{ij}^{(q)}, R)]$.

2.5 Summary

Thus, we should choose the estimator $\tilde{P}^{(q)}$.

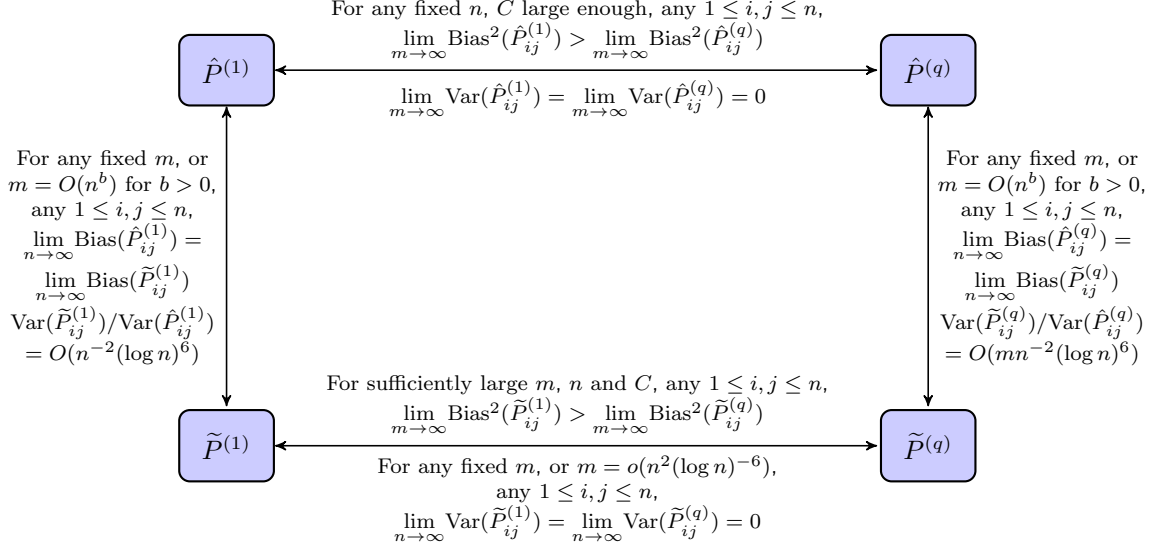


Figure 1: Relationship between four estimators.

3 Proof

3.1 $\hat{P}^{(q)}$ better than $\hat{P}^{(1)}$

Lemma 3.1 Consider the model $X_1, \dots, X_m \stackrel{iid}{\sim} \text{Exp}(P)$ with $m \geq 2$ and $E[X_1] = P$. Given any data $x = (x_1, \dots, x_m)$ such that $x_{(1)} > 0$ and not all x_i 's are the same, then no matter how the data is sampled, we have

- There exists at least one solution to the MLq equation;
- All the solutions to the MLq equation are less than the MLE.

Thus the MLqE $\hat{P}^{(q)}$, the root closest to the MLE, is well defined.

Require: Exponential distribution, MLqE

Proof: The MLE is

$$\hat{P}^{(1)}(x) = \bar{x}.$$

Consider the continuous function $g(\theta, x) = \sum_{i=1}^m e^{-\frac{(1-q)x_i}{\theta}} (x_i - \theta)$. Then the MLq equation is $g(\theta, x) = 0$.

Let $x_{(1)} \leq \dots \leq x_{(l)} \leq \bar{x} \leq x_{(l+1)} \leq \dots \leq x_{(m)}$. Define $s_i = \bar{x} - x_{(i)}$ for $1 \leq i \leq l$, and $t_i = x_{(l+i)} - \bar{x}$ for $1 \leq i \leq m-l$. Note that $\sum_{i=1}^l s_i = \sum_{i=1}^{m-l} t_i$.

Then for any $\theta \geq \bar{x}$, we have

$$\begin{aligned}
g(\theta, x) &= \sum_{i=1}^m e^{-\frac{(1-q)x_{(i)}}{\theta}} (x_{(i)} - \theta) = \sum_{i=1}^m e^{-\frac{(1-q)x_{(i)}}{\theta}} (x_{(i)} - \bar{x} + \bar{x} - \theta) \\
&= -\sum_{i=1}^l e^{-\frac{(1-q)x_{(i)}}{\theta}} s_i + \sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(i+l)}}{\theta}} t_i + \sum_{i=1}^m e^{-\frac{(1-q)x_{(i)}}{\theta}} (\bar{x} - \theta) \\
&\leq -\sum_{i=1}^l e^{-\frac{(1-q)x_{(i)}}{\theta}} s_i + \sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(i+l)}}{\theta}} t_i \\
&\leq -e^{-\frac{(1-q)x_{(l+1)}}{\theta}} \sum_{i=1}^l s_i + \sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(i+l)}}{\theta}} t_i \\
&\leq -e^{-\frac{(1-q)x_{(l+1)}}{\theta}} \sum_{i=1}^{m-l} t_i + \sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(i+l)}}{\theta}} t_i \\
&\leq -\sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(i+l)}}{\theta}} t_i + \sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(i+l)}}{\theta}} t_i \\
&= 0,
\end{aligned}$$

and equality holds if and only if all x_i 's are the same, which is excluded by the assumption. Thus $g(\theta, x) < 0$ for any $\theta \geq \bar{x}$.

Denote any solution to the MLq equation to be $\hat{P}^{(q)}(x)$, then we also know:

- $g(\hat{P}^{(q)}(x), x) = 0$;
- $\lim_{\theta \rightarrow 0^+} g(\theta, x) = 0$;
- $g(\theta, x) > 0$ when $\theta < x_{(1)}$;

Thus there exists at least one solution to the MLq equation. And all solutions to the MLq equation are between $x_{(1)}$ and \bar{x} , i.e. less than the MLE. ■

Lemma 3.2 *Consider the exponential distribution model as in Lemma 3.1 while the data is actually sampled under the contaminated model $X, X_1, \dots, X_m \stackrel{iid}{\sim} (1-\epsilon)\text{Exp}(P) + \epsilon\text{Exp}(C)$. Denote such contaminated distribution as F . Then there exists at least one solution $\theta(F)$ of the population version of MLq equation, i.e. $E_F[e^{-\frac{(1-q)X}{\theta(F)}} (X - \theta(F))] = 0$, such that $\theta(F) < E_F[\bar{X}] = (1-\epsilon)P + \epsilon C$. So we can define $\theta(F_{ij})$ to be the largest root which is less than $E_F[\bar{X}]$.*

Require: Exponential distribution, MLqE

Proof: For the MLE, i.e. \bar{X} , we have $E[\bar{X}] = (1-\epsilon)P + \epsilon C$. According to Equation (3.2) in [2], $\theta(F)$ satisfies

$$\frac{\epsilon C}{(C(1-q) + \theta)^2} - \frac{\epsilon}{C(1-q) + \theta} + \frac{(1-\epsilon)P}{(P(1-q) + \theta)^2} - \frac{(1-\epsilon)}{P(1-q) + \theta} = 0,$$

i.e.

$$\frac{\epsilon(\theta - Cq)}{(C(1-q) + \theta)^2} = \frac{(1-\epsilon)(Pq - \theta)}{(P(1-q) + \theta)^2}.$$

Define $h(\theta) = (C(1-q) + \theta)^2(1-\epsilon)(Pq - \theta) - (P(1-q) + \theta)^2\epsilon(\theta - Cq)$. Then $\lim_{\theta \rightarrow \infty} h(\theta) = -\infty$, $h(0) > 0$, and $h(Cq) < 0$. Consider q as the variable and solve the equation $h(E[\tilde{X}]) = 0$, we have three roots and one of them is $q = 1$ obviously. The other two roots are

$$\frac{(P+C)((P-C)^2\epsilon(1-\epsilon) + 2PC)}{2PC(P\epsilon + C(1-\epsilon))} \pm \sqrt{\frac{\epsilon(1-\epsilon)(C-P)^2(\epsilon(1-\epsilon)(C-P)^4 - 4P^2C^2)}{4P^2C^2(P\epsilon + C(1-\epsilon))^2}}.$$

To prove the roots are greater or equal to 1, we need to show

$$\frac{(P+C)((P-C)^2\epsilon(1-\epsilon) + 2PC)}{2PC(P\epsilon + C(1-\epsilon))} - \sqrt{\frac{\epsilon(1-\epsilon)(C-P)^2(\epsilon(1-\epsilon)(C-P)^4 - 4P^2C^2)}{4P^2C^2(P\epsilon + C(1-\epsilon))^2}} > 1.$$

For the first part,

$$\frac{(P+C)((P-C)^2\epsilon(1-\epsilon) + 2PC)}{2PC(P\epsilon + C(1-\epsilon))} > 1 + \frac{(P-C)^2\epsilon(1-\epsilon)(P+C)}{2PC(P\epsilon + C(1-\epsilon))}.$$

To prove the roots are greater or equal to 1, we just need to show

$$(P-C)^4\epsilon^2(1-\epsilon)^2(P+C)^2 \geq \epsilon^2(1-\epsilon)^2(C-P)^6.$$

Then it is sufficient to show that

$$(P+C)^2 \geq (C-P)^2,$$

which is true. Combined with the fact that when $q = 0$, $h(E[\tilde{X}]) < 0$, we have for any $0 < q < 1$, $h(E[\tilde{X}]) < 0$.

The equation $h(\theta) = 0$ is a cubic polynomial, so it has at most three real roots. Combined with the fact that $h(0) > 0$ and $h(Pq) = 0$, we have for any $0 < q < 1$, there exists at least one root of the population version of MLq equation which is less than $E[\tilde{X}] = (1-\epsilon)P + \epsilon C$. ■

Lemma 3.3 For any $a > 0$, we have

$$\sup_{\theta \in [0, R]} \left| \frac{1}{m} \sum_{i=1}^m e^{-\frac{(1-q)X_i}{\theta}} (X_i - \theta) - E_F[e^{-\frac{(1-q)X}{\theta}} (X - \theta)] \right| \xrightarrow{a.s.} 0.$$

Require: Exponential distribution, MLqE. Should be easy to extend

Proof: Define $g(x, \theta) = e^{-\frac{(1-q)x}{\theta}}(x - \theta)$ and $d(x) = e^{-\frac{(1-q)x}{R}}(x + R)$. Then $E_F[d(X)] < \infty$ and $g(x, \theta) \leq d(x)$ for all $\theta \in [0, R]$. Combined with the fact that $[0, R]$ is compact and the function $g(x, \theta)$ is continuous at each θ for all $x > 0$ and measurable function of x at each θ , we have the uniform convergence by Lemma 2.4 in [4]. ■

Lemma 3.4 $E[\hat{P}_{ij}^{(q)}] \rightarrow \theta(F_{ij})$ as $m \rightarrow \infty$, where F_{ij} is the contaminated distribution $(1-\epsilon)\text{Exp}(P_{ij}) + \epsilon\text{Exp}(C_{ij})$.

Proof: NEED PROOF HERE ■

Lemma 3.5 For any $0 < q < 1$, there exists $C_0(P_{ij}, \epsilon, q) > 0$ such that under the contaminated model with $C > C_0(P_{ij}, \epsilon, q)$,

$$\lim_{m \rightarrow \infty} |E[\hat{P}_{ij}^{(q)}] - P_{ij}| < \lim_{m \rightarrow \infty} |E[\hat{P}_{ij}^{(1)}] - P_{ij}|,$$

for $1 \leq i, j \leq n$ and $i \neq j$.

Require: Exponential distribution, MLqE

Proof: For the MLE $\hat{P}_{ij}^{(1)} = \bar{A}_{ij}$,

$$E[\hat{P}_{ij}^{(1)}] = E[\bar{A}_{ij}] = \frac{1}{m} \sum_{t=1}^m E[A_{ij}^{(t)}] = E[A_{ij}^{(1)}] = (1 - \epsilon)P_{ij} + \epsilon C_{ij}.$$

As shown in Lemma 3.2, $\theta(F)$ satisfies

$$\frac{\epsilon(\theta(F) - C_{ij}q)}{(C_{ij}(1 - q) + \theta(F))^2} = \frac{(1 - \epsilon)(P_{ij}q - \theta(F))}{(P_{ij}(1 - q) + \theta(F))^2}.$$

Thus $\theta(F) - C_{ij}q$ and $\theta(F) - P_{ij}q$ should have different signs. Combined with $C_{ij} > P_{ij}$, we have

$$qP_{ij} < \theta(F).$$

To have a smaller asymptotic bias in absolute value, combined with Lemma 3.4, we need

$$|\theta(F) - P_{ij}| < \epsilon(C_{ij} - P_{ij}).$$

Based on Lemma 3.1, we need

$$qP_{ij} > P_{ij} - \epsilon(C_{ij} - P_{ij}),$$

i.e.

$$C_{ij} > P_{ij} + \frac{(1 - q)P_{ij}}{\epsilon} = C_0(P_{ij}, \epsilon, q).$$

■

Lemma 3.6

$$\lim_{m \rightarrow \infty} \text{Var}(\hat{P}_{ij}^{(1)}) = \lim_{m \rightarrow \infty} \text{Var}(\hat{P}_{ij}^{(q)}) = 0,$$

for $1 \leq i, j \leq n$.

Proof: MLE simply follows a central limit theorem, which means the variance goes to 0 as $m \rightarrow \infty$. For MLqE, **NEED PROOF HERE** ■

3.2 $\tilde{P}^{(1)}$ better than $\hat{P}^{(1)}$

Theorem 3.7 (Matrix Bernstein: Subexponential Case). Consider a finite sequence $\{X_k\}$ of independent, random, self-adjoint matrices with dimension d . Assume that

$$E[X_k] = 0 \quad \text{and} \quad E[X_k^p] \preceq \frac{p!}{2} R^{p-2} A_k^2 \quad \text{for } p = 2, 3, 4, \dots$$

Compute the variance parameter

$$\sigma^2 := \left\| \sum_k A_k^2 \right\|.$$

Then the following chain of inequalities holds for all $t \geq 0$.

$$P \left(\lambda_{\max} \left(\sum_k X_k \right) \geq t \right) \leq d \cdot \exp \left(\frac{-t^2/2}{\sigma^2 + Rt} \right).$$

Remark: Theorem 6.2 in [9].

Theorem 3.8 Let P and C be two n -by- n symmetric matrices satisfying element-wise conditions $0 < P_{ij} \leq C_{ij} \leq R$ for some constant $R > 0$. For $0 < \epsilon < 1$, we define m symmetric and hollow matrices as

$$A^{(t)} \stackrel{iid}{\sim} (1 - \epsilon)\text{Exp}(P) + \epsilon\text{Exp}(C),$$

for $1 \leq t \leq m$. Let $\hat{P}^{(1)}$ be the element-wise MLE based on exponential distribution with m observations. Define $H_{ij}^{(1)} = E[\hat{P}_{ij}^{(1)}] = (1 - \epsilon)P_{ij} + \epsilon C_{ij}$, then for any constant $c > 0$, there exists another constant $n_0(c)$, independent of n , P , C and ϵ , such that if $n > n_0$, then for all η satisfying $n^{-c} \leq \eta \leq 1/2$,

$$P \left(\|\hat{P}^{(1)} - H^{(1)}\|_2 \leq 4R\sqrt{n \ln(n/\eta)/m} \right) \geq 1 - \eta.$$

Remark: This is the extended version of Theorem 3.1 in [5].

Require: Exponential distribution

Proof: Let $\{e_i\}_{i=1}^n$ be the canonical basis for \mathbb{R}^n . For each $1 \leq i, j \leq n$, define a corresponding matrix G_{ij} :

$$G_{ij} \equiv \begin{cases} e_i e_j^T + e_j e_i^T, & i \neq j; \\ e_i e_i^T, & i = j. \end{cases}$$

Thus

$$\hat{P}^{(1)} = \sum_{1 \leq i < j \leq n} \hat{P}_{ij}^{(1)} G_{ij} = \frac{1}{m} \sum_{t=1}^m \sum_{1 \leq i < j \leq n} A_{ij}^{(t)} G_{ij}$$

and

$$H^{(1)} = \sum_{1 \leq i < j \leq n} H_{ij}^{(1)} G_{ij}.$$

Then we have $\hat{P}^{(1)} - H^{(1)} = \frac{1}{m} \sum_{1 \leq t \leq m, 1 \leq i < j \leq n} X_{ij}^{(t)}$, where $X_{ij}^{(t)} = (A_{ij}^{(t)} - H_{ij}^{(1)}) G_{ij}$ for $1 \leq t \leq m$ and $1 \leq i < j \leq n$.

First bound the k -th moment of X_{ij} for $1 \leq i < j \leq n$ as following:

$$\begin{aligned} E[(A_{ij}^{(t)} - H_{ij}^{(1)})^k] &\leq (1 - \epsilon) \cdot \exp(-H_{ij}/P_{ij}) P_{ij}^k \Gamma(1 + k, -H_{ij}/P_{ij}) \\ &\quad + \epsilon \cdot \exp(-H_{ij}/C_{ij}) C_{ij}^k \Gamma(1 + k, -H_{ij}/C_{ij}) \\ &\leq ((1 - \epsilon) \cdot \exp(-H_{ij}/P_{ij}) P_{ij}^k + \epsilon \cdot \exp(-H_{ij}/C_{ij}) C_{ij}^k) k! \\ &\leq ((1 - \epsilon) \cdot P_{ij}^k + \epsilon \cdot C_{ij}^k) k! \\ &\leq R^k k!, \end{aligned} \tag{1}$$

Combined with

$$G_{ij}^k \equiv \begin{cases} e_i e_i^T + e_j e_j^T, & k \text{ is even;} \\ e_i e_j^T + e_j e_i^T, & k \text{ is odd,} \end{cases}$$

thus we have

1. When k is even,

$$E[(X_{ij}^{(t)})^k] = E[(A_{ij}^{(t)} - H_{ij}^{(1)})^k] G_{ij}^2 \preceq k! R^k G_{ij}^2;$$

2. When k is odd,

$$E[(X_{ij}^{(t)})^k] = E[(A_{ij}^{(t)} - H_{ij}^{(1)})^k] G_{ij} \preceq k! R^k G_{ij}^2.$$

So

$$E[(X_{ij}^{(t)})^k] \preceq k! R^k G_{ij}^2.$$

Let

$$\sigma^2 := \left\| \sum_{1 \leq t \leq m, 1 \leq i < j \leq n} (\sqrt{2} R G_{ij})^2 \right\|_2 = 2R^2 m \|(n-1)I\|_2 = 2R^2 m(n-1).$$

Notice that random matrices $X_{ij}^{(t)}$ are independent, self-adjoint and have mean zero, apply Theorem 3.7 we have

$$\begin{aligned} P\left(\lambda_{\max}(\hat{P}^{(1)} - H^{(1)}) \geq t\right) &= P\left(\lambda_{\max}\left(\frac{1}{m} \sum_{1 \leq t \leq m, 1 \leq i < j \leq n} X_{ij}^{(t)}\right) \geq t\right) \\ &= P\left(\lambda_{\max}\left(\sum_{1 \leq t \leq m, 1 \leq i < j \leq n} X_{ij}^{(t)}\right) \geq mt\right) \\ &\leq n \exp\left(-\frac{(mt)^2/2}{\sigma^2 + Rmt}\right) \\ &\leq n \exp\left(-\frac{mt^2/2}{2R^2n + Rt}\right). \end{aligned}$$

Now consider $Y_{ij}^{(t)} \equiv (H_{ij}^{(1)} - A_{ij}^{(t)}) G_{ij}$, for $1 \leq t \leq m$ and $1 \leq i < j \leq n$.

Then we have $H^{(1)} - \hat{P}^{(1)} = \frac{1}{m} \sum_{1 \leq t \leq m, 1 \leq i < j \leq n} Y_{ij}^{(t)}$. Since

$$E[(H^{(1)} - \hat{P}^{(1)})^k] = (-1)^k E[(\hat{P}^{(1)} - H^{(1)})^k],$$

1. When k is even,

$$E[(Y_{ij}^{(t)})^k] = E[(\hat{P}^{(1)} - H^{(1)})^k] G_{ij}^2 \preceq k! R^k G_{ij}^2;$$

2. When k is odd,

$$E[Y_{ij}^k] = -E[(\hat{P}^{(1)} - H^{(1)})^k] G_{ij} \preceq k! R^k G_{ij}^2.$$

Thus by similar arguments,

$$\begin{aligned} P\left(\lambda_{\min}(\hat{P}^{(1)} - H^{(1)}) \leq -t\right) &= P\left(\lambda_{\max}(H^{(1)} - \hat{P}^{(1)}) \geq t\right) \\ &\leq n \exp\left(-\frac{mt^2/2}{2R^2n + Rt}\right). \end{aligned}$$

Therefore we have

$$P\left(\|\hat{P}^{(1)} - H^{(1)}\|_2 \geq t\right) \leq n \exp\left(-\frac{mt^2/2}{2R^2n + Rt}\right).$$

Now let $c > 0$ be given and assume $n^{-c} \leq \eta \leq 1/2$. Then there exists a $n_0(c)$ independent of n , P , C and ϵ such that whenever $n > n_0(c)$,

$$t = 4R\sqrt{n \ln(n/\eta)/m} \leq 6Rn.$$

Plugging this t into the equation above, we get

$$P(\|\hat{P}^{(1)} - H^{(1)}\|_2 \geq 4R\sqrt{n \ln(n/\eta)/m}) \leq n \exp\left(-\frac{t^2}{16R^2n}\right) = \eta.$$

Define $H^{(1)} = E[\hat{P}^{(1)}] = (1 - \epsilon)P + \epsilon C$, where $P = XX^T$, $X \in \mathbb{R}^{n \times d}$, $C = YY^T$, $Y \in \mathbb{R}^{n \times d'}$. Let $d^{(1)} = \text{rank}(H^{(1)})$ be the dimension in which we are going to embed $\hat{P}^{(1)}$. Then we can define $H^{(1)} = ZZ^T$ where $Z \in \mathbb{R}^{n \times d^{(1)}}$. Since $H^{(1)} = [\sqrt{1 - \epsilon}X, \sqrt{\epsilon}Y][\sqrt{1 - \epsilon}X, \sqrt{\epsilon}Y]^T$, we have $d^{(1)} \leq d + d'$. ■

For simplicity, from now on, we will use \hat{P} to represent $\hat{P}^{(1)}$, use H to represent $H^{(1)}$ and use k to represent the dimension $d^{(1)}$ we are going to embed. Assume $H = USU^T = ZZ^T$, where $Z = [Z_1, \dots, Z_n]^T$ is a n -by- k matrix. Then our estimate for Z up to rotation is $\hat{Z} = \hat{U}\hat{S}^{1/2}$, where $\hat{U}\hat{S}\hat{U}^T$ is the rank- k spectral decomposition of $|\hat{P}| = (\hat{P}^T \hat{P})^{1/2}$.

Furthermore, we assume that the second moment matrix $E[Z_1 Z_1^T]$ is rank k and has distinct eigenvalues $\lambda_i(E[Z_1 Z_1^T])$. In particular, we assume that there exists $\delta > 0$ such that

$$\delta < \lambda_k(E[Z_1 Z_1^T])$$

Lemma 3.9 *Under the above assumptions, $\lambda_i(H) = \Theta(n)$ with high probability when $i \leq k$, i.e. the largest k eigenvalues of H is of order n . Moreover, we have $\|S\|_2 = \Theta(n)$ and $\|\hat{S}\|_2 = \Theta(n)$ with high probability.*

Remark: This is an extended version of Proposition 4.3 in [8].

Require: Theorem 3.8 and assumptions above

Proof: Note that $\lambda_i(H) = \lambda_i(ZZ^T) = \lambda_i(Z^T Z)$ when $i \leq k$. Since each entry of $Z^T Z$ is a sum of n independent random variables each in $[0, R]$, i.e. $(Z^T Z)_{ij} = \sum_{l=1}^n Z_{li} Z_{lj}$. By Hoeffding's inequality,

$$P(|(Z^T Z - nE[Z_1 Z_1^T])_{ij}| \geq t) \leq 2 \exp\left(-\frac{2t^2}{nR^2}\right).$$

Now let $c > 0$ and assume $n^{-c} \leq \eta \leq 1/2$. Let

$$t = R\sqrt{n \ln(\sqrt{2}/\eta)},$$

we have

$$P\left(|(Z^T Z - nE[Z_1 Z_1^T])_{ij}| \geq R\sqrt{n \ln(\sqrt{2}/\eta)}\right) \leq \eta.$$

By the union bound, we have

$$P\left(\|Z^T Z - nE[Z_1 Z_1^T]\|_F \geq kR\sqrt{n \ln(\sqrt{2}/\eta)}\right) \leq k^2 \eta.$$

Then by Weyl's Theorem [3], we have

$$|\lambda_i(H) - n\lambda_i(Z_1 Z_1^T)| \leq \|Z^T Z - nE[Z_1 Z_1^T]\|_2 = O(\sqrt{n \log(1/\eta)})$$

with probability at least $1 - k^2\eta$. Thus $\lambda_i(H) = S_{ii} = \Theta(n)$ with probability at least $1 - \frac{2k^2}{n^2}$ when $i \leq k$. Moreover,

$$\|H\|_2 - \|H - \hat{P}\|_2 \leq \|\hat{S}\|_2 \leq \|\hat{P} - H\|_2 + \|H\|_2.$$

Combined with Theorem 3.8, with high probability we have $\|\hat{S}\|_2 = \Theta(n)$. ■

Lemma 3.10 *Let $W_1 \Sigma W_2^T$ be the singular value decomposition of $U^T \hat{U}$. Then for sufficiently large n ,*

$$\|U^T \hat{U} - W_1 W_2^T\|_F = O(m^{-1} n^{-1} \log n)$$

with high probability.

Require: Theorem 3.8 and Lemma 3.9

Proof: Let $\sigma_1, \dots, \sigma_k$ denote the singular values of $U^T \hat{U}$. Then $\sigma_i = \cos(\theta_i)$ where the θ_i are the principal angles between the subspaces spanned by \hat{U} and U . Furthermore, by the Davis-Kahan $\sin(\Theta)$ theorem [1], combined with Theorem 3.8 and Lemma 3.9,

$$\begin{aligned} \|\hat{U} \hat{U}^T - U U^T\|_2 &= \max_i |\sin(\theta_i)| \\ &\leq \frac{\|\hat{P} - H\|_2}{\lambda_k(H)} \leq \frac{C \sqrt{n \log n/m}}{n} \\ &= O(m^{-1/2} n^{-1/2} \sqrt{\log n}) \end{aligned} \tag{2}$$

for sufficiently large n with high probability. Here $\lambda_k(H)$ denotes the k -th largest eigenvalue of H . Thus with high probability,

$$\begin{aligned} \|U^T \hat{U} - W_1 W_2^T\|_F &= \|\Sigma - I\|_F = \sqrt{\sum_{i=1}^k (1 - \sigma_i)^2} \\ &\leq \sum_{i=1}^k (1 - \sigma_i) \leq \sum_{i=1}^k (1 - \sigma_i^2) \\ &= \sum_{i=1}^k \sin^2(\theta_i) \leq k \|\hat{U} \hat{U}^T - U U^T\|_2^2 \\ &= O(m^{-1} n^{-1} \log n). \end{aligned}$$

■

We will denote the orthogonal matrix $W_1 W_2^T$ by W^* .

Lemma 3.11 *For sufficiently large n ,*

$$\begin{aligned} \|W^* \hat{S} - S W^*\|_F &= O(m^{-1/2} \log n), \\ \|W^* \hat{S}^{1/2} - S^{1/2} W^*\|_F &= O(m^{-1/2} n^{-1/2} \log n) \end{aligned}$$

and

$$\|W^* \hat{S}^{-1/2} - S^{-1/2} W^*\|_F = O(m^{-1/2} n^{-3/2} \log n)$$

with high probability.

Require: Theorem 3.8, Lemma 3.9, Lemma 3.10, Exponential distribution

Proof: By Proposition 2.1 in [7] and Equation (2), we have for some orthogonal matrix W ,

$$\begin{aligned}\|\hat{U} - UW\|_F^2 &\leq \frac{2\|\hat{U}\hat{U}^T - UU^T\|_F^2}{\delta^2} \leq \frac{8k^2\|\hat{U}\hat{U}^T - UU^T\|_2^2}{\delta^2} \\ &= O(m^{-1}n^{-1}\log n),\end{aligned}$$

with high probability. Let $Q = \hat{U} - UU^T\hat{U}$. And Q is the residual after projecting \hat{U} orthogonally onto the column space of U , we have

$$\|Q\|_F = \|\hat{U} - UU^T\hat{U}\|_F \leq \|\hat{U} - UT\|_F = O(m^{-1/2}n^{-1/2}\sqrt{\log n}). \quad (3)$$

for all $k \times k$ matrices T with high probability. Then

$$\begin{aligned}W^*\hat{S} &= (W^* - U^T\hat{U})\hat{S} + U^T\hat{U}\hat{S} = (W^* - U^T\hat{U})\hat{S} + U^T\hat{P}\hat{U} \\ &= (W^* - U^T\hat{U})\hat{S} + U^T(\hat{P} - H)\hat{U} + U^TH\hat{U} \\ &= (W^* - U^T\hat{U})\hat{S} + U^T(\hat{P} - H)Q + U^T(\hat{P} - H)UU^T\hat{U} + U^TH\hat{U} \\ &= (W^* - U^T\hat{U})\hat{S} + U^T(\hat{P} - H)Q + U^T(\hat{P} - H)UU^T\hat{U} + SU^T\hat{U}.\end{aligned}$$

Combined with Theorem 3.8, Lemma 3.9, Lemma 3.10, we have

$$\begin{aligned}\|W^*\hat{S} - SW^*\|_F &= \|(W^* - U^T\hat{U})\hat{S} + U^T(\hat{P} - H)Q + U^T(\hat{P} - H)UU^T\hat{U} + S(U^T\hat{U} - W^*)\|_F \\ &\leq \|W^* - U^T\hat{U}\|_F(\|\hat{S}\|_2 + \|S\|_2) + \|U^T\|_F\|\hat{P} - H\|_2\|Q\|_F + \|U^T(\hat{P} - H)U\|_F \\ &\leq O(m^{-1}\log n) + O(m^{-1/2}\log n) + \|U^T(\hat{P} - H)U\|_F\end{aligned}$$

with high probability. And we know $U^T(\hat{P} - H)U$ is a $k \times k$ matrix with ij -th entry to be

$$u_i^T(\hat{P} - H)u_j = \sum_{s=1}^n \sum_{t=1}^n (\hat{P}_{st} - H_{st})u_{is}u_{jt} = 2 \sum_{s < t} (\hat{P}_{st} - H_{st})u_{is}u_{jt}$$

where u_i and u_j are the i -th and j -th columns of U . Thus, conditioned on H , U is fixed and $u_i^T(\hat{P} - H)u_j$ is a sum of independent mean 0 random variables.

By Equation (1), we have

$$\begin{aligned}&E \left[\left((A_{st}^{(t')} - H_{st})u_{is}u_{jt} \right)^k \right] \\ &\leq k! R^k u_{is}^k u_{jt}^k \\ &\leq \frac{k!}{2} R^{k-2} (\sqrt{2}u_{is}u_{jt}R)^2.\end{aligned}$$

Also we have

$$\sigma^2 := \left| \sum_{t', s < t} 2R^2 u_{is}^2 u_{jt}^2 \right| \leq mR^2,$$

then by Theorem 3.7, we have

$$P \left(\left| 2 \sum_{s < t} (\hat{P}_{st} - H_{st}) u_{is} u_{jt} \right| \geq t \right) \leq \exp \left(\frac{-mt^2/8}{R^2 + Rt/2} \right).$$

Let $t = cRm^{-1/2} \log n$ for any $c > 0$, we have

$$P \left(\left| 2 \sum_{s < t} (\hat{P}_{st} - H_{st}) u_{is} u_{jt} \right| \geq Cm^{-1/2} \log n \right) \leq n^{-c}.$$

Thus each entry of $U^T(\hat{P} - H)U$ is of order $O(m^{-1} \log n)$ with high probability and

$$\|U^T(\hat{P} - H)U\|_F = O(m^{-1} \log n) \quad (4)$$

with high probability. Hence

$$\|W^* \hat{S} - SW^*\|_F = O(m^{-1/2} \log n)$$

with high probability. Also, since

$$W_{ij}^* (\lambda_j^{1/2}(\hat{P}) - \lambda_i^{1/2}(H)) = W_{ij}^* \frac{\lambda_j(\hat{P}) - \lambda_i(H)}{\lambda_j^{1/2}(\hat{P}) + \lambda_i^{1/2}(H)}$$

and the eigenvalues $\lambda_j^{1/2}(\hat{P})$ and $\lambda_i^{1/2}(H)$ are both of order $\Theta(\sqrt{n})$, we have

$$\|W^* \hat{S}^{1/2} - S^{1/2} W^*\|_F = O(m^{-1/2} n^{-1/2} \log n)$$

with high probability. Similarly, since

$$W_{ij}^* (\lambda_j^{-1/2}(\hat{P}) - \lambda_i^{-1/2}(H)) = W_{ij}^* \frac{\lambda_i(H) - \lambda_j(\hat{P})}{(\lambda_j^{-1/2}(\hat{P}) + \lambda_i^{-1/2}(H)) \lambda_j(\hat{P}) \lambda_i(H)}$$

and the eigenvalues $\lambda_j(\hat{P})$ and $\lambda_i(H)$ are both of order $\Theta(n)$, with high probability we have

$$\|W^* \hat{S}^{-1/2} - S^{-1/2} W^*\|_F = O(m^{-1/2} n^{-3/2} \log n).$$

■

Lemma 3.12 *There exists a rotation matrix W such that for sufficiently large n ,*

$$\|\hat{Z} - ZW\|_F = \|(\hat{P} - H)US^{-1/2}\|_F + O(m^{-1/2} n^{-1/2} (\log n)^{3/2})$$

with high probability.

Require: Lemma 3.10, Lemma 3.11

Proof: Let $Q_1 = UU^T \hat{U} - UW^*$, $Q_2 = W^* \hat{S}^{1/2} - S^{1/2} W^*$ and $Q_3 = \hat{U} - UW^* = \hat{U} - UU^T \hat{U} + Q_1 = Q + Q_1$. Then since $UU^T H = H$ and $\hat{U} \hat{S}^{1/2} = \hat{P} \hat{U} \hat{S}^{-1/2}$,

$$\begin{aligned} \hat{Z} - US^{1/2} W^* &= \hat{U} \hat{S}^{1/2} - UW^* \hat{S}^{1/2} + U(W^* \hat{S}^{1/2} - S^{1/2} W^*) \\ &= (\hat{U} - UU^T \hat{U}) \hat{S}^{1/2} + Q_1 \hat{S}^{1/2} + UQ_2 \\ &= (\hat{P} - H) \hat{U} \hat{S}^{-1/2} - UU^T (\hat{P} - H) \hat{U} \hat{S}^{-1/2} + Q_1 \hat{S}^{1/2} + UQ_2 \\ &= (\hat{P} - H) UW^* \hat{S}^{-1/2} - UU^T (\hat{P} - H) UW^* \hat{S}^{-1/2} \\ &\quad + (I - UU^T) (\hat{P} - H) Q_3 \hat{S}^{-1/2} + Q_1 \hat{S}^{1/2} + UQ_2. \end{aligned}$$

By Lemma 3.10, with high probability,

$$\|Q_1\|_F \leq \|U\|_F \|U^T \hat{U} - W^*\|_F = O(m^{-1} n^{-1} \log n).$$

By Lemma 3.11, with high probability,

$$\|Q_2\|_F = O(m^{-1/2} n^{-1/2} \log n).$$

By Equation (3), with high probability,

$$\|Q_3\|_F \leq \|Q\|_F + \|Q_1\|_F = O(m^{-1/2} n^{-1/2} (\log n)^{1/2}).$$

By Equation (4), with high probability,

$$\|UU^T(\hat{P} - H)UW^* \hat{S}^{-1/2}\|_F \leq \|U^T(\hat{P} - H)U\|_F \|\hat{S}^{-1/2}\|_2 = O(m^{-1} n^{-1/2} \log n).$$

By Lemma 3.11, with high probability,

$$\|W^* \hat{S}^{-1/2} - S^{-1/2} W^*\|_F = O(m^{-1/2} n^{-3/2} \log n).$$

Therefore, with high probability,

$$\begin{aligned} & \|\hat{Z} - US^{1/2}W^*\|_F \\ &= \|(\hat{P} - H)UW^* \hat{S}^{-1/2}\|_F + O(m^{-1} n^{-1/2} \log n) + \|I - UU^T\|_2 \|\hat{P} - H\|_2 O(m^{-1/2} n^{-1} (\log n)^{1/2}) \\ & \quad + O(m^{-1} n^{-1/2} \log n) + O(m^{-1/2} n^{-1/2} \log n) \\ &= \|(\hat{P} - H)UW^* \hat{S}^{-1/2}\|_F + O(m^{-1/2} n^{-1/2} \log n) \\ &\leq \|(\hat{P} - H)US^{-1/2}W^*\|_F + \|(\hat{P} - H)U(W^* \hat{S}^{-1/2} - S^{-1/2}W^*)\|_F + O(m^{-1/2} n^{-1/2} \log n) \\ &= \|(\hat{P} - H)US^{-1/2}\|_F + O(m^{-1} n^{-1} (\log n)^{3/2}) + O(m^{-1/2} n^{-1/2} (\log n)^{3/2}) \\ &= \|(\hat{P} - H)US^{-1/2}\|_F + O(m^{-1/2} n^{-1/2} (\log n)^{3/2}). \end{aligned}$$

Note that $Z = US^{1/2}W$ for some orthogonal matrix W . As W^* is also orthogonal, therefore $Z\tilde{W} = US^{1/2}W^*$ for some orthogonal \tilde{W} , which completes the proof. \blacksquare

Theorem 3.13 *There exists a rotation matrix W such that for sufficiently large n ,*

$$\max_i \|\hat{Z}_i - WZ_i\|_2 = O(m^{-1/2} n^{-1/2} (\log n)^{3/2})$$

with high probability.

Require: Lemma 3.12, Exponential distribution

Proof: By Lemma 3.12, we have

$$\|\hat{Z} - ZW\|_F = \|(\hat{P} - H)US^{-1/2}\|_F + O(m^{-1/2} n^{-1/2} (\log n)^{3/2})$$

with high probability and similarly we could have the bound for each column vector with high probability that

$$\begin{aligned} \max_i \|\hat{Z}_i - WZ_i\|_2 &\leq \frac{1}{\lambda_k^{1/2}(H)} \max_i \|((\hat{P} - H)U)_i\|_2 + O(m^{-1/2} n^{-1/2} (\log n)^{3/2}) \\ &\leq \frac{k^{1/2}}{\lambda_k^{1/2}(H)} \max_j \|(\hat{P} - H)u_j\|_\infty + O(m^{-1/2} n^{-1/2} (\log n)^{3/2}) \end{aligned}$$

where $((\hat{P} - H)U)_i$ represents the i -th row of $(\hat{P} - H)U$ and u_j denotes the j -th column of U . Now given i and j , the i -th element of the vector $(\hat{P} - H)u_j$ is of the form

$$\sum_{s=1}^n (\hat{P}_{is} - H_{is})u_{js} = \sum_{s \neq i} (\hat{P}_{is} - H_{is})u_{js}.$$

Thus, conditioned on H , the i -th element of the vector $(\hat{P} - H)u_j$ is a sum of independent mean 0 random variables. By Equation (1), we have

$$\begin{aligned} & E \left[\left((A_{is}^{(t)} - H_{is})u_{js} \right)^k \right] \\ & \leq k! R^k u_{js}^k \\ & \leq \frac{k!}{2} R^{k-2} (\sqrt{2} R u_{js})^2. \end{aligned}$$

Also we have

$$\sigma^2 := \left| \sum_{t, s \neq i} 2R^2 u_{js}^2 \right| \leq 2R^2 m,$$

then by Theorem 3.7, we have

$$P \left(\left| \sum_{s \neq i} (\hat{P}_{is} - H_{is})u_{js} \right| \geq t \right) \leq \exp \left(\frac{-mt^2/2}{2R^2 + Rt} \right).$$

Let $t = 3cRm^{-1/2} \log n$, we have

$$P \left(\left| \sum_{s \neq i} (\hat{P}_{is} - H_{is})u_{js} \right| \geq 3cRm^{-1/2} \log n \right) \leq n^{-c},$$

i.e. it is of order $O(m^{-1/2} \log n)$ with high probability. Taking the union bound over all i and j , with high probability we have,

$$\begin{aligned} \max_i \|\hat{Z}_i - WZ_i\|_2 & \leq \frac{Ck^{1/2}}{\lambda_k^{1/2}(H)} m^{-1/2} (\log n)^{3/2} + O(m^{-1/2} n^{-1/2} (\log n)^{3/2}) \\ & = O(m^{-1/2} n^{-1/2} (\log n)^{3/2}). \end{aligned}$$

■

Lemma 3.14 $\left| \hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j \right| = O(m^{-1/2} n^{-1} (\log n)^3)$ with high probability.

Require: Theorem 3.13

Proof: Let W be the rotation matrix in Theorem 3.13, then

$$\begin{aligned} \left| \hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j \right| & = \left| \hat{Z}_i^T \hat{Z}_j - \hat{Z}_i^T WZ_j + \hat{Z}_i^T WZ_j - (WZ_i)^T WZ_j \right| \\ & \leq \left| \hat{Z}_i^T (\hat{Z}_j - WZ_j) + (\hat{Z}_i^T - (WZ_i)^T) WZ_j \right| \\ & \leq \|\hat{Z}_i\|_2 \|\hat{Z}_j - WZ_j\|_2 + \|Z_j\|_2 \|\hat{Z}_i^T - (WZ_i)^T\|_2. \end{aligned}$$

Since $\|Z_i\|_2^2 = Z_i^T Z_i = H_{ii}^{(1)} = E[\hat{P}_{ii}^{(1)}] = (1 - \epsilon)P_{ij} + \epsilon C_{ij} \leq R$, we have $\|Z_i\|_2 = O(1)$. Combined with Theorem 3.13,

$$\begin{aligned} \left| \hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j \right| &= (\|\hat{Z}_i\|_2 + \|Z_j\|_2) O(n^{-1/2} (\log n)^{3/2}) \\ &\leq (\|\hat{Z}_i - W Z_i\|_2 + \|W Z_i\|_2 + \|Z_j\|_2) O(m^{-1/2} n^{-1/2} (\log n)^{3/2}) \\ &= O(m^{-1/2} n^{-1} (\log n)^3) \end{aligned}$$

with high probability. ■

Definition 3.15 Define $\tilde{P}_{ij}^{(1)} = (\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}$, our estimator for P_{ij} , to be a projection of $\hat{Z}_i^T \hat{Z}_j$ onto $[0, \min(\hat{P}_{ij}^{(1)}, R)]$.

Corollary 3.16 Assuming that $m = O(n^b)$ for any $b > 0$, then the estimator based on ASE of MLE has the same entry-wise asymptotic bias as MLE, i.e.

$$\lim_{n \rightarrow \infty} \text{Bias}(\tilde{P}_{ij}^{(1)}) = \lim_{n \rightarrow \infty} E[\tilde{P}_{ij}^{(1)}] - P_{ij} = \lim_{n \rightarrow \infty} E[\hat{P}_{ij}^{(1)}] - P_{ij} = \lim_{n \rightarrow \infty} \text{Bias}(\hat{P}_{ij}^{(1)}).$$

Require: Lemma 3.14, Exponential distribution

Proof: Fix some $a > 0$, we have

$$\begin{aligned} &E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j] \\ &= E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j | \mathbb{I}\{\hat{P}_{ij} \leq a\}] + E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j | \mathbb{I}\{\hat{P}_{ij} > a\}] \end{aligned}$$

For the first term, we have

$$\begin{aligned} &E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j | \mathbb{I}\{\hat{P}_{ij} \leq a\}] \\ &\leq E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j | \mathbb{I}\{\hat{P}_{ij} \leq a\} \mathbb{I}\{\text{Lemma 3.14 holds}\}] \\ &\quad + E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j | \mathbb{I}\{\hat{P}_{ij} \leq a\} \mathbb{I}\{\text{Lemma 3.14 does not hold}\}] \\ &\leq E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j | \mathbb{I}\{\hat{P}_{ij} \leq a\} | \text{Lemma 3.14 holds}] \\ &\quad + n^{-c} E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j | \mathbb{I}\{\hat{P}_{ij} \leq a\} | \text{Lemma 3.14 does not hold}] \\ &\leq O(m^{-1/2} n^{-1} (\log n)^3) \\ &\quad + n^{-c} E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - \hat{P}_{ij} | \mathbb{I}\{\hat{P}_{ij} \leq a\} | \text{Lemma 3.14 does not hold}] \\ &\quad + n^{-c} E[\hat{P}_{ij} - Z_i^T Z_j | \mathbb{I}\{\hat{P}_{ij} \leq a\} | \text{Lemma 3.14 does not hold}] \\ &\leq O(m^{-1/2} n^{-1} (\log n)^3) + n^{-c} E[\hat{P}_{ij} | \mathbb{I}\{\hat{P}_{ij} \leq a\} | \text{Lemma 3.14 does not hold}] \\ &\quad + n^{-c} E[(\hat{P}_{ij} + R) | \mathbb{I}\{\hat{P}_{ij} \leq a\} | \text{Lemma 3.14 does not hold}] \\ &\leq O(m^{-1/2} n^{-1} (\log n)^3) + a n^{-c} + (a + R) n^{-c} \\ &\leq O(m^{-1/2} n^{-1} (\log n)^3) + 2n^{-c} (a + R). \end{aligned}$$

Notice that

$$\begin{aligned}
E[\hat{P}_{ij} \mathbb{I}\{\hat{P}_{ij} > a\}] &= E\left[\left(\frac{1}{m} \sum_{1 \leq t \leq m} A_{ij}^{(t)}\right) \mathbb{I}\{\hat{P}_{ij} > a\}\right] \\
&= \frac{1}{m} E\left[\sum_{1 \leq t \leq m} A_{ij}^{(t)} \mathbb{I}\{\hat{P}_{ij} > a\}\right] \leq \frac{1}{m} E\left[\sum_{1 \leq t \leq m} A_{ij}^{(t)} \mathbb{I}\{\max_{1 \leq s \leq m} A_{ij}^{(s)} > a\}\right] \\
&\leq \frac{1}{m} E\left[\sum_{1 \leq t \leq m} A_{ij}^{(t)} \left(\sum_{1 \leq s \leq m} \mathbb{I}\{A_{ij}^{(s)} > a\}\right)\right] = E[A_{ij}^{(1)} \left(\sum_{1 \leq s \leq m} \mathbb{I}\{A_{ij}^{(s)} > a\}\right)] \\
&= E[A_{ij}^{(1)} \mathbb{I}\{A_{ij}^{(1)} > a\}] + (m-1)E[A_{ij}^{(1)} \mathbb{I}\{A_{ij}^{(2)} > a\}] \\
&= E[A_{ij}^{(1)} \mathbb{I}\{A_{ij}^{(1)} > a\}] + (m-1)E[A_{ij}^{(1)}]P(A_{ij}^{(1)} > a),
\end{aligned}$$

and similarly

$$\begin{aligned}
&E[(\hat{P}_{ij} + R) \mathbb{I}\{\hat{P}_{ij} > a\}] \\
&= E[\hat{P}_{ij} \mathbb{I}\{\hat{P}_{ij} > a\}] + R \cdot P(\hat{P}_{ij} > a) \\
&\leq E[A_{ij}^{(1)} \mathbb{I}\{A_{ij}^{(1)} > a\}] + (m-1)E[A_{ij}^{(1)}]P(A_{ij}^{(1)} > a) + R \cdot m \cdot P(A_{ij}^{(1)} > a).
\end{aligned}$$

Thus for the second term,

$$\begin{aligned}
&E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij} > a\}] \\
&\leq E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - \hat{P}_{ij}| \mathbb{I}\{\hat{P}_{ij} > a\}] + E[|\hat{P}_{ij} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij} > a\}] \\
&\leq E[\hat{P}_{ij} \mathbb{I}\{\hat{P}_{ij} > a\}] + E[(\hat{P}_{ij} + R) \mathbb{I}\{\hat{P}_{ij} > a\}] \\
&\leq 2E[A_{ij}^{(1)} \mathbb{I}\{A_{ij}^{(1)} > a\}] + 2(m-1)E[A_{ij}^{(1)}]P(A_{ij}^{(1)} > a) \\
&\quad + R \cdot m \cdot P(A_{ij}^{(1)} > a) \\
&\leq 2e^{-a/R}(a + 2mR).
\end{aligned}$$

Thus

$$\begin{aligned}
&E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j|] \\
&\leq O(m^{-1/2}n^{-1}(\log n)^3) + 4n^{-c}(a + R) + 2e^{-a/R}(a + 2mR).
\end{aligned}$$

Let $a = m^{-1}n^{2b}$ for any $b > 0$, and $c = 2b + 3$, combined with the assumption $m = O(n^b)$, we have

$$\begin{aligned}
&E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j|] \\
&= O(m^{-1/2}n^{-1}(\log n)^3) + O(m^{-1}n^{-3}) + O(m^{-1}n^{2b}) \cdot O(e^{-m^{-1}n^{2b}}) \\
&= O(m^{-1/2}n^{-1}(\log n)^3) + O(m^{-1}n^{-3}) + O(m^{-1}n^{2b}) \cdot O(e^{-n^b}) \\
&= O(m^{-1/2}n^{-1}(\log n)^3) + O(m^{-1}n^{-3}) + O(m^{-1}n^{2b}) \cdot O(n^{-2b-3}) \\
&= O(m^{-1/2}n^{-1}(\log n)^3) + O(m^{-1}n^{-3}) \\
&= O(m^{-1/2}n^{-1}(\log n)^3).
\end{aligned}$$

■

Theorem 3.17 *Assuming that $m = O(n^b)$ for any $b > 0$, then $\text{Var}((\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}) = O(m^{-1}n^{-2}(\log n)^6)$.*

Require: Lemma 3.14, Exponential distribution

Proof: By Lemma 3.14,

$$\begin{aligned}
\text{Var}((\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}) &= E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}]]^2 \\
&= E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j + Z_i^T Z_j - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}]]^2 \\
&= E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 + E[(Z_i^T Z_j - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}])]^2 \\
&\quad + 2E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j](Z_i^T Z_j - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}]) \\
&\leq E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 + E[(Z_i^T Z_j - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}])]^2 \\
&\quad + 2\sqrt{E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 E[(Z_i^T Z_j - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}])]^2} \\
&\leq 4E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2.
\end{aligned}$$

Fix some $a > 0$, we have

$$\begin{aligned}
&E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \\
&= E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij} \leq a\} + E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij} > a\}.
\end{aligned}$$

For the first term, we have

$$\begin{aligned}
&E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij} \leq a\} \\
&\leq E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij} \leq a\} \mathbb{I}\{\text{Lemma 3.14 holds}\} \\
&\quad + E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij} \leq a\} \mathbb{I}\{\text{Lemma 3.14 does not hold}\} \\
&\leq E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij} \leq a\} |\text{Lemma 3.14 holds}| \\
&\quad + n^{-c} E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij} \leq a\} |\text{Lemma 3.14 does not hold}| \\
&\leq O(m^{-1}n^{-2}(\log n)^6) \\
&\quad + 2n^{-c} E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - \hat{P}_{ij}]^2 \mathbb{I}\{\hat{P}_{ij} \leq a\} |\text{Lemma 3.14 does not hold}| \\
&\quad + 2n^{-c} E[(\hat{P}_{ij} - Z_i^T Z_j)^2 \mathbb{I}\{\hat{P}_{ij} \leq a\} |\text{Lemma 3.14 does not hold}| \\
&\leq O(m^{-1}n^{-2}(\log n)^6) + 2n^{-c} E[\hat{P}_{ij}^2 \mathbb{I}\{\hat{P}_{ij} \leq a\} |\text{Lemma 3.14 does not hold}| \\
&\quad + 2n^{-c} E[(\hat{P}_{ij} + R)^2 \mathbb{I}\{\hat{P}_{ij} \leq a\} |\text{Lemma 3.14 does not hold}| \\
&\leq O(m^{-1}n^{-2}(\log n)^6) + 2a^2 n^{-c} + 2(a + R)^2 n^{-c} \\
&\leq O(m^{-1}n^{-2}(\log n)^6) + 4n^{-c}(a + R)^2.
\end{aligned}$$

Notice that

$$\begin{aligned}
E[\hat{P}_{ij}^2 \mathbb{I}\{\hat{P}_{ij} > a\}] &= E\left[\left(\frac{1}{m} \sum_{1 \leq t \leq m} A_{ij}^{(t)}\right)^2 \mathbb{I}\{\hat{P}_{ij} > a\}\right] \\
&\leq \frac{1}{m} E\left[\sum_{1 \leq t \leq m} A_{ij}^{(t)2} \mathbb{I}\{\hat{P}_{ij} > a\}\right] \leq \frac{1}{m} E\left[\sum_{1 \leq t \leq m} A_{ij}^{(t)2} \mathbb{I}\{\max_{1 \leq s \leq m} A_{ij}^{(s)} > a\}\right] \\
&\leq \frac{1}{m} E\left[\sum_{1 \leq t \leq m} A_{ij}^{(t)2} \left(\sum_{1 \leq s \leq m} \mathbb{I}\{A_{ij}^{(s)} > a\}\right)\right] = E[A_{ij}^{(1)2} \left(\sum_{1 \leq s \leq m} \mathbb{I}\{A_{ij}^{(s)} > a\}\right)] \\
&= E[A_{ij}^{(1)2} \mathbb{I}\{A_{ij}^{(1)} > a\}] + (m-1)E[A_{ij}^{(1)2} \mathbb{I}\{A_{ij}^{(2)} > a\}]] \\
&= E[A_{ij}^{(1)2} \mathbb{I}\{A_{ij}^{(1)} > a\}] + (m-1)E[A_{ij}^{(1)2}]P(A_{ij}^{(1)} > a),
\end{aligned}$$

and similarly

$$\begin{aligned}
&E[(\hat{P}_{ij} + R)^2 \mathbb{I}\{\hat{P}_{ij} > a\}] \\
&= E[\hat{P}_{ij}^2 \mathbb{I}\{\hat{P}_{ij} > a\}] + 2R \cdot E[\hat{P}_{ij} \mathbb{I}\{\hat{P}_{ij} > a\}] + R^2 P(\hat{P}_{ij} > a) \\
&\leq E[A_{ij}^{(1)2} \mathbb{I}\{A_{ij}^{(1)} > a\}] + (m-1)E[A_{ij}^{(1)2}]P(A_{ij}^{(1)} > a) \\
&\quad + 2R \left(E[A_{ij}^{(1)} \mathbb{I}\{A_{ij}^{(1)} > a\}]\right) + (m-1)E[A_{ij}^{(1)}]P(A_{ij}^{(1)} > a) \\
&\quad + R^2 \cdot m \cdot P(A_{ij}^{(1)} > a).
\end{aligned}$$

Thus for the second term,

$$\begin{aligned}
&E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij} > a\}] \\
&\leq 2E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - \hat{P}_{ij}]^2 \mathbb{I}\{\hat{P}_{ij} > a\}] + 2E[(\hat{P}_{ij} - Z_i^T Z_j)^2 \mathbb{I}\{\hat{P}_{ij} > a\}] \\
&\leq 2E[\hat{P}_{ij}^2 \mathbb{I}\{\hat{P}_{ij} > a\}] + 2E[(\hat{P}_{ij} + R)^2 \mathbb{I}\{\hat{P}_{ij} > a\}] \\
&\leq 4E[A_{ij}^{(1)2} \mathbb{I}\{A_{ij}^{(1)} > a\}] + 4(m-1)E[A_{ij}^{(1)2}]P(A_{ij}^{(1)} > a) \\
&\quad + 4R \cdot E[A_{ij}^{(1)} \mathbb{I}\{A_{ij}^{(1)} > a\}] + 2R(m-1)E[A_{ij}^{(1)}]P(A_{ij}^{(1)} > a) \\
&\quad + 2R^2 \cdot m \cdot P(A_{ij}^{(1)} > a) \\
&\leq 4e^{-a/R} (a^2 + 3Ra + 3(m+1)R^2) \\
&\leq 4e^{-a/R} (a + 2m^{1/2}R)^2.
\end{aligned}$$

Thus,

$$\text{Var}((\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}) \leq O(m^{-1}n^{-2}(\log n)^6) + 16(a+R)^2n^{-c} + 16(a+2m^{1/2}R)^2e^{-a/R}.$$

Let $a = m^{-1/2}n^b$ for any $b > 0$, and $c = 2b + 3$, combined with the assumption $m = O(n^b)$, we have

$$\begin{aligned}
\text{Var}((\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}) &= O(m^{-1}n^{-2}(\log n)^6) + O(m^{-1}n^{-3}) + O(m^{-1}n^{2b}) \cdot O(e^{-m^{-1/2}n^b}) \\
&= O(m^{-1}n^{-2}(\log n)^6) + O(m^{-1}n^{-3}) + O(m^{-1}n^{2b}) \cdot O(e^{-n^{b/2}}) \\
&= O(m^{-1}n^{-2}(\log n)^6) + O(m^{-1}n^{-3}) + O(m^{-1}n^{2b}) \cdot O(n^{-2b-3}) \\
&= O(m^{-1}n^{-2}(\log n)^6) + O(m^{-1}n^{-3}) \\
&= O(m^{-1}n^{-2}(\log n)^6).
\end{aligned}$$

■

Corollary 3.18 For fixed n , $1 \leq i, j \leq n$, $\text{Var}(\hat{P}_{ij}^{(1)}) = \Theta(m^{-1})$.

Proof: Direct result from central limit theorem. ■

Theorem 3.19 For fixed m , $1 \leq i, j \leq n$ and $i \neq j$,

$$\frac{\text{Var}(\tilde{P}_{ij}^{(1)})}{\text{Var}(\hat{P}_{ij}^{(1)})} = O(n^{-2}(\log n)^6).$$

Thus

$$\text{ARE}(\hat{P}_{ij}^{(1)}, \tilde{P}_{ij}^{(1)}) = 0.$$

Furthermore, as long as m goes to infinity of order $O(n^b)$ for any $b > 0$,

$$\text{ARE}(\hat{P}_{ij}^{(1)}, \tilde{P}_{ij}^{(1)}) = 0.$$

Proof: The results are direct from Theorem 3.17 and Corollary 3.18. ■

3.3 $\tilde{P}^{(q)}$ better than $\hat{P}^{(q)}$

Theorem 3.20 Let P and C be two n -by- n symmetric and hollow matrices satisfying element-wise conditions $0 < P_{ij} \leq C_{ij} \leq R$ for some constant $R > 0$. For $0 < \epsilon < 1$, we define m symmetric and hollow matrices as

$$A^{(t)} \stackrel{iid}{\sim} (1 - \epsilon)\text{Exp}(P) + \epsilon\text{Exp}(C)$$

for $1 \leq t \leq m$. Let $\hat{P}^{(q)}$ be the entry-wise MLqE based on exponential distribution with m observations. Define $H^{(q)} = E[\hat{P}^{(q)}]$, then for any constant $c > 0$ there exists another constant $n_0(c)$, independent of n , P , C and ϵ , such that if $n > n_0$, then for all η satisfying $n^{-c} \leq \eta \leq 1/2$,

$$P \left(\|\hat{P}^{(q)} - H^{(q)}\|_2 \leq 8R\sqrt{2n \ln(n/\eta)} \right) \geq 1 - \eta.$$

Remark: This is the extended version of Theorem 3.1 in [5].

Require: Lemma 3.1, Theorem 3.8

Proof: Similar to the proof of Theorem 3.8.

By Lemma 3.1 we have

$$\begin{aligned} \left| \hat{P}_{ij}^{(q)} - H_{ij}^{(q)} \right| &= \left| \hat{P}_{ij}^{(q)} - \hat{P}_{ij}^{(1)} + \hat{P}_{ij}^{(1)} - H_{ij}^{(1)} + H_{ij}^{(1)} - H_{ij}^{(q)} \right| \\ &\leq \left| \hat{P}_{ij}^{(q)} - \hat{P}_{ij}^{(1)} \right| + \left| \hat{P}_{ij}^{(1)} - H_{ij}^{(1)} \right| + \left| H_{ij}^{(1)} - H_{ij}^{(q)} \right| \\ &\leq \hat{P}_{ij}^{(1)} + \left| \hat{P}_{ij}^{(1)} - H_{ij}^{(1)} \right| + H_{ij}^{(1)} \\ &\leq 2 \left(\left| \hat{P}_{ij}^{(1)} - H_{ij}^{(1)} \right| + H_{ij}^{(1)} \right). \end{aligned}$$

Also,

$$\begin{aligned}
E[(\hat{P}_{ij}^{(q)} - H_{ij}^{(q)})^k] &\leq E\left[\left|\hat{P}_{ij}^{(q)} - H_{ij}^{(q)}\right|^k\right] \\
&\leq 2^k E\left[\left(\left|\hat{P}_{ij}^{(1)} - H_{ij}^{(1)}\right| + H_{ij}^{(1)}\right)^k\right] \\
&\leq 2^k \sum_{s=0}^k \binom{k}{s} E\left[\left|\hat{P}_{ij}^{(1)} - H_{ij}^{(1)}\right|^s\right] \left(H_{ij}^{(1)}\right)^{k-s} \\
&\leq 2^k \sum_{s=0}^k \binom{k}{s} R^s s! \left(H_{ij}^{(1)}\right)^{k-s} \\
&\leq 2^k k! \sum_{s=0}^k \binom{k}{s} R^s \left(H_{ij}^{(1)}\right)^{k-s} \\
&= 2^k k! \left(R + H_{ij}^{(1)}\right)^k \\
&\leq 2^{2k} k! R^k.
\end{aligned} \tag{5}$$

Therefore we have

$$P\left(\|\hat{P}^{(q)} - H^{(q)}\| \geq t\right) \leq n \exp\left(-\frac{t^2/2}{32R^2n + Rt}\right).$$

Now let $c > 0$ be given and assume $n^{-c} \leq \eta \leq 1/2$. Then there exists a $n_0(c)$ independent of n , P , C and ϵ such that whenever $n > n_0(c)$,

$$t = 8R\sqrt{2n \ln(n/\eta)} \leq 32Rn.$$

Plugging this t into the equation above, we get

$$P(\|\hat{P}^{(q)} - H^{(q)}\| \geq 8R\sqrt{2n \ln(n/\eta)}) \leq n \exp\left(-\frac{t^2}{64R^2n}\right) = \eta.$$

■

As we define $H^{(q)} = E[\hat{P}^{(q)}]$, let $d^{(q)} = \text{rank}(H^{(q)})$ be the dimension in which we are going to embed $\hat{P}^{(q)}$. Notice that it is less than or equal to $K \times K'$ based on the SBM assumption. Then we can define $H^{(q)} = ZZ^T$ where $Z \in \mathbb{R}^{n \times d^{(q)}}$.

For simplicity, from now on, we will use \hat{P} to represent $\hat{P}^{(q)}$, use H to represent $H^{(q)}$ and use k to represent the dimension $d^{(q)}$ we are going to embed. Assume $H = USU^T = ZZ^T$, where $Z = [Z_1, \dots, Z_n]^T$ is a n -by- k matrix. Then our estimate for Z up to rotation is $\hat{Z} = \hat{U}\hat{S}^{1/2}$, where $\hat{U}\hat{S}\hat{U}^T$ is the rank- d spectral decomposition of $|\hat{P}| = (\hat{P}^T \hat{P})^{1/2}$.

Furthermore, we assume that the second moment matrix $E[Z_1 Z_1^T]$ is rank k and has distinct eigenvalues $\lambda_i(E[Z_1 Z_1^T])$. In particular, we assume that there exists $\delta > 0$ such that

$$\delta < \lambda_k(E[Z_1 Z_1^T])$$

Lemma 3.21 *Under the above assumptions, $\lambda_i(H) = \Theta(n)$ with high probability when $i \leq k$, i.e. the largest k eigenvalues of H is of order n . Moreover, we have $\|S\|_2 = \Theta(n)$ and $\|\hat{S}\|_2 = \Theta(n)$ with high probability.*

Remark: This is an extended version of Proposition 4.3 in [8].

Require: Theorem 3.20 and assumptions above

Proof: Exactly the same as proof for Lemma 3.9. ■

Lemma 3.22 *Let $W_1 \Sigma W_2^T$ be the singular value decomposition of $U^T \hat{U}$. Then for sufficiently large n ,*

$$\|U^T \hat{U} - W_1 W_2^T\|_F = O(n^{-1} \log n)$$

with high probability.

Require: Theorem 3.20 and Lemma 3.21

Proof: Exactly the same as proof for Lemma 3.10. ■

We will denote the orthogonal matrix $W_1 W_2^T$ by W^* .

Lemma 3.23 *For sufficiently large n ,*

$$\|W^* \hat{S} - S W^*\|_F = O(\log n),$$

$$\|W^* \hat{S}^{1/2} - S^{1/2} W^*\|_F = O(n^{-1/2} \log n)$$

and

$$\|W^* \hat{S}^{-1/2} - S^{-1/2} W^*\|_F = O(n^{-3/2} \log n)$$

with high probability.

Require: Theorem 3.20, Lemma 3.21, Lemma 3.22, Exponential distribution

Proof: Similar to the proof of Lemma 3.11. ■

Lemma 3.24 *There exists a rotation matrix W such that for sufficiently large n ,*

$$\|\hat{Z} - ZW\|_F = \|(\hat{P} - H)US^{-1/2}\|_F + O(n^{-1/2}(\log n)^{3/2})$$

with high probability.

Require: Lemma 3.22, Lemma 3.23

Proof: Exactly the same as proof for Lemma 3.12. ■

Theorem 3.25 *There exists a rotation matrix W such that for sufficiently large n ,*

$$\max_i \|\hat{Z}_i - W Z_i\|_2 = O(n^{-1/2}(\log n)^{3/2})$$

with high probability.

Require: Lemma 3.24, Exponential distribution

Proof: Similar to the proof of Theorem 3.13. ■

Lemma 3.26 $\left| \hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j \right| = O(n^{-1}(\log n)^3)$ *with high probability.*

Require: Theorem 3.25

Proof: Similar to the proof of Lemma 3.14. ■

Definition 3.27 *Define $\tilde{P}_{ij}^{(q)} = (\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}$, our estimator for P_{ij} , to be a projection of $\hat{Z}_i^T \hat{Z}_j$ onto $[0, \min(\hat{P}_{ij}^{(q)}, R)]$.*

Corollary 3.28 *Assuming that $m = O(n^b)$ for any $b > 0$, then the estimator based on ASE of MLqE has the same entry-wise asymptotic bias as MLqE, i.e.*

$$\lim_{n \rightarrow \infty} \text{Bias}(\tilde{P}_{ij}^{(q)}) = \lim_{n \rightarrow \infty} E[\tilde{P}_{ij}^{(q)}] - P_{ij} = \lim_{n \rightarrow \infty} E[\hat{P}_{ij}^{(q)}] - P_{ij} = \lim_{n \rightarrow \infty} \text{Bias}(\hat{P}_{ij}^{(q)}).$$

Proof: Fix some $a > 0$, we have

$$\begin{aligned} & E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j|] \\ &= E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\}] + E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}]. \end{aligned}$$

Note that we are thresholding according to $\hat{P}^{(1)}$ instead of $\hat{P}^{(q)}$. By Lemma 3.1, we know $\hat{P}^{(q)} < \hat{P}^{(1)}$ given any data. For the first term, we have

$$\begin{aligned} & E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\}] \\ & \leq E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 holds}\}] \\ & \quad + E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 does not hold}\}]] \\ & \leq E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} | \{\text{Lemma 3.26 holds}\}] \\ & \quad + n^{-c} E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} | \{\text{Lemma 3.26 does not hold}\}] \\ & \leq O(n^{-1}(\log n)^3) \\ & \quad + n^{-c} E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - \hat{P}_{ij}^{(q)}| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} | \{\text{Lemma 3.26 does not hold}\}] \\ & \quad + n^{-c} E[|\hat{P}_{ij}^{(q)} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} | \{\text{Lemma 3.26 does not hold}\}] \\ & \leq O(n^{-1}(\log n)^3) \\ & \quad + n^{-c} E[|\hat{P}_{ij}^{(q)}| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} | \{\text{Lemma 3.26 does not hold}\}] \\ & \quad + n^{-c} E[|(\hat{P}_{ij}^{(q)} + R)| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} | \{\text{Lemma 3.26 does not hold}\}] \\ & \leq O(n^{-1}(\log n)^3) \\ & \quad + n^{-c} E[|\hat{P}_{ij}^{(1)}| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} | \{\text{Lemma 3.26 does not hold}\}] \\ & \quad + n^{-c} E[|(\hat{P}_{ij}^{(1)} + R)| \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} | \{\text{Lemma 3.26 does not hold}\}] \\ & \leq O(n^{-1}(\log n)^3) + an^{-c} + (a + R)n^{-c} \\ & \leq O(n^{-1}(\log n)^3) + 2n^{-c}(a + R). \end{aligned}$$

For the second term, we have

$$\begin{aligned} & E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] \\ & \leq E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - \hat{P}_{ij}^{(q)}| \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] + E[|\hat{P}_{ij}^{(q)} - Z_i^T Z_j| \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] \\ & \leq E[|\hat{P}_{ij}^{(q)}| \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] + E[|(\hat{P}_{ij}^{(q)} + R)| \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] \\ & \leq E[|\hat{P}_{ij}^{(1)}| \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] + E[|(\hat{P}_{ij}^{(1)} + R)| \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] \\ & \leq 2e^{-a/R}(a + 2mR). \end{aligned}$$

Similarly, assuming $m = O(n^b)$ for any $b > 0$, we have

$$E[|(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j|] = O(n^{-1}(\log n)^3).$$

■

Theorem 3.29 Assuming that $m = O(n^b)$ for any $b > 0$, then $\text{Var}((\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}) = O(n^{-2}(\log n)^6)$.

Proof: By Lemma 3.26,

$$\begin{aligned}
\text{Var}((\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}) &= E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}]]^2 \\
&= E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j + Z_i^T Z_j - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}]]^2 \\
&= E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 + E[(Z_i^T Z_j - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}])]^2 \\
&\quad + 2E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j](Z_i^T Z_j - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}]) \\
&\leq E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 + E[(Z_i^T Z_j - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}])]^2 \\
&\quad + 2\sqrt{E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 E[(Z_i^T Z_j - E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}])]^2} \\
&\leq 4E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2.
\end{aligned}$$

Fix some $a > 0$, we have

$$\begin{aligned}
&E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \\
&= E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} + E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}.
\end{aligned}$$

Note that we are thresholding according to $\hat{P}^{(1)}$ instead of $\hat{P}^{(q)}$. By Lemma 3.1, we know $\hat{P}^{(q)} < \hat{P}^{(1)}$ given any data. For the first term, we have

$$\begin{aligned}
&E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \\
&\leq E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 holds}\} \\
&\quad + E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 does not hold}\} \\
&\leq E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 holds}\} \\
&\quad + n^{-c} E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j]^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 does not hold}\} \\
&\leq O(n^{-2}(\log n)^6) \\
&\quad + 2n^{-c} E[(\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - \hat{P}_{ij}^{(q)}]^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 does not hold}\} \\
&\quad + 2n^{-c} E[(\hat{P}_{ij}^{(q)} - Z_i^T Z_j)^2] \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 does not hold}\} \\
&\leq O(n^{-2}(\log n)^6) \\
&\quad + 2n^{-c} E[\hat{P}_{ij}^{(q)2}] \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 does not hold}\} \\
&\quad + 2n^{-c} E[(\hat{P}_{ij}^{(q)} + R)^2] \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 does not hold}\} \\
&\leq O(n^{-2}(\log n)^6) + 2n^{-c} E[\hat{P}_{ij}^{(1)2}] \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 does not hold}\} \\
&\quad + 2n^{-c} E[(\hat{P}_{ij}^{(1)} + R)^2] \mathbb{I}\{\hat{P}_{ij}^{(1)} \leq a\} \mathbb{I}\{\text{Lemma 3.26 does not hold}\} \\
&\leq O(n^{-2}(\log n)^6) + 2a^2 n^{-c} + 2(a + R)^2 n^{-c} \\
&\leq O(n^{-2}(\log n)^6) + 4n^{-c}(a + R)^2.
\end{aligned}$$

For the second term, we have

$$\begin{aligned}
& E[((\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - Z_i^T Z_j)^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] \\
& \leq 2E[((\hat{Z}_i^T \hat{Z}_j)_{\text{tr}} - \hat{P}_{ij}^{(q)})^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] + 2E[(\hat{P}_{ij}^{(q)} - Z_i^T Z_j)^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] \\
& \leq 2E[\hat{P}_{ij}^{(q)2} \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] + 2E[(\hat{P}_{ij}^{(q)} + R)^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] \\
& \leq 2E[\hat{P}_{ij}^{(1)2} \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] + 2E[(\hat{P}_{ij}^{(1)} + R)^2 \mathbb{I}\{\hat{P}_{ij}^{(1)} > a\}] \\
& \leq 4e^{-a/R}(a + 2m^{1/2}R)^2.
\end{aligned}$$

Similarly, assuming $m = O(n^b)$ for any $b > 0$, we have

$$\text{Var}((\hat{Z}_i^T \hat{Z}_j)_{\text{tr}}) = O(n^{-2}(\log n)^6).$$

■

Theorem 3.30 Let $u_q(\theta) = E_\theta[\hat{\theta}_{q,n}]$, $\phi_q(x; \theta) = \frac{\partial}{\partial \theta} L_q(f(x; \theta))$, and $\phi'_q(x; \theta) = \frac{\partial^2}{\partial \theta^2} L_q(f(x; \theta))$. Then the asymptotic distribution of $\hat{\theta}_{q,n}$ is $\sqrt{n}(\hat{\theta}_{q,n} - u_q(\theta)) \sim \mathcal{N}(0, V_q(\theta))$, where $V_q(\theta) = E[\phi_q(X; \theta)^2] / E[\phi'_q(X; \theta)]^2$.

Remark: See Theorem 1 in <http://arxiv.org/pdf/1310.7278.pdf>.

Corollary 3.31 $\text{Var}(\hat{P}_{ij}^{(q)}) = \Theta(m^{-1})$.

Proof: Direct result from Theorem 3.30. **NEED PROOF HERE.**

■

Theorem 3.32 For fixed m , $1 \leq i, j \leq n$,

$$\frac{\text{Var}(\tilde{P}_{ij}^{(q)})}{\text{Var}(\hat{P}_{ij}^{(q)})} = O(mn^{-2}(\log n)^6).$$

Thus

$$\text{ARE}(\hat{P}_{ij}^{(q)}, \tilde{P}_{ij}^{(q)}) = 0.$$

Furthermore, as long as m goes to infinity of order $o(n^2(\log n)^{-6})$,

$$\text{ARE}(\hat{P}_{ij}^{(q)}, \tilde{P}_{ij}^{(q)}) = 0.$$

Proof: The results are direct from Theorem 3.29 and Corollary 3.31.

■

3.4 $\tilde{P}^{(q)}$ better than $\tilde{P}^{(1)}$

Theorem 3.33 For sufficiently large n and C , any $1 \leq i, j \leq n$,

$$\lim_{m \rightarrow \infty} \text{Bias}(\tilde{P}_{ij}^{(1)}) > \lim_{m \rightarrow \infty} \text{Bias}(\tilde{P}_{ij}^{(q)})$$

Proof: Direct result from Lemma 3.5, Corollary 3.16 and Corollary 3.28.

■

Theorem 3.34 For any fixed m , any $1 \leq i, j \leq n$,

$$\lim_{n \rightarrow \infty} \text{Var}(\tilde{P}_{ij}^{(1)}) = \lim_{n \rightarrow \infty} \text{Var}(\tilde{P}_{ij}^{(q)}) = 0.$$

Furthermore, as long as m goes to infinity of order $o(n^2(\log n)^{-6})$, any $1 \leq i, j \leq n$,

$$\lim_{n \rightarrow \infty} \text{Var}(\tilde{P}_{ij}^{(1)}) = \lim_{n \rightarrow \infty} \text{Var}(\tilde{P}_{ij}^{(q)}) = 0$$

Proof: Direct result from Theorem 3.17 and Theorem 3.29.

■

References

- [1] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.
- [2] Davide Ferrari and Yuhong Yang. Maximum lq-likelihood estimation. *Ann. Statist.*, 38(2):753–783, 04 2010.
- [3] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.
- [4] Whitney K Newey and Daniel McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994.
- [5] Roberto Imbuzeiro Oliveira. Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*, 2009.
- [6] Yichen Qin and Carey E Priebe. Maximum l q-likelihood estimation via the expectation-maximization algorithm: A robust estimation of mixture models. *Journal of the American Statistical Association*, 108(503):914–928, 2013.
- [7] Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, pages 1878–1915, 2011.
- [8] Daniel L Sussman, Minh Tang, and Carey E Priebe. Consistent latent position estimation and vertex classification for random dot product graphs. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(1):48–57, 2014.
- [9] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.