Dear Professor Peng,

We are re-submitting a new and thoroughly revised version of our paper "Clustering of Diverse Multiplex Networks". We satisfied all reviewers' requests. Specifically, we generalized our theory to the case when all layers of the network follow the Generalized Dot Product Graph (**GDPG**) model which includes practically all known block models (Stochastic Block Model, Degree Corrected Block Model, Mixed Membership Model, Popularity Adjusted Block Model, among them) as its particular cases. We developed a novel theory for this very general case, carried out the simulation studies and demonstrated the advantages of the new approach using real data examples.

We would like to point out that even the previous versions of our paper generalized a variety of works published recently in the best statistical journals. The new version of the paper is **truly novel**, and we expect that it is accepted and published quickly. Our paper deals with the very fast moving area of research, and are concerned about our priority. For this reason, we hope that the paper will be sent to the same set of reviewers, since they are familiar with the paper and can evaluate the new version quickly.

We have attached point-by-point responses to Reviewers 1 and 3 (since Reviewer 2 has no adverse comments about the paper). Here are the main points of these responses:

1. Reviewer 1 completely misunderstood our model (DIMPLE) and the difference in complexity between DIMPLE and MMLSBM, where there are only M distinct layers (so, it is essentially the k-means clustering problem with M classes). On the other hand, our model (DIMPLE) has L different SBMs that only share the clustering matrices. Hence, the total number of distinct layers is $L >> M$. On the other hand, comparison of our paper with Lie and Lin (2022) is misleading. In the latter paper, all layers have the same communities, so averaging of layer adjacency matrices (or their adjusted squares) leads to reduction of errors. On the contrary, in our model, we need to cluster the layers first. The mistakes in this initial clustering lead to the error terms that do not tend to zero as L grows.

In addition, in the new version of the paper, we extended our model to the case where layers of the network follow the GDPG model. This setting significantly generalizes the COSIE model existing in literature, and provides better results for the subspace recovery when our model reduces to COSIE.

2. The main objections of the Reviewer 3 are that the paper focuses on the "unrealistic" Stochastic Block Model rather than the Random Dot Product Graph Model, that the real data example is analyzed when the layer networks are equipped with the SBMs, and that there is only one real data example. The new version of the paper extends the investigation to the Generalized Dot Product Graph model, which has already been present in the previous version of the paper. In fact, the GDPG-equipped network is the main focus of the new version of the paper. In addition, following reviewer's request, we added the second real data example.

Clustering of Diverse Multiplex Networks

Responses to the Reviewer 1

(Correctness of error bound) The between-layer and within-layer error bounds in Theorems 1 and 2 remain counter-intuitive in the number of layers L. The between-layer error in new equation (34) keeps the same when L increases and right hand side of within-layer error in new equation (35) is independent with L. The bounds (34) and (35) contradict to the intuition that clustering performs better with more samples.

Thank you for your comment. We have addressed all your concerns in Section 6, Discussion. Specifically, unlike in the MMLSBM model where one has M layers with **identical** SBMs, we have L different SBMs that only share the clustering matrices. Hence, the total number of parameters in our model is $O(LK^2 + nM \log K + L \log M)$ while the total number of parameters in MMLSBM is $O(MK^2 + nM \log K + L \log M)$. When you divide by the total number of observations, you see that, for fixed K, M and no sparsity, the dominant term is K^2/n while, for MMLSBM, it is $K^2/(nL)$, which is what you see in the between-layer clustering error.

The within-layer clustering errors are the sums of two components. The first component is due to the errors in the clustering of layers: indeed, **unlike** Lei and Lin (2022) paper, DIMPLE **does not have** all L identical layers, and errors in identification of layers with the similar community structures add up to errors of community detection. The second term, of the order $O(n^{-1}\,\rho_n^{-1}\,\log n\, L^{-1}$, is due to clustering in the groups of layers and, as expected by your intuition, it does tend to zero as L grows. This term is due to the clustering procedure which is similar to Lei and Lin (2022) paper (and we acknowledge this in the paper), **except** we do not assume that $n\rho_n$ is bounded above by a constant in our theoretical analysis.

In fact, authors hide the term L in the statement by assuming $L \le n^{\tau_0}$, and the revised proofs show no improvement compared to first version. The relationship between parameters, mentioned in authors response for Question 2, does not explain the error bound sub-optimality in L compared with reference Lei and Lin (2022). In addition, it should be the term $O(1/L + n)$ may not be *large* enough ... based on current results.

Sorry but there is some **misunderstanding** here. First, since we are applying a union bound over the number of layers for the between-layer clustering errors, the error bounds holds with probability $Ln^{-\tau}$ which becomes $n^{-(\tau-\tau_0)}$ under assumption $L \le n^{\tau_0}$. We placed explicit probability bound in the Theorems but we still cannot allow L to grow exponentially with respect to n. As for the "sub-optimality in L compared with reference Lei and Lin (2022)", **the reason is that you compare two completely different models**. In Lei and Lin (2022), community structures **are the same for all layers**, so the first term in the within-layer clustering error vanishes (since it is due to the errors in between-layer clustering). In fact, in the case of the SBM-equipped model with one group of layers (the case of Lei and Lin (2022)), we get results **identical** to Lei and Lin (2022). This point is extensively discussed in the introduction to the paper.

(Contribution) The between-layer clustering based on the network community membership is not new (Stanley et al., 2016), though the DIMPLE setting alters from standard model in some minor aspects. The general idea has been used, in different guises, but in similar contexts. In particular, the main method of the work follows almost the same as the standard spectral analysis and k-means results in Lei and Lin (2022). Therefore, I remain unconvinced of the technical contribution of this paper.

We are sorry but we feel that the statement that "the DIMPLE setting alters from standard model in some minor aspects" is completely inaccurate, and the reference to Stanley et al.(2016) is inappropriate. First, the Stanley et al.(2016) paper refers to the MMLSBM where there are only M distinct SBMs in L layers, and we discussed in details why the MMLSBM is **much easier** than our model. Second, the Stanley et al.(2016) paper has much weaker results than Fan et al. (2022) and Jing et al. (2021) which are reviewed extensively in our manuscript. Neither the assumptions nor the algorithms in those papers (and of Stanley et al.(2016)) work in our model. Section 7.5 in the Appendix of our paper is dedicated to the proof of the above fact.

In addition, even the previous version of the paper studied an extension of the SBM-equipped multilayer network model. Specifically, in the previous version, we showed that the between-layer clustering works in the case of a multilayer network where layers are equipped with the Generalized Dot Product Graph models, and layers can be partitioned into groups with the similar ambient subspace structures. The new version is dedicated to this, more general case and extends those results.