Minimax Rates of Community Estimation for Inhomogeneous Multilayer Stochastic Block Models

（非一様多層確率的ブロックモデルにおけるコミュニティ推定のミニマックスレート）

Dept. of Mathematical Informatics 48226221 Kota Nakagawa（中川 皓太）
Supervisor Prof. Fumiyasu Komaki（駒木 文保 教授）

1 Introduction

It has become increasingly common to observe multiple networks for the same nodes in various applications that investigate the relationships between elements. Multilayer networks are appeared, for instance, when constructing different networks from various interactions in social network analysis [2], or combining and comparing brain data collected from individual participants [6]. In such datasets, it is necessary to analyze common structures across the entire networks while also paying attention to the differences in structures between layers.

For single-layer network structure analysis, community detection using the *stochastic block model (SBM)* [3] is a well-established task. In the SBM, n vertices are partitioned into K communities. The probability of an edge being present only depends on the combination of communities to which the two vertices belong.

While the SBM is limited to handling single networks, the *inhomogeneous multilayer stochastic block model (IMLSBM)* is a natural extension of the SBM and one of the theoretical framework for multilayer networks. This framework was initially introduced in a slightly more general form by Paul and Chen [6], and Chen et al. [1] simplified and redefined it for the case where the number of communities $K = 2$.

The IMLSBM consists of two stages. In the first stage, it generates *individual community assignments* for each layer from the community structure common to the entire network, termed as *global community assignment*. Individual communities are obtained by randomly changing the elements of each global community to another community with probability ρ, which represents *inhomogeneity* and demonstrates the situation where the structures of each layer are "correlated but not identical." In the second stage, networks for each layer are generated according to SBMs from the individual communities of each layer. The model is schematically represented in Figure 1.

The goals of our study is to show the theoretical limits of estimating global community assignment for the IMLSBM and to construct an optimal and eficient estimation algorithm. This study gives insights into the inter-layer relationships of multilayer networks and significantly contributes to the analysis of real-world network data with diverse structures.

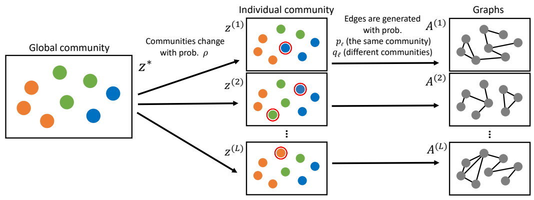

Fig. 1. Schematic diagram of the IMLSBM.

2 Related Work

Despite its practical significance, the theoretical properties of the IMLSBM are only known within a limited scope. In the homogeneous model, the minimax rates for Hamming loss in the estimation of global community is known where n tends to infinity [5], and a polynomial-time algorithm achieving this rates was proposed by Xu et al. [7]. While these results provide intriguing insights into the impact of signal strength coupling in multilayer networks, the effects of inter-layer structural differences and noise on estimation remain unclear. Chen et al. [1] gave the minimax rates and analysis of algorithms that achive this rates in the inhomogeneous case, but these results are limited to the case of the number of communities $K = 2$. Considering networks which appear in many application domains often involve numerous communities, further extensions are necessary. To the best of our knowledge, our study provides the first analysis of estimation limits and optimality in general multilayer network models with inhomogeneous structures and three or more communities.

3 Theorical Results

The theorical contributions consist of two parts.

3.1 Minimax lower bounds

The first result gives minimax lower bounds of global community estimation for the IMLSBM. Under the assumption of uniform community sizes and a fixed the number of communities K, we bound the asymptotic minimax rates for expected proportion of misclustered vertices as

$$\exp\{-(1+o(1))\mathcal{I}\} \tag{1}$$

where the number of vertices n tends to infinity and $\mathcal{I}$ is the signal-to-noise ratio (SNR) of the IMLSBM.

The signal strength depends on connection probabilities and the noise level correlates with the inhomogeneity parameter.

The proof is based on the discussion by Chen et al. [1], which provides the minimax rates for the case of $K = 2$. We extend the combinatorial arguments of their proof and replace concentration inequalities in it with tighter ones.

3.2 Algorithm and optimality

Second, we analyze a polynomial-time algorithm for global community estimation in the IMLSBM and show its minimax optimality under the similar assumptions as above and a fixed K. The algorithm is an extension of the estimation algorithm proposed by Paul et al. [1] from the case of $K = 2$ to $K \geq 3$. It is a two-stage algorithm which involve consistent initialization through spectral clustering and refinement based on MAP estimation. Spectral clustering is applied to an average adjacency matrix which is trimmed vertices with high degrees to amplify SNR. Vanilla MAP estimation faces optimization challenges due to dependencies between vertices, which also makes theoretical analysis difficult. However, these issues can be addressed through parameter fixing by the initial point.

The proof strategy of the optimality of the algorithm is also structured in two stages. In the first step, we show that the output of spectral clustering under a fixed K satisfies consistency with the global community. Subsequently, by fixing parameters with initial points that satisfy the consistency assumption, we prove that the MAP estimation yields a minimax optimal estimator.

4 Numerical Experiments

In addition to theoretical analysis, we demonstrate the performance of the algorithm in practical situations through numerical simulations. As a comparison method, we adopt the *co-regularized spectral clustering* proposed by Kumar et al. [4], which is commonly used for community detection in multilayer networks.

We compare the performance of the two algorithms in estimating global and individual communities, and demonstrate that the proposed method reduces losses in both cases (Figure 2). Additionally, comparing the computational time of two methods, we see that the proposed method exhibits high scalability with the number of layers (Figure 3).

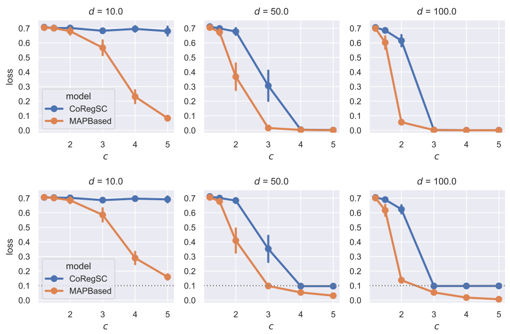

Fig. 2. Performance of co-regularized spectral clustering (blue) and the proposed algorithm (orange) for global (top) individual (bottom) community estimation.

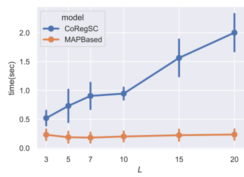

Fig. 3. Computation time of co-regularized spectral clustering (blue) and the proposed algorithm (orange) under different numbers of layers.

5 Conclusion

We theoretically revealed the limits of global community estimation error for the IMLSBM and analyzed MAP-based estimation with spectral clustering initialization. This is polynomial-time algorithm in the size of input and was shown to be minimax optimal under fixed number of communities and uniform sizes for each community. Through numerical simulations, we demonstrated that the two-stage algorithm achieves a balance between high estimation accuracy and reduction in time complexity. Our contributions provide essential insights into the theoretical properties of this model, and provide a foundation for future work.

Bibliography

[1] Chen, S., Liu, S. and Ma, Z., Global and individualized community detection in inhomogeneous multilayer networks, *The Annals of Statistics*, Vol. 50, No. 5, pp. 2664–2693 (2022).

[2] Dickison, M. E., Magnani, M. and Rossi, L., *Multilayer Social Networks*, Cambridge University Press (2016).

[3] Holland, P. W., Laskey, K. B. and Leinhardt, S., Stochastic blockmodels: First steps, *Social Networks*, Vol. 5, No. 2, pp. 109–137 (1983).

[4] Kumar, A., Rai, P. and Daume, H., Co-regularized multi-view spectral clustering, *Advances in Neural Information Processing Systems*, Vol. 24 (2011).

[5] Paul, S. and Chen, Y., Consistent community detection in multi-relational data through restricted multilayer stochastic blockmodel, *Electronic Journal of Statistics*, Vol. 10, No. 2, pp. 3807–3870 (2016).

[6] Paul, S. and Chen, Y., A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging, *The Annals of Applied Statistics*, Vol. 14, No. 2, pp. 993–1029 (2020).

[7] Xu, M., Jog, V. and Loh, P.-L., Optimal rates for community estimation in the weighted stochastic block model, *The Annals of Statistics*, Vol. 48, No. 1, pp. 183–204 (2020).