Referee Response for JCGS Submission JCGS-21-341

We thank the referees for their time in reviewing this paper and providing valuable comments and suggestions, which have all been taken into account for this revision. Most of the changes are in the simulation and application sections. In particular, we add a simulation study of "disassortative" PABMs and clarify the choices made in the data preprocessing.

The following sections are individual referee comments and our responses (highlighted in blue).

Referee 1 Comments

Main comments

The interest of this paper is to provide new perspectives on PABM (theoretical and algorithmic). The theoretical results are sound. However, the algorithmic results shown in the simulations are not really impressive since they are done in settings that do not appear very challenging. Indeed, they are done with fixed numbers of communities which are quite low (no more than 4) and only with assortative structure (larger probabilities of connections within a community than between two communities). Therefore, I think the paper needs a more challenging simulation study and the question of the choice of the number of communities needs to be addressed at least with some heuristic that works in simulation studies if theoretical results are not reachable. Moreover, some points on the exposition could be made clearer. I have specific comments below where I ask for some clarifications and detail how the simulation study could be improved in my opinion.

We thank the referee for the feedback. The concerns addressed here are answered in the following section. In particular, we make sure to address the concerns about the simulation studies.

Specific comments and questions

1. p4. What is a hollow graph?
   We thank the referee for the question. We changed the terminology a bit here and throughout the paper and specify that the analysis assumes G is an unweighted and undirected graph with no self-loops (no edge directly from itself to itself), so A is binary and symmetric with zeros on the diagonal.

2. p4. G as a graph is non-oriented or undirected and A is symmetric.
   We have made minor edits clarifying the structure of G and A.

3. p4. Please define the set $[n]$.
   This has been added. $[n] = \{1, 2, ..., n\}$.

4. p4. Definition 1. Do you consider a probability distribution on $z_1, ..., z_n$? Please give the range of variation of (i, j) when you define P_{ij}.
   We thank the referee for the question and suggestion. We have included the range of (i, j) in the definition. In regard to the probability distribution on the labels $z_1, ..., z_n$, none of the theorems require this, although some do require that each n_k be sufficiently large, so it did not factor into our definitions and analyses. One thing that was discussed was a possibility of a Bayesian analysis of the PABM, and we have some results using Variational Inference and Mean Field Approximation, but we believe that this is a topic for another paper.

5. p4. line 50. Instead of "increasing values of the community assignments", you could write "reorganized by community memberships".
   We thank the referee for the suggestion. This has been changed.

6. p5. Definition 2. I do not understand the definition of the subset $\mathcal{X}$. Is it $x \in \mathcal{X}$ if for all $y \in \mathbb{R}^d$ we have $x^\top I_{p,q} y$ or is $\mathcal{X}$ a subset on $(\mathbb{R}^d)^2$? Please clearly state that $d = p+q$ (not stated before Definition 3).
We thank the referee for the suggestion. We clarify the definition of the subset as $\mathcal{X} = \{x, y \in \mathbb{R}^d : x^\top I_{p,q} y \in [0,1]\}$ and add that $d = p + q$.

7. p6. Remark 2. What are $\hat{Z}$, V, and D?
We thank the referee for the question. We now include in the remark the D is the diagonal matrix of the p most positive and q most negative eigenvalues of A (or P), V is the matrix of corresponding eigenvectors, and $\hat{Z}$ is the embedding $V|D|^{1/2}$.

8. top of p7. $\tilde{P}$ instead of P.
We have corrected the typo.

9. p7 line 40-46. I don't understand the interest of this paragraph.

10. p8, line 14-21. I don't see why the permutation given by Π has K fixed points.
We thank the referee for the comment. The purpose of Example 1 is to illustrate this explicitly. Comparing matrices X and Y, we can see that they share three columns: the first, fifth, and ninth. The other columns are permuted. From Example 1 (as well as the proof of Theorem 1 for $K = 2$), it should be clear how this property holds for larger K. The proof of Theorem 1 for arbitrary K should then confirm the intuition based on $K = 2$ and 3.

11. Section 3.1. I found it confusing that community detection algorithms are detailed when P is assumed to be known and then methods for estimating P are recalled. A sentence in the beginning of this section to expose what is presented could help the reader.

12. p10, line 12. $c^{(i)}$ is not defined.
We thank the referee for pointing this out. $c^{(i)}$ is the i^{th} entry of column vector c. This has been added to the paper.

13. p10. I didn't understand the exposition of the SSC algorithm.

14. p11, line 31. "... SSC on the spectral embedding of A". Could you support this assertion by references?
We thank the referee for the question. We have added, "... SSC on the rows of the spectral embedding of A, since they will still lie approximately on the K subspaces [2]."

15. p12, line 13. It could help recall what the z_i, z_j are.
We have changed the wording for this bit to explicitly talk about community labels.

16. p14. Do you think that the OSC algorithm can retrieve other kinds of structure than assortative ones?
We thank the referee for the question. In response to this, we added a third simulation study with $\lambda_{kk} \sim Beta(1,2)$ and $\lambda_{kl} \sim Beta(2,1)$, resulting in an average within-community edge probability of $1/3$ and between-community edge probability of $2/3$. Results of this simulation study are very similar to that of the first two and consistent with the theoretical results.
On a related note, it is our view that the distinction between assortative and disassortative is not very important for the PABM, at least when it is viewed as a type of GRDPG. In the case of the SBM and DCBM, it can be shown that the rank of P (and therefore the number of embedding dimensions) is equal to K, but assortativity or disassortativity determines how many positive and negative eigenvalues P has [2]. In the case of the PABM, the number of positive eigenvalues is always $K(K+1)/2$ and the number of negative eigenvalues is always $K(K-1)/2$, (assuming $n_k > K$ for each $k \in [K]$).

17. Sections 4 and 5. Why do you use either the ARI criterion or the misclassification rate to assess the recovery of the nodes clustering?
We thank the referee for the question. The first to datasets in section 5 refer to datasets analyzed by other PABM papers. For the Leeds Butterfly dataset, while the paper with which we are comparing [1] provides enough details to transform and preprocess the data to match their analysis, they do not provide code to run their clustering algorithm. Therefore, we use their reported benchmark value, which is in the form of ARI. In order to make an apples-to-apples comparison, we report the ARI for OSC

and SSC-ASE here as well. For all other analyses (simulation and real data), we use misclassification error (whether rate or count).

18. Do you consider sparsity in the simulated adjacency matrices?

19. Why do you limit your simulation to assortative structure? Is it possible to see the results with a larger value of K? To what extent does the value of K impact the computational burden for the algorithms?

20. Again, in your application on the Leeds butterfly dataset, why did you limit your analysis to $K = 4$?
We thank the referee for the question. For the Leeds butterfly dataset, we wanted to compare results of OSC and SSC-ASE to an analysis performed by another group studying the PABM [1]. In their results, they describe how they removed some of the nodes and discretized the edges, which we replicate in our analysis.

21. Please specify in the text that the proofs for Theorems 3, 4, and 5 are given in the Appendix.
We have added this to the end of the introduction section.

Typos

1. p 10, line 16 "if each x_i lieS . . . "
We have corrected this typo.

Referee 2 Comments

The first part of the paper shows that any PABM is a GRDPG and gives a constructive proof of the result. In particular, the proof gives the construction of the latent positions of the corresponding GRDPG.

The second part of the paper investigates the implications of this relation to the GRDPG (and the identified latent positions) on estimation algorithms for the PABM. New algorithms for community detection and parameter estimation are proposed. In particular, a new community detection method, named orthogonal spectral clustering (OSC), is proposed that directly exploits the result of Theorem 1. Furthermore, an improvement of the sparse subspace clustering (SSC) algorithm is proposed consisting of applying SSC to the adjacency spectral embedding of the network (ASE), referred to as SSC-ASE. For both algorithms it is shown that the subspace detection property holds and that the convergence rate of the detection error is faster than for classical SSC. Likewise, it is shown that parameter estimation can be improved. A numerical study illustrates the performance of the algorithms in comparison with existing methods. It shows that OSC performs the best, and also that SSC-ASE improves the results with respect to classical SSC in some scenarios. In the application on real datasets, OSC and SSC-ASE provide good results, but do not achieve the performance of modularity maximization (MM).

The topic of the paper is interesting as it concerns a recent graph model and the results are valuable. Contributions are two-fold: theoretic and computational. The paper is well written, and the presentation of the results and the proofs is very clear.

I only have some minor comments and a couple of questions:

- I have a general issue with latent position models like the GRDPG due to the nonidentifiability of the latent space. As stated in the paper, latent positions are not unique, and not even the dimension of the latent space is identifiable. Theorem 1 (or its proof) provide one of all possible latent positions. A general question is what is the consequence of this specific choice. Does the latent space (or its dimension) have an impact on the clustering results? Indeed, I lack intuition about what type of latent space would be optimal for community detection or estimation in the PABM. The space with the smallest possible dimension? Do the authors have an idea of the properties of the latent space that would give the best results for community detection? Related to this issue, is the latent space of the proof of Theorem 1 the smallest possible latent space to relate PABM to the GRDPG?
We thank the referee for the questions.

- This is clearly beyond the scope of the paper, but I wonder if there is a straightforward way to adapt the proposed methods to directed models or weighted graphs or any other types of graphs than binary undirected ones?
- In the simulation study, SSC-ASE works well for $K \geq 3$, but not for $K = 2$. The authors explain the bad performance by the use of gaussian mixtures in the last clustering step. However, I don't understand why this has an impact on the performance for $K = 2$, but not for $K \geq 3$.
- In the application section, MM performs better than OSC and SSC-ASE, while in the simulation study the opposite is observed. What is the reason for the different behavior?
- Typo on p. 10: Xc should be $X^\top c$. Likewise, $X_{-i}c$ is to be replaced with $X_{-i}^\top c$.

References

[1] M. Noroozi, R. Rimal, and M. Pensky. Estimation and clustering in popularity adjusted block model. *Journal of the Royal Statistical Society, Series B.*, 2021+.

[2] P. Rubin-Delanchy, J. Cape, M. Tang, and C. E. Priebe. A statistical interpretation of spectral embedding: the generalised random dot product graph. arXiv: 1709.05506, 2017.