Clustering Sparse Graphs – Supplementary Material

Yudong Chen
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX 78712
ydchen@utexas.edu

Sujay Sanghavi
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX 78712
sanghavi@mail.utexas.edu

Huan Xu
Mechanical Engineering Department
National University of Singapore
Singapore 117575, Singapore
mpexuh@nus.edu.sg

In this supplementary material, we provide proofs for several results in the main paper, and discuss implementation issues.

We will use the following notation throughout this material. $M \circ N$ is the element-wise product between the matrices M and N. $\langle M, N\rangle = \text{Trace}(MN)$ is the standard matrix inner product. $\|M\|$ is the largest singular value of M and $\|M\|_\infty = \max_{i,j} |m_{ij}|$ is the matrix infinity norm.

1 Proof of Monotonicity Lemma (Lemma 1)

In this section, we prove Lemma 1 in the main paper.

Proof. Let Ω_δ denote the entries that has been changed. Notice that since A and $\tilde{A}$ are different, the respective weight matrix C and $\tilde{C}$ are also different. In particular, they differ on Ω_δ. Let Y' be an arbitrary feasible solution, then we have by the optimality of $\hat{Y}$

$$\|\hat{Y}\|_* + \|C \circ (A - \hat{Y})\|_1 \leq \|Y'\|_* + \|C \circ (A - Y')\|_1.$$

Next notice that from definition of $\tilde{A}$ we have

$$\|\hat{Y}\|_* + \|\tilde{C} \circ (\tilde{A} - \hat{Y})\|_1 = \|\hat{Y}\|_* + \|C \circ (A - \hat{Y})\|_1 - \sum_{i,j \in \Omega_\delta} C_{ij},$$

while on the other hand

$$\begin{aligned} & \|Y'\|_* + \|C \circ (A - Y')\|_1 - [\|Y'\|_* + \|\tilde{C} \circ (\tilde{A} - Y')\|_1] \\ = \quad & \sum_{\Omega_\delta} [C_{ij} |(A - Y')_{ij}| - \tilde{C}_{ij} |\tilde{A} - Y'| \\ \leq \quad & \sum_{\Omega_\delta} C_{ij}, \end{aligned}$$

where the last inequality we use $\|A - Y'\|_\infty \leq 1$, and $\|\tilde{A} - Y'\|_\infty \leq 1$. Combining all equations together establishes that

$$\|\hat{Y}\|_* + \|\tilde{C} \circ (\tilde{A} - \hat{Y})\|_1 \leq \|Y'\|_* + \|\tilde{C} \circ (\tilde{A} - Y')\|_1.$$

As Y' is arbitrary, the lemma follows. □

2 Proof of New Dual Certificate Conditions for Clustering (Proposition 1)

In this section, we prove Proposition 1 in the main paper.Recall that $P_{\Omega}(M)$ is the matrix where the $(i,j)^{th}$ entry is m_{ij} if $(i,j) \in \Omega$, and 0 else.

Proof. From the first condition we know $U_0U_0^\top + P_{T^\perp}(W)$ is a subgradient of $\|Y\|_*$ at Y^*. Consider any feasible solution $(Y^* + \Delta, S^* - \Delta)$ with $\Delta \in \mathfrak{D}$ and $\Delta \neq 0$. For this Δ, we can choose $F \in \Omega^c$, $\|F\|_\infty \leq 1$ such that $\langle C \circ F, -\Delta \rangle = \|P_{\Omega^c}(C \circ \Delta)\|_1$; in this case $C \circ (S^* + F) = C \circ (\text{sign}(S^*) + F)$ is a subgradient of $\|C \circ S\|_1$ at S^*. We have the following chain of inequalities.

$$
\begin{aligned}
& \|Y^* + \Delta\|_* - \|Y\|_* + \|C \circ (S^* - \Delta)\|_1 - \|C \circ S^*\|_1 \\
= \; & \|Y^* + \Delta\|_* - \|Y\|_* + \|C \circ (S^* - \Delta)\|_1 - \|C \circ S^*\|_1 \\
\overset{(a)}{\geq} \; & \left\langle U_0U_0^\top + P_{T^\perp}(W), \Delta \right\rangle + \langle C \circ S_0 + C \circ F, -\Delta \rangle \\
\overset{(b)}{=} \; & \left\langle U_0U_0^\top + W, \Delta \right\rangle - \langle P_T W, \Delta \rangle - \|P_\Omega (C \circ \Delta)\|_1 + \|P_{\Omega^c}(C \circ \Delta)\|_1 \\
= \; & \left\langle P_\Omega(U_0U_0^\top + W), \Delta \right\rangle + \left\langle P_{\Omega^c}(U_0U_0^\top + W), \Delta \right\rangle - \langle P_T W, \Delta \rangle - \|P_\Omega (C \circ \Delta)\|_1 + \|P_{\Omega^c}(C \circ \Delta)\|_1 \\
\overset{(c)}{\geq} \; & (1+\epsilon) \|P_\Omega (C \circ \Delta)\|_1 - (1-\epsilon) \|P_{\Omega^c}(C \circ \Delta)\|_1 - \langle P_T W, \Delta \rangle - \|P_\Omega (C \circ \Delta)\|_1 + \|P_{\Omega^c}(C \circ \Delta)\|_1 \\
\overset{(d)}{\geq} \; & \epsilon \|P_\Omega(C \circ \Delta)\|_1 + \epsilon \|P_{\Omega^c}(C \circ \Delta)\|_1 - \|P_T W\|_\infty \|\Delta\|_1 \\
\overset{(e)}{\geq} \; & \epsilon \|C \circ \Delta\|_1 - \frac{\epsilon}{2} \min\{c_{\mathcal{A}^c}, c_{\mathcal{A}}\} \|\Delta\|_1 \\
> \; & 0.
\end{aligned}
$$

Here (a) uses the definition of subgradients, (b) follows from (*) and our choice of F, (c) uses conditions 3 and 4 in the proposition, (d) uses the duality between $\|\cdot\|_\infty$ and $\|\cdot\|_1$, and (e) uses condition 2. This proves that (Y^*, S^*) is the unique optimal solution to the convex program (2). □

3 Proof of Dual Certificate (Proposition 2)

In this section we prove Proposition 2 in the main paper.

3.1 Notation and Preliminaries

For clarity of the presentation, we first review some notation used in the main paper, and introduce several new ones needed in the proof.

Let $S^* \triangleq A - Y^*$ be the true disagreement matrix. Recall that $r \triangleq$ # of cluster, $k_i \triangleq$ size of the ith cluster, and $K \triangleq \min_i k_i$. As standard, let $\Omega \triangleq \text{support}(S^*)$. In addition, we denote the singular value decomposition of Y^* (notice Y^* is symmetric) be $U_0\Sigma_0U_0^\top$, and let $P_{T^\perp}(M) \triangleq (I - U_0U_0^\top)M(I - U_0U_0^\top)$ be the projection of M onto the space of matrices whose columns and rows are orthogonal to those of Y^*, and $P_T(M) = M - P_{T^\perp}(M)$.

To exploit the special structure of the clustering setup, we introduce some new notations: $R_i \triangleq \{(l,m) : l,m \in \text{cluster } i\}$, $R \triangleq \cup_{i=1}^r R_i = \text{support}(Y^*)$. For an entry set $\Phi \subseteq [1:n] \times [1:n]$, we use $\mathbf{1}_\Phi \in \mathbb{R}^{n \times n}$ to denote the matrix which is one on entries belonging to Φ and zero elsewhere. Thus, we have $Y^* = \sum_{i=1}^r \mathbf{1}_{R_i}$ and $U_0U_0^\top = \sum_{i=1}^r \frac{1}{k_i}\mathbf{1}_{R_i}$. Also, recall that $\forall (i,j) \in \mathcal{A}^c$ we have $C_{ij} = c_{\mathcal{A}^c}$, while $\forall (i,j) \in \mathcal{A}$ we have $C_{ij} = c_{\mathcal{A}}$. Hence, we may write

$$C = c_{\mathcal{A}^c}\mathbf{1}_{\mathcal{A}^c} + c_{\mathcal{A}}\mathbf{1}_{\mathcal{A}} = c_{\mathcal{A}^c}\mathbf{1}_{R\cap\Omega + R^c\cap\Omega^c} + c_{\mathcal{A}}\mathbf{1}_{R^c\cap\Omega + R\cap\Omega^c}.$$

Notice that in the graph clustering setup, the sparse corruption matrix S^* can not be arbitrary. In particular, we have that $(S^*)_{i,j}$ =0 or -1 for $(i,j) \in R$, and $(S^*)_{i,j}$ =0 or 1 for $(i,j) \in R^c$, which implies

$$S^* = \text{sign}(S^*) = -\mathbf{1}_{\Omega\cap R} + \mathbf{1}_{\Omega\cap R^c}.$$

Let $(Y^* + \Delta, S^* - \Delta)$ be a feasible solution to the convex program (2) in the main paper. Because of the constraints (3) in the main paper, Δ must belong to the set of possible deviations $\mathfrak{D}$, defined as

$$\mathfrak{D} \triangleq \{M \in \mathbb{R}^{n\times n} | \forall (i,j) \in R: \ -1 \le m_{ij} \le 0; \quad \forall (i,j) \in R^c: \ 1 \ge m_{ij} \ge 0\}.$$

Observe that for any i, j, S^*_{ij} and Δ_{ij} either have same sign, or at least one of them is zero. Thus for any $\Delta \in \mathfrak{D}$,

$$\langle C \circ S^*, \Delta\rangle = \|P_\Omega(C \circ \Delta)\|_1 . \tag{*}$$

3.2 Proof of the Proposition

Observe that, due to the randomness of Ω, W_1 and W_2 are symmetric random matrices with independent zero-mean entries. Moreover, the magnitude and variance of the entries are bounded as in the following lemma.

Lemma 3.3. *Under the assumption of Theorem 1 in the main paper where $c_1 \ge 16$, the following holds*

1. $\epsilon \le \frac{1}{4}$.
2. *The magnitude of the entries of W_1 and W_2 is bounded by $\frac{1}{16\log^2 n}$.*
3. *The variance of the entries of W_1 and W_2 is bounded by $\frac{1}{256 n \log n}$.*

Proof of Lemma 3.3. Note that $\bar{p} - \bar{q} \ge c_1 \frac{\log^2 n \sqrt{\bar{p} n}}{K}$ implies $\bar{p} \ge c_1^2 \frac{n}{K^2}\log^4 n \ge c_1^2 \frac{\log^4 n}{n}$, which further implies $K \ge c_1 \sqrt{n}\log^2 n$ since $\bar{p} \le 1$. It follows that

$$\epsilon = \frac{2\log^2 n}{K}\sqrt{\frac{n}{\bar{p}}} \le \frac{2\log^2 n}{K}\frac{K}{c_1 \log^2 n} \le \frac{1}{4}.$$

The entries of W_1 are either $-\frac{1}{k_m}$ or $\frac{1-p}{p}\frac{1}{k_m}$. Note that $\frac{1}{k_m} \le \frac{1}{K} \le \frac{1}{c_1\sqrt{n}\log^2 n}$, and $\frac{1-p}{p}\frac{1}{k_m} \le \frac{1}{\bar{p}K} \le \frac{K}{c_1 n \log^4 n} \le \frac{1}{c_2 \log^4 n}$. So the entries of W_1 are bounded by $\frac{1}{16\log^2 n}$.

The entries of W_2 are $(1+\epsilon)c_{\mathcal{A}}$, $-(1+\epsilon)c_{\mathcal{A}^c}$, $(1+\epsilon)\frac{1-p}{p}c_{\mathcal{A}^c}$, or $-(1+\epsilon)\frac{q}{1-q}c_{\mathcal{A}}$. The magnitude of them is bounded by $\max\left\{\frac{2}{\bar{p}}c_{\mathcal{A}^c}, 2c_{\mathcal{A}}\right\} = \max\left\{\frac{1}{8}\min\left(\sqrt{\frac{1}{(1-\bar{p})n\log n}}, \sqrt{\frac{1}{\bar{p}n\log n}}\right), \frac{1}{8}\min\left(\sqrt{\frac{1-\bar{q}}{\bar{q}n\log n}}, \sqrt{\frac{1}{\log^5 n}}\right)\right\} \le \frac{1}{16\log^2 n}$.

The variance of the entries of W_1 is $(1-p)/(pk_m^2)$. Since $k_m \ge K$ and $p \ge \bar{p} \ge c_1 n \log^4 n/K^2$, the variance is upper bounded by $\frac{1}{16 n \log^4 n}$, and further upper bounded by $1/256n$ as $n \ge 4$.

The variance of the entries of W_2 are either $\frac{1-p}{p}c^2_{\mathcal{A}^c}$ or $\frac{q}{1-q}c^2_{\mathcal{A}}$. Note that $\frac{1-p}{p}c^2_{\mathcal{A}^c} \le \frac{1-\bar{p}}{\bar{p}} \cdot \frac{1}{16^2}\frac{\bar{p}}{1-\bar{p}}\frac{1}{n\log n} = \frac{1}{256 n\log n}$ and $\frac{q}{1-q}c^2_{\mathcal{A}} \le \frac{\bar{q}}{1-\bar{q}}\frac{1}{16^2}\frac{1-\bar{q}}{\bar{q}}\frac{1}{n\log n} = \frac{1}{256 n \log n}$. This completes the proof of the lemma. $\square$

We also need the following simple lemma.

Lemma 3.4. *Under the assumption of Theorem 1 in the main paper where $c_1 \ge 16$, we have*

$$\begin{aligned}(1+\epsilon)\frac{c_{\mathcal{A}^c}(1-p)}{p} &\le (1-2\epsilon)c_{\mathcal{A}},\\ (1+\epsilon)\frac{c_{\mathcal{A}}q}{1-q} &\le (1-\epsilon)c_{\mathcal{A}^c}.\end{aligned}$$

Proof of Lemma 3.4. If $\frac{\bar{q}}{1-\bar{q}} \le \frac{\log^4 n}{n}$, then $c_{\mathcal{A}} = \frac{1}{16\sqrt{n\log n}}\sqrt{\frac{n}{\log^4 n}}$. In this case, $(1+\epsilon)\frac{c_{\mathcal{A}^c}(1-p)}{p} \le 2\frac{1}{16\sqrt{n\log n}}\sqrt{\frac{1-\bar{p}}{\bar{p}}} \le \frac{1}{2}\frac{1}{16\sqrt{n\log n}}\sqrt{\frac{n}{\log^4}} \le (1-2\epsilon)c_{\mathcal{A}}$ since $\bar{p} \ge c_1^2\frac{\log^4 n}{n}$. Similarly, we have $(1+\epsilon)\frac{c_{\mathcal{A}}q}{1-q} \le 2\frac{1}{16\sqrt{n\log n}}\sqrt{\frac{\log^4 n}{n}} \le \frac{1}{2}\frac{1}{16\sqrt{n\log n}} \le (1-2\epsilon)c_{\mathcal{A}}$. since $\bar{q} \le \frac{1}{4}$.

If $\frac{\bar{p}}{1-\bar{p}} \geq 1$, then $c_{\mathcal{A}^c} = \frac{1}{16\sqrt{n\log n}}$; we also have $\epsilon \leq \frac{4\log^2 n\sqrt{n}}{K} \leq \frac{1}{8}$ since $K \geq c_1\sqrt{n}\log^2 n$. In this case, $(1+\epsilon)\frac{c_{\mathcal{A}^c}(1-p)}{p} \leq \frac{9}{8}\frac{1}{16\sqrt{n\log n}} \leq \frac{6}{8}c_{\mathcal{A}}$ since $\bar{q} \leq \frac{1}{4}$. Similarly, we have $(1+\epsilon)\frac{c_{\mathcal{A}}q}{1-q} \leq \frac{9}{8}\frac{1}{16\sqrt{n\log n}}\sqrt{\frac{\bar{q}}{1-\bar{q}}} \leq \frac{6}{8}\frac{1}{16\sqrt{n\log n}} \leq (1-2\epsilon)c_{\mathcal{A}^c}$ since $\bar{q} \leq \frac{1}{4}$.

If $\frac{\bar{q}}{1-\bar{q}} \leq \frac{\log^4 n}{n}$ and $\frac{\bar{p}}{1-\bar{p}} \geq 1$, it is easy to verify that both inequalities in the lemma are implied by

$$(1+2\epsilon)\sqrt{\frac{\bar{q}}{1-\bar{q}}} \leq (1-2\epsilon)\sqrt{\frac{\bar{p}}{1-\bar{p}}}. \quad (**)$$

By assumption of Theorem 1 in the main paper, we have $\bar{p}-\bar{q} \geq c_1\frac{\sqrt{\bar{p}n}\log^2 n}{K} \geq 8\bar{p}\epsilon$ when $c_1 \geq 16$. Notice that we have $4\epsilon \geq 4\epsilon/(1+4\epsilon^2)$ by $\epsilon > 0$, and because $\bar{p} \geq \bar{q}$, we have $2\bar{p} \geq (\bar{p}+\bar{q})$. Multiplying the two inequalities, we have

$$\bar{p}-\bar{q} \geq 8\bar{p}\epsilon \geq \frac{4\epsilon}{1+4\epsilon^2}(\bar{p}+\bar{q}),$$

which implies $(1-2\epsilon)^2\bar{p} - (1+2\epsilon)^2\bar{q} \geq 0 \geq (1-2\epsilon)^2\bar{p}\bar{q} - (1+2\epsilon)^2\bar{p}\bar{q}$. Notice that $\epsilon \leq \frac{1}{4}$ by Lemma 3.3. The desired inequality (**) follows easily. $\square$

Now we are ready to proceed with the proof of the proposition, which can be divided into four steps, corresponding to checking each of the four dual certificate conditions in Proposition 1 in the main paper.

(1) Bounding $\|P_{T^\perp}W\|$.

Recall that W_1 and W_2 are random matrices with i.i.d. entries having bounded magnitude and variance. We apply standard results on the spectral norm of random matrices (Lemma A.1 in Appendix) to obtain

$$\|P_{T^\perp}(W)\| \leq \|W_1\| + \|W_2\| \leq 12\max\left\{\frac{1}{16\log^2 n}\log^2 n, \frac{1}{16\sqrt{n\log n}}\sqrt{n\log n}\right\} \leq 1.$$

(2) Bounding $\|P_T W\|_\infty$.

Recall that $P_T W_i = U_0U_0^\top W_i + W_iU_0U_0^\top - U_0U_0^\top W_iU_0U_0^\top$, we bound each of the three terms.

Since W_i is a random matrix and $U_0U_0^\top W_i = (\sum_m \frac{1}{k_m}\mathbf{1}_{R_m})W_i$, then each entry of $U_0U_0^\top W_i$ equals $\frac{1}{k_m}$ times the sum of k_m independent zero-mean random variables, whose magnitude and variance are bounded as previously discussed, for some m. Standard Bernstein inequality (e.g., Lemma B.1 in the Appendix) yields w.h.p.

$$\begin{aligned}\left\|(U_0U_0^\top W_i)\right\|_\infty &\leq \frac{1}{K}c_3\max\left\{\frac{1}{16\log^2 n}\log n, \frac{1}{16\sqrt{n}}\sqrt{K\log n}\right\}\\ &\leq c_3\max\left\{\frac{1}{16K\log^2 n}, \frac{\sqrt{\log n}}{16\sqrt{Kn}}\right\}\\ &\leq \frac{\log n}{96K},\end{aligned}$$

where the last inequality holds for n large enough.

By an almost identical argument, we have $\left\|(W_iU_0U_0^\top)\right\|_\infty \leq \frac{\log n}{96K}$. Furthermore,

$$U_0U_0^\top W_iU_0U_0^\top = (\sum_m \frac{1}{k_m}\mathbf{1}_{R_m})[W_iU_0U_0^\top],$$

implies that

$$\|U_0U_0^\top W_iU_0U_0^\top\|_\infty \leq \|W_iU_0U_0^\top\|_\infty \leq \frac{\log n}{96K}.$$

Thus, we have

$$\|P_TW\|_\infty \leq \|P_TW_1\|_\infty + \|P_TW_2\|_\infty \leq \frac{\log n}{16K}.$$

On the other hand, we have $c_{\mathcal{A}}\epsilon \geq \frac{1}{16\sqrt{n\log n}} \cdot \frac{2}{K}\sqrt{\frac{n}{\bar{p}}}\log^2 n \geq \frac{\log n}{8K}$ and $c_{\mathcal{A}^c}\epsilon = \frac{1}{16}\sqrt{\frac{\bar{p}}{1-\bar{p}}}\frac{1}{\sqrt{n\log n}} \cdot \frac{2}{K}\sqrt{\frac{n}{\bar{p}}}\log^2 n \geq \frac{\log n}{8K}$, so $\frac{1}{2}\epsilon\min\left\{c_A, c_{A^c}\right\} \geq \frac{\log n}{16K}$. We conclude that $\left\|P_T W\right\|_\infty \leq \frac{1}{2}\epsilon\min\left\{c_A, c_{A^c}\right\}$.

(3) Computing $\langle P_\Omega(U_0 U_0 + W), \Delta\rangle$.

From Equation (*) we have $\langle C \circ S^*, \Delta\rangle = \left\|C\circ\Delta\right\|_1$. Using the definition of W, we obtain

$$\left\langle P_\Omega(U_0 U_0^\top) + P_\Omega(W), P_\Omega\Delta\right\rangle = \langle (1+\epsilon) C\circ S^*, P_\Omega\Delta\rangle = (1+\epsilon)\left\|P_\Omega(C\circ\Delta)\right\|_1.$$

(4) Bounding$\langle P_{\Omega^c}(U_0 U_0 + W), \Delta\rangle$.

Observe that

$$\begin{aligned}
\left\langle P_{R\cap\Omega^c}(U_0U_0^\top + W), \Delta\right\rangle &= \left\langle \sum_{m=1}^{r}\frac{1}{k_m}1_{R_m\cap\Omega^c} + \sum_{m=1}^{r}\frac{(1-p)}{p}\frac{1}{k_m}1_{R_m\cap\Omega^c} + (1+\epsilon)\frac{c_{\mathcal{A}^c}(1-p)}{p}1_{R\cap\Omega^c}, \Delta\right\rangle \\
&\geq -\left(\frac{1}{pK} + (1+\epsilon)\frac{c_{\mathcal{A}^c}(1-p)}{p}\right)\left\|P_{R\cap\Omega^c}(\Delta)\right\|_1 \\
&\overset{(a)}{\geq} -\left(\epsilon c_{\mathcal{A}} + (1+\epsilon)\frac{c_{\mathcal{A}^c}(1-p)}{p}\right)\left\|P_{R\cap\Omega^c}(\Delta)\right\|_1 \\
&\overset{(b)}{\geq} -\left(\epsilon c_{\mathcal{A}} + (1-2\epsilon)c_{\mathcal{A}}\right)\left\|P_{R\cap\Omega^c}(\Delta)\right\|_1 \\
&\overset{(c)}{=} -(1-\epsilon)\left\|P_{R\cap\Omega^c}(C\circ\Delta)\right\|_1,
\end{aligned}$$

where in (a) we use the fact that when n large enough and $\bar{p} \geq \bar{q}$, $\bar{p} \geq c_1^2\frac{\log^4 n}{n}$, we have

$$\begin{aligned}
\epsilon c_{\mathcal{A}} &= \frac{2\log^2 n}{K}\sqrt{\frac{n}{\bar{p}}} \cdot \frac{1}{16\sqrt{n\log n}}\min\left\{\sqrt{\frac{1-\bar{q}}{\bar{q}}}, \sqrt{\frac{n}{\log^4 n}}\right\} \\
&\geq \frac{1}{pK} \cdot \frac{\sqrt{\bar{p}}\log^{3/2} n}{8}\min\left\{\sqrt{\frac{1-\bar{q}}{\bar{q}}}, \sqrt{\frac{n}{\log^4 n}}\right\} \\
&\geq \frac{1}{pK};
\end{aligned}$$

also (b) follows from Lemma 3.4, and (c) holds since $C_{ij} = c_{\mathcal{A}}$ for $(i,j) \in R\cap\Omega^c$. Similarly, we have

$$\begin{aligned}
\langle P_{R^c\cap\Omega^c}W, \Delta\rangle &= \left\langle -(1+\epsilon)\frac{c_{\mathcal{A}}q}{1-q}1_{R^c\cap\Omega^c}, \Delta\right\rangle \\
&= -(1+\epsilon)\frac{c_{\mathcal{A}}q}{1-q}\left\|P_{R^c\cap\Omega^c}(\Delta)\right\|_1 \\
&\geq -(1-\epsilon)c_{\mathcal{A}^c}\left\|P_{R^c\cap\Omega^c}(\Delta)\right\|_1 \\
&= -(1-\epsilon)\left\|P_{R^c\cap\Omega^c}(C\circ\Delta)\right\|_1
\end{aligned}$$

where we use the Lemma 3.4 in the last inequality. Combining pieces, we conclude that

$$\left\langle P_{\Omega^c}(U_0U_0^\top + W), \Delta\right\rangle \geq -(1-\epsilon)\left\|P_{\Omega^c}(C\circ\Delta)\right\|_1.$$

This completes the proof of the proposition.

4 Implementation Issues

The convex program (2) in the main paper can be solved using a general purpose SDP solver, but this method does not scale well to problems with more than 100 nodes. To facilitate fast and efficient solution, we propose to use a family of algorithms called Augmented Lagrange Multiplier (ALM) methods (see e.g. [1]). We adapt the ALM method to our problem, given as Algorithm 1. Here

Algorithm 1 ALM for Minimizing Nuclear Norm plus Weighted ℓ_1 norm

Input: $A, C \in \mathbb{R}^{n\times n}$.
Initialize: $M^{(0)} = 0$; $Y^{(0)} = 0$;$S^{(0)} = 0$; $\mu_0 > 0$; $\alpha > 1$; $k = 0$.
while not converge **do**
  $(U, \Sigma, V) = \text{svd}(A - S^{(k)} + \mu_k^{-1} M^{(k)})$.
  $\bar{Y}^{(k+1)} = U\mathcal{S}_{\mu_k^{-1}}(\Sigma)V$.
  For all (i, j), $Y_{ij}^{(k+1)} = \max\left\{\min\left\{\bar{Y}_{ij}^{(k+1)}, 1\right\}, 0\right\}$.
  $S^{(k+1)} = \mathcal{S}_{\mu_k^{-2}C}(A - Y^{(k+1)} + \mu_k^{-1} M^{(k)})$.
  $M^{(k+1)} = M^{(k)} + \mu_k(A - Y^{(k+1)} - S^{(k+1)})$.
  $k = k + 1$.
end while
Return $Y^{(k+1)}, S^{(k+1)}$.

$\mathcal{S}_{\epsilon C}(\cdot) : \mathbb{R}^{n\times n} \mapsto \mathbb{R}^{n\times n}$ is the element-wise weighted soft-thresholding operator, defined as

$$(\mathcal{S}_{\epsilon C}(X))_{ij} \quad = \quad \begin{cases} X_{ij} - \epsilon C_{ij}, & \text{if } X_{ij} > \epsilon C_{ij} \\ X_{ij} + \epsilon C_{ij}, & \text{if } X_{ij} < -\epsilon C_{ij} \\ 0, & \text{otherwise.} \end{cases}$$

In other words, it shrinks each entry of X towards zero by ϵ. The unweighted version $\mathcal{S}_\epsilon(\cdot) \triangleq \mathcal{S}_{\epsilon I}(\cdot)$ is also used. The stopping criteria and parameters of the algorithm is chosen similarly to [1].

Appendix

A 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 A.1. *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. □

B Standard Bernstein Inequalities for Sum of Independent Variables

Lemma B.1. *([3], Proposition 5.16) 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.

References

[1] Z. Lin, M. Chen, L. Wu, and Y. Ma. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. *UIUC Technical Report UILU-ENG-09-2215*, 2009.

[2] J.A. Tropp. User-friendly tail bounds for sums of random matrices. *Arxiv preprint arXiv:1004.4389*, 2010.

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