Breaking the Small Cluster Barrier of Graph Clustering Supplementary Material

Nir Ailon Yudong Chen Huan Xu

Abstract

In this supplementary material, we present proof details.

1 Notation and Conventions

We use the following notation and conventions throughout the supplement. For a real $n \times n$ matrix M, we use the unadorned norm $\|M\|$ to denote its spectral norm. The notation $\|M\|_F$ refers to the Frobenius norm, $\|M\|_1$ is $\sum_{i,j} |M(i,j)|$ and $\|M\|_\infty$ is $\max_{ij} |M(i,j)|$.

We will also study operators on the space of matrices. To distinguish them from the matrices studied in this work, we will simply call these objects "operators", and will denote them using a calligraphic font, e.g. $\mathcal{P}$. The norm $\|\mathcal{P}\|$ of an operator is defined as

$$\|\mathcal{P}\| = \sup_{M:\|M\|_F=1} \|\mathcal{P}M\|_F \ ,$$

where the supremum is over matrices M.

For a fixed, real $n \times n$ matrix M, we define the matrix linear subspace $T(M)$ as follows:

$$T(M) := \{YM + MX : X, Y \in \mathbb{R}^{n\times n}\} \ .$$

In words, this subspace is the set of matrices spanned by matrices each row of which is in the row space of M, and matrices each column of which is in the column space of M.

For any given subspace of matrices $S \subseteq \mathbb{R}^{n\times n}$, we let $\mathcal{P}_S$ denote the orthogonal projection onto S with respect to the the inner product $\langle X, Y\rangle = \sum_{i,j=1}^n X(i,j)Y(i,j) = \operatorname{tr} X^t Y$. This means that for any matrix M,

$$\mathcal{P}_S M = \operatorname{argmin}_{X\in S} \|M - X\|_F \ .$$

For a matrix M, we let $\Gamma(M)$ denote the set of matrices supported on a subset of the support of M. Note that for any matrix X,

$$(\mathcal{P}_{\Gamma(X)}M)(i,j) = \begin{cases} M(i,j) & X(i,j) \neq 0 \\ 0 & \text{otherwise} \end{cases} .$$

It is a well known fact that $\mathcal{P}_{T(X)}$ is given as follows:

$$\mathcal{P}_{T(X)}M = P_{C(X)}M + MP_{R(X)} - P_{C(X)}MP_{R(X)} \ ,$$

where $P_{C(X)}$ is projection (of a vector) onto the column space of X, and $P_{R(X)}$ is projection onto the row space of X.

For a subspace $S \subseteq \mathbb{R}^{n\times n}$ we let $S^\perp$ denote the orthogonal subspace with respect to $\langle\cdot,\cdot\rangle$:

$$S^\perp = \{X \in \mathbb{R}^{n\times n} : \langle X, Y\rangle = 0 \ \forall Y \in S\} \ .$$

Slightly abusing notation, we will use the set complement operator $(\cdot)^c$ to formally define $\Gamma(M)^c$ to be $\Gamma(M)^\perp$ (by this we are stressing that the space $\Gamma(M)^\perp$ is given as $\Gamma(M')$ where M' is any matrix such that M and M' have complementary supports). Note that $\mathcal{P}_{T(X)^\perp} M = M - \mathcal{P}_{T(X)} M = (I - P_{C(X)}) M (I - P_{R(X)})$.

For a matrix M, $\operatorname{sgn} M$ is defined as the matrix satisfying:

$$(\operatorname{sgn} M)(i,j) = \begin{cases} 1 & M(i,j) > 0 \\ -1 & M(i,j) < 0 \\ 0 & \text{otherwise} \end{cases} \ .$$

2 Proof of Theorem 1

The proof is based on [1]. We prove it for $\kappa = 1$. The adjustment for $\kappa > 1$ is done using a padding argument, presented at the end of the proof.

Additional notation:

1. We let $V_\flat \subseteq V$ denote the set of of elements i such that $n_{\langle i\rangle} \le \ell_\flat$. (We remind the reader that $n_{\langle i\rangle} = |V_{\langle i\rangle}|$.)
2. We remind the reader that the projection $\mathcal{P}_\sharp$ is defined as follows:
$$(\mathcal{P}_\sharp M)(i,j) = \begin{cases} M(i,j) & \max\{n_{\langle i\rangle}, n_{\langle j\rangle}\} \ge \ell_\sharp \\ 0 & \text{otherwise} \end{cases} \ .$$
3. The projection $\mathcal{P}_\flat$ is defined as follows:
$$(\mathcal{P}_\flat M)(i,j) = \begin{cases} M(i,j) & \max\{n_{\langle i\rangle}, n_{\langle j\rangle}\} \le \ell_\flat \\ 0 & \text{otherwise} \end{cases} \ .$$
In words, $\mathcal{P}_\flat$ projects onto the set of matrices supported on $V_\flat \times V_\flat$. Note that by the theorem assumption, $\mathcal{P}_\sharp + \mathcal{P}_\flat = \mathcal{I}d$ (equivalently, $\mathcal{P}_\sharp$ projects onto the set of matrices supported on $(V \times V) \setminus (V_\flat \times V_\flat)$).
4. Define the set
$$\mathfrak{D} = \left\{\Delta \in \mathbb{R}^{n\times n} | \Delta_{ij} \le 0, \forall i \sim j, (i,j) \notin V_\flat \times V_\flat; 0 \le \Delta_{ij}, \forall i \not\sim j, (i,j) \notin V_\flat \times V_\flat\right\},$$
which contains all feasible deviation from $\hat{K}$.
5. For simplicity we write $T := T(\hat{K})$ and $\Gamma := \Gamma(\hat{B}), \Gamma^c := \Gamma(\hat{B})^c = \Gamma^\perp$.

We will make use of the following:

1. $\operatorname{sgn}(\hat{B}) = \hat{B}$.
2. $\mathcal{I}d = \mathcal{P}_\Gamma + \mathcal{P}_{\Gamma^c} = \mathcal{P}_{\Gamma(A)} + \mathcal{P}_{\Gamma(A)^c}$.
3. $\mathcal{P}_\sharp, \mathcal{P}_\flat, \mathcal{P}_\Gamma, \mathcal{P}_{\Gamma^c}, \mathcal{P}_{\Gamma(A)}$, and $\mathcal{P}_{\Gamma(A)^c}$ commute with each other.

2.1 Approximate Dual Certificate Condition

Proposition 1. *$(\hat{K},\ \hat{B})$ is the unique optimal solution to (CP) if there exists a matrix $Q \in \mathbb{R}^{n\times n}$ and a positive number ϵ satisfying:*

1. $\|Q\| < 1$
2. $\|\mathcal{P}_T(Q)\|_\infty \leq \frac{\epsilon}{2}\min\{c_1, c_2\}$
3. $\forall\Delta \in \mathfrak{D}$:
   (a) $\left\langle UU^\top + Q, \mathcal{P}_{\Gamma(A)}\mathcal{P}_\Gamma\mathcal{P}_\sharp\Delta\right\rangle = (1+\epsilon)c_1\left\|\mathcal{P}_{\Gamma(A)}P_\Gamma\mathcal{P}_\sharp\Delta\right\|_1$
   (b) $\left\langle UU^\top + Q, \mathcal{P}_{\Gamma(A)^c}\mathcal{P}_\Gamma\mathcal{P}_\sharp\Delta\right\rangle = (1+\epsilon)c_2\left\|\mathcal{P}_{\Gamma(A)^c}P_\Gamma\mathcal{P}_\sharp\Delta\right\|_1$
4. $\forall\Delta \in \mathfrak{D}$:
   (a) $\left\langle UU^\top + Q, \mathcal{P}_{\Gamma(A)}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\rangle \geq -(1-\epsilon)c_1\left\|\mathcal{P}_{\Gamma(A)}P_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\|_1$
   (b) $\left\langle UU^\top + Q, \mathcal{P}_{\Gamma(A)^c}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\rangle \geq -(1-\epsilon)c_2\left\|\mathcal{P}_{\Gamma(A)^c}P_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\|_1$
5. $\mathcal{P}_\Gamma\mathcal{P}_\flat(UU^\top + Q) = c_1\mathcal{P}_\flat\hat{B}$
6. $\left\|\mathcal{P}_{\Gamma^c}\mathcal{P}_\flat(UU^\top + Q)\right\|_\infty \leq c_2$

Proof. Consider any feasible solution to (CP1) $(\hat{K}+\Delta, \hat{B}-\Delta)$; we know $\Delta \in \mathfrak{D}$ due to the inequality constraints in (CP1). We will show that this solution will have strictly higher objective value than $\left\langle\hat{K},\hat{B}\right\rangle$ if $\Delta \neq 0$.

For this Δ, let G_Δ be a matrix in $T^\perp \cap \text{Range}(\mathcal{P}_\flat)$ satisfying $\|G\| = 1$ and $\langle G_\Delta, \Delta\rangle = \|\mathcal{P}_{T^\perp}\mathcal{P}_\flat\Delta\|_*$; such a matrix always exists because $\text{Range}\mathcal{P}_\flat \subseteq T^\perp$. Suppose $\|Q\| = b$. Clearly, $\mathcal{P}_{T^\perp}Q + (1-b)G \in T^\perp$ and, due to desideratum 1, we have $\|\mathcal{P}_{T^\perp}Q + (1-b)G_\Delta\| \leq \|Q\| + (1-b)\|G_\Delta\| = b + (1-b) = 1$. Therefore, $UU^\top + \mathcal{P}_{T^\perp}Q + (1-b)G_\Delta$ is a subgradient of $f(K) = \|K\|_*$ at $K = \hat{K}$. On the other hand, define the matrix $F_\Delta = -\mathcal{P}_{\Gamma^c}\text{sgn}(\Delta)$. We have $F_\Delta \in \Gamma^c$ and $\|F_\Delta\|_\infty \leq 1$. Therefore, $\mathcal{P}_{\Gamma(A)}(\hat{B}+F_\Delta)$ is a subgradient of $g_1(B) = \left\|\mathcal{P}_{\Gamma(A)}B\right\|_1$ at $B = \hat{B}$, and $\mathcal{P}_{\Gamma(A)^c}(\hat{B}+F_\Delta)$ is a subgradient of $g_2(B) = \left\|\mathcal{P}_{\Gamma(A)^c}B\right\|_1$ at $B = \hat{B}$. Using these three subgradients, the difference in the objective value can be bounded as follows:

$$
\begin{aligned}
& d(\Delta) \\
\triangleq\ & \left\|\hat{K}+\Delta\right\|_* + c_1\left\|\mathcal{P}_{\Gamma(A)}(\hat{B}-\Delta)\right\|_1 + c_2\left\|\mathcal{P}_{\Gamma(A)^c}(\hat{B}-\Delta)\right\|_1 - \left\|\hat{K}\right\|_* - c_1\left\|\mathcal{P}_{\Gamma(A)}\hat{B}\right\|_1 - c_2\left\|\mathcal{P}_{\Gamma(A)^c}\hat{B}\right\|_1 \\
\geq\ & \left\langle UU^\top + \mathcal{P}_{T^\perp}Q + (1-b)G_\Delta, \Delta\right\rangle + c_1\left\langle\mathcal{P}_{\Gamma(A)}(\hat{B}+F_\Delta), -\Delta\right\rangle + c_2\left\langle\mathcal{P}_{\Gamma(A)^c}(\hat{B}+F_\Delta), -\Delta\right\rangle \\
=\ & (1-b)\|\mathcal{P}_{T^\perp}\mathcal{P}_\flat\Delta\|_* + \left\langle UU^\top + \mathcal{P}_{T^\perp}Q, \Delta\right\rangle + c_1\left\langle\mathcal{P}_{\Gamma(A)}\hat{B}, -\Delta\right\rangle + c_2\left\langle\mathcal{P}_{\Gamma(A)^c}\hat{B}, -\Delta\right\rangle \\
& + c_1\left\langle\mathcal{P}_{\Gamma(A)}F_\Delta, -\Delta\right\rangle + c_2\left\langle\mathcal{P}_{\Gamma(A)^c}F_\Delta, -\Delta\right\rangle \\
=\ & (1-b)\|\mathcal{P}_{T^\perp}\mathcal{P}_\flat\Delta\|_* + \left\langle UU^\top + \mathcal{P}_{T^\perp}Q, \Delta\right\rangle + c_1\left\langle\mathcal{P}_\flat\mathcal{P}_{\Gamma(A)}\hat{B}, -\Delta\right\rangle + c_2\left\langle\mathcal{P}_\flat\mathcal{P}_{\Gamma(A)^c}\hat{B}, -\Delta\right\rangle \\
& + c_1\left\langle\mathcal{P}_\sharp\mathcal{P}_{\Gamma(A)}\hat{B}, -\Delta\right\rangle + c_2\left\langle\mathcal{P}_\sharp\mathcal{P}_{\Gamma(A)^c}\hat{B}, -\Delta\right\rangle + c_1\left\langle\mathcal{P}_{\Gamma(A)}F_\Delta, -\Delta\right\rangle + c_2\left\langle\mathcal{P}_{\Gamma(A)^c}F_\Delta, -\Delta\right\rangle.
\end{aligned}
$$

The last six terms of the last RHS satisfy:

1. $c_1\left\langle\mathcal{P}_\flat\mathcal{P}_{\Gamma(A)}\hat{B}, -\Delta\right\rangle + c_2\left\langle\mathcal{P}_\flat\mathcal{P}_{\Gamma(A)^c}\hat{B}, -\Delta\right\rangle = c_1\left\langle\mathcal{P}_\flat\hat{B}, -\Delta\right\rangle$, because $\mathcal{P}_\flat\hat{B} \in \Gamma(A)$.

2. $\left\langle \mathcal{P}_\sharp \mathcal{P}_{\Gamma(A)} \hat{B}, -\Delta \right\rangle \geq - \left\| \mathcal{P}_\sharp \mathcal{P}_{\Gamma(A)} \mathcal{P}_\Gamma \Delta \right\|_1$ and $\left\langle \mathcal{P}_\sharp \mathcal{P}_{\Gamma(A)^c} \hat{B}, \Delta \right\rangle \geq - \left\| \mathcal{P}_\sharp \mathcal{P}_{\Gamma(A)^c} \mathcal{P}_\Gamma \Delta \right\|_1$, because $\hat{B} \in \Gamma$ and $\left\| \hat{B} \right\|_\infty \leq 1$.

3. $\left\langle \mathcal{P}_{\Gamma(A)} F_\Delta, -\Delta \right\rangle = \left\| \mathcal{P}_{\Gamma(A)} \mathcal{P}_{\Gamma^c} \Delta \right\|_1$ and $\left\langle \mathcal{P}_{\Gamma(A)^c} F_\Delta, -\Delta \right\rangle = \left\| \mathcal{P}_{\Gamma(A)^c} \mathcal{P}_{\Gamma^c} \Delta \right\|_1$, due to the definition of F.

It follows that

$$\begin{aligned} d(\Delta) \quad \geq \quad & (1-b) \left\| \mathcal{P}_{T^\perp} \mathcal{P}_\flat \Delta \right\|_* + \left\langle UU^\top + \mathcal{P}_{T^\perp} Q, \Delta \right\rangle + c_1 \left\langle \mathcal{P}_\flat \hat{B}, -\Delta \right\rangle - c_1 \left\| \mathcal{P}_\sharp \mathcal{P}_{\Gamma(A)} \mathcal{P}_\Gamma \Delta \right\|_1 \\ & - c_2 \left\| \mathcal{P}_\sharp \mathcal{P}_{\Gamma(A)^c} \mathcal{P}_\Gamma \Delta \right\|_1 + c_1 \left\| \mathcal{P}_{\Gamma(A)} \mathcal{P}_{\Gamma^c} \Delta \right\|_1 + c_2 \left\| \mathcal{P}_{\Gamma(A)^c} \mathcal{P}_{\Gamma^c} \Delta \right\|_1 . \end{aligned} \tag{2.1}$$

Consider the second term in the last RHS, which equals $\left\langle UU^\top + \mathcal{P}_{T^\perp} Q, \Delta \right\rangle = \left\langle UU^\top + Q, \mathcal{P}_\sharp \Delta \right\rangle + \left\langle UU^\top + Q, \mathcal{P}_\flat \Delta \right\rangle - \left\langle \mathcal{P}_T Q, \Delta \right\rangle$. We bound these three separately.

First term:

$$\begin{aligned} & \left\langle UU^\top + Q, \mathcal{P}_\sharp \Delta \right\rangle \\ = \quad & \left\langle UU^\top + Q, \left(\mathcal{P}_{\Gamma(A)} \mathcal{P}_\Gamma \mathcal{P}_\sharp + \mathcal{P}_{\Gamma(A)^c} \mathcal{P}_\Gamma \mathcal{P}_\sharp + \mathcal{P}_{\Gamma(A)} \mathcal{P}_{\Gamma^c} \mathcal{P}_\sharp + \mathcal{P}_{\Gamma(A)^c} \mathcal{P}_{\Gamma^c} \mathcal{P}_\sharp \right) \Delta \right\rangle \\ \geq \quad & (1+\epsilon) c_1 \left\| \mathcal{P}_{\Gamma(A)} P_\Gamma \mathcal{P}_\sharp \Delta \right\|_1 + (1+\epsilon) c_2 \left\| \mathcal{P}_{\Gamma(A)^c} P_\Gamma \mathcal{P}_\sharp \Delta \right\|_1 - (1-\epsilon) c_1 \left\| \mathcal{P}_{\Gamma(A)} P_{\Gamma^c} \mathcal{P}_\sharp \Delta \right\|_1 \\ & - (1-\epsilon) c_2 \left\| \mathcal{P}_{\Gamma(A)^c} P_{\Gamma^c} \mathcal{P}_\sharp \Delta \right\|_1 \\ & \text{(Using properties 3 and 4)} \end{aligned}$$

Second term:

$$\begin{aligned} & \left\langle UU^\top + Q, \mathcal{P}_\flat \Delta \right\rangle \\ = \quad & \left\langle \mathcal{P}_\Gamma \mathcal{P}_\flat (UU^\top + Q), \Delta \right\rangle + \left\langle \mathcal{P}_{\Gamma^c} \mathcal{P}_\flat (UU^\top + Q), \Delta \right\rangle \\ \geq \quad & c_1 \left\langle \mathcal{P}_\flat \hat{B}, \Delta \right\rangle - c_2 \left\| \mathcal{P}_{\Gamma^c} \mathcal{P}_\flat \Delta \right\|_1 \ \text{(using properties 5 and 6)} \\ = \quad & c_1 \left\langle \mathcal{P}_\flat \hat{B}, \Delta \right\rangle - c_2 \left\| \mathcal{P}_{\Gamma(A)^c} \mathcal{P}_{\Gamma^c} \mathcal{P}_\flat \Delta \right\|_1 \ \text{(because } \mathcal{P}_{\Gamma(A)^c} \mathcal{P}_{\Gamma^c} \mathcal{P}_\flat = \mathcal{P}_{\Gamma^c} \mathcal{P}_\flat \text{)} \end{aligned}$$

Third term: Due to the block diagonal structure of the elements of T, we have $\mathcal{P}_T = \mathcal{P}_\sharp \mathcal{P}_T$

$$\begin{aligned} & \left\langle -\mathcal{P}_T Q, \Delta \right\rangle \\ = \quad & - \left\langle \mathcal{P}_T Q, \mathcal{P}_\sharp \Delta \right\rangle \\ \geq \quad & - \left\| \mathcal{P}_T Q \right\|_\infty \left\| \mathcal{P}_\sharp \Delta \right\|_1 \\ \geq \quad & - \frac{\epsilon}{2} \min \{c_1, c_2\} \left\| \mathcal{P}_\sharp \Delta \right\|_1 . \end{aligned}$$

Combining the above three bounds with Eq. (2.1), we obtain

$$\begin{aligned} & d(\Delta) \\ \geq \quad & (1-b) \left\| \mathcal{P}_{T^\perp} \mathcal{P}_\flat \Delta \right\|_* + \epsilon c_1 \left\| \mathcal{P}_\sharp \mathcal{P}_{\Gamma(A)} \mathcal{P}_\Gamma \Delta \right\|_1 + \epsilon c_2 \left\| \mathcal{P}_\sharp \mathcal{P}_{\Gamma(A)^c} \mathcal{P}_\Gamma \Delta \right\|_1 + \epsilon c_1 \left\| \mathcal{P}_{\Gamma(A)} \mathcal{P}_{\Gamma^c} \mathcal{P}_\sharp \Delta \right\|_1 \\ & + \epsilon c_2 \left\| \mathcal{P}_{\Gamma(A)^c} \mathcal{P}_{\Gamma^c} \mathcal{P}_\sharp \Delta \right\|_1 + c_1 \left\| \mathcal{P}_{\Gamma(A)} \mathcal{P}_{\Gamma^c} \mathcal{P}_\flat \Delta \right\|_1 - \frac{\epsilon}{2} \min \{c_1, c_2\} \left\| \mathcal{P}_\sharp \Delta \right\|_1 \\ = \quad & (1-b) \left\| \mathcal{P}_{T^\perp} \mathcal{P}_\flat \Delta \right\|_* + \epsilon c_1 \left\| \mathcal{P}_\sharp \mathcal{P}_{\Gamma(A)} \Delta \right\|_1 + \epsilon c_2 \left\| \mathcal{P}_\sharp \mathcal{P}_{\Gamma(A)^c} \Delta \right\|_1 - \frac{\epsilon}{2} \min \{c_1, c_2\} \left\| \mathcal{P}_\sharp \Delta \right\|_1 \\ & \text{(note that } \mathcal{P}_{\Gamma(A)} \mathcal{P}_{\Gamma^c} \mathcal{P}_\flat \Delta = 0 \text{)} \\ \geq \quad & (1-b) \left\| \mathcal{P}_\flat \Delta \right\|_* + \frac{\epsilon}{2} \min \{c_1, c_2\} \left\| \mathcal{P}_\sharp \Delta \right\|_1 , \end{aligned}$$

which is strictly greater than zero for $\Delta \neq 0$. □

2.2 Constructing Q.

We construct a matrix Q with the properties required by Proposition 1. Suppose we take $\epsilon = \frac{2\log^2 n}{\ell_\sharp}\sqrt{\frac{n}{t(1-t)}}$ and use the weights c_1 and c_2 given in Theorem 1. We specify $\mathcal{P}_\sharp Q$ and $\mathcal{P}_\flat Q$ separately.

$\mathcal{P}_\sharp Q$ is given by $\mathcal{P}_\sharp Q = \mathcal{P}_\sharp Q_1 + \mathcal{P}_\sharp Q_2 + \mathcal{P}_\sharp Q_3$, where for $(i,j) \notin V_\flat \times V_\flat$,

$$\mathcal{P}_\sharp Q_1(i,j) = \begin{cases} -\frac{1}{n_{c(i)}} & i \sim j, (i,j) \in \Gamma \\ \frac{1}{n_{c(i)}} \cdot \frac{1-p_{ij}}{p_{ij}} & i \sim j, (i,j) \in \Gamma^c \\ 0 & i \not\sim j \end{cases}$$

$$\mathcal{P}_\sharp Q_2(i,j) = \begin{cases} -(1+\epsilon)c_2 & i \sim j, (i,j) \in \Gamma \\ (1+\epsilon)c_2 \frac{1-p_{ij}}{p_{ij}} & i \sim j, (i,j) \in \Gamma^c \\ 0 & i \not\sim j \end{cases}$$

$$\mathcal{P}_\sharp Q_3(i,j) = \begin{cases} (1+\epsilon)c_1 & i \not\sim j, (i,j) \in \Gamma \\ -(1+\epsilon)c_1 \frac{q_{ij}}{1-q_{ij}} & i \not\sim j, (i,j) \in \Gamma^c \\ 0 & i \sim, j \end{cases}$$

Note that these matrices have zero-mean entries.

$\mathcal{P}_\flat Q$ as follows. For $(i,j) \in V_\flat \times V_\flat$,

$$\mathcal{P}_\flat Q = \begin{cases} c_1 & i \sim j, (i,j) \in \Gamma(A) \\ -c_2 & i \sim j, (i,j) \in \Gamma(A)^c \\ c_1 & i \not\sim j, (i,j) \in \Gamma(A) \\ c_2 W(i,j) & i \not\sim j, (i,j) \in \Gamma(A)^c \end{cases},$$

where

$$W(i,j) = \begin{cases} +1 & \text{with probability } \frac{t-q}{2t(1-q)} \\ -1 & \text{with remaining probability} \end{cases}.$$

2.3 Validating Q

Under the choice of t in Theorem 1, we have $\frac{1}{4}p \le t \le p$ and $\frac{1}{4}(1-q) \le 1-t \le 1-q$. Also under the second assumption of the theorem and $p - q \le p(1-q)$, we have $p(1-q) \gtrsim \frac{n\log^4 n}{\ell_\sharp^2} \ge \frac{\log^4 n}{\ell_\sharp}$. We will make use of these facts frequently in the proof.

It is easy to check that $\epsilon := \frac{2\log^2 n}{\ell_\sharp}\sqrt{\frac{n}{t(1-t)}} < \frac{1}{2}$ under the assumption of Theorem 1.

Property 1):

Note that $\|Q\| \le \|\mathcal{P}_\sharp Q_\sim\| + \|\mathcal{P}_\sharp Q_{\not\sim}\| + \|\mathcal{P}_\flat Q_\sim\| + \|\mathcal{P}_\flat Q_{\not\sim}\|$. We show that all four terms are upper-bounded by $\frac{1}{4}$.

(a) $\mathcal{P}_\flat Q_\sim$ is a block diagonal matrix with each block having size at most $\ell_\flat$. Moreover, $\mathcal{P}_\flat Q_\sim$ is the sum of a deterministic matrix $Q_{\sim,d}$ with all non-zero entries equal to $\frac{b_1}{\sqrt{n\log n}}\frac{p-t}{\sqrt{t(1-t)}}$ and a random matrix $Q_{\sim,r}$ whose entries are i.i.d., bounded almost surely by $\max\{c_1, c_2\}$ and have zero mean with variance

$\frac{b_1^2}{n\log n}\cdot\frac{p(1-p)}{t(1-t)}$. Therefore, we have $\|Q_\sim\| \le \|Q_{\sim,r}\| + \|Q_{\sim,d}\|$, where w.h.p.

$$\begin{aligned}\|\mathcal{P}_\flat Q_{\sim,d}\| &\le \ell_\flat \frac{b_1}{\sqrt{n\log n}}\frac{p-t}{\sqrt{t(1-t)}},\\ \|\mathcal{P}_\flat Q_{\sim,r}\| &\le 6\max\left\{\sqrt{\ell_\flat \log n}\frac{b_1}{\sqrt{n\log n}}\sqrt{\frac{p(1-p)}{t(1-t)}} + \max\{c_1,c_2\}\log^2 n\right\};\end{aligned}$$

here in the second inequality we use Lemma 5. We conclude that $\|\mathcal{P}_\flat Q_\sim\|$ is bounded by $\frac{1}{4}$ as long as $\ell_\flat \le \frac{\sqrt{t(1-t)n\log n}}{8b_1(p-t)}$ and $\max\{c_1,c_2\} \le \frac{1}{48\log^2 n}$, which holds under the assumption of Theorem 1.

(b) $\mathcal{P}_\flat Q_{\nsim}$ is a random matrix supported on $V_\flat \times V_\flat$, whose entries are i.i.d., zero mean, bounded almost surely by $\max\{c_1,c_2\}$, and have variance $\frac{b_1^2}{n\log n}\cdot\frac{t^2+q-2tq}{(1-t)t}$. It follows from Lemma 5 that

$$\|\mathcal{P}_\flat Q_{\nsim}\| \le 6\max\left\{\sqrt{n_\flat}\cdot\frac{b_1}{\sqrt{n\log n}}\sqrt{\frac{t^2+q-2tq}{(1-t)t}}, \max\{c_1,c_2\}\log^2 n\right\} \le \frac{1}{4}$$

because $n_\flat \le n$ and $\max\{c_1,c_2\} \le \frac{1}{48\log^2 n}$, which holds under the assumption of the theorem.

(c) Note that $\mathcal{P}_\sharp Q_\sim = \mathcal{P}_\sharp Q_1 + \mathcal{P}_\sharp Q_2$. By construction these two matrices are both block-diagonal, have i.i.d zero-mean entries which are bounded almost surely by $B_\sim := \max\left\{\frac{1}{\ell_\sharp p}, \frac{2c_2}{p}\right\}$ and have variance bounded by $\sigma_\sim^2 := \max\left\{\frac{1-p}{p\ell_\sharp^2}, \frac{4(1-p)}{p}c_2^2\right\}$. Lemma 5 gives $\|\mathcal{P}_\sharp Q_\sim\| \le 6\max\left\{\sqrt{\ell_\sharp}\cdot\sigma_\sim, B_\sim\log^2 n\right\} \le \frac{1}{4}$ under the assumption of Theorem 1.

(d) Note that $\mathcal{P}_\sharp Q_{\nsim} = \mathcal{P}_\sharp Q_3$ is a random matrix with i.i.d. zero-mean entries which are bounded almost surely by $B_{\nsim} := \frac{2c_1}{1-q}$ and have variance bounded by $\sigma_{\nsim}^2 := \frac{4q}{1-q}c_1^2$. Lemma 5 gives $\|\mathcal{P}_\sharp Q_{\nsim}\| \le 6\max\left\{\sqrt{n}\cdot\sigma_{\nsim}, B_{\nsim}\log^2 n\right\} \le \frac{1}{4}$.

Property 2):

Due to the structure of T, we have

$$\begin{aligned}\|\mathcal{P}_T Q\|_\infty &= \|\mathcal{P}_T\mathcal{P}_\sharp Q\|_\infty = \left\|UU^\top(\mathcal{P}_\sharp Q) + (\mathcal{P}_\sharp Q)UU^\top + UU^\top(\mathcal{P}_\sharp Q)UU^\top\right\|_\infty\\ &\le 3\left\|UU^\top\mathcal{P}_\sharp Q\right\|_\infty \le 3\sum_{m=1}^{3}\left\|UU^\top\mathcal{P}_\sharp Q_m\right\|_\infty.\end{aligned}$$

Now observe that $(UU^\top\mathcal{P}_\sharp Q_m)(i,j) = \sum_{l\in V_{c(i)}}\frac{1}{n_{c(i)}}\mathcal{P}_\sharp Q_m(l,j)$ is the sum of i.i.d. zero-mean random variables with bounded magnitude and variance. Using Lemma 7, we obtain that for $i \in V_\sharp$,

$$\begin{aligned}\left|(UU^\top\mathcal{P}_\sharp Q_1)(i,j)\right| &\lesssim \frac{1}{n_{c(i)}}\left(\sqrt{\frac{1-p}{p\ell_\sharp^2}}\cdot\sqrt{n_{c(i)}\log n} + \frac{\log n}{\ell_\sharp p}\right)\\ &\le \frac{1}{\ell_\sharp}\sqrt{\frac{\log n}{p\ell_\sharp}} \le \frac{\log n}{24^2\ell_\sharp}\sqrt{\frac{t}{p}}.\end{aligned}$$

where in the last inequality we use $t \ge \frac{p}{4} \gtrsim \frac{\log n}{\ell_\sharp}$. For $i \in V_\flat$, clearly $(UU^\top\mathcal{P}_\sharp Q_1)(i,j) = 0$. By union bound we conclude that $\left\|UU^\top\mathcal{P}_\sharp Q_1\right\|_\infty \le \frac{\log n}{24^2\ell_\sharp}\sqrt{\frac{t}{p}}$. Similarly, we can bound $\left\|UU^\top\mathcal{P}_\sharp Q_2\right\|_\infty$ and $\left\|UU^\top\mathcal{P}_\sharp Q_3\right\|_\infty$ with the same quantity (cf. [1]).

On the other hand, under the definition of c_1, c_2 and ϵ, we have

$$c_1\epsilon = b_1\sqrt{\frac{1-t}{tn\log n}}\cdot\frac{2\log^2 n}{\ell_\sharp}\sqrt{\frac{n}{t(1-t)}} = b_1\frac{\sqrt{p\log n}}{t\sqrt{t}}\cdot\frac{\log n}{24\ell_\sharp}\sqrt{\frac{t}{p}} \geq \frac{\log n}{24\ell_\sharp}\sqrt{\frac{t}{p}}$$

and similarly $c_2\epsilon \geq \frac{\log n}{24\ell_\sharp}\sqrt{\frac{t}{p}}$. It follows that $\left\|\mathcal{P}_T Q\right\|_\infty \leq 9\cdot\frac{1}{24}\epsilon\min\{c_1,c_2\}$, proving property 2).

Properties 3a) and 3b)

For 3a), by construction of Q we have

$$\begin{aligned}
\left\langle UU^\top + Q, \mathcal{P}_{\Gamma(A)}\mathcal{P}_\Gamma\mathcal{P}_\sharp\Delta\right\rangle &= \left\langle \mathcal{P}_{\Gamma(A)}\mathcal{P}_\Gamma\mathcal{P}_\sharp Q_3, \mathcal{P}_{\Gamma(A)}\mathcal{P}_\Gamma\mathcal{P}_\sharp\Delta\right\rangle \\
&= (1+\epsilon)c_1\sum_{(i,j)\in\Gamma\cap\Gamma(A)}\mathcal{P}_\sharp\Delta(i,j) \\
&= (1+\epsilon)c_1\left\|\mathcal{P}_{\Gamma(A)}\mathcal{P}_\Gamma\mathcal{P}_\sharp\Delta\right\|_1 \text{ (because } \Delta\in\mathfrak{D})
\end{aligned}$$

Property 3b) can be verified similarly.

Properties 4a) and 4b):

For 4a), we have

$$\begin{aligned}
\left\langle UU^\top + Q, \mathcal{P}_{\Gamma(A)}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\rangle &= \left\langle \mathcal{P}_{\Gamma(A)}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\left(UU^\top + \mathcal{P}_\sharp Q_1 + \mathcal{P}_\sharp Q_2\right), \mathcal{P}_{\Gamma(A)}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\rangle \\
&= \sum_{(i,j)\in\Gamma^c\cap\Gamma(A)}\left(\frac{1}{n_{c(i)}} + \frac{1}{n_{c(i)}}\frac{1-p_{ij}}{p_{ij}} + (1+\epsilon)c_2\frac{1-p_{ij}}{p_{ij}}\right)\mathcal{P}_\sharp\Delta(i,j) \\
&\geq -\left(\frac{1}{p\ell_\sharp} + (1+\epsilon)c_2\frac{1-p}{p}\right)\left\|\mathcal{P}_{\Gamma(A)}P_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\|_1, \\
&\qquad \text{(here we use } \Delta\in\mathfrak{D},\ p_{ij}\geq p, \text{ and } n_{c(i)}\geq\ell_\sharp \text{for } i\in V_\sharp).
\end{aligned}$$

Consider the two terms in the parenthesis in the last RHS. For the first term, we have

$$\frac{1}{p\ell_\sharp} = \frac{2\log^2 n}{\ell_\sharp}\sqrt{\frac{n}{t(1-t)}}\cdot\sqrt{\frac{t(1-t)}{4p^2 n\log^4 n}} \leq \frac{2\log^2 n}{\ell_\sharp}\sqrt{\frac{n}{t(1-t)}}\cdot b_1\sqrt{\frac{1-t}{tn\log n}} = \epsilon c_1.$$

For the second term, we have the following

$$\begin{aligned}
p - t \geq \frac{p-q}{4} &\geq \frac{b_3}{4}\frac{\log^2 n\sqrt{p(1-q)n}}{\ell_\sharp} \\
&= \frac{b_3}{4}\cdot\frac{\sqrt{t(1-q)}}{\sqrt{p(1-t)}}\cdot p(1-t)\cdot\frac{2\log^2 n\sqrt{n}}{\ell_\sharp\sqrt{t(1-t)}} \\
&\geq 8\cdot p(1-t)\cdot\frac{2\log^2 n\sqrt{n}}{\ell_\sharp\sqrt{t(1-t)}} = 8p(1-t)\epsilon,
\end{aligned}$$

which implies $(1+\epsilon)c_2\frac{1-p}{p} \leq (1-2\epsilon)c_1$. We conclude that

$$\left\langle UU^\top + Q, \mathcal{P}_{\Gamma(A)}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\rangle \geq -\left(\epsilon c_1 + (1-2\epsilon)c_1\right)\left\|\mathcal{P}_{\Gamma(A)}P_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\|_1,$$

proving property 4a).

For 4b), we have

$$\begin{aligned}
\left\langle UU^\top + Q, \mathcal{P}_{\Gamma(A)^c}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\rangle &= \left\langle \mathcal{P}_{\Gamma(A)^c}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp Q_3, \mathcal{P}_{\Gamma(A)^c}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\rangle \\
&= \sum_{(i,j)\in\Gamma(A)^c\cap\Gamma^c\cap\text{Range}\,\mathcal{P}_\sharp} -(1+\epsilon)\frac{c_1 q_{ij}}{1-q_{ij}}\mathcal{P}_\sharp\Delta(i,j) \\
&\geq -(1+\epsilon)\frac{c_1 q}{1-q}\left\|\mathcal{P}_{\Gamma(A)^c}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\|_1. \text{ (here we use } q_{ij}\leq q)
\end{aligned}$$

Consider the factor before the norm in the last RHS. Similarly as before, we have

$$\begin{aligned} t-q \geq \frac{p-q}{4} &\geq \frac{b_3}{4}\frac{\log^2 n\sqrt{p(1-q)n}}{\ell_\sharp} \\ &\geq 2\cdot t(1-q)\cdot\frac{2\log^2 n\sqrt{n}}{\ell_\sharp\sqrt{t(1-t)}} = 2t(1-q)\epsilon. \end{aligned}$$

This implies $(1+\epsilon)c_1\frac{q}{1-q} \leq (1-\epsilon)c_2$. We conclude that

$$\left\langle UU^\top + Q, \mathcal{P}_{\Gamma(A)^c}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\rangle \geq -(1-\epsilon)c_2\left\|\mathcal{P}_{\Gamma(A)^c}\mathcal{P}_{\Gamma^c}\mathcal{P}_\sharp\Delta\right\|_1,$$

proving property 4b).

Properties 5) and 6): It is obvious that these two properties hold by construction of Q.

Note that properties 3)-6) hold deterministically.

2.4 The $\kappa > 1$ case

Let $n' = \kappa^2 n$ and assume n' is an integer. Let $A' \in \mathbb{R}^{n'\times n'}$ be such a matrix that

$$A' = \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}.$$

Consider the following padded program

$$\begin{aligned} \text{(CP1')} \quad \min_{K',B'\in\mathbb{R}^{n'\times n'}} \quad & \|K'\|_* + c_1\left\|\mathcal{P}_{\Gamma(A')}B'\right\|_1 + c_2\left\|\mathcal{P}_{\Gamma(A')^c}B'\right\|_1 \\ \text{s.t.} \quad & K' + B' = A' \\ & 0 \leq K'_{ij} \leq 1, \forall(i,j). \end{aligned}$$

Applying Theorem 1 with $\kappa = 1$ (which we have proved) to A' and the padded program (CP1'), we conclude that the unique optimal solution $(\hat{K}', \hat{B}' = A' - \hat{K}')$ to (CP1') has the form

$$\hat{K}' = \begin{bmatrix} \mathcal{P}_\sharp K^* & 0 \\ 0 & 0 \end{bmatrix}.$$

We claim that $\hat{K} = \mathcal{P}_\sharp K^*$ is the unique optimal solution to (CP1).

Proof by contradiction: suppose an optimal solution to (CP1) is $\hat{K} = K_0 \neq \mathcal{P}_\sharp K^*$. By optimality we have

$$\|K_0\|_* + c_1\left\|\mathcal{P}_{\Gamma(A)}(A-K_0)\right\|_1 + c_2\left\|\mathcal{P}_{\Gamma(A)^c}(A-K_0)\right\|_1 \leq \|\mathcal{P}_\sharp K^*\|_* + c_1\left\|\mathcal{P}_{\Gamma(A)}(A-\mathcal{P}_\sharp K^*)\right\|_1 + c_2\left\|\mathcal{P}_{\Gamma(A)^c}(A-\mathcal{P}_\sharp K^*\right\|_1.$$

Define $K_0' = \begin{bmatrix} K_0 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n'\times n'}$. It follows that

$$\begin{aligned} & \|K_0'\|_* + c_1\left\|\mathcal{P}_{\Gamma(A')}(A'-K_0')\right\|_1 + c_2\left\|\mathcal{P}_{\Gamma(A')^c}(A'-K_0')\right\|_1 \\ = \; & \|K_0\|_* + c_1\left\|\mathcal{P}_{\Gamma(A)}(A-K_0)\right\|_1 + c_1(n'-n) + c_2\left\|\mathcal{P}_{\Gamma(A)^c}(A-K_0)\right\|_1 \\ \leq \; & \|\mathcal{P}_\sharp K^*\|_* + c_1\left\|\mathcal{P}_{\Gamma(A)}(A-\mathcal{P}_\sharp K^*)\right\|_1 + c_1(n'-n) + c_2\left\|\mathcal{P}_{\Gamma(A)^c}(A-\mathcal{P}_\sharp K^*)\right\|_1 \\ = \; & \left\|\hat{K}'\right\|_* + c_1\left\|\mathcal{P}_{\Gamma(A')}(A'-\hat{K}')\right\|_1 + c_2\left\|\mathcal{P}_{\Gamma(A')^c}(A'-\hat{K}')\right\|_1, \end{aligned}$$

contradicting the fact that $(\hat{K}', \hat{B}' = A' - \hat{K}')$ is the unique optimal to (CP1').

3 Proof of Theorem 3

Fix $\kappa \geq 1$ and t in the allowed range, let (K, B) be an optimal solution to (CP1) , and assume K is a partial clustering induced by $U_1, \ldots, U_r$ for some integer r, and also assume $\sigma_{\min}(K) = \min_{i\in[r]} |U_i|$ satisfies (3). Let $M = \sigma_{\min}(K)$.

We need a few helpful facts. First, note that any value of t in the allowed range $[\frac{1}{4}p + \frac{3}{4}q, \frac{3}{4}p + \frac{1}{4}q]$ satisfies $q + \frac{1}{4}(p-q) \leq t \leq p - \frac{1}{4}(p-q)$. Also note that from the definition of t, c_1, c_2,

$$q + \frac{1}{4}(p-q) \leq \frac{c_2}{c_1+c_2} = t \leq p - \frac{1}{4}(p-q) \ . \tag{3.1}$$

We say that a pair of sets $Y \subseteq V, Z \subseteq V$ is *cluster separated* if there is no pair $(y, z) \in Y \times Z$ satisfying $y \sim z$.

Assumption 2. *There exists a constant $C' > 0$ such that for all pairs of cluster-separated sets Y, Z of size at least $m := \frac{C' \log n}{(p-q)^2}$ each,*

$$|\hat{d}_{Y,Z} - q| < \frac{1}{4}(p-q) \ , \tag{3.2}$$

where $\hat{d}_{Y,Z} := \frac{|(Y\times Z)\cap\Omega|}{|Y|\cdot|Z|}$.

This is proven by a Hoeffding tail bound and a union bound to hold with probability at least $1 - n^{-4}$. To see why, fix the sizes m_Y, m_Z of $|Y|, |Z|$, assume $m_Y \leq m_Z$ w.l.o.g. For each such choice, there are at most $\exp\{C(m_Y + m_Z)\log n\} \leq \exp\{2Cm_Z \log n\}$ possibilities for the choice of sets Y, Z, for some $C > 0$. For each such choice, the probability that (3.2) does not hold is

$$\exp\{-C'' m_Y m_Z (p-q)^2\} \tag{3.3}$$

using Hoeffding inequality, for some $C'' > 0$. Hence, as long as $m_Y \geq m$ as defined above, for properly chosen C', using union bound (over all possibilities of m_Y, m_Z and of Y, Z) we obtain (3.2) uniformly.

If we assume also , say, that

$$M \geq 3m \ , \tag{3.4}$$

(which can be done by setting $C_1 \geq 3C'$) the implication of the assumption is that it cannot be the case that some U_i contains a subset U_i' of size in the range $[m, |U_i| - m]$ such that $U_i' = V_g \cap U_i$ for some g. Indeed, if such a set existed, then we would find a strictly better solution to (CP1), call it (K', B'), which is defined so that K' is obtained from K by splitting the block corresponding to U_i into two blocks, one corresponding to U_i' and the other to $U_i \setminus U_i'$. The difference Δ between the cost of (K, B) and (K', B') is (renaming $Y := U_i'$ and $Z := U \setminus U_i'$) $\Delta = c_1|(Y \times Z) \cap \Omega| - c_2|(Y \times Z) \cap \Omega^c| = (c_1 + c_2)\hat{d}_{Y,Z}|Y|\,|Z| - c_2|Y|\,|Z|$. But the sign of Δ is exactly the sign of $\hat{d}_{Y,Z} - \frac{c_2}{c_1+c_2}$ which is strictly negative by (3.2) and (3.1). (We also used the fact that the trace norm part of the utility function is equal for both solutions: $\|K'\|_* = \|K\|_*$).

The conclusion is that for each i, the sets $(U_i \cap V_1), \ldots, (U_i \cap V_k)$ must all be of size at most m, except maybe for at most one set of size at least $|U_i| - m$. If we now also assume that

$$M > km = (kC' \log n)/(p-q)^2 \ , \tag{3.5}$$

then we conclude that not all these sets can be of size at most m. Hence exactly one of these sets must have size at least $|U_i| - m$. From this we conclude that there is a function $\phi : [r] \mapsto [k]$ such that for all $i \in [r]$,

$$|U_i \cap V_{\phi(i)}| \geq |U_i| - m \ .$$

We now claim that this function is an injection. We will need the following assumption:

Assumption 3. *For any 4 pairwise disjoint subsets* (Y, Y', Z, Z') *such that* $(Y \cup Y') \subseteq V_i$ *for some* i, $(Z \cup Z') \subseteq [n] \setminus V_i$, $\max\{|Z|, |Z'|\} \le m$, $\min\{|Y|, |Y'|\} \ge M - m$*:*

$$|Y| \cdot |Y'|\, \hat{d}_{Y,Y'} - |Y| \cdot |Z|\, \hat{d}_{Y,Z} - |Y'| \cdot |Z'|\, \hat{d}_{Y',Z'} > \frac{c_2}{c_1 + c_2}(|Y| \cdot |Y'| - |Y| \cdot |Z| - |Y'| \cdot |Z'|) \tag{3.6}$$

The assumption holds with probability at least $1 - n^{-4}$ by using Hoeffding inequality, union bounding over all possible sets Y, Y', Z, Z' as above. Indeed, notice that for fixed $m_Y, m_{Y'}, m_Z, m_{Z'}$ (with, say, $m_Y \ge m_{Y'}$), and for each tuple Y, Y', Z, Z' such that $|Y| = m_Y, |Y'| = m_{Y'}, |Z| = m_Z, |Z'| = m_{Z'}$, the probability that (3.6) is violated is at most

$$\exp\{-C(p - q)^2(m_Y m_{Y'} + m_Y m_Z + m_{Y'} m_{Z'})\} \tag{3.7}$$

for some $C > 0$. Using (3.4), this is at most

$$\exp\{-C''(p - q)^2(m_Y m_{Y'})\}\ , \tag{3.8}$$

for some global $C'' > 0$. Now notice that the number of possibilities to choose such a 4 tuple of sets is bounded above by $\exp\{C''' m_Y \log n\}$, for some global $C''' > 0$. Assuming

$$M \ge \frac{\hat{C} \log n}{(p - q)^2} \tag{3.9}$$

for some $\hat{C}$, and applying a union bound over all possible combinations Y, Y', Z, Z' of sizes $m_Y, m_{Y'}, m_Z, m_{Z'}$ respectively, of which there are at most $\exp\{C^\circ m_Y \log n\}$ for some $C^\circ > 0$, we conclude that (3.6) is violated for some combination with probability at most

$$\exp\{-C''(p - q)^2 m_Y m_{Y'}/2\} \tag{3.10}$$

which is at most $\exp\{-20 \log n\}$ if

$$M \ge \frac{\hat{C}' \log n}{(p - q)^2}\ . \tag{3.11}$$

for some $\hat{C}' > 0$. Apply a union bound now over the possible combinations of the tuple $(m_Y, m_{Y'}, m_Z, m_{Z'})$, of which there are at most $\exp\{4 \log n\}$ to conclude that (3.6) holds uniformly for all possibilities of Y, Y', Z, Z' with probability at least $1 - n^{-4}$.

Now assume by contradiction that ϕ is not an injection, so $\phi(i) = \phi(i') =: j$ for some distinct $i, i' \in [r]$. Set $Y = U_i \cap V_j, Y' = U_{i'} \cap V_j$, $Z = U_i \setminus Y$, $Z' = U_{i'} \setminus Y'$. Note that $\max\{|Z|, |Z'|\} \le m$ and $\min\{|Y|, |Y'|\} \ge M - m$. Consider the solution (K', B') where K' is obtained from K by replacing the two blocks corresponding to $U_i, U_{i'}$ with four blocks: Y, Y', Z, Z'. Inequality (3.6) guarantees that the cost of (K', B') is strictly lower than that of (K, B), contradicting optimality of the latter. (Note that $\|K\|_* = \|K'\|_*$.)

We can now also conclude that $r \le k$. Fix $i \in [r]$. We show that not too many elements of $V_{\phi(i)}$ can be contained in $V \setminus \{U_1 \cup \cdots \cup U_r\}$. We need the following assumption.

Assumption 4. *For all pairwise disjoint sets* $Y, X, Z \subseteq V$ *such that* $|Y| \ge M - m$, $|X| \ge m$, $(Y \cup X) \subseteq V_j$ *for some* $j \in [k]$, $|Z| \le m$, $Z \cap V_j = \emptyset$*:*

$$|X| \cdot |Y| \hat{d}_{X,Y} + \binom{|X|}{2} \hat{d}_{x,x} - |Y| \cdot |Z| \hat{d}_{Y,Z} > \frac{c_2}{c_1 + c_2}(|X| \cdot |Y| + \binom{|X|}{2} - |Y| \cdot |Z|) + \frac{|X|}{c_1 + c_2}\ . \tag{3.12}$$

The assumption holds with probability at least $1-n^{-4}$. To see why, first notice that $|X|/(c_1+c_2) \leq \frac{1}{8}(p-q)|X|\cdot|Y|$ by (3), as long as C_2 is large enough. This implies that the RHS of (3.12) is upper bounded by

$$\left(p-\frac{1}{8}(p-q)\right)|X|\cdot|Y| + \frac{c_2}{c_1+c}(\binom{|X|}{2} - |Y|\cdot|Z|) \tag{3.13}$$

Proving that the LHS of (3.12) (denoted $f(X,Y,Z)$) is larger than (3.13) (denoted $g(X,Y,Z)$) uniformly w.h.p. can now be easily done as follows. By fixing $m_Y = |Y|, m_X = |X|$, the number of combinations for Y, X, Z is at most $\exp\{C(m_Y+m_X)\log n\}$ for some global $C>0$. On the other hand, the probability that $f(X,Y,Z) \leq g(X,Y,Z)$ for any such option is at most

$$\exp\{-C''(p-q)^2 m_Y m_X\} \tag{3.14}$$

for some $C'>0$. Hence, by union bounding, the probability that some tuple Y, X, Z of sizes m_Y, m_X, m_Z respectively satisfies $f(X,Y,Z) \leq g(X,Y,Z)$ is at most

$$\exp\{-C''(p-q)^2 m_Y/2\} \ , \tag{3.15}$$

which is at most $\exp\{-10\log n\}$ assuming

$$M \geq \bar{C}(\log n)/(p-q)^2 \ , \tag{3.16}$$

for some $\bar{C}>0$. Another union bound over the possible choices of m_Y, m_X, m_Z proves that (3.12) holds uniformly with probability at least $1-n^{-4}$.

Now for some $i\in[r]$ set $X := V_{\phi(i)} \cap (V \setminus \{U_1\cup\cdots\cup U_r\})$ and assume by contradiction that $|X|>m$. Set $Y := V_{\phi(i)} \cap U_i$ and $Z = U_i \setminus V_{\phi(i)}$. Define the solution (K', B') where K' is obtained from K by replacing the block corresponding to U_i in K with two blocks: $V_{\phi(i)}$ and $U_i \setminus V_{\phi(i)}$. Assumption 4 tells us that the cost of (K', B') is strictly lower than that of (K, B). Note that the expression $\frac{|X|}{c_1+c_2}$ in the RHS of (3.12) accounts for the trace norm difference $\|K'\|_* - \|K\|_* = |X|$.

We are prepared to perform the final "cleanup" step. At this point we know that for each $i\in[r]$, the set $T_i = U_i \cap V_{\phi(i)}$ satisfies

$$\begin{aligned} |T_i| &\geq |U_i| - m \\ |T_i| &\geq |V_j| - rm \ . \end{aligned}$$

(The second inequality is implied by the fact that at most m elements of $V_{\phi(i)}$ may be contained in $U_{i'}$ for $i' \neq i$, and another at most m elements in $V \setminus (U_1\cup\cdots\cup U_r)$. We are now going to conclude from this that $U_i = V_{\phi(i)}$ for all i. To that end, let (K', B') be the feasible solution to (CP1) defined so that K' is a partial clustering induced by $V_{\phi(1)}, \ldots, V_{\phi(r)}$. We would like to argue that if $K \neq K'$ then the cost of (K', B') is strictly smaller than that of (K, B). Fix the value of the collection

$$\begin{aligned} \mathcal{Y} \ := \ & ((r, \phi(1), \ldots, \phi(r), \\ & \left(m_{ij} := |V_{\phi(i)} \cap U_j|)\right)_{i,j\in[r], i\neq j} \quad , \\ & \left(m'_i := |V_{\phi(i)} \cap (V \setminus (U_1\cup\cdots\cup U_r))\right)_{i\in[r]}) \end{aligned}$$

Let $\beta(\mathcal{Y})$ denote the number of $i\neq j$ such that $m_{ij}>0$ plus the number of $i\in[r]$ such that $m_i>0$. We can assume $\beta(\mathcal{Y})>0$, otherwise $U_i = V_{\phi(i)}$ for all $i\in[r]$ as required. The number of possibilities for K and K' giving rise to $\mathcal{Y}$ is $\exp\{C(\sum_{i\neq j} m_{ij} + \sum_i m_i)\log n\}$ for some $C>0$. (Note that K' depends on $r, \phi(1), \ldots, \phi(r)$ only, while K depends on all elements of $\mathcal{Y}$). For each such possibility, the probability that the cost of (K, B) is lower than that of (K', B') is at most

$$\exp\{-C''(p-q)^2 M(\sum_{ij} m_{ij} + \sum_i m_i)\} \tag{3.17}$$

using Hoeffding inequalities, for some $C'' > 0$. (Note that special care needs to be made to account for the difference $\|K\|_* - \|K'\|_* = \sum_{i=1}^{r} m_i$ - this is similar to what we did above .) As long as

$$M \geq \hat{C}^{\dagger} k(\log n)/(p-q)^2 \tag{3.18}$$

for some $\hat{C}^{\dagger} > 0$, we conclude that the cost of (K', B') is at least that of (K, B) for some K giving rise to $\mathcal{Y}$ with probability at most $\exp\{-10(k \log n)\beta(\mathcal{Y})\}$. The number of combinations of $\mathcal{Y}$ for a fixed value of $\beta(\mathcal{Y})$ is at most $\exp\{5(k + \beta(\mathcal{Y}) \log n\}$. By union bounding, we conclude that for fixed $\beta(\mathcal{Y})$, the probability that some (K, B) has cost at most that of (K', B') is at most $\exp\{-10(k \log n)\beta(\mathcal{Y})\}$. Finally union bound over all possibilities for $\beta(\mathcal{Y})$, of which there are at most n^2.

Taking C_1, C_2 large enough to satisfy the requirements above concludes the proof.

4 Proof of Theorem 9

The proof of Theorem 3 in the previous section made repeated use of Hoeffding tail inequalities, for bounding the size of the intersection of the noise support Ω with various submatrices. This is tight for p, q which are bounded away from 0 and 1. However, if $p = \rho p', q = \rho q'$, the noise probabilities p', q' are fixed and ρ tends to 0, a sharper bound is obtained using Bernstein tail bound (see Appendix A.2, Lemma 6). Using Bernstein inequality instead of Chernoff inequality, the expression $(p - q)^2$ in (3.3), (3.5), (3.7), (3.8), (3.9), (3.10), (3.11), (3.14), (3.15), (3.16), (3.17) can be replaced with ρ. This clearly gives the required result.

5 Proof of Lemma 5

Proof. We remind the user that $g = \frac{b_3}{b_4} \log^2 n$, the multiplicative size of the interval $\ell_\flat, \ell_\sharp$. Consider the set of intervals $(n/gk_0, n/k_0), (n/g^2k_0, n/gk_0), \ldots, (n/g^{k_0+1}k_0, n/g^{k_0}k_0)$. By the pigeonhole principle, one of these intervals must not intersect the set of cluster sizes. Assume this interval is $(n/g^{i_0+1}k_0, n/g^{i_0}k_0)$, for some $0 \leq i_0 \leq k_0$. Let $\alpha = n/g^{i+1}k_0$. By setting $C_3(p, q)$ small enough and $C_4(p, q)$ large enough, one easily checks that the requirements of Corollary 4 hold with this value of α and $s = n/k_0$. This concludes the proof. □

References

[1] Y. Chen, S. Sanghavi, and H. Xu. Clustering sparse graphs. In *NIPS. Available on arXiv:1210.3335*, 2012.

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

[3] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *Arxiv preprint arxiv:1011.3027*, 2010.

A Technical Lemmas

A.1 The spectral norm of random matrices

It is well-known that the spectral norm $\lambda_1(A)$ of a zero-mean random matrix A is bounded above w.h.p. by $C\sqrt{n}$, where C is a constant that might depend on the variance and magnitude of the entries of A. Here we state and (re-)prove an upper bound of $\lambda_1(A)$ with an explicit estimate of the constant C, which is needed in the proof of the main theorem.

Lemma 5. *Let A_{ij}, $1 \le i, j \le n$ be independent random variables, each of which has mean 0 and variance at most σ^2 and is bounded in absolute value by B. Then with probability at least $1 - 2n^{-2}$*

$$\lambda_1(A) \le 6 \max\left\{\sigma\sqrt{n \log n}, B \log^2 n\right\}$$

Proof. Let e_i be the i-th standard basis in $\mathbb{R}^n$. Let $Z_{ij} = A_{ij} e_i e_j^\top$. Then Z_{ij}'s are zero-mean random matrices independent of each other, and $A = \sum_{i,j} Z_{ij}$. We have $\|Z_{ij}\| \le B$ almost surely. We also have $\| \sum_{i,j} \mathbb{E}(Z_{ij} Z_{ij}^\top)\| = \| \sum_i e_i e_i^\top \sum_j \mathbb{E}(A_{ij}^2)\| \le n\sigma^2$. Similarly $\| \sum_{i,j} \mathbb{E}(Z_{ij}^\top Z_{ij})\| \le n\sigma^2$. Applying the Non-commutative Bernstein Inequality (Theorem 1.6 in [2]) with $t = 6 \max\left\{\sigma\sqrt{n \log n}, B \log^2 n\right\}$ yields the desired bound. □

A.2 Standard Bernstein Inequality for Sum of Independent Variables

Lemma 6 (Bernstein inequality)**.** *Let $Y_1, \ldots, Y_N$ be independent random variables, each of which has variance bounded by σ^2 and is bounded in absolute value by B a.s.. Then we have that*

$$\Pr\left[\left|\sum_{i=1}^N Y_i - \mathbb{E}\left[\sum_{i=1}^N Y_i\right]\right| > t\right] \le 2 \exp\left\{\frac{t^2/2}{N\sigma^2 + Bt/3}\right\} .$$

The following well known consequence of the above lemma will also be of use.

Lemma 7. *Let $Y_1, \ldots, Y_N$ be independent random variables, each of which has variance bounded by σ^2 and is bounded in absolute value by B a.s. Then we have*

$$\left|\sum_{i=1}^N Y_i - \mathbb{E}\left[\sum_{i=1}^N Y_i\right]\right| \le C_0 \max\left\{\sigma\sqrt{N \log n}, B \log n\right\}$$

with probability at least $1 - C_1 n^{-C_2}$ where the positive constants C_0, C_1, C_2 are independent of σ, B, N and n.