Response Letter

Response to Major Comments from Editor

Reviewer 1 has identified a number of technical concerns in the proofs of some results. Please make sure that these are fully addressed in your revised manuscript.

Response: We thank editor and reviewers for the comments and suggestions. Reviewer 1 has mainly commented on the Bernoulli initialization in Section IV-C and the lower bound statement in Section III. In comment # 1, Reviewer 1 has identified the technical issue in the proof of Proposition 1. In comment # 3, Reviewer 1 has concerned the proving hardness of our impossibility statements compared with tensor block model (TBM) and pointed out the technical issues in the proof of Theorem 2.

For comment # 1, we have developed a new proof of Proposition 1 with a new lemma applicable for binary low-rank matrix estimation. The conclusion in Proposition 1 still holds. We have put the new proof in Appendix F with a more rigorous presentation.

For comment # 3, we have proved a new stronger statistical impossibility result in Theorem 2, inspired by the reviewer. With the new statement, we have added proof sketches in Section VIII to highlight the usage of dTBM-specific techniques in our proofs and revised the discussion in Section V to emphasize the comparison with TBM. We have provided the proof of the new Theorem 2 in Appendix D. We have addressed all reviewer's technical concerns in the new proof.

In addition to the aforementioned main points, we have addressed all comments from Reviewers 1 and 2. See our response in **Point-by-point response to Reviewer 1** and **Point-by-point response to Reviewer 2** for details.

Point-by-point response to Reviewer 1

The manuscript has been revised substantially and I appreciate the authors' efforts on providing additional theoretical and numeric results. Nevertheless, I think the response does not fully address my concerns and the added stuff may not be even technically sound, which I will elaborate as follow.

1. *My first comment in the previous review concerns with the applicability of the proposed framework on Bernoulli setting (hypergraph). The authors introduce the Section IV-C to discuss the initialization strategy for dense Bernoulli data (sub-algorithm 3) with guarantee (Proposition 1).*

   Response: There seem to be two sub-questions in this comment; we answer them one-by-one.

   (a) *The proposed strategy relies on the square-unfolding of tensors (i.e., Mat_{sq}). Since this is not a common tensor operation, the authors should provide the mathematical definition. My major concern is the soundness of Proposition 1, which is proved using one line by applying the result by Gao et al. (2018). The referenced paper is a benchmark work on matrix dSBM (so that the input matrix Y in their context is an exact adjacency matrix generated by dSBM). I doubt that, under tensor dSBM, is $Mat_{sq}(\mathcal{Y})$ a valid adjacency matrix? From my perspective, the $Mat_{sq}(\mathcal{Y})$ may not be a squared matrix and the independent noises may also derive from the setting studied by Gao et al. (2018). In any event, the authors should be more rigours on proving their mathematical statement instead of just saying "Following ...,".*

   Response: We thank reviewer for the suggestions and pointing out the issues in the proof of Proposition 1.

   First, we have added the definition of operation Mat_{sq} and the interpretation of the matricization $\mathrm{Mat}_{sq}(\mathcal{Y})$. The matrix $\mathrm{Mat}_{sq}(\mathcal{Y})$ is not square and symmetric, but we are able to interpret $\mathrm{Mat}_{sq}(\mathcal{Y})$ as a valid adjacency matrix for a bipartite network with connections between two groups of nodes. We quote the corresponding revision in Section IV-C:

   > "... One possible remedy is to apply singular value decomposition to the *square unfolding* (Mu et al., 2014), $\mathrm{Mat}_{sq}(\cdot)$, of Bernoulli tensor $\mathcal{Y} \in \{0,1\}^{p_1 \times \cdots \times p_K}$. Specifically, the square matricization $\mathrm{Mat}_{sq}(\mathcal{Y}) \in \{0,1\}^{p^{\lfloor K/2 \rfloor} \times p^{\lceil K/2 \rceil}}$ has entries $[\mathrm{Mat}_{sq}(\mathcal{Y})](j_1, j_2) = \mathcal{Y}(i_1, \ldots, i_K)$, where
   >
   > $$j_1 = i_1 + p_1(i_2 - 1) + \cdots + p_1 \cdots p_{\lfloor K/2 \rfloor - 1}(i_{\lfloor K/2 \rfloor} - 1),$$
   > $$j_2 = i_{\lceil K/2 \rceil} + p_{\lceil K/2 \rceil}(i_{\lceil K/2 \rceil + 1} - 1) + \cdots + p_{\lceil K/2 \rceil} \cdot p_{K-1}(i_K - 1).$$
   >
   > The matrix $\mathrm{Mat}_{sq}(\mathcal{Y})$ is asymmetric. We are able to interpret $\mathrm{Mat}_{sq}(\mathcal{Y})$ as the adjacency matrix for a bipartite network with connections between two groups of node. The two groups of nodes in the bipartite network have $p_1 \cdots p_{\lfloor K/2 \rfloor}$ and $p_{\lceil K/2 \rceil} \cdots p_K$ nodes, respectively. The entry $[\mathrm{Mat}_{sq}(\mathcal{Y})](j_1, j_2)$ refers to the presence of connection between the nodes indexed by combinations $(i_1, \ldots, i_{\lfloor K/2 \rfloor})$ and $(i_{\lceil K/2 \rceil}, \ldots, i_K)$. We summarize the procedure in Sub-algorithm 3. ... "

   Second, we agree that Gao et al. (2018, Lemma 7) is non-applicable in our case due to the asymmetry and the different estimation procedure to obtain $\hat{\mathcal{X}}'$. Instead, we have applied a new lemma for low-rank matrix estimation to upper bound the estimation error via square unfolding. The conclusion in Proposition 1 still holds with the new lemma. We have revised the proof more rigorously without "following..." terms and put the proof to Appendix F for a more detailed presentation.

For self-consistency, we attach the Proposition 1 with proof and the new lemma here:

" ...

Proposition 1 (Error for Bernoulli initialization)**.** Consider the Bernoulli dTBM in the parameter space $\mathcal{P}$ and Assumption 1 holds. Assume that $\boldsymbol{\theta}$ is balanced and $\min_{i\in[p]}\theta(i) \geq c$ for some constant $c > 0$. Let $z^{(0)}$ denote the output of Sub-algorithm 3. With probability going to 1, we have

$$\ell(z^{(0)}, z) \lesssim \frac{r^K p^{-\lfloor K/2 \rfloor}}{\mathrm{SNR}}, \quad \text{and} \quad L(z^{(0)}, z) \lesssim \sigma^2 r^K p^{-\lfloor K/2 \rfloor}.$$

... "

"...

Proof of Proposition 1. Sub-algorithm 3 shares the same algorithm strategy as Sub-algorithm 1 but with a different estimation of the mean tensor, $\hat{\mathcal{X}}'$. Hence, the proof of Proposition 1 follows the same proof idea with the proof of Theorem 4. Replacing the estimation $\hat{\mathcal{X}}$ by $\hat{\mathcal{X}}'$ in the proof of Theorem 4, we have

$$\min_{\pi\in\Pi} \sum_{i:z^{(0)}(i)\neq\pi(z(i))} \theta(i)^2 \lesssim \left(\sum_{i\in S} \|\boldsymbol{X}_{i:}\|^2 + \sum_{i\in S_0} \|\boldsymbol{X}_{i:}\|^2 \right) p^{-(K-1)} r^{K-1}. \tag{1}$$

By inequalities (44) and (46), we have

$$\sum_{i\in S} \|\boldsymbol{X}_{i:}\|^2 \leq \left(\frac{16(1+\eta)}{c_0^2 \Delta_{\min}^2} + 2 \right) \|\hat{\mathcal{X}}' - \mathcal{X}\|_F^2, \tag{2}$$

$$\sum_{i\in S_0} \|\boldsymbol{X}_{i:}\|^2 \leq \|\hat{\mathcal{X}}' - \mathcal{X}\|_F^2. \tag{3}$$

Hence, it suffices to find the upper bound of the estimation error $\|\hat{\mathcal{X}}' - \mathcal{X}\|_F^2$ to complete our proof. Note that the matricization $\mathrm{Mat}_{sq}(\mathcal{X}) \in \mathbb{R}^{p^{\lfloor K/2 \rfloor} \times p^{\lceil K/2 \rceil}}$ has $\mathrm{rank}(\mathrm{Mat}_{sq}(\mathcal{X})) \leq r^{\lceil K/2 \rceil}$, and Bernoulli random variables follow the sub-Gaussian distribution with bounded variance $\sigma^2 = 1/4$. Apply Lemma 9 to $\boldsymbol{Y} = \mathrm{Mat}_{sq}(\mathcal{Y})$, $\boldsymbol{X} = \mathrm{Mat}_{sq}(\mathcal{X})$, and $\hat{\boldsymbol{X}} = \mathrm{Mat}_{sq}(\hat{\mathcal{X}}')$. Then, with probability tending to 1 as $p \to \infty$, we have

$$\|\hat{\mathcal{X}}' - \mathcal{X}\|_F^2 = \|\mathrm{Mat}_{sq}(\hat{\mathcal{X}}') - \mathrm{Mat}_{sq}(\mathcal{X})\|_F^2 \lesssim p^{\lceil K/2 \rceil}. \tag{4}$$

Combining the estimation error (4) with inequalities (2), (3), and (1), we obtain

$$\min_{\pi\in\Pi} \sum_{i:z^{(0)}(i)\neq\pi(z(i))} \theta(i)^2 \lesssim \frac{\sigma^2 r^{K-1}}{\Delta_{\min}^2 p^{K-1}} p^{\lceil K/2 \rceil}. \tag{5}$$

Replace the inequality (47) in the proof of Theorem 4 by inequality (5). With the the same procedures to obtain $\ell(\hat{z}^{(0)}, z)$ and $L(\hat{z}^{(0)}, z)$ for Theorem 4 , we finish the proof of Proposition 1. □

Lemma 9 (Low-rank matrix estimation)**.** Let $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{E} \in \mathbb{R}^{m\times n}$, where $\boldsymbol{E}$ contains independent mean-zero sub-Gaussian entries with bounded variance σ^2. Suppose $\operatorname{rank}(\boldsymbol{X}) = r$. Consider the least square estimator

$$\hat{\boldsymbol{X}} = \underset{\boldsymbol{X}'\in\mathbb{R}^{m\times n}, \operatorname{rank}(\boldsymbol{X}')\le r}{\arg\min} \|\boldsymbol{X}' - \boldsymbol{Y}\|_F^2.$$

There exist positive constant C_1, C_2 such that

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_F^2 \le C_1\sigma^2 nr,$$

with probability at least $1 - \exp(-C_2 nr)$.

Proof of Lemma 9. Note that $\|\hat{\boldsymbol{X}} - \boldsymbol{Y}\|_F^2 \le \|\boldsymbol{X} - \boldsymbol{Y}\|_F^2$ by the definition of least square estimator. We have

$$\begin{aligned}
\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_F^2 &\le 2\left\langle \hat{\boldsymbol{X}} - \boldsymbol{X}, \boldsymbol{Y} - \boldsymbol{X}\right\rangle \\
&\le 2\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_F \sup_{\boldsymbol{T}\in\mathbb{R}^{m\times n}, \operatorname{rank}(\boldsymbol{T})\le 2r, \|\boldsymbol{T}\|_F=1} \langle \boldsymbol{T}, \boldsymbol{Y} - \boldsymbol{X}\rangle \\
&\le C\sigma\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_F\sqrt{nr},
\end{aligned}$$

with probability at least $1 - \exp(-C_2 nr)$, where the second inequality follows by re-arrangement, and the last inequality follows from Han et al. (2022, Lemma E5) under the matrix case. ” □

(b) *Also, Comparison in Remark 3 seems unfair. While the final exponential error rate developed in this paper is tight, it requires quite stronger SNR for a warm initialization. So, it does not make sense to compare the final rate without talking about under what SNR conditions these rates are achievable. Could the authors compare the SNR conditions under which both methods (the proposed one and Ke et al., 2019) achieve exact recovery?*

Response: It seems that reviewer intends to comment the Remark 8, which compares the error rates between our algorithm and Ke et al. (2019), rather than the Remark 3 for degree assumption, which is irrelevant to algorithm performance. We answer this question with focus on the Remark 8.

The method in Ke et al. (2019) adopts a signal notion based on the singular gap of the core tensor, denoted as Δ_{singular}. However, we are not able to infer the exact recovery of the spectral method by our angle-base SNR condition. There exists a case in which our algorithm achieves exact recovery but the spectral method in Ke et al. (2019) fails. Hence, for fair comparison, we compare the best performance of our algorithm and Ke et al. (2019) under the strongest signal setting of each model. We have added a new remark in Section IV-C and elaborated the comparison with Ke et al. (2019) with SNR consideration:

“...

Remark 10 (Comparison with previous methods)**.** Previous work (Ke et al., 2019) develops a spectral clustering method for Bernoulli dTBM. Ke et al. (2019) adopts a different signal notion based on the singular gap in the core tensor, denoted as Δ_{singular}. By Ke et al. (2019, Theorem 1), the spectral method achieves exact recovery with $\Delta_{\text{singular}} \gtrsim p^{-1/2}$. However, we are not able to infer the exact recovery of spectral method by our angle-base SNR condition. Consider an order-2 dTBM with $p > 2, \sigma^2 = 1$, $\boldsymbol{\theta} = \mathbf{1}$, equal size assignment $|z^{-1}(a)| = p/r$ for all $a \in [r]$, and core matrix $\boldsymbol{S} = \boldsymbol{I}_2$. The singular gap under this setting is $\Delta_{\text{singular}} = \min\{\lambda_1 - \lambda_2, \lambda_2\} = 0$, where $\lambda_1 \geq \lambda_2$ are singular values of $\boldsymbol{S}$. In contrast, our angle gap $\Delta^2_{\min} = 2$ satisfies the SNR condition in Theorem 5. Then, our algorithm achieves the exact recovery, but the spectral method in Ke et al. (2019) fails.

Hence, for fair comparison, we compare the best performance of our algorithm and Ke et al. (2019) under the strongest signal setting of each model. Since both methods contain an iteration procedure, we set the iteration number to infinity to avoid the computational error. Considering the largest angle-based SNR $\asymp 1$ in Theorem 5, our Bernoulli clustering achieves exponential error rate of order $\exp(-p^{(K-1)})$; considering the largest singular gap $\Delta_{\text{singular}} \asymp 1$ in Theorem 1 of Ke et al. (2019), the spectral clustering has a polynomial error rate of order p^{-2}. Our algorithm still shows a better theoretical accuracy than the competitive work for Bernoulli observations. "

2. *My second concern was on the imposed identifiability conditions and I think the authors have fully addressed it.*

3. *My last comment surrounding the two lower bounds is the major one, but I don't think the authors have addressed it appropriately.*

   Response: This comment concerns the proving hardness of the lower bound statements in dTBM and the flaws in the proof of Theorem 2. We separate the comment by these two aspects and address each aspect one-by-one.

   (a) *The authors argued that "Our results show the similar conclusion but under different conditions.". I agree that Δ_{ang} and Δ_{Euc} (using the notation is the response) are different quantities and Δ_{ang} can be much smaller than Δ_{Euc} in the proposed setting. However, that does NOT affect the hardness in proving the lower bound statement. For simplicity, I will just show that for statistical lower bound, which claims that*

$$\inf_{\hat{z}} \sup_{(z,S,\theta)\in P(\gamma)} \mathbb{E}[p\ell(\hat{z}, z)] \geq 1.$$

   Here $P(\gamma)$ refers to the parameter space where the core tensor S satisfies $\Delta^2_{\text{ang}}/\sigma^2 \leq p^\gamma$ and z, θ satisfy some regularity conditions (e.g., balance). To prove the lower bound, it suffices to fix particular θ and S and take supreme over the community variable z. Then, a trivial selection is to set $\theta_i = 1$ (homogeneous degree) and some S such that $\Delta_{\text{ang}}/\Delta_{\text{Euc}} = 1 + o(1)$. Such S can be constructed randomly: $S_{ijk} = 1 + \epsilon_{ijk}$ where ϵ_{ijk} are independent mean-zero Gaussian with variance tends to zero. After constructing such θ_i and S, the heterogeneity issue vanishes and Δ_{Euc} and Δ_{ang} becomes equivalent.

   The lower bounds arguments for TBM and dTBM become really different if the author considers *the following stronger statement:*

$$\inf_{S \in P(\gamma)} \inf_{\hat{z}} \sup_{(z,\theta)} \mathbb{E}[p\ell(\hat{z}, z)] \geq 1$$

This statement is stronger than the previous one, as it suggests statistical impossibility whenever the SNR condition is not met, while the current one actually suggests statistical impossibility for at least a specific S that does not meet SNR requirement.

Response: We thank reviewer for the suggestions. In short, we have provided a stronger statistical impossibility statement for dTBM, inspired by the reviewer. With the new statement, we argue that dTBM-specific techniques are required to obtain our statistical and computational impossibilities, though the proof idea shares the same spirit with reviewer's proof in comment. We have made several revisions in Section III, Section V, Section VIII, and Appendix D to address the comment. We attach each revision with explanations below.

First, we have provided a new impossibility result in Theorem 2. The new Theorem 2 suggests the statistical impossibility of dTBM whenever the core tensor $\mathcal{S}$ leads to a SNR with signal exponent $\gamma < -(K-1)$. This new impossibility statement is stronger than previous one which suggests the worst case impossibility for a particular $\mathcal{S}$. We quote the revised theorem and discussion in Section III here:

> "
>
> **Theorem 2** (Statistical critical value)**.** Consider general Gaussian dTBMs with parameter space $\mathcal{P}(\gamma)$ with $K \geq 1$. Then, we have the following statistical phase transition.
>
> • **Impossibility.** Assume $r \lesssim p^{1/3}$. Let $\mathcal{P}_{\mathcal{S}}(\gamma) := \{\mathcal{S} : c_3 \leq \|\text{Mat}(\mathcal{S})_{a:}\| \leq c_4, a \in [r]\} \cap \{\mathcal{S} : \Delta^2_{\text{min}} = p^{\gamma}\}$ denote the space for valid $\mathcal{S}$ satisfying SNR condition (5), and $\mathcal{P}_{z,\boldsymbol{\theta}} := \{\boldsymbol{\theta} \in \mathbb{R}^p_+, \frac{c_1 p}{r} \leq |z^{-1}(a)| \leq \frac{c_2 p}{r}, \left\|\boldsymbol{\theta}_{z^{-1}(a)}\right\|_1 = |z^{-1}(a)|, a \in [r]\}$ denote the space for valid $(z, \boldsymbol{\theta})$, where c_1, c_2, c_3, c_4 are the constants in parameter space (3). If the signal exponent satisfies $\gamma < -(K-1)$, then, for any true core tensor $\mathcal{S} \in \mathcal{P}_{\mathcal{S}}(\gamma)$, no estimator $\hat{z}_{\text{stat}}$ achieves exact recovery in expectation; that is,
>
> $$\gamma < -(K-1) \quad \Rightarrow \quad \liminf_{p \to \infty} \inf_{\mathcal{S} \in \mathcal{P}_{\mathcal{S}}(\gamma)} \inf_{\hat{z}_{\text{stat}}} \sup_{(z,\boldsymbol{\theta}) \in \mathcal{P}_{z,\boldsymbol{\theta}}} \mathbb{E}\left[p\ell(\hat{z}_{\text{stat}}, z)\right] \geq 1.$$
>
> ... "
>
> " ... The proofs for the two parts in Theorem 2 are in the Appendix II, Section D and Section G, respectively. The first part of Theorem 2 demonstrates impossibility of exact recovery whenever the core tensor $\mathcal{S}$ satisfies SNR condition (5) with exponent $\gamma < -(K-1)$. The proof is information-theoretical, and therefore the results apply to all statistical estimators, including but not limited to MLE and trace maximization (Ghoshdastidar and Dukkipati, 2017). The minimax bound (11) indicates the worst case impossibility for a particular core tensor $\mathcal{S}$ with signal exponent $\gamma < -(K-1)$; i.e., under the assumptions of Theorem 2,
>
> $$\gamma < -(K-1) \quad \Rightarrow \quad \liminf_{p \to \infty} \inf_{\hat{z}_{\text{stat}}} \sup_{(z,\mathcal{S},\boldsymbol{\theta}) \in \mathcal{P}(\gamma)} \mathbb{E}\left[p\ell(\hat{z}_{\text{stat}}, z)\right] \geq 1.$$

Such worst case impossibility is frequently studied in related works (Han et al., 2020; Gao et al., 2018) while our lower bound (11) provides a stronger impossibility statement for arbitrary core tensor with weak signal. ..."

Second, we have added proof sketches for the impossibility results to better highlight the dTBM-specific techniques we used. In fact, our proof idea is common in minimax analysis and shares the same spirit with reviewer's proof idea in the comment: find the lower bound of minimax error by constructing a particular set of true parameters. However, angle-based signal notion brings extra difficulties in parameter constructions, and the previous TBM construction is no longer applicable to address the arbitrariness of core tensor. We quote the new proof sketches in new Section VIII-A here:

" ...

The proofs of impossibility in Theorems 2 and 3 share the same proof idea with Han et al. (2020, Theorems 6 and 7) and Gao et al. (2018, Theorem 2). In both proofs of statistical and computational impossibilities, the key idea is to construct a particular set of parameters to lower bound the minimax rate. Specifically, for statistical impossibility in Theorem 2, we construct a particular $(z^*_{\text{stats}}, \boldsymbol{\theta}^*_{\text{stats}}) \in \mathcal{P}_{z,\boldsymbol{\theta}}$ such that for all $\mathcal{S}^* \in \mathcal{P}_{\mathcal{S}}(\gamma)$

$$\inf_{\hat{z}_{\text{stats}}} \sup_{(z,\boldsymbol{\theta})\in\mathcal{P}_{z,\boldsymbol{\theta}}} \mathbb{E}[p\ell(\hat{z}_{\text{stat}}, z)] \geq \inf_{\hat{z}_{\text{stats}}} \mathbb{E}[p\ell(\hat{z}_{\text{stat}}, z^*_{\text{stats}})|(z^*_{\text{stats}}, \mathcal{S}^*, \boldsymbol{\theta}^*_{\text{stats}})] \geq 1;$$

for computational impossibility in Theorem 3, we construct a particular $(z^*_{\text{comp}}, \mathcal{S}^*_{\text{comp}}, \boldsymbol{\theta}^*_{\text{comp}}) \in \mathcal{P}(\gamma)$ such that

$$\inf_{\hat{z}_{\text{comp}}} \sup_{(z,\mathcal{S},\boldsymbol{\theta})\in\mathcal{P}(\gamma)} \mathbb{E}[p\ell(\hat{z}_{\text{comp}}, z)] \geq \inf_{\hat{z}_{\text{comp}}} \mathbb{E}[p\ell(\hat{z}_{\text{comp}}, z^*_{\text{comp}})|(z^*_{\text{comp}}, \mathcal{S}^*_{\text{comp}}, \boldsymbol{\theta}^*_{\text{comp}})] \geq 1.$$

The constructions of $(z^*_{\text{stats}}, \boldsymbol{\theta}^*_{\text{stats}})$ and $(z^*_{\text{comp}}, \mathcal{S}^*_{\text{comp}}, \boldsymbol{\theta}^*_{\text{comp}})$ are the most critical steps. With good constructions, the lower bound "≥ 1" can be verified by classical statistical conclusions (e.g. Neyman-Pearson Lemma) or prior work (e.g. HPC Conjecture).

A notable detail in the proof of statistical impossibility is the arbitrariness of $\mathcal{S}^*$. The first infimum over $\mathcal{P}_{\mathcal{S}}(\gamma)$ in the minimax rate (11) requires that the lower bound (21) holds for any $\mathcal{S}^* \in \mathcal{P}_{\mathcal{S}}(\gamma)$. The arbitrary choice of $\mathcal{S}^*$ brings extra difficulties in the parameter construction, and consequently a non-trivial $\boldsymbol{\theta}^*_{\text{stats}} \neq \mathbf{1}$ is chosen to address the arbitrariness. Previous TBM construction in the proof of Han et al. (2020, Theorem 6) with $\boldsymbol{\theta}^*_{\text{stats}} = \mathbf{1}$ is no longer applicable in our case. Meanwhile, our construction $(z^*_{\text{comp}}, \mathcal{S}^*_{\text{comp}}, \boldsymbol{\theta}^*_{\text{comp}})$ leads to a rank-2 mean tensor to relate the HPC Conjecture while TBM Han et al. (2020, Theorem 7) constructs a rank-1 mean tensor. Hence, we emphasize that dTBM-specific techniques are required to obtain our impossibility results, though the proof idea is common for minimax lower bound analysis. ... "

Here are two details differ with reviewer's comment. First, reviewer's construction with $\boldsymbol{\theta}^* = \mathbf{1}$ is non-applicable for our new statistical impossibility. The TBM construction with $\boldsymbol{\theta}^*_{\text{stats}} = \mathbf{1}$ is not able to handle the arbitrary choice of $\mathcal{S}^*$, because the Euclidean gap condition $\Delta^2_{\text{Euc}} \leq p^{\gamma}$ is not guaranteed with arbitrary $\mathcal{S}^*$ satisfying $\Delta^2_{\text{ang}} \leq p^{\gamma}$; see Example 4 in our main text. In consequence, we design a special $\boldsymbol{\theta}^*_{\text{stats}} \neq \mathbf{1}$ in the proof. Second, we consider one particular core tensor $\mathcal{S}^*$ at a time, rather than the random tensor with entries $\mathcal{S}^*_{ijk} = 1 + \epsilon_{ijk}, \epsilon_{ijk} \sim N(0, \delta^2)$ for some $\delta \to 0$ as $p \to \infty$. The expectation in minimax error is taken with respect to the noise in observation, rather than the randomness in the true parameters.

We have also revised the discussion in Section V based on the new theorem and proof sketches:

> "...
>
> Third, from the perspective of proofs, we develop new dTBM-specific techniques to handle the extra degree heterogeneity. In our Theorem 2, we construct a special non-trivial degree heterogeneity to establish the lower bound for arbitrary core tensor with small angle gap, while, TBM (Han et al., 2020) considers the constructions without degree parameter. In our Theorem 3, we construct a rank-2 tensor to relate HPC conjecture to Δ^2_{ang}, while TBM (Han et al., 2020) constructs a rank-1 tensor to relate HPC conjecture to Δ^2_{Euc}. The asymptotic non-equivalence between Δ^2_{ang} and Δ^2_{Euc} renders our proof technically more involved. ..."

Last, we have added the comparison with TBM in terms of the stronger statistical impossibility. Though the presentation in the Theorem 6 of Han et al. (2020) suggests a weaker worst case impossibility for TBM, the proof of Theorem 6 indeed implies a stronger statistical impossibility for arbitrary core tensor with small signal. Both models show similar conclusions under different signal notions again, but we emphasize that different techniques are required to address the different signal notions. We quote the new discussion in Section V here:

> " ...
>
> Last, we discuss the statistical impossibility statements. Our Theorem 2 implies the statistical impossibility whenever the core tensor $\mathcal{S}$ leads to an angle-based SNR below the critical threshold. The Theorem 6 in Han et al. (2020) implies the worst case statistical impossibility for a particular core tensor $\mathcal{S}$ with Euclidean-based SNR below the statistical limit. Hence, our Theorem 2 shows a stronger statistical impossibility for dTBM than that presented in TBM Han et al. (2020, Theorem 6). However, inspecting the proof of Han et al. (2020), the proof of Theorem 6 indeed implies a stronger TBM impossibility statement for arbitrary core tensor; i.e.,
>
> $$\gamma < -(K-1) \quad \Rightarrow \quad \liminf_{p \to \infty} \inf_{\mathcal{S} \in \mathcal{P}_{\mathcal{S},\text{TBM}} \cap \{\Delta^2_{\text{Euc}} = p^\gamma\}} \inf_{\hat{z}_{\text{stats}}} \sup_{z \in \mathcal{P}_{z,\text{TBM}}} \mathbb{E}[p\ell(\hat{z}_{\text{stats}}, z)] \geq 1,$$
>
> where $\mathcal{P}_{\mathcal{S},\text{TBM}}$ and $\mathcal{P}_{z,\text{TBM}}$ refer to the space for core tensor $\mathcal{S}$ and assignment z under TBM, respectively. Again, in terms of the strong statistical impossibility, both models show similar conclusions but under different conditions. Since two impossibilities consider different core tensor regimes with non-equivalent Δ^2_{ang} and Δ^2_{Euc}, we emphasize that different proof techniques are required to obtain these similar conclusions. See our proof sketch in Section VIII-A, Appendices D and E for detail technical differences. "

For self-consistency, we also attach the proof of the new impossibility theorem in Appendix D:

> "...

Proof of Theorem 2 (Impossibility). Consider the general asymetric dTBM (27) in the special case that $p_k = p$ and $r_k = r$ for all $k \in [K]$. For simplicity, we show the minimax rate for the estimation on the first mode $\hat{z}_1$; the proof for other modes are essentially the same.

To prove the minimax rate (11), it suffices to take an arbitrary $\mathcal{S}^* \in \mathcal{P}_{\mathcal{S}}(\gamma)$ wih $\gamma < -(K-1)$ and construct $(z_k^*, \boldsymbol{\theta}_k^*)$ such that

$$\inf_{\hat{z}_1} \mathbb{E}\left[p\ell(\hat{z}_1, z_1^*)|(z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*)\right] \geq 1.$$

We first define a subset of indices $T_k \subset [p_k], k \in [K]$ in order to avoid the complication of label permutation. Based on Han et al. (2020, Proof of Theorem 6), we consider the restricted family of $\hat{z}_k$'s for which the following three conditions are satisfied:

$$\text{(a) } \hat{z}_k(i) = z_k(i) \text{ for all } i \in T_k; \quad \text{(b) } |T_k^c| \asymp \frac{p}{r};$$
$$\text{(c) } \min_{\pi \in \Pi} \sum_{i \in [p]} \mathbb{1}\{\hat{z}_k(i) \neq \pi \circ z_k(i)\} = \sum_{i \in [p]} \mathbb{1}\{\hat{z}_k(i) \neq z_k(i)\},$$

for all $k \in [K]$. Now, we consider the construction:

(i) $\{z_k^*\}$ satisfies properties (a)-(c) with misclassification sets T_k^c for all $k \in [K]$;

(ii) $\{\boldsymbol{\theta}_k^*\}$ such that $\boldsymbol{\theta}_k^*(i) \leq \sigma r^{(K-1)/2} p^{-(K-1)/2}$ for all $i \in T_k^c, k \in [K]$ and $\max_{k \in [K], a \in [r]} \|\boldsymbol{\theta}_{k, z_k^{*,-1}(a)}\|_2^2 \asymp p/r$.

Combining the inequalities (39) and (40) in the proof of Theorem 2 in Gao et al. (2018), we have

$$\inf_{\hat{z}_1} \mathbb{E}\left[\ell(\hat{z}_1, z_1^*)|(z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*)\right] \geq$$
$$\frac{C}{r^3|T_1^c|} \sum_{i \in T_1^c} \inf_{\hat{z}_1(i)} \{\mathbb{P}[\hat{z}_1(i) = 1 | z_1^*(i) = 2, z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*] + \mathbb{P}[\hat{z}_1(i) = 2 | z_1^*(i) = 1, z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*]\},$$

where C is some positive constant, $\hat{z}_1$ on the left hand side denote the generic assignment functions in $\mathcal{P}(\gamma)$, and the infimum on the right hand side is taken over the generic assignment function family of $\hat{z}_1(i)$ for all nodes $i \in T_1^c$. Here, the factor $r^3 = r \cdot r^2$ in (32) comes from two sources: $r^2 \asymp \binom{r}{2}$ comes from the multiple testing burden for all pairwise comparisons among r clusters; and another r comes from the number of elements $|T_k^c| \asymp p/r$ to be clustered.

Next, we need to find the lower bound of the rightmost side in (32). We consider the hypothesis test based on model (27). First, we reparameterize the model under the construction (i)-(ii).

$$\boldsymbol{x}_a^* = [\mathrm{Mat}_1 (\mathcal{S}^* \times_2 \boldsymbol{\Theta}_2^* \boldsymbol{M}_2^* \times_3 \cdots \times_K \boldsymbol{\Theta}_K^* \boldsymbol{M}_K^*)]_{a:}, \quad \text{for all } a \in [r],$$

where $\boldsymbol{x}_a^*$'s are centroids in $\mathbb{R}^{p^{K-1}}$. Without loss of generality, we consider the lower bound for the summand in (32) for $i = 1$. The analysis for other $i \in T_1^c$ are similar.

For notational simplicity, we suppress the subscript i and write $\boldsymbol{y}, \theta^*, z$ in place of $\boldsymbol{y}_1, \boldsymbol{\theta}_1^*(1)$ and $z_1(1)$, respectively. The equivalent vector problem for assessing the summand in (32) is

$$\boldsymbol{y} = \theta^* \boldsymbol{x}_z^* + \boldsymbol{e},$$

where $z \in \{1,2\}$ is an unknown parameter, $\theta^* \in \mathbb{R}_+$ is the given heterogeneity degree, $\boldsymbol{x}_1^*, \boldsymbol{x}_2^* \in \mathbb{R}^{p^{K-1}}$ are given centroids, and $\boldsymbol{e} \in \mathbb{R}^{p^{K-1}}$ consists of i.i.d. $N(0,\sigma^2)$ entries. Then, we consider the hypothesis testing under the model (33):

$$H_0 : z = 1 \text{ and } \boldsymbol{y} = \theta^* \boldsymbol{x}_1^* + \boldsymbol{e} \quad \leftrightarrow \quad H_1 : z = 2 \text{ and } \boldsymbol{y} = \theta^* \boldsymbol{x}_2^* + \boldsymbol{e},$$

The hypothesis testing (34) is a simple versus simple testing, since the assignment z is the only unknown parameter in the test. By Neyman-Pearson lemma, the likelihood ratio test is optimal with minimal Type I + II error. Under Gaussian model, the likelihood ratio test of (34) is equivalent to the least square estimator $\hat{z}_{LS} = \arg\min_{a=\{1,2\}} \|\boldsymbol{y} - \theta^* \boldsymbol{x}_a^*\|_F^2$.

Let $\boldsymbol{S} = \mathrm{Mat}_1(\mathcal{S})$. Note that

$$\begin{aligned}
\|\theta^* \boldsymbol{x}_1^* - \theta^* \boldsymbol{x}_2^*\|_F &\leq \theta^* \|\boldsymbol{S}_{1:}^* - \boldsymbol{S}_{2:}^*\|_F \prod_{k=2}^{K} \lambda_{\max}(\boldsymbol{\Theta}_k^* \boldsymbol{M}_k^*) \\
&\leq \theta^* \|\boldsymbol{S}_{1:}^* - \boldsymbol{S}_{2:}^*\|_F \max_{k \in [K]/\{1\}, a \in [r]} \|\boldsymbol{\theta}_{k, z_k^{*,-1}(a)}\|_2^{K-1} \\
&\leq \sigma r^{(K-1)/2} p^{-(K-1)/2} 2 c_4 p^{(K-1)/2} r^{-(K-1)/2} \\
&\leq 2 c_4 \sigma,
\end{aligned}$$

where $\lambda_{\max}$ denote the maximal singular value, the second inequality follows from Lemma 6, and the third inequality follows from property (ii) and the boundedness constraint in $\mathcal{P}_{\mathcal{S}}(\gamma)$ such that $\|\boldsymbol{S}_{1:}^* - \boldsymbol{S}_{2:}^*\|_F \leq \|\boldsymbol{S}_{1:}^*\|_F + \|\boldsymbol{S}_{2:}^*\|_F \leq 2c_4$.

Hence, we have

$$\begin{aligned}
&\inf_{\hat{z}_1(1)} \{\mathbb{P}[\hat{z}_1(1) = 1 | z_1^*(1) = 2, z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*] + \mathbb{P}[\hat{z}_1(1) = 2 | z_1^*(1) = 1, z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*]\} \\
&\quad = 2\mathbb{P}[\hat{z}_{LS} = 1 | z_1^*(1) = 2, z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*] \\
&\quad = 2\mathbb{P}[\|\boldsymbol{y} - \theta^* \boldsymbol{x}_1^*\|_F^2 \leq \|\boldsymbol{y} - \theta^* \boldsymbol{x}_2^*\|_F^2 | z_1^*(1) = 2, z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*] \\
&\quad = 2\mathbb{P}\left[2\left\langle \boldsymbol{e}, \theta^* \boldsymbol{x}_1^* - \theta^* \boldsymbol{x}_2^* \right\rangle \geq \|\theta^* \boldsymbol{x}_1^* - \theta^* \boldsymbol{x}_2^*\|_F^2\right] \\
&\quad = 2\mathbb{P}[N(0,1) \geq \theta^* \|\boldsymbol{x}_1^* - \boldsymbol{x}_2^*\|_F / (2\sigma)] \\
&\quad \geq 2\mathbb{P}[N(0,1) \geq c_4] \geq c,
\end{aligned}$$

where the first equation holds by symmetry, the third equation holds by rearrangement, the fourth equation holds by the fact that $\langle \boldsymbol{e}, \theta^* \boldsymbol{x}_1^* - \theta^* \boldsymbol{x}_2^* \rangle \sim N(0, \sigma \|\theta^* \boldsymbol{x}_1^* - \theta^* \boldsymbol{x}_2^*\|_F)$, and c is some positive constant in the last inequality.

Plugging the inequality (35) into the inequality (32) for all $i \in T_1^c$, then, we have

$$\liminf_{p \to \infty} \inf_{\hat{z}_1} \mathbb{E}\left[p\ell(\hat{z}_1, z_1^*) | z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*\right] \geq \liminf_{p \to \infty} \frac{Ccp}{r^3} \geq Cc,$$

where the last inequality follows by the condition $r = o(p^{1/3})$. By the discrete nature of the misclassification error, we obtain our conclusion

$$\liminf_{p\to\infty} \inf_{\mathcal{S}^*\in\mathcal{P}_{\mathcal{S}}(\gamma)} \inf_{\hat{z}_{\text{stat}}} \sup_{(z^*,\boldsymbol{\theta}^*)\in\mathcal{P}_{z,\boldsymbol{\theta}}} \mathbb{E}\left[p\ell(\hat{z}_{\text{stat}}, z)\right] \geq 1.$$

" $\square$

(b) *Finally, I checked the author's proof towards the statement. I think the proofs need to be revised substantially as I found several flaws:*

Response: We thank reviewer for pointing out these issues. We have provided a new proof of Theorem 2 for the new statistical impossibility statement. We will reply to each issue based on the related parts in the new proof.

- *Page 32, Eqn (31) suggests that θ is fixed and the decision problems are reduced to z from that point. But in latter argument θ is taken as unknown parameter.*

  Response: We agree that the inconsistent status of $\boldsymbol{\theta}$ is a flaw in previous proof of Theorem 2. Once we fix the true $\boldsymbol{\theta}$ at the beginning (Eqn (31) of previous proof), we should take true $\boldsymbol{\theta}$ as a known parameter in the rest of the proof. Specifically, in previous Eqn (33), the model of $\boldsymbol{y}$

  $$\boldsymbol{y} = \theta\boldsymbol{x}_z + \boldsymbol{e}$$

  should take θ as a known parameter. Because the distribution of observation $\boldsymbol{y}$ corresponds to the conditional probability $\mathbb{P}[\cdot|z_k(i)]$ in previous Eqn (31). When the true assignment $z_k(i)$ is known, the distribution of $\boldsymbol{y}$ should be known with true heterogeneity θ and true centroid $\boldsymbol{x}_{z_k(i)}$. To avoid this flaw in the new proof, we use new notations $z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}_k^*$ for true parameters to better distinguish the known and unknown parameters. We also re-write the conditional probability as $\mathbb{P}[\cdot|z_k^*(i), z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*]$ for better expression. We attach the related parts in the new proof in Appendix D here:

  > "... To prove the minimax rate (11), it suffices to take an arbitrary $\mathcal{S}^* \in \mathcal{P}_{\mathcal{S}}(\gamma)$ wih $\gamma < -(K-1)$ and construct $(z_k^*, \boldsymbol{\theta}_k^*)$ such that
  >
  > $$\inf_{\hat{z}_1} \mathbb{E}\left[p\ell(\hat{z}_1, z_1^*)|(z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*)\right] \geq 1.$$
  >
  > ..."
  >
  > "... we have
  >
  > $$\inf_{\hat{z}_1} \mathbb{E}\left[\ell(\hat{z}_1, z_1^*)|(z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*)\right] \geq$$
  > $$\frac{C}{r^3|T_1^c|}\sum_{i\in T_1^c} \inf_{\hat{z}_1(i)} \{\mathbb{P}[\hat{z}_1(i) = 1|z_1^*(i) = 2, z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*] + \mathbb{P}[\hat{z}_1(i) = 2|z_1^*(i) = 1, z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*]\},$$
  >
  > where C is some positive constant, ..."
  >
  > " ... The equivalent vector problem for assessing the summand in (32) is
  >
  > $$\boldsymbol{y} = \theta^*\boldsymbol{x}_z^* + \boldsymbol{e},$$

where $z \in \{1,2\}$ is an unknown parameter, $\theta^* \in \mathbb{R}_+$ is the given heterogeneity degree, $\boldsymbol{x}_1^*, \boldsymbol{x}_2^* \in \mathbb{R}^{p^{K-1}}$ are given centroids, and $\boldsymbol{e} \in \mathbb{R}^{p^{K-1}}$ consists of i.i.d. $N(0,\sigma^2)$ entries. ..."

- *Page 33, Line 9: let's say the authors may want to make θ unknown and involve it in the decision problem. The authors formulate a hypothesis testing for model (33). My question is, is this really a hypotheses testing? Eqn. (30) involves both z and θ as unknown parameters. Under null ($z = 1$), the distribution of y is still unknown.*

  Response: This is a follow-up question of last comment. We agree that involving $\boldsymbol{\theta}$ as an unknown parameter in the model of $\boldsymbol{y}$ and hypothesis testing is in appropriate. Following our last response, we consider the true heterogeneity $\boldsymbol{\theta}^*$ as known in the model of $\boldsymbol{y}$. Hence, we have a valid simple hypothesis testing problem, in which the assignment function z is the only unknown parameter in the test. Under the null hypothesis ($z = 1$), the observation $\boldsymbol{y} = \theta^* \boldsymbol{x}_1^* + \boldsymbol{e}$ follows multivariate Gaussian distribution with known mean vector $\theta^* \boldsymbol{x}_1^*$ and covariance matrix $\sigma^2 \boldsymbol{I}$. For clarification, we re-write the hypothesis testing with the model of $\boldsymbol{y}$. We attach the related parts in the new proof in Appendix D here:

  > "... The equivalent vector problem for assessing the summand in (32) is
  >
  > $$\boldsymbol{y} = \theta^* \boldsymbol{x}_z^* + \boldsymbol{e},$$
  >
  > where $z \in \{1,2\}$ is an unknown parameter, $\theta^* \in \mathbb{R}_+$ is the given heterogeneity degree, $\boldsymbol{x}_1^*, \boldsymbol{x}_2^* \in \mathbb{R}^{p^{K-1}}$ are given centroids, and $\boldsymbol{e} \in \mathbb{R}^{p^{K-1}}$ consists of i.i.d. $N(0,\sigma^2)$ entries. Then, we consider the hypothesis testing under the model (33):
  >
  > $$H_0 : z = 1 \text{ and } \boldsymbol{y} = \theta^* \boldsymbol{x}_1^* + \boldsymbol{e} \quad \leftrightarrow \quad H_1 : z = 2 \text{ and } \boldsymbol{y} = \theta^* \boldsymbol{x}_2^* + \boldsymbol{e},$$
  >
  > ... "

- *A follow-up question is why could you apply Neyman-Pearson Lemma to lower bound the decision risks? Apparently this is not a simple testing, and the authors should provide the justification on why the presented "profiled MLE" works (technically) in determining the lower bound.*

  Response: The revised hypothesis testing in new Eqn (34) is a simple ($z = 1$) versus simple ($z = 2$) testing problem. Because the assignment function z is the only unknown parameter in test. Therefore, we are able to apply Neyman-Pearson Lemma to our testing problem. By Neyman-Pearson Lemma, the likelihood ratio test, which coincides to the least square estimator of z under our Gaussian model, leads to the minimal Type I + II error. As θ^* is a known parameter in the new proof, we no longer use the "profile MLE" to determine the lower bound. We attach the related parts in the new proof in Appendix D here:

  > " ... The hypothesis testing (34) is a simple versus simple testing, since the assignment z is the only unknown parameter in the test. By Neyman-Pearson lemma, the likelihood ratio test is optimal with minimal Type I + II error. Under Gaussian model, the likelihood ratio test of (34) is equivalent to the least square estimator $\hat{z}_{LS} = \arg\min_{a=\{1,2\}} \|\boldsymbol{y} - \theta^* \boldsymbol{x}_a^*\|_F^2$. ... "

- *Although I don't understand why $\hat{\theta}_{\text{MLE}}$ is introduced here, I feel its calculation is incorrect. Does the authors consider the fact that there is a pre-specified parameter space for θ and it cannot take all values on $\mathbb{R}^+$?*

  Response: In the new proof, we no longer introduce the estimator $\hat{\theta}_{\text{MLE}}$ and $\hat{z}_{\text{MLE}}$ defined in the previous proof. Instead, we consider the least square estimator by Neyman-Pearson Lemma. In general, we agree that the MLE of $(z, \boldsymbol{\theta}, \mathcal{S})$ should lie in the pre-specified parameter space $\mathcal{P}$ to guarantee the identifiability. We emphasize this point in the formal definition of MLE (10) in the main text. We attach the definition in Section III here:

  > " We consider the Gaussian MLE, denoted as $(\hat{z}_{\text{MLE}}, \hat{\boldsymbol{\theta}}_{\text{MLE}}, \hat{\mathcal{S}}_{\text{MLE}})$, over the estimation space $\mathcal{P}$, where
  >
  > $$(\hat{z}_{\text{MLE}}, \hat{\boldsymbol{\theta}}_{\text{MLE}}, \hat{\mathcal{S}}_{\text{MLE}}) = \underset{(z, \boldsymbol{\theta}, \mathcal{S}) \in \mathcal{P}}{\arg\min} \|\mathcal{Y} - \mathcal{X}(z, \boldsymbol{\theta}, \mathcal{S})\|_F^2.$$
  >
  > ... "

Point-by-point response to Reviewer 2

I thank the authors for the significant revision that clarifies the issues that I raised. I still suggest the authors to emphasise early that this paper focuses on dense regime. For instance, they could include sparsity in Table 1 and mention that Ahn et al (2018) proves exact recovery in the sparse regime (which this paper and few others in the list don't). This would provide a clear (unbiased) comparison of the results.

Response: We thank reviewer for the reference. We have included the reference Ahn et al. (2018) in Table I:

	Gao et al. (2018)	Ahn et al. (2018)	Han et al. (2020)	Ghoshdastidar et al. (2017)	Ke et al. (2019)	**Ours**
Allow tensors of arbitrary order	$\times$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Allow degree heterogeneity	$\surd$	$\times$	$\times$	$\surd$	$\surd$	$\surd$
Singular-value gap-free clustering	$\surd$	$\surd$	$\surd$	$\times$	$\times$	$\surd$
Misclustering rate (for order K^*)	-	$p^{-(K-1)}\alpha_p^{-1**}$	$\exp(-p^{K/2})$	p^{-1}	p^{-2}	$\exp(-p^{K/2})$
Consider sparse observation	$\times$	$\surd$	$\times$	$\times$	$\times$	$\times$

Table 1: Comparison between previous methods with our method. *We list the result for order-K tensors with $K \geq 3$ and general number of communities $r = \mathcal{O}(1)$. **The parameter $\alpha = f(p) > 0$ denotes the sparsity level which is some function of dimension p.

We have added a brief clarification in the related work, Section I:

> "... Some works (Ahn et al., 2018) study the TBM with sparse observations, while, others (Wang and Zeng, 2019; Han et al., 2020) and our work focus on the dense regime. ..."

We also added the reference in the extension to Bernoulli clustering, Section IV-C:

> "The sparsity is often a popular feature in hypergraphs (Florescu and Perkins, 2016; Ke et al., 2019; Ahn et al., 2018). "

References

Ahn, K., Lee, K., and Suh, C. (2018). Hypergraph spectral clustering in the weighted stochastic block model. *IEEE Journal of Selected Topics in Signal Processing*, 12(5):959–974.

Florescu, L. and Perkins, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory*, pages 943–959. PMLR.

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.

Ghoshdastidar, D. and Dukkipati, A. (2017). Uniform hypergraph partitioning: Provable tensor methods and sampling techniques. *The Journal of Machine Learning Research*, 18(1):1638–1678.

Ghoshdastidar, D. et al. (2017). Consistency of spectral hypergraph partitioning under planted partition model. *The Annals of Statistics*, 45(1):289–315.

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*.

Han, R., Willett, R., and Zhang, A. R. (2022). An optimal statistical and computational framework for generalized tensor estimation. *The Annals of Statistics*, 50(1):1–29.

Ke, Z. T., Shi, F., and Xia, D. (2019). Community detection for hypergraph networks via regularized tensor power iteration. *arXiv preprint arXiv:1909.06503*.

Mu, C., Huang, B., Wright, J., and Goldfarb, D. (2014). Square deal: Lower bounds and improved relaxations for tensor recovery. In *International Conference on Machine Learning*, pages 73–81.

Wang, M. and Zeng, Y. (2019). Multiway clustering via tensor block models. In *Advances in Neural Information Processing Systems*, pages 713–723.