Multilayer Spectral Graph Clustering via Convex Layer Aggregation: Theory and Algorithms

Pin-Yu Chen, *Member, IEEE*, and Alfred O. Hero III, *Fellow, IEEE*

Abstract—Multilayer graphs are commonly used for representing different relations between entities and handling heterogeneous data processing tasks. Nonstandard multilayer graph clustering methods are needed for assigning clusters to a common multilayer node set and for combining information from each layer. This paper presents a multilayer spectral graph clustering (SGC) framework that performs convex layer aggregation. Under a multilayer signal-plus-noise model, we provide a phase transition analysis of clustering reliability. Moreover, we use the phase transition criterion to propose a multilayer iterative model order selection algorithm (MIMOSA) for multilayer SGC, which features automated cluster assignment and layer weight adaptation, and provides statistical clustering reliability guarantees. Numerical simulations on synthetic multilayer graphs verify the phase transition analysis, and experiments on real-world multilayer graphs show that MIMOSA is competitive or better than other clustering methods.

Index Terms—Community detection, model order selection, multilayer graphs, multiplex networks, phase transition.

I. INTRODUCTION

MULTILAYER graphs provide a framework for representing multiple types of relations between entities, represented as nodes. In a multilayer graph each layer describes a specific type of relation among pairs of nodes that are shared across layers. For example, in multi-relational social networks, two layers might correspond to friendship relations and business relations, respectively. In temporal networks, each layer might correspond to a snapshot of the entire network at a sampled time instant. Multilayer graphs can be incorporated into seems redundant many signal processing and data mining techniques, including inference of mixture models [1], [2], tensor decomposition [3], information extraction [4], multi-view learning and processing [5], graph wavelet transforms [6], principal compo-

nent analysis and dictionary learning [7], [8], anomaly detection [9], and community detection [10], [11], among others.

The objective of multilayer graph clustering is to find a consensus cluster assignment on each node in the common node set by combining connectivity patterns in each layer. Multilayer graph clustering differs from single-layer graph clustering in several respects: (1) the information about cluster membership must be aggregated from multiple layers; (2) the performance of multilayer graph clustering will depend on the proportion of noisy edges across layers. This paper proposes a multilayer spectral graph clustering (SGC) algorithm that uses convex layer aggregation. Specifically, the algorithm performs SGC on an weighted average of the adjacency matrices of the layers, where the weights are non-negative and sum to one. We establish phase transitions in multilayer graph clustering in the convex layer-aggregated graph as a function of the noisy edge connection parameters of each layer under a multilayer signal plus noise model. Our phase transition analysis shows that when one sweeps over noise levels, there exists a critical threshold below (above) which multilayer SGC will yield correct (incorrect) clusters. This critical phase transition threshold depends on the layer weights used to aggregate the multilayer graph into a single-layer graph in addition to the topology of the multilayer graph. Numerical experiments on synthetic multilayer graphs are conducted to verify the phase transitions of the proposed method. Moreover, we propose a multilayer iterative model order selection algorithm (MIMOSA) that incorporates automated layer weight adaptation and cluster assignment. Experimental results on real-world multilayer graphs show that MIMOSA has competitive clustering performance to (1) the baseline approach of assigning uniform layer weights, (2) the greedy multilayer modularity maximization method [12], and (3) the subspace approach [13].

This paper makes two principal contributions. First, under a general multilayer signal plus noise model, we establish a phase transition on the performance of multilayer SGC. Fixing the within-cluster edges (signals) and varying the parameters governing the between-cluster edges (noises), we show that the clustering accuracy of multilayer SGC can be separated into two regimes: a reliable regime where high clustering accuracy can be guaranteed, and an unreliable regime where high clustering accuracy is impossible. Moreover, we specify upper and lower bounds on the critical noise value that separates these two regimes, which is an analytical function of the signal strength, the number of clusters, the cluster size distributions, and the layer weights for convex layer aggregation. The bounds become

exact in the case of identical cluster sizes. The analysis specifies the interplay between the layer weights, the multilayer graph connectivity structure in terms of eigenspectrum, and the performance of multilayer SGC via convex layer aggregation. The analysis also provides a criterion for assessing the quality of clustering results, which leads to the second contribution: the introduction of a new multilayer clustering algorithm with automated model order selection (number of clusters). This algorithm, called the multilayer iteration model order selection algorithm (MIMOSA), selects both the model order and the layer weights and results in improved performance. MIMOSA incrementally increases the number of clusters, adapts layer weight assignment, and adopts a series of statistical clustering reliability tests. To illustrate the proposed MIMOSA approach, we apply it to several real-world multilayer graphs pertaining to social, biological, collaboration and transportation networks.

The rest of this paper is organized as follows. Section II reviews related work on multilayer graph clustering. Section III introduces the multilayer signal plus noise model for multilayer graphs, and presents the mathematical formulation of multilayer SGC via convex layer aggregation. Section IV provides performance analysis of the proposed multilayer SGC under a multilayer signal plus noise model. We specify a breakdown condition for the success of multilayer SGC, and establish a phase transition on the clustering accuracy of multilayer SGC under a block-wise identical noise model and a block-wise non-identical noise model, respectively. Section V describes the proposed MIMOSA approach for automated multilayer graph clustering. Section VI presents numerical results that verify the phase transition analysis. Section VII compares the performance of MIMOSA with two other automated multilayer graph clustering methods on 9 real-world multilayer graph datasets. Finally, Section VIII concludes this paper.

II. RELATED WORK

Graph clustering, also known as community detection, on multilayer graphs aims to find a consensus cluster assignment on each node in the common node set shared by different layers. Layer aggregation has been a principal method for processing and mining multilayer graphs [14]–[20], as it transforms a multilayer graph into a single aggregated graph, facilitating application of data analysis techniques designed for single-layer graphs. Extending the stochastic block model (SBM) for graph clustering in single-layer graphs [21], a multilayer SBM has been proposed for graph clustering on multilayer graphs [19], [22]–[26]. Under the assumption of two equally-sized clusters, the authors in [19] show that if layer aggregation is used and if each layer is an independent realization of a common SBM, the inferential limit for cluster detectability decays at rate $O(L^{-\frac{1}{2}})$, where L is the number of layers. In [26], a layer selection method based on a multilayer SBM is proposed to improve the performance of graph clustering by identifying a subset of coherent layers. However, the multilayer SBM assumes homogeneous connectivity structure for within-cluster and between-cluster edges in each layer, and it also assumes layer-wise independence.

In addition to inference approaches based on the multilayer SBM, other methods have been proposed for graph clustering on

multilayer graphs, including information-theoretic approaches [27], [28], k-nearest neighbor method [29], nonnegative matrix factorization [30], flow-based approach [31], linked matrix factorization [32], random walk [33], tensor decomposition [3], subspace methods [13], [34], subgraph mining with edge labels [35], and greedy multilayer modularity maximization [12]. More details on multilayer graph models can be found in the recent survey papers for graph clustering on multilayer graphs [10], [11].

It is worth mentioning that the methods proposed in many of the aforementioned publications require the knowledge of the number of clusters (model order) for graph clustering, especially for matrix decomposition-based methods [3], [13], [30], [32], [34] and multilayer SBM [19], [22]–[26]. However, in many practical cases the model order is not known. Although many model order selection methods have been proposed for single-layer graphs [36]–[39], little has been developed for multilayer graphs. Moreover, many layer aggregation methods assign uniform weights over layers such that the aggregated graph is insensitive to the quality of clusters in each layer [15], [17], [19]. This paper studies the sensitivity of the clustering accuracy to layer weights under a multilayer signal plus noise model. We then propose a model order selection algorithm featuring layer weight adaptation that automatically finds the minimal model order that meets statistical clustering reliability guarantees.

III. MULTILAYER GRAPH MODEL AND SPECTRAL GRAPH CLUSTERING VIA CONVEX LAYER AGGREGATION

A. Multilayer Graph Model

Throughout this paper, we consider a multilayer graph model consisting of L layers representing different relationships among a common node set $\mathcal{V}$ of n nodes. The graph in the ℓ-th layer is an undirected graph with nonnegative edge wights, which is denoted by $G_\ell = (\mathcal{V}, \mathcal{E}_\ell)$, where $\mathcal{E}_\ell$ is the set of weighted edges in the ℓ-th layer. The $n \times n$ binary symmetric adjacency matrix $\mathbf{A}^{(\ell)}$ is used to represent the connectivity structure of G_ℓ. The entry $[\mathbf{A}^{(\ell)}]_{uv} = 1$ if nodes u and v are connected in the ℓ-th layer, and $[\mathbf{A}^{(\ell)}]_{uv} = 0$ otherwise. Similarly, the $n \times n$ nonnegative symmetric weight matrix $\mathbf{W}^{(\ell)}$ is used to represent the edge weights in G_ℓ, where $\mathbf{W}^{(\ell)}$ and $\mathbf{A}^{(\ell)}$ have the same zero structure.

We assume each layer in the multilayer graph is a (possibly correlated) representation of a common set of disjoint K clusters that partitions the node set $\mathcal{V}$, where the k-th cluster has cluster size n_k such that $\sum_{k=1}^{K} n_k = n$, and $n_{\min} = \min_{k \in \{1, \ldots, K\}} n_k$ and $n_{\max} = \max_{k \in \{1, \ldots, K\}} n_k$ denote the smallest and largest cluster size, respectively. Specifically, the adjacency matrix $\mathbf{A}^{(\ell)}$ of G_ℓ in the ℓ-th layer can be represented as

$$\mathbf{A}^{(\ell)} = \begin{bmatrix} \mathbf{A}_1^{(\ell)} & \mathbf{C}_{12}^{(\ell)} & \mathbf{C}_{13}^{(\ell)} & \cdots & \mathbf{C}_{1K}^{(\ell)} \\ \mathbf{C}_{21}^{(\ell)} & \mathbf{A}_2^{(\ell)} & \mathbf{C}_{23}^{(\ell)} & \cdots & \mathbf{C}_{2K}^{(\ell)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{K1}^{(\ell)} & \mathbf{C}_{K2}^{(\ell)} & \cdots & \cdots & \mathbf{A}_K^{(\ell)} \end{bmatrix}, \quad (1)$$

where $\mathbf{A}_k^{(\ell)}$ is an $n_k \times n_k$ binary symmetric matrix denoting the adjacency matrix of within-cluster edges of the k-th cluster in the ℓ-th layer, and $\mathbf{C}_{ij}^{(\ell)}$ is an $n_i \times n_j$ binary rectangular matrix denoting the adjacency matrix of between-cluster edges of clusters i and j in the ℓ-th layer, $1 \le i, j \le K$, $i \ne j$, and $\mathbf{C}_{ij}^{(\ell)} = \mathbf{C}_{ji}^{(\ell)T}$.

Similarly, the edge weight matrix $\mathbf{W}^{(\ell)}$ of the ℓ-th layer can be represented as

$$\mathbf{W}^{(\ell)} = \begin{bmatrix} \mathbf{W}_1^{(\ell)} & \mathbf{F}_{12}^{(\ell)} & \mathbf{F}_{13}^{(\ell)} & \cdots & \mathbf{F}_{1K}^{(\ell)} \\ \mathbf{F}_{21}^{(\ell)} & \mathbf{W}_2^{(\ell)} & \mathbf{F}_{23}^{(\ell)} & \cdots & \mathbf{F}_{2K}^{(\ell)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}_{K1}^{(\ell)} & \mathbf{F}_{K2}^{(\ell)} & \cdots & \cdots & \mathbf{W}_K^{(\ell)} \end{bmatrix}, \quad (2)$$

where $\mathbf{W}_k^{(\ell)}$ is an $n_k \times n_k$ nonnegative symmetric matrix denoting the edge weights of within-cluster edges of the k-th cluster in the ℓ-th layer, and $\mathbf{F}_{ij}^{(\ell)}$ is an $n_i \times n_j$ nonnegative rectangular matrix denoting the edge weights of between-cluster edges of clusters i and j in the ℓ-th layer, $1 \le i, j \le K$, $i \ne j$, and $\mathbf{F}_{ij}^{(\ell)} = \mathbf{F}_{ji}^{(\ell)T}$.

B. Multilayer Signal Plus Noise Model

Using the cluster-wise block representations of the adjacency and edge weight matrices for the multilayer graph model described in (1) and (2), we propose a signal-plus-noise model for $\mathbf{A}^{(\ell)}$ and $\mathbf{W}^{(\ell)}$ to analyze the effect of convex layer aggregation on graph clustering. Specifically, for each layer we assume the connectivity structure and edge weight distributions follow the random interconnection model (RIM) [39]. In RIM, the signal of the k-th cluster in the ℓ-th layer is the connectivity structure in terms of eigenspectrum and weights of the within-cluster edges represented by the matrices $\mathbf{A}_k^{(\ell)}$ and $\mathbf{W}_k^{(\ell)}$, respectively. The RIM imposes no distributional assumption on the within-cluster edges. The noise between clusters i and j in the ℓ-th layer is caused by random between-cluster edges, which are represented by the matrices $\mathbf{C}_{ij}^{(\ell)}$ and $\mathbf{F}_{ij}^{(\ell)}$, respectively.

Throughout this paper, we assume the connectivity of a between-cluster edge (i.e., the noise) in each layer is independently drawn from a layer-wise and block-wise independent Bernoulli distribution. Specifically, each entry in $\mathbf{C}_{ij}^{(\ell)}$ representing the existence of an edge between clusters i and j in the ℓ-th layer is an independent realization of a Bernoulli random variable with edge connection probability $p_{ij}^{(\ell)} \in [0, 1]$ that is layer-wise and block-wise independent. In addition, given the existence of an edge (u, v) between clusters i and j in the ℓ-th layer, the entry $[\mathbf{F}_{ij}^{(\ell)}]_{uv}$ representing the corresponding edge weight is independently drawn from a nonnegative distribution with mean $\overline{W}_{ij}^{(\ell)}$ and bounded fourth moment that is layer-wise and block-wise independent. The assumption of bounded fourth moment is required for the phase transition analysis established in Section IV.

For the ℓ-th layer, the noise accounting for the between-cluster edges is said to be *block-wise identical* if the noise parameters $p_{ij}^{(\ell)} = p^{(\ell)}$ and $\overline{W}_{ij}^{(\ell)} = \overline{W}^{(\ell)}$ for every cluster pair i and j, $i \ne j$. Otherwise it is said to be *block-wise non-identical*. The effect of these two noise models on multilayer spectral graph clustering will be studied in Section IV.

C. Multilayer Spectral Graph Clustering via Convex Layer Aggregation

Let $\mathbf{w} = [w_1, \ldots, w_L]^T \in \mathcal{W}_L$ be an $L \times 1$ column vector representing the layer weight vector for convex layer aggregation, where $\mathcal{W}_L = \{\mathbf{w} : w_\ell \ge 0, \sum_{\ell=1}^{L} w_\ell = 1\}$ is the set of feasible layer weight vectors. The single-layer graph obtained via convex layer aggregation with layer weight vector $\mathbf{w}$ is denoted by $G^{\mathbf{w}}$. The (weighted) adjacency matrix and the edge weight matrix of $G^{\mathbf{w}}$ are denoted by $\mathbf{A}^{\mathbf{w}}$ and $\mathbf{W}^{\mathbf{w}}$, respectively, where $\mathbf{A}^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell \mathbf{A}^{(\ell)}$ and $\mathbf{W}^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell \mathbf{W}^{(\ell)}$. The graph Laplacian matrix $\mathbf{L}^{\mathbf{w}}$ of $G^{\mathbf{w}}$ is defined as $\mathbf{L}^{\mathbf{w}} = \mathbf{S}^{\mathbf{w}} - \mathbf{W}^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell \mathbf{L}^{(\ell)}$, where $\mathbf{S}^{\mathbf{w}} = \mathrm{diag}(\mathbf{s}^{\mathbf{w}})$ is a diagonal matrix, $\mathbf{s}^{\mathbf{w}} = \mathbf{W}^{\mathbf{w}} \mathbf{1}_n$ is the vector of nodal strength of $G^{\mathbf{w}}$, $\mathbf{1}_n$ is the $n \times 1$ column vector of ones, and $\mathbf{L}^{(\ell)}$ is the graph Laplacian matrix of G_ℓ. Similarly, the graph Laplacian matrix $\mathbf{L}_k^{\mathbf{w}}$ accounting for the within-cluster edges of the k-th cluster in $G^{\mathbf{w}}$ is defined as $\mathbf{L}_k^{\mathbf{w}} = \mathbf{S}_k^{\mathbf{w}} - \mathbf{W}_k^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell \mathbf{L}_k^{(\ell)}$, where $\mathbf{W}_k^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell \mathbf{W}_k^{(\ell)}$, $\mathbf{S}_k^{\mathbf{w}} = \mathrm{diag}(\mathbf{W}_k^{\mathbf{w}} \mathbf{1}_{n_k})$, and $\mathbf{L}_k^{(\ell)} = \mathbf{S}_k^{(\ell)} - \mathbf{W}_k^{(\ell)}$. The i-th smallest eigenvalue of $\mathbf{L}^{\mathbf{w}}$ is denoted by $\lambda_i(\mathbf{L}^{\mathbf{w}})$. Based on the definition of $\mathbf{L}^{\mathbf{w}}$, the smallest eigenvalue $\lambda_1(\mathbf{L}^{\mathbf{w}})$ of $\mathbf{L}^{\mathbf{w}}$ is 0, since $\mathbf{L}^{\mathbf{w}} \mathbf{1}_n = \mathbf{0}_n$, where $\mathbf{0}_n$ is the $n \times 1$ column vector of zeros.

Spectral graph clustering (SGC) [40] partitions the nodes in $G^{\mathbf{w}}$ into K ($K \ge 2$) clusters based on the K eigenvectors associated with the K smallest eigenvalues of $\mathbf{L}^{\mathbf{w}}$. Specifically, SGC first transforms each node in $G^{\mathbf{w}}$ to a K-dimensional vector in the subspace spanned by these eigenvectors, and then implements K-means clustering [41] on these vectors to group the nodes in $G^{\mathbf{w}}$ into K clusters. For analysis purposes, throughout this paper we assume $G^{\mathbf{w}}$ is a connected graph. If $G^{\mathbf{w}}$ is disconnected, SGC can be applied to each connected component in $G^{\mathbf{w}}$. Moreover, if $G^{\mathbf{w}}$ is connected, $\lambda_i(\mathbf{L}^{\mathbf{w}}) > 0$ for all $i \ge 2$. That is, the second to the n-th smallest eigenvalue of $\mathbf{L}^{\mathbf{w}}$ are all positive [42]. In addition, the eigenvector associated with the smallest eigenvalue $\lambda_1(\mathbf{L}^{\mathbf{w}})$ provides no information about graph clustering since it is proportional to a constant vector, the vector of ones $\mathbf{1}_n$.

Let $\mathbf{Y} \in \mathbb{R}^{n \times (K-1)}$ denote the eigenvector matrix where its k-th column is the $(k+1)$-th eigenvector associated with $\lambda_{k+1}(\mathbf{L}^{\mathbf{w}})$, $1 \le k \le K-1$. By the Courant-Fischer theorem [43], $\mathbf{Y}$ is the solution of the minimization problem

$$S_{2:K}(\mathbf{L}^{\mathbf{w}}) = \min_{\mathbf{X} \in \mathbb{R}^{n \times (K-1)}} \mathrm{trace}(\mathbf{X}^T \mathbf{L}^{\mathbf{w}} \mathbf{X}),$$

$$\text{subjec to } \mathbf{X}^T \mathbf{X} = \mathbf{I}_{K-1}, \ \mathbf{X}^T \mathbf{1}_n = \mathbf{0}_{K-1}, \quad (3)$$

where the optimal value $S_{2:K}(\mathbf{L}^{\mathbf{w}}) = \mathrm{trace}(\mathbf{Y}^T \mathbf{L}^{\mathbf{w}} \mathbf{Y})$ in (3) is the partial eigenvalue sum $S_{2:K}(\mathbf{L}^{\mathbf{w}}) = \sum_{k=2}^{K} \lambda_k(\mathbf{L}^{\mathbf{w}})$, $\mathbf{I}_{K-1}$ is the $(K-1) \times (K-1)$ identity matrix, and the constraints in (3) impose orthonormality and centrality on the eigenvectors. In

summary, multilayer SGC via convex layer aggregation works by computing the eigenvector matrix $\mathbf{Y}$ from $\mathbf{L^w}$ of $G^\mathbf{w}$, and implementing K-means clustering on the rows of $\mathbf{Y}$ to group the nodes into K clusters.

IV. PERFORMANCE ANALYSIS OF MULTILAYER SPECTRAL GRAPH CLUSTERING VIA CONVEX LAYER AGGREGATION

In this section, we establish three theorems on the performance of multilayer spectral graph clustering (SGC) via convex layer aggregation, which generalizes the phase transition analysis established in [39] for single-layer graphs. The novelty of the analysis presented in this section is the incorporation of the effect of layer weights into multilayer SGC. One obtains the results in [39] as a special case of the analysis presented in this section when there is only one layer (i.e., $L = 1$) and therefore the layer weight vector $\mathbf{w}$ reduces to a unit scalar. To assist comparison, in this section we use a similar, but abbreviated, presentation structure as in [39] for our phase transition analysis.[1] The proofs are given in the supplementary file. The analysis provides a theoretical framework for multilayer SGC and allows us to evaluate the quality of clustering results in terms of a signal-to-noise (SNR) ratio that falls out of the established theorems. This SNR is then used for determining the number of clusters and selecting layer weights in the algorithm proposed in Section V.

The first theorem (Theorem 1) specifies the interplay between layer weights and the success of multilayer SGC by establishing a condition under which multilayer SGC fails to correctly identify clusters under the multilayer signal plus noise model in Section III-B due to inconsistent rows in the eigenvector matrix $\mathbf{Y}$. The condition is called a "breakdown condition" and can be used as a test for identifiability of a given cluster configuration in the multilayer SGC problem.

The second theorem (Theorem 2) establishes phase transitions on the clustering performance of multilayer SGC under the block-wise identical noise model for a given layer weight vector $\mathbf{w}$. Under the block-wise identical noise model, define $t^{(\ell)} = p^{(\ell)} \cdot \overline{W}^{(\ell)}$ to be the noise level of the ℓ-th layer and let $t^\mathbf{w} = \sum_{\ell=1}^{L} w_\ell \cdot t^{(\ell)}$ be the aggregated noise level via convex layer aggregation. We show that for each $\mathbf{w} \in \mathcal{W}_L$ there exists a critical value $t^{\mathbf{w}*}$ of $t^\mathbf{w}$ such that if $t^\mathbf{w} < t^{\mathbf{w}*}$, multi-layer SGC can correctly identify the clusters, and if $t^\mathbf{w} > t^{\mathbf{w}*}$, reliable multi-layer SGC is not possible.

The third theorem (Theorem 3) extends the phase transition analysis of the block-wise identical noise model to the block-wise non-identical noise model. Under the block-wise non-identical noise model, define $t_{\max}^{(\ell)} = \max_{i,j,i\neq j} p_{ij}^{(\ell)} \cdot \overline{W}_{ij}^{(\ell)}$ as the maximum noise level of the ℓ-th layer and let $t_{\max}^\mathbf{w} = \sum_{\ell=1}^{L} w_\ell \cdot t_{\max}^{(\ell)}$. Then for each $\mathbf{w} \in \mathcal{W}_L$ we show that reliable clustering results can be guaranteed provided that $t_{\max}^\mathbf{w} < t^{\mathbf{w}*}$,

where $t^{\mathbf{w}*}$ is the critical value for phase transition under the block-wise identical noise model.

A. Breakdown Condition for Multilayer SGC via Convex Layer Aggregation

Under the multilayer signal plus noise model in Section III-B, let $t_{ij}^{(\ell)} = \overline{W}_{ij}^{(\ell)} \cdot p_{ij}^{(\ell)}$ be the noise level between clusters i and j in the ℓ-th layer, $1 \leq i, j \leq K$, $i \neq j$, and $1 \leq \ell \leq L$. The following theorem establishes a general breakdown condition under which multilayer SGC fails to correctly identify the clusters.

Theorem 1 (general breakdown condition).

Let $\widetilde{\mathbf{W}}^\mathbf{w}$ be the $(K-1) \times (K-1)$ matrix with (i,j)-th entry

$$
\left[\widetilde{\mathbf{W}}^\mathbf{w}\right]_{ij} = \begin{cases} \sum_{\ell=1}^{L} w_\ell \left[(n_i + n_K)\, t_{iK}^{(\ell)} + \sum_{z=1, z\neq i}^{K-1} n_z t_{iz}^{(\ell)} \right], \\ \quad \text{if } i = j; \\ \sum_{\ell=1}^{L} w_\ell n_i \cdot \left(t_{iK}^{(\ell)} - t_{ij}^{(\ell)} \right), \text{ if } i \neq j. \end{cases}
$$

The following holds almost surely as $n_k \to \infty \,\, \forall\, k$ and $\frac{n_{\min}}{n_{\max}} \to c > 0$. If for any layer weight vector $\mathbf{w} \in \mathcal{W}_L$, $\lambda_i \left(\frac{\widetilde{\mathbf{W}}^\mathbf{w}}{n} \right) \neq \lambda_j \left(\frac{\mathbf{L}^\mathbf{w}}{n} \right)$ for all $i = 1, 2, \dots, K-1$ and $j = 2, 3, \dots, K$, then multilayer SGC cannot be successful.

Theorem 1 specifies the interplay between the layer weight vector $\mathbf{w}$ and the accuracy of multilayer SGC. Different from the case of single-layer graphs (i.e., $L = 1$ and hence $\mathbf{w} = 1$) such that the layer weight has no effect on the performance of SGC, Theorem 1 states that multilayer SGC cannot be successful if every possible layer weight vector $\mathbf{w} \in \mathcal{W}_L$ leads to distinct $K - 1$ smallest nonzero eigenvalues of the matrices $\frac{\widetilde{\mathbf{W}}^\mathbf{w}}{n}$ and $\frac{\mathbf{L}^\mathbf{w}}{n}$. It also suggests that the selection of layer weight vector affects the performance of multilayer SGC.

B. Phase Transitions in Multilayer SGC Under Block-Wise Identical Noise

Under the multilayer signal plus noise model in Section III-B, if we further assume the between-cluster edges in each layer follow a block-wise identical distribution, then the noise level in the ℓ-th layer can be characterized by the parameter $t^{(\ell)} = p^{(\ell)} \cdot \overline{W}^{(\ell)}$, where $p^{(\ell)} \in [0, 1]$ is the edge connection parameter and $\overline{W}^{(\ell)} > 0$ is the mean of the between-cluster edge weights in the ℓ-th layer. Under the block-wise identical noise model and given a layer weight vector $\mathbf{w} \in \mathcal{W}_L$, let $t^\mathbf{w} = \sum_{\ell=1}^{L} w_\ell t^{(\ell)}$ denote the aggregated noise level of the graph $G^\mathbf{w}$. Theorem 2 below establishes phase transitions in the eigendecomposition of the graph Laplacian matrix $\mathbf{L}^\mathbf{w}$ of the graph $G^\mathbf{w}$. We show that there exists a critical value $t^{\mathbf{w}*}$ such that the K smallest eigenpairs of $\mathbf{L}^\mathbf{w}$ that are used for multilayer SGC have different characteristics when $t^\mathbf{w} < t^{\mathbf{w}*}$ and $t^\mathbf{w} > t^{\mathbf{w}*}$. In particular, we show that the solution to the minimization problem in (3), the eigenvector matrix $\mathbf{Y} = [\mathbf{Y}_1^T, \mathbf{Y}_2^T, \dots, \mathbf{Y}_K^T]^T \in \mathbb{R}^{n \times (K-1)}$, where its rows $\mathbf{Y}_k \in \mathbb{R}^{n_k \times (K-1)}$ index the nodes in cluster k, has cluster-wise separability when $t^\mathbf{w} < t^{\mathbf{w}*}$. This means that, under this condition, the rows of each $\mathbf{Y}_k$ are identical with columns

---

[1] In Section IV, there are a number of limit theorems stated about the behavior of random matrices and vectors whose dimensions go to infinity as the sizes $\{n_k\}_{k=1}^{K}$ of the clusters go to infinity while their relative sizes $n_k/n_{k'}$ are held constant. For simplicity and convenience, the limit theorems are often stated in terms of the finite, but arbitrarily large, dimensions n_k, $k = 1, 2, \dots, K$. For any two matrices $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ of the same dimension, The notation $\mathbf{X} \to \widetilde{\mathbf{X}}$ means convergence in the spectral norm [44]. The notation $\mathbf{X} \xrightarrow{\text{a.s.}} \widetilde{\mathbf{X}}$ means $\mathbf{X} \to \widetilde{\mathbf{X}}$ almost surely.

that are cluster-wise distinct. On the other hand, when $t^{\mathbf{w}} > t^{\mathbf{w}*}$ the row-wise average of each matrix $\mathbf{Y}_k$ is a zero vector and hence the clusters cannot be perfectly separated by inspecting the eigenvector matrix $\mathbf{Y}$.

Theorem 2 (block-wise identical noise).

Let $\mathbf{Y} = [\mathbf{Y}_1^T, \mathbf{Y}_2^T, \ldots, \mathbf{Y}_K^T]^T$ be the solution of the minimization problem in (3) and let $c^{\mathbf{w}*} = \min_{k \in \{1,2,\ldots,K\}}$ $\{\frac{S_{2:K}(\mathbf{L}_k^{\mathbf{w}})}{n}\}$, where $\mathbf{L}_k^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell \mathbf{L}_k^{(\ell)}$. Given a layer weight vector $\mathbf{w} \in \mathcal{W}_L$, under the block-wise identical noise model with aggregated noise level $t^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell t^{(\ell)} = \sum_{\ell=1}^{L} w_\ell$ $p^{(\ell)} \overline{W}^{(\ell)}$, there exists a critical value $t^{\mathbf{w}*}$ such that the following holds almost surely as $n_k \to \infty \; \forall k$ and $\frac{n_{\min}}{n_{\max}} \to c > 0$:

(a) $\begin{cases} \text{If } t^{\mathbf{w}} \leq t^{\mathbf{w}*}, \; \frac{S_{2:K}(\mathbf{L}^{\mathbf{w}})}{n} = (K-1)t^{\mathbf{w}}; \\ \text{If } t^{\mathbf{w}} > t^{\mathbf{w}*}, \; c^{\mathbf{w}*} + (K-1)\left(1 - \frac{n_{\max}}{n}\right)t^{\mathbf{w}} \leq \frac{S_{2:K}(\mathbf{L}^{\mathbf{w}})}{n} \\ \qquad \leq c^{\mathbf{w}*} + (K-1)\left(1 - \frac{n_{\min}}{n}\right)t^{\mathbf{w}}. \end{cases}$

In particular, if $t^{\mathbf{w}} > t^{\mathbf{w}*}$ and $c = 1$, $\frac{S_{2:K}(\mathbf{L}^{\mathbf{w}})}{n} = c^{\mathbf{w}*} + \frac{(K-1)^2}{K}t^{\mathbf{w}}$.

(b) $\begin{cases} \text{If } t^{\mathbf{w}} < t^{\mathbf{w}*}, \; \mathbf{Y}_k = \mathbf{1}_{n_k}\mathbf{1}_{K-1}^T\mathbf{V}_k \\ \quad = [v_1^k \mathbf{1}_{n_k}, v_2^k \mathbf{1}_{n_k}, \ldots, v_{K-1}^k \mathbf{1}_{n_k}], \\ \quad \forall k \in \{1, 2, \ldots, K\}; \\ \text{If } t^{\mathbf{w}} > t^{\mathbf{w}*}, \; \mathbf{Y}_k^T \mathbf{1}_{n_k} = \mathbf{0}_{K-1}, \; \forall k \in \{1, 2, \ldots, K\}; \\ \text{If } t^{\mathbf{w}} = t^{\mathbf{w}*}, \; \forall k \in \{1, 2, \ldots, K\}, \; \mathbf{Y}_k = \mathbf{1}_{n_k}\mathbf{1}_{K-1}^T\mathbf{V}_k \\ \quad \text{or } \mathbf{Y}_k^T \mathbf{1}_{n_k} = \mathbf{0}_{K-1}, \end{cases}$

where $\mathbf{V}_k = \mathrm{diag}(v_1^k, v_2^k, \ldots, v_{K-1}^k) \in \mathbb{R}^{(K-1) \times (K-1)}$ is a diagonal matrix.

In particular, when $t^{\mathbf{w}} < t^{\mathbf{w}*}$, $\mathbf{Y}$ has the following properties:

(b-1) The columns of $\mathbf{Y}_k$ are constant vectors.

(b-2) Each column of $\mathbf{Y}$ has at least two nonzero cluster-wise constant components, and these constants have alternating signs such that their weighted sum equals 0 (i.e., $\sum_k n_k v_j^k = 0$, $\forall j \in \{1, 2, \ldots, K-1\}$).

(b-3) No two columns of $\mathbf{Y}$ have the same sign on the cluster-wise nonzero components.

Finally, $t^{\mathbf{w}*}$ satisfies:

(c) $t_{\mathrm{LB}}^{\mathbf{w}} \leq t^{\mathbf{w}*} \leq t_{\mathrm{UB}}^{\mathbf{w}}$, where
$\begin{cases} t_{\mathrm{LB}}^{\mathbf{w}} = \frac{\min_{k \in \{1,2,\ldots,K\}} S_{2:K}(\mathbf{L}_k^{\mathbf{w}})}{(K-1)n_{\max}}; \\ t_{\mathrm{UB}}^{\mathbf{w}} = \frac{\min_{k \in \{1,2,\ldots,K\}} S_{2:K}(\mathbf{L}_k^{\mathbf{w}})}{(K-1)n_{\min}}. \end{cases}$
In particular, $t_{\mathrm{LB}}^{\mathbf{w}} = t_{\mathrm{UB}}^{\mathbf{w}}$ when $c = 1$.

Theorem 2(a) establishes a phase transition in the increase of the normalized partial eigenvalue sum $\frac{S_{2:K}(\mathbf{L}^{\mathbf{w}})}{n}$ with respect to the aggregated noise level $t^{\mathbf{w}}$. When $t^{\mathbf{w}} \leq t^{\mathbf{w}*}$ the quantity $\frac{S_{2:K}(\mathbf{L}^{\mathbf{w}})}{n}$ is exactly $(K-1)t^{\mathbf{w}}$. When $t^{\mathbf{w}} > t^{\mathbf{w}*}$ the slope in $t^{\mathbf{w}}$ of $\frac{S_{2:K}(\mathbf{L})}{n}$ changes and the intercept $c^* = \min_{k \in \{1,2,\ldots,K\}}$ $\{\frac{S_{2:K}(\mathbf{L}_k^{\mathbf{w}})}{n}\} = \min_{k \in \{1,2,\ldots,K\}}\{\frac{\sum_{\ell=1}^{L} w_\ell S_{2:K}(\mathbf{L}_k^{(\ell)})}{n}\}$ depends on the cluster having the smallest aggregated partial eigenvalue sum given a layer weight vector $\mathbf{w}$. In particular, when all clusters have the same size (i.e., $n_{\max} = n_{\min} = \frac{n}{K}$) so that $c = 1$, $\frac{S_{2:K}(\mathbf{L})}{n}$ undergoes a slope change from $K-1$ to $\frac{(K-1)^2}{K}$ at the critical value $t^{\mathbf{w}} = t^{\mathbf{w}*}$. The visual illustration of Theorem 2(a) is displayed in Fig. S1 of the supplementary material.

Theorem 2(b) establishes a phase transition in cluster-wise separability of the eigenvector matrix $\mathbf{Y}$ for multilayer SGC. When $t^{\mathbf{w}} < t^{\mathbf{w}*}$, the conditions (b-1) to (b-3) imply that the rows of the cluster-wise components $\{\mathbf{Y}_k\}_{k=1}^{K}$ are coherent, and hence the row vectors in $\mathbf{Y}$ possess cluster-wise separability. On the other hand, when $t^{\mathbf{w}} > t^{\mathbf{w}*}$, the row sum of each $\mathbf{Y}_k$ is a zero vector, making $\mathbf{Y}_k$ incoherent. This means that the entries of each column in $\mathbf{Y}_k$ have alternating signs and the centroid of the row vectors in $\mathbf{Y}_k$ is centered at the origin. Therefore, K-means clustering on the rows of $\mathbf{Y}$ yields incorrect clusters.

Theorem 2(c) establishes upper and lower bounds on the critical threshold value $t^{\mathbf{w}*}$ of the aggregated noise level $t^{\mathbf{w}}$ given a layer weight vector $\mathbf{w}$. These bounds are determined by the cluster having the smallest aggregated partial eigenvalue sum $S_{2:K}(\mathbf{L}_k^{\mathbf{w}}) = \sum_{\ell=1}^{L} w_\ell S_{2:K}(\mathbf{L}_k^{(\ell)})$, the number of clusters K, and the largest and smallest cluster size ($n_{\max}$ and $n_{\min}$). When all cluster sizes are identical (i.e., $c = 1$), these bounds become tight (i.e., $t_{\mathrm{LB}}^{\mathbf{w}} = t_{\mathrm{UB}}^{\mathbf{w}}$). Moreover, by the nonnegativity of the layer weights we can obtain a universal lower bound on $t_{\mathrm{LB}}^{\mathbf{w}}$ for any $\mathbf{w} \in \mathcal{W}_L$, which is

$$t_{\mathrm{LB}}^{\mathbf{w}} = \frac{\min_{k \in \{1,2,\ldots,K\}} S_{2:K}(\mathbf{L}_k^{\mathbf{w}})}{(K-1)n_{\max}}$$

$$\geq \frac{\min_{k \in \{1,2,\ldots,K\}} \min_{\ell \in \{1,2,\ldots,L\}} S_{2:K}(\mathbf{L}_k^{(\ell)})}{(K-1)n_{\max}}. \quad (4)$$

Since $S_{2:K}(\mathbf{L}_k^{(\ell)})$ is a measure of connectivity for cluster k in the ℓ-th layer, the lower bound of $t_{\mathrm{LB}}^{\mathbf{w}}$ in (4) implies that the performance of multilayer SGC is indeed affected by the least connected cluster among all K clusters and across L layers. Specifically, if the graph in each layer is unweighted and $K = 2$, then $S_{2:K}(\mathbf{L}_k^{(\ell)}) = \lambda_2(\mathbf{L}_k^{(\ell)})$ reduces to the algebraic connectivity of cluster k in the ℓ-th layer. Similarly, a universal upper bound on $t_{\mathrm{UB}}^{\mathbf{w}}$ for any $\mathbf{w} \in \mathcal{W}_L$ is

$$t_{\mathrm{UB}}^{\mathbf{w}} \leq \frac{\min_{k \in \{1,2,\ldots,K\}} \max_{\ell \in \{1,2,\ldots,L\}} S_{2:K}(\mathbf{L}_k^{(\ell)})}{(K-1)n_{\min}}. \quad (5)$$

C. Phase Transitions in Multilayer SGC Under Block-Wise Non-Identical Noise

Under the block-wise non-identical noise model, the noise level of between-cluster edges between clusters i and j in the ℓ-th layer is characterized by the parameter $t_{ij}^{(\ell)} = p_{ij}^{(\ell)} \cdot \overline{W}_{ij}^{(\ell)}$, $1 \leq i, j \leq K$, $i \neq j$, and $1 \leq \ell \leq L$. Let $t_{\max}^{(\ell)} = \max_{1 \leq i,j \leq K, \, i \neq j} t_{ij}^{(\ell)}$ be the maximum noise level in the ℓ-th layer and let $t_{\max}^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell t_{\max}^{(\ell)}$ denote the aggregated maximum noise level given a layer weight vector $\mathbf{w} \in \mathcal{W}_L$.

Let $\mathbf{Y} \in \mathbb{R}^{n \times (K-1)}$ be the eigenvector matrix of $\mathbf{L}^{\mathbf{w}}$ under the block-wise non-identical noise model, and let $\widetilde{\mathbf{Y}} \in \mathbb{R}^{n \times (K-1)}$ be the eigenvector matrix of the graph Laplacian $\widetilde{\mathbf{L}}^{\mathbf{w}}$ of another graph generated by the block-wise identical noise model with aggregated noise level $t^{\mathbf{w}}$, which is independent of $\mathbf{L}$. Theorem 3 below specifies the distance between the subspaces spanned by the columns of $\mathbf{Y}$ and $\widetilde{\mathbf{Y}}$ by inspecting their principal angles [40]. Specifically, since $\mathbf{Y}$ and $\widetilde{\mathbf{Y}}$ both have orthonormal columns, the vector $\mathbf{a}$ of

$K - 1$ principal angles between their column spaces is $\mathbf{a} = [\cos^{-1} \sigma_1(\mathbf{Y}^T \widetilde{\mathbf{Y}}), \ldots, \cos^{-1} \sigma_{K-1}(\mathbf{Y}^T \widetilde{\mathbf{Y}})]^T$, where $\sigma_k(\mathbf{M})$ is the k-th largest singular value of a real rectangular matrix $\mathbf{M}$. Let $\mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}}) = \mathrm{diag}(\mathbf{a})$, and let $\sin \mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}})$ be defined entrywise. When $t^{\mathbf{w}} < t^{\mathbf{w}*}$, Theorem 3 provides an upper bound on the Frobenius norm of $\sin \mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}})$, which is denoted by $\| \sin \mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}}) \|_F$. Moreover, if $t^{\mathbf{w}}_{\max} < t^{\mathbf{w}*}$, where $t^{\mathbf{w}*}$ is the critical threshold value for the block-wise identical noise model as specified in Theorem 2, then $\| \sin \mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}}) \|_F$ can be further bounded.

Theorem 3 (block-wise non-identical noise).

Under the multilayer signal plus noise model in Section III-B with maximum noise level $\{t^{(\ell)}_{\max}\}_{\ell=1}^L$ for each layer, given a layer weight vector $\mathbf{w} \in \mathcal{W}_L$, let $t^{\mathbf{w}*}$ be the critical threshold value for the block-wise identical noise model specified by Theorem 2, and define $\delta_{t^{\mathbf{w}},n} = \min\{t^{\mathbf{w}}, |\lambda_{K+1}(\frac{\mathbf{L}^{\mathbf{w}}}{n}) - t^{\mathbf{w}}|\}$. For a fixed $t^{\mathbf{w}}$, if $t^{\mathbf{w}} < t^{\mathbf{w}*}$ and $\delta_{t^{\mathbf{w}},n} \to \delta_{t^{\mathbf{w}}} > 0$ as $n_k \to \infty \; \forall k$, the following statement holds almost surely as $n_k \to \infty \; \forall k$ and $\frac{n_{\min}}{n_{\max}} \to c > 0$:

$$\| \sin \mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}}) \|_F \leq \frac{\| \mathbf{L}^{\mathbf{w}} - \widetilde{\mathbf{L}}^{\mathbf{w}} \|_F}{n \delta_{t^{\mathbf{w}}}}. \tag{6}$$

Furthermore, let $t^{\mathbf{w}}_{\max} = \sum_{\ell=1}^L w_\ell t^{(\ell)}_{\max}$. If $t^{\mathbf{w}}_{\max} < t^{\mathbf{w}*}$,

$$\| \sin \mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}}) \|_F \leq \min_{t^{\mathbf{w}} \leq t^{\mathbf{w}}_{\max}} \frac{\| \mathbf{L}^{\mathbf{w}} - \widetilde{\mathbf{L}}^{\mathbf{w}} \|_F}{n \delta_{t^{\mathbf{w}}}}. \tag{7}$$

Theorem 3 shows that the subspace distance $\| \sin \mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}}) \|_F$ is upper bounded by (6), where $\widetilde{\mathbf{Y}}$ is the eigenvector matrix of $\widetilde{\mathbf{L}}^{\mathbf{w}}$ under the block-wise identical noise model when its aggregated noise level $t^{\mathbf{w}} < t^{\mathbf{w}*}$. Furthermore, if the aggregated maximum noise level $t^{\mathbf{w}}_{\max} < t^{\mathbf{w}*}$, then a tight upper bound on $\| \sin \mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}}) \|_F$ can be obtained by (7). Therefore, using the phase transition results of the cluster-wise separability in $\widetilde{\mathbf{Y}}$ as established in Theorem 2(b), when $t^{\mathbf{w}}_{\max} < t^{\mathbf{w}*}$, cluster-wise separability in $\mathbf{Y}$ can be expected provided that $\| \sin \mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}}) \|_F$ is small.

V. MIMOSA: MULTILAYER ITERATIVE MODEL ORDER SELECTION ALGORITHM

The phase transition analysis established in Section IV shows that under the multilayer signal plus noise model in Section III-B, the performance of multilayer spectral graph clustering (SGC) via convex layer aggregation can be separated into two regimes: a reliable regime where high clustering accuracy is guaranteed, and an unreliable regime where high clustering accuracy is impossible. We have specified the critical threshold value of the aggregated noise level that separates these two regimes, and have shown that the assigned layer weight vector $\mathbf{w}$ for convex layer aggregation indeed affects the accuracy of multilayer SGC.

In this section, we use the established phase transition criterion to propose a multilayer SGC algorithm, which we call multilayer iterative model order selection algorithm (MIMOSA). MIMOSA is a multilayer SGC algorithm that features automated model order selection for determining the number of clusters (K) and the layer weight vector $\mathbf{w}$. It works by
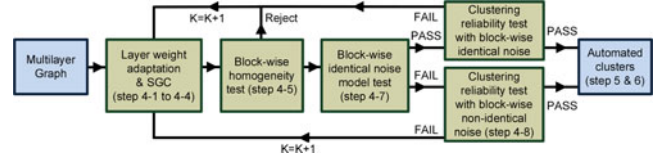

Fig. 1. Flow diagram of the proposed multilayer iterative model order selection algorithm (MIMOSA) for multilayer spectral graph clustering (SGC).

incrementally partitioning the aggregated graph $G^{\mathbf{w}}$ into K clusters, adjusting the layer weight vector, and finding the minimal number of clusters such that the output clusters are estimated to be in the reliable regime. The flow diagram of MIMOSA is displayed in Fig. 1, and the complete algorithm is summarized in Algorithm 1. Since part of MIMOSA uses the same statistical testing methods developed for single-layer graphs in [39], the details on the V-test and Wilk's test are omitted. The interested reader can refer to [39, Sec. V].

A. Input Data

The input data for MIMOSA is summarized as follows. (1) a multilayer graph $\{G_\ell\}_{\ell=1}^L$ of L layers, where each layer G_ℓ is an undirected weighted graph. (2) an initial layer weight vector $\mathbf{w}^{\mathrm{ini}} \in \mathcal{W}_L$. $\mathbf{w}^{\mathrm{ini}}$ can be specified according to domain knowledge, or it can be a uniform vector such that $w_\ell = \frac{1}{L} \; \forall \ell$. (3) a layer weight adaptation coefficient set $\mathcal{T} = \{\tau_z\}_{z=1}^{|\mathcal{T}|}$. The coefficients in $\mathcal{T}$ play a role in the process of layer weight adaptation in Section V-B. (4) a p-value significance level η that is used for the block-wise homogeneity test in Section V-C. (5) confidence interval parameters $\{\alpha_\ell\}_{\ell=1}^L$ of each layer under the block-wise identical noise model for clustering reliability evaluation in Section V-D. (6) confidence interval parameters $\{\alpha'_\ell\}_{\ell=1}^L$ of each layer under the block-wise non-identical noise model for clustering reliability evaluation in Section V-E.

B. Layer Weight Adaptation

Given an initial layer weight vector $\mathbf{w}^{\mathrm{ini}}$ and the number of clusters K in the iterative process (step 4) of MIMOSA, we propose to adjust the layer weight vector $\mathbf{w}$ for convex layer aggregation by estimating the noise level $\{\widehat{t}^{(\ell)}_{\mathrm{ini}}\}_{\ell=1}^L$ under the block-wise identical noise model in Section III-B. Specifically, given K clusters $\{\mathcal{C}_k^{\mathbf{w}^{\mathrm{ini}}}\}_{k=1}^K$ of size $\{\widehat{n}_k\}_{k=1}^K$ via multilayer SGC with $\mathbf{w}^{\mathrm{ini}}$, let $\{\widehat{\mathbf{C}}_{ij}^{(\ell)}\}$ and $\{\widehat{\mathbf{F}}_{ij}^{(\ell)}\}$ be the interconnection matrix and edge weight matrix of $\{\mathcal{C}_k^{\mathbf{w}^{\mathrm{ini}}}\}_{k=1}^K$, respectively, for $1 \leq i, j \leq K$, $i \neq j$, and $1 \leq \ell \leq L$. Then the noise level estimator under the block-wise identical noise model is

$$\widehat{t}^{(\ell)}_{\mathrm{ini}} = \widehat{p}^{(\ell)} \cdot \widehat{\overline{W}}^{(\ell)}, \tag{8}$$

for $\ell \in \{1, 2, \ldots, L\}$, where $\widehat{p}^{(\ell)} = \frac{\sum_{i=1}^K \sum_{j=i+1}^K \widehat{m}_{ij}^{(\ell)}}{\sum_{i=1}^K \sum_{j=i+1}^K \widehat{n}_i \widehat{n}_j}$ is the maximum likelihood estimator (MLE) of $p^{(\ell)}$, $\widehat{m}_{ij}^{(\ell)} = \mathbf{1}_{\widehat{n}_i}^T \widehat{\mathbf{C}}_{ij}^{(\ell)} \mathbf{1}_{\widehat{n}_j}$ is the number of between-cluster edges of clusters i and j in the ℓ-th layer, and $\widehat{\overline{W}}^{(\ell)}$ is the average of between-cluster edge weights in the ℓ-th layer.

Since the estimates $\{\widehat{t}_{\text{ini}}^{(\ell)}\}_{\ell=1}^L$ reflect the noise level in each layer, we propose to adjust the layer weight vector $\mathbf{w} \in \mathcal{W}_L$ with a nonnegative regularization parameter $\tau \in \mathcal{T}$. The adjusted $\mathbf{w}$ layer weight vector is inversely proportional to the estimated noise level, which is defined as

$$w_\ell \propto \frac{w_\ell^{\text{ini}}}{1 + \tau \cdot \widehat{t}_{\text{ini}}^{(\ell)}}, \qquad (9)$$

for $\ell \in \{1, 2, \ldots, L\}$. Note that if $\tau = 0$, then $\mathbf{w}$ reduces to $\mathbf{w}^{\text{ini}}$. In addition, larger τ further penalizes the layers of high noise level by assigning less weight for convex layer aggregation. In addition, to enable the computation of the function $\min_{k \in \{1,2,\ldots,K\}} S_{2:K}(\sum_{\ell=1}^L w_\ell \cdot \mathbf{L}_k^{(\ell)})$ for clustering reliability test in the following step of MIMOSA, the detected clusters $\{\mathcal{C}_k^{\mathbf{w}}\}_{k=1}^K$ are deemed unreliable if the size of any detected cluster is less than K.

C. Block-Wise Homogeneity Test

Given K clusters $\{\mathcal{C}_k^{\mathbf{w}}\}_{k=1}^K$ with respect to a layer weight vector $\mathbf{w}$ in the iterative process (step 4) of MIMOSA, we implement a block-wise homogeneity test for each block $\widehat{\mathbf{C}}_{ij}^{(\ell)}$ accounting for the interconnection matrix of clusters i and j in the ℓ-th layer, in order to test the assumption of the block-wise homogeneity noise model as assumed in Section III-B, which is the cornerstone of the phase transition results established in Section IV.

In particular, we use the V-test developed in [39, Algorithm 1] to test the assumption of block-wise homogeneity noise model. Given x independent binomial random variables, the V-test tests that they are all identically distributed [45]. Here we apply the V-test to the row sums of $\widehat{\mathbf{C}}_{ij}^{(\ell)}$. The block-wise homogeneity test on $\widehat{\mathbf{C}}_{ij}^{(\ell)}$ rejects the block-wise homogeneous hypothesis if its p-value$(i, j, \ell) \leq \eta$, where η is the desired single comparison significance level.

In step 4-5 of MIMOSA, the layer weight vector $\mathbf{w}$ and the corresponding clusters $\{\mathcal{C}_k^{\mathbf{w}}\}_{k=1}^K$ are deemed unreliable if there exists some $\widehat{\mathbf{C}}_{ij}^{(\ell)}$ such that its p-value does not exceed the significance level.

D. Clustering Reliability Test Under the Block-Wise Identical Noise Model

In the iterative process of step 4 in MIMOSA, if every interconnection matrix $\widehat{\mathbf{C}}_{ij}^{(\ell)}$ passes the block-wise homogeneity test in Section V-C, the identified clusters $\{\mathcal{C}_k^{\mathbf{w}}\}_{k=1}^K$ are then used to test the clustering reliability under the block-wise identical noise model in Section III-B. In particular, for each layer ℓ, we first estimate the noise level parameter $\widehat{p}_{ij}^{(\ell)}$ for every cluster pair i and j as $\widehat{p}_{ij}^{(\ell)} = \frac{\widehat{m}_{ij}^{(\ell)}}{\widehat{n}_i \widehat{n}_j}$, where $\widehat{p}_{ij}^{(\ell)}$ is an MLE of $p_{ij}^{(\ell)}$. We then use the generalized log-likelihood ratio test (GLRT) developed in [39, Sec. V-C] to specify an asymptotic $100(1 - \alpha_\ell)\%$ confidence interval for $p^{(\ell)}$ accounting for the block-wise identical noise level parameter for each layer. In particular, the GLRT is a test statistic of the null hypothesis *all block-wise noises are*

independent and identical versus the alternative hypothesis all block-wise noises are independent but not identical.

If the estimated block-wise identical noise level parameter $\widehat{p}^{(\ell)} = \frac{\sum_{i=1}^K \sum_{j=i+1}^K \widehat{m}_{ij}^{(\ell)}}{\sum_{i=1}^K \sum_{j=i+1}^K \widehat{n}_i \widehat{n}_j}$ is within the $100(1 - \alpha_\ell)\%$ confidence interval for every ℓ, then the clusters $\{\mathcal{C}_k^{\mathbf{w}}\}_{k=1}^K$ satisfy the block-wise identical noise model, and therefore we can apply the phase transition results in Theorem 2 to evaluate the clustering reliability. In particular, we compare the estimated aggregated noise level $\widehat{t}^{\mathbf{w}}$ with the estimated phase transition lower bound $\widehat{t}_{\text{LB}}^{\mathbf{w}}$ of $t_{\text{LB}}^{\mathbf{w}}$ in Theorem 2 (c), where $\widehat{t}^{\mathbf{w}} = \sum_{\ell=1}^L w_\ell \widehat{t}^{(\ell)} = \sum_{\ell=1}^L w_\ell \cdot \widehat{p}^{(\ell)} \cdot \overline{\widehat{W}}^{(\ell)}$, and

$$\widehat{t}_{\text{LB}}^{\mathbf{w}} = \frac{\min_{k \in \{1,2,\ldots,K\}} S_{2:K}(\sum_{\ell=1}^L w_\ell \cdot \widehat{\mathbf{L}}_k^{(\ell)})}{(K-1) \cdot \widehat{n}_{\max}}, \qquad (10)$$

where $\widehat{\mathbf{L}}_k^{(\ell)}$ is the graph Laplacian matrix of within-cluster edges of cluster $\mathcal{C}_k^{\mathbf{w}}$ in the ℓ-th layer, $S_{2:K}(\sum_{\ell=1}^L w_\ell \cdot \widehat{\mathbf{L}}_k^{(\ell)}) = \sum_{z=2}^K \lambda_z(\sum_{\ell=1}^L w_\ell \cdot \widehat{\mathbf{L}}_k^{(\ell)})$, and $\widehat{n}_{\max} = \max_{k \in \{1,2,\ldots,K\}} \widehat{n}_k$. Therefore, using Theorem 2, the clusters $\{\mathcal{C}_k^{\mathbf{w}}\}_{k=1}^K$ are deemed reliable if $\widehat{t}^{\mathbf{w}} < \widehat{t}_{\text{LB}}^{\mathbf{w}}$, since the eigenvector matrix $\mathbf{Y}$ used for multilayer SGC possesses cluster-wise separability. The lower bound in (10) also specifies the effect of cluster size on clustering reliability test. Ignoring the term in the numerator, a set of imbalanced clusters having larger $\widehat{n}_{\max}$ leads to smaller $\widehat{t}_{\text{LB}}^{\mathbf{w}}$ and hence implies a more difficult clustering problem.

E. Clustering Reliability Test Under the Block-Wise Non-Identical Noise Model

In the iterative process of step 4 in MIMOSA, if every interconnection matrix $\widehat{\mathbf{C}}_{ij}^{(\ell)}$ passes the block-wise homogeneity test in Section V-C, but some layers fail the clustering reliability test under the block-wise identical noise model in Section V-D, the identified clusters $\{\mathcal{C}_k^{\mathbf{w}}\}_{k=1}^K$ are then used to test the clustering reliability under the block-wise non-identical noise model in Section III-B based on Theorem 3. Given a layer weight vector $\mathbf{w}$, the noise level estimates $\{\widehat{t}_{ij}^{(\ell)}\}$, and the estimate $\widehat{t}_{\text{LB}}^{\mathbf{w}}$ of the phase transition lower bound in (10), we compare the maximum noise level $\widehat{t}_{\max}^{(\ell)} = \max_{1 \leq i,j \leq K, i \neq j} \widehat{t}_{ij}^{(\ell)}$ with $\widehat{t}_{\text{LB}}^{\mathbf{w}}$ for each layer ℓ. In the supplementary file we show that if the estimated maximum noise level $\widehat{t}_{\max}^{(\ell)}$ of each layer ℓ satisfies a certain condition (condition (S39) in the supplementary file), then if the aggregated maximum noise level $\widehat{t}_{\max}^{\mathbf{w}} = \sum_{\ell=1}^L w_\ell \widehat{t}_{\max}^{(\ell)} < \widehat{t}_{\text{LB}}^{\mathbf{w}}$, by Theorem 3 the identified clusters $\{\mathcal{C}_k\}_{k=1}^K$ are deemed reliable with high probability.

F. A Signal-to-Noise Ratio Criterion for Final Clustering Results

In step 4 of MIMOSA, given the number of clusters K, if MIMOSA finds any feasible layer weight vector that passes the clustering reliability tests in Sections V-D or V-E, it then stores the vector in the set $\mathcal{W}_{\text{feasible}}$, and stops increasing K. This means that MIMOSA has identified a set of reliable clustering results of the same number of clusters K based on the clustering reliability tests. To select the best clustering result from the feasible set, in step 5 we use the phase transition results established

---

Algorithm 1: Multilayer iterative model order selection algorithm (MIMOSA) for multilayer SGC.

Input:
(1) a multilayer graph $\{G_\ell\}_{\ell=1}^L$
(2) an initial layer weight vector $\mathbf{w}^{\mathrm{ini}} \in \mathcal{W}_L$
(3) a layer weight adaptation coefficient set $\mathcal{T} = \{\tau_z\}_{z=1}^{|\mathcal{T}|}$
(4) a p-value significance level η
(5) confidence interval parameters $\{\alpha_\ell\}_{\ell=1}^L$ under the block-wise identical noise model for each layer
(6) confidence interval parameters $\{\alpha'_\ell\}_{\ell=1}^L$ under the block-wise non-identical noise model for each layer

Output: K clusters $\{\mathcal{C}_k\}_{k=1}^K$

Initialization: $K = 2$. Flag $= 1$. $\mathcal{W}_{\mathrm{reliable}} = \varnothing$.

while Flag$= 1$ **do**
   1. Compute $\mathbf{Y} \in \mathbb{R}^{n \times (K-1)}$ of $\mathbf{L^{w^{\mathrm{ini}}}}$
   2. Obtain K clusters $\{\mathcal{C}_k^{\mathbf{w}^{\mathrm{ini}}}\}_{k=1}^K$ by implementing K-means algorithm on the rows of $\mathbf{Y}$
   3. Estimate the noise level $\{\widehat{t}_{\mathrm{ini}}^{(\ell)}\}_{\ell=1}^L$ from (8)
   4. Layer weight adaptation and multilayer SGC reliability tests:
   for $z = 1$ to $|\mathcal{T}|$ **do**
      4-1. Layer weight adaptation: $w_\ell \leftarrow w_\ell^{\mathrm{ini}} \cdot (1 + \tau_z \cdot \widehat{t}^{(\ell)})^{-1}, \forall \ell \in \{1, 2, \dots, L\}$
      4-2. Layer weight normalization: $w_\ell \leftarrow \frac{w_\ell}{\sum_{\ell'=1}^L w_{\ell'}}, \forall \ell \in \{1, 2, \dots, L\}$
      4-3. Compute $\mathbf{Y} \in \mathbb{R}^{n \times (K-1)}$ of $\mathbf{L^w}$
      4-4. Obtain K clusters $\{\mathcal{C}_k^{\mathbf{w}}\}_{k=1}^K$ by implementing K-means algorithm on the rows of $\mathbf{Y}$
      4-5. **Block-wise homogeneity test:** calculate p-value$(i, j, \ell), \forall i, j, \ell, 1 \le i, j \le K, i \ne j$, and $1 \le \ell \le L$
      if p-value$(i, j, \ell) \le \eta$ for some (i, j, ℓ) **then**
         Go back to step 4-1 with $z = z + 1$
      end if
      4-6. Estimate the noise level $\{\widehat{t}_{ij}^{(\ell)}\}$ for all i, j, ℓ and estimate $\widehat{t}_{\mathrm{LB}}^{\mathbf{w}}$ from (10)
      4-7. **Block-wise identical noise test:** estimate the aggregated noise level $\widehat{t}^{\mathbf{w}} = \sum_{\ell=1}^L w_\ell \cdot \widehat{t}^{(\ell)}$
      if $\widehat{t}^{(\ell)}$ lies in the $100(1 - \alpha_\ell)\%$ confidence interval $\forall \ell$ **then**
        if $\widehat{t}^{\mathbf{w}} < \widehat{t}_{\mathrm{LB}}^{\mathbf{w}}$ **then**
           Flag$= 0$. $\mathcal{W}_{\mathrm{reliable}} = \mathcal{W}_{\mathrm{reliable}} \cup \{\mathbf{w}\}$.
        end if
      else if $\widehat{t}^{(\ell)}$ does not lie in the $100(1 - \alpha_\ell)\%$ confidence interval for some ℓ
        4-8. **Block-wise non-identical noise test:** estimate the aggregated maximum noise level $\widehat{t}_{\max}^{\mathbf{w}} = \sum_{\ell=1}^L w_\ell \widehat{t}_{\max}^{(\ell)}$
        if $\prod_{i=1}^K \prod_{j=i+1}^K F_{ij}(\frac{\widehat{t}_{\mathrm{LB}}^{\mathbf{w}}}{\widehat{\overline{W}}_{ij}^{(\ell)}}, \widehat{p}_{ij}^{(\ell)}) \ge 1 - \alpha'_\ell \; \forall \ell$ **then**
           if $\widehat{t}_{\max}^{\mathbf{w}} < \widehat{t}_{\mathrm{LB}}^{\mathbf{w}}$ **then**
              Flag$= 0$. $\mathcal{W}_{\mathrm{reliable}} = \mathcal{W}_{\mathrm{reliable}} \cup \{\mathbf{w}\}$.
           end if
        end if
      end if
      Go back to step 4-1 with $z = z + 1$
   end for
   if Flag$= 1$ **then**
      Go back to step 1 with $K = K + 1$
   end if
end while
5. SNR criterion: select $\mathbf{w}^* = \arg\max_{\mathbf{w} \in \mathcal{W}_{\mathrm{reliable}}} \frac{\widehat{t}_{\mathrm{LB}}^{\mathbf{w}}}{\widehat{t}^{\mathbf{w}}}$
6. Output final clustering result: $\{\mathcal{C}_k\}_{k=1}^K \leftarrow \{\mathcal{C}_k^{\mathbf{w}^*}\}_{k=1}^K$

---

in Section IV to define a signal-to-noise ratio (SNR) for each clustering result, which is

$$\mathrm{SNR}^{\mathbf{w}} = \frac{\widehat{t}_{\mathrm{LB}}^{\mathbf{w}}}{\widehat{t}^{\mathbf{w}}}. \tag{11}$$

$\widehat{t}_{\mathrm{LB}}^{\mathbf{w}}$ can be viewed as the aggregated signal strength of within-cluster edges, and $\widehat{t}^{\mathbf{w}}$ is the the aggregated noise level across layers. Therefore, the final clustering result is the clusters $\{\mathcal{C}_k^{\mathbf{w}^*}\}_{k=1}^K$, where $\mathbf{w}^* = \arg\max_{\mathbf{w} \in \mathcal{W}_{\mathrm{feasible}}} \mathrm{SNR}^{\mathbf{w}}$ is the layer weight vector having the largest SNR in the set $\mathcal{W}_{\mathrm{feasible}}$.

G. Computational Complexity Analysis

The overall computational complexity of MIMOSA is $O(|\mathcal{T}|K^3(\widetilde{m} + n))$, where K is the number of output clusters, n is the number of nodes, and $\widetilde{m} = \sum_{\ell=1}^L |\mathcal{E}_\ell|$ is the sum of total number of edges in each layer. The analysis is as follows.
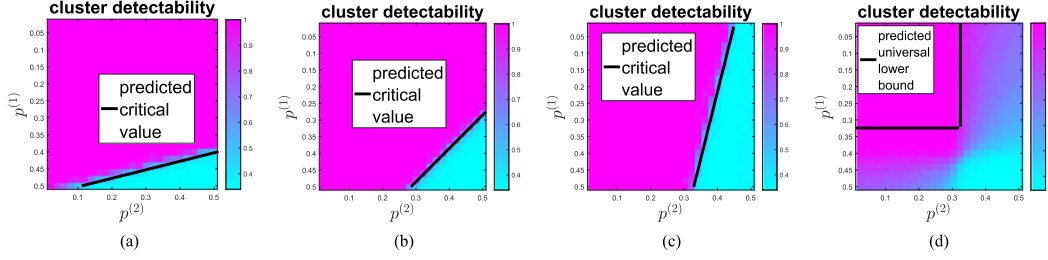
Fig. 2. Phase transitions in the accuracy of multilayer SGC with respect to different layer weight vector $\mathbf{w} = [w_1\ w_2]^T$ for the two-layer correlated graph model. $n_1 = n_2 = n_3 = 1000$, $q_{11} = 0.3$, $q_{10} = 0.2$, $q_{01} = 0.1$, and $q_{00} = 0.4$. The results are averaged over 10 runs. (a)–(c) for a given $\mathbf{w}$, the variations in the noise level $\{p^{(\ell)}\}_{\ell=1}^2$ indeed separates the accuracy of multilayer SGC into a reliable regime and an unreliable regime. Furthermore, the critical value that separates these two regimes is successfully predicted by Theorem 2. (d) The geometric mean over $\mathbf{w}$, where w_1 is uniformly drawn from $[0, 1]$ with unit interval 0.1. There is a universal region of perfect cluster detectability that includes the region specified by the universal phase transition lower bound in (4). (a) $(w_1, w_2) = (0.8, 0.2)$ (b) $(w_1, w_2) = (0.5, 0.5)$ (c) $(w_1, w_2) = (0.2, 0.8)$ (d) average over $\mathbf{w}$.

Fixing model order K and regularization parameter $\tau \in \mathcal{T}$ in the MIMOSA iteration, as displayed in Fig. 1, there are three main contributions to the computational complexity of MIMOSA: (i) Incremental eigenpair computation - acquiring an additional smallest eigenvector for augmenting $\mathbf{Y}$ of $\mathbf{L}^{\mathbf{w}}$ takes $O(\widetilde{m} + n)$ operations via power iteration [46], [47], since the maximum number of nonzero entries in $\mathbf{L}^{\mathbf{w}}$ is $\widetilde{m} + n$. (ii) Parameter estimation - estimating the RIM parameters $\{p_{ij}^{(\ell)}\}$ and $\{\overline{W}_{ij}^{(\ell)}\}$ takes $O(\widetilde{m})$ operations since they only depend on the number of edges and edge weights in each layer. Estimating t_{LB} takes $O(K(\widetilde{m} + n) \cdot K) = O(K^2(\widetilde{m} + n))$ operations for computing the numerator in (10). (iii) K-means clustering - $O(nK^2)$ operations [48] for clustering n data points of dimension $K - 1$ into K groups. Unfixing τ, iterating this process over the elements in $\mathcal{T}$ takes $O(|\mathcal{T}|K^2(\widetilde{m} + n))$ operations. Finally, if MIMOSA outputs K clusters, then the overall computational complexity is $O(|\mathcal{T}|K^3(\widetilde{m} + n))$.

VI. NUMERICAL EXPERIMENTS

To validate the phase transition results in the accuracy of multilayer SGC via convex layer aggregation established in Section IV, we generate synthetic multilayer graphs from a two-layer correlated multilayer graph model. Specifically, we generate edge connections within and between $K = 3$ equally-sized ground-truth clusters on $L = 2$ layers G_1 and G_2. The two layers G_1 and G_2 are correlated since their edge connections are generated in the following manner. For every node pair (u, v) of the same cluster, with probability q_{11} there is a within-cluster edge (u, v) in G_1 and G_2, with probability q_{10} there is a within-cluster edge (u, v) in G_1 but not in G_2, with probability q_{01} there is a within-cluster edge (u, v) in G_2 but not in G_1, and with probability q_{00} there is no edge (u, v) in G_1 and G_2. These four parameters $\{q_{xy}\}_{x,y\in\{0,1\}}$ are nonnegative and sum to 1. For between-cluster edges, we adopt the block-wise identical noise model in Section III-B such that for each layer ℓ, the edge connection between every node pair from different clusters is an i.i.d. Bernoulli random variable with parameter $p^{(\ell)}$.

A. Phase Transitions in Multilayer SGC via Convex Layer Aggregation

By varying the noise level $\{p^{(\ell)}\}_{\ell=1}^2$, Fig. 2 shows the accuracy of multilayer SGC with respect to different layer weight vector $\mathbf{w} = [w_1\ w_2]^T$ and the averaged result over $\mathbf{w}$, where

the accuracy is evaluated in terms of cluster detectability. Let $\{\mathcal{C}_k\}_{k=1}^K$ and $\{\mathcal{C}_k'\}_{k=1}^K$ denote the detected and ground-truth clusters, respectively, and let $|\mathcal{C}_k \cap \mathcal{C}_k'|$ denote the number of common nodes in $\mathcal{C}_k$ and $\mathcal{C}_k'$. Cluster detectability is defined as $\max_{\{\mathcal{C}_{\widetilde{k}}\}\in\mathrm{Perm}(\{\mathcal{C}_k\})} \frac{1}{n} \sum_{k=1}^K |\mathcal{C}_{\widetilde{k}} \cap \mathcal{C}_k'|$, where $\mathrm{Perm}(\{\mathcal{C}_k\})$ is the set of all possible cluster label permutations of the detected clusters. In other words, cluster detectability requires consistency between the detected and ground-truth clusters. Given a fixed $\mathbf{w}$, as proved in Theorem 2, Fig. 2(a)–(c) show that there is indeed a phase transition in cluster detectability that separates the noise level $\{p^{(\ell)}\}_{\ell=1}^2$ into two regimes: a reliable regime where high clustering accuracy is guaranteed, and an unreliable regime where high clustering accuracy is impossible. Furthermore, the critical value of $\{p^{(\ell)}\}_{\ell=1}^2$ that separates these two regimes are successfully predicted by Theorem 2(c), which validates the phase transition analysis. Fig. 2(d) shows the geometric mean of cluster detectability from different layer weight vectors. There is a universal region of perfect cluster detectability that includes the region specified by the universal phase transition lower bound in (4).

B. The Effect of Layer Weight Vector on Multilayer SGC via Convex Layer Aggregation

Next we investigate the effect of layer weight vector $\mathbf{w}$ on multilayer SGC via convex layer aggregation given fixed noise levels $\{p^{(\ell)}\}_{\ell=1}^2$. In the two-layer graph setting, since by definition $w_2 = 1 - w_1$, it suffices to study the effect of w_1 on clustering accuracy. Fig. 3 shows the clustering accuracy by varying w_1 under the two-layer correlated graph model. As shown in Fig. 3(a), if each layer has low noise level, then any layer weight vector $\mathbf{w} \in \mathcal{W}_2$ can lead to correct clustering result. If one layer has high noise level, Fig. 3(b) and (c) show that there exists a critical value $w_1^\star \in [0, 1]$ that separates the cluster detectability into a reliable regime and an unreliable regime. In particular, Theorem 2 implies that the critical value $w_1^\star$, if existed, satisfies the condition $t^{\mathbf{w}} = t^{\mathbf{w}^*}$ when $\mathbf{w} = [w_1^\star, 1 - w_1^\star]^T = \mathbf{w}^*$, which is equivalent to

$$\frac{K-1}{K}\left[w_1^\star p^{(1)} + (1 - w_1^\star)p^{(2)}\right] = w_1^\star \cdot \min_{k\in\{1,2,\ldots,K\}} S_{2:K}\left(\frac{\mathbf{L}_k^{(1)}}{n}\right)$$

$$+ (1 - w_1^\star) \cdot \min_{k\in\{1,2,\ldots,K\}} S_{2:K}\left(\frac{\mathbf{L}_k^{(2)}}{n}\right). \tag{12}$$
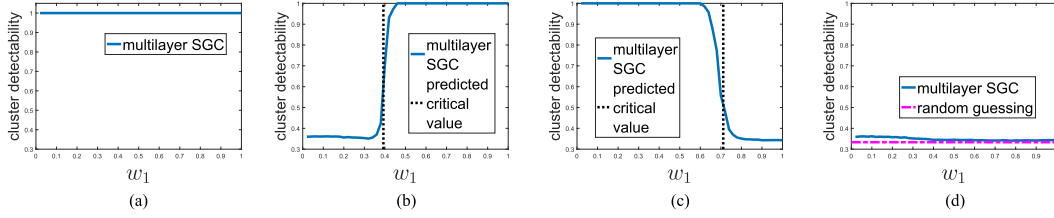
Fig. 3. The effect of the layer weight vector $\mathbf{w} = [w_1 \; w_2]^T$ on the accuracy of multilayer SGC with respect to different noise levels $\{p^{(\ell)}\}_{\ell=1}^{2}$ for the two-layer correlated graph model. $n_1 = n_2 = n_3 = 1000$, $q_{11} = 0.3$, $q_{10} = 0.2$, $q_{01} = 0.1$, and $q_{00} = 0.4$. The results are averaged over 50 runs. (a) In the case of low noise level for each layer, any layer weight vector $\mathbf{w} \in \mathcal{W}_2$ can lead to correct clustering result. (b) and (c)If one layer has high noise level, then there may exist a critical value $w_1^* \in [0,1]$ that separates the cluster detectability into a reliable regime and an unreliable regime. Furthermore, the critical value w_1^* is shown to satisfy the equation in (12) derived from Theorem 2. (d) In the case of high noise level for each layer, no layer weight vector can lead to correct clustering result, and the cluster detectability is similar to random guessing of clustering accuracy 33.33%. (a) $(p^{(1)}, p^{(2)}) = (0.2, 0.2)$ (b)$(p^{(1)}, p^{(2)}) = (0.2, 0.5)$ (c) $(p^{(1)}, p^{(2)}) = (0.5, 0.2)$ (d) $(p^{(1)}, p^{(2)}) = (0.5, 0.5)$.

TABLE I
SUMMARY OF REAL-WORLD MULTILAYER GRAPH DATASETS

Dataset	# of layers	ground-truth cluster labels and cluster sizes
VC 7th grader	3	boys (12) girls (17)
Leskovec-Ng collaboration network	4	Leskovec's collaborator (87) Ng's collaborator (104)
109th Congress votes - Budget	4	Democratic (45) Republican (55)
109th Congress votes - Energy	2	Democratic (45) Republican (55)
109th Congress votes- Security	2	Democratic (45) Republican (55)
Reality mining	2	None
London transportation network	2	None
Human H1V1 genetic interaction	5	None
Pierre Auger coauthorship	16	None

It is observed in Fig. 3(b) and (c) that the empirical critical value $w_1^\star$ matches the predicted value from (12). Lastly, as shown in Fig. 3(d), if each layer has high noise level, then no layer weight vector can lead to correct clustering result, and the corresponding cluster detectability is similar to random guessing of clustering accuracy $\frac{1}{K} \approx 33.33\%$.

VII. MIMOSA ON REAL-WORLD MULTI-LAYER GRAPHS

A. Dataset Descriptions

In this section, we apply MIMOSA to 9 real-world multilayer graphs and compute the external and internal clustering metrics for quality assessment. The statistics of the 9 real-world multilayer graphs are summarized in Table I, and the details are described as follows.

1) *VC 7th grader social network* [49]: This dataset is based on a survey of social relations among 29 7th grade students in Victoria, Australia, including 12 boys and 17 girls. A 3-layer graph is created based on different relationships, including "friends you get on with", "your best friends", and "friends you prefer to work with" in the class. For each layer we only retain the edges where there is mutual agreement between every student pair.

2) *Leskovec-Ng collaboration network*[2]: We collected the coauthors of Prof. Jure Leskovec or Prof. Andrew Ng at Stanford University from ArnetMiner [50] from year 1995 to year 2014. In total, there are 191 researchers
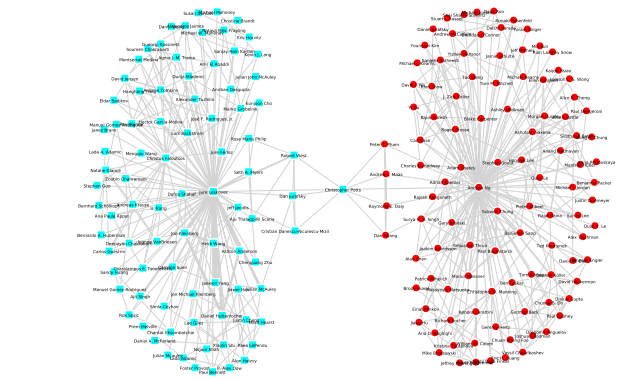
Fig. 4. Ground-truth clusters of the collected Leskovec-Ng collaboration network. Nodes represent researchers, edges represent the strength of coauthorship [51], [52], and colors and shapes represent two clusters - "Leskovec's collaborator" (cyan square) or "Ng's collaborator" (red circle).

in this dataset. We partition coauthorship over a 20-year period into 4 different 5-year intervals and hence create a 4-layer multilayer graph. For each layer, there is an edge between two researchers if they coauthored at least one paper in the 5-year interval. For every edge in each layer, we adopt the temporal collaboration strength as the edge weight [51], [52]. Notably, while Prof. Leskovec and Prof. Ng both were members of the same department, there is no record of coauthorship between them on Arnet-Miner. However, they are connected through a common co-author, Christopher Potts. As a result the full collaboration network among 191 researchers is a connected graph. We manually label each researcher by either "Leskovec's collaborator" or "Ng's collaborator" based on the collaboration frequency, and use the labels as the ground-truth cluster assignment. The ground-truth clusters with researcher names are displayed in Fig. 4.

3) *109th Congress votes*: We collected the votes of 100 senators of the 109th U.S. Congress to create 3 multilayer graph datasets based on the topic area of each bill on which they voted, including "Budget", "Energy", and "Security". Only bills on which every senator has voting records are considered in these datasets. For each bill topic (a multilayer graph) we create a layer for each bill. In each layer, there is an edge between two senators if they vote the same way. We use the party (Democratic or Republican) as the ground-truth cluster label. In addition, we label the

one independent senator as Democratic since he caucused with the Democrats.

4) *Reality mining* [53]: The reality mining dataset contains mobile and social traces among 94 MIT students. We extract the largest connected component of students from this dataset to form a 2-layer graph, where one layer represents user connection via text messaging, and the other layer represents user connection via proximity (Bluetooth). For each layer we only retain edges for which there is mutual contact between student pairs.

5) *London transportation network* [54]: The London transportation network dataset contains different transportation routes through Tube stations in London. We extract the largest connected component of stations that are either connected by Overground transportation or by Docklands Light Railway (DLR) to form a 2-layer graph, where one layer represents overground connectivity, and the other layer represents DLR connectivity.

6) *Human H1V1 genetic interaction [55]*: The human H1V1 genetic interaction dataset contains different types of genetic interactions among 1005 proteins. We extract the largest connected genetic interaction network from this dataset to form a 5-layer graph, where each interaction type corresponds to one layer and for each layer we only retain the edge of mutual interaction.

7) *Pierre Auger coauthorship [31]*: The Pierre Auger coauthorship dataset contains the coauthorship among 514 researchers between 2010 and 2012 associated with the Pierre Auger Observatory, which involves 16 working research tasks (layers) related to studies of ultra-high energy cosmic rays. We extract the largest connected component from this network to form a 16-layer graph.

Since MIMOSA allows the input multilayer graph to be weighted, for each layer G_ℓ, if G_ℓ is unweighted, we adopt the degree normalization [40] such that the (u, v)-th entry in the weight matrix $\mathbf{W}^{(\ell)}$ is $[\mathbf{W}^{(\ell)}]_{uv} = \frac{[\mathbf{A}^{(\ell)}]_{uv}}{\sqrt{d_u^{(\ell)} \cdot d_v^{(\ell)}}}$ if $d_u^{(\ell)}, d_v^{(\ell)} > 0$, and $[\mathbf{W}^{(\ell)}]_{uv} = 0$ otherwise, where $\mathbf{A}^{(\ell)}$ is the adjacency matrix of G_ℓ and $d_u^{(\ell)}$ is the degree of node u in G_ℓ.

B. Performance Evaluation

Using the multilayer graph datasets described in Table I, we compare the clustering performance of MIMOSA with four other methods. The first method is the baseline approach that assigns uniform weight to each layer in the convex layer aggregation (i.e., $w_\ell = \frac{1}{L} \forall \ell$). Since this baseline approach is equivalent to MIMOSA with the setting $\mathbf{w}^{\mathrm{ini}} = \frac{1_L}{L}$ and $\mathcal{T} = \{0\}$, we call this method *MIMOSA-uniform*. The second method is a greedy multilayer modularity maximization approach that extends the Louvain method for clustering in single-layer graphs to multilayer graphs, which is called *GenLouvain*.[3] GenLouvain aims to merge the nodes to maximize the multilayer modularity defined in [12] in a greedy manner. The third method is the multilayer graph clustering algorithm proposed in [13], called SC-ML. The fourth method is the Self-Tuning algorithm [36]

for graph clustering in single-layer graphs, where the single-layer graph is obtained by summing the edge weights across all layers.

For GenLouvain, we set the resolution parameter $\gamma \in \{0.5, 1, 2\}$ and the latent inter-layer coupling parameter $\omega = 1$. For MIMOSA, we set $\mathbf{w}^{\mathrm{ini}} = \frac{1_L}{L}$ to be a uniform vector, $\eta = 10^{-5}$, $\alpha_\ell = \alpha'_\ell = 0.05 \ \forall \ell$, and the regularization set $\mathcal{T} = \{0, 10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5\}$. The effect of the parameters in MIMOSA on the output clusters are summarized as follows. If one has some prior knowledge of the noise level in each layer, then adjusting $\mathbf{w}^{\mathrm{ini}}$ by assigning more weights to less noisy layers may yield better clustering results. Increasing η or decreasing $\{\alpha_\ell\}$ and $\{\alpha'_\ell\}$ tightens the clustering reliability constraint and may increase the number of output clusters. Expanding $\mathcal{T}$ may yield better clustering results. Like MIMOSA, GenLouvain and Self-Tuning are automated clustering algorithms that do not require specifying the number of clusters K *a priori*. SC-ML requires the knowledge of K, and for performance comparison we set the value of K in SC-ML to be the number of clusters found by MIMOSA.

We use the following external and internal clustering metrics to evaluate the performance of different methods. External metrics can be computed only when ground-truth cluster labels are known, whereas internal metrics can be computed in the absence of ground-truth cluster labels. In particular, since these internal metrics are designed for single-layer graphs, in the evaluation we extend these internal metrics to multilayer graphs by summing the metrics defined at each layer. The clustering metrics are summarized as follows. Specifically, we denote the K clusters identified by a graph clustering algorithm by $\{\mathcal{C}_k\}_{k=1}^K$, and denote the K' ground-truth clusters by $\{\mathcal{C}'_k\}_{k=1}^{K'}$.

1) *External clustering metrics:*

a) normalized mutual information (NMI) [56]: NMI is defined as

$$\mathrm{NMI}(\{\mathcal{C}_k\}_{k=1}^K, \{\mathcal{C}'_k\}_{k=1}^{K'}) = \frac{2 \cdot I(\{\mathcal{C}_k\}, \{\mathcal{C}'_k\})}{|H(\{\mathcal{C}_k\}) + H(\{\mathcal{C}'_k\})|}, \quad (13)$$

where I is the mutual information between $\{\mathcal{C}_k\}_{k=1}^K$ and $\{\mathcal{C}'_k\}_{k=1}^{K'}$, and H is the entropy of clusters. Larger NMI means better clustering performance.

b) Rand index (RI) [57]: RI is defined as

$$\mathrm{RI}(\{\mathcal{C}_k\}_{k=1}^K, \{\mathcal{C}'_k\}_{k=1}^{K'}) = \frac{TP + TN}{TP + TN + FP + FN}, \quad (14)$$

where TP, TN, FP and FN represent true positive, true negative, false positive, and false negative decisions, respectively. Larger RI means better clustering performance.

c) F-measure [58]: F-measure is the harmonic mean of the precision and recall values for each cluster, which is defined as

$$\mathrm{F-measure}(\{\mathcal{C}_k\}_{k=1}^K, \{\mathcal{C}'_k\}_{k=1}^{K'}) = \frac{1}{K} \sum_{k=1}^K$$

$$\mathrm{F-measure}_k, \quad (15)$$

---

[3]http://netwiki.amath.unc.edu/GenLouvain/GenLouvain

TABLE II
SUMMARY OF THE NUMBER OF IDENTIFIED CLUSTERS (K) AND THE EXTERNAL AND INTERNAL CLUSTERING METRICS

Dataset	Method	K	NMI	RI	F-measure	conductance	NC
VC 7th grader social network	MIMOSA	2	**0.8123**	**0.9310**	**0.9317**	**0.2649**	**0.4330**
	MIMOSA-uniform	NA	NA	NA	NA	NA	NA
	GenLouvain ($\gamma = 0.5$)	3	0.6495	0.7833	0.7333	0.4487	0.6051
	GenLouvain ($\gamma = 1$)	3	0.6495	0.7833	0.7333	0.4487	0.6051
	GenLouvain ($\gamma = 2$)	8	0.4418	0.5911	0.3197	1.4295	1.6081
	SC-ML	2	0.6119	0.8079	0.8040	0.2756	0.4618
	Self-Tuning	6	0.5345	0.6995	0.5764	0.4329	0.5510
Leskovec-Ng collaboration network	MIMOSA	2	**1**	**1**	**1**	**0.0213**	**0.0415**
	MIMOSA-uniform	NA	NA	NA	NA	NA	NA
	GenLouvain ($\gamma = 0.5$)	7	0.6824	0.8488	0.8243	0.1989	0.2663
	GenLouvain ($\gamma = 1$)	16	0.4972	0.7156	0.6055	0.3054	0.3702
	GenLouvain ($\gamma = 2$)	29	0.3553	0.5586	0.2173	0.4874	0.5569
	SC-ML	2	**1**	**1**	**1**	**0.0213**	**0.0415**
	Self-Tuning	2	**1**	**1**	**1**	**0.0213**	**0.0415**
109th Congress votes - Budget	MIMOSA	2	0.7959	0.9224	0.9220	0.2713	0.4975
	MIMOSA-uniform	2	**0.8778**	**0.9604**	**0.9603**	0.2702	0.5055
	GenLouvain ($\gamma = 0.5$)	2	0.7959	0.9224	0.9220	0.2713	0.4978
	GenLouvain ($\gamma = 1$)	2	0.7959	0.9224	0.9220	0.2713	0.4978
	GenLouvain ($\gamma = 2$)	55	0.3822	0.6915	0.5539	**0.1500**	**0.1959**
	SC-ML	2	0.7610	0.9040	0.9036	0.2742	0.5089
	Self-Tuning	3	0.8488	0.9164	0.9087	1.5046	1.8011
109th Congress votes - Energy	MIMOSA	2	**0.7290**	**0.8861**	**0.8855**	**0.1151**	**0.2086**
	MIMOSA-uniform	2	0.6716	0.8508	0.8508	0.1154	0.2178
	GenLouvain ($\gamma = 0.5$)	2	0.5403	0.8182	0.8173	**0.1151**	**0.2086**
	GenLouvain ($\gamma = 1$)	2	0.5403	0.8182	0.8173	**0.1151**	**0.2086**
	GenLouvain ($\gamma = 2$)	7	0.6371	0.8521	0.8422	0.3145	0.3593
	SC-ML	2	0.6716	0.8513	0.8508	0.1154	0.2178
	Self-Tuning	4	0.6310	0.8521	0.8424	1.0204	1.0970
109th Congress votes - Security	MIMOSA	2	0.6105	0.8513	0.8506	0.0400	0.0785
	MIMOSA-uniform	2	0.6304	0.8513	0.8506	0.0400	0.0785
	GenLouvain ($\gamma = 0.5$)	2	0.5816	0.8345	0.8337	0.0400	0.0770
	GenLouvain ($\gamma = 1$)	2	**0.6598**	**0.8685**	**0.8678**	0.0400	0.0770
	GenLouvain ($\gamma = 2$)	4	0.6181	0.8515	0.8477	**0.0204**	**0.0492**
	SC-ML	2	0.6304	0.8513	0.8506	0.0400	0.0785
	Self-Tuning	2	0.6304	0.8513	0.8506	0.0400	0.0785
Reality mining	MIMOSA	2	-	-	-	**0.0819**	**0.1573**
	MIMOSA-uniform	2	-	-	-	**0.0819**	**0.1573**
	GenLouvain ($\gamma = 0.5$)	3	-	-	-	0.2239	0.3165
	GenLouvain ($\gamma = 1$)	3	-	-	-	0.2239	0.3165
	GenLouvain ($\gamma = 2$)	6	-	-	-	0.1240	0.2011
	SC-ML	2	-	-	-	**0.0819**	**0.1573**
	Self-Tuning	4	-	-	-	0.4267	0.5247
London transportation network	MIMOSA	5	-	-	-	0.0553	0.0801
	MIMOSA-uniform	5	-	-	-	0.0553	0.0801
	GenLouvain ($\gamma = 0.5$)	9	-	-	-	0.1046	0.1286
	GenLouvain ($\gamma = 1$)	14	-	-	-	0.1558	0.1763
	GenLouvain ($\gamma = 2$)	21	-	-	-	0.2001	0.2181
	SC-ML	5	-	-	-	0.1044	0.1425
	Self-Tuning	26	-	-	-	**0.0154**	**0.0798**
Human H1V1 genetic interaction	MIMOSA	2	-	-	-	**0.0346**	**0.0666**
	MIMOSA-uniform	2	-	-	-	**0.0346**	**0.0666**
	GenLouvain ($\gamma = 0.5$)	4	-	-	-	0.1822	0.2292
	GenLouvain ($\gamma = 1$)	4	-	-	-	0.1822	0.2292
	GenLouvain ($\gamma = 2$)	5	-	-	-	0.1458	0.3167
	SC-ML	2	-	-	-	0.1161	0.2027
	Self-Tuning	7	-	-	-	0.5627	0.8722
Pierre Auger coauthorship	MIMOSA	2	-	-	-	**0.0113**	**0.1888**
	MIMOSA-uniform	NA	-	-	-	NA	NA
	GenLouvain ($\gamma = 0.5$)	9	-	-	-	1.5207	1.8423
	GenLouvain ($\gamma = 1$)	13	-	-	-	1.2655	1.4699
	GenLouvain ($\gamma = 2$)	61	-	-	-	0.5717	0.6356
	SC-ML	2	-	-	-	1.2939	2.5181
	Self-Tuning	63	-	-	-	0.8400	0.9321

"NA" means "not applicable", and "−" means "not available" due to lack of ground-truth cluster labels. For each dataset, the method that leads to the highest clustering metric is highlighted in bold face.
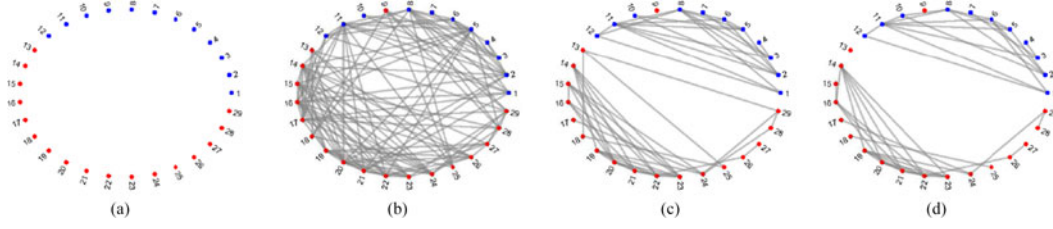
Fig. 5. Illustration of the ground-truth clusters and the clusters found by MIMOSA for the VC 7th grader social network dataset. Fig. 5(a) displays the ground-truth clusters, where nodes 1 to 12 are boys (labeled by blue color) and nodes 13 to 29 are girls (labeled by red color). (b) to (d) The clusters (labeled by different colors) found by MIMOSA in each layer. Comparing to the ground-truth clusters, MIMOSA correctly group all nodes into 2 clusters except for node 9, since node 9 has no edge connections in (c) and (d), and has more connections to girls than boys in (b). Enlarged plots are displayed in the supplementary material. (a) Ground-truth clusters (b) Friends you get on with (c) Your best friends (d) Friends you work with.

where $\text{F} - \text{measure}_k = \frac{2 \cdot PREC_k \cdot RECALL_k}{PREC_k + RECALL_k}$, and $PREC_k$ and $RECALL_k$ are the precision and recall values for cluster $\mathcal{C}_k$. Larger F-measure means better clustering performance.

2) *Internal clustering metrics:*

a) conductance [59]: conductance is defined as

$$\text{conductance}(\{\mathcal{C}_k\}_{k=1}^K) = \frac{1}{K} \sum_{k=1}^K \text{conductance}_k, \tag{16}$$

where $\text{conductance}_k = \frac{W_k^{out}}{2 \cdot W_k^{in} + W_k^{out}}$, and W_k^{in} and W_k^{out} are the sum of within-cluster and between-cluster edge weights of cluster $\mathcal{C}_k$, respectively. Lower conductance means better clustering performance.

b) normalized cut (NC) [59]: NC is defined as

$$\text{NC}(\{\mathcal{C}_k\}_{k=1}^K) = \frac{1}{K} \sum_{k=1}^K \text{NC}_k, \tag{17}$$

where $\text{NC}_k = \frac{W_k^{out}}{2 \cdot W_k^{in} + W_k^{out}} + \frac{W_k^{out}}{2 \cdot (W_k^{all} - W_k^{in}) + W_k^{out}}$, and W_k^{in}, W_k^{out} and W_k^{all} are the sum of within-cluster, between-cluster and total edge weights of cluster $\mathcal{C}_k$, respectively. Lower NC means better clustering performance.

Table II summarizes the external and internal clustering metrics obtained after multilayer graph clustering by the four methods for the datasets listed in Table I. For MIMOSA and MIMOSA-uniform, we terminate the iterative process and report the clustering result as "not applicable" (NA) when the number of clusters K exceeds $\frac{n}{2}$, where n is the number of nodes. As a result, NA means that before termination no clustering results have passed the clustering reliability tests.

It is observed from Table II that MIMOSA has the best clustering performance among 6 out of 9 datasets. For the Congress-votes-Budget and Congress-votes-Security datasets, MIMOSA performs somewhat worse than MIMOSA-uniform. For the VC 7th grader social network, Leskovec-Ng collaboration network and Pierre Auger coauthorship datasets, MIMOSA-uniform fails to find a reliable clustering result, whereas MIMOSA has superior clustering metrics. The robustness of MIMOSA implies the utility of layer weight adaptation, and it also suggests that assigning uniform weight to every layer regardless of the noise

level may lead to unreliable clustering results. Comparing MIMOSA to SC-ML with the same number of clusters, the clusters found by MIMOSA have better clustering metrics. MIMOSA also outperforms Self-Tuning in most of the datasets, suggesting that simply summing a multilayer graph to create a single-layer graph does not necessarily benefit multilayer graph clustering. In addition, we also observe that GenLouvain tends to identify more clusters than the number of ground-truth clusters. The fact that MIMOSA-uniform and Self-Tuning outperform MIMOSA in some cases is likely due to the fact that these particular datasets have similar connectivity in each layer. For example, in the Congres-votes-Budget dataset almost every senator voted along party lines on all budget related legislation.

As a visual illustration, Fig. 5 displays the ground-truth clusters and the clusters identified by MIMOSA for each layer of the VC 7th grader social network dataset. The number of clusters identified by MIMOSA is 2, which is consistent with the ground truth. The optimal layer weight vector obtained from step 5 of MIMOSA in Algorithm 1 is $\mathbf{w}^* = [0.0531\,0.1608\,0.7861]^T$. Comparing each layer with the ground-truth clusters, it can be observed that the connectivity patterns in Fig. 5(c) and (d) are more consistent with the ground truth, whereas the connectivity pattern in Fig. 5(b) is less informative, which explains why MIMOSA adapts more weights to the second and the third layers. Furthermore, Fig. 5 also explains why MIMOSA-uniform does not yield reliable clustering results, since it assigns uniform weight to each layer and is insensitive to the noise distribution. It is worth noting that MIMOSA correctly groups all nodes into 2 clusters except for node 9. However, we also observe that node 9 has no edge connections in the two informative layers as shown in Fig. 5(c) and (d), and indeed has more connections to girls than boys in the first layer as shown in Fig. 5(b), which leads to the misclassification of node 9 when compared with the ground-truth clusters.

VIII. CONCLUSION

We have characterized the phase transition that governs the accuracy of a convex aggregation method of multilayer spectral graph clustering (SGC). By varying the noise level, as measured by the edge connection probability of spurious between-cluster edges, we specified the critical value that separates the performance of multilayer SGC into a reliable regime and an unreliable regime. The phase transition was validated via numerical experiments. Furthermore, based on the phase transition

analysis, we proposed MIMOSA, a multilayer SGC algorithm that provides automated model order selection for cluster assignment and layer weight adaptation with statistical clustering reliability guarantees. Applying MIMOSA to real-world multilayer graphs shows competitive or better clustering performance with respect to several baseline methods, including the uniform weight assignment, a greedy multilayer modularity maximization method, and a subspace approach. Our future work will include extending the phase transition analysis and MIMOSA to other multilayer block models.

ACKNOWLEDGMENT

REFERENCES

[1] B. Oselio, A. Kulesza, and A. O. Hero, "Multi-layer graph analysis for dynamic social networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 4, pp. 514–523, Aug. 2014.

[2] K. S. Xu and A. O. Hero, "Dynamic stochastic blockmodels for time-evolving social networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 4, pp. 552–562, Aug. 2014.

[3] M. De Domenico et al., "Mathematical formulation of multilayer networks," *Phys. Rev. X*, vol. 3, Dec. 2013, Art. no. 041022.

[4] B. Oselio, A. Kulesza, and A. Hero, "Information extraction from large multi-layer social networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2015, pp. 5451–5455.

[5] D. Zhou and C. J. Burges, "Spectral clustering and transductive learning with multiple views," in *Proc. Int. Conf. Mach. Learn.*, 2007, pp. 1159–1166.

[6] N. Leonardi and D. Van De Ville, "Tight wavelet frames on multislice graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3357–3367, Jul. 2013.

[7] K. Benzi, B. Ricaud, and P. Vandergheynst, "Principal patterns on graphs: Discovering coherent structures in datasets," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 2, pp. 160–173, Jun. 2016.

[8] P.-Y. Chen, S. Choudhury, and A. O. Hero, "Multi-centrality graph spectral decompositions and their application to cyber intrusion detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016, pp. 4553–4557.

[9] Y. Park, C. E. Priebe, and A. Youssef, "Anomaly detection in time series of graphs using fusion of graph invariants," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 67–75, Feb. 2013.

[10] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *J. Complex Netw.*, vol. 2, no. 3, pp. 203–271, 2014.

[11] J. Kim and J.-G. Lee, "Community detection in multi-layer graphs: A survey," *ACM SIGMOD Rec.*, vol. 44, no. 3, pp. 37–48, 2015.

[12] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *Science*, vol. 328, no. 5980, pp. 876–878, 2010.

[13] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov, "Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 905–918, Feb. 2014.

[14] D. Cai, Z. Shao, X. He, X. Yan, and J. Han, "Community mining from multi-relational networks," in *European Conference on Principles of Data Mining and Knowledge Discovery*. New York, NY, USA: Springer, 2005, pp. 445–452.

[15] L. Tang, X. Wang, and H. Liu, "Uncoverning groups via heterogeneous interaction analysis," in *Proc. IEEE Int. Conf. Data Mining*, 2009, pp. 503–512.

[16] Z. Wu, Z. Bu, J. Cao, and Y. Zhuang, "Discovering communities in multi-relational networks," in *User Community Discovery*. New York, NY, USA: Springer, 2015, pp. 75–95.

[17] L. Tang, X. Wang, and H. Liu, "Community detection via heterogeneous interaction analysis," *Data Mining Knowl. Discovery*, vol. 25, no. 1, pp. 1–33, 2012.

[18] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, "Structural reducibility of multilayer networks," *Nature Commun.*, vol. 6, 2015, Art. no. 6864.

[19] D. Taylor, S. Shai, N. Stanley, and P. J. Mucha, "Enhanced detectability of community structure in multilayer networks through layer aggregation," *Phys. Rev. Lett.*, vol. 116, Jun. 2016, Art. no. 228301.

[20] J. Kim, J.-g. Lee, and S. Lim, "Differential flattening: A novel framework for community detection in multi-layer graphs," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 2, 2016, Art. no. 27.

[21] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Netw.*, vol. 5, no. 2, pp. 109–137, 1983.

[22] Q. Han, K. Xu, and E. Airoldi, "Consistent estimation of dynamic and multi-layer block models," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1511–1520.

[23] S. Paul and Y. Chen, "Community detection in multi-relational data with restricted multi-layer stochastic blockmodel," arXiv:1506.02699, 2015. [Online]. Available: https://projecteuclid.org/euclid.ejs/1480993455

[24] P. Barbillon, S. Donnet, E. Lazega, and A. Bar-Hen, "Stochastic block models for multiplex networks: An application to a multilevel network of researchers," *J. Roy. Statist. Soc., Ser. A*, vol. 180, no. 1, pp. 295–314, 2016.

[25] T. Vallès-Català, F. A. Massucci, R. Guimerà, and M. Sales-Pardo, "Multilayer stochastic block models reveal the multilayer structure of complex networks," *Phys. Rev. X*, vol. 6, Mar. 2016, Art. no. 011036.

[26] N. Stanley, S. Shai, D. Taylor, and P. J. Mucha, "Clustering network layers with the strata multilayer stochastic block model," *IEEE Trans. Netw. Sci. Eng.*, vol. 3, no. 2, pp. 95–105, Apr. 2016.

[27] E. E. Papalexakis, L. Akoglu, and D. Ience, "Do more views of a graph help? community detection and clustering in multi-graphs," in *Proc. Int. Conf. Inf. Fusion*, 2013, pp. 899–905.

[28] J. Iacovacci, Z. Wu, and G. Bianconi, "Mesoscopic structures reveal the network between the layers of multiplex data sets," *Phys. Rev. E*, vol. 92, Oct. 2015, Art. no. 042806.

[29] D. Greene and P. Cunningham, "Producing a unified graph representation from multiple social network views," in *Proc. ACM Web Sci. Conf.*, 2013, pp. 118–121.

[30] J. Ni, H. Tong, W. Fan, and X. Zhang, "Flexible and robust multi-network clustering," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 835–844.

[31] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, "Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems," *Phys. Rev. X*, vol. 5, Mar. 2015, Art. no. 011027.

[32] W. Tang, Z. Lu, and I. S. Dhillon, "Clustering with multiple graphs," in *Proc. IEEE Int. Conf. Data Mining*, 2009, pp. 1016–1021.

[33] Z. Kuncheva and G. Montana, "Community detection in multiplex networks using locally adaptive random walks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, 2015, pp. 1308–1315.

[34] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov, "Clustering with multi-layer graphs: A spectral perspective," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5820–5831, Nov. 2012.

[35] B. Boden, S. Günnemann, H. Hoffmann, and T. Seidl, "Mining coherent subgraphs in multi-layer graphs with edge labels," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 1258–1266.

[36] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 1601–1608.

[37] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Statist. Mech., Theory Exp.*, no. 10, p. P10008, 2008. [Online]. Available: http://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008/meta

[38] F. Krzakala et al., "Spectral redemption in clustering sparse networks," *Proc. Nat. Acad. Sci. USA*, vol. 110, pp. 20 935–20 940, 2013.

[39] P.-Y. Chen and A. O. Hero, "Phase transitions and a model order selection criterion for spectral graph clustering," arXiv:1604.03159, 2016.

[40] U. Luxburg, "A tutorial on spectral clustering," *Statist.Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[41] J. A. Hartigan and M. A. Wong, "A k-means clusterin algorithm," *Appl. Statist.*, vol. 28, pp. 100–108, 1979.

[42] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Math. J.*, vol. 23, no. 98, pp. 298–305, 1973.

[43] A. Jennings and J. J. McKeown, *Matrix computation*. New York, NY, USA: Wiley, 1992.

[44] J. A. Tropp, "An introduction to matrix concentration inequalities," *Foundations Trends Mach. Learn.*, vol. 8, no. 1–2, pp. 1–230, 2015. [Online]. Available: http://dx.doi.org/10.1561/2200000048

[45] R. F. Potthoff and M. Whittinghill, "Testing for homogeneity: I. The binomial and multinomial distributions," *Biometrika*, vol. 53, no. 1/2, pp. 167–182, 1966.

[46] P.-Y. Chen, B. Zhang, M. A. Hasan, and A. O. Hero, "Incremental method for spectral clustering of increasing orders," in *Proc. ACM Int. Conf. Knowl. Discovery Data Mining/Workshop Mining Learn. Graphs*, arXiv:1512.07349, 2016. [Online]. Available: http://www.mlgworkshop.org/2016/

[47] L. Wu, E. Romero, and A. Stathopoulos, "Primme_svds: A high-performance preconditioned SVD solver for accurate large-scale computations," *SIAM J. Sci. Comput.*, arXiv:1607.01404, 2016.

[48] M. J. Zaki and W. M. Jr, *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge UK: Cambridge Univ. Press, 2014.

[49] M. Vickers and S. Chan, "Representing classroom social structure," Victoria Inst. Secondary Educ., Melbourn, Vic, Australia, 1981. [Online]. Available: http://deim.urv.cat/ manlio.dedomenico/data.php

[50] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Arnetminer: Extraction and mining of academic social networks," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2008, pp. 990–998.

[51] B. Zhang, T. K. Saha, and M. Al Hasan, "Name disambiguation from link data in a collaboration graph," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, 2014, pp. 81–84.

[52] T. K. Saha, B. Zhang, and M. Al Hasan, "Name disambiguation from link data in a collaboration graph using temporal and topological features," *Social Netw. Anal. Mining*, vol. 5, pp. 1–14, 2015.

[53] A. Pentland, N. Eagle, and D. Lazer, "Inferring social network structure using mobile phone data," *Proc. Nat Acad. Sci. USA*, vol. 106, no. 36, pp. 15 274–15 278, 2009. [Online]. Available: http://realitycommons.media.mit.edu

[54] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, "Navigability of interconnected networks under random failures," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 23, pp. 8351–8356, 2014. [Online]. Available: http://deim.urv.cat/ manlio.dedomenico/data.php

[55] M. De Domenico, M. A. Porter, and A. Arenas, "Muxviz: A tool for multilayer analysis and visualization of networks," *J. Complex Netw.*, vol. 3, pp. 159–176, 2015.

[56] A. Strehl and J. Ghosh, "Cluster ensembles-a knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, 2002.

[57] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Statist. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.

[58] C. J. V. Rijsbergen, *Information Retrieval*, 2nd ed. Newton, MA, USA: Butterworth-Heinemann, 1979.

[59] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[60] W. M. K. Chandler Davis, "The rotation of eigenvectors by a perturbation. III," *SIAM J. Numer. Anal.*, vol. 7, no. 1, pp. 1–46, 1970.

[61] F. J. Anscombe, "The transformation of poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.

[62] Y.-P. Chang and W.-T. Huang, "Generalized confidence intervals for the largest value of some functions of parameters under normality," *Statist. Sin.*, pp. 1369–1383, 2000.

Pin-Yu Chen (S'10–M'16) received the B.S. degree in electrical engineering and computer science (undergraduate honors program) from the National Chiao Tung University, Hsinchu, Taiwan, in 2009, the M.S. degree in communication engineering from the National Taiwan University, Taipei, Taiwan, in 2011, and the Ph.D. degree in electrical engineering and computer science and M.A. degree in statistics from the University of Michigan, Ann Arbor, MI, USA, in 2016.

He is currently a Research Scientist in the AI Foundations Group, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA. His research interest includes graph data analytics and their applications to data mining, machine learning, artificial intelligence, and cyber security. He is a member of the Tau Beta Pi Honor Society and the Phi Kappa Phi Honor Society, and was the recipient of the Chia-Lun Lo Fellowship from the University of Michigan. He received the IEEE GLOBECOM 2010 GOLD Best Paper Award and several travel grants, including IEEE ICASSP 2014 (NSF), IEEE ICASSP 2015 (SPS), IEEE Security and Privacy Symposium 2016, Graph Signal Processing Workshop 2016, and ACM KDD 2016.

Alfred O. Hero, III (S'79–M'84–SM'96–F'98) received the B.S. degree (*summa cum laude*) from Boston University, Boston, MA, USA, in 1980, and the Ph.D. degree from Princeton University, Princeton, NJ, USA, in 1984, both in electrical engineering.

Since 1984, he has been with the University of Michigan, Ann Arbor, MI, USA, where he is the John H. Holland Distinguished University Professor of electrical engineering and computer science and the R. Jamison and Betty Williams Professor of Engineering. He is also the Co-Director of the University's Michigan Institute for Data Science. At the University of Michigan, his primary appointment is in the Department of Electrical Engineering and Computer Science and he also has appointments, by courtesy, in the Department of Biomedical Engineering and the Department of Statistics. From 2008 to 2013, he held the Digiteo Chaire d'Excellence, sponsored by Digiteo Research Park in Paris, located at the Ecole Superieure d'Electricite, Gif-sur-Yvette, France. He has held other visiting positions at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology (2006), Boston University (2006), I3S University of Nice, Sophia-Antipolis, France (2001), Ecole Normale Supérieure de Lyon (1999), Ecole Nationale Supérieure des Télécommunications, Paris (1999), Lucent Bell Laboratories (1999), Scientific Research Labs of the Ford Motor Company, Dearborn, MI, USA (1993), Ecole Nationale Superieure des Techniques Avancees, Ecole Superieure d'Electricite, Paris, France (1990), and M.I.T. Lincoln Laboratory (1987–1989). His recent research interests include high-dimensional spatio-temporal data, multimodal data integration, statistical signal processing, and machine learning. Of particular interest are applications to social networks, network security and forensics, computer vision, and personalized health.

Dr. O. Hero, III has served on the editorial boards of the IEEE TRANSACTIONS ON INFORMATION THEORY (1995–1998, 1999), the IEEE TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS (2004–2006), and the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2002, 2004). He was the Chairman of the Statistical Signal and Array Processing Technical Committee (1997–1998) and the Treasurer of the Conference Board of the IEEE Signal Processing Society. He was the Chairman for Publicity of the 1986 IEEE International Symposium on Information Theory, Ann Arbor, MI, USA, and the General Chairman of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing, Detroit, MI, USA. He was the Co-Chair of the 1999 IEEE Information Theory Workshop on Detection, Estimation, Classification and Filtering, Santa Fe, NM, USA, and the 1999 IEEE Workshop on Higher Order Statistics, Caesaria, Israel. He was the General Chair of the 2012 IEEE Workshop on Statistical Signal Processing, Ann Arbor, MI, USA. He co-chaired the 2002 Workshop on Genomic Signal Processing and Statistics. He was the Vice President (Finance) of the IEEE Signal Processing Society (1999–2002), and the Chair of Commission C (Signals and Systems) of the US National Commission of the International Union of Radio Science (URSI) (1999–2002). He was a member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society (1999–2004). He was the President of the IEEE Signal Processing Society (2006–2007) and during his term he served on the TAB Periodicals Committee (2006) and the IEEE Technical Activities Board (TAB). He was a member of the IEEE TAB Society Review Committee (2009), the IEEE Awards Committee (2010–2011), and served on the Board of Directors of the IEEE (2009–2011) as the Director of Division IX (Signals and Applications). He served on the IEEE TAB Nominations and Appointments Committee (2012–2014). He is currently a member of the Big Data Special Interest Group of the IEEE Signal Processing Society. Since 2011, he has been a member of the Committee on Applied and Theoretical Statistics of the US National Academies of Science. He is a member of Tau Beta Pi, the American Statistical Association, the Society for Industrial and Applied Mathematics, and the US National Commission (Commission C) of the International URSI. He received the University of Michigan Distinguished Faculty Achievement Award in 2011. He has been a plenary and keynote speaker at several workshops and conferences. He has received several best paper awards including the IEEE Signal Processing Society Best Paper Award (1998), the Best Original Paper Award from the *Journal of Flow Cytometry* (2008), the Best Magazine Paper Award from the IEEE Signal Processing Society (2010), the SPIE Best Student Paper Award (2011), the IEEE ICASSP Best Student Paper Award (2011), the AISTATS Notable Paper Award (2013), and the IEEE ICIP Best Paper Award (2013). He received the IEEE Signal Processing Society Meritorious Service Award (1998), the IEEE Third Millenium Medal (2000), the IEEE Signal Processing Society Distinguished Lecturership (2002), and the IEEE Signal Processing Society Technical Achievement Award (2014). In 2015, he received the Society Award, which is the highest career award bestowed by the IEEE Signal Processing Society.