

Initialization convergence

Jiaxin Hu

August 4, 2021

1 Model&Algorithm

Suppose we have p nodes from r communities and observe the adjacent tensor $\mathcal{Y} \in \{0, 1\}^{p \times p \times p}$ whose entry \mathcal{Y}_{ijk} refers to the connection of the triplet (i, j, k) . Let $\theta = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$ denote the degree-corrected parameters and $z = (z_1, \dots, z_p) \in [r]^p$ denote the clustering assignment. Consider the hDCBM model

$$\mathbb{E}[\mathcal{Y}] = \mathcal{X} = \mathcal{S} \times_1 \Theta \mathbf{M} \times_2 \Theta \mathbf{M} \times_3 \Theta \mathbf{M},$$

where $\mathcal{S} \in \mathbb{R}^{r \times r \times r}$ is the symmetric core tensor, $\Theta = \text{diag}(\theta) \in \mathbb{R}^{p \times p}$ and $\mathbf{M} \in \mathbb{R}^{p \times r}$ is the hard membership matrix based on z . Consider the parameter space $\mathcal{P}_z(r, \beta)$ for $\beta > 1$ where

$$\mathcal{P}_z(r, \beta) = \left\{ z : \frac{p}{\beta r} \leq \sum_{j=1}^p \mathbf{1}\{z_j = a\} = p_a \leq \frac{\beta p}{r}, \quad \text{for all } a \in [r] \right\},$$

and the parameter space $\mathcal{P}(\delta, \Delta_{\min}, \alpha)$ for (Θ, \mathcal{S}) , where

$$\mathcal{P}(\delta, \Delta_{\min}, \alpha) = \left\{ (\mathcal{S}, \Theta) : \begin{aligned} & \min_{u \neq u' \in [r]} \min_{v, w \in [p]} (\mathcal{S}_{uvw} - \mathcal{S}_{u'vw})^2 = \Delta_{\min}^2, \\ & \frac{1}{p_a} \sum_{i: z_i = a} \theta_i^2 \in [1 - \delta, 1 + \delta], \text{ for all } a \in [r], \\ & \max_{u, v, w \in [r]} \mathcal{S}_{uvw} \leq \alpha, \mathcal{S}_{uvw} > 0, \text{ for all } u, v, w \in [p] \end{aligned} \right\}.$$

Notations

1. Let $\bar{\theta} = \max_{j \in [p]} \theta_j$.
2. Let $\mathbf{X}_k = \mathcal{M}_k(\mathcal{X})$ and \mathbf{X}_{kj} denote the j -th row of \mathbf{X}_k .

See Algorithm 1 for detailed algorithm. The convergence of the initialization is stated below:

Theorem 1.1 (Initialization convergence(conjecture)). Suppose $\delta = o(1)$, $\Delta_{\min} > 0$, $\|\theta\|_{\max} = o(p/r)$. Let \hat{z} denote the output of Algorithm 1. With probability at least $1 - C \exp(-cp)$, we have

$$\sum_{j: \hat{z}_j \neq z_j^*} \theta_j^2 \leq C \frac{(2\beta^2 + 1)\beta^2 r^2}{\Delta_{\min}^2 p^2 (1 - \delta)^2} \left[\frac{16\alpha^2(1 + M)}{\Delta_{\min}^2} + 1 \right] \bar{\theta}^3 \alpha (p^{3/2} r + pr + r^3).$$

Corollary 1. Suppose $\min_{j \in [p]} \theta_j > a$. Then, we have

$$h(\hat{z}, z^*) \leq C \frac{(2\beta^2 + 1)\beta^2 r^2}{\Delta_{\min}^2 p^2 (1 - \delta)^2} \left[\frac{16\alpha^2(1 + M)}{\Delta_{\min}^2} + 1 \right] \bar{\theta}^3 \alpha (p^{3/2} r + pr + r^3).$$

Remark 1. Here are few remarks:

1. The bound in Corollary 1 consists with the Theorem 2 in Han et al. (2020). Note that the definition of the misclassification error is

$$h(z, z') = \sum_{j \in [p]} \mathbf{1}\{z_j \neq z'_j\},$$

and the error in Han et al. (2020), $h_k^{(0)} = ph(z, z')$.

2. The bound in Theorem 1.1 in the matrix case, $m = 2$, $\mathcal{O}(\frac{1}{\Delta_{\min}^2})$, is better than the initialization bound in Gao et al. (2018), $\mathcal{O}(\frac{\sqrt{p}}{\Delta_{\min}^2})$. I believe the difference comes from the difference between k -means and k -median. Both Gao et al. (2018) and Han et al. (2020) use the gap-free estimation error $\|\hat{\mathcal{Y}} - \mathcal{X}\|_F = \mathcal{O}(\sqrt{p})$. However, in Gao et al. (2018), the k -median leads to the misclassification error bounded by $\|\hat{\mathcal{Y}}_j - \mathcal{X}_j\|_1 \leq p \|\hat{\mathcal{Y}}_j - \mathcal{X}_j\|_F = \mathcal{O}(p^{3/2})$ while k -means leads to the misclassification error bounded by $\|\hat{\mathcal{Y}}_j - \mathcal{X}_j\|_F^2 = \mathcal{O}(p)$. Thus, the extra \sqrt{p} occurs.
3. The bound for h in Ke et al. (2019) is $\mathcal{O}(\frac{1}{p^2})$ if we suppose all the degree-corrected parameter $\theta_i = \mathcal{O}(1)$, which is not optimal in the sense of convergence rate. Our bound in Corollary 1, $\mathcal{O}(\frac{1}{\Delta_{\min}^2 p^{m/2-1}})$, is much better in this sense.
4. On the other hand, Ke et al. (2019) uses θ_i to capture the sparsity in the network, and thus the condition $\theta_i = \mathcal{O}(1)$ for all $i \in [p]$ does not reflect the sparsity. Therefore, the average of θ_i , $\bar{\theta}$, is a function of p and they care more about the rate of $\bar{\theta}$ to let the error degenerate. In this sense, their Theorem 1 implies that the error degenerates if $p^{m-1} \bar{\theta}^m \geq \log n$, which allows a pretty small $\bar{\theta}$ and thus a large range of sparsity. In our results, we may not reflect such “lower bound” for the θ since we already assume $\theta_i = \mathcal{O}(1)$ in the parameter space. To reflect such lower bound, we need to relax the assumption $\frac{1}{p_a} \sum_{i: z_i=a} \theta_i^2 \in [1 - \delta, 1 + \delta]$ in the proof.

2 Proof of Theorem 1

For simplicity, we ignore the permutation of assignment here.

2.1 Gap-free estimation error bound

Note that our model is equal to

$$\mathcal{Y} = \mathcal{X} + \mathcal{Z}.$$

The \mathcal{Z} has independent entries following $\text{subG}(\bar{\theta}^3\alpha)$, which follows by the fact that $\|\mathcal{X}\|_{\max} = \bar{\theta}^3\alpha$. By Proposition 1 in [Han et al. \(2020\)](#), we have

$$\left\| \hat{\mathcal{Y}} - \mathcal{X} \right\|_F^2 \leq C\bar{\theta}^3\alpha(p^{3/2}r + pr + r^3),$$

with probability at least $1 - C\exp(-cp)$ and C is a constant.

2.2 Measurement of misclassification

For any set $S \subset [p]$, note that

$$\begin{aligned} \sum_{j \in S} \|\mathbf{X}_{kj}\|_F^2 &= \sum_{j \in S} \theta_j^2 \left(\sum_{k, l \in [p]} [\theta_k \theta_l \mathcal{S}_{z_j^* z_k^* z_l^*}]^2 \right) \\ &\geq \sum_{j \in S} \theta_j^2 \frac{p^2(1-\delta)^2}{\beta^2 r^2} \Delta_{\min}^2, \end{aligned}$$

where the second inequality follows by the fact that at most 1 $\mathcal{S}_{uvw} \leq \Delta_{\min}$ (otherwise, the minimal gap would smaller than Δ_{\min}) and the minimal size of the cluster is $\frac{p}{\beta r}$. Hence, to bound $\sum_{j \in S} \theta_j^2$, it is sufficient to bound $\sum_{j \in S} \|\mathbf{X}_{kj}\|_F^2$.

2.3 Weighted k -means

Further, let $\mathbf{X}_{kj}^s = \mathbf{X}_{kj} / \|\mathbf{X}_{kj}\|_F$. Note that $\mathbf{X}_{kj}^s = \mathbf{X}_{ki}^s$ is $z_i^* = z_j^*$. Note that the weighted k -means implies that

$$\sum_{j=1}^p \left\| \hat{\mathbf{Y}}_{kj} \right\|_F^2 \left\| (\hat{\mathbf{Y}}_{kj}^s) - \hat{x}_{\hat{z}_j} \right\|_F^2 \leq M \sum_{j=1}^p \left\| \hat{\mathbf{Y}}_{kj} \right\|_F^2 \left\| (\hat{\mathbf{Y}}_{kj}^s) - \mathbf{X}_{kj}^s \right\|_F^2,$$

where \hat{z}, \hat{x} are the estimated assignment and centroids. Hence, we may bound the term $\left\| (\hat{\mathbf{Y}}_{kj}^s)^T - \hat{x}_{\hat{z}_j} \right\|_F^2$ by the easier term $\left\| (\hat{\mathbf{Y}}_{kj}^s)^T - \mathbf{X}_{kj}^s \right\|_F^2$. Particularly, we have

$$\sum_{j=1}^p \left\| \hat{\mathbf{Y}}_{kj} \right\|_F^2 \left\| (\hat{\mathbf{Y}}_{kj}^s) - \mathbf{X}_{kj}^s \right\|_F^2 \leq 2 \sum_{j=1}^p \left\| \mathbf{Y}_{kj} - \mathbf{X}_{kj} \right\|_F^2 = 2 \left\| \hat{\mathcal{Y}} - \mathcal{X} \right\|_F^2,$$

where the inequality follows by the triangle inequality

$$\left\| \frac{x}{\|x\|_F} - \frac{y}{\|y\|_F} \right\|_F \leq \frac{2\|x - y\|_F}{\|x\| \vee \|y\|}.$$

2.4 Quantify the number of misclassification

Similarly with Lemma 6 in [Gao et al. \(2018\)](#), let $S = \{j \in [p] : \|\hat{x}_{\hat{z}_j} - \mathbf{X}_{kj}^s\|_F \geq \frac{1}{2\alpha}\Delta_{\min}\}$. First, for $z_j^* \neq z_i^*$, we know that

$$\|\mathbf{X}_{kj}^s - \mathbf{X}_{ki}^s\|_F^2 \geq \frac{\sum_{k,l \in [p]} \Delta_{\min}^2}{\sum_{k,l \in [p]} \alpha^2} = \frac{\Delta_{\min}^2}{\alpha^2}.$$

Next, we partition the clusters in three subsets based on the sets C_u , where

$$C_u = \left\{ j \in [p] : z_j^* = u, \|\hat{x}_{\hat{z}_j} - \mathbf{X}_{kj}^s\|_F \leq \frac{1}{2\alpha}\Delta_{\min} \right\}$$

and

$$\begin{aligned} R_1 &= \{u \in [r] : C_u = \emptyset\} \\ R_2 &= \{u \in [r] : C_u \neq \emptyset, \text{ for all } i, j \in C_u, \hat{z}_i = \hat{z}_j\} \\ R_3 &= \{u \in [r] : C_u \neq \emptyset, \text{ exist } i, j \in C_u, \hat{z}_i \neq \hat{z}_j\}. \end{aligned}$$

Note that

$$\sum_{j: \hat{z}_j \neq z_j^*} \theta_j^2 \leq \sum_{j \in S} \theta_j^2 + \sum_{j \in \cup_{u \in R_3} C_u} \theta_j^2.$$

Also, note that $|R_2| + 2|R_3| \leq r = |R_1| + |R_2| + |R_3|$, which implies that $|R_3| \leq |R_1|$. Then, we have

$$\begin{aligned} \sum_{j \in \cup_{u \in R_3} C_u} \theta_j^2 &\leq |R_3|(1 + \delta) \frac{\beta p}{r} \\ &\leq |R_1|(1 + \delta) \frac{\beta p}{r} \\ &\leq \frac{1 + \delta}{1 - \delta} \beta^2 \sum_{j \in \cup_{u \in R_1} C_u} \theta_j^2 \\ &\leq 2\beta^2 \sum_{j \in S} \theta_j^2, \end{aligned}$$

where the third inequality follows by the fact that $\sum_{j \in \cup_{u \in R_1} C_u} \theta_j^2 \geq |R_1|(1 - \delta) \frac{p}{\beta r}$, and the last inequality follows by the fact that $\cup_{u \in R_1} C_u \subset \{j \in S\}$ and $\delta = o(1)$. Therefore, we obtain

$$\sum_{j: \hat{z}_j \neq z_j^*} \theta_j^2 \leq (2\beta^2 + 1) \sum_{j \in S} \theta_j^2.$$

2.5 Assemble

Now, to bound the desire misclassification rate $\sum_{j: \hat{z}_j \neq z_j^*} \theta_j^2$, we only need to bound the $\sum_{j \in S} \|(\mathcal{M}_k(\mathcal{X}))_j\|_F^2$.

Note that

$$\begin{aligned} \sum_{j \in S} \|\mathbf{X}_{kj}\|_F^2 &\leq 2 \sum_{j \in S} \|\hat{\mathbf{Y}}_{kj}\|_F^2 + 2 \sum_{j \in S} \|\hat{\mathbf{Y}}_{kj} - \mathbf{X}_{kj}\|_F^2 \\ &\leq 2 \sum_{j \in S} \|\hat{\mathbf{Y}}_{kj}\|_F^2 + 2 \|\hat{\mathbf{Y}}_k - \mathbf{X}_k\|_F^2, \end{aligned}$$

where the second term is bounded by first part. For the first term, note that

$$\begin{aligned} \sum_{j \in S} \|\hat{\mathbf{Y}}_{kj}\|_F^2 &\leq \frac{4\alpha^2}{\Delta_{\min}^2} \sum_{j \in S} \|\hat{\mathbf{Y}}_{kj}\|_F^2 \|\hat{x}_{z_j} - \mathbf{X}_{kj}^s\|_F^2 \\ &\leq \frac{8\alpha^2}{\Delta_{\min}^2} \sum_{j \in S} \|\hat{\mathbf{Y}}_{kj}\|_F^2 \left[\left\| (\hat{\mathbf{Y}}_{kj}^s) - \hat{x}_{z_j} \right\|_F^2 + \left\| (\hat{\mathbf{Y}}_{kj}^s) - \mathbf{X}_{kj}^s \right\|_F^2 \right] \\ &\leq \frac{8\alpha^2(1+M)}{\Delta_{\min}^2} \sum_{i=1}^n \|\hat{\mathbf{Y}}_{kj}\|_F^2 \left\| (\hat{\mathbf{Y}}_{kj}^s) - \mathbf{X}_{kj}^s \right\|_F^2 \\ &\leq \frac{16\alpha^2(1+M)}{\Delta_{\min}^2} \left\| \hat{\mathbf{Y}}_k - (\mathcal{M}_k(\mathcal{X})) \right\|_F^2, \end{aligned}$$

where the first inequality follows by the definition of S , third and fourth inequality follow by the statements of k -mean in Section 2.3. Then, we know that the misclassification is bounded by the estimation error with polynomial rate,

$$\begin{aligned} \sum_{j: \hat{z}_j \neq z_j^*} \theta_j^2 &\leq (2\beta^2 + 1) \sum_{j \in S} \theta_j^2 \\ &\leq \frac{(2\beta^2 + 1)\beta^2 r^2}{\Delta_{\min}^2 p^2 (1 - \delta)^2} \sum_{j \in S} \|\mathbf{X}_{kj}\|_F^2 \\ &\leq \frac{2(2\beta^2 + 1)\beta^2 r^2}{\Delta_{\min}^2 p^2 (1 - \delta)^2} \left[\frac{16\alpha^2(1+M)}{\Delta_{\min}^2} + 1 \right] \left\| \hat{\mathbf{Y}}_k - (\mathcal{M}_k(\mathcal{X})) \right\|_F^2 \\ &\leq C \frac{(2\beta^2 + 1)\beta^2 r^2}{\Delta_{\min}^2 p^2 (1 - \delta)^2} \left[\frac{16\alpha^2(1+M)}{\Delta_{\min}^2} + 1 \right] \bar{\theta}^3 \alpha (p^{3/2} r + pr + r^3) \end{aligned}$$

with probability at least $1 - C \exp(-cp)$.

Algorithm 1 High-order weighted k -means clustering

Input: Observation $\mathcal{Y} \in \{0, 1\}^{p \times \dots \times p}$, r , relaxation factor in k -means $M > 1$, SCORE normalization function h .

- 1: Compute $\hat{U}_k = \text{SVD}_{r_k}(\mathcal{M}_k(\mathcal{Y}))$ for $k \in [d]$
- 2: **for** $k \in [3]$ **do**
- 3: Estimate the singular space \hat{U}_k via

$$\hat{U}_k = \text{SVD}_r(\mathcal{M}_k(\mathcal{Y} \times_1 \tilde{U}_1^T \times \dots \times_{k-1} \tilde{U}_{k-1}^T \times_{k+1} \tilde{U}_{k+1}^T \times \dots \times_3 \tilde{U}_3^T))$$

- 4: **end for**
- 5: Obtain

$$\hat{\mathcal{Y}} = \mathcal{Y} \times_1 \hat{U}_1 \hat{U}_1^T \times_2 \hat{U}_2 \hat{U}_2^T \times_3 \hat{U}_3 \hat{U}_3^T$$

- 6: **for** $k \in [3]$ **do**
- 7: Let $\mathbf{Y}_k = \mathcal{M}_k(\hat{\mathcal{Y}})$ and let $\hat{\mathbf{Y}}_{kj}$ denote the rows of $\hat{\mathbf{Y}}_k$ for $j \in [p]$. Obtain the SCORE normalized $\hat{\mathbf{Y}}_k^s$ via $\hat{\mathbf{Y}}_{kj}^s = \frac{\hat{\mathbf{Y}}_{kj}}{h(\hat{\mathbf{Y}}_{kj})}$ for $j \in [p]$.
- 8: Find the initial assignment $z_k^{(0)} \in [r]^p$ and centroids $\hat{x}_1, \dots, \hat{x}_r \in \mathbb{R}^{p^2}$ such that

$$\sum_{j=1}^p h(\hat{\mathbf{Y}}_{kj})^2 \left\| (\hat{\mathbf{Y}}_{kj}^s)^T - \hat{x}_{(z_k^{(0)})_j} \right\|_F^2 \leq M \min_{x_1, \dots, x_{r_k}, z_k} \sum_{j=1}^p h(\hat{\mathbf{Y}}_{kj})^2 \left\| (\hat{\mathbf{Y}}_{kj}^s)^T - \hat{x}_{(z_k^{(0)})_j} \right\|_F^2$$

- 9: **end for**
 - 10: Find the average of $z_k^{(0)}, k \in [3], z^{(0)}$.
- Output:** $\{z^{(0)} \in [r]^p\}$
-

3 Possible Improvement for Han et al. (2020)

In Theorem 2 of Han et al. (2020), we need the signal condition (15) to obtain bound of the misclassification rate $h_k^{(0)}$. Here, I would like to show another way to obtain the misclassification bound without the signal condition (15). However, condition (15) is required to ensure the misclassification error degenerates as $p \rightarrow \infty$. Also, the prove would be very similar with Gao et al. (2018) for Theorem 1, and we won't bound the $\ell_k^{(0)}$ without the signal condition.

Proof. The changes start from the definition of S . Define the S and \mathcal{C}_u as

$$S = \left\{ j \in [p_1] : \left\| \hat{\theta}_{(z_1^{(0)})_j} - \theta_{(z_1)_j}^* \right\|_2 \geq \sqrt{\frac{p-1}{r-1}} \frac{c_0 \Delta_1}{2} \right\},$$

$$\mathcal{C}_u = \{ j \in [p_1] : (z_1)_j = u, j \in S^c \}.$$

By the last inequality in Page 31, we know that $(z_1^{(0)})_j \neq (z_1^{(0)})_i$ if $j \in \mathcal{C}_a$ and $i \in \mathcal{C}_b$ for some $a \neq b$. However, it may not be true that for all $j \in \mathcal{C}_a$, $(z_1^{(0)})_j$ shares the same label. It is possible that for $i, j \in \mathcal{C}_a$ and $(z_1^{(0)})_j \neq (z_1^{(0)})_i$. Similarly with the proof of Lemma 6 in Gao et al. (2018), we separate the $u \in [r_1]$ in three subsets

$$R_1 = \{ u \in [r_1] : \mathcal{C}_u = \emptyset \}$$

$$R_2 = \left\{ u \in [r_1] : \mathcal{C}_u \neq \emptyset, \text{ for all } i, j \in \mathcal{C}_u, (z_1^{(0)})_i = (z_1^{(0)})_j \right\}$$

$$R_3 = \left\{ u \in [r_1] : \mathcal{C}_u \neq \emptyset, \text{ exist } i, j \in \mathcal{C}_u, (z_1^{(0)})_i \neq (z_1^{(0)})_j \right\}.$$

Note that $\cup_{u \in [r_1]} \mathcal{C}_u = S^c$. So, we have

$$\sum_{j \in [p_1]} \mathbf{1} \left\{ (z_1^{(0)})_j \neq \pi_1 \circ (z_1)_j \right\} \leq |\{ j \in \cup_{u \in R_3} \mathcal{C}_u \}| + |S|.$$

Note that $(z_1^{(0)})_j \neq (z_1^{(0)})_i$ if $j \in \mathcal{C}_a$ and $i \in \mathcal{C}_b$ for some $a \neq b$. We have $|R_2|$ different labels for the nodes $j \in \cup_{u \in R_2} \mathcal{C}_u$. Also, for nodes $j \in \cup_{u \in R_3} \mathcal{C}_u$, we have at least $2|R_3|$ different labels given by $(z_1^{(0)})_j$. Therefore, we have $|R_2| + 2|R_3| \leq |R_1| + |R_2| + |R_3|$, which implies $|R_3| \leq |R_1|$. Then

$$\begin{aligned} |\{ j \in \cup_{u \in R_3} \mathcal{C}_u \}| &\leq |R_3| \frac{\beta p_1}{r_1} \\ &\leq |R_1| \frac{\beta p_1}{r_1} \\ &\leq \frac{\beta}{\alpha} |\{ j \in \cup_{u \in R_1} \{ (z_1)_j = u \} \}| \\ &\leq \frac{\beta}{\alpha} |S|, \end{aligned}$$

where the third inequality follows by $|\{j \in \cup_{u \in R_1} \{(z_1)_j = u\}\}| \geq |R_1| \frac{\alpha p_1}{r_1}$. Hence, we have

$$\begin{aligned} h_1^{(0)} &= \frac{1}{p_1} \sum_{j \in [p_1]} \mathbf{1} \left\{ (z_1^{(0)})_j \neq \pi_1 \circ (z_1)_j \right\} \\ &\leq \frac{1}{p_1} \left(1 + \frac{\beta}{\alpha} \right) |S| \\ &\leq \frac{CMr_{-1}}{\Delta_{\min}^2 p_*} (r_* + \bar{p}\bar{r} + p_*^{1/2}\bar{r}), \end{aligned}$$

where the last inequality follows by the definition of S . This result is the same as current Theorem 2, and the signal condition (15) is not required. \square

References

- Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2018). Community detection in degree-corrected block models. The Annals of Statistics, 46(5):2153–2185.
- Han, R., Luo, Y., Wang, M., and Zhang, A. R. (2020). Exact clustering in tensor block model: Statistical optimality and computational limit. arXiv preprint arXiv:2012.09996.
- Ke, Z. T., Shi, F., and Xia, D. (2019). Community detection for hypergraph networks via regularized tensor power iteration. arXiv preprint arXiv:1909.06503.