Response Letter

Response to Major Comments from Editor

1. *The similarity of the manuscript to the previous paper "Exact clustering in tensor block model: Statistical optimality and computational limit". Please clearly highlight the connections with* Han et al. (2020) *and explain the differences.*

Response: We have added a new Section **??**, "Comparison with Non-degree Tensor Block Model", to discuss the connections and difference between our work and Han et al. (2020) from three aspects: signal notions, theoretical results, and algorithms. We quote the new section below:

> "...We discuss the connections and difference between our dTBM and TBM (Han et al., 2020) from three aspects: signal notions, theoretical results, and algorithms. Without loss of generality, let $\sigma^2 = 1$.
>
> • *Signal notion.* The signal levels in both TBM (Han et al., 2020) and our dTBM are functions of the core tensor $\mathcal{S}$. We emphasize that the signal notions are different between the two models. In particular, the Euclidean-based signal notion in TBM Han et al. (2020) fails to accurately describe the phase transition in our dTBM due to the possible heterogeneity in degree $\boldsymbol{\theta}$. To compare, we denote our angle-based signal notion in (**??**) and the Euclidean-based SNR in Han et al. (2020) as Δ_{ang}^2 and Δ_{Euc}^2, respectively:
>
> $$\Delta_{\mathrm{ang}}^2 = 2(1 - \max_{a \neq b \in [r]} \cos{(\boldsymbol{S}_{a:}, \ \boldsymbol{S}_{b:})}), \quad \Delta_{\mathrm{Euc}}^2 = \min_{a \neq b \in [r]} \|\boldsymbol{S}_{a:} - \boldsymbol{S}_{b:}\|^2.$$
>
> By Lemma **??** in the Appendix II, we have
>
> $$\Delta_{\mathrm{ang}}^2 \max_{a \in [r]} \|\boldsymbol{S}_{a:}\|^2 \leq \Delta_{\mathrm{Euc}}^2.$$
>
> The above inequality indicates the condition $\Delta_{\mathrm{Euc}}^2 \leq p^\gamma$ is **sufficient but not necessary** for $\Delta_{\mathrm{ang}}^2 \leq p^\gamma$. In fact, if we were to use Δ_{Euc}^2 for both models, then the phase transition of dTBM can be arbitrarily worse than that for TBM.
>
> Here, we provide an example to illustrate the dramatical difference between TBM and dTBM with the same core tensor.
>
> **Example 1** (Comparison with Euclidean-based signal notion)**.** Consider a biclustering model with $\boldsymbol{\theta} = 1$ and an order-2 core matrix
>
> $$\boldsymbol{S} = \begin{pmatrix} p^{(\gamma+1)/2} + 2 & 2p^{(\gamma+1)/2} + 4 \\ 2 & 4 \end{pmatrix}, \quad \text{with } \gamma \leq -1.$$
>
> The core matrix $\boldsymbol{S}$ lies in the parameter spaces of TBM and our dTBM. Here, the constraint $\gamma \leq -1$ is added to ensure the bounded condition of $\boldsymbol{S}$ in our parameter space in (**??**). The angle-based and Euclidean-based signal levels of $\boldsymbol{S}$ are
>
> $$\Delta_{\mathrm{ang}}^2 (\boldsymbol{S}) = 0 \ (\leq p^\gamma), \quad \Delta_{\mathrm{Euc}}^2(\boldsymbol{S}) = 5p^{\gamma+1} \ (\geq p^\gamma).$$

We conclude that TBM with $\boldsymbol{S}$ achieves exact recovery with a polynomial-time algorithm; see Han et al. (2020, Theorem 4). By contrast, the dTBM with the same $\boldsymbol{S}$ and input $r = 2$ violets the identifiability condition, and thus fails to be solved by all estimators; see our Theorem **??**.

• *Theoretical results.* In both works, we study the phase transition of TBM and dTBM with respect to the Euclidean and angle-based SNRs. We briefly summarize the results in Han et al. (2020) and compare with ours.

Statistical critical value:

Ours: $\Delta^2_{\mathrm{ang}} \lesssim p^{-(K-1)} \Rightarrow$ statistically impossible; $\quad \Delta^2_{\mathrm{ang}} \gtrsim p^{-(K-1)} \Rightarrow$ MLE achieves exact recovery;

Han's: $\Delta^2_{\mathrm{Euc}} \lesssim p^{-(K-1)} \Rightarrow$ statistically impossible; $\quad \Delta^2_{\mathrm{Euc}} \gtrsim p^{-(K-1)} \Rightarrow$ MLE achieves exact recovery.

Computational critical value:

Ours: $\Delta^2_{\mathrm{ang}} \lesssim p^{-K/2} \Rightarrow$ computationally impossible; $\quad \Delta^2_{\mathrm{ang}} \gtrsim p^{-K/2} \Rightarrow$ computationally efficient;

Han's: $\Delta^2_{\mathrm{Euc}} \lesssim p^{-K/2} \Rightarrow$ computationally impossible; $\quad \Delta^2_{\mathrm{Euc}} \gtrsim p^{-K/2} \Rightarrow$ computationally efficient.

The above comparison reveals three major differences.

First, none of our results in Section **??** are corollaries of Han et al. (2020). Both models show the similar conclusion **but under different conditions**. While the TBM impossibility (Han et al., 2020) provides a necessary condition for our dTBM impossibility, we find that such a condition is often loose. There exists a regime of $\mathcal{S}$ in which TBM problems are computationally efficient but dTBM problems are statistically impossible; see Example 5. This observation has motivated us to develop the new signal notion Δ^2_{ang} for sharp dTBM phase transition conditions.

Second, to find the phase transition, we need to show both the impossibility **and** achievability when SNR is below **and** above the critical value, respectively. While the TBM impossibility can serve as a loose condition of our dTBM impossibility, more efforts are required to show the achievability. In particular, since TBM is a more restrictive model than dTBM, the achievability in Han et al. (2020) does not imply the achievability of dTBM in a larger parameter space. The latter requires us to develop new MLE and polynomial algorithms for dTBM achievability.

Third, from the perspective of proofs, we develop new tools to handle the extra degree heterogeneity. In our Theorem **??**, we consider the profile MLE by maximizing out the nuisance degree parameter, while TBM (Han et al., 2020) considers the usual MLE without degree parameter. In our Theorem **??**, 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.

- *Algorithms.* Both Han et al. (2020) and our work propose the two-step algorithm, which combines warm initialization and iterative refinement to achieve exact recovery. This local-to-global strategy is not new in clustering literature (Gao and Zhang, 2022; Chien et al., 2019). The highlight of our algorithm is the angle-based update in lines 10-14, Sub-algorithm 2, which is specifically designed for dTBM to avoid the estimation of $\boldsymbol{\theta}$. This angle-based update brings new proof challenges. We develop polar-coordinate based techniques to establish the error rate for the proposed algorithm. "

2. *Guarantees for the MLE and the claim that it matches the statistical lower bound. Please include an analysis of the MLE and clarify the claim.*

Response: We have added the MLE guarantee in the revised Theorem **??** in Section **??**. The MLE error rate matches the statistical lower bound. Detailed proofs have been added to Appendix II, Section **??**. Inspired by this comment, we have also revised the exposition of Theorems **??**-**??** to better present the stat-computational lower and upper bounds. We quote the revision in Section **??** here:

New Section **??**:

"... We consider the Gaussian MLE, denoted as $(\hat{z}_{\mathrm{MLE}}, \hat{\boldsymbol{\theta}}_{\mathrm{MLE}}, \hat{\mathcal{S}}_{\mathrm{MLE}})$, over the estimation space $\mathcal{P}(\gamma)$, where

$$(\hat{z}_{\mathrm{MLE}}, \hat{\boldsymbol{\theta}}_{\mathrm{MLE}}, \hat{\mathcal{S}}_{\mathrm{MLE}}) = \underset{(z, \boldsymbol{\theta}, \mathcal{S}) \in \mathcal{P}(\gamma)}{\arg\min} \|\mathcal{Y} - \mathcal{X}(z, \boldsymbol{\theta}, \mathcal{S})\|_F^2. \tag{1}$$

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 $r \lesssim p^{1/3}$. If the signal exponent satisfies $\gamma < -(K-1)$, then, no estimator $\hat{z}_{\mathrm{stat}}$ achieves exact recovery in expectation; that is,

$$\gamma < -(K-1) \quad \Rightarrow \underset{\hat{z}_{\mathrm{stat}}}{\inf} \underset{(z, \mathcal{S}, \boldsymbol{\theta}) \in \mathcal{P}(\lambda)}{\sup} \mathbb{E}\left[p\ell(\hat{z}_{\mathrm{stat}}, z)\right] \geq 1.$$

- **MLE achievability.** If the signal exponent satisfies $\gamma > -(K-1) + c$ for an arbitrary constant $c > 0$, then, the MLE in (1) achieves exact recovery in high probability; that is,

$$\gamma > -(K-1) + c \quad \Rightarrow \quad \ell(\hat{z}_{\mathrm{MLE}}, z) \lesssim \mathrm{SNR}^{-1} \exp\left(-\frac{p^{K-1}\mathrm{SNR}}{r^{K-1}}\right) \to 0,$$

with probability going to 1.

The proofs for the two parts in Theorem **??** are in the Appendix II, Section **??** and Section **??**, respectively. The first part of Theorem **??** demonstrates the impossibility of exact recovery when $\gamma < -(K-1)$. Inspection of the proof shows that the impossibility holds with or without the balanced degree assumption. 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 second part of Theorem **??** shows the exact recovery of MLE when the $\gamma > -(K-1) + c$ for an arbitrary constant $c > 0$. Combining the impossibility and achievability results,

we conclude that the boundary $\gamma_{\text{stat}} := -(K-1)$ is the critical value for statistical performance of dTBM with respect to our SNR. ”

New Section **??**:

“...The proof of MLE error shares the similar idea as Theorems **??-??**. we first show a weaker polynomial rate for MLE, and then improve the rate from polynomial to exponential throughout the iterations. The only difference is that the MLE remains the same over iterations due to its global optimality. See Appendix II, Section **??** for the detailed proof.”

New Section **??**:

“...

Theorem 3 (Computational critical value). Consider general Gaussian dTBMs under the parameter space $\mathcal{P}$ with $K \geq 2$. Then, we have the following computational phase transition.

- **Impossibility.** Assume HPC conjecture holds. If the signal exponent satisfies $\gamma < -K/2$, then, no *polynomial-time estimator* $\hat{z}_{\text{comp}}$ achieves exact recovery in expectation as $p \to \infty$; that is,

$$\gamma < -K/2 \;\Rightarrow\; \liminf_{p \to \infty} \sup_{(z, \mathcal{S}, \boldsymbol{\theta}) \in \mathcal{P}(\gamma)} \mathbb{E}\left[p\ell(\hat{z}_{\text{comp}}, z) \right] \geq 1.$$

- **Polynomial-time algorithm achievability.** Suppose the parameter space satisfies $\gamma > -K/2 + c$ for an arbitrary constant $c > 0$. Furthermore, assume the degree is lower bounded in that $\min_{i \in [p]} \theta_i \geq c$ for some constant $c > 0$, and $\boldsymbol{\theta}$ satisfies the local linear stability in Definition **??** in the neighborhood $\mathcal{N}(z, \varepsilon)$ for $\varepsilon \geq \log^{-1}(p)$. Then, there exists a polynomial-time algorithm $\hat{z}_{\text{ploy}}$ that achieves exact recovery in high probability; that is,

$$\gamma > -K/2 + c \;\Rightarrow\; \ell(\hat{z}_{\text{poly}}, z) \lesssim \text{SNR}^{-1} \exp\left(-\frac{p^{K-1} \text{SNR}}{r^{K-1}} \right) \to 0,$$

with probability going to 1.

The proofs for the two parts in Theorem **??** are in the Appendix II, Section **??** and Section **??**, respectively. The first part of Theorem **??** indicates the impossibility of exact recovery by polynomial-time algorithms when $\gamma < -K/2$, and the second part shows the existence of such algorithm when $\gamma > -K/2 + c$ under extra technical assumptions. In Section **??**, we will present an efficient polynomial-time algorithm in this setting. Therefore, we conclude that $\gamma_{\text{comp}} := -K/2$ is the critical value for computational performance of dTBM with respect to our SNR. ”

Point-by-point Response to Reviewer 1

The paper studies the problem of finding the planted block structure in a tensor, generated from the degree-corrected tensor block model (dTBM). This problem arises in different contexts, such as community detection in hypergraph, higher-order clustering etc. A simpler version of the problem (non-degree corrected TBM) has been studied in previous works. The main contribution of the paper is to extend the guarantees for recovering planted blocks in TBM to the case of dTBM. In particular, a suitable SNR is proposed and statistical lower bound (Thm 2) and computational lower bound under HPC (Thm 3) are derived. Thm 4-5 provide error rate for a poly-time algorithm.

Overall, the paper is well-written and comes with additional numerical studies, also on real data sets. The contribution of the paper is quite good. There are few aspects that are not addressed in the paper (listed below). It is to be noted that all the mentioned may not be easily resolved, but brief discussions on these aspects would improve the paper.

1. The statistical upper bound (guarantees for MLE) seem to be missing since Thm 4-5 focuses on a poly-time approach. Please explicitly describe if (and how) one can derived error rate for MLE from Thm 5.

Response: Thank you for the suggestion. We have added the MLE guarantee in the revised Theorem **??**, Section **??**. Detailed proof has been added in the Appendix II, Section **??**. The proof shares the similar idea as Theorems **??**-**??**; the only difference is that the MLE remains the same over iterations due to its global optimality. See our response to **# 2 Major Comments from Editor** for details.

2. While the dTBM model is general, it ignores the effect of sparsity in the context of hypergraphs. In case of community detection in sparse hypergraphs, $\mathbb{E}[\mathcal{Y}]$ includes a scalar that vanishes with number of nodes/tensor dimension. As in the case of graph SBM, the analysis gets a bit tricky for sparse hypergraphs (which is not covered in this paper). The authors could look at Florescu and Perkins (2016) which studies this setting and the proposed method also has similarities with the algorithm presented here.

Response: We thank reviewer for the suggestions. Our current work focuses on *dense* dTBM. While sparse dTBM is an interesting application, the theory and algorithm require different techniques. Inspired by the comment, we have extended our algorithm and experiment to sparse dTBM with a decaying sparse parameter. We adapt diagonal-deleted HOSVD as a new initialization strategy under extremely sparse settings.

We have added the following discussion in Section **??**:

> "... *Extension to sparse binary dTBMs.* The sparsity is often a popular feature in hypergraphs (Florescu and Perkins, 2016; Ke et al., 2019). Specifically, the sparse binary dTBM assumes that, the entries of $\mathcal{Y}$ follow independent Bernoulli distributions with the mean
>
> $$\mathbb{E}\mathcal{Y} = \alpha_p \mathcal{S} \times_1 \boldsymbol{\Theta M} \times_2 \cdots \times_K \boldsymbol{\Theta M}, \tag{2}$$
>
> where the extra scalar parameter $\alpha_p \in (0, 1]$ is function of p that controls the sparsity. A smaller α_p indicates a higher level of sparsity. Our current work focuses on dense dTBM with $\alpha_p = 1$. While sparse dTBM is an interesting application, the algorithm and

its analysis require different techniques. Below, we discuss possible modifications of the algorithm.

The sparsity affects our initialization guarantee in our Theorem **??**. In our initialization, the spectral denoising step (lines 1-2 in Sub-algorithm 1) implements matrix SVD to unfolded tensors. However, SVD-based methods are believed to fail in extremely sparse SBM due to the localization phenomenon in the singular vectors (Florescu and Perkins, 2016). Inspired by Florescu and Perkins (2016), we adapt the diagonal-deleted HOSVD (D-HOSVD) (Ke et al., 2019) as the initialization in our higher-order clustering.

The sparsity also affects the iteration guarantee in our Theorem **??**. The decaying mean tensor leads to a worse statistical error of order $\mathcal{O}(-\alpha_p p^{K-1})$ on $\hat{\mathcal{X}}$. The theoretical analyses for sparse binary dTBM and algorithms are left as future direction. Instead, we add numerical experiments to evaluate the robustness of our algorithm and the improvement of D-HOSVD initialization in the sparse dTBM; see Appendix I."

We have added a numerical experiment and Figure 1 under the sparse binary dTBM in Appendix I:
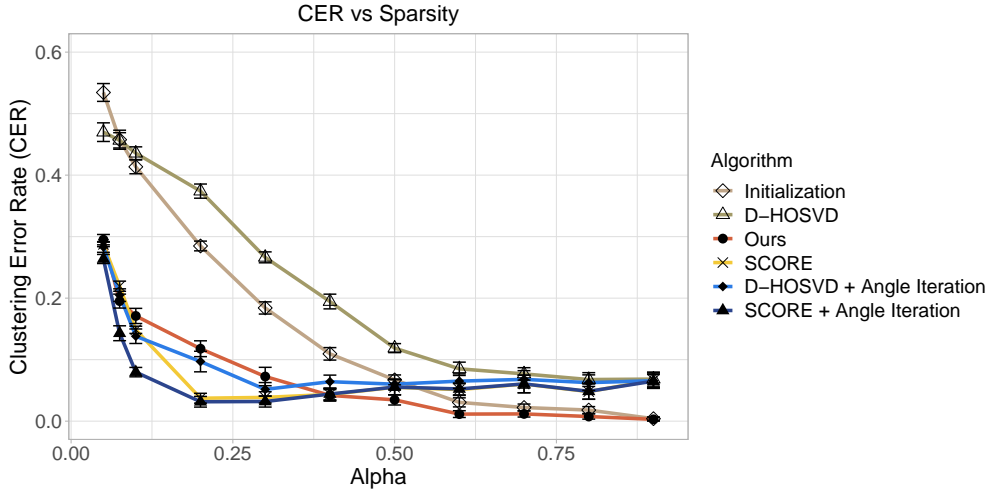

Figure 1: CER comparison versus sparsity parameter α_p in $[0.05, 0.9]$. We set $p = 100, r = 5$, and $\gamma = -1.2$ under sparse binary dTBM.

" ...The second additional experiment evaluates the algorithm performances under the sparse binary dTBM (2). We fix the signal exponent $\gamma = -1.2$ and vary the sparsity parameter $\alpha_p \in [0.05, 0.9]$. A smaller α_p leads to a higher probability of zero entries in the observation. In addition to the three algorithms mentioned in Section **??** (denoted **Initialization**, **dTBM**, and **SCORE**), we consider other three algorithms based on the discussion in Section **??**:

- **D-HOSVD**, the diagonal-deleted HOSVD in Ke et al. (2019);

- **D-HOSVD + Angle**, the combined algorithm of our angle-based iteration with initialization from **D-HOSVD**;

- **SCORE + Angle**, the combined algorithms of our angle-based iteration with initial-

6

ization from **SCORE**.

Figure 1 shows a slightly larger error in **dTBM** than that in **SCORE**, **D-HOSVD + Angle**, and **SCORE + Angle** under the sparse setting with $\alpha_p < 0.3$. The small gap between **dTBM** and other sparse-specific methods implies the robustness of our algorithm. In addition, comparing **SCORE** versus **SCORE + Angle** (or **D-HOSVD** versus **D-HOSVD + Angle**) indicates the benefit of our angle iterations under the sparse dTBM. In the intermediate and dense cases with $\alpha_p \geq 0.3$, our proposed **dTBM** has a clear improvement over others, which again verifies the success of our algorithm in dense settings."

3. *(Critical) The statistical lower bound is derived under non-degree TBM (since all $\theta(i) = 1$). From this, it is still not clear whether the lower bound would significantly increase when there is a skewed degree correction. The authors' claim that the statistical lower bound matches MLE is difficult to verify (since I could not find an explicit result for MLE). Moreover, the stated upper bound (Thm 4-5) assume some uniformity in θ (see top of page-11), and so the (matched) upper/lower bounds could only be for the case of relatively homogeneous degrees. A more clear discussion on this would be useful. The present results leaves open the possibility that phase-transition SNR for dTBM could be arbitrarily worse than that of TBM, in which case making some assumption on balance is reasonable.*

Response:

We thank reviewer for the suggestions. There seem to be three questions in this comment. We address them one by one.

(a) Regarding "whether the lower bound would significantly increase when there is a skewed degree correction". We have revised the proof of statistical minimax bound with relaxed degree assumptions. Briefly, we relax the previous non-degree construction $\theta(i) \equiv 1$ to the skewed-degree construction under assumption (3). The latter allows moderate degree heterogeneity; for example, within each of the cluster, we allow the highest degree to be $\theta(i) = \Omega(p)$, whereas the lowest degree to be $\theta(i) = \mathcal{O}(1)$. The revised minimax bound does not increase in the presence of moderately skewed degree. See our new Theorem **??** in Section **??** and new proof in Appendix II, Section **??**.

(b) Regarding "explicit result for MLE". We have added the MLE guarantee in new Theorem **??**. The MLE error matches with statistical lower bound and ensures the dTBM achievability above the statistical critical value. We have also revised the exposure of Theorems **??**-**??** to include both impossibility and achievability results. See our new Section **??** and Section **??**. See our response to **# 2 Major Comments from Editor** for details.

(c) Regarding "The stated upper bound assume some uniformity in θ". We have added a new Section **??** to discuss the rational for the balanced degree assumption (3). Briefly, we focus on the regime with balanced heterogeneity due to the relationship between dTBM performance and core tensor signal $\Delta_{\min}^2$. Indeed, without the balance assumption, the phase transition of dTBM can be arbitrarily worse than that of TBM. We quote our new discussion below:

New Section **??**:

"...In our work, we focus on the regime of balanced degree heterogeneity. We call the

degree $\boldsymbol{\theta}$ *balanced* if

$$\min_{a\in[r]}\|\boldsymbol{\theta}_{z^{-1}(a)}\| = (1+o(1))\max_{a\in[r]}\|\boldsymbol{\theta}_{z^{-1}(a)}\|. \tag{3}$$

The following lemma provides the rational of balanced degree assumption. We show the close relation between angle gaps in the mean tensor $\mathcal{X}$ and the core tensor $\mathcal{S}$ under balanced degree heterogeneity.

Lemma 1 (Angle gaps in $\mathcal{X}$ and $\mathcal{S}$). Consider the dTBM model (**??**) under the parameter space $\mathcal{P}$ in (**??**). Suppose Assumption **??** holds and $\boldsymbol{\theta}$ is balanced satisfying (3). Then, for all i, j such that $z(i) \neq z(j)$, we have

$$\cos(\boldsymbol{X}_{i:}, \boldsymbol{X}_{j:}) \asymp \cos(\boldsymbol{S}_{z(i):}, \boldsymbol{S}_{z(j):}),$$

where $\boldsymbol{X} = \mathrm{Mat}(\mathcal{X})$ and $\boldsymbol{S} = \mathrm{Mat}(\mathcal{S})$.

In practice, an estimation algorithm has access to a noisy version of $\mathcal{X}$ but not $\mathcal{S}$. Our goal is to establish the algorithm performance with respect to the signal $\Delta_{\min}^2$ in the core tensor. By Lemma 1, the mapping from the core tensor $\boldsymbol{S}_{z(i):}$ to the mean tensor $\boldsymbol{X}_{z(i):}$ preserves the angle information $\Delta_{\min}^2$ under balanced degree heterogeneity (3). Therefore, the balanced degree assumption helps to exclude the cases in which the degree heterogeneity distorts the algorithm garantees.

Here, we provide an example to illustrate the insufficiency of $\Delta_{\min}^2$ in the absence of balanced degrees.

Example 2 (Insufficiency of $\Delta_{\min}^2$ in the absence of balanced degrees). Consider an order-2 (p, p)-dimensional dTBM with core matrix

$$\boldsymbol{S} = \begin{pmatrix} 1 & a \\ 1 & -a \end{pmatrix}, \quad \text{and } \boldsymbol{\theta} \text{ such that } \|\boldsymbol{\theta}_{z^{-1}(1)}\|^2 = p^m \|\boldsymbol{\theta}_{z^{-1}(2)}\|^2, \tag{4}$$

where $m \geq 0$ is a scalar parameter controlling the skewness of degrees. Let $\Delta_{\boldsymbol{X}}^2$ denote the angle gap of the mean tensor, defined by

$$\Delta_{\boldsymbol{X}}^2 := \min_{i,j\in[p], z(i)\neq z(j)} \left\| \frac{\boldsymbol{X}_{i:}}{\|\boldsymbol{X}_{i:}\|} - \frac{\boldsymbol{X}_{j:}}{\|\boldsymbol{X}_{j:}\|} \right\|, \quad \boldsymbol{X} = \mathrm{Mat}(\mathcal{X}).$$

Take $a = p^{-1/4}$ in the model setup (4). We have

$$\Delta_{\min}^2 = \frac{2a^2}{1+a^2} \asymp p^{-1/2}, \quad \Delta_{\boldsymbol{X}}^2 = \frac{2\|\boldsymbol{\theta}_{z^{-1}(2)}\|^2 a^2}{\|\boldsymbol{\theta}_{z^{-1}(1)}\|^2 + \|\boldsymbol{\theta}_{z^{-1}(2)}\|^2 a^2} \asymp p^{-1/2-m}.$$

Based on our theory in later Sections, the stat-computational achievability of dTBM occurs when $\Delta_{\boldsymbol{X}}^2 > p^{-1}$; that is, the dTBM estimation depends on the relative magnitude of m vs. $1/2$. In such a setting, the proposed signal notion $\Delta_{\min}^2$ alone fails to fully characterize dTBM.

Remark 1 (Flexibility in balanced degree assumption). One important note is that our balance assumption (3) does not preclude the mild degree heterogeneity. In fact, within each of the clusters, we allow the highest degree at the order $\mathcal{O}(p)$, whereas the lowest degree at the order $\Omega(1)$. This range is more relaxed than previous work (Gao et al., 2018) that restricts the highest degree in the sub-linear regime $o(p)$ and the lowest degree at the order $\Omega(1)$.

Remark 2 (Similar assumptions in literature). Similar degree regulations are not rare in literature. In higher-order tensor model (Ke et al., 2019), the degree assumption $\max_{a \in [r]} \|\boldsymbol{\theta}_{z^{-1}(a)}\| \leq C \min_{a \in [r]} \|\boldsymbol{\theta}_{z^{-1}(a)}\|$ is made to ensure balancedness across communities. In Gao et al. (2018), the degree distribution is restricted to $\frac{1}{|z^{-1}(a)|} \sum_{i \in z^{-1}(a)} \theta_i = 1 + o(1)$ for all communities.

4. *Please provide some intuition on Definition 2, particularly Eqn (10).*

Response: We have added more explanations for Definition ?? and equation (??) in Section ??:

"... Roughly speaking, the vector $\boldsymbol{p}(\bar{z})$ represents the raw cluster sizes, and $\boldsymbol{p}_\theta(\bar{z})$ represents the relative cluster sizes weighted by degrees. ...The condition (??) controls the impact of node degree to the $\boldsymbol{p}_\theta(\cdot)$ with respect to the misclassification rate ε and angle gap $\Delta_{\min}$. Intuitively, the condition (??) controls the skewness of degree so that the angle between raw cluster size and degree-weighted cluster size is well controlled. The stability assumption is proposed for technical convenience, and we relax this condition in numerical studies; see Section ??."

Point-by-point Response to Reviewer 2

Comments to the Author: This manuscript studies clustering under degree-corrected tensor block models (dTBMs). The dTBM is a generalization of the degree-corrected stochastic block model (DCSBM) for matrices/networks. It is also a generalization of the (non-degree) tensor block model (TBM) which is a higher-order extension of the stochastic block model (SBM).

Contributions: The identifiability for the uniqueness of clustering under dTBMs is studied. Statistical and computational lower bounds for clustering under dTBMs are established. A two-stage, angle-based algorithm is proposed to achieve exact clustering in polynomial time under mild conditions. The last contribution is the most important one. The algorithm proposed uses the "global-to-local strategy" which is commonly seen in community detection literature. An angle-based clustering is developed to handle the degree heterogeneity of dTBMs. Then theoretical analysis shows that it provably achieves the exact clustering. I feel this is a solid contribution, making the manuscript worth publication.

Nevertheless, I have a few concerns and comments.

1. One concern I have is that this manuscript is very similar to one of the authors' previous paper Han et al. (2020) on clustering under the non-degree tensor block model. They share overall the same spirit, taste, and structure. For example: The sessions for the computational and statistical lower bounds in this manuscript are very similar to the corresponding sessions in Han et al. (2020) and they have the same results. Han et al. (2020) also gives a two-stage algorithm using the "global-to-local strategy" and also achieves the exact clustering. Han et al. (2020) gives the same threshold for the exact clustering as in this manuscript. I think this manuscript should point out its connection with Han et al. (2020). The current version mentions that the method of Han et al. (2020) fails under dTBMs but I think Han et al. (2020) deserves more discussion given the similarity.

Response: Thank you for your suggestion. We intentionally make the paper structure similar to Han et al. (2020) in order for easier comparison with non-degree models. Our results (e.g. the stat-comp lower bounds, the threshold for exact clustering) are all different from Han et al. (2020). The confusion may come from the choice of our exposition: we chose the same notations as in Han et al. (2020) but actually they represent different quantities.

We have added a new Section **??**, "Comparison with Non-degree Tensor Block Model", to discuss the connection and difference between our dTBM work and TBM in Han et al. (2020) from three aspects: signal notions, theoretical results, and algorithms. Please see our response to **# 1, Major Comments from Editor** for details.

Briefly, we separate the discussion from three aspects: signal notion, theoretical results, and algorithm. We develop an angle-based signal notion to address the degree heterogeneity in dTBM. Our theoretical results are different with, and not easy extension of, the result in Han et al. (2020). Our algorithm share the classical local-to-global idea, but our angle-based iteration is specifically designed for dTBM and requires new proof techniques.

2. My next comment is about whether Session III B (computational lower bound) can be much simplified. My understanding is that the non-degree tensor block model is a special case of dTBMs, and hence lower bounds givens in Han et al. (2020) also hold under dTBMs. As a result, I also feel the 2nd contribution this manuscript claim (ie, statistical and computational lower bounds) are

relatively minor.

Response: Thank you for your suggestion. The Session III B (computational lower bound) cannot be derived from earlier non-degree work (Han et al., 2020). The short answer is that, our results show the similar conclusion but **under different conditions.** While the TBM impossibility (Han et al., 2020) provides a necessary condition for our dTBM impossibility, we find that **such a condition is loose** in our contexts. This has motivated to develop new techniques to derive sharp lower bounds.

Inspired by your comment, we have revised the exposition of our theory. See our response to **# 1 Major Comments from Editor** for details.

We quote the newly-added texts here:

"...*Theoretical results.* In both works, we study the phase transition of TBM and dTBM with respect to the Euclidean and angle-based SNRs. We briefly summarize the results in Han et al. (2020) and compare with ours.

Statistical critical value:

Ours: $\Delta^2_{\text{ang}} \lesssim p^{-(K-1)} \Rightarrow$ statistically impossible; $\quad \Delta^2_{\text{ang}} \gtrsim p^{-(K-1)} \Rightarrow$ MLE achieves exact recovery;

Han's: $\Delta^2_{\text{Euc}} \lesssim p^{-(K-1)} \Rightarrow$ statistically impossible; $\quad \Delta^2_{\text{Euc}} \gtrsim p^{-(K-1)} \Rightarrow$ MLE achieves exact recovery.

Computational critical value:

Ours: $\Delta^2_{\text{ang}} \lesssim p^{-K/2} \Rightarrow$ computationally impossible; $\quad \Delta^2_{\text{ang}} \gtrsim p^{-K/2} \Rightarrow$ computationally efficient;

Han's: $\Delta^2_{\text{Euc}} \lesssim p^{-K/2} \Rightarrow$ computationally impossible; $\quad \Delta^2_{\text{Euc}} \gtrsim p^{-K/2} \Rightarrow$ computationally efficient.

The above comparison reveals three major differences.

First, none of our results in Section **??** are corollaries of Han et al. (2020). Both models show the similar conclusion **but under different conditions**. While the TBM impossibility (Han et al., 2020) provides a necessary condition for our dTBM impossibility, we find that such a condition is often loose. There exists a regime of $\mathcal{S}$ in which TBM problems are computationally efficient but dTBM problems are statistically impossible; see Example 5. This observation has motivated us to develop the new signal notion Δ^2_{ang} for sharp dTBM phase transition conditions.

Second, to find the phase transition, we need to show both the impossibility **and** achievability when SNR is below **and** above the critical value, respectively. While the TBM impossibility can serve as a loose condition of our dTBM impossibility, more efforts are required to show the achievability. In particular, since TBM is a more restrictive model than dTBM, the achievability in Han et al. (2020) does not imply the achievability of dTBM in a larger parameter space. The latter requires us to develop new MLE and polynomial algorithms for dTBM achievability.

Third, from the perspective of proofs, we develop new tools to handle the extra degree heterogeneity. In our Theorem **??**, we consider the profile MLE by maximizing out the nuisance degree parameter, while TBM (Han et al., 2020) considers the usual MLE without degree parameter. In our Theorem **??**, we construct a rank-2 tensor to relate HPC conjecture to Δ_{ang}^2, while TBM (Han et al., 2020) constructs a rank-1 tensor to relate HPC conjecture to Δ_{Euc}^2. The asymptotic non-equivalence between Δ_{ang}^2 and Δ_{Euc}^2 renders our proof technically more involved..."

"... *Signal Notion.* We emphasize that the signal notions are different between the two models. In particular, the Euclidean-based signal notion in TBM Han et al. (2020) fails to accurately describe the phase transition in our dTBM due to the possible heterogeneity in degree $\boldsymbol{\theta}$. To compare, we denote our angle-based signal notion in (**??**) and the Euclidean-based SNR in Han et al. (2020) as Δ_{ang}^2 and Δ_{Euc}^2, respectively:

$$\Delta_{\mathrm{ang}}^2 = 2(1 - \max_{a \neq b \in [r]} \cos(\boldsymbol{S}_{a:}, \ \boldsymbol{S}_{b:})), \quad \Delta_{\mathrm{Euc}}^2 = \min_{a \neq b \in [r]} \|\boldsymbol{S}_{a:} - \boldsymbol{S}_{b:}\|^2.$$

By Lemma **??** in the Appendix II, we have

$$\Delta_{\mathrm{ang}}^2 \max_{a \in [r]} \|\boldsymbol{S}_{a:}\|^2 \leq \Delta_{\mathrm{Euc}}^2.$$

The above inequality indicates the condition $\Delta_{\mathrm{Euc}}^2 \leq p^\gamma$ is **sufficient but not necessary** for $\Delta_{\mathrm{ang}}^2 \leq p^\gamma$. In fact, if we were to use Δ_{Euc}^2 for both models, then the phase transition of dTBM can be arbitrarily worse than that for TBM.

Here, we provide an example to illustrate the dramatical difference between TBM and dTBM with the same core tensor.

Example 3 (Comparison with Euclidean-based signal notion). Consider a biclustering model with $\boldsymbol{\theta} = 1$ and an order-2 core matrix

$$\boldsymbol{S} = \begin{pmatrix} p^{(\gamma+1)/2} + 2 & 2p^{(\gamma+1)/2} + 4 \\ 2 & 4 \end{pmatrix}, \quad \text{with } \gamma \leq -1.$$

The core matrix $\boldsymbol{S}$ lies in the parameter spaces of TBM and our dTBM. Here, the constraint $\gamma \leq -1$ is added to ensure the bounded condition of $\boldsymbol{S}$ in our parameter space in (**??**). The angle-based and Euclidean-based signal levels of $\boldsymbol{S}$ are

$$\Delta_{\mathrm{ang}}^2(\boldsymbol{S}) = 0 \ (\leq p^\gamma), \quad \Delta_{\mathrm{Euc}}^2(\boldsymbol{S}) = 5p^{\gamma+1} \ (\geq p^\gamma).$$

We conclude that TBM with $\boldsymbol{S}$ achieves exact recovery with a polynomial-time algorithm; see Han et al. (2020, Theorem 4). By contrast, the dTBM with the same $\boldsymbol{S}$ and input $r = 2$ violates the identifiability condition, and thus fails to be solved by all estimators; see our Theorem **??**."

We have revised our Section **??** to better present the lower bounds. See our response to **# 1 Major Comments from Editor**.

3. *My third comment is on Condition (7), i.e., "balance" of θ. This is quite artificial and impractical and is needed only for technical reasons in the theoretical analysis of the proposed algorithm. In*

fact, as far as I know, this (or similar condition) is not needed in DCSBM literature for community detection. I suspect (7) can be removed by some better proof technique or modification of the proposed method.

Response: Thank you for your suggestion. See our response to **#3(a) and 3(c), Reviewer 1** (page 7 of this response letter) for the detailed response.

While we agree the assumption is more for technical reasons, we suspect some "balanced" assumption or its variation seems necessary (and common in literature) for the achievability of computational critical values. Our constraints in $\mathcal{P}$ are mild compared with previous literature. We have also added following comparison in Section **??**:

Assumptions in parameter space	Gao et al. (2018)	Han et al. (2020)	Ke et al. (2019)	Ours
Balanced community size	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Balanced degree assumption	$\checkmark$	-	$\checkmark$	$\checkmark$
Bounded core tensor assumption	$\checkmark$	$\times$	$\checkmark$	$\checkmark$
Flexible in-group connections	$\times$	$\checkmark$	$\checkmark$	$\checkmark$
Gap among cluster centers	In-between cluster difference	Euclidean gap	Eigen gap	Angle gap

Table 1: Parameter space comparison between previous work with our assumption.

Point-by-point Response to Reviewer 3

In general, this paper is well written and organized, numeric studies including simulation and real data application illustrate that the proposed model and method do have some merits. But I have several concerns regarding the theory and technical contribution.

1. Firstly, the mathematical formulation lacks motivation. From my knowledge, most existing degree-corrected block models are proposed for network data and Bernoulli model. While the proposed paper can be applied to Bernoulli as it could be taken as a special sub-Gaussian but I believe such treatment would not be suitable and/or optimal if one only cares about Bernoulli setting. Also, I notice several formulation inconsistencies. For example, the author said they focus on symmetric tensor, but they assume the noise tensor $\mathcal{E}$ consists of independent entries, which does not make sense.

Response: Thank you for your suggestions. There seem to be three questions in this comment; we address them one-by-one.

(a) Regarding "motivation of general sub-Gaussian model". Our proposal of sub-Gaussian dTBM is intended to cover multiple types of data in applications. Continuous tensors arise commonly in fields such as genetics and medical image studies. From theoretical perspective, our sub-Gaussian framework allows both binary and continuous observations. Conversely, binary-specific methods may not be suitable for general (continuous) sub-Gaussian data.

(b) Regarding "sub-optimality of Bernoulli as a special sub-Gaussian model". We discuss below which of our results are still optimal for Bernoulli data, and which are not.

- **sub-Algorithm I.** Our initialization gaurantee is not suitable for Bernoulli data. To address this, we have proposed a remedy for Bernoulli initialization with square unfolding; see Section **??**. The Bernoulli initialization error rate $\mathcal{O}(p^{-\lfloor K/2 \rfloor})$ is slightly weaker than Gaussian initialization error rate $\mathcal{O}(p^{-K/2})$. Combining the Bernoulli initialization with sub-Algorithm II, our method achieves an exponential error rate for Bernoulli data.

- **sub-Algorithm II.** Our angle-based iteration guarantee is suitable for Bernoulli data. With Bernoulli data and a proper initialization, our iteration reduces the misclustering error from a polynomial rate to an exponential rate $\mathcal{O}(\exp(-p^{K-1}))$, while the state-of-art algorithm for binary hypergraph (Ke et al., 2019) leads to a polynomial error rate $\mathcal{O}(p^{-2})$.

- **Theorems ??-??.** The phase transitions are not applicable for Bernoulli data. The analyses of statistical phase transition and computational impossibility are based on Gaussian noise, and the polynomial-time algorithm achievability requires a larger SNR of order $\mathcal{O}(p^{-\lfloor K/2 \rfloor})$ for Bernoulli data. We have added a numerical experiment in Appendix I to verify the existence of statistical-computational gap in Bernoulli tensor clustering.

New Section **??**:

"...

Initialization guarantee for Bernoulli data. The main difficulty to establish the initialization guarantee for Bernoulli observations lies in the denoising step (lines 1-2 in Sub-algorithm 1). We now provide a high-level explanation for the technical difficulty when applying Theorem **??** to Bernoulli observations.

The derivation of Theorem **??** relies on the upper bound of the estimation error for the mean tensor in Lemma **??**; i.e., with high probability

$$\|\hat{\mathcal{X}} - \mathcal{X}\|_F^2 \lesssim p^{K/2}, \tag{5}$$

where $\mathcal{X} = \mathbb{E}\mathcal{Y}$ and $\hat{\mathcal{X}}$ is defined in Step 2 of Sub-algorithm 1. Unfortunately, the inequality (5) holds only for i.i.d. sub-Gaussian observations, while Bernoulli observations are generally not identically distributed.

One possible remedy is to apply singular value decomposition to the *square unfolding* (Mu et al., 2014) of Bernoulli tensor $\mathcal{Y}$. We summarize the procedure in sub-algorithm 3.

Sub-algorithm 3: Weighted higher-order initialization for Bernoulli observation

---

Input: Bernoulli tensor $\mathcal{Y} \in \{0,1\}^{p \times \cdots \times p}$, cluster number r, relaxation factor $\eta > 1$ in k-means clustering.
1: Let the matrix $\mathrm{Mat}_{sq}(\mathcal{Y}) \in \{0,1\}^{\lfloor p^{K/2} \rfloor \times \lceil p^{K/2} \rceil}$ denote the nearly square unfolded tensor. Compute the estimate $\mathcal{X}'$, where

$$\hat{\mathcal{X}}' = \underset{\mathrm{rank}(\mathrm{Mat}_{sq}(\mathcal{X})) \leq r^{\lceil K/2 \rceil}}{\arg\min} \|\mathrm{Mat}_{sq}(\mathcal{X}) - \mathrm{Mat}_{sq}(\mathcal{Y})\|_F^2. \tag{6}$$

2: Implement lines 3-5 of Sub-algorithm 1 with $\hat{\mathcal{X}}$ replaced by $\hat{\mathcal{X}}'$ in (6).
Output: Initial clustering $z^{(0)} \leftarrow \hat{z}$.

Proposition 1 (Error for Bernoulli initialization)**.** Consider the Bernoulli dTBM in the parameter space $\mathcal{P}$ and Assumption **??** 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. Following Lemma 7 in Gao et al. (2018), with high probability, the new estimate satisfies
$$\|\hat{\mathcal{X}}' - \mathcal{X}\|_F^2 \lesssim p^{\lceil K/2 \rceil}.$$
We obtain the upper bound by using $\hat{\mathcal{X}}'$ in place of $\hat{\mathcal{X}}$ in Theorem **??**. $\qquad\square$

Remark 3 (Comparison with previous methods)**.** The Bernoulli bound $\mathcal{O}(p^{-\lfloor K/2 \rfloor})$ in Proposition 1 is relatively looser than the Gaussian bound $\mathrm{O}(p^{-K/2})$ in Theorem **??**. The gap between Bernoulli and Gaussian error decreases as the order K increases. Nevertheless, combining with angle iteration Sub-algorithm 2, Bernoulli clustering still achieves exponential error rate $\exp\left(-p^{(K-1)}\right)$ at a price of a larger SNR. The exponential rate is tighter than the rate error $\mathcal{O}(p^{-2})$ in previous work (Ke et al., 2019). The investigation of the gap between upper bound $p^{-\lfloor K/2 \rfloor}$ and the lower bound $p^{-K/2}$ for Bernoulli tensors will be left as future work. In numerical experiments, we will use our original initialization, Sub-algorithm 1, to verify the robustness to Bernoulli observations."

New Appendix I:

" **Bernoulli phase transition.** The first additional experiment verifies the statistical-computational gap in Section **??** under the Bernoulli model. Consider the Bernoulli model with $p = \{80, 100\}$, $r = 5$. We vary γ in $[-1.2, -0.4]$ and $[-2.1, -1.4]$ for matrix ($K = 2$) and tensor ($K = 3$) clustering, respectively. We approximate MLE using an oracle estimator, i.e., the output of Sub-algorithm 2 initialized from the true assignment. Figure 2 shows a similar pattern as Figure **??**. The algorithm and oracle estimators have no gap in the matrix case, while an error gap emerges between the critical values $\gamma_{\mathrm{stat}} = -2$ and $\gamma_{\mathrm{comp}} = -1.5$ in the tensor case. Figure **??** suggests the statistical-computational gap in Bernoulli models."
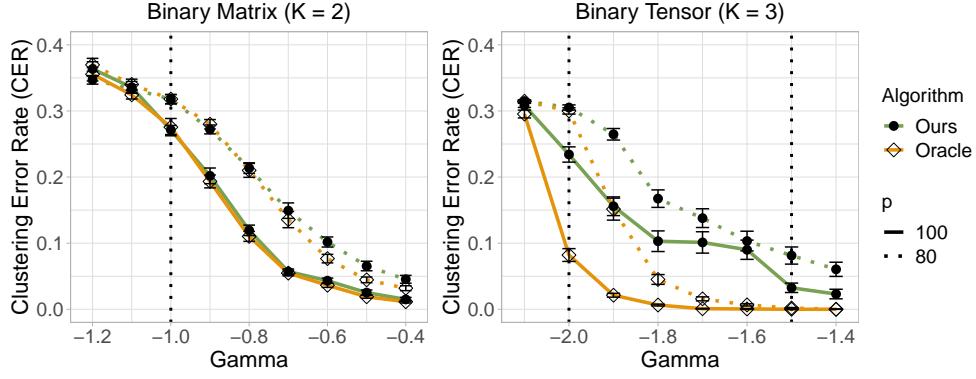

Figure 2: SNR phase transitions for Bernoulli dTBM with $p = \{80, 100\}, r = 5$ under (a) matrix case with $\gamma \in [-1.2, -0.4]$ and (b) tensor case with $\gamma \in [-2.1, -1.4]$.

(c) Regarding "Formulation inconsistency". We agree that the original exposition is confusing and also unnecessary. We were attempting to simplify the notation by focusing on symmetric signal tensors (not necessarily symmetric data). Based on your comment, the revised paragraph in Section **??** now reads

"... For ease of notation, we focus on the case with symmetric mean tensor $\mathbb{E}\mathcal{Y}$. This assumption simplifies the notation because all modes have the same (Θ, M, z); the noise tensor $\mathcal{E}$ and the data tensor $\mathcal{Y}$ are still possibly asymmetric. In general, we allow asymmetric mean tensor with $\{(\Theta_k, M_k, z_k)\}_{k=1}^K$, one for each mode. The extension can be found in Appendix II...."

In Section **??**:

"...we provide...for general Gaussian dTBMs (**??**) without symmetric assumptions. For general (asymmetric) Gaussian dTBMs, ...we extend the parameter space (**??**) to allow K clustering functions $(z_k)_{k \in [K]}$, one for each mode. For notational simplicity, we still use z and $\mathcal{P}(\gamma)$ for this general (asymmetric) model. All results should be interpreted as the worst-case results across K modes...."

Consequently, we have revised our algorithm to reflect the asymmetry of data. We quote the revised algorithm in the last page of the response letter. We have also revised the notation

in Theorems **??-??** by adding subscripts; that is, we now use $z_k^{(0)}$, $z_k^{(t+1)}$, ... in place of $z^{(0)}$, $z^{(t+1)}$, ..., throughout the paper.

2. *I also found some technical assumptions strong. For example, in Eq. (3), the authors restricted the matricization of core tensor has norm bounded from below and above. I don't see why that is necessary. The author argues "...so the core tensor $\mathcal{S}$ is not trivially reduced to a lower rank", but the tensor block model should be identifiable even when $\mathcal{S}$ has degenerated rank. This leads to another question on the defined SNR in Eq. (4) (which requires minimal $\mathcal{S}$ sliced norm). It is doubtful that this is a true quantity to work on.*

Response: This comment concerns the lower and upper bounds for the core tensor. We address them separately.

(a) Lower bound. Regarding "...but tensor block model should be identifiable even when $\mathcal{S}$ has degenerated rank." This is in fact what we intended to convey. Indeed, our remark of Theorem **??** reads: "...the identifiability guarantee for the dTBM is stronger than classical Tucker model. ." To avoid the confusion, we have revised the sentence into

> "...Third, the lower bound constant $c_3 > 0$ requires no purely zero slide in $\mathcal{S}$, and thus the model (**??**) is not trivially non-identifiable; see Example **??**. ...

The lower boundness of core tensor slices is **necessary** for the model identifiability. We have added following example in Section **??** to illustrate the necessity of the lower bound c_3.

> " ...
>
> **Example 4** (Non-identifiability with purely zero core slice)**.** Consider an order-2 dTBM with core tensor $\boldsymbol{S} = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}$ and mean tensor
>
> $$\mathcal{X} = \boldsymbol{\Theta}_1 \boldsymbol{M} \boldsymbol{S} \boldsymbol{M}^T \boldsymbol{\Theta}_2, \quad \text{with } \boldsymbol{M} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2 = \mathrm{diag}(1, 1, 1, 1).$$
>
> Replacing $\boldsymbol{\Theta}_1$ by $\boldsymbol{\Theta}_1' = (3/2, 1/2, 1, 1)$ leads to the same mean tensor $\mathcal{X}$. "

We note that the purely zero slice in $\mathcal{S}$ always makes the degree heterogeneity indistinguishable. Therefore, we exclude such case for model identifiability.

(b) Upper bound. The upper bound $c_4 > 0$ is a technical constraint to avoid unbounded entries in the core tensor; in practices, the constraint $\max_{a \in [r]} \|\mathrm{Mat}(\mathcal{S})_{a:}\| \leq c_4$ would likely never be active with a large choice of $c_4 \geq \|\mathcal{Y}\|$. We have changed the wording in Section **??** into

> "... The upper bound $c_4 > 0$ is a technical constraint to avoid unbounded entries in the core tensor. We introduce this assumption for proof convenience; in practical algorithms, the constraint $\max_{a \in [r]} \|\mathrm{Mat}(\mathcal{S})_{a:}\| \leq c_4$ would likely never be active with a large choice of $c_4 \geq \|\mathcal{Y}\|_F$. "

(c) Comparison of assumptions with previous work. Our constraints in $\mathcal{P}$ are mild compared with previous literature. We have also added following comparison in Section **??**:

Assumptions in parameter space	Gao et al. (2018)	Han et al. (2020)	Ke et al. (2019)	Ours
Balanced community sizes	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Bounded core tensors	$\checkmark$	$\times$	$\checkmark$	$\checkmark$
Balanced degrees	$\checkmark$	-	$\checkmark$	$\checkmark$
Flexible in-group connections	$\times$	$\checkmark$	$\checkmark$	$\checkmark$
Gaps among cluster centers	In-between cluster difference	Euclidean gap	Eigen gap	Angle gap

Table 2: Parameter space comparison between previous work with our assumption.

3. *Another big concern I have is the whole Section 3. The statistical and computational lower bounds are exactly the same as the tensor block model in Han et al. (2020). Therefore, the whole results in Section 3 can be simply obtained as corollary from Han et al. (2020). since the authors are working on a more restrictive model and the lower bounds must be larger than simpler model.*

Response:

In fact, none of our results in Section 3 can be derived from earlier non-degree work (Han et al., 2020). The short answer is that, our results show the similar conclusion but **under different conditions.** While the TBM impossibility (Han et al., 2020) provides a necessary condition for our dTBM impossibility, we find that **such a condition is loose** in our contexts. This has motivated to develop new techniques to derive sharp lower bounds. Inspired by your comment, we have revised the exposition of our theory. See our response to **# 1 Major Comments from Editor** for details.

We quote the newly-added texts here:

"...*Theoretical results.* In both works, we study the phase transition of TBM and dTBM with respect to the Euclidean and angle-based SNRs. We briefly summarize the results in Han et al. (2020) and compare with ours.

Statistical critical value:

Ours: $\Delta^2_{\text{ang}} \lesssim p^{-(K-1)} \Rightarrow$ statistically impossible; $\quad \Delta^2_{\text{ang}} \gtrsim p^{-(K-1)} \Rightarrow$ MLE achieves exact recovery;

Han's: $\Delta^2_{\text{Euc}} \lesssim p^{-(K-1)} \Rightarrow$ statistically impossible; $\quad \Delta^2_{\text{Euc}} \gtrsim p^{-(K-1)} \Rightarrow$ MLE achieves exact recovery.

Computational critical value:

Ours: $\Delta^2_{\text{ang}} \lesssim p^{-K/2} \Rightarrow$ computationally impossible; $\quad \Delta^2_{\text{ang}} \gtrsim p^{-K/2} \Rightarrow$ computationally efficient;

Han's: $\Delta^2_{\text{Euc}} \lesssim p^{-K/2} \Rightarrow$ computationally impossible; $\quad \Delta^2_{\text{Euc}} \gtrsim p^{-K/2} \Rightarrow$ computationally efficient.

The above comparison reveals three major differences.

First, none of our results in Section **??** are corollaries of Han et al. (2020). Both models show the similar conclusion **but under different conditions**. While the TBM

impossibility (Han et al., 2020) provides a necessary condition for our dTBM impossibility, we find that such a condition is often loose. There exists a regime of $\mathcal{S}$ in which TBM problems are computationally efficient but dTBM problems are statistically impossible; see Example 5. This observation has motivated us to develop the new signal notion Δ_{ang}^2 for sharp dTBM phase transition conditions.

Second, to find the phase transition, we need to show both the impossibility **and** achievability when SNR is below **and** above the critical value, respectively. While the TBM impossibility can serve as a loose condition of our dTBM impossibility, more efforts are required to show the achievability. In particular, since TBM is a more restrictive model than dTBM, the achievability in Han et al. (2020) does not imply the achievability of dTBM in a larger parameter space. The latter requires us to develop new MLE and polynomial algorithms for dTBM achievability.

Third, from the perspective of proofs, we develop new tools to handle the extra degree heterogeneity. In our Theorem **??**, we consider the profile MLE by maximizing out the nuisance degree parameter, while TBM (Han et al., 2020) considers the usual MLE without degree parameter. In our Theorem **??**, we construct a rank-2 tensor to relate HPC conjecture to Δ_{ang}^2, while TBM (Han et al., 2020) constructs a rank-1 tensor to relate HPC conjecture to Δ_{Euc}^2. The asymptotic non-equivalence between Δ_{ang}^2 and Δ_{Euc}^2 renders our proof technically more involved..."

"... *Signal notion.* We emphasize that the signal notions are different between the two models. In particular, the Euclidean-based signal notion in TBM Han et al. (2020) fails to accurately describe the phase transition in our dTBM due to the possible heterogeneity in degree $\boldsymbol{\theta}$. To compare, we denote our angle-based signal notion in (**??**) and the Euclidean-based SNR in Han et al. (2020) as Δ_{ang}^2 and Δ_{Euc}^2, respectively:

$$\Delta_{\text{ang}}^2 = 2(1 - \max_{a \neq b \in [r]} \cos{(\boldsymbol{S}_{a:},\ \boldsymbol{S}_{b:})}), \quad \Delta_{\text{Euc}}^2 = \min_{a \neq b \in [r]} \|\boldsymbol{S}_{a:} - \boldsymbol{S}_{b:}\|^2.$$

By Lemma **??** in the Appendix II, we have

$$\Delta_{\text{ang}}^2 \max_{a \in [r]} \|\boldsymbol{S}_{a:}\|^2 \leq \Delta_{\text{Euc}}^2.$$

The above inequality indicates the condition $\Delta_{\text{Euc}}^2 \leq p^\gamma$ is **sufficient but not necessary** for $\Delta_{\text{ang}}^2 \leq p^\gamma$. In fact, if we were to use Δ_{ang}^2 for both models, then the phase transition of dTBM can be arbitrarily worse than that for TBM.

Here, we provide an example to illustrate the dramatical difference between TBM and dTBM with the same core tensor.

Example 5 (Comparison with Euclidean-based signal notion)**.** Consider a biclustering model with $\boldsymbol{\theta} = 1$ and an order-2 core matrix

$$\boldsymbol{S} = \begin{pmatrix} p^{(\gamma+1)/2} + 2 & 2p^{(\gamma+1)/2} + 4 \\ 2 & 4 \end{pmatrix}, \quad \text{with } \gamma \leq -1.$$

The core matrix $\boldsymbol{S}$ lies in the parameter spaces of TBM and our dTBM. Here, the constraint $\gamma \leq -1$ is added to ensure the bounded condition of $\boldsymbol{S}$ in our parameter space

in (**??**). The angle-based and Euclidean-based signal levels of $\boldsymbol{S}$ are

$$\Delta^2_{\text{ang}}(\boldsymbol{S}) = 0 \;\; (\leq p^\gamma), \quad \Delta^2_{\text{Euc}}(\boldsymbol{S}) = 5p^{\gamma+1} \;\; (\geq p^\gamma).$$

We conclude that TBM with $\boldsymbol{S}$ achieves exact recovery with a polynomial-time algorithm; see Han et al. (2020, Theorem 4). By contrast, the dTBM with the same $\boldsymbol{S}$ and input $r = 2$ violets the identifiability condition, and thus fails to be solved by all estimators; see our Theorem **??**."

We have revised our Section **??** to better present the lower bounds. See our response to **# 1 Major Comments from Editor**.

References

Chien, I. E., Lin, C.-Y., and Wang, I.-H. (2019). On the minimax misclassification ratio of hypergraph community detection. *IEEE Transactions on Information Theory*, 65(12):8095–8118.

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.

Gao, C. and Zhang, A. Y. (2022). Iterative algorithm for discrete structure recovery. *The Annals of Statistics*, 50(2):1066–1094.

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.

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*.

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.

Algorithm: Multiway spherical clustering for degree-corrected tensor block model

Sub-algorithm 1: Weighted higher-order initialization

Input: Observation $\mathcal{Y} \in \mathbb{R}^{p \times \cdots \times p}$, cluster number r, relaxation factor $\eta > 1$ in k-means clustering.

1: Compute factor matrix $\boldsymbol{U}_{\mathrm{pre},k} = \mathrm{SVD}_r(\mathrm{Mat}_k(\mathcal{Y})), k \in [K]$ and the $(K-1)$-mode projection

$$\mathcal{X}_{\mathrm{pre},k} = \mathcal{Y} \times_1 \boldsymbol{U}_{\mathrm{pre},1}\boldsymbol{U}_{\mathrm{pre},1}^T \times_2 \cdots \times_{k-1} \boldsymbol{U}_{\mathrm{pre},k-1}\boldsymbol{U}_{\mathrm{pre},k-1}^T \times_{k+1} \boldsymbol{U}_{\mathrm{pre},k-1}\boldsymbol{U}_{\mathrm{pre},k+1}^T \times_{K+1} \times_K \boldsymbol{U}_{\mathrm{pre},K}\boldsymbol{U}_{\mathrm{pre},K}^T.$$

2: Compute factor matrix $\hat{\boldsymbol{U}}_k = \mathrm{SVD}_r(\mathrm{Mat}_k(\mathcal{X}_{\mathrm{pre},k})), k \in [K]$ and denoised tensor

$$\hat{\mathcal{X}} = \mathcal{Y} \times_1 \hat{\boldsymbol{U}}_1\hat{\boldsymbol{U}}_1^T \times_2 \cdots \times_K \hat{\boldsymbol{U}}_K\hat{\boldsymbol{U}}_K^T.$$

3: **for** $k \in [K]$ **do**
4:     Let $\hat{\boldsymbol{X}} = \mathrm{Mat}_k(\hat{\mathcal{X}})$ and $S_0 = \{i \in [p] : \|\hat{\boldsymbol{X}}_{i:}\| = 0\}$. Set $\hat{z}(i)$ randomly in $[r]$ for $i \in S_0$.
5:     For all $i \in S_0^c$, compute normalized rows $\hat{\boldsymbol{X}}_{i:}^s := \|\hat{\boldsymbol{X}}_{i:}\|^{-1}\hat{\boldsymbol{X}}_{i:}$.
6:     Solve the clustering $\hat{z}_k : [p] \to [r]$ and centroids $(\hat{\boldsymbol{x}}_j)_{j \in [r_k]}$ using weighted k-means, such that

$$\sum_{i \in S_0^c} \|\hat{\boldsymbol{X}}_{i:}\|^2 \|\hat{\boldsymbol{X}}_{i:}^s - \hat{\boldsymbol{x}}_{\hat{z}_k(i)}\|^2 \leq \eta \min_{\bar{\boldsymbol{x}}_j, j \in [r], \bar{z}_k(i), i \in S_0^c} \sum_{i \in S^c} \|\hat{\boldsymbol{X}}_{i:}\|^2 \|\hat{\boldsymbol{X}}_{i:}^s - \bar{\boldsymbol{x}}_{\hat{z}_k(i)}\|^2.$$

7: **end for**
Output: Initial clustering $z_k^{(0)} \leftarrow \hat{z}_k, k \in [K]$.

Sub-algorithm 2: Angle-based iteration

Input: Observation $\mathcal{Y} \in \mathbb{R}^{p \times \cdots \times p}$, initialization $z_k^{(0)} : [p] \to [r], k \in [K]$ from Sub-algorithm 1, iteration number T.

8: **for** $t = 0$ to $T - 1$ **do**
9:     Update the block tensor $\mathcal{S}^{(t)}$ via $\mathcal{S}^{(t)}(a_1, ..., a_K) = \mathrm{Ave}\{\mathcal{Y}(i_1, \ldots, i_K) : z_k^{(t)}(i_k) = a_k, k \in [K]\}$.
10:     **for** $k \in [K]$ **do**
11:         Calculate reduced tensor $\mathcal{Y}_k^{\mathrm{d}} \in \mathbb{R}^{r \times r \times p \times r \times \cdots \times r}$ via

$$\mathcal{Y}_k^{\mathrm{d}}(a_1, \ldots, a_{k-1}, i, a_{k+1}, \ldots, a_K) = \mathrm{Ave}\{\mathcal{Y}(i_1, \ldots, i_{k-1}, i, i_{k+1}, \ldots, i_K) : z^{(t)}(i_j) = a_j, j \neq k\}$$

12:         Let $\boldsymbol{Y}_k^{\mathrm{d}} = \mathrm{Mat}_k(\mathcal{Y}^{\mathrm{d}})$ and $J_0 = \{i \in [p] : \|\boldsymbol{Y}_{i:}^{\mathrm{d}}\| = 0\}$. Set $z_k^{(t+1)}(i)$ randomly in $[r]$ for $i \in J_0$.
13:         Let $\boldsymbol{S}^{(t)} = \mathrm{Mat}(\mathcal{S}^{(t)})$. For all $i \in J_0^c$ update the cluster assignment by

$$z(i)_k^{(t+1)} = \underset{a \in [r]}{\arg\max}\cos\left(\boldsymbol{Y}_{k,i:}^{\mathrm{d}}, \ \boldsymbol{S}_{a:}^{(t)}\right).$$

14:     **end for**
15: **end for**
Output: Estimated clustering $z_k^{(T)} \in [r]^p, k \in [K]$.