



18 Jan 2022

Dear Editors,

I am writing to submit our manuscript “Multiway Spherical Clustering via Degree-Corrected Tensor Block Models” to the IEEE Transactions on Information Theory.

In this manuscript, we investigate the degree-corrected tensor block model (dTBM) for multiway clustering. The goal of multiway clustering is to identify the checkerboard structure in a noisy data tensor collected from social sciences, neuroscience, genetics, and computer vision applications. Compared with classical tensor block models, dTBM allows both block and individual effects and assumes a more flexible structure in multiway clustering. Studying the statistical and computational limits of dTBM and developing an optimal algorithm are of necessity and importance to the field of multiway clustering.

We present the phase transition of clustering performance under dTBM based on the notion of angle separability. We characterize three signal-to-noise regimes corresponding to different statistical-computational behaviors. In particular, we demonstrate that an intrinsic statistical-to-computational gap emerges only for tensors of order three or greater. Further, we develop an efficient polynomial-time algorithm that provably achieves exact clustering under mild signal conditions. The efficacy of our procedure is verified by two applications, one on human brain connectome project, and another on Peru Legislation network dataset.

Part of this work was presented at the **25th International Conference on Artificial Intelligence and Statistics (AISTATS)**. However, this IEEE submission covers much greater depth and extends the earlier publication in a substantial way. Newly added results are summarized below.

- We have added four new motivating problems to highlight the applicability of our proposed models. The new applications include community detection in hypergraphes, clustering in multi-layer weighted networks, and multi-tissue multi-individual gene expression analysis. See Section II.B.
- The earlier publication addresses only the special symmetric tensors. We have substantially extended our algorithm to more general models, including asymmetric tensors and non-Gaussian observations. New theoretical results are provided to justify the extension. We have also added two initialization schemes, one for independent and identically distributed sub-Gaussian noise, and another for independent but non-identically distributed Bernoulli noise. Theories are extended correspondingly. See Section V.A.
- We have proposed a new procedure for hyperparameter selection. We have added new simulation studies to assess the efficacy in Section V.A and Table II.

- We have added an entirely new real data application on Human Brain Connection Project (HCP). The extensive data analysis in Section VI.A and new Figures 9-11 demonstrate the potential of our method in neuroscience applications.

We believe our results will be of interest to a very broad readership – from those interested in new theoretical results to those interested in efficient algorithms in multiway clustering with social networks and neuroscience applications. Our method will help the practitioners efficiently analyze tensor datasets in various areas. Toward this end, we also submit the developed software package accompanying this paper.

We appreciate your consideration.

Sincerely,

Miaoyan Wang,
Assistant Professor of Statistics
University of Wisconsin–Madison
Website: <http://www.stat.wisc.edu/~miaoyan/>
Email: miaoyan.wang@wisc.edu