A Divide and Conquer Framework for Distributed Graph Clustering

Wenzhuo Yang A0096049@NUS.EDU.SG

Department of Mechanical Engineering, National University of Singapore, Singapore 117576

Huan Xu MPEXUH@NUS.EDU.SG

Department of Mechanical Engineering, National University of Singapore, Singapore 117576

1. Notations

Recall that the graph we analyzed contains n nodes and r clusters, and is generated according to the generalized stochastic blockmodel. We let K_i be the size of the ith cluster, K be the minimum cluster size, i.e., $K = \min_i K_i$, and K^* be the size of the smallest cluster that contains at least one ordinary node. Therefore, edge (i, j) is present in the graph with probability $p_{ij} \geq p$ for every pair of nodes i, j that belong to the same cluster, and edge (i, j) is present in the graph with probability $q_{ij} \leq q$ for every pair of nodes i, j that are in different clusters. Note that the outliers in the graph do not belong to any cluster.

Let $\mathbf{U\Sigma U}^\top$ be the singular value decomposition of $\mathbf{Y}^*$ and $P_T(\mathbf{M}) = \mathbf{UU}^\top\mathbf{M} + \mathbf{MUU}^\top - \mathbf{UU}^\top\mathbf{MUU}^\top$ be the projection of $\mathbf{M}$ onto the row and column spaces of $\mathbf{Y}^*$, and let $P_{T^\perp}(\mathbf{M}) = \mathbf{M} - P_T(\mathbf{M})$. Let $\mathcal{R}$ be the support of $\mathbf{Y}^*$, i.e., $\mathcal{R} = \{(i, j) : Y^*_{ij} = 1\}$, $\mathcal{C}$ be the set of the edges connecting to the high confidence nodes, i.e., $\mathcal{C} = \{(i, j) : i \text{ or } j \text{ is a high confidence node}\}$ and $\mathcal{A}$ be the support of $\mathbf{A}$, i.e., $\mathcal{A} = \{(i, j) : A_{ij} = 1\}$. For a set of matrix indices Ω, we let $P_\Omega(\mathbf{M})$ be the matrix whose (i, j)th entry equals M_{ij} if $(i, j) \in \Omega$ or 0 otherwise. We let $\mathbf{E}$ be the matrix whose entries are all ones.

2. Proof of Theorem 1

For clarity, we let

$$\lambda = \frac{c_0}{\sqrt{\max\{n - s, K^*\}\log n}},\ c_{\mathcal{A}} = \sqrt{\frac{1-t}{t}},\ c_{\mathcal{A}^c} = \sqrt{\frac{t}{1-t}},\ \tau = \min\{\tau_1, \tau_2\}.$$

In other words, "$c_{\mathcal{A}}$" and "$c_{\mathcal{A}^c}$" in Equation (1) are replaced by $\lambda c_{\mathcal{A}}$ and $\lambda c_{\mathcal{A}^c}$, respectively. Recall that $K = \min_i K_i$ and $K^* = \min\{K_i : \text{cluster } i \text{ contains at least one ordinary node}\}$. If all the nodes are high confident, we let $K^* = K$ without loss of generality. Clearly, $K^* > K$ means that there exist some clusters whose nodes are all high confident. We denote the set of the nodes in these clusters by $\mathcal{N}$. Let $\mathcal{H} = \{(i, j) : i \text{ or } j \in \mathcal{N}\}$. Obviously, $\mathcal{H} \subseteq \mathcal{C}$. Let $\mathcal{E} = \{(i, i) : i = 1, \cdots, n\}$ be the set of the diagonal entries. We first explore the sufficient conditions such that the true adjacent matrix $\mathbf{Y}^*$ is the unique optimal solution of (1). In the proofs, the constants may vary from line to line.

Lemma A-1. *For any matrix* $\mathbf{X} \in \mathbb{R}^{n\times n}$, $P_{\mathcal{H}^c}P_TP_{\mathcal{H}}\mathbf{X} = 0$ *and* $P_{\mathcal{H}}P_TP_{\mathcal{H}^c}\mathbf{X} = 0$.

Proof. We need to show that support$(P_TP_{\mathcal{H}}\mathbf{X}) \subseteq \mathcal{H}$ and support$(P_TP_{\mathcal{H}^c}\mathbf{X}) \subseteq \mathcal{H}^c$. Recall that $P_T(\mathbf{M}) = \mathbf{UU}^\top\mathbf{M} + \mathbf{MUU}^\top - \mathbf{UU}^\top\mathbf{MUU}^\top$, since $\mathcal{H}$ is "symmetric", namely, $(i, j) \in \mathcal{H}$ implies $(j, i) \in \mathcal{H}$, we only need to show that support($\mathbf{U}\mathbf{U}^\top P_{\mathcal{H}}\mathbf{X}$) $\subseteq \mathcal{H}$ and support($\mathbf{U}\mathbf{U}^\top P_{\mathcal{H}^c}\mathbf{X}$) $\subseteq \mathcal{H}^c$. For any $(i,j) \notin \mathcal{H}$, suppose that i, j belong to clusters $R(i)$ and $R(j)$, respectively. From the definition of $\mathcal{H}$, we know that $R(i), R(j) \nsubseteq \mathcal{H}$, which implies that $(k,j) \notin \mathcal{H}$ for all k such that $(i,k) \in R(i)$. Thus, we have

$$(\mathbf{U}\mathbf{U}^\top P_{\mathcal{H}}\mathbf{X})_{ij} = \sum_{k:(i,k)\in R(i)} (\mathbf{U}\mathbf{U}^\top)_{ik}(P_{\mathcal{H}}\mathbf{X})_{kj} = 0, \text{ for } (i,j) \notin \mathcal{H}.$$

Similarly, we can prove that $(\mathbf{U}\mathbf{U}^\top P_{\mathcal{H}^c}\mathbf{X})_{ij} = 0$ for all $(i,j) \in \mathcal{H}$. □

Lemma A-2. $(\mathbf{Y}^*, A - \mathbf{Y}^*)$ *is the unique optimal solution of* (1)*, if there exist matrices* $\mathbf{W}_1, \mathbf{W}_2$ *and a positive number* $\epsilon < 0.5$ *such that (a)* $P_{\mathcal{H}}\mathbf{W}_1 = 0, P_{\mathcal{H}^c}\mathbf{W}_2 = 0$*, (b)* $\|\mathbf{W}_1\| \le \frac{1}{2}$, $\|\mathbf{W}_2\| \le \frac{1}{2}$*, (c)* $\|P_T\mathbf{W}_1\|_\infty \le \frac{1}{2}\epsilon \min\{\lambda c_{\mathcal{A}}, \lambda c_{\mathcal{A}^c}, c_{\mathcal{C}}\}$, $\|P_T\mathbf{W}_2\|_\infty \le \frac{1}{4}c_{\mathcal{C}}$ *and (d)*

$$\begin{aligned}
\textit{I.} \quad & (1-\epsilon)\lambda c_{\mathcal{A}} - P_{\mathcal{R}\cap\mathcal{A}\cap\mathcal{C}^c\cap\mathcal{H}^c}(\mathbf{U}\mathbf{U}^\top + \mathbf{W}_1) \ge 0,\\
\textit{II.} \quad & -(1+\epsilon)\lambda c_{\mathcal{A}^c} - P_{\mathcal{R}\cap\mathcal{A}^c\cap\mathcal{C}^c\cap\mathcal{H}^c}(\mathbf{U}\mathbf{U}^\top + \mathbf{W}_1) \ge 0,\\
\textit{III.} \quad & -(1+\epsilon)\lambda c_{\mathcal{A}} + P_{\mathcal{R}^c\cap\mathcal{A}\cap\mathcal{C}^c\cap\mathcal{H}^c}(\mathbf{W}_1) \ge 0,\\
\textit{IV.} \quad & (1-\epsilon)\lambda c_{\mathcal{A}^c} + P_{\mathcal{R}^c\cap\mathcal{A}^c\cap\mathcal{C}^c\cap\mathcal{H}^c}(\mathbf{W}_1) \ge 0,\\
\textit{V.} \quad & (1-\epsilon)c_{\mathcal{C}} - P_{\mathcal{R}\cap\mathcal{A}\cap\mathcal{C}\cap\mathcal{H}^c}(\mathbf{U}\mathbf{U}^\top + \mathbf{W}_1) \ge 0,\\
\textit{VI.} \quad & -(1+\epsilon)c_{\mathcal{C}} - P_{\mathcal{R}\cap\mathcal{A}^c\cap\mathcal{C}\cap\mathcal{H}^c}(\mathbf{U}\mathbf{U}^\top + \mathbf{W}_1) \ge 0,\\
\textit{VII.} \quad & -(1+\epsilon)c_{\mathcal{C}} + P_{\mathcal{R}^c\cap\mathcal{A}\cap\mathcal{C}\cap\mathcal{H}^c}(\mathbf{W}_1) \ge 0,\\
\textit{VIII.} \quad & (1-\epsilon)c_{\mathcal{C}} + P_{\mathcal{R}^c\cap\mathcal{A}^c\cap\mathcal{C}\cap\mathcal{H}^c}(\mathbf{W}_1) \ge 0,\\
\textit{IX.} \quad & \tfrac{1}{2}c_{\mathcal{C}} - P_{\mathcal{R}\cap\mathcal{A}\cap\mathcal{H}}(\mathbf{U}\mathbf{U}^\top + \mathbf{W}_2) \ge 0,\\
\textit{X.} \quad & -\tfrac{3}{2}c_{\mathcal{C}} - P_{\mathcal{R}\cap\mathcal{A}^c\cap\mathcal{H}}(\mathbf{U}\mathbf{U}^\top + \mathbf{W}_2) \ge 0,\\
\textit{XI.} \quad & -\tfrac{3}{2}c_{\mathcal{C}} + P_{\mathcal{R}^c\cap\mathcal{A}\cap\mathcal{H}}(\mathbf{W}_2) \ge 0,\\
\textit{XII.} \quad & \tfrac{1}{2}c_{\mathcal{C}} + P_{\mathcal{R}^c\cap\mathcal{A}^c\cap\mathcal{H}}(\mathbf{W}_2) \ge 0,
\end{aligned}$$

Proof. When the conditions above are satisfied, we need to show that the following inequality holds for any $\mathbf{Y}$ and $\mathbf{S}$ such that $\mathbf{Y} \ne \mathbf{Y}^*$, $0 \le \mathbf{Y} \le 1$ and $\mathbf{Y} + \mathbf{S} = \mathbf{A}$:

$$\begin{aligned}
\text{Opt} \triangleq & \|\mathbf{Y}^*\|_* + \lambda c_{\mathcal{A}}\|P_{\mathcal{A}\cap\mathcal{C}^c}\mathbf{S}^*\|_1 + \lambda c_{\mathcal{A}^c}\|P_{\mathcal{A}^c\cap\mathcal{C}^c}\mathbf{S}^*\|_1 + c_{\mathcal{C}}\|P_{\mathcal{C}}\mathbf{S}^*\|_1\\
< & \|\mathbf{Y}\|_* + \lambda c_{\mathcal{A}}\|P_{\mathcal{A}\cap\mathcal{C}^c}\mathbf{S}\|_1 + \lambda c_{\mathcal{A}^c}\|P_{\mathcal{A}^c\cap\mathcal{C}^c}\mathbf{S}\|_1 + c_{\mathcal{C}}\|P_{\mathcal{C}}\mathbf{S}\|_1.
\end{aligned}$$

So we need to prove that

$$\begin{aligned}
\Delta \triangleq\ & [\|\mathbf{Y}\|_* + \lambda c_{\mathcal{A}}\|P_{\mathcal{A}\cap\mathcal{C}^c}\mathbf{S}\|_1 + \lambda c_{\mathcal{A}^c}\|P_{\mathcal{A}^c\cap\mathcal{C}^c}\mathbf{S}\|_1 + c_{\mathcal{C}}\|P_{\mathcal{C}}\mathbf{S}\|_1] -\\
& [\|\mathbf{Y}^*\|_* + \lambda c_{\mathcal{A}}\|P_{\mathcal{A}\cap\mathcal{C}^c}\mathbf{S}^*\|_1 + \lambda c_{\mathcal{A}^c}\|P_{\mathcal{A}^c\cap\mathcal{C}^c}\mathbf{S}^*\|_1 + c_{\mathcal{C}}\|P_{\mathcal{C}}\mathbf{S}^*\|_1]\\
=\ & \|\mathbf{Y}\|_* - \|\mathbf{Y}^*\|_* + \lambda c_{\mathcal{A}}(\|P_{\mathcal{A}\cap\mathcal{C}^c}\mathbf{S}\|_1 - \|P_{\mathcal{A}\cap\mathcal{C}^c}\mathbf{S}^*\|_1) +\\
& \lambda c_{\mathcal{A}^c}(\|P_{\mathcal{A}^c\cap\mathcal{C}^c}\mathbf{S}\|_1 - \|P_{\mathcal{A}^c\cap\mathcal{C}^c}\mathbf{S}^*\|_1) + c_{\mathcal{C}}(\|P_{\mathcal{C}}\mathbf{S}\|_1 - \|P_{\mathcal{C}}\mathbf{S}^*\|_1)\\
=\ & \lambda c_{\mathcal{A}} \sum_{(i,j)\in\mathcal{A}\cap\mathcal{C}^c}(Y^*_{ij} - Y_{ij}) + \lambda c_{\mathcal{A}^c}\sum_{(i,j)\in\mathcal{A}^c\cap\mathcal{C}^c}(Y_{ij} - Y^*_{ij}) +\\
& c_{\mathcal{C}}\sum_{(i,j)\in\mathcal{A}\cap\mathcal{C}}(Y^*_{ij} - Y_{ij}) + c_{\mathcal{C}}\sum_{(i,j)\in\mathcal{A}^c\cap\mathcal{C}}(Y_{ij} - Y^*_{ij}) + \|\mathbf{Y}\|_* - \|\mathbf{Y}^*\|_* > 0
\end{aligned}$$

Let $\mathbf{W} = \mathbf{W}_1 + \mathbf{W}_2$, then $\mathbf{U}\mathbf{U}^\top + \mathbf{W} - P_T(\mathbf{W})$ is a subgradient of $f(\mathbf{X}) = \|\mathbf{X}\|_*$ at $\mathbf{X} = \mathbf{Y}^*$, which implies that

$\|\mathbf{Y}\|_* - \|\mathbf{Y}^*\|_* \geq \langle \mathbf{U}\mathbf{U}^\top + \mathbf{W} - P_T(\mathbf{W}), \mathbf{Y} - \mathbf{Y}^* \rangle$. Hence we have

$$\begin{aligned}\Delta \geq \lambda c_{\mathcal{A}} \sum_{(i,j)\in\mathcal{A}\cap\mathcal{C}^c} (Y^*_{ij} - Y_{ij}) + \lambda c_{\mathcal{A}^c} \sum_{(i,j)\in\mathcal{A}^c\cap\mathcal{C}^c} (Y_{ij} - Y^*_{ij}) + \\ c_{\mathcal{C}} \sum_{(i,j)\in\mathcal{A}\cap\mathcal{C}} (Y^*_{ij} - Y_{ij}) + c_{\mathcal{C}} \sum_{(i,j)\in\mathcal{A}^c\cap\mathcal{C}} (Y_{ij} - Y^*_{ij}) + \\ \langle \mathbf{U}\mathbf{U}^\top + \mathbf{W}, \mathbf{Y} - \mathbf{Y}^* \rangle + \langle -P_T(\mathbf{W}), \mathbf{Y} - \mathbf{Y}^* \rangle\end{aligned}$$

By Lemma A-1 and Condition (d), we have

$$\begin{aligned}\Delta \geq & \langle \lambda c_{\mathcal{A}} \mathbf{E} - (\mathbf{U}\mathbf{U}^\top + \mathbf{W}), P_{\mathcal{A}\cap\mathcal{C}^c}(\mathbf{Y}^* - \mathbf{Y}) \rangle + \langle -\lambda c_{\mathcal{A}^c} \mathbf{E} - (\mathbf{U}\mathbf{U}^\top + \mathbf{W}), P_{\mathcal{A}^c\cap\mathcal{C}^c}(\mathbf{Y}^* - \mathbf{Y}) \rangle + \\ & \langle c_{\mathcal{C}} \mathbf{E} - (\mathbf{U}\mathbf{U}^\top + \mathbf{W}), P_{\mathcal{A}\cap\mathcal{C}}(\mathbf{Y}^* - \mathbf{Y}) \rangle + \langle -c_{\mathcal{C}} \mathbf{E} - (\mathbf{U}\mathbf{U}^\top + \mathbf{W}), P_{\mathcal{A}^c\cap\mathcal{C}}(\mathbf{Y}^* - \mathbf{Y}) \rangle + \\ & \langle -P_T(\mathbf{W}_1) - P_T(\mathbf{W}_2), \mathbf{Y} - \mathbf{Y}^* \rangle \\ \geq & \langle \epsilon\lambda c_{\mathcal{A}} \mathbf{E}, P_{\mathcal{R}\cap\mathcal{A}\cap\mathcal{C}^c}(\mathbf{Y}^* - \mathbf{Y}) \rangle + \langle \epsilon\lambda c_{\mathcal{A}} \mathbf{E}, P_{\mathcal{R}^c\cap\mathcal{A}\cap\mathcal{C}^c}(\mathbf{Y} - \mathbf{Y}^*) \rangle + \langle \epsilon\lambda c_{\mathcal{A}^c} \mathbf{E}, P_{\mathcal{R}\cap\mathcal{A}^c\cap\mathcal{C}^c}(\mathbf{Y}^* - \mathbf{Y}) \rangle + \\ & \langle \epsilon\lambda c_{\mathcal{A}^c} \mathbf{E}, P_{\mathcal{R}^c\cap\mathcal{A}^c\cap\mathcal{C}^c}(\mathbf{Y} - \mathbf{Y}^*) \rangle + \langle \epsilon c_{\mathcal{C}} \mathbf{E}, P_{\mathcal{R}\cap\mathcal{C}\cap\mathcal{H}^c}(\mathbf{Y}^* - \mathbf{Y}) \rangle + \langle \epsilon c_{\mathcal{C}} \mathbf{E}, P_{\mathcal{R}^c\cap\mathcal{C}\cap\mathcal{H}^c}(\mathbf{Y} - \mathbf{Y}^*) \rangle + \\ & \langle \frac{1}{2} c_{\mathcal{C}} \mathbf{E}, P_{\mathcal{R}\cap\mathcal{H}}(\mathbf{Y}^* - \mathbf{Y}) \rangle + \langle \frac{1}{2} c_{\mathcal{C}} \mathbf{E}, P_{\mathcal{R}^c\cap\mathcal{H}}(\mathbf{Y} - \mathbf{Y}^*) \rangle + \langle -P_T(\mathbf{W}_1) - P_T(\mathbf{W}_2), \mathbf{Y} - \mathbf{Y}^* \rangle \\ = & \epsilon\lambda c_{\mathcal{A}} \|P_{\mathcal{A}\cap\mathcal{C}^c}(\mathbf{Y} - \mathbf{Y}^*)\|_1 + \epsilon\lambda c_{\mathcal{A}^c} \|P_{\mathcal{A}^c\cap\mathcal{C}^c}(\mathbf{Y} - \mathbf{Y}^*)\|_1 + \epsilon c_{\mathcal{C}} \|P_{\mathcal{C}\cap\mathcal{H}^c}(\mathbf{Y} - \mathbf{Y}^*)\|_1 + \\ & \frac{1}{2} c_{\mathcal{C}} \|P_{\mathcal{H}}(\mathbf{Y} - \mathbf{Y}^*)\|_1 + \langle -P_T(\mathbf{W}_1) - P_T(\mathbf{W}_2), \mathbf{Y} - \mathbf{Y}^* \rangle \\ \geq & \lambda\epsilon \min\{c_{\mathcal{A}}, c_{\mathcal{A}^c}, c_{\mathcal{C}}\} \|P_{\mathcal{H}^c}(\mathbf{Y} - \mathbf{Y}^*)\|_1 + \frac{1}{2} c_{\mathcal{C}} \|P_{\mathcal{H}}(\mathbf{Y} - \mathbf{Y}^*)\|_1 - \langle P_T(\mathbf{W}_1), \mathbf{Y} - \mathbf{Y}^* \rangle - \langle P_T(\mathbf{W}_2), \mathbf{Y} - \mathbf{Y}^* \rangle \\ = & \lambda\epsilon \min\{c_{\mathcal{A}}, c_{\mathcal{A}^c}, c_{\mathcal{C}}\} \|P_{\mathcal{H}^c}(\mathbf{Y} - \mathbf{Y}^*)\|_1 + \frac{1}{2} c_{\mathcal{C}} \|P_{\mathcal{H}}(\mathbf{Y} - \mathbf{Y}^*)\|_1 - \langle P_T(\mathbf{W}_1), P_{\mathcal{H}^c}(\mathbf{Y} - \mathbf{Y}^*) \rangle - \langle P_T(\mathbf{W}_2), P_{\mathcal{H}}(\mathbf{Y} - \mathbf{Y}^*) \rangle \\ \geq & (\lambda\epsilon \min\{c_{\mathcal{A}}, c_{\mathcal{A}^c}, c_{\mathcal{C}}\} - \|P_T(\mathbf{W}_1)\|_\infty) \|P_{\mathcal{H}^c}(\mathbf{Y} - \mathbf{Y}^*)\|_1 + (\frac{1}{2} c_{\mathcal{C}} - \|P_T(\mathbf{W}_2)\|_\infty) \|P_{\mathcal{H}}(\mathbf{Y} - \mathbf{Y}^*)\|_1 \\ \geq & \frac{1}{2} \lambda\epsilon \min\{c_{\mathcal{A}}, c_{\mathcal{A}^c}, c_{\mathcal{C}}\} \|P_{\mathcal{H}^c}(\mathbf{Y} - \mathbf{Y}^*)\|_1 + \frac{1}{4} c_{\mathcal{C}} \|P_{\mathcal{H}}(\mathbf{Y} - \mathbf{Y}^*)\|_1 > 0.\end{aligned}$$

Hence we obtain this lemma. □

From the conditions in Theorem 1, we know that

$$\lambda = \frac{c_0}{\sqrt{\max\{n - s, K^*\} \log n}}, c_{\mathcal{C}} = \frac{c_0}{\sqrt{K \log n}} \geq \frac{c_0}{\sqrt{K^* \log n}} \text{ and } \frac{p - q}{\sqrt{p(1 - q)}} \geq \frac{c_1}{\lambda K^*}.$$

Let $\epsilon = \frac{c_2}{\sqrt{t(1-t)}} \cdot \frac{1}{\lambda K^*}$, we have the following two lemmas. For simplicity, we do not provide the explicit values for the constants $c, c_0, c_1, c_2, c_3, c_\tau, c_K$ used in the following proofs. One can easily verify that such constants exist, e.g., $c_1 = 200c, c_2 = 1, c_0 = \frac{1}{2048c^2}, c_3 \geq 512c^2, c_\tau \leq \frac{1}{4096c^2}, c_K \geq \frac{1}{8}$ for $c \geq 1$.

Lemma A-3. *(a)* $t(1 - t) \geq \frac{c_3}{\lambda^2 K^{*2}}$*; (b)* $0 < \epsilon < 0.5$*; (c)* $(1 + \epsilon)\frac{1-p}{p} \leq (1 - 2\epsilon)\frac{1-t}{t}$*; (d)* $(1 + \epsilon)\frac{q}{1-q} \leq (1 - \epsilon)\frac{t}{1-t}$.

Proof. Since $\frac{1}{4}p + \frac{3}{4}q \leq t \leq \frac{3}{4}p + \frac{1}{4}q$, $t(1 - t) \geq \frac{1}{2} \min\{t, 1 - t\} \geq \frac{1}{8}(p - q) \geq \frac{1}{8}\sqrt{p(1 - q)}\frac{c_1}{\lambda K^*} \geq \frac{1}{8}\sqrt{t(1 - t)}\frac{c_1}{\lambda K^*}$, (a) holds when $\frac{c_1^2}{64} \geq c_3$. By choosing proper constants, e.g., $\frac{c_2^2}{c_3} \leq \frac{1}{4}$, (b) follows from (a) directly. For (c), note that $p - t \geq \frac{p-q}{4} \geq \sqrt{p(1 - q)}\frac{c_1}{4\lambda K^*} \geq \frac{c_1}{4c_2}p(1 - t)\epsilon$. It can be easily verified that this implies (c) when $\frac{c_1}{c_2} \geq 32$. Similarly, $t - q \geq \frac{p-q}{4} \geq \frac{c_1}{4c_2}p(1 - t)\epsilon \geq \frac{c_1}{16c_2}t(1 - q)\epsilon$ since $1 - t \geq \frac{1}{4}(1 - q)$, which implies (d) when $\frac{c_1}{c_2} \geq 32$. □

Lemma A-4. $p \geq \frac{c_3}{\lambda^2 K^{*2}} \geq c_3 \max\{\frac{\log n}{K^*}, \frac{(n-s)\log n}{K^{*2}}\}$.

Proof. By Lemma A-3, $p \geq t(1-t) \geq \frac{c_3}{\lambda^2 K^{*2}} \geq c_3 \max\{\frac{\log n}{K^*}, \frac{(n-s)\log n}{K^{*2}}\}$. □

In the following parts, we will construct $\mathbf{W}_1$ and $\mathbf{W}_2$ to meet the conditions in Lemma A-2.

2.1. Construct $\mathbf{W}_1$

We now construct $\mathbf{W}_1$ such that the conditions in Lemma A-2 are satisfied.

Step 1. Construct the dual certificate $\mathbf{W}_1$: We let $\mathbf{W}_1 = \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3 + \mathbf{Q}_4$, where $\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3, \mathbf{Q}_4$ are defined as follows:

$$\mathbf{Q}_1(i,j) = \begin{cases} -(\mathbf{U}\mathbf{U}^\top)_{ij}, & (i,j) \in \mathcal{R} \cap \mathcal{A}^c \cap \mathcal{H}^c \\ \frac{1-p_{ij}}{p_{ij}}(\mathbf{U}\mathbf{U}^\top)_{ij}, & (i,j) \in \mathcal{R} \cap \mathcal{C}^c \cap \mathcal{A} \cap \mathcal{H}^c \\ \frac{1-\tau_1}{\tau_1}(\mathbf{U}\mathbf{U}^\top)_{ij}, & (i,j) \in \mathcal{R} \cap \mathcal{C} \cap \mathcal{A} \cap \mathcal{H}^c \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbf{Q}_2(i,j) = \begin{cases} -(1+\epsilon)\lambda c_{\mathcal{A}^c}, & (i,j) \in \mathcal{R} \cap \mathcal{C}^c \cap \mathcal{A}^c \cap \mathcal{H}^c \\ \frac{1-p_{ij}}{p_{ij}}(1+\epsilon)\lambda c_{\mathcal{A}^c}, & (i,j) \in \mathcal{R} \cap \mathcal{C}^c \cap \mathcal{A} \cap \mathcal{H}^c \\ -(1+\epsilon)c_{\mathcal{C}}, & (i,j) \in \mathcal{R} \cap \mathcal{C} \cap \mathcal{A}^c \cap \mathcal{H}^c \\ \frac{1-\tau_1}{\tau_1}(1+\epsilon)c_{\mathcal{C}}, & (i,j) \in \mathcal{R} \cap \mathcal{C} \cap \mathcal{A} \cap \mathcal{H}^c \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbf{Q}_3(i,j) = \begin{cases} (1+\epsilon)\lambda c_{\mathcal{A}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{C}^c \cap \mathcal{A} \cap \mathcal{H}^c \cap \mathcal{E}^c \\ -\frac{q_{ij}}{1-q_{ij}}(1+\epsilon)\lambda c_{\mathcal{A}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{C}^c \cap \mathcal{A}^c \cap \mathcal{H}^c \cap \mathcal{E}^c \\ (1+\epsilon)c_{\mathcal{C}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{C} \cap \mathcal{A} \cap \mathcal{H}^c \\ -\frac{1-\tau_2}{\tau_2}(1+\epsilon)c_{\mathcal{C}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{C} \cap \mathcal{A}^c \cap \mathcal{H}^c \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbf{Q}_4(i,j) = \begin{cases} (1+\epsilon)\lambda c_{\mathcal{A}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{E} \\ 0, & \text{otherwise} \end{cases}$$

It can be easily verified that $\mathbb{E}[\mathbf{Q}_1] = \mathbb{E}[\mathbf{Q}_2] = \mathbb{E}[\mathbf{Q}_3] = 0$, and

$$|\mathbf{Q}_1(i,j)| \leq \frac{1}{pK^*},\ |\mathbf{Q}_2(i,j)| \leq \max\{\frac{2\lambda c_{\mathcal{A}^c}}{p}, 2c_{\mathcal{C}}\},\ |\mathbf{Q}_3(i,j)| \leq \max\{\frac{2\lambda c_{\mathcal{A}}}{1-q}, 2c_{\mathcal{C}}\}.$$

Note that $\tau = \min\{\tau_1, \tau_2\} \geq \frac{4}{5}$ and $q \leq t \leq p$, by simple calculation, we have

$$\begin{aligned}
&\mathrm{Var}[\mathbf{Q}_1(i,j)] \leq \tfrac{1-p}{pK^{*2}} \leq \tfrac{1}{pK^{*2}}, && (i,j) \in \mathcal{C}^c \\
&\mathrm{Var}[\mathbf{Q}_1(i,j)] \leq \tfrac{2(1-\tau)}{K^{*2}}, && (i,j) \in \mathcal{C} \\
&\mathrm{Var}[\mathbf{Q}_2(i,j)] \leq \tfrac{4\lambda^2 c_{\mathcal{A}^c}^2(1-p)}{p} \leq \tfrac{4\lambda^2 c_{\mathcal{A}^c}^2(1-t)}{p}, && (i,j) \in \mathcal{C}^c \\
&\mathrm{Var}[\mathbf{Q}_2(i,j)] \leq 8c_{\mathcal{C}}^2(1-\tau), && (i,j) \in \mathcal{C} \\
&\mathrm{Var}[\mathbf{Q}_3(i,j)] \leq \tfrac{4\lambda^2 c_{\mathcal{A}}^2 q}{1-q} \leq \tfrac{4\lambda^2 c_{\mathcal{A}}^2 t}{1-q}, && (i,j) \in \mathcal{C}^c \\
&\mathrm{Var}[\mathbf{Q}_3(i,j)] \leq 8c_{\mathcal{C}}^2(1-\tau), && (i,j) \in \mathcal{C}
\end{aligned}$$

Step 2. Bound $\|\mathbf{W}_1\|$: From Lemma A-5, the following inequalities hold with high probability:

$$\|\mathbf{Q}_1\| \le c\left[\frac{\log n}{pK^*} + \sqrt{\frac{2n(1-\tau)}{{K^*}^2} + \frac{n-s}{p{K^*}^2}}\cdot\sqrt{\log n}\right]$$
$$\|\mathbf{Q}_2\| \le c\left[\max\{\frac{2\lambda c_{\mathcal{A}^c}}{p}, 2c_{\mathcal{C}}\}\log n + \sqrt{8nc_{\mathcal{C}}^2(1-\tau) + (n-s)\frac{4\lambda^2 c_{\mathcal{A}^c}^2(1-t)}{p}}\cdot\sqrt{\log n}\right]$$
$$\|\mathbf{Q}_3\| \le c\left[\max\{\frac{2\lambda c_{\mathcal{A}}}{1-q}, 2c_{\mathcal{C}}\}\log n + \sqrt{8nc_{\mathcal{C}}^2(1-\tau) + (n-s)\frac{4\lambda^2 c_{\mathcal{A}}^2 t}{1-q}}\cdot\sqrt{\log n}\right]$$

Recall that $K^* \ge K \ge c_K \log n$, $\lambda = \frac{c_0}{\sqrt{\max\{n-s, K^*\}\log n}}$, $1-\tau \le c_\tau \frac{K}{n}$ and $c_{\mathcal{C}} = \frac{c_0}{\sqrt{K\log n}}$. From Lemma A-4, $p \ge c_3\frac{\log n}{K^*}$, which implies that $c\max\{\frac{\log n}{pK^*}, \frac{\log n}{K}\} \le \frac{1}{16}$. On the other hand, $p \ge c_3\frac{(n-s)\log n}{{K^*}^2}$, so $c\sqrt{\frac{2n(1-\tau)}{{K^*}^2} + \frac{n-s}{p{K^*}^2}}\cdot\sqrt{\log n} \le c\sqrt{\frac{c_\tau}{c_K\log n} + \frac{1}{c_3\log n}}\cdot\sqrt{\log n} \le \frac{1}{16}$. Hence $\|\mathbf{Q}_1\| \le \frac{1}{8}$.

To bound $\|\mathbf{Q}_2\|$, note that $\frac{\lambda c_{\mathcal{A}^c}}{p} = \lambda\frac{1}{p}\sqrt{\frac{t}{1-t}} \le \lambda\frac{1}{\sqrt{t(1-t)}} \le \frac{\lambda^2 K^*}{c_3} \le \frac{1}{c_3\log n}$ and $c_{\mathcal{C}}\log n = \sqrt{\frac{c_0\log n}{K}}$, so $c\max\{\frac{2\lambda c_{\mathcal{A}^c}}{p}, 2c_{\mathcal{C}}\}\log n \le \frac{1}{16}$. We also have $(n-s)\frac{\lambda^2 c_{\mathcal{A}^c}^2(1-t)}{p} = (n-s)\frac{\lambda^2}{p}\cdot\frac{t}{1-t}\cdot(1-t) \le (n-s)\lambda^2 \le \frac{c_0}{\log n}$ and $nc_{\mathcal{C}}^2(1-\tau) \le \frac{c_0^2 c_\tau}{\log n}$ which implies $c\sqrt{8nc_{\mathcal{C}}^2(1-\tau) + (n-s)\frac{4\lambda^2 c_{\mathcal{A}^c}^2(1-t)}{p}}\cdot\sqrt{\log n} \le \frac{1}{16}$, so that $\|\mathbf{Q}_2\| \le \frac{1}{8}$. Similarly, we can prove that $\|\mathbf{Q}_3\| \le \frac{1}{8}$. For $\|\mathbf{Q}_4\|$, note that $(1+\epsilon)\lambda c_{\mathcal{A}} \le 2\lambda c_{\mathcal{A}} = 2\lambda\sqrt{\frac{1-t}{t}} \le 2\lambda\frac{1}{\sqrt{t(1-t)}} \le \frac{2}{c_3\log n} \le \frac{1}{8}$. Hence $\|\mathbf{W}\| \le \|\mathbf{Q}_1\| + \|\mathbf{Q}_2\| + \|\mathbf{Q}_3\| + \|\mathbf{Q}_4\| \le \frac{1}{2}$.

Step 3. Bound $\|P_T\mathbf{W}_1\|_\infty$: Since $\|P_T\mathbf{W}_1\|_\infty = \|\mathbf{U}\mathbf{U}^\top\mathbf{W}_1 + \mathbf{W}_1\mathbf{U}\mathbf{U}^\top - \mathbf{U}\mathbf{U}^\top\mathbf{W}_1\mathbf{U}\mathbf{U}^\top\|_\infty \le 3\|\mathbf{U}\mathbf{U}^\top\mathbf{W}_1\|_\infty$, we only need to bound $\|\mathbf{U}\mathbf{U}^\top\mathbf{W}_1\|_\infty$. By Lemma A-6, the following inequalities hold with high probability

$$|(\mathbf{U}\mathbf{U}^\top\mathbf{Q}_1)_{ij}| \le c\left(\frac{\sqrt{(2s(1-\tau)/{K^*}^2 + (n-s)/(p{K^*}^2))\log n}}{K^*} + \frac{\log n}{p{K^*}^2}\right)$$
$$|(\mathbf{U}\mathbf{U}^\top\mathbf{Q}_2)_{ij}| \le c\left(\frac{\sqrt{(8sc_{\mathcal{C}}^2(1-\tau) + 4(n-s)\lambda^2 c_{\mathcal{A}^c}^2(1-t)/p)\log n}}{K^*} + \max\{\frac{2\lambda c_{\mathcal{A}^c}}{p}, 2c_{\mathcal{C}}\}\cdot\frac{\log n}{K^*}\right)$$
$$|(\mathbf{U}\mathbf{U}^\top\mathbf{Q}_3)_{ij}| \le c\left(\frac{\sqrt{(8sc_{\mathcal{C}}^2(1-\tau) + 4(n-s)\lambda^2 c_{\mathcal{A}^c}^2 t/(1-q))\log n}}{K^*} + \max\{\frac{2\lambda c_{\mathcal{A}}}{1-q}, 2c_{\mathcal{C}}\}\cdot\frac{\log n}{K^*}\right)$$

We now show that these upper bounds are less than $\frac{1}{6}\epsilon\min\{\lambda c_{\mathcal{A}}, \lambda c_{\mathcal{A}^c}, c_{\mathcal{C}}\}$. Since $c_{\mathcal{C}} \ge \lambda$ and $\min\{c_{\mathcal{A}}, c_{\mathcal{A}^c}\} \le 1$, $\frac{1}{6}\epsilon\min\{\lambda c_{\mathcal{A}}, \lambda c_{\mathcal{A}^c}, c_{\mathcal{C}}\} = \frac{1}{6}\epsilon\min\{\lambda c_{\mathcal{A}}, \lambda c_{\mathcal{A}^c}\}$. Note that

$$\epsilon\lambda c_{\mathcal{A}} = \lambda\cdot\frac{c_2}{\sqrt{t(1-t)}}\cdot\frac{1}{\lambda K^*}\cdot\sqrt{\frac{1-t}{t}} = \frac{c_2}{tK^*} \ge \frac{c_2}{K^*},$$
$$\epsilon\lambda c_{\mathcal{A}^c} = \lambda\cdot\frac{c_2}{\sqrt{t(1-t)}}\cdot\frac{1}{\lambda K^*}\cdot\sqrt{\frac{t}{1-t}} = \frac{c_2}{(1-t)K^*} \ge \frac{c_2}{K^*}.$$

We now verify that all the terms in $|(\mathbf{U}\mathbf{U}^\top\mathbf{Q}_1)_{ij}|$, $|(\mathbf{U}\mathbf{U}^\top\mathbf{Q}_2)_{ij}|$ and $|(\mathbf{U}\mathbf{U}^\top\mathbf{Q}_3)_{ij}|$ are less than $\frac{c_2}{6K^*}$. Since $1-\tau \le c_\tau\frac{K}{n}$ and $K^* \ge K \ge c_K\log n$, we have $\frac{2c_{\mathcal{C}}\log n}{K^*} = \sqrt{\frac{\log n}{K}}\cdot\frac{2c_0}{K^*} \le \frac{c_2}{K^*}$, $\frac{s(1-\tau)}{{K^*}^2} \le \frac{c_\tau}{K^*}$ and $sc_{\mathcal{C}}^2(1-\tau) \le \frac{c_0^2 c_\tau}{\log n}$, which implies that $\frac{\sqrt{2s(1-\tau)\log n/{K^*}^2}}{K^*} \le \sqrt{\frac{2c_\tau}{{K^*}^3}} \le \frac{c_2}{18K^*}$ and $\frac{\sqrt{8sc_{\mathcal{C}}^2(1-\tau)\log n}}{K^*} \le \frac{\sqrt{8c_0^2 c_\tau}}{K^*} \le \frac{c_2}{18K^*}$.

For $|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_1)_{ij}|$,

$$\frac{\log n}{p{K^*}^2} \leq \frac{\log n}{\frac{c_3 \log n}{K^*} \cdot {K^*}^2} = \frac{1}{c_3 K^*} \leq \frac{c_2}{18K^*},$$

$$\frac{\sqrt{(n-s)\log n/(p{K^*}^2)}}{K^*} \leq \frac{\sqrt{(n-s)\log n/(c_3(n-s)\log n)}}{K^*} = \frac{1}{\sqrt{c_3}K^*} \leq \frac{c_2}{18K^*}.$$

For $|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_2)_{ij}|$,

$$\frac{\lambda c_{\mathcal{A}^c} \log n}{pK^*} = \lambda \log n \cdot \sqrt{\frac{t}{1-t}} \cdot \frac{1}{pK^*} \leq \frac{\lambda \log n}{K^*}\sqrt{\frac{1}{t(1-t)}} \leq \frac{\lambda \log n}{K^*}\sqrt{\frac{\lambda^2 {K^*}^2}{c_3}} = \frac{\lambda^2 \log n}{\sqrt{c_3}} \leq \frac{c_0^2}{\sqrt{c_3}K^*} \leq \frac{c_2}{18K^*},$$

$$\frac{\sqrt{(n-s)\lambda^2 c_{\mathcal{A}^c}^2 (1-t)\log n/p}}{K^*} \leq \frac{\sqrt{\lambda^2 (n-s)\log n \cdot \frac{t}{1-t} \cdot \frac{1-t}{p}}}{K^*} \leq \frac{\sqrt{\lambda^2(n-s)\log n}}{K^*} \leq \frac{c_0}{K^*} \leq \frac{c_2}{18K^*}.$$

For $|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_3)_{ij}|$,

$$\frac{\lambda c_{\mathcal{A}} \log n}{(1-q)K^*} = \lambda \log n \cdot \sqrt{\frac{1-t}{t}} \cdot \frac{1}{(1-q)K^*} \leq \frac{\lambda \log n}{K^*}\sqrt{\frac{1}{t(1-t)}} \leq \frac{\lambda \log n}{K^*}\sqrt{\frac{\lambda^2 {K^*}^2}{c_3}} = \frac{\lambda^2 \log n}{\sqrt{c_3}} \leq \frac{c_0^2}{\sqrt{c_3}K^*} \leq \frac{c_2}{18K^*},$$

$$\frac{\sqrt{(n-s)\lambda^2 c_{\mathcal{A}}^2 t \log n/(1-q)}}{K^*} \leq \frac{\sqrt{\lambda^2 (n-s)\log n \cdot \frac{1-t}{t} \cdot \frac{t}{1-q}}}{K^*} \leq \frac{\sqrt{\lambda^2(n-s)\log n}}{K^*} \leq \frac{c_0}{K^*} \leq \frac{c_2}{18K^*}.$$

For $|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_4)_{ij}|$, we know that $|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_4)_{ij}| = 0$. Hence we conclude that $\|P_T \mathbf{W}_1\|_\infty \leq \frac{1}{2}\epsilon \min\{\lambda c_{\mathcal{A}}, \lambda c_{\mathcal{A}^c}, c_{\mathcal{C}}\}$.

Step 4. Verify Condition (c): From the construction of $\mathbf{W}_1$, we know that the inequalities (II)(III)(VI)(VII) hold. We now show that the other inequalities also hold. From Lemma A-3(c),

$$(1+\epsilon)\frac{1-p}{p} \leq (1-2\epsilon)\frac{1-t}{t} \iff (1+\epsilon)\lambda\frac{c_{\mathcal{A}^c}(1-p)}{p} \leq (1-2\epsilon)\lambda c_{\mathcal{A}}.$$

Thus, for $(i,j) \in \mathcal{R} \cap \mathcal{A} \cap \mathcal{C}^c \cap \mathcal{H}^c$,

$$(\mathbf{U}\mathbf{U}^\top + \mathbf{W}_1)_{ij} = \frac{1}{p}(\mathbf{U}\mathbf{U}^\top)_{ij} + (1+\epsilon)\lambda\frac{c_{\mathcal{A}^c}(1-p)}{p} \leq \frac{1}{pK^*} + (1-2\epsilon)\lambda c_{\mathcal{A}}$$

Recall that $\epsilon\lambda c_{\mathcal{A}} \geq \frac{c_2}{tK^*} \geq \frac{1}{pK^*}$, hence (I) holds. From Lemma A-3(d),

$$(1+\epsilon)\frac{q}{1-q} \leq (1-\epsilon)\frac{t}{1-t} \iff -(1+\epsilon)\lambda\frac{c_{\mathcal{A}}q}{1-q} \geq -(1-\epsilon)\lambda c_{\mathcal{A}^c},$$

which implies (IV). Since $\epsilon < 0.5$,

$$(1-\epsilon)c_{\mathcal{C}} \geq \frac{1}{2}c_{\mathcal{C}} = \frac{c_0}{2\sqrt{K\log n}} \geq \frac{5}{K^*}.$$

On the other hand, since $\tau \geq \frac{4}{5}$, for $(i,j) \in \mathcal{R} \cap \mathcal{A} \cap \mathcal{C} \cap \mathcal{H}^c$,

$$(\mathbf{U}\mathbf{U}^\top + \mathbf{W})_{ij} \leq \frac{1}{\tau K^*} + (1+\epsilon)c_{\mathcal{C}}\frac{1-\tau}{\tau} \leq \frac{5}{4K^*} + \frac{3}{8}c_{\mathcal{C}} \leq \frac{5}{4K^*} + \frac{15}{4K^*} = \frac{5}{K^*},$$

and for $(i,j) \in \mathcal{R}^c \cap \mathcal{A}^c \cap \mathcal{C} \cap \mathcal{H}^c$,

$$\mathbf{W}_1(i,j) + (1-\epsilon)c_{\mathcal{C}} \geq (1-\epsilon)c_{\mathcal{C}} - (1+\epsilon)c_{\mathcal{C}}\frac{1-\tau}{\tau} \geq \frac{1}{2}c_{\mathcal{C}} - \frac{3}{2}c_{\mathcal{C}}\frac{1-\tau}{\tau} \geq \frac{1}{8}c_{\mathcal{C}} \geq 0,$$

so (V) and (VIII) hold.

2.2. Construct $\mathbf{W}_2$

Step 1. Construct the dual certificate $\mathbf{W}_2$: We let $\mathbf{W}_2 = \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3$, where $\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3$ are defined as follows:

$$\mathbf{Q}_1(i,j) = \begin{cases} -(\mathbf{U}\mathbf{U}^\top)_{ij}, & (i,j) \in \mathcal{R} \cap \mathcal{A}^c \cap \mathcal{H} \\ \frac{1-\tau_1}{\tau_1}(\mathbf{U}\mathbf{U}^\top)_{ij}, & (i,j) \in \mathcal{R} \cap \mathcal{A} \cap \mathcal{H} \\ 0, & (i,j) \in \mathcal{H}^c \end{cases}$$

$$\mathbf{Q}_2(i,j) = \begin{cases} -\frac{3}{2}c_{\mathcal{C}}, & (i,j) \in \mathcal{R} \cap \mathcal{A}^c \cap \mathcal{H} \\ \frac{3(1-\tau_1)}{2\tau_1}c_{\mathcal{C}}, & (i,j) \in \mathcal{R} \cap \mathcal{A} \cap \mathcal{H} \\ 0, & (i,j) \in \mathcal{H}^c \end{cases}$$

$$\mathbf{Q}_3(i,j) = \begin{cases} \frac{3}{2}c_{\mathcal{C}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{A} \cap \mathcal{H} \\ -\frac{3(1-\tau_2)}{2\tau_2}c_{\mathcal{C}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{A}^c \cap \mathcal{H} \\ 0, & (i,j) \in \mathcal{H}^c \end{cases}$$

It can be easily verified that $\mathbb{E}[\mathbf{Q}_1] = \mathbb{E}[\mathbf{Q}_2] = \mathbb{E}[\mathbf{Q}_3] = 0$, and

$$|\mathbf{Q}_1(i,j)| \leq \frac{1}{K},\ |\mathbf{Q}_2(i,j)| \leq 2c_{\mathcal{C}},\ |\mathbf{Q}_3(i,j)| \leq 2c_{\mathcal{C}},$$

$$\mathrm{Var}[\mathbf{Q}_1(ij)] \leq \frac{2(1-\tau)}{K^2},\ \mathrm{Var}[\mathbf{Q}_2(ij)] \leq 5c_{\mathcal{C}}^2(1-\tau),\ \mathrm{Var}[\mathbf{Q}_3(ij)] \leq 5c_{\mathcal{C}}^2(1-\tau).$$

Step 2. Bound $\|\mathbf{W}_2\|$ and $\|P_T\mathbf{W}_2\|_\infty$: From Lemma A-5, the following inequalities hold with high probability:

$$\|\mathbf{Q}_1\| \leq c\left[\frac{\log n}{K} + \sqrt{\frac{2n(1-\tau)\log n}{K^2}}\right]$$
$$\|\mathbf{Q}_2\| \leq c\left[2c_{\mathcal{C}}\log n + \sqrt{5nc_{\mathcal{C}}^2(1-\tau)\log n}\right]$$
$$\|\mathbf{Q}_3\| \leq c\left[2c_{\mathcal{C}}\log n + \sqrt{5nc_{\mathcal{C}}^2(1-\tau)\log n}\right]$$

Recall that $1-\tau \leq c_\tau\frac{K}{n}$, $K \geq c_K \log n$ and $c_{\mathcal{C}} = \frac{c_0}{\sqrt{K\log n}}$. Thus, $\|\mathbf{W}\| \leq \|\mathbf{Q}_1\| + \|\mathbf{Q}_2\| + \|\mathbf{Q}_3\| \leq \frac{1}{2}$. Since $\|P_T\mathbf{W}_2\|_\infty = \|\mathbf{U}\mathbf{U}^\top\mathbf{W}_2 + \mathbf{W}_2\mathbf{U}\mathbf{U}^\top - \mathbf{U}\mathbf{U}^\top\mathbf{W}_2\mathbf{U}\mathbf{U}^\top\|_\infty \leq 3\|\mathbf{U}\mathbf{U}^\top\mathbf{W}_2\|_\infty$, we only need to bound $\|\mathbf{U}\mathbf{U}^\top\mathbf{W}_2\|_\infty$.

By Lemma A-6, the following inequalities hold with high probability

$$|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_1)_{ij}| \le c\left(\frac{\sqrt{2n(1-\tau)\log n}}{K^2} + \frac{\log n}{K^2}\right)$$
$$|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_2)_{ij}| \le c\left(\frac{\sqrt{5nc_{\mathcal{C}}^2(1-\tau)\log n}}{K} + \frac{2c_{\mathcal{C}}\log n}{K}\right)$$
$$|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_3)_{ij}| \le c\left(\frac{\sqrt{5nc_{\mathcal{C}}^2(1-\tau)\log n}}{K} + \frac{2c_{\mathcal{C}}\log n}{K}\right)$$

Since $K \ge c_K \log n$, $|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_1)_{ij}|$, $|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_2)_{ij}|$ and $|(\mathbf{U}\mathbf{U}^\top \mathbf{Q}_3)_{ij}|$ are all less than $\frac{1}{4}c_{\mathcal{C}}$ when c_K is large enough.

Step 3. Verify Condition (d): From the construction of $\mathbf{W}_2$, we know that the inequalities (X)(XI) hold. We now show that the other inequalities also hold. Observe that

$$\frac{1}{2}c_{\mathcal{C}} = \frac{c_0}{2\sqrt{K\log n}} \ge \frac{5}{K}.$$

On the other hand, since $\tau \ge \frac{4}{5}$, for $(i,j) \in \mathcal{R}\cap\mathcal{A}\cap\mathcal{H}$,

$$(\mathbf{U}\mathbf{U}^\top + \mathbf{W})_{ij} \le \frac{1}{\tau K} + \frac{3}{2}c_{\mathcal{C}}\frac{1-\tau}{\tau} \le \frac{5}{4K^*} + \frac{3}{8}c_{\mathcal{C}} \le \frac{5}{4K^*} + \frac{15}{4K^*} = \frac{5}{K^*},$$

and for $(i,j) \in \mathcal{R}^c\cap\mathcal{A}^c\cap\mathcal{H}$,

$$\mathbf{W}_{ij} + \frac{1}{2}c_{\mathcal{C}} = \frac{1}{2}c_{\mathcal{C}} - \frac{3}{2}c_{\mathcal{C}}\frac{1-\tau}{\tau} \ge \frac{1}{2}c_{\mathcal{C}} - \frac{3}{8}c_{\mathcal{C}} = \frac{1}{8}c_{\mathcal{C}} \ge 0,$$

so (IX) and (XII) hold.

2.3. The "Outlier-free" Case

The proofs in this setup are almost the same as above. Recall that K_i is the size of the ith cluster and s_i is the number of high confidence nodes in the ith cluster. In this case, we just need to let

$$\lambda = \frac{c_0}{\sqrt{\max\{K^*, \max_i\{\sum_{j\neq i}(K_i - s_i)\}\}\log n}}.$$

For the dual certificate $\mathbf{W}_1$, we let $\mathbf{W}_1 = \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3$, where $\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3$ are defined as follows:

$$\mathbf{Q}_1(i,j) = \begin{cases} -(\mathbf{U}\mathbf{U}^\top)_{ij}, & (i,j) \in \mathcal{R} \cap \mathcal{A}^c \cap \mathcal{H}^c \\ \frac{1-p_{ij}}{p_{ij}}(\mathbf{U}\mathbf{U}^\top)_{ij}, & (i,j) \in \mathcal{R} \cap \mathcal{C}^c \cap \mathcal{A} \cap \mathcal{H}^c \\ \frac{1-\tau_1}{\tau_1}(\mathbf{U}\mathbf{U}^\top)_{ij}, & (i,j) \in \mathcal{R} \cap \mathcal{C} \cap \mathcal{A} \cap \mathcal{H}^c \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbf{Q}_2(i,j) = \begin{cases} -(1+\epsilon)\lambda c_{\mathcal{A}^c}, & (i,j) \in \mathcal{R} \cap \mathcal{C}^c \cap \mathcal{A}^c \cap \mathcal{H}^c \\ \frac{1-p_{ij}}{p_{ij}}(1+\epsilon)\lambda c_{\mathcal{A}^c}, & (i,j) \in \mathcal{R} \cap \mathcal{C}^c \cap \mathcal{A} \cap \mathcal{H}^c \\ -(1+\epsilon)c_{\mathcal{C}}, & (i,j) \in \mathcal{R} \cap \mathcal{C} \cap \mathcal{A}^c \cap \mathcal{H}^c \\ \frac{1-\tau_1}{\tau_1}(1+\epsilon)c_{\mathcal{C}}, & (i,j) \in \mathcal{R} \cap \mathcal{C} \cap \mathcal{A} \cap \mathcal{H}^c \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbf{Q}_3(i,j) = \begin{cases} (1+\epsilon)\lambda c_{\mathcal{A}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{C}^c \cap \mathcal{A} \cap \mathcal{H}^c \\ -\frac{q_{ij}}{1-q_{ij}}(1+\epsilon)\lambda c_{\mathcal{A}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{C}^c \cap \mathcal{A}^c \cap \mathcal{H}^c \\ (1+\epsilon)c_{\mathcal{C}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{C} \cap \mathcal{A} \cap \mathcal{H}^c \\ -\frac{1-\tau_2}{\tau_2}(1+\epsilon)c_{\mathcal{C}}, & (i,j) \in \mathcal{R}^c \cap \mathcal{C} \cap \mathcal{A}^c \cap \mathcal{H}^c \\ 0, & \text{otherwise} \end{cases}$$

The only difference is that we remove $\mathbf{Q}_4$ since there are no outliers. Similar to Lemma A-5 and Lemma A-6, from the matrix Bernstein inequality, the followings hold with probability at least $1 - n^{-10}$:

$$\|\mathbf{Q}_1\| \leq c \left[\frac{\log n}{pK^*} + \sqrt{\frac{2n(1-\tau)}{{K^*}^2} + \frac{\max_i\{K_i - s_i\}}{p{K^*}^2}} \cdot \sqrt{\log n} \right]$$

$$\|\mathbf{Q}_2\| \leq c \left[\max\{\frac{2\lambda c_{\mathcal{A}^c}}{p}, 2c_{\mathcal{C}}\} \log n + \sqrt{8nc_{\mathcal{C}}^2(1-\tau) + \max_i\{K_i - s_i\} \frac{4\lambda^2 c_{\mathcal{A}^c}^2 (1-t)}{p}} \cdot \sqrt{\log n} \right]$$

$$\|\mathbf{Q}_3\| \leq c \left[\max\{\frac{2\lambda c_{\mathcal{A}}}{1-q}, 2c_{\mathcal{C}}\} \log n + \sqrt{8nc_{\mathcal{C}}^2(1-\tau) + \max_i\{\sum_{j\neq i}(K_j - s_j)\} \frac{4\lambda^2 c_{\mathcal{A}}^2 t}{1-q}} \cdot \sqrt{\log n} \right],$$

and

$$|(\mathbf{U}\mathbf{U}^\top\mathbf{Q}_1)_{ij}| \leq c \left(\frac{\sqrt{(2s(1-\tau)/{K^*}^2 + \max_i\{K_i - s_i\}/(p{K^*}^2)) \log n}}{K^*} + \frac{\log n}{p{K^*}^2} \right)$$

$$|(\mathbf{U}\mathbf{U}^\top\mathbf{Q}_2)_{ij}| \leq c \left(\frac{\sqrt{(8sc_{\mathcal{C}}^2(1-\tau) + 4\max_i\{K_i - s_i\}\lambda^2 c_{\mathcal{A}^c}^2(1-t)/p) \log n}}{K^*} + \max\{\frac{2\lambda c_{\mathcal{A}^c}}{p}, 2c_{\mathcal{C}}\} \cdot \frac{\log n}{K^*} \right)$$

$$|(\mathbf{U}\mathbf{U}^\top\mathbf{Q}_3)_{ij}| \leq c \left(\frac{\sqrt{(8sc_{\mathcal{C}}^2(1-\tau) + 4\max_i\{\sum_{j\neq i}(K_j - s_j)\}\lambda^2 c_{\mathcal{A}^c}^2 t/(1-q)) \log n}}{K^*} + \max\{\frac{2\lambda c_{\mathcal{A}}}{1-q}, 2c_{\mathcal{C}}\} \cdot \frac{\log n}{K^*} \right).$$

Since $\max_i\{K_i - s_i\} \leq \max_i\{\sum_{j\neq i}(K_j - s_j)\}$, the terms $\max_i\{K_i - s_i\}$ in these inequalities can be replaced by $\max_i\{\sum_{j\neq i}(K_j - s_j)\}$. Then one can prove the desired result easily by following the same calculation in Section 2.1.

3. Proof of Theorem 2

Recall that the graph has n nodes, r clusters and $n-\sum_{i=1}^{r} K_i$ outliers. K is the minimum cluster size, i.e., $K=\min_i K_i$. For clarity, the constants may vary from line to line.

Step 1. The n nodes are uniformly randomly separated into m groups which form m small graphs $\{g_1,\cdots,g_m\}$. For each $i\in[n]$ and $j\in[m]$, node i is assigned to graph g_j with probability $\frac{1}{m}$. For $g\in\{g_1,\cdots,g_m\}$, let K_i^g be the number of the nodes in the ith cluster that are assigned to graph g and let n^g be the number of nodes in g. Clearly, K_i^g and n^g are two random variables whose expected values are $\mathbb{E}[K_i^g]=\frac{K_i}{m}$ and $\mathbb{E}[n^g]=\frac{n}{m}$, respectively. From the Hoeffding's inequality,

$$\mathbb{P}[|K_i^g-\mathbb{E}[K_i^g]|\geq t]\leq 2\exp\left(-\frac{2t^2}{K_i}\right).$$

For constant $\rho<1$, let $t=\frac{1-\rho}{2(1+\rho)m}K_i$, then we have

$$\mathbb{P}\left[\left|K_i^g-\frac{K_i}{m}\right|\geq\frac{1-\rho}{2(1+\rho)m}K_i\right]\leq 2\exp\left(-\frac{(1-\rho)^2K_i}{2(1+\rho)^2m^2}\right)\leq 2\exp\left(-\frac{(1-\rho)^2K}{2(1+\rho)^2m^2}\right).$$

In other words, $\frac{1+3\rho}{2(1+\rho)m}K_i\leq K_i^g\leq\frac{3+\rho}{2(1+\rho)m}K_i$ holds with probability at least $1-2\exp\left(-\frac{(1-\rho)^2K}{2(1+\rho)^2m^2}\right)$. Similarly, $\frac{1+3\rho}{2(1+\rho)m}n\leq n^g\leq\frac{3+\rho}{2(1+\rho)m}n$ holds with probability at least $1-2\exp\left(-\frac{(1-\rho)^2n}{2(1+\rho)^2m^2}\right)$. By the union bound, we have

$$\frac{1+3\rho}{2(1+\rho)m}K_i\leq K_i^g\leq\frac{3+\rho}{2(1+\rho)m}K_i \text{ for } i\in[r], g\in\{g_1,\cdots,g_m\} \text{ and } \frac{1+3\rho}{2(1+\rho)m}n\leq n^g\leq\frac{3+\rho}{2(1+\rho)m}n \quad \text{(A-1)}$$

hold with probability at least $1-2(mr+1)\exp\left(-\frac{(1-\rho)^2K}{2(1+\rho)^2m^2}\right)$. Since $m\leq\frac{1-\rho}{4(1+\rho)}\sqrt{\frac{K}{\log n}}$ and $mr+1\leq\frac{mn}{K}+1\leq n$, (A-1) holds with probability at least $1-n^{-6}$.

Step 2. After all the subgraphs are generated, we perform algorithm $\mathfrak{A}$ on each subgraph $g\in\{g_1,\cdots,g_m\}$. Let $\mathcal{S}_g$ be the set of the recovered clusters in g. Since $\mathfrak{A}$ is $\boldsymbol{\lambda}$-workable and $\frac{1+3\rho}{2(1+\rho)m}K_i\leq K_i^g\leq\frac{3+\rho}{2(1+\rho)m}K_i$ for $i\in[r]$ holds with high probability, we know that when (p,q) is in $\mathfrak{C}(n/m,K_1/m,\cdots,K_r/m,\boldsymbol{\lambda},\mathcal{I})$, $\mathcal{S}_g$ satisfies that 1) for each $i\in\mathcal{I}$, there exists $\mathcal{C}_i\in\mathcal{S}_g$ such that $\mathcal{C}_i$ a subset of the ith cluster and $|\mathcal{C}_i|\geq\lambda_iK_i^g\geq\frac{1+3\rho}{2(1+\rho)m}\lambda_iK_i$, and 2) for each cluster $\mathcal{C}\in\mathcal{S}_g\setminus\bigcup_{i\in\mathcal{I}}\mathcal{C}_i$, we have $|\mathcal{C}|<\min_{i\in\mathcal{I}}\rho\lambda_iK_i^g\leq\frac{3\rho+\rho^2}{2(1+\rho)m}\min_{i\in\mathcal{I}}\lambda_iK_i$, with probability at least $1-n^{-2}$. By the union bound, with probability at least $1-n^{-1}$, all of $\mathcal{S}_{g_1},\cdots,\mathcal{S}_{g_m}$ satisfy these two properties.

In the "breaking up small clusters" step, note that threshold T satisfies $\frac{T}{\min_{i\in\mathcal{I}}\lambda_iK_i}\in(\frac{3\rho+\rho^2}{2(1+\rho)m},\frac{1+3\rho}{2(1+\rho)m})$. For each $\mathcal{S}_g\in\{\mathcal{S}_{g_1},\cdots,\mathcal{S}_{g_m}\}$, after breaking up the clusters in $\mathcal{S}_g$ whose size is less than T, $\mathcal{S}_g$ becomes

$$\mathcal{S}_g^0=\bigcup_{i\in\mathcal{I}}\mathcal{C}_i\cup\left\{\{u\}:\forall u\in\mathcal{C},\forall\mathcal{C}\in\mathcal{S}_g\setminus\bigcup_{i\in\mathcal{I}}\mathcal{C}_i\right\}.$$

Then for each $\mathcal{C}_i\in\mathcal{S}_g^0$, $\mathcal{C}_i$ is uniformly randomly divided into l clusters, namely, $\{\mathcal{C}_i^1,\cdots,\mathcal{C}_i^l\}$. Since w.h.p

$$|\mathcal{C}_i|\geq\frac{1+3\rho}{2(1+\rho)m}\min_{j\in\mathcal{I}}\lambda_jK_j,\ \forall i\in\mathcal{I},$$

by the Hoeffding's inequality and the union bound, one can easily verify that for all $\mathcal{S}_g^0\in\{\mathcal{S}_{g_1}^0,\cdots,\mathcal{S}_{g_m}^0\}$ and $\mathcal{C}_i\in\mathcal{S}_g^0$, the following inequality holds with probability at least $1 - n^{-6}$ when $l \le \frac{1}{4}\sqrt{\frac{(1+3\rho)\min_{i\in\mathcal{I}}\lambda_i K_i}{2(1+\rho)m\log n}}$ or $l = 1$:

$$|\mathcal{C}_i^k| \ge \frac{1+3\rho}{4(1+\rho)ml}\min_{j\in\mathcal{I}}\lambda_j K_j,\ \forall i \in \mathcal{I}, k \in [l].$$

Therefore, after the "breaking up small clusters" step, $\mathcal{S}_g$ becomes

$$\mathcal{S}_g^1 = \bigcup_{i\in\mathcal{I}}\bigcup_{k\in[l]}\mathcal{C}_i^k \cup (\mathcal{S}_g^0 \setminus \bigcup_{i\in\mathcal{I}}\mathcal{C}_i).$$

For simplicity, we use $\mathcal{S}_g$ instead of $\mathcal{S}_g^1$ in the following parts.

Step 3. We now analyze the properties of the fused graph. We view each cluster $\mathcal{U}_i$ in $\bigcup_{i=1}^m \mathcal{S}_{g_i}$ as a super node V_i. If $|\mathcal{U}_i| > 1$, V_i is added into the "high confidence node" set $\mathcal{H}$, which means V_i is a high confidence node in the fused graph. Otherwise, V_i is an ordinary node. For two nodes V_i and V_j, we say "V_i and V_j are in the same cluster" if the nodes in $\mathcal{U}_i$ and $\mathcal{U}_j$ belong to the same cluster. From the construction of the edge between V_i and V_j, we know that when V_i and V_j are both ordinary nodes, $E_{ij} = 1$ with probability at least p if V_i and V_j are in the same cluster or $E_{ij} = 1$ with probability at most q otherwise. If one of V_i and V_j is a high confidence node, we compute

$$\hat{E}(V_i, V_j) := \frac{\sum_{u\in\mathcal{U}_i}\sum_{v\in\mathcal{U}_j} A_{uv}}{\sum_{u\in\mathcal{U}_i}\sum_{v\in\mathcal{U}_j} 1}.$$

We let $X \triangleq \hat{E}(V_i, V_j)$ and $Z \triangleq \sum_{u\in\mathcal{U}_i}\sum_{v\in\mathcal{U}_j} 1$. Clearly, V_i and V_j being in the same cluster means that $\mathbb{E}[A_{uv}] \ge p$ for any $u \in \mathcal{U}_i$ and $v \in \mathcal{U}_j$, which implies that $\mathbb{E}[X] \ge p$. From the Hoeffding's inequality, we have

$$\mathbb{P}[|X - \mathbb{E}[X]| \ge \theta] \le 2\exp\left(-2Z\theta^2\right) \le 2\exp\left(-\frac{1+3\rho}{2(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i \theta^2\right).$$

Thus, $X \ge p - \theta$ holds with probability at least $1 - 2\exp\left(-\frac{1+3\rho}{2(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i\theta^2\right)$. Similarly, V_i and V_j being in different clusters means $\mathbb{E}[A_{uv}] \le q$ for any $u \in \mathcal{U}_i$ and $v \in \mathcal{U}_j$, which implies that $\mathbb{E}[X] \le q$. From the Hoeffding's inequality, we have $X \le q + \theta$ holds with probability at least $1 - 2\exp\left(-\frac{1+3\rho}{2(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i\theta^2\right)$.

In Algorithm 2, we set $E_{ij} = 1$ if $X \ge t$ or $E_{ij} = 0$ otherwise. Hence from the analysis above, we know that $E_{ij} = 1$ with probability at least $1 - 2\exp\left(-\frac{1+3\rho}{2(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i(p-t)^2\right)$ if V_i and V_j are in the same cluster, while $E_{ij} = 1$ with probability at most $2\exp\left(-\frac{1+3\rho}{2(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i(t-q)^2\right)$ if V_i and V_j are in different clusters. Recall that $t \in (\frac{1}{4}p + \frac{3}{4}q, \frac{3}{4}p + \frac{1}{4}q)$. Since $p - q \ge c_2\sqrt{\frac{(1+\rho)ml\log\frac{n}{K}}{(1+3\rho)\min_{i\in\mathcal{I}}\lambda_i K_i}}$, we have

$$\tau \triangleq 1 - 2\exp\left(-\frac{1+3\rho}{32(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i(p-q)^2\right) \ge 1 - c_\tau\frac{K}{n},$$

where c_τ and c_2 are universal constants. Then we have

- $E_{ij} = 1$ with probability at least p if V_i and V_j are ordinary and in the same cluster;
- $E_{ij} = 1$ with probability at most q if V_i and V_j are ordinary and in different clusters;
- $E_{ij} = 1$ with probability at least τ if V_i or V_j is high confident and they are in the same cluster;
- $E_{ij} = 1$ with probability at most $1 - \tau$ if V_i or V_j is high confident and they are in different clusters;

Step 4. We perform the graph clustering algorithm (1) on the fused graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. From the analysis above, we know that the number of the high confidence nodes in $\mathcal{G}$ is at least $ml|\mathcal{I}|$, the size of the smallest cluster in $\mathcal{G}$ that contains no ordinary nodes is at least ml, the total number of the ordinary nodes in $\mathcal{G}$ is at most $n - \sum_{i\in\mathcal{I}} \lambda_i K_i$, and the total number of the nodes is at least mr. Let $\mathcal{J}$ be the set $\{i \in \mathcal{I} : \lambda_i \neq 1\}$, then the size of the smallest cluster that contains at least one ordinary node is at least $S(m,l) = \min\{\min_{i\in\mathcal{J}}\{ml + (1-\lambda_i)K_i\}, \min_{i\in\mathcal{I}^c} K_i\}$. From Theorem 1, if $ml \geq c_3 \log n$ and

$$\frac{p-q}{\sqrt{p(1-q)}} \geq c_1 \max\left\{\frac{\sqrt{(n-\sum_{i\in\mathcal{I}}\lambda_i K_i)\log n}}{S(m,l)}, \sqrt{\frac{\log n}{S(m,l)}}\right\},$$

then the clusters in graph $\mathcal{G}$ can be correctly recovered with probability at least $1-(mr)^{-10}$.

Overall, if $c_3 \log n \leq m \leq \frac{1-\rho}{4(1+\rho)}\sqrt{\frac{K}{\log n}}$ and

$$p - q \geq \max\left\{c_1\sqrt{p(1-q)}\max\left\{\frac{\sqrt{(n-\sum_{i\in\mathcal{I}}\lambda_i K_i)\log n}}{S(m,l)}, \sqrt{\frac{\log n}{S(m,l)}}\right\}, c_2\sqrt{\frac{(1+\rho)ml\log\frac{n}{K}}{(1+3\rho)\min_{i\in\mathcal{I}}\lambda_i K_i}}\right\}$$

hold, Algorithm 1 outputs the true clusters w.h.p. By minimizing the right hand side over l, we obtain this theorem.

4. Proof of Theorem 3

We use the same notation as that in the proof of Theorem 2.

Step 1. This step is similar to Step 1 in the proof of Theorem 2. The n nodes are uniformly randomly separated into m groups which form m subgraphs $\{g_1, \cdots, g_m\}$. As shown above, we can prove that

$$\frac{1+3\rho}{2(1+\rho)m}K_i \leq K_i^g \leq \frac{3+\rho}{2(1+\rho)m}K_i \text{ for } i \in [r], g \in \{g_1,\cdots,g_m\} \text{ and } \frac{1+3\rho}{2(1+\rho)m}n \leq n^g \leq \frac{3+\rho}{2(1+\rho)m}n \quad \text{(A-2)}$$

hold with probability at least $1-n^{-6}$ since $m \leq \frac{1-\rho}{4(1+\rho)}\sqrt{\frac{K}{\log n}}$.

Step 2. After the subgraphs are obtained, we perform algorithm $\mathfrak{A}$ on each subgraph $g \in \{g_1,\cdots,g_m\}$. Let $\mathcal{S}_g$ be the set of the recovered clusters in g. Since algorithm $\mathfrak{A}$ is $(\boldsymbol{\lambda}, \mathcal{I}, \boldsymbol{\epsilon})$-pseudo-workable and $\frac{1+3\rho}{2(1+\rho)m}K_i \leq K_i^g \leq \frac{3+\rho}{2(1+\rho)m}K_i$ for $i \in [r]$ holds with high probability, when (p,q) is in $\mathfrak{C}(n/m, K_1/m, \cdots, K_r/m, \boldsymbol{\lambda}, \mathcal{I}, \boldsymbol{\epsilon})$, we know that with probability at least $1-n^{-2}$, $\mathcal{S}_g$ satisfies that 1) for each $i \in \mathcal{I}$, there exists $\mathcal{C}_i \in \mathcal{S}_g$ so that $\mathcal{C}_i$ contains at least $\lambda_i K_i^g$ nodes in the ith cluster and at most $\epsilon_i K_i^g$ nodes not in the ith cluster, which implies that $|\mathcal{C}_i| \geq \lambda_i K_i^g \geq \frac{1+3\rho}{2(1+\rho)m}\lambda_i K_i$, 2) for each cluster $\mathcal{C} \in \mathcal{S}_g \setminus \bigcup_{i\in\mathcal{I}}\mathcal{C}_i$, we have $|\mathcal{C}| < \min_{i\in\mathcal{I}} \rho\lambda_i K_i^g \leq \frac{3\rho+\rho^2}{2(1+\rho)m}\min_{i\in\mathcal{I}}\lambda_i K_i$. By the union bound, with probability at least $1-n^{-1}$, all of $\mathcal{S}_{g_1},\cdots,\mathcal{S}_{g_m}$ satisfy these two properties.

In the "breaking up small clusters" step, note that $\frac{T}{\min_{i\in\mathcal{I}}\lambda_i K_i} \in (\frac{3\rho+\rho^2}{2(1+\rho)m}, \frac{1+3\rho}{2(1+\rho)m})$, and each $\mathcal{C}_i \in \mathcal{S}_g$ is divided into l clusters $\{\mathcal{C}_i^1, \cdots, \mathcal{C}_i^l\}$ while the clusters in $\mathcal{S}_g \setminus \bigcup_{i\in\mathcal{I}}\mathcal{C}_i$ are broken up to single nodes. By the Hoeffding's inequality and the union bound, we have for all $\mathcal{S}_g \in \{\mathcal{S}_{g_1},\cdots,\mathcal{S}_{g_m}\}$ and $\mathcal{C}_i \in \mathcal{S}_g$

$$|\mathcal{C}_i^k| \geq \frac{1+3\rho}{4(1+\rho)ml}\lambda_i K_i, \ \forall i \in \mathcal{I}, k \in [l] \quad \text{(A-3)}$$

holds with probability at least $1-n^{-6}$ when $l \leq \frac{1}{4}\sqrt{\frac{(1+3\rho)\min_{i\in\mathcal{I}}\lambda_i K_i}{2(1+\rho)m\log n}}$ or $l=1$. Then after this step, $\mathcal{S}_g$ becomes

$$\mathcal{S}_g = \bigcup_{i\in\mathcal{I}}\bigcup_{k\in[l]}\mathcal{C}_i^k \cup \left\{\{u\} : \forall u \in \mathcal{C}, \forall \mathcal{C} \in \mathcal{S}_g \setminus \bigcup_{i\in\mathcal{I}}\mathcal{C}_i\right\}.$$

Step 3. In the "building the fused graph" step, we view each cluster $\mathcal{U}_i$ in $\bigcup_{i=1}^m \mathcal{S}_{g_i}$ as a super node V_i. If $|\mathcal{U}_i| > 1$, V_i

is added into the "high confidence node" set $\mathcal{H}$, which means V_i is a high confidence node. Otherwise, V_i is an ordinary node. For two nodes V_i and V_j, from the construction of the edge between V_i and V_j, we know that if V_i and V_j are both ordinary nodes, then $E_{ij} = 1$ with probability at least p if V_i and V_j are in the same cluster while $E_{ij} = 1$ with probability at most q if V_i and V_j are in different clusters. If one of V_i and V_j is high confident, we compute

$$\hat{E}(V_i, V_j) = \frac{\sum_{u\in\mathcal{U}_i}\sum_{v\in\mathcal{U}_j} A_{uv}}{\sum_{u\in\mathcal{U}_i}\sum_{v\in\mathcal{U}_j} 1}.$$

Note that because $\mathfrak{A}$ is $(\boldsymbol{\lambda}, \mathcal{I}, \boldsymbol{\epsilon})$-pseudo-workable, $\mathcal{U}_i$ may contain some outliers when $|\mathcal{U}_i| > 1$. We denote the inlier and outlier nodes in $\mathcal{U}_i$ by $\bar{\mathcal{U}}_i$ and $\hat{\mathcal{U}}_i$, respectively. Suppose that the inlier nodes belong to the kth cluster, then from Inequality (A-2) and (A-3), we know that $|\bar{\mathcal{U}}_i| \geq \frac{1+3\rho}{4(1+\rho)ml}\lambda_k K_k$ and $|\hat{\mathcal{U}}_i| \leq \frac{3+\rho}{2(1+\rho)m}\epsilon_k K_k$ hold with high probability.

We first consider the case that V_i and V_j are in the same cluster, e.g., V_i, V_j belong to the kth cluster. Then

$$\hat{E}(V_i, V_j) \geq \frac{\sum_{u\in\bar{\mathcal{U}}_i}\sum_{v\in\bar{\mathcal{U}}_j} A_{uv}}{(|\bar{\mathcal{U}}_i| + |\hat{\mathcal{U}}_i|)(|\bar{\mathcal{U}}_j| + |\hat{\mathcal{U}}_j|)} \geq \frac{\sum_{u\in\bar{\mathcal{U}}_i}\sum_{v\in\bar{\mathcal{U}}_j} A_{uv}}{|\bar{\mathcal{U}}_i||\bar{\mathcal{U}}_j|}\left(1 - \frac{|\hat{\mathcal{U}}_i|}{|\bar{\mathcal{U}}_i|} - \frac{|\hat{\mathcal{U}}_j|}{|\bar{\mathcal{U}}_j|} - \frac{|\hat{\mathcal{U}}_i||\hat{\mathcal{U}}_j|}{|\bar{\mathcal{U}}_i||\bar{\mathcal{U}}_j|}\right).$$

Since $l \leq \frac{p-q}{72}\min_{k\in\mathcal{I}}\frac{\lambda_k}{\epsilon_k}$, we have $\frac{|\hat{\mathcal{U}}_i|}{|\bar{\mathcal{U}}_i|}, \frac{|\hat{\mathcal{U}}_j|}{|\bar{\mathcal{U}}_j|} \leq \frac{2(3+\rho)}{1+3\rho}\cdot\frac{\epsilon_k l}{\lambda_k} \leq \frac{6\epsilon_k l}{\lambda_k} \leq 1$, which implies that

$$\hat{E}(V_i, V_j) \geq \frac{\sum_{u\in\bar{\mathcal{U}}_i}\sum_{v\in\bar{\mathcal{U}}_j} A_{uv}}{|\bar{\mathcal{U}}_i||\bar{\mathcal{U}}_j|}\left(1 - \max_{k\in\mathcal{I}}\frac{18\epsilon_k l}{\lambda_k}\right).$$

From the Hoeffding's inequality, one can easily verify that $\hat{E}(V_i, V_j) \geq (p-\theta)\left(1 - \max_{k\in\mathcal{I}}\frac{18\epsilon_k l}{\lambda_k}\right)$ holds with probability at least $1 - 2\exp\left(-\frac{1+3\rho}{2(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i\theta^2\right)$.

Similarly, when V_i and V_j are in different clusters, we have

$$\hat{E}(V_i, V_j) \leq \frac{\sum_{u\in\bar{\mathcal{U}}_i}\sum_{v\in\bar{\mathcal{U}}_j} A_{uv} + |\bar{\mathcal{U}}_i||\hat{\mathcal{U}}_j| + |\hat{\mathcal{U}}_i||\bar{\mathcal{U}}_j| + |\hat{\mathcal{U}}_i||\hat{\mathcal{U}}_j|}{|\bar{\mathcal{U}}_i||\bar{\mathcal{U}}_j|} \leq \frac{\sum_{u\in\bar{\mathcal{U}}_i}\sum_{v\in\bar{\mathcal{U}}_j} A_{uv}}{|\bar{\mathcal{U}}_i||\bar{\mathcal{U}}_j|} + \max_{k\in\mathcal{I}}\frac{18\epsilon_k l}{\lambda_k}.$$

From the Hoeffding's inequality, we know that $\hat{E}(V_i, V_j) \leq q + \theta + \max_{k\in\mathcal{I}}\frac{18\epsilon_k l}{\lambda_k}$ holds with probability at least $1 - 2\exp\left(-\frac{1+3\rho}{2(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i\theta^2\right)$.

Let $\varphi \triangleq \max_{k\in\mathcal{I}}\frac{18\epsilon_k l}{\lambda_k}$. Since $l \leq \frac{p-q}{72}\min_{k\in\mathcal{I}}\frac{\lambda_k}{\epsilon_k}$, $\varphi \leq \frac{1}{4}(p-q)$, which implies that the inequality $\frac{1}{4}p + \frac{3}{4}q + \varphi \leq (\frac{3}{4}p + \frac{1}{4}q)(1-\varphi)$ hold. Therefore, there exists t such that $\frac{1}{4}p + \frac{3}{4}q + \varphi \leq t \leq (\frac{3}{4}p + \frac{1}{4}q)(1-\varphi)$. In Algorithm 2, we set $E_{ij} = 1$ if $X \geq t$ or $E_{ij} = 0$ otherwise. Hence $E_{ij} = 1$ with probability at least $1 - 2\exp\left(-\frac{1+3\rho}{2(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i(p - \frac{t}{1-\varphi})^2\right)$ if V_i and V_j are in the same cluster, while $E_{ij} = 1$ with probability at most $2\exp\left(-\frac{1+3\rho}{2(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i(t - q - \varphi)^2\right)$ if V_i and V_j are in different clusters. Since $\frac{1}{4}p + \frac{3}{4}q + \varphi \leq t \leq (\frac{3}{4}p + \frac{1}{4}q)(1-\varphi)$, we have

$$\frac{1}{4}p + \frac{3}{4}q \leq \frac{t}{1-\varphi} \leq \frac{3}{4}p + \frac{1}{4}q, \text{ and } \frac{1}{4}p + \frac{3}{4}q \leq t - c \leq \frac{3}{4}p + \frac{1}{4}q.$$

When $p - q \geq c_2\sqrt{\frac{(1+\rho)ml\log\frac{n}{K}}{(1+3\rho)\min_{i\in\mathcal{I}}\lambda_i K_i}}$, we have

$$\tau \triangleq 1 - 2\exp\left(-\frac{1+3\rho}{32(1+\rho)ml}\min_{i\in\mathcal{I}}\lambda_i K_i(p-q)^2\right) \geq 1 - c_\tau\frac{K}{n},$$

where c_τ and c_2 are universal constants. Hence we have

- $E_{ij} = 1$ with probability at least p if V_i and V_j are ordinary and in the same cluster;
- $E_{ij} = 1$ with probability at most q if V_i and V_j are ordinary and in different clusters;
- $E_{ij} = 1$ with probability at least τ if V_i or V_j is high confident and the inlier nodes of $\mathcal{U}_i$ and $\mathcal{U}_j$ are in the same cluster;
- $E_{ij} = 1$ with probability at most $1 - \tau$ if V_i or V_j is high confident and the inlier nodes of $\mathcal{U}_i$ and $\mathcal{U}_j$ are in different clusters;

Step 4. We run the graph clustering algorithm (1) on the fused graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. From the analysis above, we know that the number of the high confidence nodes in $\mathcal{G}$ is at least $ml|\mathcal{I}|$, the size of the smallest cluster in $\mathcal{G}$ that contains no ordinary nodes is at least ml, the total number of the ordinary nodes in $\mathcal{G}$ is at most $n - \sum_{i \in \mathcal{I}} \lambda_i K_i$, and the total number of the nodes is at least mr. Let $\mathcal{J}$ be the set $\{i \in \mathcal{I} : \lambda_i \neq 1\}$, then the size of the smallest cluster that contains at least one ordinary node is at least $S(m,l) = \min\{\min_{i \in \mathcal{J}}\{ml + [(1-\lambda_i)K_i - \sum_{j \in \mathcal{I}, j \neq i} \epsilon_j K_j]_+\}, \max\{\min_{i \in \mathcal{I}^c} K_i - \sum_{j \in \mathcal{I}} \epsilon_j K_j, 1\}\}$. From Theorem 1, if $ml \geq c_3 \log n$ and

$$\frac{p-q}{\sqrt{p(1-q)}} \geq c_1 \max\left\{\frac{\sqrt{(n - \sum_{i=1}^r \lambda_i K_i)\log n}}{S(m,l)}, \sqrt{\frac{\log n}{S(m,l)}}\right\},$$

then the clusters in graph $\mathcal{G}$ can be correctly recovered with probability at least $1 - (mr)^{-10}$.

Overall, if $c_3 \log n \leq m \leq \frac{1-\rho}{4(1+\rho)}\sqrt{\frac{K}{\log n}}$ and

$$p - q \geq \max\left\{c_1\sqrt{p(1-q)} \max\left\{\frac{\sqrt{(n - \sum_{i=1}^r \lambda_i K_i)\log n}}{S(m,l)}, \sqrt{\frac{\log n}{S(m,l)}}\right\}, c_2\sqrt{\frac{(1+\rho)ml\log\frac{n}{K}}{(1+3\rho)\min_{i\in\mathcal{I}}\lambda_i K_i}}, 72\max_{i\in\mathcal{I}}\frac{\epsilon_i}{\lambda_i}\right\}$$

hold, the output of Algorithm 1 contains at most $\sum_{i=1}^r \epsilon_i K_i$ misclassified nodes.

5. Proof of Corollary 1

Since algorithm $\mathfrak{A}$ recovers clusters by solving (1) with $\mathcal{C} = \emptyset$, we have that $\mathfrak{A}$ is $(\mathbf{1}, [r])$-workable with $\rho = 0$ and set $\mathfrak{C}$ defined by

$$\mathfrak{C} = \left\{(p,q) : \frac{p-q}{\sqrt{p(1-q)}} \geq c_1 \frac{\sqrt{n\log n}}{K}\right\}$$

where K is the size of the smallest cluster in the graph and c_1 is a universal constant.

Then from Theorem 2,we know that in order to recover the true clusters, (p,q) should satisfy

$$\frac{p-q}{\sqrt{p(1-q)}} \geq c_1 \frac{\sqrt{mn\log n}}{K}.$$

and

$$\begin{aligned} p - q &\geq \min_{\bar{l} \geq l \geq 1} \max\left\{c_1\sqrt{p(1-q)}\max\left\{\frac{\sqrt{(n - \sum_{i\in\mathcal{I}}\lambda_i K_i)\log n}}{S(m,l)}, \sqrt{\frac{\log n}{S(m,l)}}\right\}, c_2\sqrt{\frac{(1+\rho)ml\log\frac{n}{K}}{(1+3\rho)\min_{i\in\mathcal{I}}\lambda_i K_i}}\right\} \\ &= \min_{\bar{l}\geq l\geq 1} c_2\sqrt{\frac{(1+\rho)ml\log\frac{n}{K}}{(1+3\rho)\min_{i\in\mathcal{I}}\lambda_i K_i}} = c_2\sqrt{\frac{m\log\frac{n}{K}}{K}}. \end{aligned}$$

Hence we obtain Corollary 1.

6. Proof of Corollary 2

Recall that algorithm $\mathfrak{A}$ recovers clusters by solving (1) with $\mathcal{C} = \emptyset$. For a graph containing n nodes and r clusters with size $\{K_1, \cdots, K_r\}$, we define

$$u = c_3 \frac{\sqrt{p(1-q)n}}{p-q} \log^2 n, \text{ and } l = c_4 \frac{\sqrt{p(1-q)n}}{p-q}.$$

Let $\mathcal{K}_u$ be the set of the clusters whose sizes are greater than or equal to u and $\mathcal{K}_l$ be the set of the clusters whose sizes are less than or equal to l. Let $\mathbf{Y}^*$ be the true adjacent matrix, then by Theorem 1 in (Ailon et al., 2013), if each cluster is included in either $\mathcal{K}_u$ or $\mathcal{K}_l$, then $(\hat{\mathbf{Y}}, \mathbf{A} - \hat{\mathbf{Y}})$ is an optimal solution of (1) with probability at least $1 - n^{-3}$, where $\hat{\mathbf{Y}}$ is defined as

$$\hat{\mathbf{Y}}(i,j) = \begin{cases} \mathbf{Y}^*(i,j), & \text{node } i \text{ and } j \text{ belongs to the same cluster in } \mathcal{K}_u \\ 0, & \text{otherwise.} \end{cases}$$

Let $\mathcal{I} = \{i : K_i \geq u\}$ and $\boldsymbol{\lambda}$ be a vector whose entry $\lambda_i = 1$ if $i \in \mathcal{I}$ or 0 otherwise. The conditions above related to (p, q) is denoted by $\mathfrak{C}(n, K_1, \cdots, K_r, \boldsymbol{\lambda}, \mathcal{I})$. Clearly, $\mathfrak{A}$ is $(\boldsymbol{\lambda}, \mathcal{I})$-workable with $\rho = 0$ and set $\mathfrak{C}(n, K_1, \cdots, K_r, \boldsymbol{\lambda}, \mathcal{I})$.

From Theorem 2, in order to recover the true clusters, (p, q) should be in $\mathfrak{C}(n, K_1/m, \cdots, K_r/m, \boldsymbol{\lambda}, \mathcal{I})$, which means that for all $i \in [r]$, either $K_i \geq u$ or $K_i \leq l$ where

$$u = c_3 \frac{\sqrt{p(1-q)mn}}{p-q} \log^2 n, \text{ and } l = c_4 \frac{\sqrt{p(1-q)mn}}{p-q}.$$

Besides, (p, q) should also satisfy

$$p - q \geq \min_{\bar{l} \geq l \geq 1} \max \left\{ c_1 \sqrt{p(1-q)} \max \left\{ \frac{\sqrt{(n - \sum_{i \in \mathcal{I}} \lambda_i K_i) \log n}}{S(m,l)}, \sqrt{\frac{\log n}{S(m,l)}} \right\}, c_2 \sqrt{\frac{(1+\rho) ml \log \frac{n}{K}}{(1+3\rho) \min_{i \in \mathcal{I}} \lambda_i K_i}} \right\}, \tag{A-4}$$

where $S(m,l) = \min\{\min_{i \in \mathcal{I}: \lambda_i \neq 1} \{ml + (1-\lambda_i)K_i\}, \min_{i \in \mathcal{I}^c} K_i\}$. Since $\mathfrak{A}$ is $(\boldsymbol{\lambda}, \mathcal{I})$-workable, (A-4) becomes

$$\begin{aligned} p - q &\geq \max \left\{ c_1 \sqrt{p(1-q)} \max \left\{ \frac{\sqrt{\sum_{i \in \mathcal{I}^c} K_i \log n}}{\min_{i \in \mathcal{I}^c} K_i}, \sqrt{\frac{\log n}{\min_{i \in \mathcal{I}^c} K_i}} \right\}, c_2 \sqrt{\frac{m \log \frac{n}{K}}{\min_{i \in \mathcal{I}} K_i}} \right\} \\ &= \max \left\{ c_1 \sqrt{p(1-q)} \max \left\{ \frac{\sqrt{\sum_{i \in \mathcal{I}^c} K_i \log n}}{K}, \sqrt{\frac{\log n}{K}} \right\}, c_2 \sqrt{\frac{m \log \frac{n}{K}}{\min_{i \in \mathcal{I}} K_i}} \right\} \end{aligned}$$

which implies that

$$K \geq \max \left\{ c_1 \frac{\sqrt{p(1-q) \sum_{i \in \mathcal{I}^c} K_i \log n}}{p-q}, c_1^2 \frac{p(1-q) \log n}{(p-q)^2} \right\}, \text{ and } m \leq \frac{(p-q)^2 \min_{i \in \mathcal{I}} K_i}{c_2^2 \log \frac{n}{K}}.$$

Besides, m should also satisfy $c_3 \log n \leq m \leq \frac{1}{4} \sqrt{\frac{K}{\log n}}$. Hence, by combining these inequalities together, we obtain this corollary.

7. Proof of Theorem 4

It requires $O(f(\frac{n}{m})m)$ computation and $O(g(\frac{n}{m})m)$ memory for $\mathfrak{A}$ recovering the clusters in the subgraphs. From the proof of Theorem 3, we know that the size of the fused graph is $O(mrl + n - \sum_{i \in \mathcal{I}} \lambda_i K_i)$. Thus, recovering clusters in the fused graph by solving (1) needs $O((mrl + n - \sum_{i\in\mathcal{I}} \lambda_i K_i)^3)$ computation and $O((mrl + n - \sum_{i\in\mathcal{I}} \lambda_i K_i)^2)$ memory. Hence we obtain this theorem.

8. Useful Lemmas

The following two lemmas are derived from the matrix Bernstein inequality (Tropp, 2012).

Theorem A-1. *(Matrix Bernstein, (Tropp, 2012)) Let $\mathbf{X}_1, \cdots, \mathbf{X}_n$ be independent random matrices with common dimension $d_1 \times d_2$. Assume that each matrix has bounded deviation from its mean:*

$$\|\mathbf{X}_k - \mathbb{E}\mathbf{X}_k\| \leq R \text{ for each } k = 1, \cdots, n.$$

Form the sum $\mathbf{Z} = \sum_{k=1}^{n} \mathbf{X}_k$, and introduce a variance parameter

$$\sigma^2 = \max\{\|\mathbb{E}[(\mathbf{Z} - \mathbb{E}\mathbf{Z})(\mathbf{Z} - \mathbb{E}\mathbf{Z})^\top]\|, \|\mathbb{E}[(\mathbf{Z} - \mathbb{E}\mathbf{Z})^\top(\mathbf{Z} - \mathbb{E}\mathbf{Z})]\|\},$$

then

$$\mathbb{P}[\|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| \geq t] \leq (d_1 + d_2) \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right).$$

Lemma A-5. *Suppose $\mathbf{W}$ is a $n \times n$ random matrix whose entries are independent random variables satisfying that $\mathbb{E}[\mathbf{W}] = 0$, $\|\mathbf{W}\|_\infty \leq b$, Var$[W_{ij}] \leq \sigma_0^2$ for $(i,j) \in \mathcal{C}$ and Var$[W_{ij}] \leq \sigma_1^2$ for $(i,j) \in \mathcal{C}^c$, then the following inequality holds with probability at least $1 - n^{-10}$:*

$$\|\mathbf{W}\| \leq c\left(b \log n + \sqrt{(n\sigma_0^2 + (n-s)\sigma_1^2)\log n}\right)$$

where c is a universal constant.

Proof. Let $\mathbf{e}_i$ be the ith standard basis vector, then

$$\mathbf{W} - \mathbb{E}\mathbf{W} = \sum_{i,j} W_{ij}\mathbf{e}_i\mathbf{e}_j^\top \triangleq \sum_{i,j} \mathbf{X}_{ij}.$$

Thus, $\|\mathbf{X}_{ij}\| = |W_{ij}| \leq b$ for all (i,j). Since the entries of $\mathbf{W}$ are independent,

$$\begin{aligned} &\|\mathbb{E}[(\mathbf{W} - \mathbb{E}\mathbf{W})(\mathbf{W} - \mathbb{E}\mathbf{W})^\top]\| = \|\mathbb{E}\sum_{i,j} W_{ij}^2\mathbf{e}_i\mathbf{e}_j^\top\mathbf{e}_j\mathbf{e}_i^\top\| \\ \leq &\|\mathbb{E}\sum_{(i,j)\in\mathcal{C}^c} W_{ij}^2\mathbf{e}_i\mathbf{e}_j^\top\mathbf{e}_j\mathbf{e}_i^\top\| + \|\mathbb{E}\sum_{(i,j)\in\mathcal{C}} W_{ij}^2\mathbf{e}_i\mathbf{e}_j^\top\mathbf{e}_j\mathbf{e}_i^\top\| \leq (n-s)\sigma_1^2 + n\sigma_0^2, \end{aligned}$$

where the last inequality follows from the definition of the high confidence nodes and the fact that the number of the high confidence nodes is s. Then from the matrix Bernstein inequality, there exists a universal constant c such that

$$\|\mathbf{W} - \mathbb{E}\mathbf{W}\| \leq c\left(b \log n + \sqrt{(n\sigma_0^2 + (n-s)\sigma_1^2)\log n}\right)$$

holds with probability at least $1 - n^{-10}$. □

Lemma A-6. *Suppose $\mathbf{W}$ is a $n \times n$ random matrix whose entries are independent random variables satisfying that 1) $\mathbb{E}[\mathbf{W}] = 0$; 2) $\max_{ij} |W_{ij}| \leq b_0$ and Var$[W_{ij}] \leq \sigma_0^2$ for $(i,j) \in \mathcal{C}$; 3) $\max_{ij} |W_{ij}| \leq b_1$ and Var$[W_{ij}] \leq \sigma_1^2$ for*

$(i,j) \in \mathcal{C}^c$, then the following inequality holds with probability at least $1 - n^{-10}$:

$$|(\mathbf{U}\mathbf{U}^\top \mathbf{W})_{ij}| \leq \begin{cases} \frac{\sqrt{n\sigma_0^2 \log n}}{K} + \frac{b_0 \log n}{K}, & \text{all the nodes in } R(i) \text{ are high confident} \\ \frac{\sqrt{(s\sigma_0^2 + (n-s)\sigma_1^2)\log n}}{K^*} + \frac{\max\{b_0, b_1\}\log n}{K^*}, & \text{otherwise} \end{cases}$$

where c is a universal constant and $R(i)$ is the cluster that node i belongs to.

Proof. Suppose cluster $R(i)$ contains $K(i)$ nodes, then

$$(\mathbf{U}\mathbf{U}^\top \mathbf{W})_{ij} = \frac{1}{K(i)} \sum_{j':(i,j') \in R(i)} \mathbf{W}_{ij'}.$$

If all the nodes in cluster $R(i)$ are high confident, then

$$\sum_{j':(i,j') \in R(i)} \mathbb{E}[\mathbf{W}_{ij'}^2] = K(i)\sigma_0^2 \leq n\sigma_0^2.$$

Otherwise, suppose that cluster $R(i)$ contains $c(i)$ high confidence nodes, then

$$\sum_{j':(i,j') \in R(i)} \mathbb{E}[\mathbf{W}_{ij'}^2] = (K(i) - c(i))\sigma_1^2 + c(i)\sigma_0^2 \leq s\sigma_0^2 + (n-s)\sigma_1^2.$$

By the standard Bernstein inequality, we can obtain this theorem. □

References

Ailon, N., Chen, Y., and Xu, H. Breaking the small cluster barrier of graph clustering. In *International Conference on Machine Learning*, ICML '13, 2013.

Tropp, J. A. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4): 389–434, 2012.