We thank the reviewers for the helpful feedback. Our responses are detailed below.

- **(R4, Prior work)** The comparisons were provided in the original Table 1. We will add more discussions as follows. (1) **Comparison to Ke et al'20**. We emphasize that our contribution is not to propose a new tensor model. The degree-corrected model and its variant have been proposed in literature (Ke et al'20, Gao et al'18, Chi et al'21), as we pointed out in Contributions and Section 2.1. For this popular model, however, the optimal statistical-computational rates remain an open problem. Our main contribution is to provide a sharp statistical and computational critical phase transition in dTBM literature. In addition, our algorithm results in a faster *exponential* error rate, in contrast to the *polynomial* rate in Ke et al. We believe our results are novel and helpful to the community. (2) **Comparison with Han et al'21**. Extending results from non-degree to degree-corrected model is highly challenging. Our dTBM parameter space is equipped with angle-based similarity and nuisance degree parameters. The extra complexity makes the Cartesian coordinates based analysis [Han et al] non-applicable to our setting. Towards this goal, we have developed a new polar coordinates based analysis to control the model complexity. We also develop a new angle-based iteration algorithm to achieve optimal clustering rates *without the need of estimating nuisance degree parameters*.

- **(R4, Gaussian vs. non-Gaussian)** The boundedness constraint in binary data incur no extra implementation challenges to our algorithm. Both initialization and iteration use scale-invariant angle-base similarity for clustering. After we obtain the final clustering, our mean tensor estimation is based on sample average, which is guaranteed to be in the valid range $[0, 1]$ for binary data. Numerical results also confirm the improvement of our algorithm under binary data. The theoretical behavior for general non-Gaussian data (and mixture thereof) is a direction for future extension.

– We thank R2, R3, R6 and R8 for the positive comments! We will improve the exposition as suggested.

- **(R2, R6, Clarification)** (1) We will change the wording: *individual/block effects* $\rightarrow$ *individual/block parameters*, *degree heretogeneity* $\rightarrow$ *non-equal degrees*. (2) The extension to asymmetrical tensors is obtained via replacing $(\boldsymbol{M}, \boldsymbol{\Theta})$ in eq (1) by mode-specific parameters $(\boldsymbol{M}_k, \boldsymbol{\Theta}_k)$ for every mode $k \in [K]$. (3) We will remove Remark 1 as suggested.

- **(R6, Motivation of dTBM)** The dTBM is well-motivated from practical applications. First, the higher-order clustering provides new insights of underlying similarity among entities. In evolution biology, for example, the higher-order clustering is useful for explaining stable coexistence of diverse species competitors [Grilli et al, 2017]. Second, the allowance of degree makes the model more realistic. Degree-corrected method captures both the individual heterogeneity and community structure. Non-degree model ignoring the individual heterogeneity may seriously mislead the clustering results. See the Karate club network and political blogs networks analyses [Karrer et al, 2012] for detailed discussions.

- **(R6, Matrix vs. Tensor)** Unique challenges arise for higher-order tensors. We show that the statistical-computational gap is *provably* unavoidable for tensors of order 3 or higher; see Fig 2 and Remark 2. This result reveals the intrinsic distinctions between matrices ($K = 2$) and tensors ($K \geq 3$). The difference stems from the algebraic *multilinearity* properties of tensors. We develop multilinear techniques to prove this fundamental distinction from matrices to tensors.

- **(R6, Fair Comparison)** The numerical comparison is fair and reasonable. We use the same simulation set-up as in the past literature for fair comparison with *competing* methods **SCORE** and **HOSVD+**. The data is generated from the generative models underlying the competing methods, so no model misspecification issue arises for *all* considered methods. Non-degree methods are applied as benchmarks to assess the necessarity of degrees correction.

- **(R8, Parameter Space)** Our parameter space is flexible compared to existing work. The balanced community assumption is mild in literature. In fact, our simulation requires only the positive community size (i.e., $\min_{a\in[r]} |z^{-1}(a)| \geq 1$) but no balanced community assumptions; see Section 5. The outperformance of dTBM shows the robustness of our algorithm to the imbalanced communities. We will add the table below to the final version, as suggested by the reviewer.

Assumptions in parameter space	[Gao et al'18]	[Han et al'21]	[Ke et al'20]	Ours
Balanced community size	$\surd$	$\surd$	$\surd$	$\surd$
Balanced degree parameters	$\surd$	-	$\surd$	$\surd$
Flexible in-group connections	$\times$	$\surd$	$\times$	$\surd$
Gap among cluster centers	In-between cluster difference	Euclidean gap	Eigen gap	Angle gap

- **(R8, Growth of K)** Our results in numerical experiments are still valid as the number of modes K grows larger. In fact, our theoretical results in Sections 3-4 show that the statistical-theoretical gap and outperformance of our algorithm will be more significant with larger K.

- **(R2, R3, R6, Real data)** We will add two more real data applications, space permitting. (1) human brain connectivity analysis. We will apply our algorithm to multi-modal connectivity networks to exploit clustering among brain nodes and among measurements; (2) recommendation system. We will use InCarMusic as our dataset, where the data is a tensor of user $\times$ song $\times$ context. We will use our method to identify three-way clustering of users, songs, and contexts.