A Proof

To prove Theorem 1, we introduce the following lemma first.

Lemma 1. *Let $C_2, C_3, C_4, C_5 > 0$ be arbitrary global constants and assume n is large enough so that $1/n^2 + (C_4+8)^{1/2}/n \le \left(C_5 - (C_2+3)^{1/2} - (C_3+3)^{1/2}\right)\left(n^{-3}\log n\right)^{1/4}$, then for any neighborhood sets $\mathbf{S} = \{S_1, \dots, S_n\}$, with probability $1 - 2n^{-2C_2/3} - 2n^{-2C_3/3} - 2n^{-C_4/2}$,*

$$\frac{1}{n^2 s^4}\sum_i \sum_j \left(\sum_{i' \in S_i}\sum_{i' \in S_i}(P_{i'j'} - A_{i'j'})\right)^2 \le C_5\left(\frac{\log n}{n^3}\right)^{1/4}.$$

Proof of Lemma 1. The summand satisfies

$$\begin{aligned}
&\left(\sum_{i' \in S_i}\sum_{j' \in S_j}(P_{i'j'} - A_{i'j'})\right)^2 \\
=&\sum_{i' \in S_i}\sum_{j' \in S_j}(P_{i'j'} - A_{i'j'})^2 + \\
&\sum_{i' \in S_i}\sum_{i'' \in S_i, i'' \neq i'}\sum_{j' \in S_j}(P_{i'j'} - A_{i'j'})(P_{i''j'} - A_{i''j'}) + \\
&\sum_{i' \in S_i}\sum_{j' \in S_j}\sum_{j'' \in S_j, j'' \neq j'}(P_{i'j'} - A_{i'j'})(P_{i'j''} - A_{i'j''}) + \\
&\sum_{i' \in S_i}\sum_{i'' \in S_i, i'' \neq i'}\sum_{j' \in S_j}\sum_{j'' \in S_j, j'' \neq j'}(P_{i'j'} - A_{i'j'})(P_{i''j''} - A_{i''j''}) \\
=&E_1(i,j) + E_2(i,j) + E_3(i,j) + E_4(i,j).
\end{aligned} \tag{1}$$

The first term in (1) satisfies

$$\sum_{i' \in S_i}\sum_{j' \in S_j}(P_{i'j'} - A_{i'j'})^2 \le \sum_{i' \in S_i}\sum_{j' \in S_j} 1 = s^2,$$

so $(n^2s^2)^{-1}\sum_i\sum_j E_1(i,j) \le 1/n^2$. The term $(n^2s^2)^{-1}\sum_i\sum_j E_2(i,j)$ can be bounded by

$$\begin{aligned}
&\frac{1}{n^2 s_i^2 s_j^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{i' \in S_i}\sum_{i'' \in S_i, i'' \neq i'}\sum_{j' \in S_j}(P_{i'j'} - A_{i'j'})(P_{i''j'} - A_{i''j'}) \\
\le&\frac{1}{n^2 s_i^2 s_j}\sum_{i=1}^n\sum_{j=1}^n\sum_{i', i'' \in S_i : i' \neq i''}\frac{1}{s_j}\left|\sum_{j' \in S_j}(P_{i'j'} - A_{i'j'})(P_{i''j'} - A_{i''j'})\right|.
\end{aligned} \tag{2}$$

Note that we have assume that $P_{ii} = 0$ for all $i \in V$, so there is no need to consider the cases where $j = i'$ of $j = i''$. To bound (2), for any $i_1 \neq i_2$ and $0 < \varepsilon < 1$, by Bernstein's inequality we have

$$\Pr\left\{\frac{1}{s_j}\left|\sum_{j'}(P_{i_1j'} - A_{i_1j'})(P_{i_2j'} - A_{i_2j'})\right| \ge \varepsilon\right\} \le 2\exp\left\{-\frac{\frac{1}{2}s_j\varepsilon^2}{1 + \frac{1}{3}\varepsilon}\right\} \le 2e^{-\frac{s_j\varepsilon^2}{3}} \le 2e^{-\frac{(n\log n)^{1/2}\varepsilon^2}{3}},$$

due to $s_j = s > (n\log n)^{1/2}$. Then, with arbitrary global constants C_2 and $n > 9 > e^2$, by taking $\varepsilon = \sqrt{(C_2+3)\log n (n^{-1}\log n)^{1/2}}$ and a union bound over all $i_1 \neq i_2$, we have

$$\Pr\left\{\max_{i_1,i_2}\frac{1}{s_j}\left|\sum_{j'}(P_{i_1j'} - A_{i_1j'})(P_{i_1j'} - A_{i_2j'})\right| \ge \varepsilon\right\} \le 2n^2\exp\left\{-\frac{(n\log n)^{1/2}\varepsilon^2}{3}\right\} < 2n^{-2C_2/3}.$$

Then, with probability $1-2n^{-2C_2/3}$, for all (i,j) simultaneously, we have

$$
\begin{aligned}
&\frac{1}{n^2 s_i^2 s_j^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{i'\in S_i}\sum_{i''\in S_i, i''\neq i'}\sum_{j'\in S_j}(P_{i'j'}-A_{i'j'})(P_{i''j'}-A_{i''j'})\\
\leq&\frac{1}{n^2 s_i^2 s_j} n^2 s_i(s_i-1)\sqrt{(C_2+3)\log n (n^{-1}\log n)^{1/2}}\\
\leq&\frac{\sqrt{(C_2+3)\log n (n^{-1}\log n)^{1/2}}}{(n\log n)^{1/2}}=(C_2+3)^{1/2}\left(\frac{\log n}{n^3}\right)^{1/4}.
\end{aligned}
\tag{3}
$$

The bound of the term $(n^2 s^2)^{-1}\sum_i\sum_j E_3(i,j)$ can be derived in the same way. That is, with probability $1-2n^{-2C_3/3}$, for all (i,j) simultaneously, we have

$$
\frac{1}{n^2 s_i^2 s_j^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{i'\in S_i}\sum_{j'\in S_j}\sum_{j''\in S_j, j''\neq j'}(P_{i'j'}-A_{i'j'})(P_{i'j''}-A_{i'j''})\leq (C_3+3)^{1/2}\left(\frac{\log n}{n^3}\right)^{1/4}.
\tag{4}
$$

As to the fourth term $E_4(i,j)$ which consists of $s_i(s_i-1)s_j(s_j-1)$ summands, for any $(i_1,j_1)\neq(i_2,j_2)$ and $0<\varepsilon<1$, if $n>9$ so that $(n\log n)^{1/2}>3$ and $\log n>2$, by Bernstein's inequality, we have

$$
\begin{aligned}
&\Pr\left\{\frac{1}{s_i(s_i-1)s_j(s_j-1)}\left|\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}(P_{i_1j_1}-A_{i_1j_1})(P_{i_2j_2}-A_{i_2j_2})\right|\right\}\\
\leq&2\exp\left\{-\frac{\frac{1}{2}s_i(s_i-1)s_j(s_j-1)\varepsilon^2}{1+\frac{1}{3}\varepsilon}\right\}\\
\leq&2\exp\left\{-\frac{s_i^4\varepsilon^2}{4}\right\}\\
\leq&2\exp\left\{-\frac{n^2(\log n)^2\varepsilon^2}{4}\right\}.
\end{aligned}
\tag{5}
$$

Then, with arbitrary global constants C_4, by taking $\varepsilon=(C_4+8)^{1/2}/n$ and a union bound over all $(i_1,j_1)\neq(i_2,j_2)$, we have

$$
\begin{aligned}
&\Pr\left\{\max_{(i_1,j_1),(i_2,j_2)}\frac{1}{s_i(s_i-1)s_j(s_j-1)}\left|\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}(P_{i_1j_1}-A_{i_1j_1})(P_{i_2j_2}-A_{i_2j_2})\right|\geq\varepsilon\right\}\\
\leq&2n^4\exp\left\{-\frac{n^2(\log n)^2\varepsilon^2}{4}\right\}<2n^{-C_4/2}.
\end{aligned}
\tag{6}
$$

Then, with probability $1-2n^{-C_4/2}$, for all (i,j) simultaneously, we have

$$
\begin{aligned}
&\frac{1}{n^2 s_i^2 s_j^2}\sum_{i}\sum_{j}\sum_{i',i''\in S_i: i'\neq i''}\sum_{j',j''\in S_j: j'\neq j''}(P_{i'j'}-A_{i'j'})(P_{i''j''}-A_{i''j''})\\
\leq&\frac{1}{n^2}\sum_{i}\sum_{j}\frac{1}{s_i(s_i-1)s_j(s_j-1)}\left|\sum_{i',i''\in S_i: i'\neq i''}\sum_{j',j''\in S_j: j'\neq j''}(P_{i'j'}-A_{i'j'})(P_{i''j''}-A_{i''j''})\right|\\
\leq&\frac{(C_4+8)^{1/2}}{n}.
\end{aligned}
\tag{7}
$$

Then by plugging (3), (4) and (7) into (1), with probability $1-2n^{-2C_2/3}-2n^{-2C_3/3}-2n^{-C_4/2}$, we have

$$\begin{aligned}
&\frac{1}{n^2s^2}\sum_i\sum_j\left(\sum_{i'\in S_i}\sum_{i'\in S_i}\left(P_{i'j'}-A_{i'j'}\right)\right)^2\\
\le&\frac{1}{n^2}+\left((C_2+3)^{1/2}+(C_3+3)^{1/2}\right)\left(\frac{\log n}{n^3}\right)^{1/4}+\frac{(C_4+8)^{1/2}}{n} \qquad (8)\\
\le&C_5\left(\frac{\log n}{n^3}\right)^{1/4}.
\end{aligned}$$

□

Proof of Theorem 1. We begin the proof with the following decomposition of the error term:

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(\widetilde{P}_{ij}-P_{ij}\right)^2\\
=&\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(\widetilde{P}_{ij}-P^*_{ij}+P^*_{ij}-P_{ij}\right)^2 \qquad (9)\\
\le&\frac{2}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(\widetilde{P}_{ij}-P^*_{ij}\right)^2+\frac{2}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(P^*_{ij}-P_{ij}\right)^2,
\end{aligned}$$

where P^*_{ij} is defined as

$$P^*_{ij}=\frac{\sum_{i'\in S^*_i}\sum_{j'\in S^*_j}P_{i'j'}}{s_is_j}.$$

For the first term, according to Lemma 1, with probability $1-2n^{-2C_2/3}-2n^{-2C_3/3}-2n^{-C_4/2}$, we have

$$\begin{aligned}
&\frac{2}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(\widetilde{P}_{ij}-P^*_{ij}\right)^2\\
=&\frac{2}{n}\sum_{i=1}^n\sum_{j=1}^n\left(\frac{\sum_{i'\in S^*_i}\sum_{j'\in S^*_j}A_{i'j'}}{|S^*_i||S^*_j|}-\frac{\sum_{i'\in S^*_i}\sum_{j'\in S^*_j}P_{i'j'}}{|S^*_i||S^*_j|}\right)^2\\
=&\frac{2}{n^2s^4}\sum_{i=1}^n\sum_{j=1}^n\left(\sum_{i'\in S^*_i}\sum_{j'\in S^*_j}\left(A_{i'j'}-P_{i'j'}\right)\right)^2 \qquad (10)\\
\le&2C_5\left(\frac{\log n}{n^3}\right)^{1/4}.
\end{aligned}$$

For the second term, we have

$$
\begin{aligned}
&\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(P_{ij}^{*}-P_{ij}\right)^2 \\
=&\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\sum_{i'\in S_i^*}\sum_{j'\in S_j^*}P_{i'j'}}{|S_i^*||S_j^*|}-\frac{\sum_{i'\in S_i^*}\sum_{j'\in S_j^*}P_{ij}}{|S_i^*||S_j^*|}\right)^2 \\
=&\frac{2}{n^2 s^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{i'\in S_i^*}\sum_{j'\in S_j^*}(P_{i'j'}-P_{ij})\right)^2 \\
\leq&\frac{2s^2}{n^2 s^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{i'\in S_i^*}\sum_{j'\in S_j^*}(P_{i'j'}-P_{ij})^2 \\
\leq&\frac{2}{n^2 s^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{i'\in S_i^*}\sum_{j'\in S_j^*}4L^2\Delta_n^2 \\
=&8L^2\left[1+(C_1+4)^{1/2}\right]\frac{\log n}{n}.
\end{aligned}
\tag{11}
$$

Then, combining with (10) and (11), with probability $1-2n^{-C_1/4}-2n^{-2C_2/3}-2n^{-2C_3/3}-2n^{-C_4/2}$, we have

$$
\frac{||\widetilde{\mathbf{P}}-\mathbf{P}||_F^2}{n^2}\leq C_6\left(\frac{\log n}{n^3}\right)^{1/4}.
$$

□

Proof of Theorem 2. Again, we decompose the error term as follows:

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widehat{P}_{ij}-P_{ij}\right)^2 \\
=&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widehat{P}_{ij}-P_{ij}^{S}+P_{ij}^{S}-P_{ij}\right)^2 \\
\leq&\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widehat{P}_{ij}-P_{ij}^{S}\right)^2+\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(P_{ij}^{S}-P_{ij}\right)^2,
\end{aligned}
\tag{12}
$$

where P_{ij}^{S} here is defined as

$$
P_{ij}^{S}=\frac{\sum_{i'\in S_i}\sum_{j'\in S_j}P_{i'j'}}{s_i s_j}.
$$

For the first term, according to Lemma 1, with probability $1-2n^{-2C_2/3}-2n^{-2C_3/3}-2n^{-C_4/2}$, we have

$$
\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widehat{P}_{ij}-P_{ij}^{S}\right)^2\leq 2C_5\left(\frac{\log n}{n^3}\right)^{1/4}. \tag{13}
$$

For the second term, we have

$$
\begin{aligned}
& \frac{2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(P_{ij}^S - P_{ij}\right)^2 \\
=& \frac{2}{n^2 s^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\sum_{i' \in S_i} \sum_{j' \in S_j} (P_{i'j'} - P_{ij}) \right)^2 \\
=& \frac{2}{n^2 s^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\sum_{i' \in S_i^*, j' \in S_j^*} (P_{i'j'} - P_{ij}) + \sum_{i' \notin S_i^* \vee j' \notin S_j^*} (P_{i'j'} - P_{ij}) \right]^2 \\
\leq& \frac{4}{n^2 s^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\left(\sum_{i' \in S_i^*, j' \in S_j^*} (P_{i'j'} - P_{ij}) \right)^2 + \left(\sum_{i' \notin S_i^* \vee j' \notin S_j^*} (P_{i'j'} - P_{ij}) \right)^2 \right] \\
\leq& \frac{4s^2}{n^2 s^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i' \in S_i^*, j' \in S_j^*} (P_{i'j'} - P_{ij})^2 + \frac{4e(n)}{n^2 s^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i' \notin S_i^* \vee j' \notin S_j^*} (P_{i'j'} - P_{ij})^2 \\
\leq& \frac{4}{n^2 s^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i' \in S_i^*, j' \in S_j^*} 4L^2 \Delta_n^2 + \frac{4e(n)}{n^2 s^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i' \notin S_i^* \vee j' \notin S_j^*} (b-a)^2 \\
\leq& 16L^2 \left[1 + (C_1+4)^{1/2}\right] \frac{\log n}{n} + \frac{4e^2(n)}{s^4} (b-a)^2 \\
\leq& 16L^2 \left[1 + (C_1+4)^{1/2}\right] \frac{\log n}{n} + \frac{4e^2(n)}{n^2 (\log n)^2} (b-a)^2 \\
=& 16L^2 \left[1 + (C_1+4)^{1/2}\right] \frac{\log n}{n} + 4C_7 \left(\frac{\log n}{n^3} \right)^{1/4} .
\end{aligned}
\tag{14}
$$

Then, combining with (13) and (14), with probability $1 - 2n^{-C_1/4} - 2n^{-2C_2/3} - 2n^{-2C_3/3} - 2n^{-C_4/2}$, we have

$$
\frac{||\widehat{\mathbf{P}} - \mathbf{P}||_F^2}{n^2} \leq C_8 \left(\frac{\log n}{n^3} \right)^{1/4} .
$$

$\square$

According to Theorem 2, for each pair (i, j), since $|S_i \times S_j| = s^2$, the max error rate allowed is $e(n)/s^2 \approx e(n)/n \log n = \sqrt{C_7} (n^{-3} \log n)^{1/8} / (b-a)$, which is much larger then $C_8 (n^{-3} \log n)^{1/4}$. That is, we can obtain an estimate with low error rate even with relatively high error rate on neighborhood selection. And the smaller $b-a$ is, the larger the error rate is allowed. Let $C_5 = 4$, $C_7 = 1$ and $C_8 = 15$, we display the curves of the error rate on neighborhood selection against the error rate on estimation with n varying from $1,000$ to $1000,000$ in Figure 1. It is obvious that the former is always much larger then the latter. For example, if $n = 1000$ and $b - a = 0.5$, to achieve an estimate with 0.137 error rate, it allows $S_i \times S_j$ to include about 20% wrongly assigned pairs for each (i, j).

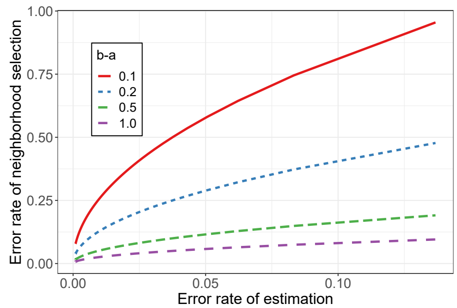


Figure 1: *The curves of the error rate on neighborhood selection against the error rate on estimation.*

Proof of Theorem 3. We first calculate the pairwise distances based on $\widehat{\mathbf{P}}^{(0)}$. For $i' \in S_i^*$, with probability $1 - 2n^{-C_1/4} - n^{-C_{10}}$, we have

$$\begin{aligned} d(i,i') &= \frac{1}{n}\sum_{j=1}^{n}(\widehat{P}_{ij}^{(0)} - \widehat{P}_{i'j}^{(0)})^2 \\ &\le \frac{4}{n}\sum_{j=1}^{n}(P_{ij} - P_{i'j})^2 + \frac{2}{n}\sum_{j=1}^{n}(P_{ij} - \widehat{P}_{ij}^{(0)})^2 + \frac{4}{n}\sum_{j=1}^{n}(P_{i'j} - \widehat{P}_{i'j}^{(0)})^2 \\ &\le 4L^2\Delta_n^2 + 6\max_{i\in V}\frac{1}{n}\sum_{j=1}^{n}(P_{ij} - \widehat{P}_{ij}^{(0)})^2 \\ &\le 4L^2\left[1 + (C_1+4)^{1/2}\right]^2\frac{\log n}{n} + 6C_9E(n). \end{aligned} \tag{15}$$

For any $i'' \notin S_i^*$, with probability $1 - 2n^{-C_1/4}$, we have

$$\begin{aligned} d(i,i'') &= \frac{1}{n}\sum_{j=1}^{n}(\widehat{P}_{ij}^{(0)} - \widehat{P}_{i''j}^{(0)})^2 \\ &\ge \frac{1}{2n}\sum_{j=1}^{n}(P_{ij} - P_{i''j})^2 - \frac{2}{n}\sum_{j=1}^{n}(P_{ij} - \widehat{P}_{ij}^{(0)})^2 - \frac{2}{n}\sum_{j=1}^{n}(P_{i''j} - \widehat{P}_{i''j}^{(0)})^2 \\ &\ge \frac{1}{2}C^2(n) - 4\max_{i\in V}\frac{1}{n}\sum_{j=1}^{n}(P_{ij} - \widehat{P}_{ij}^{(0)})^2 \\ &\ge \frac{1}{2}C^2(n) - 4C_9E(n). \end{aligned} \tag{16}$$

Then, due to $C^2(n) \ge 8L^2\left[1 + (C_1+4)^{1/2}\right]^2(n^{-1}\log n) + 20C_9E(n)$, we can deduce that $d(i,i') \le d(i,i'')$ for any $i \in V, i' \in S_i^*, i'' \notin S_i^*$. That is, with probability $1 - 2n^{-C_1/4} - n^{-C_{10}}$, one can select all the true neighbors for each vertex i based on $\widehat{\mathbf{P}}$. Then, combining with Theorem 1, with probability $1 - 2n^{-C_1/4} - 2n^{-2C_2/3} - 2n^{-2C_3/3} - 2n^{-C_4/2} - n^{-C_{10}}$, we have

$$\frac{||\widehat{\mathbf{P}}_{new} - \mathbf{P}||_F^2}{n^2} \le C_6\left(\frac{\log n}{n^3}\right)^{1/4}.$$

□

It should be noted that the lower bound $C(n)$ we define is expected to be small. Indeed, if it is equal to $L\Delta_n$, then we are able to differentiate the true neighbors of each vertex i even from all the vertexes in $V \setminus S_i^*$. However, because $E(n)$ is always much greater then $n^{-1}\log n$, the pairwise distances calculated on the estimate is also larger than that defined on $\mathbf{P}$, which leads to a large $C(n)$. Nevertheless, the lower bound of $C(n)$ is allowed to go to 0 as $n \to 0$, making $C(n)$ get close to $L\Delta_n$ as we expect.

Proof of Theorem 4. We begin the proof with the following decomposition of the error term:

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widehat{P}_{ij}^{(m)}-\widehat{P}_{ij}^{(m+1)}\right)^2\\
=&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widehat{P}_{ij}^{(m)}-\frac{\sum_{i'\in S_i}\sum_{j'\in S_j}A_{i'j'}}{s_is_j}\right)^2\\
=&\frac{1}{n^2s_i^2s_j^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\sum_{i'\in S_i}\sum_{j'\in S_j}\left(\widehat{P}_{ij}^{(m)}-A_{i'j'}\right)\right]^2\\
=&\frac{1}{n^2s_i^2s_j^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\sum_{i'\in S_i}\sum_{j'\in S_j}\left(\widehat{P}_{ij}^{(m)}-\widehat{P}_{i'j'}^{(m)}+\widehat{P}_{i'j'}^{(m)}-P_{i'j'}+P_{i'j'}-A_{i'j'}\right)\right]^2.
\end{aligned}
\tag{17}
$$

We can bound the summand by

$$
\begin{aligned}
&\left[\sum_{i'\in S_i}\sum_{j'\in S_j}\left(\widehat{P}_{ij}^{(m)}-\widehat{P}_{i'j'}^{(m)}+\widehat{P}_{i'j'}^{(m)}-P_{i'j'}+P_{i'j'}-A_{i'j'}\right)\right]^2\\
\leq&4\left\{\sum_{i'\in S_i}\sum_{j'\in S_j}\left(\widehat{P}_{ij}^{(m)}-\widehat{P}_{i'j'}^{(m)}\right)\right\}^2+2\left\{\sum_{i'\in S_i}\sum_{j'\in S_j}\left(\widehat{P}_{i'j'}^{(m)}-P_{i'j'}\right)\right\}^2+4\left\{\sum_{i'\in S_i}\sum_{j'\in S_j}\left(P_{i'j'}-A_{i'j'}\right)\right\}^2\\
=&4E_5(i,j)+2E_6(i,j)+4E_7(i,j).
\end{aligned}
$$

Our goal is to bound $(n^3\log n)^{-1}\sum_i\sum_j\{4E_5(i,j)+2E_6(i,j)+4E_7(i,j)\}$. For the first term, due to $|\widehat{P}_{ij}^{(m)}-\widehat{P}_{i'j'}^{(m)}|\leq 2L\Delta_n$, we have

$$
\begin{aligned}
&\frac{4}{n^2s_i^2s_j^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\sum_{i'\in S_i}\sum_{j'\in S_j}\left(\widehat{P}_{ij}^{(m)}-\widehat{P}_{i'j'}^{(m)}\right)\right]^2\\
\leq&\frac{4}{n^2s_is_j}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{i'\in S_i}\sum_{j'\in S_j}\left(\widehat{P}_{ij}^{(m)}-\widehat{P}_{i'j'}^{(m)}\right)^2\\
\leq&\frac{4}{n^2s_is_j}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{i'\in S_i}\sum_{j'\in S_j}4L^2\Delta_n^2\\
=&16L^2\left[1+(C_1+4)^{1/2}\right]^2\frac{\log n}{n}.
\end{aligned}
\tag{18}
$$

For the second term, it is obvious that

$$
\begin{aligned}
&\frac{2}{n^2s_i^2s_j^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\sum_{i'\in S_i}\sum_{j'\in S_j}\left(\widehat{P}_{i'j'}^{(m)}-P_{i'j'}\right)\right]^2\\
\leq&\frac{2}{n^2s_is_j}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{i'\in S_i}\sum_{j'\in S_j}\left(\widehat{P}_{i'j'}^{(m)}-P_{i'j'}\right)^2\\
=&\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widehat{P}_{ij}^{(m)}-P_{ij}\right)^2\\
\leq&2C_{11}\left(\frac{\log n}{n^3}\right)^{1/4}.
\end{aligned}
\tag{19}
$$

As to the third term, according to Lemma 1, with probability $1 - 2n^{-2C_2/3} - 2n^{-2C_3/3} - 2n^{-C_4/2}$, we have

$$\frac{4}{n^2 s_i^2 s_j^2} \sum_i \sum_j \left(\sum_{i' \in S_i} \sum_{i' \in S_i} (P_{i'j'} - A_{i'j'}) \right)^2 \leq 4C_5 \left(\frac{\log n}{n^3} \right)^{1/4}. \tag{20}$$

Finally, plugging (18), (19) and (20) into (17) and combining with Lemma 1, with probability $1 - 2n^{C_1/4} - 2n^{-2C_2/3} - 2n^{-2C_3/3} - 2n^{-C_4/2}$, we have

$$\begin{aligned} &\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\widehat{P}_{ij}^{(m)} - \widehat{P}_{ij}^{(m+1)} \right)^2 \\ \leq& 32L^2 \left[1 + (C_1 + 4)^{1/2} \right]^2 \frac{\log n}{n} + (4C_5 + 2C_{11}) \left(\frac{\log n}{n^3} \right)^{1/4} \\ \leq& C_{12} \left(\frac{\log n}{n^3} \right)^{1/4}. \end{aligned} \tag{21}$$

□

Recall that we define $\delta_{\mathbf{P}} = ||\widehat{\mathbf{P}}^{(m+1)} - \widehat{\mathbf{P}}^{(m)}||_F / ||\widehat{\mathbf{P}}^{(m)}||_F$ in Algorithm 1 of the main paper, assume that $\widehat{P}_{ij}^{(m)} \geq \widehat{a}$ for all $(i,j) \in V \times V$, then under the condition of Theorem 4, with high probability, we have

$$\delta_{\mathbf{P}}^2 = \frac{||\widehat{\mathbf{P}}^{(m+1)} - \widehat{\mathbf{P}}^{(m)}||_F^2}{||\widehat{\mathbf{P}}^{(m)}||_F^2} \leq \frac{C_{12}}{\widehat{a}^2} \left(\frac{\log n}{n^3} \right)^{1/4}.$$

B Variations of Our Method

The Algorithm 1 in the main paper gives the framework of our proposed method but allows many variations. In fact, the ways of updating the probability matrix, the pairwise distances and the neighborhood sets all can be modified to adapt specific network structure.

- **Weighted Neighborhood Averaging.** When estimating $\mathbf{P}_{i\cdot}$ by neighborhood averaging, we use all the vertexes in S_i with equal weights. However, the vertexes more similar to vertex i may be more helpful. Therefore, we could expand S_i and let the weights of $A_{i'j'}$ be proportional to $1/d_{ii'}$. In this way, we assign larger weights to those close to vertex i and smaller weights to those relatively farther away from it.
- **Distance Measurement.** When updating $\mathbf{D}$, besides ℓ_2 distance, other distance measurement can also be considered. Furthermore, because we only care about the vertexes which are possible to be the neighbors for a given vertexes, we can only update d_{ij} smaller than a threshold and speed up the process.
- **Vertex-specific Neighborhood Size.** For networks with unknown complicated structure, the number of useful neighbors may vary from vertex to vertex. A more natural idea is to assign different numbers of neighbors to different vertexes. With unequal neighborhood sizes, the only difference in Algorithm 1 is to replace s in $S_i = \{i' : 0 < d_{ii'} \leq d_s\}$ with s_i, where s_i is the specific neighborhood size of vertex i.

Here we focus on the last variation. Consider a network generated by SBM with some blocks in different sizes. As is shown in Figure 2, in this SBM network with 1000 vertexes, there are 4 blocks whose sizes are $100, 200, 300, 400$ respectively. The connecting probability inner a block is set 0.4 while that between two different blocks is in $\{0.15, 0.20, 0.25\}$. In this case, it is reasonable to assign different numbers of neighbors to vertexes in different blocks. And s_i, the size of neighbors for vertex i, is expected to be equal to the size of its corresponding block, as the blue dashed curve in Figure 3(a) shows.

Let $\mathbf{S} = \{s_1, \ldots, s_n\}$ denote the size vector. As the block structure is unknown, we have to estimate $\mathbf{S}$ first. Since estimating n parameters $s_1, \ldots, s_n$ simultaneously is impracticable, we use a threshold for the pairwise distances to get a adaptive size vector. For each vertex i, let $S_i = \{i' : 0 < d_{ii'} \leq d_{\text{thre}}\}$, where d_{thre} is the threshold. In this way, we transform the issue of estimating $\mathbf{S}$ into the selection of

Figure 2: *Network generated by SBM with blocks in different sizes.*

appropriate d_{thre}. Algorithm 1 gives the details of this procedure. A previous connecting probability estimate $\widehat{\mathbf{P}}$ is required for two reasons. First, the pairwise distances are calculated on $\widehat{\mathbf{P}}$. Second, we use $\widehat{\mathbf{P}}$ to evaluate the performance of a given distance threshold by network bootstrap. A series of adjacency matrices are generated with $\widehat{\mathbf{P}}$, then for a given distance threshold and its corresponding size vector, we average the RMSE on estimating $\widehat{\mathbf{P}}$ based on these bootstrap samples to evaluate its performance. Finally we select the best threshold and obtain the size vector.

Algorithm 1 Neighborhood size assignment

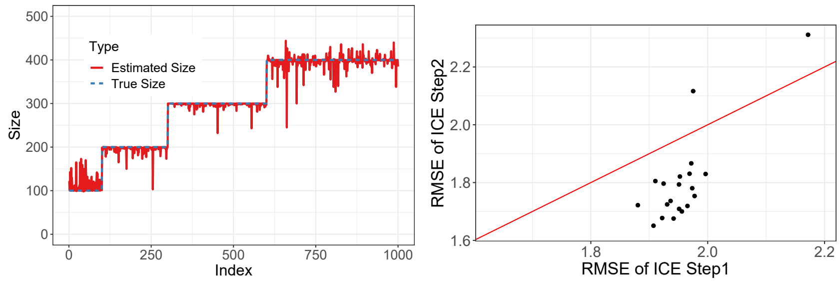
Input: connecting probability estimate $\widehat{\mathbf{P}}$; a series of distance thresholds $\{d_1, \ldots, d_T\}$; number of bootstrap samples B.

Output: size vector estimate $\widehat{\mathbf{S}} = \{\hat{s}_1, \ldots, \hat{s}_n\}$.

```
For each vertex pair i, j ∈ V, obtain their distance d_ij = ||P̂_i· − P̂_j·||_2^2 / n.
for t = 1; t <= T; t++ do
    For each vertex i ∈ V, obtain its neighborhood set S_i = {i' : 0 < d_ii' ≤ d_t} and its size
    s_i = |S_i|.
    Obtain the size vector S^t = {s_1, ..., s_n} with threshold d_t.
end for
Generate a series of adjacency matrices A_1, ..., A_B with P̂ as expectation.
for t = 1; t <= T; t++ do
    for b = 1; b <= B; b++ do
        Based on A_b and S^t, apply Algorithm 1 to obtain P̂_tb, estimate of P̂.
        Calculate RMSE_tb = ||P̂_tb − P̂||_F.
    end for
    Let RMSE_t = Σ_{b=1}^{B} RMSE_tb / B.
end for
Let t* = arg min_{t∈{1,...,T}} RMSE_t.
return Ŝ = S^{t*}.
```

In the SBM case discussed above, we use ICE method with equal neighborhood sizes to get $\widehat{\mathbf{P}}$. Then we estimate the vertex-specific size vector via Algorithm 1. As the red solid curve in Figure 3(a) shows, we successfully assign appropriate numbers of neighbors to most of the vertexes according to the network structure. Then with the estimated size vector, we apply Algorithm 1 in the main paper again, start from random selected neighbors, update the pairwise distances, the neighborhood sets and the estimate iteratively until they converge. Figure 3(b) presents the RMSE of ICE with equal neighborhood sizes and specific neighborhood sizes in 20 repetitions. It is obvious that using specific neighborhood sizes significantly improves the precision.

Figure 3(b) also implies another phenomenon that the precision of these two versions of ICE method are highly correlated. Because both the estimation of the size vectors and the selection procedure in Algorithm 1 rely on $\widehat{\mathbf{P}}$, the performance of ICE with vertex-specific neighborhood sizes also depends on the precision of $\widehat{\mathbf{P}}$. Indeed, Algorithm 1 tends to select a size vector that most appropriate for



(a) The true size vector and estimated size vector.

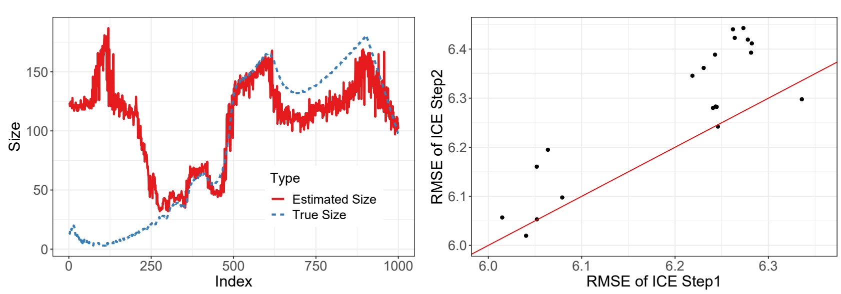
(b) The RMSE by ICE with equal neighborhood sizes (Step1) and specific neighborhood sizes (Step2).

Figure 3: *Result on SBM network with unequal block sizes.*

estimating $\widehat{\mathbf{P}}$, instead of $\mathbf{P}$. Therefore, if $\widehat{\mathbf{P}}$ is a good estimate, then the estimated size vector will be close to the true size vector and ICE with vertex-specific neighborhood sizes will outperform that with equal sizes. However, if $\widehat{\mathbf{P}}$ is a poor estimate, it is hard to estimate the size vector well even with an appropriate distance threshold. And ICE method that starts from the wrongly selected size vector will repeat the mistakes made by $\widehat{\mathbf{P}}$ without improvement.

A vivid instance comes from the simulated network with complicated local structure that we have mentioned in the main paper. The blue dashed curve in Figure 4(a) presents the most appropriate size vector obtained via Algorithm 1 based on $\mathbf{P}$. It is easily seen that the neighborhood size s_i is proportional to the smoothness of $\mathbf{P}_{i\cdot}$. Then, based on $\widehat{\mathbf{P}}$ estimated by ICE with equal neighborhood sizes, we get the estimated size vector, as shown by the red solid curve. These two curves coincide well for vertexes indices ranging from 250 to 1000. However, for vertexes corresponding to the local structure, whose indices under 250, the neighborhood sizes are heavily overestimated. The reason is that the local structure in $\mathbf{P}$ is hard to estimate and the corresponding part in $\widehat{\mathbf{P}}$ is over-smoothing. Consequently, for vertex $i, j \in \{1, \dots, 250\}$, the pairwise distances d_{ij} calculated on $\widehat{\mathbf{P}}$ is much smaller than the true distance on $\mathbf{P}$. Then the estimated size of neighborhood set for vertex $i \in \{1, \dots, 250\}$ is much larger because it include more non-neighbors.

Figure 4 displays the RMSE in 20 repetitions. Again, the RMSE by ICE with equal and vertex-specific neighborhood sizes are linearly correlated. As the estimated size vector is far away from the truth, ICE with vertex-specific neighborhood sizes shows no advantage. How to deal with this problem is worthy of further investigation.



(a) The true size vector and estimated size vector.

(b) The RMSE by ICE with equal neighborhood sizes (Step1) and specific neighborhood sizes (Step2).

Figure 4: *Results on Graphon 4.*

C Broader Impact

This paper provides an iterative method based on neighborhood averaging for estimating the connecting probabilities between pairs of vertexes in networks. This work may benefit the analysis of the relationship between subjects in networks, such as user friendship in social network. Although we focus on the task of connecting probability estimation, we believe that the iterative procedure may be applied in other learning tasks solved by methods based on neighborhood averaging, like KNN. Indeed, if the output (e.g., estimate, prediction) obtained by neighborhood averaging is helpful to construct a more reliable distance measurement, it is natural to update the neighbors, thus improving the performance of the output. We should also be aware of the unintended usage for our method. For instance, advertisers may apply our method to discover potential friends of users in social media and recommend products to them, which may cause privacy violations and overflowing of advertising information.

D Code

Implementation of our proposed method ICE is available online at https://github.com/Siva-47/ICE. We have provided the main steps with some necessary details to reproduce the results in our paper. It should be noted that we narrow the range of the tuning parameters in grid search to save time for reproduction. In practice, one should use a large range first and then narrow it.