Probability of detecting co-clusters and setting parameters

1 Notation

- $A \in \mathbb{R}^{M\times N}$ is a matrix with K co-clusters (co-cluster set $C = \{C_k\}_{k=1}^{K}$);
- A is partitioned into $m \times n$ blocks, each block has size $\phi_i \times \psi_j$, that is, $M = \sum_{i=1}^{m} \phi_i$ and $N = \sum_{j=1}^{n} \psi_j$;
- thus block set $B = \{B_{(i,j)}\}_{i=1}^{m}{}_{,j=1}^{n}$;
- the size of co-cluster $C_k \in \mathbb{R}^{M^{(k)}\times N^{(k)}}$ that falls into block $B_{(i,j)}$ is $M_{(i,j)}^{(k)} \times N_{(i,j)}^{(k)}$;
- T_m is the minimum number of rows, T_n is the minimum number of columns.

2 Probability

Consider co-cluster C_k,

$$P(M_{(i,j)}^{(k)} = \alpha) = \frac{\binom{M^{(k)}}{\alpha}\binom{M-M^{(k)}}{\phi_i-\alpha}}{\binom{M}{\phi_i}}$$

$$P(N_{(i,j)}^{(k)} = \beta) = \frac{\binom{N^{(k)}}{\beta}\binom{N-N^{(k)}}{\psi_j-\beta}}{\binom{N}{\psi_j}}$$

The tail probability of $M_{(i,j)}^{(k)}$ and $N_{(i,j)}^{(k)}$ are

$$\begin{aligned} P(M_{(i,j)}^{(k)} < T_m) &= \sum_{\alpha=1}^{T_m-1} P(M_{(i,j)}^{(k)} = \alpha) \\ &\leq \exp(-2(s_i^{(k)})^2\phi_i) \end{aligned}$$

where $s_i^{(k)} = \dfrac{M^{(k)}}{M} - \dfrac{T_m-1}{\phi_i}$, and

$$\begin{aligned} P(N_{(i,j)}^{(k)} < T_n) &= \sum_{\beta=1}^{T_n-1} P(N_{(i,j)}^{(k)} = \beta) \\ &\leq \exp(-2(t_j^{(k)})^2\psi_j) \end{aligned}$$

where $t_j^{(k)} = \dfrac{N^{(k)}}{N} - \dfrac{T_n-1}{\psi_j}$.

The joint probability of $M_{(i,j)}^{(k)}$ and $N_{(i,j)}^{(k)}$ are

$$\begin{aligned} P(M_{(i,j)}^{(k)} < T_m, N_{(i,j)}^{(k)} < T_n) &= \sum_{\alpha=1}^{T_m-1} \sum_{\beta=1}^{T_n-1} P(M_{(i,j)}^{(k)} = \alpha) P(N_{(i,j)}^{(k)} = \beta) \\ &\leq \exp[-2(s_i^{(k)})^2 \phi_i + -2(t_j^{(k)})^2 \psi_j] \end{aligned}$$

If $\phi_i = p$ and $\psi_j = q$ for all i and j, then

Suppose event ω_k is that co-cluster C_k can't be find in any block $B_{(i,j)}$, then

$$\begin{aligned} P(\omega_k) &= \prod_{i=1}^{m} \prod_{j=1}^{n} P(M_{(i,j)}^{(k)} < T_m, N_{(i,j)}^{(k)} < T_n) \\ &\leq \prod_{i=1}^{m} \prod_{j=1}^{n} \exp\{-2 \left[(s_i^{(k)})^2 \phi_i + (t_j^{(k)})^2 \psi_j \right] \} \\ &= \exp\{-2 \sum_{i=1}^{m} \sum_{j=1}^{n} \left[(s_i^{(k)})^2 \phi_i + (t_j^{(k)})^2 \psi_j \right] \} \end{aligned}$$

If $\phi_i = \phi$ and $\psi_j = \psi$ for all i and j, then

$$\begin{aligned} s_i^{(k)} &= s^{(k)} = \frac{M^{(k)}}{M} - \frac{T_m - 1}{\phi} \\ t_j^{(k)} &= t^{(k)} = \frac{N^{(k)}}{N} - \frac{T_n - 1}{\psi} \end{aligned}$$

$$P(\omega_k) \leq \exp\left\{-2[\phi m (s^{(k)})^2 + \psi n (t^{(k)})^2]\right\}$$

And if we do T_p times of random sampling, the Probability of detecting the co-cluster is

$$\begin{aligned} P &= 1 - P(\omega_k)^{T_p} \\ &\geq 1 - \exp\left\{-2T_p[\phi m (s^{(k)})^2 + \psi n (t^{(k)})^2]\right\} \end{aligned}$$

according to which, we can set $m, n, \phi, \psi, T_m, T_n$ and T_p to ensure the probability of detecting the co-cluster is larger than a given threshold.

3 Nosiy case

3.1 Assumption

Assume each noise n_{ij} complies with a normal distribution $N(0, \sigma^2)$, i.i.d. for all i and j. Suppose $\exists \lambda > 0$, such that

$$\lambda \leq \max(||B||_1, ||B^\top||_1)/\sigma^2.$$

3.2 Score

The score of a submatrix $A_{I,J}$ is defined as

$$S(I,J) = \min(S_{row}(I,J), S_{col}(I,J)) \tag{1}$$

$$S_{row}(I,J) = \min_{i_1,i_2\in I}\left(1 - \frac{1}{|I|-1}\sum_{i_2\in I, i_2\neq i_1}\langle x_{i_1,J}, x_{i_2,J}\rangle\right) \tag{2}$$

$$S_{col}(I,J) = \min_{j_1,j_2\in J}\left(1 - \frac{1}{|J|-1}\sum_{j_2\in J, j_2\neq j_1}\langle x_{I,j_1}, x_{I,j_2}\rangle\right) \tag{3}$$

where $x_{i,J}$ is the i-th row of $A_{I,J}$. Here we define the inner product of two vectors x and y as

$$\langle x, y\rangle = \exp(-\frac{||x-y||_1^2}{2\alpha||x||_1||y||_1})$$

Suppose A is a hidden co-cluster matrix, and E is the noise matrix. Then the observed matrix is $B = A + E$. Consider co-cluster $B_{I,J}$, denote $1 - \frac{1}{|I|-1}\sum_{i_2\in I, i_2\neq i_1}\langle x_{i_1,J}, x_{i_2,J}\rangle$ as $s_{row}(i_1, i_2, J)$, then

$$\begin{aligned}
\mathbb{E}(s_{row}(i_1,i_2,J)) &= 1 - \frac{1}{|I|-1}\sum_{i_2\in I, i_2\neq i_1}\mathbb{E}(\langle x_{i_1,J}, x_{i_2,J}\rangle) \\
&= 1 - \frac{1}{|I|-1}\sum_{i_2\in I, i_2\neq i_1}\exp(-\frac{||x_{i_1,J}-x_{i_2,J}||_1^2}{2\alpha||x_{i_1,J}||_1||x_{i_2,J}||_1}) \\
&\geq 1 - \exp(-\frac{2}{\alpha\min(||x_{i_1,J}||_1, ||x_{i_2,J}||_1)}) \\
&\geq 1 - \exp(-\frac{2}{\alpha\max(||B||_1, ||B^\top||_1)}) \\
\sigma^2(s_{row}(i_1,i_2,J)) &= \frac{1}{|I|-1}\sum_{i_2\in I, i_2\neq i_1}\sigma^2(\langle x_{i_1,J}, x_{i_2,J}\rangle) \\
&= \frac{1}{|I|-1}\sum_{i_2\in I, i_2\neq i_1}\exp(-\frac{||x_{i_1,J}-x_{i_2,J}||_1^2}{2\alpha||x_{i_1,J}||_1||x_{i_2,J}||_1}) \\
&\leq \sigma^2\exp(-\frac{2}{\alpha\min(||x_{i_1,J}||_1, ||x_{i_2,J}||_1)}) \\
&\leq \sigma^2\exp(-\frac{2}{\alpha\max(||B||_1, ||B^\top||_1)})
\end{aligned}$$

Thus the expected value of $S_{row}(I,J)$ satisfies

$$\begin{aligned}
\mathbb{E}(S_{row}(I,J)) &\geq |J||I|\left(1 - \exp(-\frac{2}{\alpha\max(||B||_1, ||B^\top||_1)})\right) \\
\sigma^2(S_{row}(I,J)) &\leq |I||J|\sigma^2\exp(-\frac{2}{\alpha\max(||B||_1, ||B^\top||_1)})
\end{aligned}$$

Similarly, we can get

$$\mathbb{E}(S_{col}(I,J)) \geq |J||I| \left(1 - \exp(-\frac{2}{\alpha \max(||B||_1, ||B^\top||_1)})\right)$$
$$\sigma^2(S_{col}(I,J)) \leq |I||J|\sigma^2 \exp(-\frac{2}{\alpha \max(||B||_1, ||B^\top||_1)})$$

Then since $x^2 \leq x$ for $x \in (0,1)$, we have

$$\mathbb{E}(S(I,J)) \geq 1 - \exp(-\frac{2}{\alpha \max(||B||_1, ||B^\top||_1)})$$
$$\sigma^2(S(I,J)) \leq |I||J|\sigma^2 \exp(-\frac{2}{\alpha \max(||B||_1, ||B^\top||_1)})$$

According to the Chernoff bound, we have

$$\begin{aligned} P(S(I,J) \leq \mathbb{E}(S(I,J)) - \epsilon) &\leq \exp(-\frac{\epsilon^2}{2\sigma^2(S(I,J))}) \\ &\leq \exp(-\frac{\epsilon^2}{2|I||J|\sigma^2 \exp(-\frac{2}{\alpha \max(||B||_1, ||B^\top||_1)})}) \\ &\leq \exp(-\frac{\epsilon^2}{2|I||J|\sigma^2 \exp(-\frac{2}{\alpha\lambda})}) \end{aligned}$$

Thus if we set $\epsilon = \sqrt{2|I||J|\sigma^2 \exp(-\frac{2}{\alpha\lambda}) \log(1/\delta)}$, then

$$P(S(I,J) \leq \mathbb{E}(S(I,J)) - \epsilon) \leq \delta$$

Combine with the probability control of T_p, we can select parameters to ensure the probability of detecting the co-cluster is larger than a given threshold.