Consistency between ordering and clustering methods for graphs

Kawamoto, Tatsuro

Ochi, Masaki

Kobayashi, Teruyoshi

Consistency between ordering and clustering methods for graphs

Tatsuro Kawamoto ⬤

Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, 135-0064 Tokyo, Japan

Masaki Ochi ⬤

Department of Physics, The University of Tokyo, 277-8574 Chiba, Japan

Teruyoshi Kobayashi ⬤

Department of Economics, Kobe University, 657-8501 Kobe, Japan

A relational dataset is often analyzed by optimally assigning a label to each element through clustering or ordering. While similar characterizations of a dataset would be achieved by both clustering and ordering methods, the former has been studied much more actively than the latter, particularly for the data represented as graphs. This study fills this gap by investigating methodological relationships between several clustering and ordering methods, focusing on spectral techniques. Furthermore, we evaluate the resulting performance of the clustering and ordering methods. To this end, we propose a measure called the *label continuity error*, which generically quantifies the degree of consistency between a sequence and partition for a set of elements. Based on synthetic and real-world datasets, we evaluate the extents to which an ordering method identifies a module structure and a clustering method identifies a banded structure.

I. INTRODUCTION

Identifying macroscopic connection patterns in graphs is a major challenge in network science. A number of algorithms have been proposed to extract different features, such as community structure [1–3], hierarchical community structure [4,5], core-periphery structure [6], nested structure [7], and banded structure [8–10], to name a few.

When a graph consists of subgraphs in each of which vertices are densely connected, the graph structure is referred to as a community structure. A common approach for extracting a community structure is the partitioning of graphs, termed community detection [2,3] or graph clustering [1]. In this approach, an algorithm assigns a group label to each vertex such that vertices with the same group label are densely connected. Alternatively, we may also identify densely connected vertices through an *ordering* method that infers the optimal ordering of vertices such that vertices close to each other in the sequence are densely connected. The corresponding optimization problems are collectively termed the minimum linear arrangement [11–13] or envelope reduction [14,15], and the inferred structural property is called a banded structure or sequentially local structure [10]. As exemplified in Fig. 1, the densely connected vertices are clearly detected by appropriately visualizing the graph and the adjacency matrix based on an appropriate vertex ordering.

Despite the similarity between these two approaches, the clustering problem has received considerable attention in the literature. Figure 2 shows the number of articles with keywords that represent ordering (pink bars) or clustering (blue bars) problems. Most of the keywords for ordering problems represent more general matrix ordering problems rather than vertex ordering problems for graphs (i.e., adjacency matrices), whereas the keywords for clustering problems mostly capture problems for graphs. Clustering methods have been studied and applied much more actively than ordering methods.

Spectral methods are popular in both ordering and clustering problems; the former and the latter are respectively termed spectral ordering [8,14,15] and spectral clustering [17–19]. In both methods, the leading eigenvector(s) of a Laplacian or its variant is used to identify the optimal ordering or clustering of vertices. Specifically, when a graph is partitioned into two groups based on the sorting of the eigenvector elements [17], the result of spectral clustering is generally consistent with the vertex sequence inferred by spectral ordering.

However, spectral ordering and clustering algorithms are not generally consistent. For instance, when graphs are partitioned into more than two groups, it is common to employ the K-means algorithm [20] on $K(>2)$ leading eigenvectors to achieve a K-way partitioning [19]. By contrast, to identify the optimal vertex sequence using the spectral ordering method, we always use the eigenvector associated with the second leading eigenvalue. Therefore, it is nontrivial to determine the extent to which the two methods are quantitatively consistent. Even when we partition a graph into two groups, the result of spectral clustering may not be consistent with the vertex sequence obtained by spectral ordering when the K-means algorithm is used to obtain a partition.
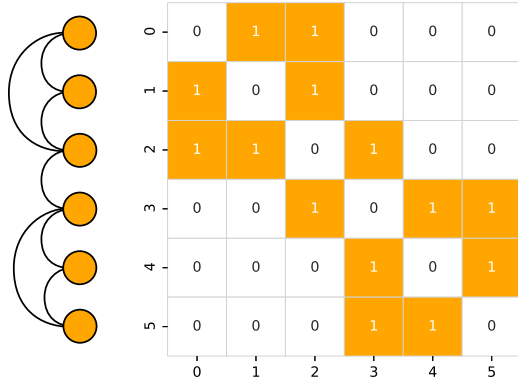
FIG. 1. Simple example of a graph (left) and its adjacency matrix (right) for identifying a community structure through the optimal ordering of vertices without specifying the group labels. The (i, j) element of the adjacency matrix is one (highlighted) when vertices i and j are connected, and zero (not highlighted) otherwise.

We conduct a systematic investigation to evaluate the consistency between the spectral ordering and clustering methods. We first introduce a generic measure, referred to as the *label continuity error* (LCE), to quantify the difference between a sequence and partition for a set of elements (e.g., vertices of graphs). Intuitively, a sequence and partition are more consistent with each other if, for a given number of groups, the group label flips less often when following the elements in the specified order. We provide a more precise definition in the next section. Although we use this measure throughout the study, it is not the only method of quantifying consistency; we will revisit this point in Sec. V.

There are also several modern spectral clustering algorithms with unexplored ordering counterparts. These include the methods based on the modularity matrix [21,22], Bethe Hessian [23,24], and regularized Laplacian [25–29]. To fill this gap, we show how spectral ordering algorithms can be derived from optimization problems using the matrices on which these modern spectral clustering methods are based. Spectral ordering problems based on these matrices are formulated as variants of the classical spectral ordering problem [14,15] with different penalty terms and/or constraints.
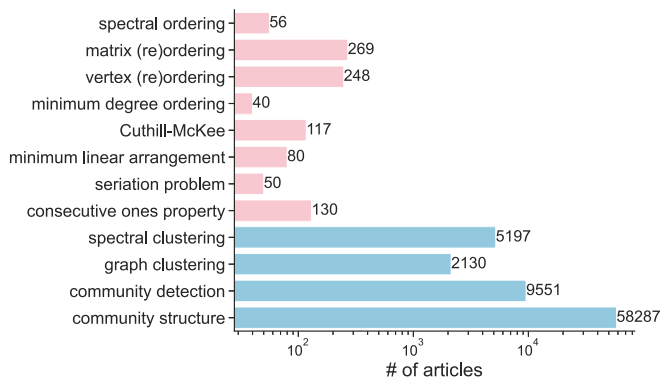

FIG. 2. Number of articles with a keyword related to the ordering (pink bar) or clustering (blue bar) problems in the title or abstract. The data was collected from Dimensions [16] on May 30, 2022.

The remainder of this paper is organized as follows. Section II formally introduces the LCE to quantitatively evaluate the consistency between ordering and clustering methods and examine its properties. Section III formulates spectral ordering methods corresponding to existing spectral clustering methods for graphs. Using the LCE introduced in Sec. II and the methods formulated in Sec. III, we analyze the consistency between spectral ordering and clustering methods using synthetic and real-world networks in Sec. IV. Finally, Sec. V discusses the results of this study.

II. LABEL CONTINUITY ERROR

Let $G(V, E)$ be a graph, where $V = \{v_1, \ldots, v_N\}$ is the vertex set and E is the edge set. We assume that every vertex in the graph is distinguishable and let $\mathcal{I} = \{1, \ldots, N\}$ be the ordered set indicating the original sequence of the vertices which corresponds to the subscripts in $\{v_1, \ldots, v_N\}$. For vertex $v_i \in V$ ($i \in \mathcal{I}$), we denote $\pi(i) = \pi_i \in \{1, \ldots, N\}$ as the index after permutation (i.e., we use π as both a mapping and a variable) and $\boldsymbol{\pi} = \{\pi(i)|i \in \mathcal{I}\}$ as the reordered sequence of the vertices. Similarly, we denote $\sigma(i) = \sigma_i \in \{1, \ldots, K\}$ as the group label of vertex v_i and $\boldsymbol{\sigma} = \{\sigma(i)|i \in \mathcal{I}\}$ as the partition of the vertex set. We also denote $V_k = \{v_i|\sigma_i = k, i \in \mathcal{I}\}$ and $N_k = |V_k|$ for group k (we let $\{N_1, \ldots, N_K\} =: \{N_k\}$). Throughout this study, $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\sigma}}$ represent the inferred sequence and partition using algorithms, respectively. We denote d_i for the degree of vertex v_i.

A. Definition

We introduce a measure to quantify the consistency between a sequence $\boldsymbol{\pi}$ and partition $\boldsymbol{\sigma}$. We define the sequence $\boldsymbol{\pi}$ as consistent with $\boldsymbol{\sigma}$ if vertices with the same group label are maximally adjacent to each other in the sequence $\boldsymbol{\pi}$. For instance, if the original indices $\mathcal{I}$ are consistent with group labels $\boldsymbol{\sigma}$,

$$\sum_{i=1}^{N-1} \delta(\sigma(i), \sigma(i+1)) \tag{1}$$

is maximized, where $\delta(a, b)$ represents the Kronecker delta; Fig. 3(a) presents an example. To evaluate the consistency between $\boldsymbol{\pi}$ and $\boldsymbol{\sigma}$, we introduce a measure that we refer to as the *label continuity*, defined by

$$\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma}) = \frac{1}{N-1} \sum_{i'=1}^{N-1} \delta(\sigma(\pi^{-1}(i')), \sigma(\pi^{-1}(i'+1))), \tag{2}$$

where π^{-1} is the inverse mapping of π and i' is the index label after the permutation; that is, $\pi^{-1}(i')$ is the label i in the original indices satisfying $\pi(i) = i'$. The number of times that the group labels are flipped when following the vertices in the order of $\boldsymbol{\pi}$ is expressed as $(N-1)(1 - \mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma}))$, for which the group labels must be flipped at least $K - 1$ times. Considering this feature, we define the *label continuity error* (LCE) as

$$\Delta(\boldsymbol{\pi}, \boldsymbol{\sigma}) = 1 - \frac{K-1}{N-1} - \mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma}). \tag{3}$$

Hereafter, we abbreviate $\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma})$ and $\Delta(\boldsymbol{\pi}, \boldsymbol{\sigma})$ as $\mathcal{C}$ and Δ, respectively, as long as there is no possibility of confusion.
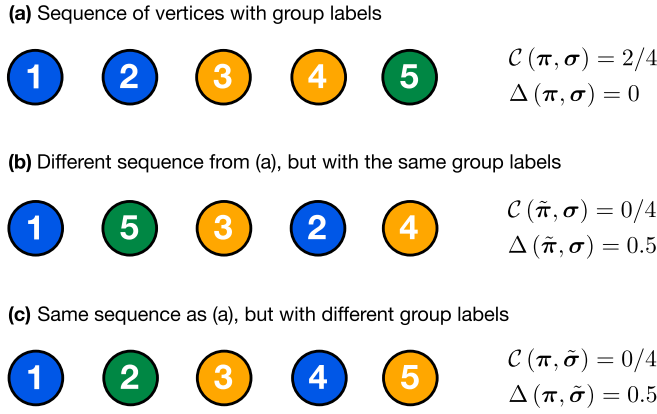
(a) Sequence of vertices with group labels

$\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma}) = 2/4$

$\Delta(\boldsymbol{\pi}, \boldsymbol{\sigma}) = 0$

(b) Different sequence from (a), but with the same group labels

$\mathcal{C}(\tilde{\boldsymbol{\pi}}, \boldsymbol{\sigma}) = 0/4$

$\Delta(\tilde{\boldsymbol{\pi}}, \boldsymbol{\sigma}) = 0.5$

(c) Same sequence as (a), but with different group labels

$\mathcal{C}(\boldsymbol{\pi}, \tilde{\boldsymbol{\sigma}}) = 0/4$

$\Delta(\boldsymbol{\pi}, \tilde{\boldsymbol{\sigma}}) = 0.5$

FIG. 3. Examples of the label continuity $\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma})$ and the label continuity error $\Delta(\boldsymbol{\pi}, \boldsymbol{\sigma})$ for different sequences and partitions for the same vertex set. The number on each vertex represents the original index of the vertex.

For a given partition $\boldsymbol{\sigma}$ and different vertex sequences, we can evaluate which vertex sequence is more consistent with $\boldsymbol{\sigma}$ using the LCE [e.g., Figs. 3(a) and 3(b)]. Similarly, for a given sequence $\boldsymbol{\pi}$ and different partitions, we can also evaluate which partition is more consistent with $\boldsymbol{\pi}$, keeping the group sizes $\{N_k\}$ fixed [e.g., Figs. 3(a) and 3(c)].

B. Properties of the LCE

The LCE can take only small values when the number of groups K is very small or large. For example, it is obvious that Δ is zero when $K = 1$ or $K = N$. In other words, the resolution of the LCE is low in such regions. Moreover, this property would depend on the distribution of the group sizes $\{N_k\}$. In this section, we quantify these intuitions.

The minimum value of Δ is zero by construction. The maximum value of Δ is obtained when labels are flipped the maximum number of times. The maximization of Δ by optimizing the sequence $\boldsymbol{\pi}$, given an arbitrary partition $\boldsymbol{\sigma}$ with $\{N_k\}$, is equivalent to maximizing Δ by optimizing partition $\boldsymbol{\sigma}$ (constrained to $\{N_k\}$) for a given sequence $\boldsymbol{\pi}$. We denote the maximum by max Δ as

$$\max_{\boldsymbol{\pi}} \Delta(\boldsymbol{\pi}, \boldsymbol{\sigma}) = \max_{\boldsymbol{\sigma}(\{N_k\})} \Delta(\boldsymbol{\pi}, \boldsymbol{\sigma}) = \max \Delta. \quad (4)$$

As derived in Appendix A, we have

$$\max \Delta = \begin{cases} \frac{2(N - \max_k N_k)}{N-1} - \frac{K-1}{N-1} & \left(\max_k N_k > \left\lceil \frac{N}{2} \right\rceil\right) \\ 1 - \frac{K-1}{N-1} & \text{(otherwise)} \end{cases}, \quad (5)$$

where $\lceil \cdot \rceil$ denotes the ceiling function.

We next investigate statistical properties of the LCE. First, we calculate the probability $\mathcal{P}(m)$ that the number of times that two consecutive vertices in a sequence have the same group label is m, where $m = (N-1)\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma})$. When any sequence realizes at random, we have

$$\mathcal{P}(m) = \frac{1}{N!} \sum_{\boldsymbol{\pi}'} \delta(m, (N-1)\mathcal{C}(\boldsymbol{\pi}', \boldsymbol{\sigma})), \quad (6)$$

where $\boldsymbol{\sigma}$ is an arbitrary partition with group sizes $\{N_k\}$ and the sum is over all possible sequences ($|\{\boldsymbol{\pi}'\}| = N!$). Note that Eq. (6) is also a distribution in which each distinct partition

realizes at random. This equivalence might sound peculiar because there are only $N! / \prod_{k=1}^{K} N_k!$ distinct partitions, whereas there are $N!$ possible sequences. However, because every distinct partition is overcounted exactly $\prod_{k=1}^{K} N_k!$ times in the summation of Eq. (6), the distribution $\mathcal{P}(m)$ is identical for both random sequences and random partitions.

Although Eq. (6) is a straightforward expression, a strict constraint on $\{N_k\}$ makes analytical calculations complicated. Therefore, we instead calculate the distribution of bootstrapped group labels $\boldsymbol{\sigma}^*$ as an approximation. That is, we generate a random group assignment $\boldsymbol{\sigma}^*$ by sampling independently from the empirical distribution $\text{Prob}[k] = N_k/N$ ($k \in \{1, \ldots, K\}$); in other words, we randomly resample group labels from $\boldsymbol{\sigma}$ with replacement. The distribution of group labels $\boldsymbol{\sigma}^*$ is

$$P(\boldsymbol{\sigma}^*) = \prod_{i=1}^{N} \frac{N_{\sigma^*(i)}}{N}. \quad (7)$$

This approximation for random group labels is expected to be accurate if each element in $\{N_k\}$ is sufficiently large.

Using the bootstrapped group labels, the mean value of $\mathcal{C}$ is obtained as

$$\mathbb{E}[\mathcal{C}] = \sum_{\boldsymbol{\sigma}^*} P(\boldsymbol{\sigma}^*) \frac{\sum_{i=1}^{N-1} \delta(\sigma^*(i), \sigma^*(i+1))}{N-1}$$

$$= \frac{1}{N-1} \sum_{i=1}^{N-1} \left[\sum_{\sigma_i^*, \sigma_{i+1}^*} P(\sigma_i^*) P(\sigma_{i+1}^*) \delta(\sigma^*(i), \sigma^*(i+1)) \right.$$

$$\left. \times \prod_{j(\neq i, i+1)} \left(\sum_{\sigma_j^*} P(\sigma_j^*) \right) \right]$$

$$= \sum_{k=1}^{K} \left(\frac{N_k}{N} \right)^2. \quad (8)$$

Therefore, the mean value of LCE under random partitioning is

$$\overline{\Delta}(\{N_k\}) := \mathbb{E}[\Delta] = \frac{N-K}{N-1} - \sum_{k=1}^{K} \left(\frac{N_k}{N} \right)^2. \quad (9)$$

As the LCE does not practically become greater than $\overline{\Delta}(\{N_k\})$, this mean value is a more meaningful reference value than Eq. (5) as the upper bound.

We can also derive the variance $\text{Var}[\Delta]$ (the derivation is shown in Appendix B) as

$$\text{Var}[\Delta] = \frac{1}{N-1} \sum_{k=1}^{K} \left(\frac{N_k}{N} \right)^2 + \frac{2(N-2)}{(N-1)^2} \sum_{k=1}^{K} \left(\frac{N_k}{N} \right)^3$$

$$- \frac{3N-5}{(N-1)^2} \left(\sum_{k=1}^{K} \left(\frac{N_k}{N} \right)^2 \right)^2, \quad (10)$$

showing that Δ converges to $\mathbb{E}[\Delta]$ by the law of large numbers. Furthermore, in Appendix C, we show that the probability distribution is asymptotically normal when the group sizes are equal and $K = O(1)$, implying that higher-order moments will vanish.
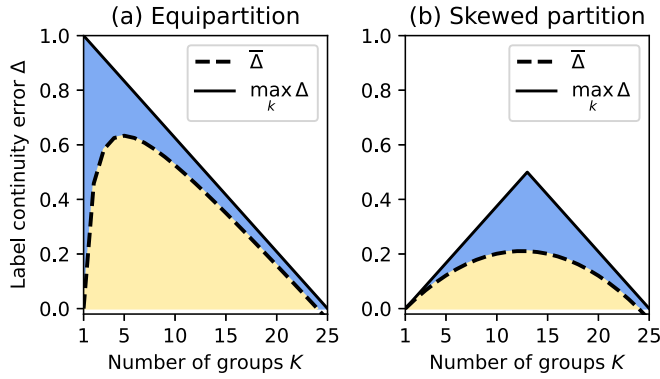
FIG. 4. Upper bound max Δ (solid line) of the LCE and bootstrap estimate of the mean values $\overline{\Delta}$ (dashed line) of the LCE under a random sequence as functions of the number of groups K: (a) an equipartition (i.e., $N_k/N = 1/K$ for any $k \in \{1, \dots, K\}$) and (b) a skewed partition [i.e., $N_k = 1$ $(2 \leqslant k \leqslant K - 1)$ and $N_1 = N - K + 1$]. We set $N = 25$ in these examples. Continuous approximations of max Δ and $\overline{\Delta}$ are shown to highlight their dependency on K.

Let us summarize the results we obtained in this section. As the number of groups $K \, (> 1)$ increases, the upper bound of the LCE [max Δ in Eq. (5)] decreases monotonically as long as the partitions are not highly skewed, i.e., $\max_k N_k < \lceil N/2 \rceil$. However, as illustrated in Fig. 4, the LCE for a random sequence [$\overline{\Delta}$ in Eq. (9)] is a convex function with respect to K. When K is small, the LCE increases because the chance for label flips increases, while the LCE decreases owing to the increase in the minimum number of label flips. For equipartitioning [Fig. 4(a)], the mean LCE $\overline{\Delta}$ is peaked at an integer of approximately $K = \sqrt{N - 1}$. As a partition becomes more skewed [Fig. 4(b)], max Δ and $\overline{\Delta}$ are peaked at larger values of K. Therefore, when evaluating the LCE, we must implement appropriate normalizations.

In this study, we focus on comparing partitions with the same number of groups K. In Appendix D, however, we discuss nested partitions (subpartitions of another partition) as an example in which different partitions have different numbers of groups.

III. SPECTRAL ORDERING METHODS

In this section, we describe variants of spectral ordering methods using different matrices. After reviewing the derivation of the standard methods based on the unnormalized and normalized Laplacians, we show how spectral ordering problems can be formulated with the modularity matrix, regularized Laplacian, and Bethe Hessian.

A. Unnormalized Laplacian

Spectral ordering is derived as a continuous relaxation of the discrete optimization problem called envelope reduction [14]. This problem optimizes the vertex sequence π such that each connected pair of vertices is located close to each other in the sequence. To this end, the following objective function

is considered:

$$H_2(\boldsymbol{\pi}; A) = \frac{1}{2} \sum_{i,j} A_{ij}(\pi_i - \pi_j)^2, \tag{11}$$

which is the sum of squared distances $(\pi_i - \pi_j)^2$ with respect to the set of connected vertices. The sequence that minimizes this function is the solution to envelope reduction.

As the minimization of Eq. (11) is not computationally feasible, we consider its continuous relaxation. That is, we represent $\boldsymbol{\pi}$ using a continuous vector $\boldsymbol{x} \in \mathbb{R}^N$. However, if we simply replace $\boldsymbol{\pi}$ with $\boldsymbol{x}$, $\boldsymbol{x} = \boldsymbol{0}$ would be the trivial minimizer of $H_2(\boldsymbol{x}; A)$. Thus, we constrain $\boldsymbol{x}$ such that $\sum_{i=1}^{N} x_i^2$ is a positive constant (i.e., the spherical constraint) to reflect the fact that $\sum_{i=1}^{N} \pi_i^2$ is positive regardless of the choice of sequence. Therefore, we consider the minimization of the following function:

$$\frac{1}{2} \sum_{i,j} A_{ij}(x_i - x_j)^2 - \lambda \left(\sum_{i=1}^{N} x_i^2 - 1 \right), \tag{12}$$

where λ is the Lagrange multiplier. The extremum condition in Eq. (12) yields the following eigenvalue equation with an eigenvector $\boldsymbol{v}$,

$$L\boldsymbol{v} = \lambda\boldsymbol{v}. \tag{13}$$

Here, $L \equiv D - A$ is the unnormalized (or combinatorial) Laplacian, where $D = \mathrm{diag}(d_1, \dots, d_N)$ is the degree matrix $(d_i = \sum_{j=1}^{N} A_{ij})$. Although we would have a vector proportional to $\boldsymbol{1}$ (a vector of ones) as the minimizer of Eq. (12), which is also the eigenvector associated with the smallest eigenvalue of L, we cannot infer the optimal sequence from $\boldsymbol{1}$ because all the elements are identical. Therefore, we exclude vectors proportional to $\boldsymbol{1}$, which is equivalent to imposing a perpendicular constraint to $\boldsymbol{1}$ in Eq. (12), i.e., $\sum_{i=1}^{N} x_i = 0$. Then, the minimizer of the objective function is the eigenvector $\boldsymbol{v}_2$ of L associated with the second-smallest eigenvalue.

The estimate of the optimal sequence $\hat{\boldsymbol{\pi}}$ using the spectral ordering method is

$$\hat{\boldsymbol{\pi}} = \{\mathrm{rank}(v_{2i}) | i \in \mathcal{I}\}, \tag{14}$$

where v_{2i} is the ith element of $\boldsymbol{v}_2$, and $\mathrm{rank}(v_{2i})$ is the index of v_{2i} in an array in which the vector elements of $\boldsymbol{v}_2$ are sorted in the ascending or descending order.

B. Normalized Laplacian

A spectral ordering method with the normalized Laplacian was derived in [15]. Note that the objective function (11) does not have a periodic boundary condition. Therefore, while the distance from one vertex at the end of the sequence to another vertex ranges from 1 to $N - 1$, the distance from the vertex at the middle of the sequence ranges from 1 to $\lfloor N/2 \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. This implies that when a graph has a vertex with a considerably large degree (i.e., a hub), it is typically more beneficial for the minimization objective to assign such a vertex near the middle of the sequence. To incorporate this feature, we replace the spherical constraint

in Eq. (12) with the following ellipsoidal constraint:

$$\sum_{i=1}^{N} d_i x_i^2 = \text{const.}, \tag{15}$$

which tends to restrict x_i with a large d_i to be relatively small (recall that a variable with a large coefficient typically has relatively small values on an ellipsoid). Therefore, Eq. (15) constrains $\boldsymbol{x}$ such that x_i of a hub vertex v_i is near the origin, and when $\boldsymbol{x}$ is discretized, the hub vertices are likely to be located near the middle of the sequence. Note also that the mean of $\{x_i\}$ is located at the origin because of the perpendicular constraint $\sum_{i=1}^{N} x_i = 0$.

Consequently, Eq. (13) is replaced with the following generalized eigenvalue equation with respect to its second-smallest eigenvalue λ_2:

$$L \boldsymbol{v}_2 = \lambda_2 D \boldsymbol{v}_2. \tag{16}$$

This is equivalent to

$$\mathcal{L} \boldsymbol{z}_2 = \lambda_2 \boldsymbol{z}_2, \tag{17}$$

where $\mathcal{L} \equiv D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ is the normalized Laplacian and $\boldsymbol{z}_2 \equiv D^{\frac{1}{2}} \boldsymbol{v}_2$. As $\boldsymbol{v}_2$ is a continuous relaxation of the sequence $\boldsymbol{\pi}$, we estimate the optimal sequence $\hat{\boldsymbol{\pi}}$ as

$$\hat{\boldsymbol{\pi}} = \left\{ \text{rank}\left(d_i^{-1/2} z_{2i} \right) \middle| i \in \mathcal{I} \right\}. \tag{18}$$

C. Modularity matrix

The modularity matrix Q appears in the spectral clustering method for modularity maximization in community detection [21]. The matrix element is commonly defined as

$$Q_{ij} = A_{ij} - \frac{d_i d_j}{2M}, \tag{19}$$

where M is the total number of edges in the graph.

To formulate the spectral ordering problem with the modularity matrix, we again consider the objective function $H_2(\boldsymbol{\pi}; A)$ in the envelope reduction problem and its continuous relaxation with the spherical constraint $\sum_{i=1}^{N} x_i^2 = 1$. Herein, we add the following penalty terms to the objective function:

$$\frac{\left(\sum_i d_i x_i \right)^2}{2M} - \sum_i d_i x_i^2. \tag{20}$$

The penalty terms ensure that $\{x_i\}$ are "balanced" around the origin. The first term prohibits $\{x_i\}$ for hub vertices from being located only on the positive or negative side of the real interval $[-1, 1]$. Owing to the second term, $\{x_i\}$ associated with hub vertices also tend to be away from the origin. Therefore, the penalty term Eq. (20) decreases when $\{x_i\}$ are more symmetrically distributed around the origin.

Using Lagrange multipliers, the objective function to be minimized is then

$$\frac{1}{2} \sum_{i,j} A_{ij} (x_i - x_j)^2 + \frac{\left(\sum_i d_i x_i \right)^2}{2M} - \sum_i d_i x_i^2 + \lambda \left(\sum_{i=1}^{N} x_i^2 - 1 \right)$$

$$= -\sum_{i,j} x_i Q_{ij} x_j + \lambda \left(\sum_{i=1}^{N} x_i^2 - 1 \right). \tag{21}$$

Here, we do not impose the perpendicular constraint in Eq. (21), because a vector proportional to **1** is not a trivial minimizer. The extremum conditions in Eq. (21) yield

$$Q \boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1, \tag{22}$$

where λ_1 represents the largest eigenvalue of Q and $\boldsymbol{v}_1$ is the associated eigenvector. $\boldsymbol{v}_1$ is the minimizer in Eq. (21) provided that it is not a vector proportional to **1**. Analogously to Eq. (14), we estimate the optimal sequence $\hat{\boldsymbol{\pi}}$ as

$$\hat{\boldsymbol{\pi}} = \{\text{rank}(v_{1i}) | i \in \mathcal{I}\}. \tag{23}$$

The ellipsoidal constraint enforces $\{x_i\}$ for hub vertices to be concentrated around the origin, whereas the penalty terms (20) enforce them to be evenly distributed at both the positive and negative ends of the real line. Therefore, the results of the spectral ordering methods using the normalized Laplacian and modularity matrix are expected to be quite distinct for graphs with heterogeneous degree distributions.

D. Bethe Hessian

Bethe Hessian is also a matrix that is originally formulated to perform spectral clustering [23,24]. This method is inspired by the statistical inference of the stochastic block model, which will be explained in Sec. IV A. This section considers a spectral ordering method using the Bethe Hessian.

The derivation of spectral ordering with the Bethe Hessian is analogous to that with the normalized Laplacian. However, instead of imposing an ellipsoidal constraint (15), we introduce $\sum_{i=1}^{N} d_i x_i^2$ as a penalty term. Thus, we consider the following objective function:

$$\frac{1}{2} \sum_{i,j} A_{ij} (x_i - x_j)^2 + \tau \sum_{i=1}^{N} d_i x_i^2, \tag{24}$$

where τ is an arbitrary constant (hyperparameter) that can be either positive or negative. To avoid the trivial minimizer $\boldsymbol{x} = \boldsymbol{0}$, we impose the spherical constraint $\sum_{i=1}^{N} x_i^2 = 1$.

Using Lagrange multipliers, the objective function to be minimized is then

$$\frac{1}{2} \sum_{i,j} A_{ij} (x_i - x_j)^2 + \tau \sum_i d_i x_i^2 - \lambda \left(\sum_{i=1}^{N} x_i^2 - 1 \right)$$

$$= (1 + \tau) \sum_{i,j} x_i B_{ij} x_j - \lambda \left(\sum_{i=1}^{N} x_i^2 - 1 \right), \tag{25}$$

where

$$B_{ij} = D_{ij} - r A_{ij} \quad \left(r = \frac{1}{1 + \tau} \right) \tag{26}$$

is a matrix element of Bethe Hessian. The extremum conditions in Eq. (25) yield an eigenvalue equation with respect to B.

We estimate the optimal sequence $\hat{\boldsymbol{\pi}}$ as follows:

$$\hat{\boldsymbol{\pi}} = \{\text{rank}(v_{2i}) | i \in \mathcal{I}\}, \tag{27}$$

where $\boldsymbol{v}_2$ is the eigenvector associated with the second-smallest eigenvalue λ_2, i.e.,

$$B\,\boldsymbol{v}_2 = \lambda_2 \boldsymbol{v}_2. \tag{28}$$

Note that there is no guarantee that $\boldsymbol{v}_2$ always provides the best estimate in terms of $H_2(\hat{\boldsymbol{\pi}}; A)$ among all the eigenvectors. In fact, we confirmed that the eigenvector that yields the best estimate in terms of $H_2(\hat{\boldsymbol{\pi}}; A)$ [when we employ the rounding rule in Eq. (27)] depends sensitively on the value of r, particularly when r is small (see Sec. S1 in Supplemental Material [30] for details). However, we employ Eq. (27) because the estimate with $\boldsymbol{v}_2$ offers the smallest value of $H_2(\hat{\boldsymbol{\pi}}; A)$ as long as r is sufficiently large.

Throughout this study, we set $r = \sqrt{\sum_i d_i^2 / \sum_i d_i - 1}$ (>1), because it is a commonly employed value in spectral clustering. The hyperparameter τ is negative when $r > 1$. Thus, $\{x_i\}$ for hub vertices are aligned near the ends of the real line $[-1, 1]$ so that a sequence achieves a lower value of Eq. (24) using the penalty term. By contrast, when $r < 1$ $(\tau > 0)$, $\{x_i\}$ for hub vertices are likely to be located near the origin, implying that the resulting sequence is similar to that obtained by the spectral ordering method based on the normalized Laplacian.

E. Regularized Laplacian

During the past decade, it has been found that the performance of the Laplacian-based spectral clustering can be considerably improved by adding a constant value to every element in the adjacency matrix [26,28] or the diagonal elements in the degree matrix [25,27]. Although the two variants of the Laplacian are often termed differently, we collectively refer to them as the regularized Laplacian [28] for simplicity, and we denote the former version of the regularized Laplacian as $\mathsf{L}^{(\tau)}$ and the latter version as L. The spectral clustering method based on L can also be interpreted as a continuous relaxation of the minimization of the *core cut* function [29]. This section considers the spectral ordering method using a regularized Laplacian.

Similar to the formulation of the spectral ordering method with the modularity matrix, we consider the continuous relaxation of $H_2(\boldsymbol{\pi}; A)$ with a penalty term. We consider the following objective function:

$$\frac{1}{2} \sum_{i,j} A_{ij}(x_i - x_j)^2 + \tau N \operatorname{Var}[\boldsymbol{x}], \tag{29}$$

which is minimized with respect to the continuous vector $\boldsymbol{x}$. τ is an arbitrary positive constant (hyperparameter) and

$$\operatorname{Var}[\boldsymbol{x}] = \left(\frac{1}{N} \sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^2 \right) \tag{30}$$

is the variance with respect to the elements in $\boldsymbol{x}$. To ensure that $\boldsymbol{x}$ is not a vector of zeros, we impose the following ellipsoidal constraint:

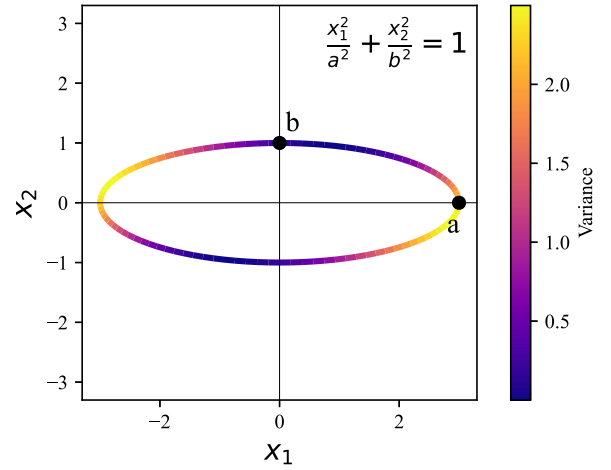$$\sum_{i=1}^{N} (d_i + \tau) x_i^2 = 1. \tag{31}$$

FIG. 5. Ellipse equation $x_1^2/a^2 + x_2^2/b^2 = 1$ ($a = 3$ and $b = 1$), where x_2 corresponds to the variable for a hub vertex. The color depth represents the variance $\operatorname{Var}[\boldsymbol{x}]$ for $\boldsymbol{x} = (x_1, x_2)$. Although most of the coordinates on the ellipse have $x_1 > x_2$, the variance is smaller when x_1 and x_2 are closer.

By incorporating this constraint, the objective function to be minimized is given by

$$\frac{1}{2} \sum_{i,j} A_{ij}(x_i - x_j)^2 + \tau \sum_{i=1}^{N} x_i^2 - \frac{\tau}{N} \left(\sum_{i=1}^{N} x_i \right)^2$$
$$- \lambda \left(\sum_{i=1}^{N} (d_i + \tau) x_i^2 - 1 \right). \tag{32}$$

Because a vector proportional to $\mathbf{1}$ is a trivial minimizer of Eq. (32), we also impose the constraint that $\boldsymbol{x}$ is perpendicular to $\mathbf{1}$, i.e., $\sum_{i=1}^{N} x_i = 0$. Then, the extremum conditions in Eq. (32) yield

$$A^{(\tau)} \boldsymbol{v}_2 = (1 - \lambda_2)(D + \tau I) \boldsymbol{v}_2, \quad \left(A_{ij}^{(\tau)} = A_{ij} + \frac{\tau}{N} \right), \tag{33}$$

where I is the identity matrix. λ_2 is the second-smallest eigenvalue of the generalized eigenvalue equation and $\boldsymbol{v}_2$ is the associated generalized eigenvector. Equation (33) is equivalent to

$$\mathsf{L}^{(\tau)} \boldsymbol{z}_2 = \lambda_2 \boldsymbol{z}_2, \tag{34}$$

where

$$\mathsf{L}^{(\tau)} = I - (D + \tau I)^{-1/2} A^{(\tau)} (D + \tau I)^{-1/2} \tag{35}$$

is the regularized Laplacian and $\boldsymbol{z}_2 = (D + \tau I)^{1/2} \boldsymbol{v}_2$. Similar to the spectral ordering method with the normalized Laplacian, we estimate the optimal sequence $\hat{\boldsymbol{\pi}}$ as

$$\hat{\boldsymbol{\pi}} = \{ \operatorname{rank}((d_i + \tau)^{-1/2} z_{2i}) | i \in \mathcal{I} \}. \tag{36}$$

The contribution of hub vertices is more complicated than that of other methods. As shown in Fig. 5, whereas $\{x_i\}$ for the hub vertices tend to be relatively small because of the ellipsoidal constraint, the variance $\operatorname{Var}[\boldsymbol{x}]$ is minimized when all $\{x_i\}$ have the same value. Therefore, when the hyperparameter τ is small, the result is similar to that obtained by using the spectral ordering method based on the normalized Laplacian.

TABLE I. Summary of constraints and penalty terms in spectral methods, and their effect on hub location in vertex sequence.

Matrix	Constraints	Penalty terms	Effects on hub locations
Unnormalized Laplacian $L = D - A$	$\sum_{i=1}^{N} x_i^2 = 1$ $\sum_{i=1}^{N} x_i = 0$		
Normalized Laplacian $\mathcal{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$	$\sum_{i=1}^{N} d_i x_i^2 = 1$ $\sum_{i=1}^{N} x_i = 0$		Concentrate around the middle
Modularity matrix $Q = A + \frac{d^\top d}{2M}$ $(d = (d_1, \ldots, d_N))$	$\sum_{i=1}^{N} x_i^2 = 1$	$\frac{(\sum_i d_i x_i)^2}{2M} - \sum_i d_i x_i^2$	Distribute at both ends
Bethe Hessian $B = D - rA$	$\sum_{i=1}^{N} x_i^2 = 1$ perpendicular to v_1	$\tau \sum_{i=1}^{N} d_i x_i^2$	$\tau > 0$: Concentrate around the middle $\tau < 0$: Distribute at either/both ends
Regularized Laplacian $\mathsf{L}^{(\tau)} = I - D_{(\tau)}^{-1/2} A^{(\tau)} D_{(\tau)}^{-1/2}$ $(D_{(\tau)} = D + \tau I)$	$\sum_{i=1}^{N} (d_i + \tau) x_i^2 = 1$ $\sum_{i=1}^{N} x_i = 0$	$\tau N \, \mathrm{Var}[x]$	small τ: Concentrate around the middle large τ: Avoid concentration around the middle
Regularized Laplacian $\mathsf{L} = D_{(\tau)}^{-1/2} A D_{(\tau)}^{-1/2}$	$\sum_{i=1}^{N} (d_i + \tau) x_i^2 = 1$ perpendicular to v_1	$\tau \sum_{i=1}^{N} x_i^2$	small τ: Concentrate around the middle large τ: Avoid concentration around the middle

As τ increases, the hub vertices are less likely to be located in the middle of the sequence because of the penalty term.

As mentioned above, we also consider

$$\mathsf{L} = (D + \tau I)^{-1/2} A (D + \tau I)^{-1/2} \qquad (37)$$

as the definition of a regularized Laplacian. Unlike $\mathsf{L}^{(\tau)}$, only the degree matrix is perturbed by a constant value in L. If we consider

$$\frac{1}{2} \sum_{i,j} A_{ij} (x_i - x_j)^2 + \tau \sum_{i=1}^{N} x_i^2 \qquad (38)$$

as the objective function to be minimized, and impose the ellipsoidal constraint (31), we obtain the eigenvalue equation with respect to L as a result of the extremum conditions.

If we also impose the constraint $\sum_{i=1}^{N} x_i = 0$ in Eq. (38), this objective function becomes equivalent to Eq. (29). We do not have such a constraint because L does not have $\mathbf{1}$ as a trivial eigenvector unlike $\mathsf{L}^{(\tau)}$. The eigenvectors of $\mathsf{L}^{(\tau)}$ and L are therefore distinct. As a spectral ordering method with the regularized Laplacian L, we replace z_2 in Eq. (36) with the eigenvector associated with the second-smallest eigenvalue of L. Here, we use the second-smallest value because L approaches $\mathcal{L}$ as $\tau \to 0$, and $\mathcal{L}$ has $z_1 \propto \mathbf{1}$. Hereafter, when we refer to the spectral ordering method with the regularized Laplacian, we employ L because it is more computationally efficient. Throughout this study, we set τ as the average degree of the graph, as it is a commonly employed value [27].

The constraints and penalty terms for each method are summarized in Table I. Compared to the classical method based on the normalized Laplacian, where hub vertices are concentrated around the middle of the sequence ("hub-centered"), the spectral ordering methods obtained with the modularity matrix, Bethe Hessian, and the regularized Laplacian may assign hub vertices at both ends of the sequence ("hub-at-the-corner"). Particularly, for the Bethe Hessian and the regularized Laplacian, we can choose the "hub-centered" or "hub-at-the-corner" alignment by tuning the hyperparameter.

Although we found the penalties and constraints that provide the spectral ordering methods corresponding to the ones considered in spectral clustering, we have not confirmed whether these choices of penalties and constraints are unique. In addition, there is no guarantee that the resulting spectral ordering methods exhibit high performance in practice. The next section investigates the practical performance of these spectral ordering and clustering methods using synthetic and real-world datasets. Although one might expect all the methods to work similarly when the graph is close to regular, it is not trivial to determine whether this always holds; this is investigated using synthetic datasets in Secs. IV A and IV B. The effect of heterogeneous degree distribution in each method is examined using real-world datasets in Sec. IV C.

IV. PERFORMANCE ANALYSIS

We conduct a numerical performance analysis of the spectral ordering and clustering methods using synthetic graphs and real-world networks. For experiments on synthetic graphs, we consider a random graph model with a prespecified module structure, which is referred to as the stochastic block model (SBM) [31–33], and a random graph model with a prespecified sequentially local structure, which is referred to as the ordered random graph model (ORGM) [10].

A. Stochastic block model

The SBM is often used as a generative model for the inference of module structures in graphs [34,35] and in several theoretical studies in the community detection literature [36]. In the SBM, each vertex has a "planted" (or preassigned) group assignment; we denote the corresponding partition as σ^B. Each vertex pair is connected by an edge, independently and randomly, based on the planted group assignments. The probabilities for the upper-right elements of the adjacency matrix are given as follows:

$$\text{Prob}[\{A_{ij}\}_{i<j}] = \prod_{i<j} p_{\sigma_i^B \sigma_j^B}^{A_{ij}} \left(1 - p_{\sigma_i^B \sigma_j^B}\right)^{1-A_{ij}}, \quad (39)$$

where $p_{k\ell}$ is the probability that a vertex in group k and vertex in group ℓ are connected [in Eq. (39), $k = \sigma_i^B$ and $\ell = \sigma_j^B$]. We have $A_{ij} = A_{ji}$ for any pair of elements because we consider undirected graphs. In general, the SBM can generate graphs with complex module structures. Herein, however, we focus on the SBM with a community structure that is characterized by the following group-wise connection probability:

$$p_{k\ell} = \begin{cases} p_{\text{in}} & (k = \ell) \\ p_{\text{out}} & (k \neq \ell) \end{cases}, \quad (40)$$

where $0 < p_{\text{out}} \leqslant p_{\text{in}} \leqslant 1$, that is, vertices are more densely connected within the same planted group than between different groups. In particular, when the group sizes are equal, it is common to parametrize the model using the average degree c and the fuzziness parameter ϵ, which are related to p_{in} and p_{out} as

$$c = \frac{N}{K}(p_{\text{in}} + (K-1)p_{\text{out}}), \quad \epsilon = \frac{p_{\text{out}}}{p_{\text{in}}}. \quad (41)$$

As ϵ approaches unity, the planted community structure becomes less clear. This particular case of the SBM is known as the planted partition model [37]. For a given average degree c, the critical value of ϵ above which an algorithm cannot detect the planted block structure better than chance is called the (algorithmic) detectability limit [38–43].

Using the SBM and spectral ordering methods, we investigate the following questions:

(1) How would the reordered adjacency matrix look like? Can we visually identify the community structure through the matrix?

(2) When and how would the spectral ordering methods lose their correlations with the planted partition in the SBM? Are the spectral ordering methods superior or inferior to their clustering counterparts in detecting the planted partition? Does the choice of matrix matter?

To answer these questions, we apply both spectral ordering and clustering methods to graphs generated by the SBM.

We first investigate the former question. Figure 6 shows the results of spectral ordering applied to instances of SBM. Vertices in the same planted group are indeed located closely in the inferred sequence when the community structure is strong. Even when the community structure is weak, the planted group labels and the inferred sequence are correlated. In both examples, the boundaries of the groups are ambiguous. Therefore, if we do not know the planted group labels (and without the coloring of the adjacency matrix elements), it is not clear whether the identified structure is a community structure or
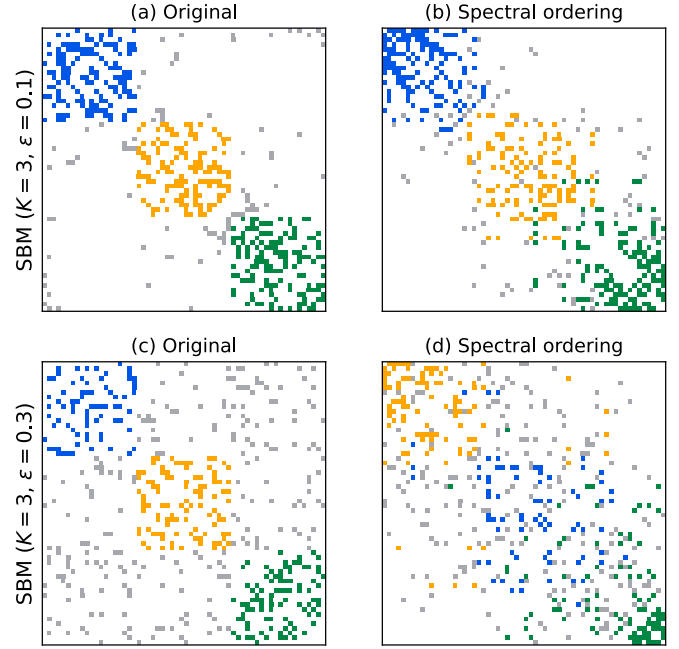

FIG. 6. Spectral ordering with the regularized Laplacian for graphs generated by the SBM ($N = 60$ and $c = 8$). The top panels show the adjacency matrices of a graph with a strong community structure where vertices are aligned by (a) the sequence based on the planted partition σ^B and (b) an inferred sequence $\hat{\pi}$. The bottom panels show the adjacency matrices of a graph with a weak community structure where vertices are aligned based on (c) σ^B and (d) $\hat{\pi}$. The matrix elements indicating the connections within the same planted group are represented in the same color; otherwise, the elements are colored in gray.

a banded structure from the reordered adjacency matrix. Note that, as discussed in [10], even when we generate a graph from a uniformly random graph model, one could identify a weak banded structure owing to the ordering of vertices.

Next, we address the latter question. The consistency between the sequence inferred by a spectral ordering method $\hat{\pi}$ and planted partition σ^B is measured with the normalized LCE $\Delta(\hat{\pi}, \sigma^B)/\overline{\Delta}(\{N_k^B\})$. Here, $\{N_k^B\}$ is the set of group sizes in σ^B and $\overline{\Delta}(\{N_k^B\})$ is the LCE under a random sequence defined in Eq. (9). When $\Delta(\hat{\pi}, \sigma^B)$ saturates (i.e., the normalized LCE equals unity) as ϵ increases, the spectral ordering method does not infer σ^B better than random; it is deemed that the algorithm has reached the detectability limit.

The consistency between the inferred partition $\hat{\sigma}$ by a spectral clustering method and planted partition σ^B is measured using the normalized mutual information (NMI) [44], which is defined as

$$\text{NMI}(\hat{\sigma}, \sigma^B) = \frac{2I(\hat{\sigma}; \sigma^B)}{H(\hat{\sigma}) + H(\sigma^B)}, \quad (42)$$

where

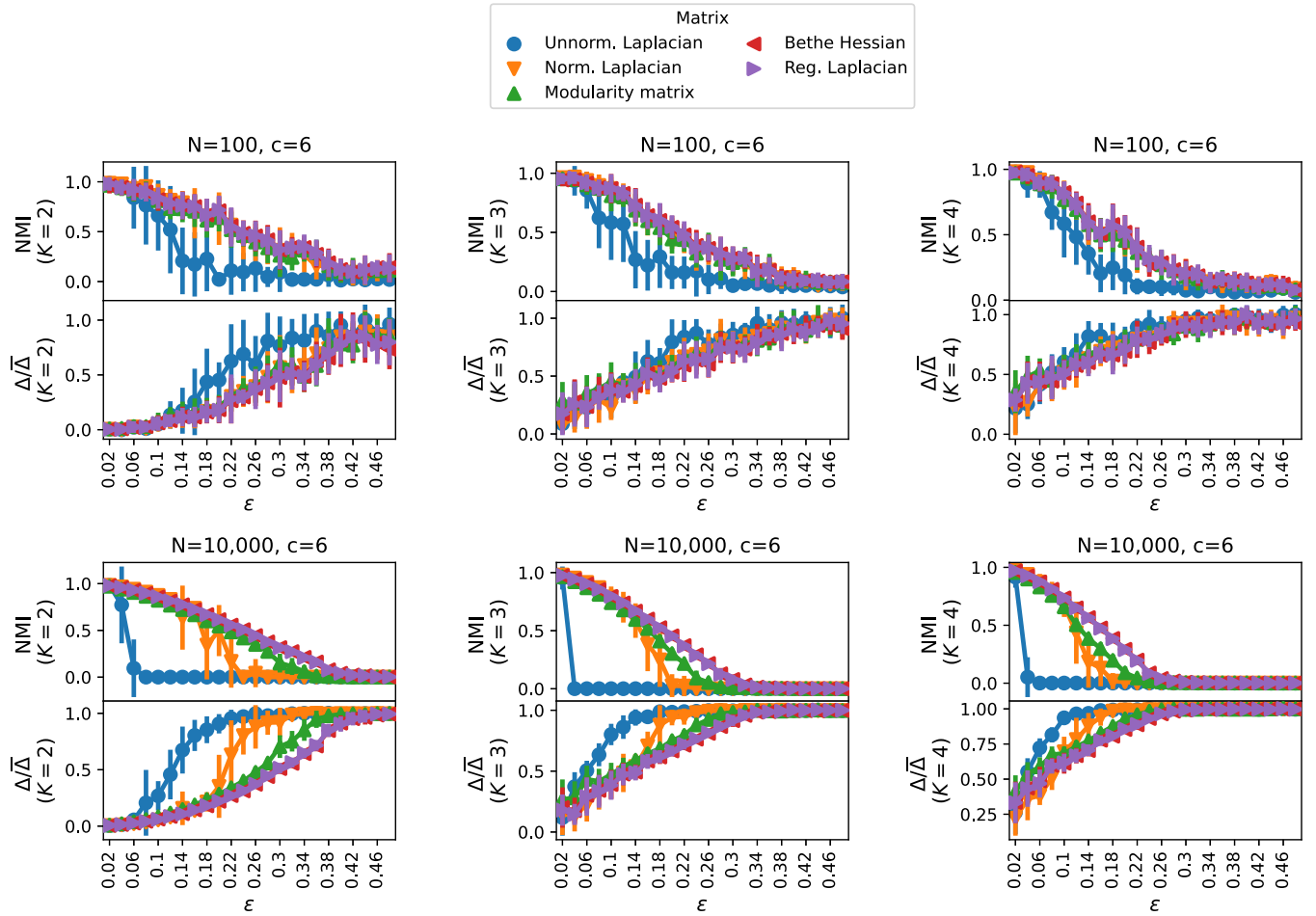$$H(\sigma) = -\sum_{k \in \{1,\dots,K\}} q(k) \log q(k) \quad \left(q(k) = \frac{N_k}{N}\right) \quad (43)$$

FIG. 7. Detectability of the SBM for the spectral ordering and spectral clustering methods. The top panels show the result for small graphs with $K = 2$, $K = 3$, and $K = 4$. The bottom panels show the result for large graphs with $K = 2$, $K = 3$, and $K = 4$. In each panel, the values of the NMI obtained by the spectral clustering methods (top) and the values of the LCE obtained by the spectral ordering methods (bottom) are shown. The horizontal axis represents the fuzziness of community structure ϵ. Each symbol and error bar represents the mean and the standard deviation of 30 samples that are obtained with the same SBM parameters.

is the entropy with respect to the frequency of group labels, and

$$I(\boldsymbol{\sigma}_1; \boldsymbol{\sigma}_2) = \sum_{k \in \{1,...,K\}} \sum_{k' \in \{1,...,K\}} q(k, k') \log \frac{q(k, k')}{q(k)q(k')} \quad (44)$$

is the mutual information. Here, $q(k, k')$ is the fraction of co-occurrences that a vertex belonging to group k in partition $\boldsymbol{\sigma}_1$ belongs to group k' in partition $\boldsymbol{\sigma}_2$. The NMI is unity when a pair of partitions coincides perfectly. When NMI($\hat{\boldsymbol{\sigma}}, \boldsymbol{\sigma}^B$) reaches (nearly) zero as ϵ increases, the spectral clustering method does not infer $\boldsymbol{\sigma}^B$ better than random, which again represents the detectability limit. The detectability analysis of spectral clustering methods is not new and has been analyzed in several theoretical and benchmark studies [41–43,45,46]. We evaluate $\Delta(\hat{\boldsymbol{\pi}}, \boldsymbol{\sigma}^B)$ and NMI($\hat{\boldsymbol{\sigma}}, \boldsymbol{\sigma}^B$) to compare the performances of the spectral ordering and clustering methods for each of the matrices considered in the previous section.

Figure 7 shows the performances of the ordering and clustering methods based on the SBM for different graph sizes N, numbers of blocks K, and fuzziness parameter ϵ. When graphs are small (and thus relatively dense), there is no clear

saturation in the curves of the LCE and the NMI, and it is difficult to evaluate whether the ordering methods or clustering methods exhibit superior performance in terms of the detectability limit. Moreover, the differences in performances are not noticeable among the different matrices, except for the unnormalized Laplacian. When graphs are large, we can clearly identify the saturation. For the unnormalized and normalized Laplacians, the values of LCE gradually decrease, even when the values of NMI saturate, indicating that the spectral ordering methods are superior to their clustering counterparts. By contrast, the detectability limits of the modularity matrix, regularized Laplacian, and Bethe Hessian are not very different between the ordering and clustering methods. In addition, the methods with the regularized Laplacian and Bethe Hessian perform similarly and are superior to the other matrices, whereas the methods with the unnormalized Laplacian are clearly inferior.

B. Ordered random graph model

We have observed how and to what extent the community structure can be inferred using spectral ordering methods. This

section discusses the opposite scenario. That is, we analyze whether the spectral clustering methods can infer banded structures. To this end, we conduct a performance analysis using the ORGM. This section uses the K-means method to determine the group labels in spectral clustering.

The vertex set in the ORGM has a planted sequence, as the vertex set in the SBM has a planted partition. We let the planted sequence coincide with the original sequence $\mathcal{I}$. The edges in the ORGM are generated independently and randomly by referring to the planted sequence. We divide the space of the adjacency matrix elements into two regions, Ω_{in} and Ω_{out}. Ω_{in} (resp. Ω_{out}) is the set of elements in which an edge connects two vertices that are deemed to be "close" (resp. "not close") to each other. An edge is generated between a vertex pair with probability p_{in} if they are "close" and with probability p_{out} otherwise. Therefore, the probabilities of the upper-right elements of the adjacency matrix are given as follows:

$$P(\{A_{ij}\}_{i<j}|\{p_{ij}\}) = \prod_{i<j} p_{ij}^{A_{ij}}(1-p_{ij})^{1-A_{ij}}, \qquad (45)$$

$$p_{ij} = \begin{cases} p_{in} & (i,j) \in \Omega_{in} \\ p_{out} & (i,j) \in \Omega_{out} \end{cases}. \qquad (46)$$

We set the boundary of Ω_{in} and Ω_{out} as

$$\Omega_{in} = \{(i,j)||i-j| \leqslant r\}$$
$$\Omega_{out} = \{(i,j)||i-j| > r\}, \qquad (47)$$

where r is the bandwidth that specifies the boundary of the regions. Although Eq. (47) is a simple one, we note that the boundary in the ORGM can be more complex in general. In the following, instead of p_{in} and p_{out}, we specify the edge density by the average degree c and the strength of the banded structure $\epsilon = p_{out}/p_{in}$; when $\epsilon = 0$, nonzero elements in the adjacency matrix are completely confined within Ω_{in}, whereas the model is uniformly random when $\epsilon = 1$. [See Fig. 8(a) for an example of the resulting adjacency matrix of the ORGM.] In summary, except for the number of vertices N, which is a nuisance parameter, the parameters in the ORGM are the average degree c, strength of the banded structure ϵ, and bandwidth ratio r/N.

Using the ORGM and the spectral clustering methods, we investigate the following questions:

(1) How would the reordered adjacency matrix look like? Can we visually identify the banded structure through the matrix?

(2) How and when would the spectral clustering algorithms lose their correlations with the planted ordering in the ORGM?

We first investigate the former question. Figure 8 shows the results of a spectral clustering method with different values of K applied to a graph generated by the ORGM. A graph tends to be partitioned into equally-sized groups (see also Fig. S2 in the Supplemental Material [30] for a quantitative evidence). Recall that we observe a banded structure through a spectral ordering method even when the graph is generated from the SBM (Fig. 6). Analogously, we can identify block-diagonal structures in Fig. 8 although the graph is generated from the ORGM. This is an interesting observation
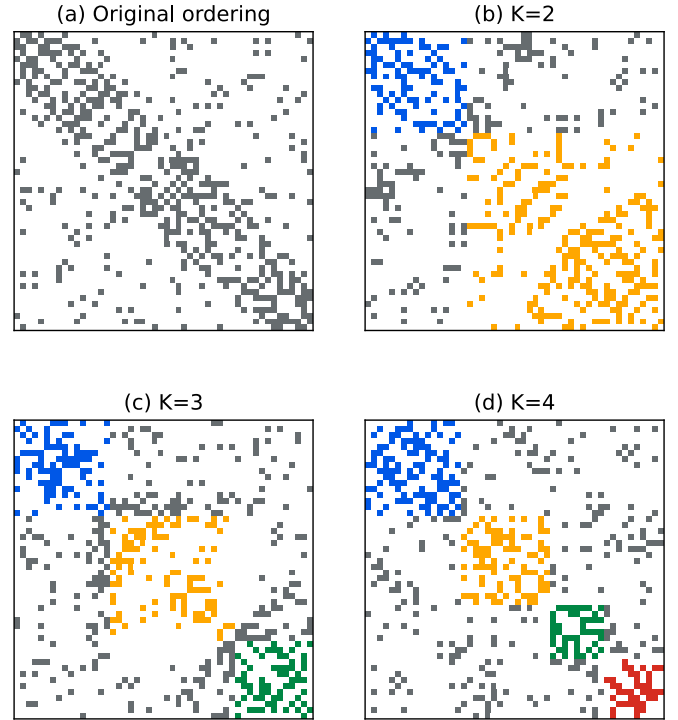

FIG. 8. Results of the spectral clustering methods using the normalized Laplacian with different numbers of groups K, applied to a graph generated by the ORGM. The parameters of the ORGM are $N = 50$, $c = 10$, $\epsilon = 0.2$, and $r/N = 0.16$. (a) The vertices of the adjacency matrix are ordered based on the original ordering in the ORGM. For panels (b)–(d), the vertices are ordered such that the vertices in the same inferred group are close to each other: (b) $K = 2$, (c) $K = 3$, and (d) $K = 4$. The nonzero matrix elements are represented by the same color when they are edges connecting the vertices within an inferred group; otherwise, the nonzero matrix elements are colored in gray.

because it implies that some of the community structures identified in the literature may be better described by banded structures.

Figure 9 shows the normalized LCE $\Delta(\mathcal{I}, \hat{\boldsymbol{\sigma}})/\overline{\Delta}(\{\hat{N}_k\})$ between the planted sequence $\mathcal{I}$ and inferred partition $\hat{\boldsymbol{\sigma}}$, where $\{\hat{N}_k\}$ is the set of group sizes in $\hat{\boldsymbol{\sigma}}$. The normalized LCE is generally low when r/N is not too small or large and ϵ is small.

The existence of detectability limits is implied from Figure 9. In the limit of $N \to \infty$, there exists a critical value of ϵ above which the normalized LCE is unity for any value of r/N. Moreover, for a given ϵ, there also exists an upper limit (and possibly a lower limit) of the bandwidth ratio r/N beyond which a spectral clustering method is not correlated with the planted sequence better than a random guess. These critical values depend on the average degree (see Fig. S3 in the Supplemental Material [30] for the numerical phase diagrams).

Analogous to the analysis for the SBM, the performance of the unnormalized Laplacian is notably inferior in terms of the normalized LCE; in most of the parameter sets, it does not perform better than a random guess. The behaviors of the modularity matrix, Bethe Hessian, and regularized Laplacian
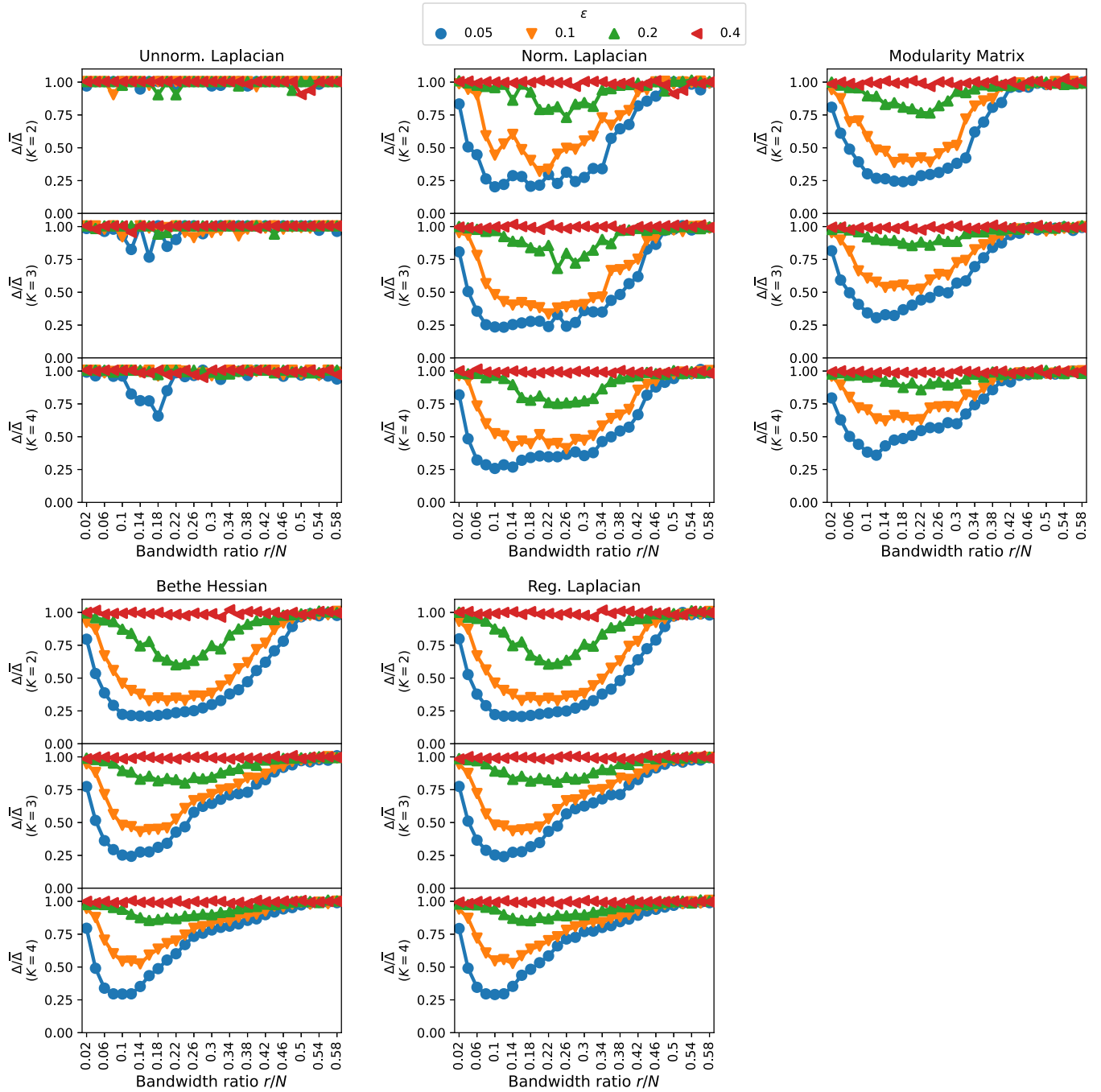
FIG. 9. Performance of the spectral clustering method for the graphs generated by the ORGM. The graphs are generated by the ORGM with $N = 1,000$ and $c = 6$. Each panel shows the normalized LCE $\Delta(\mathcal{I}, \hat{\boldsymbol{\sigma}})/\overline{\Delta}(\{\hat{N}_k\})$ for various parameter sets of the ORGM for a matrix used in the spectral clustering method. Each point represents the 10-sample average of the normalized LCE under the same parameter set.

are similar. Moreover, the results for the latter two matrices are apparently identical. In contrast to the analysis of graphs generated by the SBM, the performance of the normalized Laplacian is as good as or even better than that of the Bethe Hessian and regularized Laplacian.

The inferior performance of the spectral clustering with the unnormalized Laplacian can also be characterized by the distribution of the group sizes $\{N_k\}$. The fraction of the largest group $\max_k N_k/N$ is nearly unity, i.e., most of the vertices belong to the same group (see Fig. S2 in the Supplemental Material [30] for the experimental results). In such a case, the

result of clustering contains very little information about the inherent ordering in the graph; as shown in Fig. 4(b), the upper bound $\max \Delta$ and the mean value under the random sequence $\overline{\Delta}$ are small when a partition is highly skewed, reflecting the fact that the group labels tend to be aligned consecutively for any sequence. A possible mechanism for such skewed distributions of group sizes is the emergence of localized eigenvectors [42,47], which deteriorates the performance of spectral clustering. However, we do not pursue the detailed mechanisms that could have caused the outcome obtained in this study.
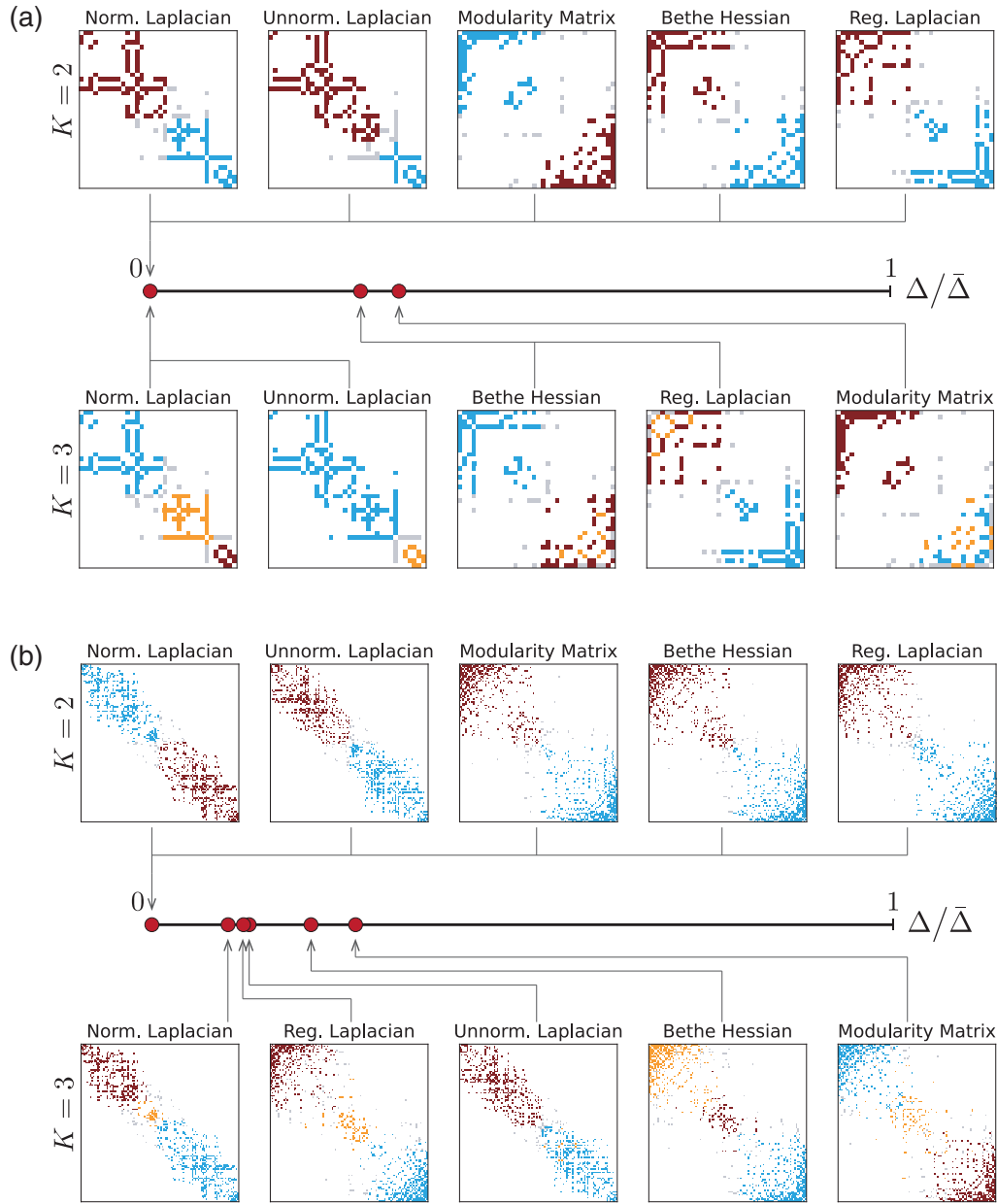
FIG. 10. Adjacency matrix aligned with spectral ordering based on a matrix annotated at the top and its corresponding LCE, $\Delta/\bar{\Delta}$. Colors denote vertex groups inferred by the K-means method. (a) Zachary's karate club network [48] and (b) a network of political books [49].

In summary, we have confirmed that some spectral clustering methods detect community structures that are correlated to the inherent sequential structure of the ORGM, and that there are nontrivial limits of detectability.

C. Real-world networks

We now apply the spectral ordering and clustering methods to five empirical adjacency matrices. Descriptions of the empirical datasets examined are provided in Table S1 in the Supplemental Material [30]. Note that many empirical datasets exhibit a high degree heterogeneity, whereas the synthetic graphs in Secs. IV A and IV B do not. As discussed in Sec. III, spectral orderings with different matrices are characterized as the minimization problem of H_2 with different constraints and penalty terms (Table I), and these differences become prominent when vertex degrees are heterogeneous.

In Fig. 10, we see a banded structure for the vertex orderings based on the normalized and unnormalized Laplacians, L and $\mathcal{L}$, for the karate club [Fig. 10(a)] and political books datasets [Fig. 10(b)], where the hub vertices tend to be located around the middle of the optimized sequence. In contrast, the ordering method with the modularity matrix Q locates vertices with large degrees at both ends of the sequence, as expected from the penalty term in the objective function (20). A similar observation applies to the methods using the Bethe Hessian B and regularized Laplacian L (Figs. 10, S4, and S7 in the Supplemental Material [30]). Importantly, however, vertex orderings based on these matrices is critically influenced by the regularization parameter τ.

For many empirical graphs, the hyperparameter $r = \sqrt{\sum_i d_i^2 / \sum_i d_i - 1}$ in the Bethe Hessian takes a large positive value, i.e., $\tau < 0$ [Eq. (24)]. Thus, the penalty term $\tau \sum d_i x_i^2$ contributes to reducing the objective function. Hence, hub vertices tend to be aligned at the ends of the vertex sequence. However, the spectral ordering with the regularized Laplacian has an exogenous regularization parameter τ in its constraint and penalty terms [see Eqs. (32) and (38)], where we set τ as the average degree. As discussed in Sec. III E, a larger value of τ tends to avoid locating hub vertices around the middle of the sequence. Although the validation analysis based on synthetic graphs suggested that the sequences inferred based on these matrices are fairly similar (Sec. IV), they do not necessarily coincide in general. Note also that, as $\tau \to 0$, the Bethe Hessian B approaches the unnormalized Laplacian L, and the regularized Laplacian L approaches the normalized Laplacian $\mathcal{L}$ (Table I). Therefore, when τ is small in absolute value, the optimal vertex sequences based on the Bethe Hessian and regularized Laplacian are close to those obtained from the unnormalized and normalized Laplacians, respectively (Figs. S8 and S9 in the Supplemental Material [30]). Indeed, the location of vertices with large degrees in optimal vertex sequence can be tuned by varying τ, from a "hub-centered" alignment to a "hub-at-the-corner" alignment.

Figure 10 also shows the normalized LCE representing the consistency between the inferred sequence $\hat{\pi}$ and group labels $\hat{\sigma}$ for each matrix used in the spectral ordering and clustering methods. When we set $K = 2$ in the clustering method, as shown in Figs. 10(a) and 10(b), $\hat{\pi}$ and $\hat{\sigma}$ are perfectly consistent in terms of the LCE. For $K \geqslant 3$, the LCEs are mostly lower than 0.8 and are typically approximately 0.5 (Fig. 10 and Figs. S10–S14 in the Supplemental Material [30]), suggesting that the optimized vertex sequences using spectral ordering convey some information about a nonrandom structure. We also find that some methods yield similar LCEs for all datasets, whereas the LCEs obtained with the (un)normalized Laplacian exhibit different behaviors (Figs. S10–S14 in the Supplemental Material [30]). This is consistent with the previous numerical observation that the spectral ordering based on the (un)normalized Laplacian is quite distinct from those obtained from the modularity matrix, Bethe Hessian, and regularized Laplacian (Fig. 10).

Interestingly, despite having distinct optimized sequences using different objective functions, the value of the normalized LCE can be very close to each other. Therefore, adjacency matrices may exhibit the same or similar structures from the perspective of community structure, and they are differentiated only by detailed orderings within each group.

V. SUMMARY AND DISCUSSION

This study analyzed the relationship between the ordering and clustering methods for graphs by quantifying the extent to which vertices close to each other in the optimized sequence have the same group label through the LCE. To obtain analytical insight into spectral ordering, we first showed that the spectral ordering problem is formulated as a minimization of the squared sequential distance H_2 subject to a particular
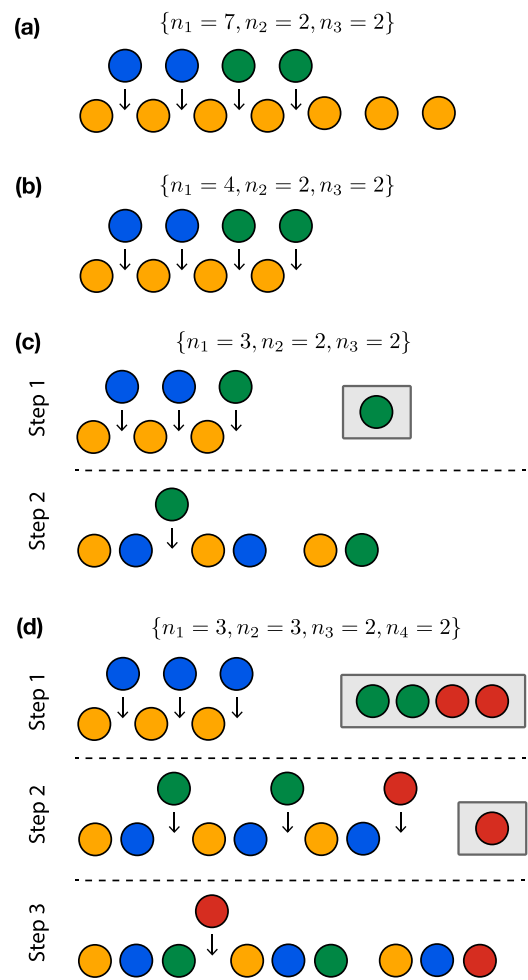

FIG. 11. Vertex sequence yielding the maximum LCE for a given group sizes $\{N_k\}$. Panels (a) and (b) are cases where $N_1 > \lceil N/2 \rceil$ and $N_1 = N/2$, respectively, and panels (c) and (d) are cases where $N_1 < N/2$. The sequence with the maximum LCE can be constructed by aligning the vertices with different labels alternately in the procedure shown in each step. Vertices in the box indicate that they are to be aligned in the following steps. The vertex indices are omitted because they are not essential for the construction of a sequence.

penalty function and constraints, depending on the matrix representation of a graph (i.e., normalized Laplacian, modularity matrix, etc). The numerical results suggested that the spectral ordering methods, except that based on unnormalized Laplacian, often yield optimized sequences such that vertices in the same group are close to each other; that is, the normalized LCEs are considerably below 1 as long as strong community structures exist.

Several issues remain to be addressed in future studies. First, we defined LCE to quantify the continuity of group labels for a given vertex sequence. The consistency between ordering and clustering can also be measured in other ways; for example, one can quantify the continuity of indices in a vertex sequence for given group labels on the vertices, whereas the LCE quantifies the continuity of group labels for a given vertex sequence. Second, we focused on unipartite graphs for which the connectivities are represented by square

matrices (i.e., adjacency matrices). In principle, the proposed method can also be applied to study non-square matrices, such as bipartite graphs. Third, we implemented ordering and clustering methods independently and examined their consistency. Given that we found some consistency between the two, it would be possible to develop a clustering method that incorporates information about the inherent vertex sequence. Analogously, the spectral ordering method can be adjusted in such a way that the obtained vertex sequence reflects group labels. We expect our paper will stimulate further research in these directions.

APPENDIX A: UPPER BOUND OF THE LABEL CONTINUITY ERROR

We derive the upper bound of the LCE by explicitly constructing a worst-case sequence. We assume that a partition σ is given (i.e., the number of groups K and group sizes $\{N_k\}$ are given), and the first group is the largest group (i.e., $\max_k N_k = N_1 = |V_1|$). When N_1 satisfies $N_1 > \lceil N/2 \rceil$, some vertices in V_1 must be aligned consecutively. As exemplified in Fig. 11(a), the LCE is maximized when the vertices in V_1 and those in $\cup_{k>1} V_k$ are aligned alternately as possible. In this case, there are $2(N - N_1)$ vertices that are aligned alternately with different group labels, and the label continuity is $\mathcal{C} = (2N_1 - N - 1)/(N - 1)$. Therefore, the maximum LCE

leads to

$$
\Delta = 1 - \frac{K-1}{N-1} - \frac{2N_1 - N - 1}{N-1}
$$
$$
= \frac{2(N - N_1)}{N-1} - \frac{K-1}{N-1}, \tag{A1}
$$

which corresponds to the upper case of Eq. (5).

When N_1 is less than or equal to the sum of the vertices in all other groups [Figs. 11(b), 11(c), and 11(d)], vertices can be aligned such that no group labels are consecutive. Such a sequence is constructed as follows. We first align the vertices in V_1 and the vertices in $\cup_{k>1} V_k$ as alternately as possible. In this step, all the vertices in V_1 are aligned, and there are $\sum_{k>1} N_k - N_1$ vertices that are not yet aligned; here, in $\cup_{k>1} V_k$, we preferentially consume the labels with larger N_k [Fig. 11(b) and Step 1 in Figs. 11(c) and 11(d)]. When there are remaining vertices, we regard a set of alternately-aligned vertices as a fundamental unit and treat all such sets as "super vertices" with the same labels. We then align the super vertices and the remaining vertices in the same manner as in the previous step. We repeat this procedure until all vertices are aligned. We can always align vertices and super vertices alternately because the number of remaining vertices with the same label never exceeds the number of already aligned vertices or super vertices. Therefore, we can establish a sequence for which the label continuity $\mathcal{C}$ is zero, and the upper bound of the LCE leads to

$$
\Delta = 1 - \frac{K-1}{N-1}. \tag{A2}
$$

APPENDIX B: VARIANCE OF THE LABEL CONTINUITY ERROR IN RANDOM PARTITIONS

The second moment of Δ is

$$
\mathbb{E}[\Delta^2] = \sum_{\sigma^*} P(\sigma^*) \left(\frac{N-K}{N-1} - \frac{\sum_{i=1}^{N-1} \delta(\sigma^*(i), \sigma^*(i+1))}{N-1} \right)^2
$$

$$
= \left(\frac{N-K}{N-1} \right)^2 - 2\frac{N-K}{N-1}\mathbb{E}[\mathcal{C}] + \frac{1}{(N-1)^2} \sum_{i=1}^{N-1} \sum_{\sigma_i^*, \sigma_{i+1}^*} P(\sigma_i^*)P(\sigma_{i+1}^*)\delta(\sigma^*(i), \sigma^*(i+1))
$$

$$
+ \frac{2}{(N-1)^2} \sum_{i=1}^{N-2} \sum_{\sigma_i^*, \sigma_{i+1}^*, \sigma_{i+2}^*} P(\sigma_i^*)P(\sigma_{i+1}^*)P(\sigma_{i+2}^*)\delta(\sigma^*(i), \sigma^*(i+1))\delta(\sigma^*(i+1), \sigma^*(i+2))
$$

$$
+ \frac{1}{(N-1)^2} \sum_{\substack{i,j \\ (|i-j|>2)}} \sum_{\substack{\sigma_i^*, \sigma_{i+1}^*, \\ \sigma_j^*, \sigma_{j+1}^*}} P(\sigma_i^*)P(\sigma_{i+1}^*)P(\sigma_j^*)P(\sigma_{j+1}^*)\delta(\sigma^*(i), \sigma^*(i+1))\delta(\sigma^*(j), \sigma^*(j+1)) \tag{B1}
$$

$$
= \left(\frac{N-K}{N-1} \right)^2 - 2\frac{N-K}{N-1}\sum_{k=1}^{K} \left(\frac{N_k}{N} \right)^2 + \frac{1}{N-1}\sum_{k=1}^{K} \left(\frac{N_k}{N} \right)^2
$$

$$
+ \frac{2(N-2)}{(N-1)^2}\sum_{k=1}^{K} \left(\frac{N_k}{N} \right)^3 + \frac{(N-2)(N-3)}{(N-1)^2}\left(\sum_{k=1}^{K} \left(\frac{N_k}{N} \right)^2 \right)^2. \tag{B2}
$$

Thus, the variance $\mathrm{Var}[\Delta]$ is

$$\mathrm{Var}[\Delta] = \mathbb{E}[\Delta^2] - \mathbb{E}[\Delta]^2$$

$$= \frac{1}{N-1} \sum_{k=1}^{K} \left(\frac{N_k}{N}\right)^2 + \frac{2(N-2)}{(N-1)^2} \sum_{k=1}^{K} \left(\frac{N_k}{N}\right)^3 - \frac{3N-5}{(N-1)^2} \left(\sum_{k=1}^{K} \left(\frac{N_k}{N}\right)^2\right)^2. \tag{B3}$$

APPENDIX C: PROBABILITY DISTRIBUTION OF THE LABEL CONTINUITY ERROR IN RANDOM PARTITIONS

This section derives the probability distribution of the label continuity error $\Delta(\boldsymbol{\pi}, \boldsymbol{\sigma})$ when group labels are assigned randomly based on bootstrapped group labels $\boldsymbol{\sigma}^*$.

To derive the probability distribution of $\Delta(\boldsymbol{\pi}, \boldsymbol{\sigma})$, it is sufficient to calculate that of label continuity $\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma})$. The probability of $(N-1)\mathcal{C} = m$ is

$$P[(N-1)\mathcal{C} = m]$$

$$= \sum_{\boldsymbol{\sigma}^*} P(\boldsymbol{\sigma}^*) \delta\left(m, \sum_{i=1}^{N-1} \delta(\sigma^*(i), \sigma^*(i+1))\right)$$

$$= \sum_{\boldsymbol{\sigma}^*} \prod_{i=1}^{N} \frac{N_{\sigma^*(i)}}{N} \oint \frac{dz}{2\pi i} z^{\sum_{i=1}^{N-1} \delta(\sigma^*(i), \sigma^*(i+1)) - m - 1}$$

$$= \oint \frac{dz}{2\pi i} z^{-(1+m)} \sum_{\boldsymbol{\sigma}^*} \frac{N_{\sigma^*(N)}}{N} \prod_{i=1}^{N-1} \left(\frac{N_{\sigma^*(i)}}{N} z^{\delta(\sigma^*(i), \sigma^*(i+1))}\right)$$

$$= \oint \frac{dz}{2\pi i} z^{-(1+m)} \mathbf{1}^\top D(FD)^{N-1} \mathbf{1}, \tag{C1}$$

where

$$D \equiv \mathrm{diag}\left(\frac{N_1}{N}, \dots, \frac{N_K}{N}\right), \quad F \equiv \mathbf{1}\mathbf{1}^\top + (z-1)I. \tag{C2}$$

Here, I is the identity matrix. In Eq. (C1), we used the identity

$$\delta(x, y) = \oint \frac{dz}{2\pi i} \frac{1}{z^{x-y+1}}, \tag{C3}$$

which is an integral around the origin of the complex plane.

Using the eigenvalue decomposition, F can be expressed as

$$\frac{z-1+K}{K} \mathbf{1}\mathbf{1}^\top + (z-1) \sum_{k=2}^{K} \boldsymbol{u}_k \boldsymbol{u}_k^\top, \tag{C4}$$

where $\boldsymbol{u}_k$ ($2 \leqslant k \leqslant K$) is an eigenvector of F that is perpendicular to $\mathbf{1}$, and we have

$$FD\mathbf{1} = \frac{z-1+K}{K} \mathbf{1} + (z-1) \sum_{k=2}^{K} \boldsymbol{u}_k \boldsymbol{u}_k^\top D\mathbf{1}. \tag{C5}$$

Because the second term in Eq. (C5) vanishes when the group sizes are equal, the exact probability distribution can be derived as follows:

$$P[(N-1)\mathcal{C} = m]$$

$$= \frac{1}{K^{N-1}} \oint \frac{dz}{2\pi i} z^{-(1+m)} (z-1+K)^{N-1}$$

$$= \frac{1}{K^{N-1}} \oint \frac{dz}{2\pi i} z^{-(1+m)} \sum_{k=0}^{N-1} \binom{N-1}{k} z^k (K-1)^{N-1-k}$$

$$= \binom{N-1}{m} \left(\frac{1}{K}\right)^m \left(1 - \frac{1}{K}\right)^{N-1-m}. \tag{C6}$$

Equivalently,

$$P(\mathcal{C}) = \binom{N-1}{(N-1)\mathcal{C}} \left(\frac{1}{K}\right)^{(N-1)\mathcal{C}} \left(1 - \frac{1}{K}\right)^{(N-1)(1-\mathcal{C})}. \tag{C7}$$

Therefore, $(N-1)\mathcal{C}$ follows a binomial distribution. This result can be interpreted as follows. We suppose that there are N elements that are linearly aligned, and we assign group labels from one end. As we focus only on the consecutive property of the group labels, the label of the first element can be arbitrary. For the next $N-1$ elements, the probability that the label is consecutive to the previous one is $1/K$, whereas the complement probability is $1 - 1/K$ because the group label can be arbitrary as long as it is not identical to the previous one. We sum over all possible patterns that have consecutive labels m times to obtain $P[\mathcal{C} = \frac{m}{N-1}]$.

Even when the group sizes are not equal, Eq. (C6) is close to the actual distribution as long as the second term in Eq. (C5) is negligible. When $N \gg 1$ and the size of each group is of constant order, i.e., $N/K = O(1)$, Eq. (C6) is well approximated as a Poisson distribution. Furthermore, when $N/K \gg 1$, the distribution is nearly normal. The distribution of $\Delta(\boldsymbol{\pi}, \boldsymbol{\sigma})$ is obtained by shifting the distribution (C7) by a constant factor.

APPENDIX D: LABEL CONTINUITY ERRORS FOR NESTED PARTITIONS

As an example of partitions with different numbers of groups, here we investigate the difference in the LCEs between a partition $\boldsymbol{\sigma}$ with K groups and its nested partition $\boldsymbol{\sigma}'$. The partition $\boldsymbol{\sigma}'$ is obtained by subpartitioning the vertices V_K having Kth group label in $\boldsymbol{\sigma}$ into $V_{K,1}$ and $V_{K,2}$ ($V_{K,1} \cup V_{K,2} = V_K$); we denote the sizes of these two groups as $N'_{K,1}$ and $N'_{K,2}$ ($N'_{K,1} + N'_{K,2} = N_K$), and also denote $\{N'_k\} = \{N_1, \dots, N_{K-1}, N'_{K,1}, N'_{K,2}\}$. The partitions $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ are only locally different. The difference in the LCEs for $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ with the same sequence $\boldsymbol{\pi}$ is bounded as

$$-\frac{1}{N-1} \leqslant \Delta(\boldsymbol{\pi}, \boldsymbol{\sigma}') - \Delta(\boldsymbol{\pi}, \boldsymbol{\sigma})$$

$$\leqslant \frac{2\min\{N'_{K,1}, N'_{K,2}\} - \delta(N'_{K,1}, N'_{K,2}) - 1}{N-1}. \tag{D1}$$
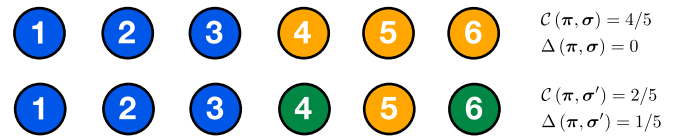
The lower bound is trivial because the label continuity $\mathcal{C}$ is a nonnegative quantity and $\mathcal{C}$ cannot be smaller when $\mathcal{C} = 0$ before subpartition. The upper bound of Eq. (D1) can be

derived as follows. The difference in the LCE is maximized when the difference in $\mathcal{C}$ is maximized. Note that the number of flips of the labels can be maximized when the labels before the subpartition are aligned completely consecutively, e.g., the case in Fig. 12(a). In this case, we can maximize the difference in $\mathcal{C}$ by aligning the vertices in $V_{K,1}$ and $V_{K,2}$ as alternately as possible. The achieved difference is

$$
\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma'}) - \mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma}) = \begin{cases} -\frac{N_K - 1}{N - 1} & (N'_{K,1} = N'_{K,2}) \\ -\frac{2\min\{N'_{K,1}, N'_{K,2}\}}{N - 1} & (N'_{K,1} \neq N'_{K,2}) \end{cases}
$$
$$
= \frac{\delta(N'_{K,1}, N'_{K,2}) - 2\min\{N'_{K,1}, N'_{K,2}\}}{N - 1}.
$$
(D2)

Equation (D1) indicates that the LCE is a local quantity, that is, the bound of variation in the LCE is characterized by $N'_{K,1}$ and $N'_{K,2}$; the variation tends to be small when $\min\{N'_{K,1}, N'_{K,2}\}$ is small. However, when it comes to the specific difference, not bounds, it depends not only on the subsequence within V_K, but also on the position V_K in the entire sequence $\boldsymbol{\pi}$ (see Fig. 12 for specific examples). The present result implies that comparison of the LCEs is generally complicated when partitions have different numbers of groups.

(a) Subpartition of vertices that changes the label continuity

$\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma}) = 4/5$
$\Delta(\boldsymbol{\pi}, \boldsymbol{\sigma}) = 0$

$\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\sigma'}) = 2/5$
$\Delta(\boldsymbol{\pi}, \boldsymbol{\sigma'}) = 1/5$

(b) Subpartition of vertices that does not change the label continuity

$\mathcal{C}(\tilde{\boldsymbol{\pi}}, \boldsymbol{\sigma}) = 0$
$\Delta(\tilde{\boldsymbol{\pi}}, \boldsymbol{\sigma}) = 4/5$

$\mathcal{C}(\tilde{\boldsymbol{\pi}}, \boldsymbol{\sigma'}) = 0$
$\Delta(\tilde{\boldsymbol{\pi}}, \boldsymbol{\sigma'}) = 3/5$

FIG. 12. Effect of subpartitioning on the label continuity $\mathcal{C}$ and label continuity error Δ. The partition $\boldsymbol{\sigma'}$ is a nested partition of $\boldsymbol{\sigma}$; the yellow label in $\boldsymbol{\sigma}$ is subpartitioned into the yellow and green labels in $\boldsymbol{\sigma'}$. (a) When the group labels are maximally consecutive (sequence $\boldsymbol{\pi}$), $\mathcal{C}$ becomes smaller by subpartitioning. (b) When the group labels are not consecutive at all (sequence $\tilde{\boldsymbol{\pi}}$), $\mathcal{C}$ does not change regardless of the choice of the nested partition. Although $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma'}$ are the same between (a) and (b), the value of $\mathcal{C}$ is affected by the locations of the blue labels.

[1] S. E. Schaeffer, Graph clustering, Comput. Sci. Rev. **1**, 27 (2007).

[2] S. Fortunato, Community detection in graphs, Phys. Rep. **486**, 75 (2010).

[3] S. Fortunato and D. Hric, Community detection in networks: A user guide, Phys. Rep. **659**, 1 (2016).

[4] A. Clauset, C. Moore, and M. E. J. Newman, Hierarchical structure and the prediction of missing links in networks, Nature (London) **453**, 98 (2008).

[5] T. P. Peixoto, Hierarchical Block Structures and High-Resolution Model Selection in Large Networks, Phys. Rev. X **4**, 011047 (2014).

[6] M. P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, Core-periphery structure in networks, SIAM J. Appl. Math. **74**, 167 (2014).

[7] M. S. Mariani, Z.-M. Ren, J. Bascompte, and C. J. Tessone, Nestedness in complex networks: Observation, emergence, and implications, Phys. Rep. **813**, 1 (2019).

[8] I. Liiv, Seriation and matrix reordering methods: An historical overview, Stat. Anal. Data Min. **3**, 70 (2010).

[9] M. Behrisch, B. Bach, N. H. Riche, T. Schreck, and J.-D. Fekete, Matrix reordering methods for table and network visualization, Comput. Graph Forum **35**, 693 (2016).

[10] T. Kawamoto and T. Kobayashi, Sequential locality of graphs and its hypothesis testing, arXiv:2111.11267.

[11] L. H. Harper, Optimal assignments of numbers to vertices, SIAM J. Appl. Math. **12**, 131 (1964).

[12] F. R. K. Chung, On optimal linear arrangements of trees, Comput. Math. with Appl. **10**, 43 (1984).

[13] H. Seitz, Contributions to the minimum linear arrangement problem, PhD. thesis, Ruprecht-Karls-Universität, Heidelberg, 2010.

[14] S. T. Barnard, A. Pothen, and H. Simon, A spectral algorithm for envelope reduction of sparse matrices, Numer. Linear Algebra Appl. **2**, 317 (1995).

[15] C. Ding and X. He, Linearized cluster assignment via spectral ordering, in *Proceedings of the Twenty-First International Conference on Machine Learning, ICML '04* (ACM, New York, 2004), p. 30.

[16] https://app.dimensions.ai/discover/publication?search_mode=content.

[17] L. Hagen and A. B. Kahng, New spectral methods for ratio cut partitioning and clustering, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **11**, 1074 (1992).

[18] J. Shi and J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Machine Intell. **22**, 888 (2000).

[19] U. von Luxburg, A tutorial on spectral clustering, Stat. Comput. **17**, 395 (2007).

[20] D. Arthur and S. Vassilvitskii, K-means++: The advantages of careful seeding, In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07* (SIAM, Philadelphia, 2007), pp. 1027–1035.

[21] M. E. J. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E **74**, 036104 (2006).

[22] M. E. J. Newman, Modularity and community structure in networks, Proc. Natl. Acad. Sci. U.S.A. **103**, 8577 (2006).

[23] A. Saade, F. Krzakala, and L. Zdeborová, Spectral clustering of graphs with the Bethe Hessian, in *Advances in Neural Information Processing Systems*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (Curran Associates, New York, 2014), Vol. 27.

[24] L. Dall'Amico, R. Couillet, and N. Tremblay, Revisiting the Bethe-Hessian: Improved community detection in sparse

heterogeneous graphs, in *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett (Curran Associates, New York, 2019), Vol. 32.

[25] K. Chaudhuri, F. Chung, and A. Tsiatas, Spectral clustering of graphs with general degrees in the extended planted partition model, in *Proceedings of the 25th Annual Conference on Learning Theory*, edited by S. Mannor, N. Srebro, and R. C. Williamson (PMLR, Edinburgh, Scotland, 2012), Vol. 23, pp. 35.1–35.23.

[26] A. A. Amini, A. Chen, P. J. Bickel, and E. Levina, Pseudo-likelihood methods for community detection in large sparse networks, Ann. Statist. **41**, 2097 (2013).

[27] T. Qin and K. Rohe, Regularized spectral clustering under the degree-corrected stochastic blockmodel, in *Advances in Neural Information Processing Systems*, edited by in C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Curran Associates, New York, 2013), Vol. 26.

[28] A. Joseph and B. Yu, Impact of regularization on spectral clustering, Ann. Statist. **44**, 1765 (2016).

[29] Y. Zhang and K. Rohe, Understanding regularized spectral clustering via graph conductance, in *Advances in Neural Information Processing System*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, New York, 2018), Vol. 31.

[30] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.023006 for the results of additional experiments and details of the real-world datasets.

[31] P. W. Holland, K. B. Laskey, and S. Leinhardt, Stochastic blockmodels: First steps, Soc. Networks **5**, 109 (1983).

[32] Y. J. Wang and G. Y. Wong, Stochastic blockmodels for directed graphs, J. Am. Stat. Assoc. **82**, 8 (1987).

[33] T. P. Peixoto, Entropy of stochastic blockmodel ensembles, Phys. Rev. E **85**, 056122 (2012).

[34] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi, A survey of statistical network models, Found. Trends Mach. Learn. **2**, 129 (2010).

[35] T. P. Peixoto, *Bayesian Stochastic Blockmodeling* (John Wiley & Sons, Hoboken, NJ, 2019), Chap. 11, pp. 289–332.

[36] E. Abbe, Community detection and stochastic block models: Recent developments, J. Mach. Learn. Res. **18**, 1 (2018).

[37] A. Condon and R. M. Karp, Algorithms for graph partitioning on the planted partition model, Random Struct. Algorithms **18**, 116 (2001).

[38] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Inference and Phase Transitions in the Detection of Modules in Sparse Networks, Phys. Rev. Lett. **107**, 065701 (2011).

[39] E. Mossel, J. Neeman, and A. Sly, Reconstruction and estimation in the planted partition model, Probab. Theory Relat. Fields **162**, 431 (2015).

[40] L. Massoulié, Community detection thresholds and the weak Ramanujan property, in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing, STOC '14* (ACM, New York, 2014), pp. 694–703.

[41] R. R. Nadakuditi and M. E. J. Newman, Graph Spectra and the Detectability of Community Structure in Networks, Phys. Rev. Lett. **108**, 188701 (2012).

[42] T. Kawamoto and Y. Kabashima, Limitations in the spectral method for graph partitioning: Detectability threshold and localization of eigenvectors, Phys. Rev. E **91**, 062803 (2015).

[43] T. Kawamoto and Y. Kabashima, Detectability of the spectral method for sparse graph partitioning, Europhys. Lett. **112**, 40007 (2015).

[44] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, Comparing community structure identification, J. Stat. Mech. (2005) P09008.

[45] R. K. Darst, Z. Nussinov, and S. Fortunato, Improving the performance of algorithms to find communities in networks, Phys. Rev. E **89**, 032809 (2014).

[46] Z. Yang, R. Algesheimer, and C. J. Tessone, A comparative analysis of community detection algorithms on artificial networks, Sci. Rep. **6**, 1 (2016).

[47] U. Von Luxburg, M. Belkin, and O. Bousquet, Consistency of spectral clustering, Ann. Statist. **36**, 555 (2008).

[48] W. W. Zachary, An information flow model for conflict and fission in small groups, J. Anthropol. Res. **33**, 452 (1977).

[49] V. Krebs, The political books network, http://www-personal.umich.edu/~mejn/netdata/.