A. Proofs

A.1. Proof of Theorem 1

Let $R = \{(i,j) : Y^*_{ij} = b_1\}$. Let $\mathcal{P}_R$ be the projection operator on matrices such that $(\mathcal{P}_R Z)_{ij} = Z_{ij}$ if $(i,j) \in R$ and $(\mathcal{P}_R Z)_{ij} = 0$ otherwise.

Recall that $Y^* = USV^\top$ where $U = U_C U_B$ and $V = V_C V_B$. We use the index p_i to refer to the cluster to which node i belong. Note that

$$(UU^\top)_{ij} = \frac{\langle U_B^{p_i}, U_B^{p_j}\rangle}{\sqrt{K_{p_i} K_{p_j}}} \le \frac{u_1}{K}, \tag{2}$$

$$(VV^\top)_{ij} = \frac{\langle V_B^{q_i}, V_B^{q_j}\rangle}{\sqrt{L_{q_i} L_{q_j}}} \le \frac{u_2}{L} \tag{3}$$

and

$$(UV^\top)_{ij} = \frac{\langle U_B^{p_i}, V_B^{q_j}\rangle}{\sqrt{K_{p_i} L_{q_j}}} \le \max\left\{\frac{u_1}{K}, \frac{u_2}{L}\right\}. \tag{4}$$

Furthermore, if U_B is orthogonal, we have $(UU^\top)_{ij} = \frac{1}{K_{p_i}}$ if $p_i = p_j$ and $(UU^\top)_{ij} = 0$ otherwise. Similarly, if V_B is orthogonal, we have $(VV^\top)_{ij} = \frac{1}{L_{q_i}}$ if $q_i = q_j$ and $(VV^\top)_{ij} = 0$ otherwise.

Define the projection operators

$$\mathcal{P}_T Z := UU^\top Z + ZVV^\top - UU^\top ZVV^\top \tag{5}$$

and

$$\mathcal{P}_{T^\perp} Z := Z - \mathcal{P}_T Z.$$

For any matrix X such that $\|X\| \le \lambda$, $UV^\top + \frac{1}{\lambda}\mathcal{P}_{T^\perp} X$ is a subgradient of $\|\cdot\|_*$ at Y^*. For any feasible Y, we therefore have

$$\|Y\|_* \ge \|Y^*\|_* + \langle UV^\top + \frac{1}{\lambda}\mathcal{P}_{T^\perp} X, Y - Y^*\rangle$$

which gives

$$\langle X, Y^* - Y\rangle + \lambda(\|Y\|_* - \|Y^*\|_*) \ge \langle \mathcal{P}_T X - \lambda UV^\top, Y^* - Y\rangle. \tag{6}$$

Lemma 1. *Suppose that B is full-rank. For any $n_1 \times n_2$ matrix Z, we have*

$$(\mathcal{P}_T Z)_{ij} = \frac{1}{K_{p_i}} \sum_{i' \in C_{p_i}} Z_{i'j} + \sum_{t=1}^{n_2} \frac{\langle V_B^{q_t}, V_B^{q_j}\rangle}{\sqrt{L_{q_t} L_{q_j}}} Z_{it} - \frac{1}{K_{p_i}} \sum_{i' \in C_{p_i}} \left(\sum_{t=1}^{n_2} \frac{\langle V_B^{q_t}, V_B^{q_j}\rangle}{\sqrt{L_{q_t} L_{q_j}}} Z_{i't} \right) \tag{7}$$

if $r_1 \le r_2$ and

$$(\mathcal{P}_T Z)_{ij} = \sum_{t=1}^{n_1} \frac{\langle U_B^{p_i}, U_B^{p_t}\rangle}{\sqrt{K_{p_i} K_{p_t}}} Z_{tj} + \frac{1}{L_{q_j}} \sum_{j' \in D_{q_j}} Z_{ij'} - \frac{1}{L_{q_j}} \sum_{j' \in D_{q_j}} \left(\sum_{t=1}^{n_1} \frac{\langle U_B^{p_i}, U_B^{p_t}\rangle}{\sqrt{K_{p_i} K_{p_t}}} Z_{tj'} \right) \tag{8}$$

if $r_2 \le r_1$.

Proof. By the assumption that B is full rank, it has to be that either U_B is $r_1 \times r_1$ orthogonal (when $r_1 \le r_2$) or V_B is $r_2 \times r_2$ orthogonal (when $r_2 \le r_1$). Suppose $r_1 \le r_2$, then we obtain (7) from (2), (3) and (5). Similarly, for the case when $r_2 \le r_1$, we obtain (8). □

Lemma 2. *With probability at least $1 - n^{-\beta}$ the followings hold:*

$$\|W - \mathbb{E}W\| \le \lambda \tag{9}$$

and for all i, j

$$|(\mathcal{P}_T(W - \mathbb{E}W))_{ij}| \le \lambda \max\left\{\frac{u_1}{K}, \frac{u_2}{L}\right\} \tag{10}$$

where

$$\lambda = c_1 \left(b\beta \log n + \sqrt{\beta V n \log n} \right).$$

Proof. For (9), consider $W - \mathbb{E}W$ as the sum of independent, zero-mean random matrices:

$$W - \mathbb{E}W = \sum_{i,j} X_{i,j}$$

where

$$X_{i,j} = W_{ij} e_i e_j^\top - \mathbb{E}W_{ij} e_i e_j^\top$$

and e_i denotes the i-th vector of the standard basis. Note that

$$\|X_{i,j}\| \le 2b \quad \forall i, j$$

and

$$\max\left\{ \left\| \sum_{i,j} \mathbb{E} X_{ij} X_{ij}^\top \right\|, \left\| \sum_{i,j} \mathbb{E} X_{ij}^\top X_{ij} \right\| \right\} \le Vn$$

By applying matrix Bernstein inequality, we obtain (9).

For (10), let $Z = W - \mathbb{E}W$ be a zero-mean random matrix where we have $|Z_{ij}| \le 2b, \forall (i,j)$. Suppose $r_1 \le r_2$, then $(\mathcal{P}_T Z)_{ij}$ is given by (7). Note that $u_1 = 1$ in this case. The first summation term in (7) is the average of K_{p_i} independent random variables $Z_{i'j}$. By applying standard Bernstein inequality, we have that

$$\left| \frac{1}{K_{p_i}} \sum_{i' \in C_{p_i}} Z_{i'j} \right| \le c \frac{b\beta \log n + \sqrt{KV\beta \log n}}{K} \le \lambda \frac{u_1}{K}$$

with probability at least $1 - n^{-\beta}$. The second summation term in (7) is the sum of n_2 zero-mean random variables $\tilde{Z}_t$ with

$$\tilde{Z}_t = \frac{\langle V_B^{q_t}, V_B^{q_j} \rangle}{\sqrt{L_{q_t} L_{q_j}}} Z_{it}.$$

We have that $|\tilde{Z}_t| \le 2\frac{u_2 b}{L}$ and $\mathrm{Var}\tilde{Z}_t \le \frac{u_2^2}{L^2} V$. Again, by standard Bernstein inequality, we have that

$$\left| \sum_{t=1}^{n_2} \tilde{Z}_t \right| \le c \frac{u_2 b \beta \log n + u_2 \sqrt{\beta V n_2 \log n}}{L} \le \lambda \frac{u_2}{L}$$

with probability at least $1 - n^{-\beta}$. The last summation term of (7) can be bounded similarly by noting that the magnitude of the average is no larger than the magnitude of the individual terms.

The case of $r_2 \le r_1$ can be bounded similarly for each corresponding term of (8). The proof is completed by applying a union bound over all i, j. □

We show that with probability at least $1 - n^{-\beta}$ the following holds for all feasible $Y \neq Y^*$:

$$\langle W, Y^* - Y + \lambda(\|Y\|_* - \|Y^*\|_*) \rangle > 0$$

which implies that Y^* is the unique solution of program (1):

$$\begin{aligned}
&\langle W, Y^* - Y \rangle + \lambda(\|Y\|_* - \|Y^*\|_*) \\
&= \langle \mathbb{E}W, Y^* - Y \rangle + \langle W - \mathbb{E}W, Y^* - Y \rangle + \lambda(\|Y\|_* - \|Y^*\|_*) \\
&\ge \min\{E_1, E_0\} \|Y^* - Y\|_1 + \langle W - \mathbb{E}W, Y^* - Y \rangle + \\
&\quad \lambda(\|Y\|_* - \|Y^*\|_*) \\
&\overset{(a)}{\ge} \min\{E_1, E_0\} \|Y^* - Y\|_1 + \\
&\quad \langle \mathcal{P}_T(W - \mathbb{E}W) - \lambda U V^\top, Y^* - Y \rangle \\
&\overset{(b)}{\ge} \min\{E_1, E_0\} \|Y^* - Y\|_1 - \lambda \max\left\{ \frac{u_1}{K}, \frac{u_2}{L} \right\} \|Y^* - Y\|_1 \\
&= \left(\min\{E_1, E_0\} - \lambda \max\left\{ \frac{u_1}{K}, \frac{u_2}{L} \right\} \right) \|Y^* - Y\|_1 \\
&> 0
\end{aligned}$$

where we apply (6) and (9) in (a). In (b), we apply (10) and (4).

A.2. Proof of Theorem 2

We shall apply Theorem 1 and establish an upper-bound for $\max\{u_1/K, u_2/L\}$.

Since $r_1 \le r_2$, B is at most rank-r_1. By Theorem 1.1 of (Rudelson & Vershynin, 2009), the smallest singular value of B is at least $c_1(\sqrt{r_2} - \sqrt{r_1 - 1})$ for some universal constant $c_1, (0 < c_1 < 1)$ with probability at least $1 - \frac{1}{2} n^{-\beta}$ provided $r_1 \ge \frac{\beta \log r_1}{c}$. This implies that B is full rank with $\mathrm{rank}(B) = r_1$.

Recall that we defined $U_B S_B V_B^\top$ as an SVD of $\mathcal{K} B \mathcal{L}$. Since $\mathcal{K}$ and $\mathcal{L}$ do not change the rank of B, we have that U_B is orthogonal and therefore $u_1 = 1$. Furthermore, the smallest singular value of S_B is at least $c_1 \sqrt{KL}(\sqrt{r_2} - \sqrt{r_1 - 1})$.

We upper-bound u_2 by using the lower-bound on the singular value of $\mathcal{K} B \mathcal{L}$. First, note that the (p, q)-entry of $\mathcal{K} B \mathcal{L}$ is either $+\sqrt{K_p L_q}$ or $-\sqrt{K_p L_q}$, we therefore have

$$\begin{aligned}
n_1 L_q &= \sum_{p=1}^{r_1} K_p L_q \\
&= \|(\mathcal{K} B \mathcal{L}) e_q\|^2 \\
&= \|U_B S_B V_B^\top e_q\|^2 \\
&= \|S_B V_B^\top e_q\|^2 \\
&\ge c_1^2 KL (\sqrt{r_2} - \sqrt{r_1 - 1})^2 \|V_B^q\|^2.
\end{aligned}$$

Define γ such that $L = \gamma K$. Continuing from the above, we have

$$\begin{aligned}
\|V_B^q\|^2 &\le \frac{n_1 L_q}{c_1^2 KL (\sqrt{r_2} - \sqrt{r_1 - 1})^2} \\
&\le \frac{r_1}{c_1^2 \phi \psi (\sqrt{r_2} - \sqrt{r_1 - 1})^2} \\
&\overset{(a)}{\le} \frac{2\gamma}{c_1^2 \phi \psi \left(\sqrt{\frac{\psi n_2}{n_1}} - \sqrt{\gamma} \right)^2}
\end{aligned}$$

where in (a) we use the relationship $r_2 \ge \frac{\psi n_2}{L} = \frac{\psi n_2}{\gamma K} \ge \frac{\psi n_2}{\gamma n_1} r_1 \ge \frac{\psi n_2}{\gamma n_1}(r_1 - 1)$ and that $\frac{r_1}{r_1 - 1} \le 2$. Since the above bound for $\|V_B^q\|^2$ holds for all q, we now have

$$\begin{aligned}
\frac{u_2}{L} = \frac{u_2}{\gamma K} &\le \frac{1}{K} \min\left\{ \frac{1}{\gamma}, \frac{2}{c_1^2 \phi \psi \left(\sqrt{\frac{\psi n_2}{n_1}} - \sqrt{\gamma} \right)^2} \right\} \\
&\le \left(\frac{8 n_1}{c_1^2 \phi \psi^2 n_2} \right) \frac{1}{K}
\end{aligned}$$

where the last inequality can be obtained by considering the case of $\gamma < \frac{c_1^2 \phi \psi^2 n_2}{8 n_1}$ and $\gamma \ge \frac{c_1^2 \phi \psi^2 n_2}{8 n_1}$ respectively.

Applying Theorem 1 with failure probability $\frac{1}{2} n^{-\beta}$ and a union bound completes the proof.

A.3. Proofs of Theorem 3 and Theorem 4

We refer the reader to (Lim et al., 2014) for the proofs of the analogous results in clustering. Theorem 3 corresponds to Theorem 2 and Corollary 1 in (Lim et al., 2014), while Theorem 4 corresponds to Theorem 4 in (Lim et al., 2014). An additional element of Theorem 3 in the present paper is the introduction of u_1 and u_2. These present no difficulty by simply observing their respective range $\frac{1}{r_1} \le u_1 \le 1$ and $\frac{1}{r_2} \le u_2 \le 1$.