Functional edged network model based on functional tensor decomposition

1. **Summary and Contributions: Briefly summarize the paper and its contributions**

   This work proposes dynamic network model in which edges are smooth continuous functions. The problem is formulated to estimation the low-rank tucker tensor decomposition from incomplete and noisy observations. A conjugate gradient algorithm under Riemann optimization is provided. Basic theoretical guarantees for the noisy tensor completion, numerical comparisons with related methods, and real data analysis are also provided to support the efficacy of proposed method.

   As far as I am concerned, the main contributions lie in the novel functional-edged network model and the corresponding optimization problem with smoothness penalty.

2. **Strengths: Please describe the strengths of the work according (but not limited) to the following criteria: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the AISTATS community.**

   The functional-edged network is novel and interesting. It is also interesting to re-formulate the infinite-dimensional estimation problem (6) to a regular low-rank tensor completion problem with smoothness penalty (10).

3. **Weaknesses: Please describe the limitations of this work according (but not limited) to the following criteria: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the AISTATS community.**

   As the functional-edges network estimation reduces to the regular tensor completion problem, I feel the main novelty should be reflected on how to address the smoothness penalty in (10); otherwise, the problem degenerates to a pure tensor completion finding the least square estimator. However, the theoretical analysis avoids the discussion about the smoothness penalty, and the numerical experiments do not highlight the necessity of the smoothness penalty under certain scenario. Further, the theoretical results in Section 5 seem not new in the tensor completion studies, and the error bounds may not be sharp. Guarantees for the algorithm empirical performance are also not provided.

   See additional comments for details and other comments.

4. **Correctness: Are the method and claims correct? Is the empirical methodology correct?**

   Most parts of the result seem sound. Some details in the algorithm and theorem are unclear for me; see additional comments for detail.

5. **Clarity: Is the paper well written? Does it clearly state its contributions, notation and results?**

   Most parts of the manuscript are well-written. Some typos and minor writing problems could be fixed.

   - The word “irregular” is not well-explained or defined at the beginning of the introduction. Does “irregular” mean “incomplete”?

   - Page 3, right column, “obay” -¿ “obey”;

   - Page 4, left column, “$\mathbf{B} \times \Psi$” -¿ $\mathbf{B} \times \Phi$;

   - Page 5, right column, the matricization $\mathbf{X}_{(i)}$ is not defined;

6. **Relation to prior work: Is it clearly discussed how this work differs from or relates to prior work in the literature?**

   Note that the proposed functional-edged network recovery problem is reformulated to a tensor completion problem in the practical scenario. It is worthwhile to compare with more advanced tensor completion methods [1], [2], and other methods. Particularly, both [1] and [2] adopt the tucker decomposition structure, which have similar interpretability with the proposed method. The theoretical results for the least square estimator in [1] may also be comparable with the theoretical results in Section 5.

   [1]Xia, Dong, and Ming Yuan. ”On polynomial time methods for exact low-rank tensor completion.” Foundations of Computational Mathematics 19, no. 6 (2019): 1265-1313.

   [2] Xia, Dong, Ming Yuan, and Cun-Hui Zhang. ”Statistically optimal and computationally efficient low rank tensor completion from noisy entries.” The Annals of Statistics 49, no. 1 (2021).

7. **Additional Comments: Add your additional comments, feedback and suggestions for improvement, as well as any further questions for the authors.**

   - Theorem 2 and Corollary 1 indicate the MSE ($||\hat{\mathbf{X}} - \mathbf{X}||_F^2$) upper bound is of order $\mathcal{O}(\prod_{i=1}^4 n_i)$, which is not optimal in many denoising works. For example, in the denoising method [3] with missing tensor entries, similar MSE is of order $\mathcal{O}(\sum_{i=1}^4 r_i n_i)$, which is much smaller than that in Section 5 and coincides with the number of parameters in low-rank tensor decomposition model. More discussions about the theoretical sharpness may be helpful.

   [3] Lee, Chanwoo, and Miaoyan Wang. ”Tensor denoising and completion based on ordinal observations.” In International Conference on Machine Learning, pp. 5778-5788. PMLR, 2020.

   - The symmetrization Algorithm 2 is not clear for me. The output tensor is $\tilde{\mathbf{B}} \times_1 \hat{U}_1 \times_2 \hat{U}_1 (\hat{U}_1)^T \hat{U}_2 \times_3 \hat{U}_3$. The symmetry on the first and second modes of this output is not straightforward. Are there any underlying relationship between $\tilde{\mathbf{B}}$ and factors $\hat{U}_1, \hat{U}_2$? More explanations would be helpful to understanding Algorithm 2.

   - The algorithm and theorem focus on the estimation of the dicretilized signal $\mathbf{X}$. The second last paragraph in Section 3 states that it is easy to recover the infinite-dimensional signal $\mathcal{X}$. However, obtaining continuous signal from discrete signal is not a trivial procedure for me. Adding more discussions on this point would be beneficial.

   - The interpretation of Theorem 1 is unclear. Given two different signals may minimize the least square sum, why can we say the objective function is identifiable?

- The interpretation of the community structure in model (1) is not valid for me. First, how can we interpret the orthogonality of the factor matrix? Second, should we have any constraint like $\sum_k \phi_{ik}^2 = 1$ for all i? Otherwise, the community assignment possibilities of every node i are not of the same scale. Last, the factor matrices Φ and Θ are not identifiable with respect to orthogonal transformation in the tucker model. Then, how can explain the non-identifiable factor?

References