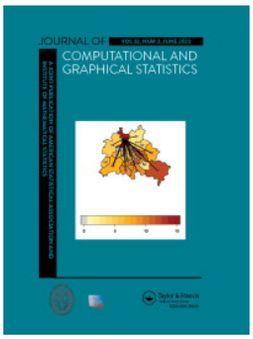


Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

Spectral Clustering via Adaptive Layer Aggregation for Multi-Layer Networks

Sihan Huang, Haolei Weng & Yang Feng

To cite this article: Sihan Huang, Haolei Weng & Yang Feng (2022): Spectral Clustering via Adaptive Layer Aggregation for Multi-Layer Networks, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2022.2134874

To link to this article: https://doi.org/10.1080/10618600.2022.2134874

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2022, VOL. 00, NO. 0, 1–15
https://doi.org/10.1080/10618600.2022.2134874

Taylor & Francis
Taylor & Francis Group

Spectral Clustering via Adaptive Layer Aggregation for Multi-Layer Networks

Sihan Huang[a], Haolei Weng[b], and Yang Feng[c]

[a]Department of Statistics, Columbia University, New York, NY; [b]Department of Statistics and Probability, Michigan State University, East Lansing, MI; [c]Department of Biostatistics, New York University, New York, NY

ABSTRACT

One of the fundamental problems in network analysis is detecting community structure in multi-layer networks, of which each layer represents one type of edge information among the nodes. We propose integrative spectral clustering approaches based on effective convex layer aggregations. Our aggregation methods are strongly motivated by a delicate asymptotic analysis of the spectral embedding of weighted adjacency matrices and the downstream *k*-means clustering, in a challenging regime where community detection consistency is impossible. In fact, the methods are shown to estimate the optimal convex aggregation, which minimizes the misclustering error under some specialized multi-layer network models. Our analysis further suggests that clustering using Gaussian mixture models is generally superior to the commonly used *k*-means in spectral clustering. Extensive numerical studies demonstrate that our adaptive aggregation techniques, together with Gaussian mixture model clustering, make the new spectral clustering remarkably competitive compared to several popularly used methods. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received September 2021
Accepted October 2022

KEYWORDS
Asymptotic misclustering error; Community detection; Convex aggregation; Eigenvalue ratio; Gaussian mixture distributions; *k*-means; Multi-layer networks; Spectral clustering

1. Introduction

Clustering network data is termed community detection, where communities are understood as groups of nodes that share more similarities with each other than with other nodes. Examples include social groups in a social network and papers on the same research topic in a citation network. Community detection is one of the fundamental problems in network analysis to understand the network structure and functionality (Newman 2018). With enormous effort from a broad spectrum of disciplines, a large set of methodologies have been proposed and can be roughly classified into algorithmic and model-based ones (Zhao, Levina, and Zhu 2012). Examples of algorithmic methods include divisive algorithms using edge betweenness (Girvan and Newman 2002), network random walk (Zhou 2003), spectral method (Hagen and Kahng 1992; Shi and Malik 2000), modularity optimization (Newman 2006), and information-theoretic approaches (Rosvall and Bergstrom 2008). We refer to Fortunato (2010) for a thorough and in-depth review of algorithmic community detection techniques. The class of model-based methods relies on fitting probabilistic models and applying statistical inference tools. Several widely studied models include stochastic block model (SBM) (Holland, Laskey, and Leinhardt 1983), degree-corrected stochastic block model (DSBM) (Karrer and Newman 2011), mixed membership stochastic block model (Airoldi et al. 2008), and latent variable models (Handcock, Raftery, and Tantrum 2007; Hoff 2008), among others.

In the past decade, there have been increasingly active researches toward understanding the theoretical performance of community detection methods under different types of models. The seminal work by Bickel and Chen (2009) introduced an asymptotic framework for the study of community detection consistency, and developed a general theory for checking the consistency properties of a range of methods including modularity and profile likelihood maximization, under SBM. Their consistency framework and results have been generalized to DSBM (Zhao, Levina, and Zhu 2012), to allow the number of communities to diverge with network size for maximum likelihood estimator under SBM (Choi, Wolfe, and Airoldi 2012), and to scalable pseudo-likelihood method (Amini et al. 2013). The consistency and asymptotic normality of maximum likelihood and variational estimators for other model parameters were also established under both SBM and DSBM (Celisse, Daudin, and Pierre 2012; Bickel et al. 2013). Another important line of research focuses on the analysis of spectral clustering. Consistency or misclustering error rate of different variants of spectral clustering approaches have been investigated under SBM (Rohe, Chatterjee, and Yu 2011; Lei and Rinaldo 2015; Yu, Wang, and Samworth 2015; Le, Levina, and Vershynin 2017), DSBM (Qin and Rohe 2013; Jin 2015), mixed membership models (Jin, Ke, and Luo 2017; Mao, Sarkar, and Chakrabarti 2020; Zhang, Levina, and Zhu 2020), and SBM with covariates (Zhang, Levina, and Zhu 2016; Weng and Feng 2022). A class of semidefinite optimization approaches with established strong performance guarantees has been developed (Cai and Li 2015; Guédon and Vershynin 2016; Amini and Levina 2018). Moreover, there is one major stream of research focused on characterizing the fundamental limits of community detection, under the minimax framework (Zhang and Zhou 2016; Gao et al. 2017; Xu, Jog, and

CONTACT Yang Feng *yang.feng@nyu.edu* Department of Biostatistics, New York University, New York, NY.
Sihan Huang and Haolei Weng have equally contributed to this work.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JCGS*.

Loh 2020), and with respect to sharp information-theoretic and computational thresholds (Decelle et al. 2011; Krzakala et al. 2013; Abbe, Bandeira, and Hall 2015; Mossel, Neeman, and Sly 2015; Abbe 2017).

All the aforementioned methods perform community detection based on a single-layer network that only retains one specific type of edge information. However, multi-layer networks with multiple types of edge information are ubiquitous in the real world. For example, employees in a company can have various types of relations such as Facebook friendship and coworkership; genes in a cell can have both physical and coexpression interactions; stocks in the U.S. financial market can have different levels of stock price correlations in different time periods. See Kim and Lee (2015) for a comprehensive report of multi-layer network data. Each type of relationship among the nodes forms one layer of the network. Each layer can carry potentially useful information about the underlying communities. Integrating the edge information from all the layers to obtain a more accurate community detection is of great importance. This problem of community detection for multi-layer networks has received significant interest over the last decade. The majority of existing algorithmic methods base on either spectral clustering (Long et al. 2006; Zhou and Burges 2007; Kumar, Rai, and Daume 2011; Dong et al. 2012) or low-rank matrix factorization (Singh and Gordon 2008; Tang, Lu, and Dhillon 2009; Nickel, Tresp, and Kriegel 2011; Liu et al. 2013), and combine information from different layers via some form of regularization. An alternative category of methods relies on fitting probabilistic generative models. Various extensions of single-layer network models to multi-layer settings have been proposed, including multi-layer stochastic block models (Han, Xu, and Airoldi 2015; Stanley et al. 2016; Valles-Catala et al. 2016; Paul and Chen 2016), multi-layer mixed-membership stochastic block model (De Bacco et al. 2017), Poisson Tucker decomposition models (Schein et al. 2015, 2016), and Bayesian latent factor models (Jenatton et al. 2012), among others. However, the theoretical understanding of multi-layer community detection methods has been rather limited. Under multi-layer stochastic block models, Han, Xu, and Airoldi (2015) proved the consistency of maximum likelihood estimation (MLE) as the number of layers goes to infinity and the number of nodes is fixed. Paul and Chen (2016) established the consistency of MLE under much more general conditions that allow both the numbers of nodes and layers to grow. The authors further obtained the minimax rate over a large parameter space. Paul and Chen (2020) derived the asymptotic results for several spectral and matrix factorization based methods in the high-dimensional setting where the numbers of layers, nodes, and communities can all diverge. Bhattacharyya and Chatterjee (2018) proposed a spectral clustering method that can consistently detect communities even if each layer of the network is extremely sparse. They then generalized their method and results to multi-layer degree-corrected block models.

In this article, we consider spectral clustering after convex layer aggregation, a two-step framework for multi-layer network community detection. Our contribution is 2-fold. First, with a sharp asymptotic characterization of the misclustering error, we reveal the impact of a given convex aggregation on the community detection performance. This motivates us to develop two adaptive aggregation methods that can effectively use community structure information from different layers. Second, our study of the spectral embedding suggests using clustering with Gaussian mixture models as a substitute for the commonly adopted k-means in spectral clustering. Together, these two proposed recipes strengthen the two-step procedure to be an efficient community detection approach that outperforms several popular methods, especially for networks with heterogeneous layers. Throughout the article, our treatment will be mainly focused on networks with assortative community structures in all layers. We discuss the application of our proposed methods to networks of a mixed community structure (with both assortative and disassortative structures) in Section 4. We refer the reader to Bhattacharyya and Chatterjee (2020), Paul and Chen (2020), Lei, Chen, and Lynch (2020), and Lei and Lin (2022) for recent developments toward effectively combining both assortative and disassortative community structures in multi-layer networks. We should also point out that our asymptotic analysis considers the partial recovery regime (Abbe 2017) where the node degrees diverge to infinity sufficiently fast while the gap between the within and between community probabilities remains small. Such asymptotics yields precise error characterization that motivates the proposed approaches. We discuss this asymptotic regime in detail in Section 2.2.

2. Detecting Communities in Multi-Layer Networks using New Spectral Methods

2.1. Definitions and Problem Statement

We focus on undirected networks throughout the article. The observed edge information of a single-layer network with n nodes can be represented by the symmetric adjacency matrix $A = (A_{ij}) \in \{0,1\}^{n\times n}$, where $A_{ij} = A_{ji} = 1$ if and only if there exists a connection between nodes i and j. Suppose the network can be divided into K nonoverlapping communities, and let $\vec{c} = (c_1, \ldots, c_n)^T$ be latent community membership vector corresponding to nodes $1, \ldots, n$, taking values in $[K] := \{1, \ldots, K\}$. Arguably, the most studied network model for community detection is the stochastic block model (SBM) (Holland, Laskey, and Leinhardt 1983).

Definition 1 (Stochastic Block Model). The latent community labels $\{c_i\}_{i=1}^n$ are independently sampled from a multinomial distribution, that is, for $i \in [n]$ and $k \in [K]$, $\mathrm{pr}(c_i = k) = \pi_k$ with constraint $\sum_{k=1}^K \pi_k = 1$, $0 < \pi_k < 1$. Define a symmetric connectivity matrix $\Omega = (\Omega_{ab}) \in (0,1)^{K\times K}$. Conditioning on $\vec{c}$, the adjacency matrix A has independent entries with $A_{ij} \sim$Bernoulli$(\Omega_{c_i c_j})$ for all $i \le j$. We denote the model by $A \sim \mathrm{SBM}(\Omega, \vec{\pi})$.

Under SBM, the distribution of the edge between nodes i and j only depends on their community assignments c_i and c_j. Nodes from the same community are stochastically equivalent. The parameters $\vec{\pi} = (\pi_1, \ldots, \pi_K)^T$ control sizes of the K communities. We call the network balanced when $\pi_1 = \cdots = \pi_K = 1/K$. The symmetric matrix Ω represents the connectivity probabilities among the communities. A special case of SBM that has been widely studied in theoretical computer science is called the planted partition model (Bui et al. 1987; Dyer

and Frieze 1989), where the values of the probability matrix Ω are one constant on the diagonal and another constant off the diagonal.

Definition 2 (Planted Partition Model). A *planted partition model (PPM)* is a special homogeneous SBM, of which the connectivity matrix is $\Omega = (p-q)I_K + qJ_K \in (0,1)^{K\times K}$, where I_K is the identity matrix and J_K is the matrix of ones. This means the within-community connectivity probabilities of PPM are all p while the between-community probabilities are all q. Given that our focus is on assortative networks, we assume $p > q$ throughout this article. The model is written as $A \sim \text{PPM}(p, q, \overrightarrow{\pi})$.

In this article, we consider a multi-layer network of L layers to be a collection of L single-layer networks that share the same nodes but with different edges. For each $\ell \in [L]$, the adjacency matrix $A^{(\ell)} = (A_{ij}^{(\ell)}) \in \{0,1\}^{n\times n}$ represents edge information from the ℓth layer. The multi-layer stochastic block model (MSBM) (Han, Xu, and Airoldi 2015; Paul and Chen 2016; Bhattacharyya and Chatterjee 2018) is a natural extension of the standard SBM to the multi-layer case.

Definition 3 (Multi-layer Stochastic Block Model). The layers of a multi-layer stochastic block model share the common community assignments $\overrightarrow{c} = (c_1, \ldots, c_n)^T \in [K]^n$ which are independently sampled from a multinomial distribution with parameters $\overrightarrow{\pi} = (\pi_1, \ldots, \pi_K)^T$. Conditioning on $\overrightarrow{c}$, all the adjacency matrices have independent entries with $A_{ij}^{(\ell)} \sim \text{Bernoulli}(\Omega_{c_i c_j}^{(\ell)})$ for all $\ell \in [L]$ and $i \le j$. We write the model as $A^{[L]} \sim \text{MSBM}(\Omega^{[L]}, \overrightarrow{\pi})$ for short.

Under MSBM, each layer follows an SBM with consensus community assignments $\overrightarrow{c}$, but with possibly different connectivity patterns as characterized by the set of parameters $\Omega^{[L]} = \{\Omega^{(\ell)}\}_{\ell=1}^L$. A multi-layer planted partition model is a special type of MSBM when each layer follows a planted partition model.

Definition 4 (Multi-layer Planted Partition Model). A multi-layer planted partition model (MPPM) is a special MSBM with $A^{(\ell)} \sim \text{PPM}(p^{(\ell)}, q^{(\ell)}, \overrightarrow{\pi})$ for each $\ell \in [L]$, that is, $\Omega^{(\ell)} = (p^{(\ell)} - q^{(\ell)})I_K + q^{(\ell)}J_K \in (0,1)^{K\times K}$, $\ell \in [L]$. The model is written as $A^{[L]} \sim \text{MPPM}(p^{[L]}, q^{[L]}, \overrightarrow{\pi})$.

We consider the following two-step framework for multi-layer community detection:

1. *Convex layer aggregation*. Form the weighted adjacency matrix $A^{\overrightarrow{w}} = \sum_{\ell=1}^L w_\ell A^{(\ell)}$, for some $\overrightarrow{w} \in \mathcal{W} = \{\overrightarrow{w} : \sum_{\ell=1}^L w_\ell = 1, w_\ell \ge 0, \ell \in [L]\}$.
2. *Spectral clustering*. Suppose the spectral decomposition of $A^{\overrightarrow{w}}$ is given by $A^{\overrightarrow{w}} = \sum_{i=1}^n \lambda_i u_i u_i^T$, where $\{\lambda_i\}_{i=1}^n$ are the eigenvalues and $\{u_i\}_{i=1}^n$ are the corresponding orthonormal eigenvectors. The eigenvalues are ordered in magnitude so that $|\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_n|$. Form the eigenvector matrix $U = (u_1, u_2, \ldots, u_K) \in \mathcal{R}^{n\times K}$. Treat each row of U as a data point in $\mathcal{R}^K$ and run the k-means clustering on the n data points. The cluster label outputs from the k-means are the community membership estimates for the n nodes.

In the first step, the community structure information from different layers is integrated by a simple convex aggregation of the adjacency matrices. The second step runs spectral clustering on the weighted adjacency matrix. A similar two-step spectral method has been considered in Chen and Hero (2017), where the authors presented a phase transition analysis of clustering reliability. With a distinctly different focus, we will provide novel solutions to refine the framework toward a better community detection approach. In the case when $\overrightarrow{w} = (1/L, \ldots, 1/L)$, the two-step procedure is considered as a generally effective baseline method in the literature (Tang, Lu, and Dhillon 2009; Kumar, Rai, and Daume 2011; Dong et al. 2012; Bhattacharyya and Chatterjee 2018; Paul and Chen 2020). However, it is common that different levels of signal-to-noise ratios exist across the layers; hence, aggregation with equal weights may not be the optimal choice. Thusly motivated, we propose two approaches that can adaptively choose a favorable weight vector $\overrightarrow{w} \in \mathcal{W}$ leading to superior community detection results. The two methods are simple and intuitive while having strong theoretical motivations. We provide analytical calculations to reveal that the two approaches, in fact, are estimating the optimal weight vector that minimizes the misclustering error (see formal definition of the error in Section 2.2) under balanced multi-layer planted partition models. Moreover, we present convincing arguments to show that, the k-means clustering in the spectral clustering step should be replaced by clustering using Gaussian mixture models to effectively capture the shape of spectral embedded data thus yielding improved community detection performances. In a nutshell, the proposed weight selection and change of the clustering method make the two-step framework a highly competitive multi-layer community detection procedure, as will be demonstrated by extensive numerical experiments in Sections 3.1 and 3.2.

2.2. Asymptotic Misclustering Error under Balanced MPPM

We first provide some asymptotic characterization of the two-step framework introduced in Section 2.1. The asymptotic analysis paves the way to develop two adaptive layer aggregation methods in Sections 2.3 and 2.4. Toward this end, let $\delta : [K] \to [K]$ denote a permutation of $[K]$, and $\overrightarrow{c} = (c_1, \ldots, c_n)^T \in [K]^n$ be the latent community assignments. The misclustering error for a given community estimator $\hat{\overrightarrow{c}} = (\hat{c}_1, \ldots, \hat{c}_n)^T$ is defined as $r(\hat{\overrightarrow{c}}) = \inf_\delta \sum_{i=1}^n 1(\delta(\hat{c}_i) \ne c_i)/n$, which is the proportion of nodes that are misclustered, modulo permutations of the community labels. We consider a sequence of balanced multi-layer planted partition models indexed by the network size n: $A_n^{[L]} \sim \text{MPPM}(p_n^{[L]}, q_n^{[L]}, \overrightarrow{\pi})$ with both K and L fixed. For a given weight $\overrightarrow{w} \in \mathcal{W}$, we introduce the following quantity that plays a critical role in the error characterization:

$$\tau_n^{\overrightarrow{w}} = \frac{n\big[\sum_{\ell=1}^L w_\ell (p_n^{(\ell)} - q_n^{(\ell)})\big]^2}{\sum_{\ell=1}^L w_\ell^2 \big[p_n^{(\ell)}(1-p_n^{(\ell)}) + (K-1)q_n^{(\ell)}(1-q_n^{(\ell)})\big]}. \quad (1)$$

Theorem 1. Recall the two-step procedure in Section 2.1. For a given weight $\overrightarrow{w} \in \mathcal{W}$, let $\hat{\overrightarrow{c}}_{\overrightarrow{w}}$ be the corresponding community estimator. Assume the following conditions

(i) $\sum_{\ell=1}^{L} w_\ell^2[p_n^{(\ell)}(1-p_n^{(\ell)}) + (K-1)q_n^{(\ell)}(1-q_n^{(\ell)})] = \Omega(n^{-1}\log^4 n)$,
(ii) $[\sum_{\ell=1}^{L} w_\ell^2 p_n^{(\ell)}(1-p_n^{(\ell)})] \cdot [\sum_{\ell=1}^{L} w_\ell^2 q_n^{(\ell)}(1-q_n^{(\ell)})]^{-1} \to 1$,
(iii) $\tau_\infty^{\vec{w}} \equiv \lim_{n\to\infty} \tau_n^{\vec{w}} < \infty$,

where for two sequences a_n and b_n, $a_n = \Omega(b_n)$ means that $\limsup_{n\to\infty} |a_n/b_n| > 0$. It holds that for $\tau_\infty^{\vec{w}} \in (K, \infty)$, as $n \to \infty$,

$$E[r(\hat{\vec{c}}, \vec{c}_{\vec{w}})] \to 1 - \mathrm{pr}(a_i \geq 0, i = 1, 2, \ldots, K-1), \quad (2)$$

where $\vec{a} = (a_1, \ldots, a_{K-1})^T \sim \mathcal{N}(\vec{\mu}, \Sigma)$ with $\vec{\mu} = \sqrt{\tau_\infty^{\vec{w}} - K} \cdot (1, 1, \ldots, 1)^T, \Sigma = I_{K-1} + J_{K-1}$. Moreover, the asymptotic error $1 - \mathrm{pr}(a_i \geq 0, i = 1, 2, \ldots, K-1)$ is a strictly monotonically decreasing function of $\tau_\infty^{\vec{w}}$ over (K, ∞).

Remark 1. In light of the asymptotic error characterization in Theorem 1, $\tau_\infty^{\vec{w}}$ can be interpreted as the signal-to-noise ratio (SNR) for the two-step procedure with a given weight $\vec{w}$. The assumption $\tau_\infty^{\vec{w}} > K$ is critical. As will be clear from Proposition 2, if the SNR is below the threshold K, the informative eigenvalues and eigenvectors of the weighted adjacency matrix $A^{\vec{w}}$ cannot be separated from the noisy ones. As a result, the spectral clustering in the second step will completely fail. Some supporting simulations are shown in Figure 1. As the left plot demonstrates, the two-step method performs like random guess when $\tau_n^{\vec{w}}$ is smaller than the cutoff point $K = 2$.

Remark 2. A general scenario where the conditions in Theorem 1 will hold for all $\vec{w} \in \mathcal{W}$ is $p_n^{(\ell)} = \Omega(n^{-1}\log^4 n), n(p_n^{(\ell)} - q_n^{(\ell)})^2 (p_n^{(\ell)})^{-1} = \Theta(1)$. This implies that $p_n^{(\ell)} - q_n^{(\ell)} \ll q_n^{(\ell)} \propto p_n^{(\ell)}$, that is, the gap between within-community and between-community connectivity probabilities is of smaller order compared to the connectivity probabilities themselves. This is in marked contrast to the equal order assumption that is typically made in the statistical community detection literature (Zhao 2017; Abbe 2017). The minimax rate result in Zhang and Zhou (2016) shows that for finite number of communities, the sufficient and necessary condition for consistent single-layer community detection is $n(p_n^{(\ell)} - q_n^{(\ell)})^2 (p_n^{(\ell)})^{-1} \to \infty$. As a result, under our conditions the misclustering error will not vanish asymptotically even though the sequence of networks are sufficiently dense $p_n^{(\ell)} = \Omega(n^{-1}\log^4 n)$. We believe this asymptotic set-up where community detection consistency is unattainable, is a more appropriate analytical platform to understand real large-scale networks which are dense but not necessarily have strong community structure signals. Furthermore, it enables us to obtain the *asymptotically* exact error formula that reveals the precise impact of the weight $\vec{w}$ on community detection.

Remark 3. Because of the distinct asymptotic regime as explained in the last paragraph, existing works on the asymptotic properties of eigenvectors of random matrices (Tang and Priebe 2018; Cape, Tang, and Priebe 2019; Fan et al. 2019; Abbe et al. 2020) cannot be directly applied or adapted to the current setting. Our asymptotic analysis is motivated by the study of eigenvectors of large Wigner matrices in Bai and Pan (2012). A similar asymptotic setting to ours was adopted in Deshpande, Abbe, and Montanari (2017) and Deshpande et al. (2018). However, notably different from our study, these two works focused on characterizing the information-theoretical limit of community detection for single-layer networks via message passing or belief propagation algorithms. We should also point out that community detection inconsistency can also arise for sparse networks where the degree of nodes remains bounded (see Decelle et al. 2011; Krzakala et al. 2013; Abbe 2017 and references therein). Nevertheless, to our best knowledge, the *asymptotically* exact error characterization in this regime is largely unknown.

Remark 4. We conduct a small simulation to evaluate the results of Theorem 1. As is clear from the left panel of Figure 1, the asymptotic error is a rather accurate prediction of the finite-sample error over a wide range of SNR when the network size is large. Moreover, using the parameter values specified in the caption of Figure 1, it is straightforward to compute $\tau_n^{(1,0)^T} = 0.64, \tau_n^{(0,1)^T} = 9.07$. Hence, the second layer is much more informative than the first one. The right panel reveals that the optimal weight is located around $\vec{w} = (0.2, 0.8)^T$. It is

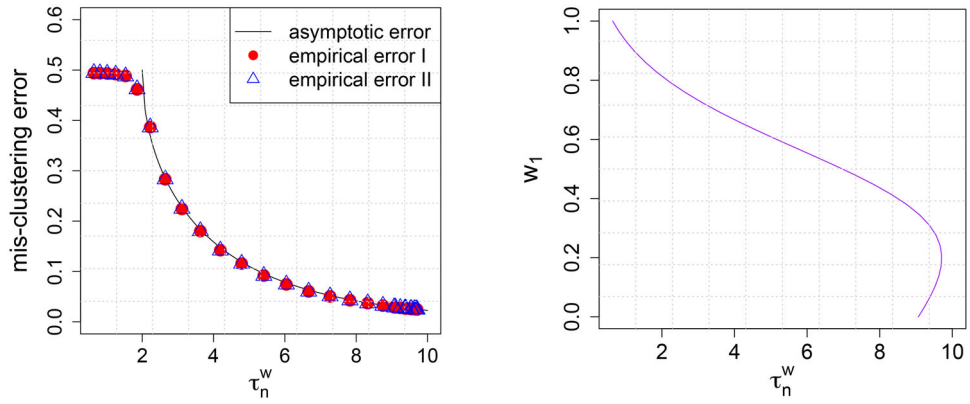


Figure 1. Consider a balanced multi-layer planted partition model with $K = 2, L = 2, n = 6000, p^{(1)} = 0.02, q^{(1)} = 0.018, p^{(2)} = 0.02, q^{(2)} = 0.013$. Left panel: "asymptotic error" is calculated according to the limit formula (2) in Theorem 1; "empirical error I" denotes the finite-sample error of the two-step procedure described in Section 2.1; "empirical error II" represents the finite-sample error of the modified two-step procedure with *k*-means replaced by clustering using Gaussian mixture models; both empirical errors are calculated for the procedures without making use of the oracle information of parameters. Finite-sample errors are the averages over five repetitions. Right panel: the y-axis refers to the first component of the weight vector.

interesting to observe that although the first layer alone acts like random noise to spectral clustering (since $\tau_n^{(1,0)^T} < K = 2$; see also the left panel), appropriately combined with the second layer it provides useful community structure information that contributes to a decent performance boost compared to the result solely based on the second layer. We, therefore, see the importance of weight tuning in the two-step procedure.

2.3. Iterative Spectral Clustering

We are in the position to introduce the first adaptive layer aggregation method. Theorem 1 reveals that the asymptotic misclustering error depends on $\tau_\infty^{\vec{w}}$ in a strictly monotonically decreasing fashion. Hence, the optimal $\vec{w}$ that will minimize the asymptotic error can be found by solving $\max_{\vec{w}\in\mathcal{W}} \tau_\infty^{\vec{w}}$. Naturally, in the realistic finite-sample scenario where a multi-layer network with n nodes is given, we would like to use the weight vector that maximize $\tau_n^{\vec{w}}$.

Proposition 1. Recall $\tau_n^{\vec{w}}$ defined in (1). Denote $\vec{w}^* = (w_1^*, \ldots, w_L^*)^T = \arg\max_{\vec{w}\in\mathcal{W}} \tau_n^{\vec{w}}$. Then $\vec{w}^*$ exists and is unique, admitting the explicit expression:

$$w_l^* \propto \frac{p_n^{(l)} - q_n^{(l)}}{p_n^{(l)}(1-p_n^{(l)}) + (K-1)q_n^{(l)}(1-q_n^{(l)})}, \quad \text{for } l \in [L], \tag{3}$$

where a normalization constant is taken to ensure $\vec{w}^* \in \mathcal{W}$.

Formula (3) has a simple and intuitive interpretation. Each w_ℓ^* is determined by the layer's own parameters. It measures the standardized difference between the within and between community connectivity probability, where the standardization (ignoring the smaller order terms $(p_n^{(\ell)})^2, (q_n^{(\ell)})^2$ in the denominator) is essentially the averaged connectivity probability over all pairs of nodes. A larger standardized difference implies a stronger community structure signal, deserving a larger weight on the corresponding layer.

The weight vector $\vec{w}^*$ cannot be directly used in the two-step procedure as it depends on unknown parameters $\{(p_n^{(\ell)}, q_n^{(\ell)})\}_{\ell=1}^L$. To address this issue, we propose an iterative spectral clustering (ISC) method. For each $\ell \in [L]$, we first run spectral clustering on the layer's own adjacency matrix $A^{(l)}$ to obtain an initial community estimator $\hat{\vec{c}}^{(\ell)}$ and compute the weight estimates:

$$\hat{w}_\ell^* \propto \frac{\hat{p}_n^{(\ell)} - \hat{q}_n^{(\ell)}}{\hat{p}_n^{(\ell)}(1-\hat{p}_n^{(\ell)}) + (K-1)\hat{q}_n^{(\ell)}(1-\hat{q}_n^{(\ell)})}, \tag{4}$$

where

$$\hat{p}_n^{(\ell)} = \frac{\sum_{ij} A_{ij}^{(\ell)} 1(\hat{c}_i^{(\ell)} = \hat{c}_j^{(\ell)})}{\sum_{ij} 1(\hat{c}_i^{(\ell)} = \hat{c}_j^{(\ell)})}, \quad \hat{q}_n^{(\ell)} = \frac{\sum_{ij} A_{ij}^{(\ell)} 1(\hat{c}_i^{(\ell)} \neq \hat{c}_j^{(\ell)})}{\sum_{ij} 1(\hat{c}_i^{(\ell)} \neq \hat{c}_j^{(\ell)})}. \tag{5}$$

We then run spectral clustering on the weighted adjacency matrix $A^{\hat{\vec{w}}^*} = \sum_{l=1}^L \hat{w}_l^* A^{(l)}$ to obtain a refined community estimate $\hat{\vec{c}}$. Such refinement can be repeatedly applied until convergence. We summarize the outlined method as Algorithm 1.

Algorithm 1 Iterative spectral clustering [ISC].

Require: L layers of adjacency matrices $[A^{(l)}, l = 1, \ldots, L]$, the number of communities K, and the precision parameter ϵ_0.

Ensure: $\hat{\vec{w}}_{\text{new}}$ and the community estimate $\hat{\vec{c}}$ by apply spectral clustering on $A^{\hat{\vec{w}}_{\text{new}}}$.

1: Initialization: Apply spectral clustering on every single $A^{(l)}$, and compute the initial weight estimates according to (4) and (5). Denote it by $\hat{\vec{w}}_{\text{old}}$. Set $\epsilon = \epsilon_0 + 1$.
2: **while** $\epsilon > \epsilon_0$ **do**
3: Apply spectral clustering on $A^{\hat{\vec{w}}_{\text{old}}}$ and compute updated weights $\hat{\vec{w}}_{\text{new}}$ as in (4) and (5).
4: Assign $\epsilon \leftarrow \|\hat{\vec{w}}_{\text{old}} - \hat{\vec{w}}_{\text{new}}\|$ and $\hat{\vec{w}}_{\text{old}} \leftarrow \hat{\vec{w}}_{\text{new}}$.
5: **end while**

Remark 5. Algorithm 1 is motivated by the asymptotic analysis of the misclustering error under balanced multi-layer planted partition models. However, as will be demonstrated by extensive numerical experiments in Sections 3.1 and 3.2, it works well for a much larger family of multi-layer stochastic block models. An intuitive explanation is that for general stochastic block models, (5) is estimating the averaged within and between community probabilities; and the weight formula (4) represents a certain normalized gap between the aforementioned two probabilities which can be considered as a measure of the community signal strength, thus, providing useful aggregation information. That being said, the weight formula (4) is not necessarily estimating the optimal weights for general multi-layer stochastic block models. Deriving the optimal weight formulas in such a general setting is an interesting and important future research. On a related note, since our focus is on the partial recovery regime (see Section 2.2 for detailed discussions), it would be interesting to investigate how well the optimal weight is estimated under MPPM cases. To our best knowledge, in the partial recovery regime when the node degrees diverge with n, the fundamental limits for parameter estimation have not been established in the literature. We defer a thorough evaluation and discussion of Algorithm 1 to Sections 3.1 and 3.2.

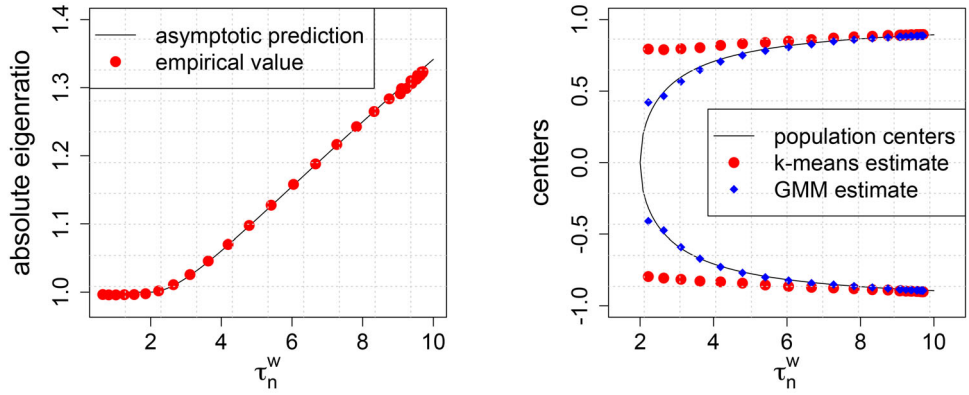
2.4. Spectral Clustering with Maximal Eigenratio

We now present the second method to select the weight $\vec{w}$. Let $\lambda_i^{\vec{w}}$ be the ith largest (in magnitude) eigenvalue of the weighted adjacency matrix $A^{\vec{w}}$. The spectral clustering in the two-step procedure is implemented using the eigenvectors corresponding to the first K eigenvalues $\{\lambda_i^{\vec{w}}, i = 1, \ldots, K\}$. We will show that in addition to these eigenvectors, the eigenvalues of $A^{\vec{w}}$ can be used for community detection. In particular, the eigenvalue ratio $\lambda_K^{\vec{w}}/\lambda_{K+1}^{\vec{w}}$ holds critical information about how $\vec{w}$ affects misclustering error, as shown in following proposition.

Proposition 2. Under the same conditions of Theorem 1, it holds that

$$\frac{|\lambda_K^{\vec{w}}|}{|\lambda_{K+1}^{\vec{w}}|} \xrightarrow{a.s.} \begin{cases} \frac{1}{2}\left(\sqrt{\frac{\tau_\infty^{\vec{w}}}{K}} + \sqrt{\frac{K}{\tau_\infty^{\vec{w}}}}\right), & \text{if } \tau_\infty^{\vec{w}} > K, \\ 1, & \text{if } \tau_\infty^{\vec{w}} \leq K. \end{cases}$$

Here, we have suppressed the dependence of $\lambda_K^{\vec{w}}$ and $\lambda_{K+1}^{\vec{w}}$ on n to simplify the notation.



(i) Absolute Eigenratio (ii) Centers

Figure 2. The same setting as in Figure 1. (i). The asymptotic prediction is calculated via the limit in Proposition 2. The empirical value is computed over one repetition without using any oracle information of the parameters. (ii). Proposition 3 shows that the first eigenvector is uninformative for clustering. We thus focus on the second one. There are two population centers that are calculated according to Proposition 3. The center estimates from *k*-means ("*k*-means estimate") and MLE under Gaussian mixture distribution ("GMM estimate") are computed over five repetitions.

Proposition 2 reveals that the absolute eigenratio $|\lambda_K^{\vec{w}}|/|\lambda_{K+1}^{\vec{w}}|$ undergoes a phase transition: it remains a constant one when the SNR $\tau_\infty^{\vec{w}}$ is smaller than K; the ratio will be strictly increasing in $\tau_\infty^{\vec{w}}$ once $\tau_\infty^{\vec{w}} > K$. This phenomenon is consistent with Theorem 1. Indeed, when $\tau_\infty^{\vec{w}}$ is below the threshold K, the informative eigenvalue $\lambda_K^{\vec{w}}$ is indistinguishable from the noisy one $\lambda_{K+1}^{\vec{w}}$. Spectral clustering on $A^{\vec{w}}$ will thus fail. See Remark 1 for more details. On the other hand, as the SNR $\tau_\infty^{\vec{w}}$ increases over the range (K, ∞), the eigenratio becomes larger so that $\lambda_K^{\vec{w}}$ is better separated from $\lambda_{K+1}^{\vec{w}}$ and the misclustering error of spectral clustering on $A^{\vec{w}}$ is decreased. Figure 2(a) depicts both the finite-sample and asymptotic values of the eigenratio from the same simulation study as described in Figure 1. Clearly the asymptotic values are fine predictions of the empirical ones.

Proposition 2 together with Theorem 1 tells us that asymptotically the optimal weight achieving the minimum misclustering error maximizes the absolute eigenratio $|\lambda_K^{\vec{w}}|/|\lambda_{K+1}^{\vec{w}}|$. This motivates us to maximize the eigenratio to obtain the weight. Define the objective function $g(\vec{w}) \equiv (\lambda_K^{\vec{w}})^2/(\lambda_{K+1}^{\vec{w}})^2$. We aim to solve the optimization problem: $\max_{\vec{w}\in\mathcal{W}} g(\vec{w})$.

When the eigenvalues $\lambda_K^{\vec{w}}$ and $\lambda_{K+1}^{\vec{w}}$ are simple (in magnitude), it is well known that they are differentiable at $A^{\vec{w}}$ and admit closed-form gradients (Magnus 1985). Using the chain rule, we can derive the gradient of $g(\vec{w})$: for each $\ell \in [L]$

$$\begin{aligned}\frac{\partial g(\vec{w})}{\partial w_\ell} &= 2\lambda_K^{\vec{w}}(\lambda_{K+1}^{\vec{w}})^{-3}(\lambda_{K+1}^{\vec{w}}\nabla\lambda_K^{\vec{w}} - \lambda_K^{\vec{w}}\nabla\lambda_{K+1}^{\vec{w}}) \\ &= 2\lambda_K^{\vec{w}}(\lambda_{K+1}^{\vec{w}})^{-3}(\lambda_{K+1}^{\vec{w}}u_K^T A^{(\ell)}u_K - \lambda_K^{\vec{w}}u_{K+1}^T A^{(\ell)}u_{K+1}),\end{aligned}$$

where u_K and u_{K+1} are the eigenvectors associated with the eigenvalues $\lambda_K^{\vec{w}}, \lambda_{K+1}^{\vec{w}}$, respectively. Hence, we can perform projected gradient descent to update $\vec{w}$ using

$$\vec{w}_{t+1} = \mathcal{P}_{\mathcal{W}}\left(\vec{w}_t + \gamma_t \nabla g(\vec{w}_t)\right), \tag{6}$$

where $\mathcal{P}_{\mathcal{W}}(\cdot)$ denotes the projection onto unit simplex; $\gamma_t = \gamma_0/(1+rt)$ is the learning rate that decays with time, r is the decay rate and t is the number of iterations. The above update is only feasible when $\lambda_K^{\vec{w}}$ and $\lambda_{K+1}^{\vec{w}}$ are simple. Nevertheless, we found empirical evidence that $A^{\vec{w}}$ does not have repeated eigenvalues (in magnitude) for a wide range of $\vec{w}$. In fact, it has been proved that certain types of random matrices have simple spectrum with high probability (Tao and Vu 2017; Luh and Vu 2018). For completeness, to handle the rare scenario when $\lambda_K^{\vec{w}}$ or $\lambda_{K+1}^{\vec{w}}$ is not simple, we resort to coordinate descent update with one-dimensional line search. Whenever $\lambda_K^{\vec{w}}$ and $\lambda_{K+1}^{\vec{w}}$ become simple at the current update, the projected gradient descent is resumed. We implement the algorithm with random initialization. To achieve a better convergence, we run it independently multiple times and choose the output weight that gives the largest value of $g(\vec{w})$. The method is summarized as Algorithm 2.

Algorithm 2 Spectral clustering with maximal eigenratio [SCME].

Require: L layers of adjacency matrices $[A^{(l)}, l = 1, \ldots, L]$, the number of communities K, initial learning rate γ_0, decay rate r, maximum number of iterations T, number of random initializations M, and the precision parameter ϵ_0.

Ensure: $\hat{\vec{w}}$ and the community estimate $\hat{\vec{c}}$ by apply spectral clustering on $A^{\hat{\vec{w}}}$.

1: Initialize $m = 1$.
2: **while** $m \le M$ **do**
3: Obtain random initialization and denote it by $\hat{\vec{w}}_{\text{old}}$. Set $\epsilon = \epsilon_0 + 1$ and $t = 1$.
4: **while** $\epsilon > \epsilon_0$ and $t \le T$, **do**
5: Compute the update $\hat{\vec{w}}_{\text{new}}$
$$\begin{cases}\text{using (6)}, & \text{if } \lambda_K^{\vec{w}_{\text{old}}} \text{ and } \lambda_{K+1}^{\vec{w}_{\text{old}}} \text{ are simple,}\\ \text{via coordinate descent}, & \text{otherwise.}\end{cases}$$
6: Assign $\epsilon \leftarrow \|\hat{\vec{w}}_{\text{old}} - \hat{\vec{w}}_{\text{new}}\|$, $\hat{\vec{w}}_{\text{old}} \leftarrow \hat{\vec{w}}_{\text{new}}$, and $t \leftarrow t + 1$.
7: **end while**
8: Set $\hat{\vec{w}}^m_{\text{new}} = \hat{\vec{w}}_{\text{new}}$ and $m \leftarrow m + 1$.
9: **end while**
10: Set $\hat{\vec{w}} \leftarrow \hat{\vec{w}}^{m^*}_{\text{new}}$ where $m^* = \arg\max_m g(\hat{\vec{w}}^m_{\text{new}})$.

Remark 6. This second adaptive layer aggregation method has an intuitive explanation as well. For networks with K communities, the ratio between the last informative eigenvalue λ_K and the first noisy eigenvalue λ_{K+1} is a reasonable measure of the community structure signal strength. Such intuition is well pronounced under the balanced MPPM: the maximization of the eigen-ratio leads to the optimal layer aggregation. Extensive numerical studies in Section 3 will further demonstrate the robustness and effectiveness of the method working beyond balanced MPPM.

2.5. k-means Clustering Versus Gaussian Mixture Model Clustering

We have presented two novel methods tailored for adaptive layer aggregation. We now turn to discuss the spectral clustering step of the two-step framework that is introduced in Section 2.1. Specifically, we will provide convincing evidence to support clustering using Gaussian mixture models (GMM) as a substitute for k-means in spectral clustering. Toward this goal, let $U \in \mathcal{R}^{n\times K}$ be the eigenvector matrix of which the ith column is the eigenvector of $A^{\vec{w}}$ associated with the ith largest (in magnitude) eigenvalue.

Proposition 3. Under the same conditions as in Theorem 1, there exists an orthogonal matrix $\mathcal{O} \in \mathcal{R}^{K\times K}$ such that for any pair $i, j \in [n]$ conditioning on the community labels c_i, c_j, it holds that as $n \to \infty$

$$\begin{pmatrix} \sqrt{n}\mathcal{O}U^T e_i \\ \sqrt{n}\mathcal{O}U^T e_j \end{pmatrix} \xrightarrow{d} \mathcal{N}(\mu, \Sigma), \quad \mu = \begin{pmatrix} \mu^{(c_i)} \\ \mu^{(c_j)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Theta & 0 \\ 0 & \Theta \end{pmatrix},$$

where $\{e_i\}_{i=1}^n$ is the standard basis in $\mathcal{R}^n$. We have omitted the dependence of $\mathcal{O}, U$ on n for simplicity. The mean and covariance matrix in the multivariate Gaussian distribution take the following expressions:

$$\mu^{(c_i)} = \begin{pmatrix} 1 \\ \sqrt{\frac{K(\tau_\infty^{\vec{w}} - K)}{\tau_\infty^{\vec{w}}}} v_{c_i} \end{pmatrix}, \quad \mu^{(c_j)} = \begin{pmatrix} 1 \\ \sqrt{\frac{K(\tau_\infty^{\vec{w}} - K)}{\tau_\infty^{\vec{w}}}} v_{c_j} \end{pmatrix},$$
$$\Theta = \frac{K}{\tau_\infty^{\vec{w}}} \cdot \begin{pmatrix} 0 & 0 \\ 0 & I_{K-1} \end{pmatrix}.$$

Here, the $\mathcal{V}_{K\times K-1} = [v_1, v_2, \ldots, v_K]^T$ such that the $K \times K$ matrix $[1/\sqrt{K}1_K, \mathcal{V}_{K\times K-1}]$ is orthogonal, where 1_K is a length-K vector of 1.

Remark 7. We believe that it is possible to have a nontrivial generalization of our analysis to obtain similar convergence results under more general multi-layer stochastic block models. However, the limiting Gaussian mixture distribution will be much more complicated. Some limiting results have been obtained under single-layer stochastic block models in Tang and Priebe (2018). Levin et al. (2017) derives the central limiting theorem for an omnibus embedding of multiple random graphs for graph comparison inference. However, as explained in Remarks 2 and 3, we are considering a more challenging asymptotic regime, which requires notably different analytical techniques. Given that k-means is invariant under scaling and orthogonal transformation, the vectors $\{\sqrt{n}\mathcal{O}U^T e_i\}_{i=1}^n$ can be considered as the input data points for k-means in spectral clustering. Proposition 3 establishes that asymptotically all the spectral embedded data points follow a Gaussian mixture distribution, and they are pairwise independent. These properties turn out to be sufficient for the convergence result of k-means under classical iid settings in Pollard (1981) to carry over to the current case. With the convergence of estimated centers of k-means, the misclustering error in Theorem 1 can be derived in a direct way. See the proof of Theorem 1 for details.

Remark 8. Given a sample of independent observations from a mixture distribution: $x \mid c = k \sim f(x; \theta_k)$, it is known that the optimal *asymptotic* misclustering error is

$$\min_g \text{pr}(g(x) \neq c) = \text{pr}(g^*(x) \neq c),$$
$$\text{where } g^*(x) = \underset{k}{\text{argmax}}\, \text{pr}(c = k \mid x).$$

It is straightforward to verify that the *asymptotic* error in Theorem 1 coincides with the above optimal error when the mixture distribution is the Gaussian mixture distribution given in Proposition 3. Combining this fact with Remark 7, we can conclude that k-means achieves the best possible outcome and is the proper clustering method to use in spectral clustering (if the conditions of Theorem 1 hold). However, we would like to emphasize that the k-means in spectral clustering, in general, does not consistently estimate the population centers of the Gaussian mixture distribution. Such a phenomenon is well recognized in the classical iid scenario (see, e.g., Bryant and Williamson 1978). It continues to occur in the present case. A formal justification of the statement can be found in Lemma 1 of the Appendix, supplementary materials. Some supportive simulation results are shown in Figure 2(b). Given the inconsistency of k-means for estimating the population centers, it is intriguing to ask why k-means is able to obtain the optimal misclustering error. This is uniquely due to population centers' symmetry and the simple structure of the covariance matrix in this specific Gaussian mixture distribution. Simple calculations show that the k-means error will be optimal as long as the estimated centers (ignoring the irrelevant first coordinate) after a common scaling converge to the population centers. This is verified numerically in Figure 2(b): at each $\tau_n^{\vec{w}}$, the two center estimates from the k-means will be compatible with the population centers after a common shrinkage.

In light of Proposition 3, it is appealing to consider using Gaussian mixture model (GMM) clustering as an alternative in the spectral clustering, since the spectral embedded data follows a Gaussian mixture distribution asymptotically. Indeed, Figures 2(b) and 1 from a simulation study have demonstrated that GMM clustering is able to recover the population centers and obtain the optimal misclustering error. Therefore, both k-means and GMM clustering are optimal clustering methods that can be implemented in spectral clustering under balanced multi-layer planted partition models. However, as discussed in Remark 8, it is the special parameter structures in the Gaussian mixture distribution that make k-means attain optimal error even though it is inconsistent for estimating population centers. Under a general multi-layer stochastic block model, it is likely that the limiting Gaussian mixture distribution (if exists) will

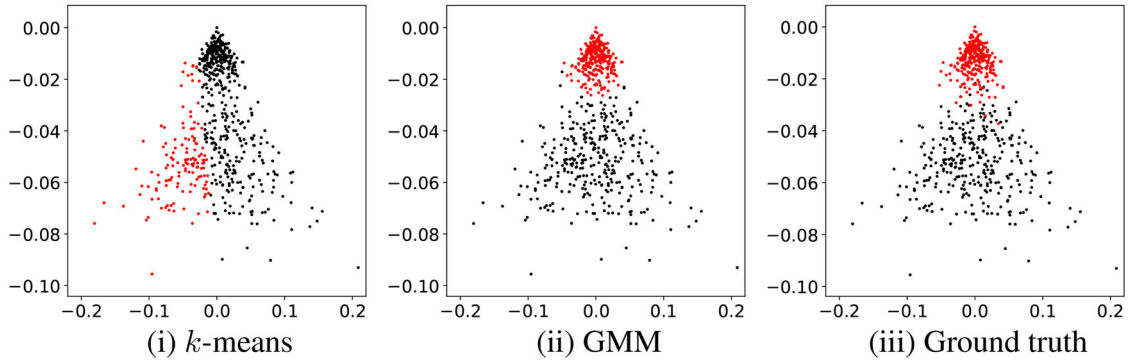


Figure 3. Consider a balanced multi-layer stochastic block model with $K = 2, L = 1, n = 600, \Omega_{11}^{(1)} = 0.053, \Omega_{12}^{(1)} = 0.011, \Omega_{22}^{(1)} = 0.016$. The spectral embedding in $\mathcal{R}^2$ along with the clustering results of k-means, GMM, and ground truth is shown.

have different parameter configurations so that the inconsistency of k-means eventually degrades the clustering performance. In contrast, GMM clustering specifies the asymptotically correct model and is able to better capture various shapes of embedded data. We therefore expect it to outperform k-means in general settings. A quick numerical comparison is shown in Figure 3. We observe that as the model deviates from balanced MPPM, the shapes of mixture components can be fairly different. The k-means fails to identify the shapes, resulting in poor clustering performance, while GMM provides an excellent fit thus gaining significant improvement in clustering. We provide more numerical experiments regarding the comparison in Sections 3.1 and 3.2.

3. Numerical Experiments

We have proposed iterative spectral clustering (ISC) and spectral clustering with maximal eigenratio (SCME) for community detection. With the use of k-means or GMM clustering in spectral clustering, we in fact have four different methods: ISC with k-means clustering [ISC_km], ISC with GMM clustering [ISC_gm], SCME with k-means clustering [SCME_km], and SCME with GMM clustering [SCME_gm]. In this section, we present a systematic numerical study of the community detection performance of the four methods. Moreover, we compare our methods with three standard spectral methods in the literature, which are based on mean adjacency matrix [Mean adj.] (Han, Xu, and Airoldi 2015), aggregate spectral kernel [SpecK] (Paul and Chen 2017) and module allegiance matrix [Module alleg.] (Braun et al. 2015). We use the adjusted rand index (ARI) (Hubert and Arabie 1985) to evaluate the community detection performance of all the methods. ARI is a common measure of the similarity between two data clusterings. It is bounded by 1, with the value of 1 indicating perfect recovery while 0 implying the estimation is no better than a random guess (Steinhaeuser and Chawla 2010). Finally, to verify that our algorithms can reach (nearly) optimal layer aggregation under balanced MPPM, we include the oracle two-step procedure [Grid Search] which uses the weight that minimizes the empirical ARI (over a grid).

3.1. Simulation Results

In the simulations, we consider four different cases corresponding to balanced MPPM, imbalanced MPPM, balanced MSBM, and imbalanced MSBM, respectively. Our methods are motivated by the asymptotic analysis under balanced MPPM. We thus would like to verify their effectiveness under the assumed models. In addition, we aim to study to what extent our methods work well for more general models. Each experiment is repeated 100 times.

Table 1. Balanced MPPM.

Case	n	K	L	c_ρ	$\overrightarrow{p}$	$\overrightarrow{q}$
1a	600	2	2	1.5	(4, 4)	(2, 0–4)
1b	600	2–6	2	1.5	(4, 4)	(0, 3)
1c	600	2	1-5	1.5	(4, . . ., 4)	(0, 4, . . ., 4)
1d	600	2	2	0.4–1.2	(4, 4)	(0.5, 2.5)
1e	200–1000	2	2	0.6	(4, 4)	(1, 3)

NOTE: The notation "s-t" denotes that a parameter is changed over the range $[s, t]$.

1. *Balanced MPPM case.* For a balanced model $A^{[L]} \sim \text{MPPM}(p^{[L]}, q^{[L]}, \overrightarrow{\pi})$, we consider $p^{[L]} \equiv (p^{(1)}, \ldots, p^{(L)})^T = c_\rho \frac{\log n}{n} \overrightarrow{p}, q^{[L]} \equiv (q^{(1)}, \ldots, q^{(L)})^T = c_\rho \frac{\log n}{n} \overrightarrow{q}$. We vary values of the five parameters $n, K, L, c_\rho, \overrightarrow{q}$ in the model to investigate settings with different sample sizes, numbers of communities, numbers of layers, sparsity levels, and connectivity probabilities. Values of model parameters are summarized in Table 1.

   Simulation results are presented in Figure 4. We observe that our four methods show competitive or superior performances compared with the other three consistently under Cases 1a-1e. This lends further strong support to the use of proposed methods. Moreover, the four methods yield uniformly comparable results across all the settings. Such phenomena are consistent with the discussion on the equivalence of k-means and GMM clustering under balanced MPPM in Section 2.5.

   We now discuss each of the scenarios to shed more light on the performance of our methods. In Case 1a, we vary the between-community connectivity probability in the second layer while keeping all other parameters fixed. As this probability increases, the second layer becomes less informative for community detection, so the ARI decreases for all the methods. However, compared with the three existing methods, ISC and SCME are both robust to the increased noises in the second layer, thanks to the adaptive layer aggregation mechanism in the new methods. Figure 4(b) depicts the weight for the first layer selected by ISC_gm, ISC_km, SCME_gm, and

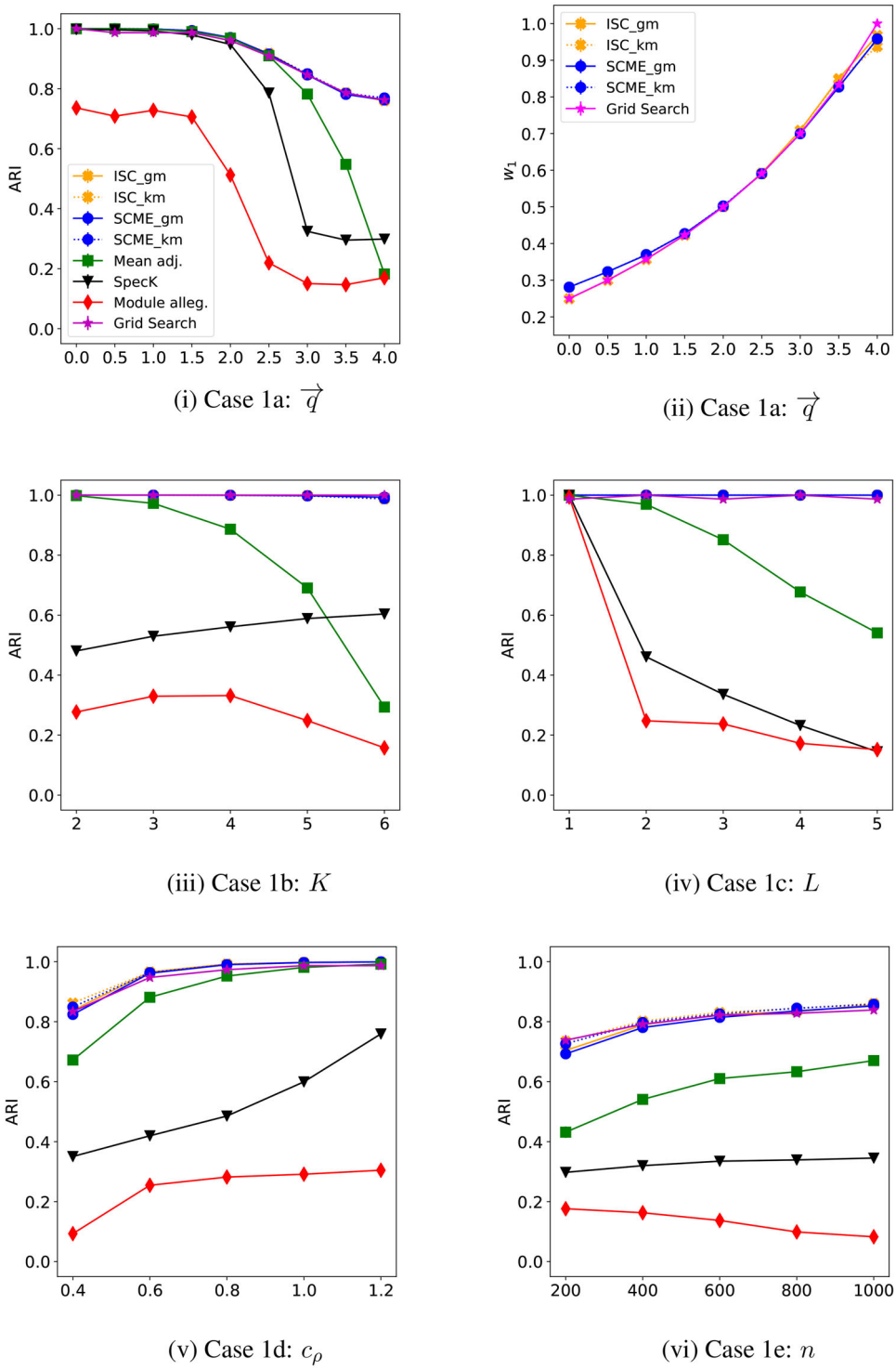


Figure 4. Balanced MPPM case (maximum standard error in the figure is 0.022).

SCME_km, along with the (empirical) optimal weights. It is clear that all the four methods successfully found the (nearly) optimal weights, thus, being able to adaptively down-weigh the second layer when it becomes noisier. Case 1b changes the number of communities. As K varies with everything else fixed, the effective information in each layer for every method tends to change. Take our methods for example. As can be verified from the optimal weight formula (3), a layer with a larger ratio of within-community to between-community connectivity probability should be weighted higher when the number of communities increases. Referring to Figure 4(c), we see that our methods provide stable and decent results due to adaptive layer aggregation, while the other three methods are comparatively sensitive to the change of K. In Case 1c, we fix the first layer as an informative layer and set other layers to be random noises. Figure 4(d) clearly shows that our methods detect communities effectively and are remarkably robust to added noise layers. In contrast, the three existing methods

perform worse with more noise layers. Case 1d generates networks with different sparsity levels by simply scaling all the connectivity probabilities in the model. As the scaling c_ρ is getting larger, the network becomes denser and more informative, so all the methods perform better as seen in Figure 4(e). Our methods outperform Mead adj. for sparse networks, even though they are similar in dense settings. Case 1e examines the performance under different sample sizes. We notice from Figure 4(f) that our methods can be advantageous for moderately small-scale networks.

2. *Imbalanced* MPPM *case*. For an imbalanced $A^{[L]} \sim \text{MPPM}(p^{[L]}, q^{[L]}, \vec{\pi})$, the parameter $\vec{\pi}$ characterizing the size of each community is of primary interest. Adopting the same notation from the balanced MPPM case, we list the parameter values in Table 2. The results are shown in Figure 5. In all scenarios, our four methods outperform other methods. ISC_km (ISC_gm) and SCME_km (SCME_gm) are comparable. One notably different observation from the balanced MPPM case is that GMM clustering leads to superior performances than k-means in both ISC and SCME. This validates our arguments regarding the comparison between k-means and GMM in Section 2.5: when the model deviates from balanced MPPM, GMM clustering is expected to better account for the heterogeneity of spectral embedded data. As illustrated in Figure 5(a), the improvement of GMM over k-means becomes more significant as the communities get more imbalanced. In Case 2b, we keep the communities imbalanced at a given level and change the sparsity level by varying c_ρ. Figure 5(b) shows that SCME is slightly better than ISC in some settings. Our methods outperform Mean adj. by a larger margin for sparser networks.
3. *Balanced* MSBM *case*. We set $\Omega^{[L]} = c_\rho \frac{\log n}{n} \bar{\Omega}$ in the balanced $A^{[L]} \sim \text{MSBM}(\Omega^{[L]}, \vec{\pi})$. Table 3 summarizes the parameter values. We consider two different balanced MSBM scenarios, both varying the sparsity level by changing c_ρ.

Table 2. Imbalanced MPPM.

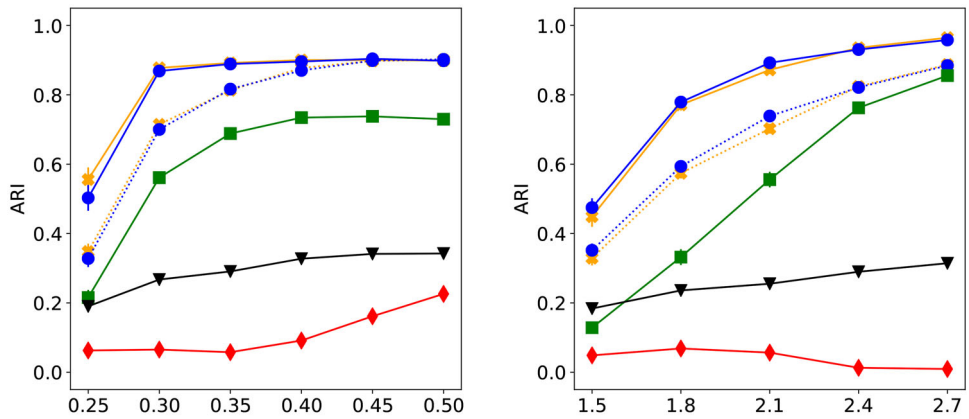
Case	n	K	L	c_ρ	$\vec{p}$	$\vec{q}$	$\vec{\pi}$
2a	600	2	2	2	(4, 4)	(2, 3.5)	(0.25–0.5, 0.75–0.5)
2b	600	2	2	1.5–2.7	(4, 4)	(2, 3.5)	(0.3, 0.7)

In Case 3a, we consider a two-layer MSBM deviating moderately from MPPM. As illustrated in Figure 6(a), the comparison results are similar to what we observe in the balanced MPPM case: our four methods perform alike and are uniformly better than the other three methods. In Case 3b, we make the MSBM adequately different from MPPM. Figure 6(b) shows the results. First, we see that for both ISC and SCME, GMM outperforms k-means. The same outcomes occur in the imbalanced MPPM case. We have well discussed the reason in Section 2.5. Second, ISC_gm (ISC_km) works better than SCME_gm (SCME_km) in sparse settings while the latter wins in the dense cases. The result indicates that eigenratio is more robust to model misspecification as long as the network is sufficiently dense, while the weight update (4) used in ISC is less variable when the network is sparse. We leave a thorough investigation of the distinct impact of sparsity level on ISC and SCME for future research. We also observe that SCME_gm continues to outperform the three standard methods, and ISC_gm does so in all the sparse settings.

4. *Imbalanced* MSBM *case*. We set $\Omega^{[L]} = c_\rho \frac{\log n}{n} \bar{\Omega}$ in the imbalanced $A^{[L]} \sim \text{MSBM}(\Omega^{[L]}, \vec{\pi})$. We list the parameter values in Table 4, where for Case 4a we fix the second community's proportion. Results are shown in Figure 7. As expected, we observe again that GMM performs better than k-means for both ISC and SCME. Case 4a can be seen as an imbalanced continuation of Case 3b at a given c_ρ. Figure 7(a) illustrates the effectiveness of SCME_gm and ISC_gm compared to the three existing methods over a wide range of imbalanced

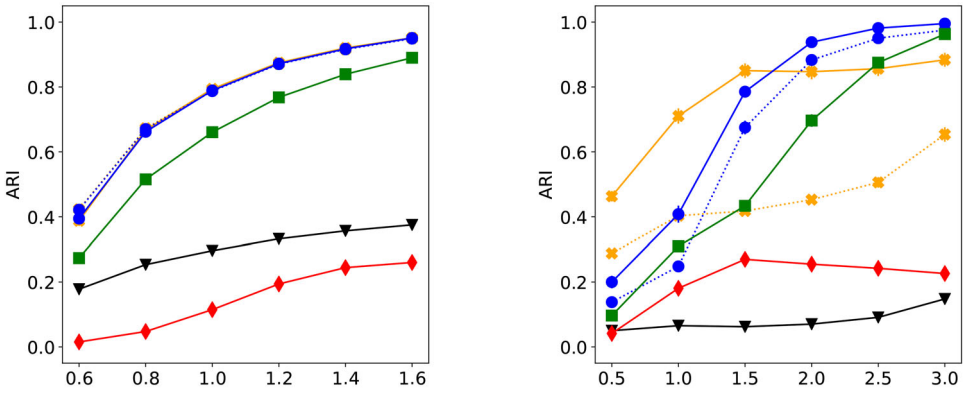
Table 3. Balanced MSBM.

Case	n	K	L	c_ρ	$\bar{\Omega}$
3a	600	2	2	0.6–1.6	$\begin{pmatrix} 5 & 2 \\ 2 & 4 \end{pmatrix}, \begin{pmatrix} 4 & 3.5 \\ 3.5 & 5 \end{pmatrix}$
3b	600	3	5	0.5–3	$\begin{pmatrix} 9 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 9 \end{pmatrix}, \begin{pmatrix} 2 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 2 \end{pmatrix}, 2J_3, 2J_3, 2J_3$



(i) Case 2a: π_1 (ii) Case 2b: c_ρ

Figure 5. Imbalanced MPPM case (maximum standard error in the figure is 0.037).

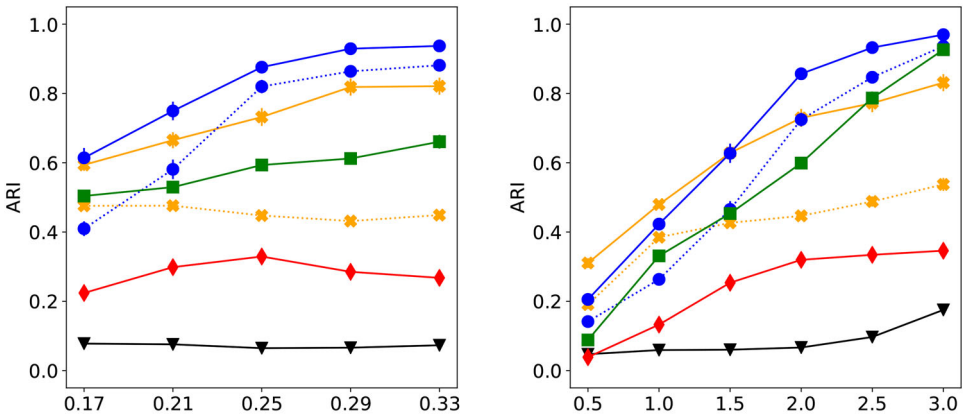


(i) Case 3a: c_ρ (ii) Case 3b: c_ρ

Figure 6. Balanced MSBM case (maximum standard error in the figure is 0.027).

Table 4. Imbalanced MSBM.

Case	n	K	L	c_ρ	$\bar{\Omega}$	$\vec{\pi}$
4a	600	3	5	2	$\begin{pmatrix}9&2&2\\2&2&2\\2&2&9\end{pmatrix}, \begin{pmatrix}2&2&2\\2&4&2\\2&2&2\end{pmatrix}, 2J_3, 2J_3, 2J_3$	(0.17–0.33, 0.33, 0.5–0.34)
4b				0.5–3		(0.25, 0.33, 0.42)



(i) Case 4a: π_1 (ii) Case 4b: c_ρ

Figure 7. Imbalanced MSBM case (maximum standard error in the figure is 0.028).

community sizes. Case 4b is an imbalanced variant of Case 3b. We see from Figure 7(b) that the same comparison outcomes observed in Figure 6(b) for Case 3b remain valid in the imbalanced case.

5. *Summary*. The simulations under balanced MPPM demonstrate that both ISC and SCME are able to select (nearly) optimal weight for layer aggregation and thus achieve superior community detection results, which is consistent with the asymptotic analysis presented in Section 2. Additional simulations under general MSBM cases support our claim made in Section 2.5 that GMM is a generally better method than k-means in spectral clustering. Moreover, ISC_gm and SCME_gm are rather robust and continue to work well under general multi-layer stochastic block models. In practice, we recommend SCME_gm for dense networks and ISC_gm for sparse ones.

3.2. Real Data Example

In this section, we use S&P 1500 data to evaluate our methods further and compare them with the three existing spectral methods. S&P 1500 index is a good representative of the U.S. economy as it covers 90% of the market value of U.S. stocks. We obtain the daily adjusted close price of stocks from Yahoo! Finance (Aroussi 2019) for the period from January 01, 2001 to June 30, 2019, including 4663 trading days in total. We keep only stocks with less than 50 days' missing data and forward fill the price. This leaves us with 1020 stocks. According to the newest Global Industry Classification Standard [GICS], there are in total 11 stock market sectors, of which each consists of a group of stocks that share some common economic characteristics. Therefore, we treat each sector as a community and use the sector information as the ground truth for community detection. We aim to discover the sectors or communities from

stock prices. One more step of data pre-processing we did is to remove the sector "Communication Services" due to its small size, and the sectors "Industrials" and "Materials" because of their similar performances during economic cycles. The final dataset contains 770 stocks from 8 sectors.

We use the logarithmic return, a standard measure used in stock price analysis, to construct the network for stocks. Specifically, for a pair of stocks, we compute the Pearson correlation coefficient of log-returns (during a period of time) between the two stocks as the edge among them to measure their similarity. Note that the constructed network is a weighted graph whose adjacency matrix takes continuous values in $[-1, 1]$. A binary network would have been easily obtained by thresholding. However, the conversion can result in a substantial loss of community information. Hence, we will keep the weighted networks for community detection. We will see that our methods work well for this type of network as well.

The most straightforward idea is to use the whole time window to calculate the Pearson correlation among the stocks and create a single-layer network. However, since financial data is usually nonstationary, correlation tends to change over time. As a result, the within-community and between-community connectivity patterns may vary with time. To tackle such heterogeneity, we split the data into four time periods, according to the National Bureau of Economic Research [NBER], which are respectively recession I (2001/03–2001/11), expansion I (2001/12–2007/12), recession II (2008/01–2009/06) and expansion II (2009/07–2019/06). The intuition is that the economy cycle is a determinant of sector performance over the intermediate term. Different sectors tend to perform differently compared with the market in different phases of the economy. We then use the Pearson correlation computed within each time period to construct one layer. We end up having a four-layer network of stocks.

Table 5 summarizes the community detection results of different methods. Again, GMM works better than k-means in both ISC and SCME. Our method SCME_gm outperforms all the others by a large margin. ISC_gm is also competitive compared with the three spectral methods. For a closer comparison, Figure 8 shows the confusion matrix from Mean adj. and SCME_gm. We can clearly see the significant accuracy improvements for the consumer discretionary, financials, information technology, and real estate sectors. Table 5 also shows the weights for the four layers learned by our methods. In the Appendix, supplementary materials we further explain the interesting implications of the weights learned by our best method SCME_gm.

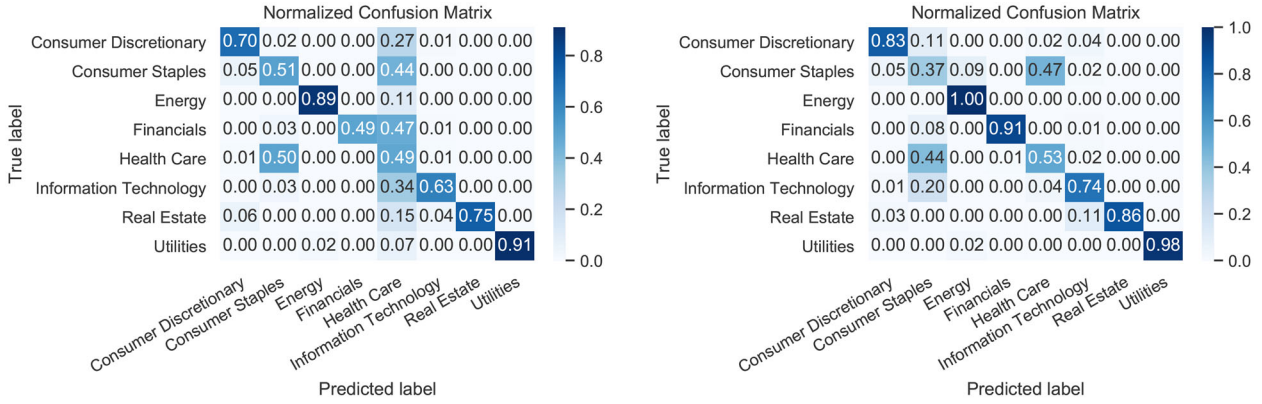
4. Conclusions and Discussions

This article presents a thorough study of the community detection problem for multi-layer networks via spectral clustering with adaptive layer aggregation. We develop two eigensystem-based methods for adaptive selection of the layer weight, and further provide a detailed discussion on the superiority of GMM clustering to k-means in the spectral clustering. Extensive numerical experiments demonstrate the impressive community detection performance of our algorithms. The proposed ISC and SCME algorithms are implemented on GitHub (*https://github.com/sihanhuang/Multi-layer-Network*). Several important directions are left open for future research.

- The theory and methods are primarily developed for multi-layer networks with assortative communities structures. It is interesting to further study how our layer aggregation methods can be used to discover both assortative and disassortative community structures in multi-layer networks of a mixed structure, especially given the fact that simply adding the adjacency matrices from assortative layers and disassortative layers can potentially wash out the community signals in the data. We have applied the ISC algorithm (without

Table 5. Community detection results for the stock market data.

Method	ARI	Weights	Method	ARI	Weights
ISC_gm	0.44	[0.52, 0.23, 0.07, 0.18]	Mean adj.	0.37	[0.25, 0.25, 0.25, 0.25]
ISC_km	0.35	[0.57, 0.21, 0.07, 0.15]	SpecK	0.45	–
SCME_gm	0.65	[0.08, 0.33, 0.01, 0.60]	Module alleg.	0.29	–
SCME_km	0.43	[0.08, 0.33, 0.01, 0.60]			



(i) Mean adj. (ii) SCME_gm

Figure 8. Normalized confusion matrix of community detection results.

any modification) to balanced MPPM of a mixed structure in some new simulations (see Section 2 in the Appendix, supplementary materials for details). The preliminary results show that like in assortative cases of Section 3.1, our algorithm is able to find the (nearly) optimal weight for layer aggregation. On the other hand, we expect that the SCME algorithm cannot be directly applied. This is because the algorithm is looking for maximum eigenratio between the Kth and $(K+1)$th eigenvalues, while aggregating assortative layers (with positive weights) and disassortative layers (with negative weights) can potentially reduce the number of informative eigenvalues. Some modifications such as taking into account the first K eigenratios $|\lambda_1|/|\lambda_2|, \ldots, |\lambda_K|/|\lambda_{K+1}|$ are needed to make it work. We leave a full investigation of detecting mixed community structures in a separate project.

- The current article focuses on spectral clustering based on adjacency matrices. It is well known that variations of spectral clustering using matrices such as normalized or regularized graph Laplacian (Rohe, Chatterjee, and Yu 2011; Qin and Rohe 2013; Amini et al. 2013; Sarkar and Bickel 2015; Joseph and Yu 2016; Le, Levina, and Vershynin 2017) can be useful for community detection under different scenarios. It is of great interest and feasible to generalize our framework to incorporate various spectral clustering forms.
- The effectiveness of our framework under general MSBM was illustrated by a wide array of synthetic and real datasets. A generalization of our asymptotic analysis to general MSBM will enable us to refine the proposed adaptive layer aggregation methods to achieve even higher accuracy. Such a generalization will rely on more sophisticated random matrix analysis and is considered as an important yet challenging future work.
- Our work assumes the number of communities K is given, which may be unknown in certain applications. Some recent efforts for estimating K in single-layer networks include Le and Levina (2015), Lei (2016), Saldana, Yu, and Feng (2017), Wang and Bickel (2017), among others. Extending our framework along these lines is interesting and doable.

Supplementary Materials

1. Appendix: Additional simulation results under balanced MPPM, more real data analysis results, and all proofs. (pdf)
2. Package: Python package for implementing the proposed algorithms. *https://github.com/sihanhuang/Multi-layer-Network*
3. Code: Python codes for reproducing the simulation and real data analysis results. (zip)

Acknowledgments

We thank the editor, the AE, and the referees for their insightful comments which greatly improved the article.

Funding

Haolei Weng was partially supported by NSF Grant DMS-1915099. Yang Feng was partially supported by NSF CAREER Grant DMS-2013789 and NIH Grant 1R21AG074205-01.

ORCID

Yang Feng https://orcid.org/0000-0001-7746-7598

References

Abbe, E. (2017), "Community Detection and Stochastic Block Models: Recent Developments," *The Journal of Machine Learning Research*, 18, 6446–6531. [2,4]

Abbe, E., Bandeira, A. S., and Hall, G. (2015), "Exact Recovery in the Stochastic Block Model," *IEEE Transactions on Information Theory*, 62, 471–487. [2]

Abbe, E., Fan, J., Wang, K., and Zhong, Y. (2020), "Entrywise Eigenvector Analysis of Random Matrices with Low Expected Rank," *Annals of Statistics*, 48, 1452–1474. [4]

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008), "Mixed Membership Stochastic Blockmodels," *Journal of Machine Learning Research*, 9, 1981–2014. [1]

Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. (2013), "Pseudo-likelihood Methods for Community Detection in Large Sparse Networks," *The Annals of Statistics*, 41, 2097–2122. [1,13]

Amini, A. A., and Levina, E. (2018), "On Semidefinite Relaxations for the Block Model," *The Annals of Statistics*, 46, 149–179. [1]

Aroussi, R. (2019), "Yahoo! Finance market data, version 0.1.37." Available from *https://pypi.org/project/fix-yahoo-finance/*. [11]

Bai, Z., and Pan, G. (2012), "Limiting Behavior of Eigenvectors of Large Wigner Matrices," *Journal of Statistical Physics*, 146, 519–549. [4]

Bandeira, A. S., and Van Handel, R. (2016), "Sharp Nonasymptotic Bounds on the Norm of Random Matrices with Independent Entries," *The Annals of Probability*, 44, 2479–2506.

Bhattacharyya, S., and Chatterjee, S. (2018), "Spectral Clustering for Multiple Sparse Networks: I," arXiv preprint arXiv:1805.10594. [2,3]

——— (2020), "General Community Detection with Optimal Recovery Conditions for Multi-Relational Sparse Networks with Dependent Layers," arXiv preprint arXiv:2004.03480. [2]

Bickel, P., Choi, D., Chang, X., and Zhang, H. (2013), "Asymptotic Normality of Maximum Likelihood and its Variational Approximation for Stochastic Blockmodels," *The Annals of Statistics*, 41, 1922–1943. [1]

Bickel, P. J., and Chen, A. (2009), "A Nonparametric View of Network Models and Newman–Girvan and other Modularities," *Proceedings of the National Academy of Sciences*, 106, 21068–21073. [1]

Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., Schweiger, J. I., Grimm, O., Heinz, A., and Tost, H. (2015), "Dynamic Reconfiguration of Frontal Brain Networks During Executive Cognition in Humans," *Proceedings of the National Academy of Sciences*, 112, 11678–11683. [8]

Bryant, P., and Williamson, J. A. (1978), "Asymptotic Behaviour of Classification Maximum Likelihood Estimates," *Biometrika*, 65, 273–281. [7]

Bui, T. N., Chaudhuri, S., Leighton, F. T., and Sipser, M. (1987), "Graph Bisection Algorithms with Good Average Case Behavior," *Combinatorica*, 7, 171–191. [2]

Cai, T. T., and Li, X. (2015), "Robust and Computationally Feasible Community Detection in the Presence of Arbitrary Outlier Nodes," *The Annals of Statistics*, 43, 1027–1059. [1]

Cape, J., Tang, M., and Priebe, C. E. (2019), "The Two-to-Infinity Norm and Singular Subspace Geometry with Applications to High-Dimensional Statistics," *The Annals of Statistics*, 47, 2405–2439. [4]

Celisse, A., Daudin, J.-J., and Pierre, L. (2012), "Consistency of Maximum-Likelihood and Variational Estimators in the Stochastic Block Model," *Electronic Journal of Statistics*, 6, 1847–1899. [1]

Chen, P.-Y., and Hero, A. O. (2017), "Multilayer Spectral Graph Clustering via Convex Layer Aggregation: Theory and Algorithms," *IEEE Transactions on Signal and Information Processing over Networks*, 3, 553–567. [3]

Choi, D. S., Wolfe, P. J., and Airoldi, E. M. (2012), "Stochastic Blockmodels with a Growing Number of Classes," *Biometrika*, 99, 273–284. [1]

De Bacco, C., Power, E. A., Larremore, D. B., and Moore, C. (2017), "Community Detection, Link Prediction, and Layer Interdependence in Multilayer Networks," *Physical Review E*, 95, 042317. [2]

Decelle, A., Krzakala, F., Moore, C., and Zdeborová, L. (2011), "Asymptotic Analysis of the Stochastic Block Model for Modular Networks and its Algorithmic Applications," *Physical Review E*, 84, 066106. [2,4]

Deshpande, Y., Abbe, E., and Montanari, A. (2017), "Asymptotic Mutual Information for the Balanced Binary Stochastic Block Model," *Information and Inference: A Journal of the IMA*, 6, 125–170. [4]

Deshpande, Y., Sen, S., Montanari, A., and Mossel, E. (2018), "Contextual Stochastic Block Models," in *Advances in Neural Information Processing Systems*, pp. 8581–8593. [4]

Dong, X., Frossard, P., Vandergheynst, P., and Nefedov, N. (2012), "Clustering with Multi-Layer Graphs: A Spectral Perspective," *IEEE Transactions on Signal Processing*, 60, 5820–5831. [2,3]

Dyer, M. E., and Frieze, A. M. (1989), "The Solution of Some Random NP-hard Problems in Polynomial Expected Time," *Journal of Algorithms*, 10, 451–489. [3]

Fan, J., Fan, Y., Han, X., and Lv, J. (2019), "Asymptotic Theory of Eigenvectors for Large Random Matrices," arXiv preprint arXiv:1902.06846. [4]

Fortunato, S. (2010), "Community Detection in Graphs," *Physics Reports*, 486, 75–174. [1]

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2017), "Achieving Optimal Misclassification Proportion in Stochastic Block Models," *The Journal of Machine Learning Research*, 18, 1980–2024. [1]

Girvan, M., and Newman, M. E. (2002), "Community Structure in Social and Biological Networks," *Proceedings of the National Academy of Sciences*, 99, 7821–7826. [1]

Guédon, O., and Vershynin, R. (2016), "Community Detection in Sparse Networks via Grothendieck's Inequality," *Probability Theory and Related Fields*, 165, 1025–1049. [1]

Hagen, L., and Kahng, A. B. (1992), "New Spectral Methods for Ratio Cut Partitioning and Clustering," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 11, 1074–1085. [1]

Han, Q., Xu, K., and Airoldi, E. (2015), "Consistent Estimation of Dynamic and Multi-Layer Block Models," in *International Conference on Machine Learning*, pp. 1511–1520. [2,3,8]

Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007), "Model-based Clustering for Social Networks," *Journal of the Royal Statistical Society*, Series A, 170, 301–354. [1]

Hoff, P. (2008), "Modeling Homophily and Stochastic Equivalence in Symmetric Relational Data," in *Advances in Neural Information Processing Systems*, pp. 657–664. [1]

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983), "Stochastic Blockmodels: First Steps," *Social Networks*, 5, 109–137. [1,2]

Hubert, L., and Arabie, P. (1985), "Comparing Partitions," *Journal of classification*, 2, 193–218. [8]

Jenatton, R., Roux, N. L., Bordes, A., and Obozinski, G. R. (2012), "A Latent Factor Model for Highly Multi-Relational Data," in *Advances in Neural Information Processing Systems*, pp. 3167–3175. [2]

Jin, J. (2015), "Fast Community Detection by SCORE," *The Annals of Statistics*, 43, 57–89. [1]

Jin, J., Ke, Z. T., and Luo, S. (2017), "Estimating Network Memberships by Simplex Vertex Hunting," arXiv preprint arXiv:1708.07852. [1]

Joseph, A., and Yu, B. (2016), "Impact of Regularization on Spectral Clustering," *The Annals of Statistics*, 44, 1765–1791. [13]

Karrer, B., and Newman, M. E. (2011), "Stochastic Blockmodels and Community Structure in Networks," *Physical Review E*, 83, 016107. [1]

Kim, J., and Lee, J.-G. (2015), "Community Detection in Multi-Layer Graphs: A Survey," *ACM SIGMOD Record*, 44, 37–48. [2]

Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L., and Zhang, P. (2013), "Spectral Redemption in Clustering Sparse Networks," *Proceedings of the National Academy of Sciences*, 110, 20935–20940. [2,4]

Kumar, A., Rai, P., and Daume, H. (2011), "Co-Regularized Multi-View Spectral Clustering," in *Advances in Neural Information Processing Systems*, pp. 1413–1421. [2,3]

Le, C. M., and Levina, E. (2015), "Estimating the Number of Communities in Networks by Spectral Methods," arXiv preprint arXiv:1507.00827. [13]

Le, C. M., Levina, E., and Vershynin, R. (2017), "Concentration and Regularization of Random Graphs," *Random Structures & Algorithms*, 51, 538–561. [1,13]

Lei, J. (2016), "A Goodness-of-Fit Test for Stochastic Block Nodels," *The Annals of Statistics*, 44, 401–424. [13]

Lei, J., Chen, K., and Lynch, B. (2020), "Consistent Community Detection in Multi-Layer Network Data," *Biometrika*, 107, 61–73. [2]

Lei, J., and Lin, K. Z. (2022), "Bias-Adjusted Spectral Clustering in Multi-Layer Stochastic Block Models," *Journal of the American Statistical Association*, 1–37. [2]

Lei, J., and Rinaldo, A. (2015), "Consistency of Spectral Clustering in Stochastic Block Models," *The Annals of Statistics*, 43, 215–237. [1]

Levin, K., Athreya, A., Tang, M., Lyzinski, V., and Priebe, C. E. (2017), "A Central Limit Theorem for an Omnibus Embedding of Multiple Random Dot Product Graphs," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, IEEE, pp. 964–967. [7]

Liu, J., Wang, C., Gao, J., and Han, J. (2013), "Multi-View Clustering via Joint Nonnegative Matrix Factorization," in *Proceedings of the 2013 SIAM International Conference on Data Mining*, SIAM, pp. 252–260. [2]

Long, B., Zhang, Z., Wu, X., and Yu, P. S. (2006), "Spectral Clustering for Multi-Type Relational Data," in *Proceedings of the 23rd International Conference on Machine Learning*, pp. 585–592. [2]

Luh, K., and Vu, V. (2018), "Sparse Random Matrices have Simple Spectrum," Manuscript. [6]

Magnus, J. R. (1985), "On Differentiating Eigenvalues and Eigenvectors," *Econometric Theory*, 1, 179–191. [6]

Mao, X., Sarkar, P., and Chakrabarti, D. (2020), "Estimating Mixed Memberships with Sharp Eigenvector Deviations," *Journal of the American Statistical Association*, 116, 1928–1940. [1]

Mossel, E., Neeman, J., and Sly, A. (2015), "Reconstruction and Estimation in the Planted Partition Model," *Probability Theory and Related Fields*, 162, 431–461. [2]

Newman, M. E. (2006), "Modularity and Community Structure in Networks," *Proceedings of the National Academy of Sciences*, 103, 8577–8582. [1]

——— (2018), *Networks*, Oxford: Oxford University Press. [1]

Nickel, M., Tresp, V., and Kriegel, H.-P. (2011), "A Three-Way Model for Collective Learning on Multi-Relational Data." in *Icml* (Vol. 11), pp. 809–816. [2]

Paul, S., and Chen, Y. (2016), "Consistent Community Detection in Multi-Relational Data through Restricted Multi-Layer Stochastic Blockmodel," *Electronic Journal of Statistics*, 10, 3807–3870. [2,3]

——— (2017), "Consistency of Community Detection in Multi-Layer Networks using Spectral and Matrix Factorization Methods," arXiv preprint arXiv:1704.07353. [8]

——— (2020), "Spectral and Matrix Factorization Methods for Consistent Community Detection in Multi-Layer Networks," *The Annals of Statistics*, 48, 230–250. [2,3]

Pollard, D. (1981), "Strong Consistency of k-means Clustering," *The Annals of Statistics*, 9, 135–140. [7]

Qin, T., and Rohe, K. (2013), "Regularized Spectral Clustering under the Degree-Corrected Stochastic Blockmodel," in *Advances in Neural Information Processing Systems*, pp. 3120–3128. [1,13]

Rohe, K., Chatterjee, S., and Yu, B. (2011), "Spectral Clustering and the High-Dimensional Stochastic Blockmodel," *The Annals of Statistics*, 39, 1878–1915. [1,13]

Rosvall, M., and Bergstrom, C. T. (2008), "Maps of Random Walks on Complex Networks Reveal Community Structure," *Proceedings of the National Academy of Sciences*, 105, 1118–1123. [1]

Saldana, D. F., Yu, Y., and Feng, Y. (2017), "How Many Communities are there?" *Journal of Computational and Graphical Statistics*, 26, 171–181. [13]

Sarkar, P., and Bickel, P. J. (2015), "Role of Normalization in Spectral Clustering for Stochastic Blockmodels," *The Annals of Statistics*, 43, 962–990. [13]

Schein, A., Paisley, J., Blei, D. M., and Wallach, H. (2015), "Bayesian Poisson Tensor Factorization for Inferring Multilateral Relations from Sparse Dyadic Event Counts," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1045–1054. [2]

——— (2016), "Bayesian Poisson Tucker Decomposition for Learning the Structure of International Relations," in *International Conference on Machine Learning*, pp. 2810–2819. [2]

Shi, J., and Malik, J. (2000), "Normalized Cuts and Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 888–905. [1]

Singh, A. P., and Gordon, G. J. (2008), "Relational Learning via Collective Matrix Factorization," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 650–658. [2]

Stanley, N., Shai, S., Taylor, D., and Mucha, P. J. (2016), "Clustering Network Layers with the Strata Multilayer Stochastic Block Model," *IEEE Transactions on Network Science and Engineering*, 3, 95–105. [2]

Steinhaeuser, K., and Chawla, N. V. (2010), "Identifying and Evaluating Community Structure in Complex Networks," *Pattern Recognition Letters*, 31, 413–421. [8]

Tang, M., and Priebe, C. E. (2018), "Limit Theorems for Eigenvectors of the Normalized Laplacian for Random Graphs," *The Annals of Statistics*, 46, 2360–2415. [4,7]

Tang, W., Lu, Z., and Dhillon, I. S. (2009), "Clustering with Multiple Graphs," in *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*, IEEE, pp. 1016–1021. [2,3]

Tao, T., and Vu, V. (2017), "Random Matrices have Simple Spectrum," *Combinatorica*, 37, 539–553. [6]

Valles-Catala, T., Massucci, F. A., Guimera, R., and Sales-Pardo, M. (2016), "Multilayer Stochastic Block Models Reveal the Multilayer Structure of Complex Networks," *Physical Review X*, 6, 011036. [2]

Weng, H., and Feng, Y. (2022), "Community Detection with Nodal Information: Likelihood and its Variational Approximation," *Stat*, 11, e428. [1]

Wang, Y. R., and Bickel, P. J. (2017), "Likelihood-based Model Selection for Stochastic Block Models," *The Annals of Statistics*, 45, 500–528. [13]

Xu, M., Jog, V., and Loh, P.-L. (2020), "Optimal Rates for Community Estimation in the Weighted Stochastic Block Model," *The Annals of Statistics*, 48, 183–204. [2]

Yu, Y., Wang, T., and Samworth, R. J. (2015), "A Useful Variant of the Davis–Kahan Theorem for Statisticians," *Biometrika*, 102, 315–323. [1]

Zhang, A. Y., and Zhou, H. H. (2016), "Minimax Rates of Community Detection in Stochastic Block Models," *The Annals of Statistics*, 44, 2252–2280. [1,4]

Zhang, Y., Levina, E., and Zhu, J. (2016), "Community Detection in Networks with Node Features," *Electronic Journal of Statistics*, 10, 3153–3178. [1]

——— (2020), "Detecting Overlapping Communities in Networks Using Spectral Methods," *SIAM Journal on Mathematics of Data Science*, 2, 265–283. [1]

Zhao, Y. (2017), "A Survey on Theoretical Advances of Community Detection in Networks," *Wiley Interdisciplinary Reviews: Computational Statistics*, 9, e1403. [4]

Zhao, Y., Levina, E., and Zhu, J. (2012), "Consistency of Community Detection in Networks under Degree-Corrected Stochastic Block Models," *The Annals of Statistics*, 40, 2266–2292. [1]

Zhou, D., and Burges, C. J. (2007), "Spectral Clustering and Transductive Learning with Multiple Views," in *Proceedings of the 24th International Conference on Machine Learning*, pp. 1159–1166. [2]

Zhou, H. (2003), "Distance, Dissimilarity Index, and Network Community Structure," *Physical review e*, 67, 061901. [1]