Higher-Order Spectral Clustering for Geometric Graphs

Konstantin Avrachenkov, Andrei Bobu, Maximilien Dreveton

HAL Id: hal-03169834

https://hal.inria.fr/hal-03169834

Submitted on 15 Mar 2021

Higher-Order Spectral Clustering for Geometric Graphs

Konstantin Avrachenkov[1] · Andrei Bobu[1] · Maximilien Dreveton[1]

Abstract

The present paper is devoted to clustering geometric graphs. While the standard spectral clustering is often not effective for geometric graphs, we present an effective generalization, which we call higher-order spectral clustering. It resembles in concept the classical spectral clustering method but uses for partitioning the eigenvector associated with a higher-order eigenvalue. We establish the weak consistency of this algorithm for a wide class of geometric graphs which we call Soft Geometric Block Model. A small adjustment of the algorithm provides strong consistency. We also show that our method is effective in numerical experiments even for graphs of modest size.

Keywords Spectral clustering · Random geometric graphs · Block models

1 Introduction

Graph clustering—the task of identifying groups of tightly connected nodes in a graph—is a widely studied unsupervised learning problem, with applications in computer science, statistics, biology, economy or social sciences [7].

In particular, spectral clustering is one of the key graph clustering methods [15]. In its most basic form, this algorithm consists in partitioning a graph into two communities using the eigenvector associated with the second smallest eigenvalue of the graph's Laplacian matrix (the so-called Fiedler vector [6]). Spectral clustering is popular, as it is an efficient relaxation of the NP-hard problem of cutting the graph into two balanced clusters so that the weight between the two clusters is minimal [15].

✉ Konstantin Avrachenkov
    k.avrachenkov@inria.fr

    Andrei Bobu
    a.v.bobu@gmail.com

    Maximilien Dreveton
    maximilien.dreveton@inria.fr

[1] Inria Sophia-Antipolis, 2004 Route des Lucioles, 06902 Valbonne, France

In particular, spectral clustering is consistent in the Stochastic Block Model (SBM) for a large set of parameters [1,11]. The SBM is a natural basic model with community structure. It is also the most studied one [1]. In this model each node is assigned to one cluster, and edges between node pairs are drawn independently and with probability depending only on the community assignment of the edge endpoints.

However, in many situations, nodes also have geometric attributes (a position in a metric space). Thus, the interaction between a pair of nodes depends not only on the community labelling, but also on the distance between the two nodes. We can model this by assigning to each node a position, chosen in a metric space. Then, the probability of an edge appearance between two nodes will depend both on the community labelling and on the positions of these nodes. Recent proposals of random geometric graphs with community structure include the Geometric Block Model (GBM) [8] and Euclidean random geometric graphs [2]. The nodes' interactions in geometric models are no longer independent: two interacting nodes are likely to have many common neighbors. While this is more realistic ('friends of my friends are my friends'), this also renders the theoretical study more challenging.

Albeit spectral clustering was shown to be consistent in some specific geometric graphs [13], the geometric structure can also heavily handicap a cut-based approach. Indeed, one could partition space into regions such that nodes between two different regions interact very sparsely. Thus, the Fiedler vector of a geometric graph might be associated only with a geometric configuration, and bear no information about the latent community labelling. Moreover, the common technique of regularization [18], which aims to penalize small size communities in order to bring back the vector associated with the community structure in the second position, will not work in geometric graphs as the regions of space can contain a balanced number of nodes. Nonetheless, this observation does not automatically render spectral clustering useless. Indeed, as we shall see, in some situations there is still one eigenvector associated with the community labelling. Thus, it is now necessary to distinguish the eigenvectors corresponding to a geometric cut—hence potentially useless for cluster recovery— from the one corresponding to the community labelling. In other words, to achieve a good performance with spectral clustering in such a setting, one needs to select carefully the correct eigenvector, which may no longer be associated with the second smallest eigenvalue.

Our working model of geometric graphs with clustering structure will be the Soft Geometric Block Model (SGBM). It is a block generalization of soft random geometric graphs and includes as particular cases the SBM and the GBM. Another important example is the Waxman Block Model (WBM) where the edge probabilities decrease exponentially with the distance. The SGBM is similar to the model of [2], but importantly we do not assume the knowledge of nodes' positions.

In this paper, we propose a generalization of standard spectral clustering based on a higher-order eigenvector of the adjacency matrix. This eigenvector is selected using the average intra- and inter-community degrees, and is not necessarily the Fiedler vector. The goal of the present work is to show that this algorithm performs well both theoretically and practically on SGBM graphs.

Our specific contributions are as follows. We establish weak consistency of higher-order spectral clustering on the SGBM in the dense regime, where the average degrees

are proportional to the number of nodes. With a simple additional step, we also establish strong consistency. One important ingredient of the proof is the characterization of the spectrum of the clustered geometric graphs, and can be of independent interest. In particular, it shows that the limiting spectral measure can be expressed in terms of the Fourier transform of the connectivity probability functions. Additionally, our numerical simulations show that our method is effective and efficient even for graphs of modest size. Besides, we also illustrate by a numerical example the unsuitability of the Fiedler vector for community recovery in some situations.

Let us describe the structure of the paper. We introduce in Section 2 the Soft Geometric Block Model and the main notations. The characterization of the limiting spectrum is given in Section 3. This characterization will be used in Section 4 to establish the consistency of higher-order spectral clustering in dense SGBM graphs. Finally, Section 5 shows numerical results and Section 6 concludes the paper with a number of interesting future research directions.

2 Model Definition and Notations

2.1 Notations

Let $\mathbf{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ be the flat unit torus in dimension d represented by $\left[-\frac{1}{2}, \frac{1}{2}\right]^d$. The norm ℓ_∞ in $\mathbb{R}^d$ naturally induces a norm on $\mathbf{T}^d$ such that for any vector $x = (x_1, \ldots, x_d) \in \mathbf{T}^d$ we have $\|x\| = \max_{1 \leq i \leq d} |x_i|$.

For a measurable function $F : \mathbf{T}^d \to \mathbb{R}$ and $k \in \mathbb{Z}^d$, we denote $\widehat{F}(k) = \int_{\mathbf{T}^d} F(x) e^{-2i\pi \langle k, x \rangle} \, dx$ the Fourier transform of F. The Fourier series of F is given by

$$\sum_{k \in \mathbb{Z}^d} \widehat{F}(k) e^{2i\pi \langle k, x \rangle}.$$

For two integrable functions $F, G : \mathbf{T}^d \to \mathbb{R}$, we define the convolution operation $F * G(y) = \int_{\mathbf{T}^d} F(y - x) G(x) \, dx$ and $F^{*m} = F * F * \cdots * F$ (m times). We recall that $\widehat{F * G}(k) = \widehat{F}(k)\widehat{G}(k)$.

2.2 Soft Geometric Block Model

A Soft Geometric Block Model (SGBM) is defined by a dimension d, a number of nodes n, a set of blocks K and a connectivity probability function $F : \mathbf{T}^d \times K \times K \to [0, 1]$. The node set is taken as $V = [n]$. The model is parametrized by a *node labelling* $\sigma : V \to K$ and *nodes' positions* $X = (X_1, \ldots, X_n) \in \left(\mathbf{T}^d\right)^n$. We suppose that $F(\cdot, \sigma, \sigma') = F(\cdot, \sigma', \sigma)$ and for any $X \in \mathbf{T}^d$, $F(X)$ depends only on the norm $\|X\|$. The probability of appearance of an edge between nodes i and j is defined by $F\left(X_i - X_j, \sigma_i, \sigma_j\right)$. Note that this probability depends only on the distance between X_i and X_j and the labels σ_i, σ_j. Consequently, the model parameters specify the

distribution

$$\mathbb{P}_{\sigma,X}(A) = \prod_{1 \le i < j \le n} \left(F\left(X_i - X_j, \sigma_i, \sigma_j\right) \right)^{A_{ij}} \left(1 - F\left(X_i - X_j, \sigma_i, \sigma_j\right) \right)^{1-A_{ij}} \quad (1)$$

of the adjacency matrix $A = (A_{ij})$ of a random graph.

Furthermore, for this work we assume that the model has only two equal size blocks, i.e., $K = \{1, 2\}$, and $\sum_{i=1}^{n} 1(\sigma_i = 1) = \sum_{i=1}^{n} 1(\sigma_i = 2) = \frac{n}{2}$. The labels are assigned randomly, that is, the set $\{i \in [n]: \sigma_i = 1\}$ is chosen randomly over all the $\frac{n}{2}$-subsets of $[n]$. We assume that the entries of X and σ are independent and $\forall i \in V$, X_i is uniformly distributed over $\mathbf{T}^d$. Finally, suppose that for any $x \in \mathbf{T}^d$

$$F(x, \sigma, \sigma') = \begin{cases} F_{\text{in}}(x), & \text{if } \sigma = \sigma', \\ F_{\text{out}}(x), & \text{otherwise,} \end{cases} \quad (2)$$

where $F_{\text{in}}, F_{\text{out}} : \mathbf{T}^d \to [0, 1]$ are two measurable functions. We call these functions *connectivity probability functions*.

The average intra- and inter-community *edge densities* are denoted by μ_{in} and μ_{out}, respectively. Their expressions are given by the first Fourier modes of F_{in} and F_{out}:

$$\mu_{\text{in}} = \int_{\mathbf{T}^d} F_{\text{in}}(x)dx \quad \text{and} \quad \mu_{\text{out}} = \int_{\mathbf{T}^d} F_{\text{out}}(x)dx.$$

These quantities will play an important role in the following, as they represent the intensities of interactions between nodes in the same community and nodes in different communities. In particular, the average inside community degree is $\left(\frac{n}{2} - 1\right)\mu_{\text{in}}$, and the average outside community degree is $\frac{n}{2}\mu_{\text{out}}$.

Example 1 An SGBM where $F_{\text{in}}(x) = p_{\text{in}}$ and $F_{\text{out}}(x) = p_{\text{out}}$ with $p_{\text{in}}, p_{\text{out}}$ being constants is an instance of the Stochastic Block Model.

Example 2 An SGBM where $F_{\text{in}}(x) = 1(\|x\| \le r_{\text{in}})$, $F_{\text{out}}(x) = 1(\|x\| \le r_{\text{out}})$ with $r_{\text{in}}, r_{\text{out}} \in \mathbb{R}_+$ is an instance of the Geometric Block Model introduced in [8].

Example 3 We call Waxman Block Model (WBM) an SGBM with $F_{\text{in}}(x) = \min(1, q_{\text{in}}e^{-s_{\text{in}}\|x\|})$, $F_{\text{out}}(x) = \min(1, q_{\text{out}}e^{-s_{\text{out}}\|x\|})$. This is a clustered version of the Waxman model [16], which is a particular case of soft geometric random graphs [12].

Formally, *clustering* or *community recovery problem* is the following problem: given the observation of the adjacency matrix A and the knowledge of $F_{\text{in}}, F_{\text{out}}$, we want to recover the latent community labelling σ. Given an estimator $\widehat{\sigma}$ of σ, we define the loss ℓ as the ratio of misclassified nodes, up to a global permutation π of the labels: $\ell(\sigma, \widehat{\sigma}) = \frac{1}{n} \min_{\pi \in \mathcal{S}_2} \sum_i 1(\sigma_i \ne \pi \circ \widehat{\sigma}_i)$. Then, $\widehat{\sigma}$ is said to be weakly consistent (or equivalently, achieves almost exact recovery) if

$$\forall \epsilon > 0 : \lim_{n \to \infty} \mathbb{P}(\ell(\sigma, \widehat{\sigma}) > \epsilon) = 0,$$

and strongly consistent (equivalently, achieves exact recovery) if

$$\lim_{n\to\infty} \mathbb{P}\left(\ell\left(\sigma, \widehat{\sigma}\right) > 0\right) = 0.$$

3 The Analysis of Limiting Spectrum

3.1 Limit of the Spectral Measure

Theorem 1 *Consider an SGBM defined by* (1)-(2). *Assume that* $F_{\mathrm{in}}(0)$, $F_{\mathrm{out}}(0)$ *are equal to the Fourier series of* $F_{\mathrm{in}}(\cdot)$, $F_{\mathrm{out}}(\cdot)$ *evaluated at 0. Let* $\lambda_1, \ldots, \lambda_n$ *be the eigenvalues of A, and*

$$\mu_n = \sum_{i=1}^{n} \delta_{\lambda_i/n}$$

be the spectral measure of the matrix $\frac{1}{n}A$. *Then, for all Borel sets B with* $\mu(\partial B) = 0$ *and* $0 \notin \bar{B}$, *a.s.,*

$$\lim_{n\to\infty} \mu_n(B) = \mu(B),$$

where μ *is the following measure:*

$$\mu = \sum_{k\in\mathbb{Z}^d} \delta_{\frac{\widehat{F}_{\mathrm{in}}(k)+\widehat{F}_{\mathrm{out}}(k)}{2}} + \delta_{\frac{\widehat{F}_{\mathrm{in}}(k)-\widehat{F}_{\mathrm{out}}(k)}{2}}.$$

Remark 1 The limiting measure μ is composed of two terms. The first term, $\sum_{k\in\mathbb{Z}^d} \delta_{\frac{\widehat{F}_{\mathrm{in}}(k)+\widehat{F}_{\mathrm{out}}(k)}{2}}$ corresponds to the spectrum of a random graph with no community structure, and where edges between two nodes at distance x are drawn with probability $\frac{F_{\mathrm{in}}(x)+F_{\mathrm{out}}(x)}{2}$. In other words, it is the null-model of the considered SGBM. Hence, the eigenvectors associated with those eigenvalues bear no community information, but only geometric features.

On the contrary, the second term $\sum_{k\in\mathbb{Z}^d} \delta_{\frac{\widehat{F}_{\mathrm{in}}(k)-\widehat{F}_{\mathrm{out}}(k)}{2}}$ corresponds to the difference between intra- and inter-community edges. In particular, as we shall see later, the ideal eigenvector for clustering is associated with the eigenvalue $\widetilde{\lambda}$ closest to $\lambda_* = n\frac{\widehat{F}_{\mathrm{in}}(0)-\widehat{F}_{\mathrm{out}}(0)}{2}$. Other eigenvectors might mix some geometric and community features and hence are harder to analyze.

Last, the eigenvalue $\widetilde{\lambda}$ is not necessarily the second largest eigenvalue, as the ordering of eigenvalues here depends on the Fourier coefficients $\widehat{F}_{\mathrm{in}}(k)$ and $\widehat{F}_{\mathrm{out}}(k)$, and is in general non trivial.

Remark 2 The assumptions on $F_{\mathrm{in}}(0)$ and $F_{\mathrm{out}}(0)$ are validated for a wide range of reasonable connectivity functions. For instance, by Dini's criterion, all the functions that are differentiable at 0 satisfy these conditions. Another appropriate class consists

of piecewise C^1 functions that are continuous at 0 (this follows from the Dirichlet conditions).

Proof The outline of the proof of Theorem 1 follows closely [4]. First, we show that $\forall m \in \mathbb{N}, \lim_{n \to \infty} \mathbb{E} \mu_n(P_m) = \mu(P_m)$ where $P_m(t) = t^m$. Second, we use Talagrand's concentration inequality to prove that $\mu_n(P_m)$ is not far from its mean, and conclude with Borel–Cantelli lemma.

(i) By Lemma 1 in the Appendix, in order to establish the desired convergence it is enough to show that $\lim_{n \to \infty} \mathbb{E} \mu_n(P_m) = \mu(P_m)$ for any $m \in \mathbb{N}$. First,

$$\mathbb{E} \mu_n(P_m) = \frac{1}{n^m} \sum_{i=1}^{n} \mathbb{E} \lambda_i^m = \frac{1}{n^m} \mathbb{E} \operatorname{Tr} A^m. \tag{3}$$

By definition,

$$\operatorname{Tr} A^m = \sum_{\alpha \in [n]^m} \prod_{j=1}^{m} A_{i_j, i_{j+1}},$$

with $\alpha = (i_1, \ldots, i_m) \in [n]^m$ and $i_{m+1} = i_1$. We denote $\mathcal{A}_n^m$ the set of m-permutations of $[n]$, that is $\alpha \in \mathcal{A}_n^m$ iff α is an m-tuple without repetition. We have,

$$\operatorname{Tr} A^m = \sum_{\alpha \in \mathcal{A}_n^m} \prod_{j=1}^{m} A_{i_j, i_{j+1}} + R_m, \tag{4}$$

where $R_m = \sum_{\alpha \in [n]^m \setminus \mathcal{A}_n^m} \prod_{j=1}^{m} A_{i_j, i_{j+1}}$.

We first bound the quantity R_m. Since $|A_{ij}| \leq 1$, we have

$$|R_m| \leq \left| [n]^m \setminus \mathcal{A}_n^m \right| = n^m - \frac{n!}{(n-m)!} = \frac{m(m-1)}{2} n^{m-1} + o(n^{m-1}),$$

where we used $\frac{n!}{(n-m)!} = n^m - n^{m-1} \sum_{i=0}^{m-1} i + o(n^{m-1})$. Hence

$$\lim_{n \to \infty} \frac{1}{n^m} R_m = 0. \tag{5}$$

Moreover,

$$\mathbb{E} \sum_{\alpha \in \mathcal{A}_n^m} \prod_{j=1}^{m} A_{i_j, i_{j+1}} = \sum_{\alpha \in \mathcal{A}_n^m} \int_{(\mathbf{T}^d)^m} \prod_{j=1}^{m} F\left(x_{i_j} - x_{i_{j+1}}, \sigma_{i_j}, \sigma_{i_{j+1}}\right) dx_{i_1} \ldots dx_{i_m}$$

$$= \sum_{\alpha \in \mathcal{A}_n^m} G(\alpha)$$

where $G(\alpha) = \int_{(\mathbf{T}^d)^m} \prod_{j=1}^m F(x_{i_j} - x_{i_{j+1}}, \sigma_{i_j}, \sigma_{i_{j+1}}) dx_{i_1} \dots dx_{i_m}$ for $\alpha \in \mathcal{A}_n^m$.

Let us first show that the value of $G(\alpha)$ depends only on the number of consecutive indices corresponding to the nodes from the same community. More precisely, let us define the set $\mathcal{S}(\alpha) = \{ j \in [m] : \sigma_{i_j} = \sigma_{i_{j+1}} \}$. Using Lemma 2 in the Appendix and the fact that the convolution is commutative, we have

$$G(\alpha) = F_{\text{in}}^{*|\mathcal{S}(\alpha)|} * F_{\text{out}}^{*(m-|\mathcal{S}(\alpha)|)}(0).$$

We introduce the following equivalence relationship in $\mathcal{A}_n^m$: $\alpha \sim \alpha'$ if $|\mathcal{S}(\alpha)| = |\mathcal{S}(\alpha')|$. We notice than $G(\cdot)$ is constant on each equivalence class, and equals to $F_{\text{in}}^{*p} * F_{\text{out}}^{*(m-p)}(0)$ for any $\alpha \in \mathcal{A}_n^m$ such that $|\mathcal{S}(\alpha)| = p$.

Then, let us calculate the cardinal of each equivalence class with $|\mathcal{S}(\alpha)| = p$. First of all, we choose the set $\mathcal{S}(\alpha)$ which can be done in $\binom{m}{p}$ ways if $m - p$ is even and cannot be done if $m - p$ is odd. The latter follows from the fact that p (the number of 'non-changes' in the consecutive community labels) has the same parity as m (the total number of indices) since $i_{m+1} = i_1$. The set $\mathcal{S}(\alpha)$ defines the community labels up to the flip of communities since $\sigma_{i_j} = \sigma_{i_{j+1}}$ for any $j \in \mathcal{S}(\alpha)$ and $\sigma_{i_j} \neq \sigma_{i_{j+1}}$ for $j \in [m] \setminus \mathcal{S}(\alpha)$.

Let $N_1(\alpha)$ be the number of indices i_j with $\sigma_{i_j} = 1$. Consider first the case $\sigma_{i_1} = 1$ and note that $N_1(\alpha)$ is totally defined by the set $\mathcal{S}(\alpha)$. There are $\frac{n}{2}$ possible choices for i_1. Now we have two possibilities. If $\sigma_{i_1} = \sigma_{i_2}$ then we have $\frac{n}{2} - 1$ possible choices for the index i_2 (since $\alpha \in \mathcal{A}_n^m$). Otherwise, if $\sigma_{i_1} \neq \sigma_{i_2}$ then the index i_2 can be chosen in $\frac{n}{2}$ ways. Resuming the above operation, we choose $N_1(\alpha)$ indices from the first community, and it can be done in $n/2(n/2 - 1) \dots (n/2 - N_1(\alpha))$ ways. The indices from the second community can be chosen in $n/2(n/2 - 1) \dots (n/2 - (m - N_1(\alpha)))$ ways. Thus in total the number of possible choices of indices is

$$\frac{n}{2} \left(\frac{n}{2} - 1 \right) \dots \left(\frac{n}{2} - N_1(\alpha) \right) \cdot \frac{n}{2} \left(\frac{n}{2} - 1 \right) \dots \left(\frac{n}{2} - (m - N_1(\alpha)) \right)$$
$$= \frac{n^m}{2^m} + O(n^{m-1}), \quad n \to \infty.$$

The same reasoning applies if $\sigma_{i_1} = 2$. Hence, when n goes to infinity, the cardinal of each equivalence class is

$$\left| \{ \alpha \in \mathcal{A}_n^m : |\mathcal{S}(\alpha)| = p \} \right| = \begin{cases} 0 & \text{if } m - p \text{ is odd,} \\ 2 \binom{m}{p} \frac{n^m}{2^m} + O\left(n^{m-1} \right) & \text{otherwise.} \end{cases}$$

This can be rewritten as

$$\left| \{ \alpha \in \mathcal{A}_n^m : |\mathcal{S}(\alpha)| = p \} \right| = \binom{m}{p} \left(1 + (-1)^{m-p} \right) \frac{n^m}{2^m} + O\left(n^{m-1} \right), \quad n \to \infty.$$

Hence,

$$\mathbb{E} \sum_{\alpha \in \mathcal{A}_n^m} \prod_{j=1}^m A_{i_j, i_{j+1}} = \sum_{p=0}^n \left| \{ \alpha \in \mathcal{A}_n^m : |\mathcal{S}(\alpha)| = p \} \right| F_{\text{in}}^{*p} * F_{\text{out}}^{*(m-p)}(0)$$

$$= \frac{n^m}{2^m} \sum_{p=0}^m \binom{m}{p} \left(1 + (-1)^{m-p} \right) F_{\text{in}}^{*p} * F_{\text{out}}^{*(m-p)}(0) + O\left(n^{m-1}\right)$$

$$= n^m \left(\left(\frac{F_{\text{in}} + F_{\text{out}}}{2} \right)^{*m}(0) + \left(\frac{F_{\text{in}} - F_{\text{out}}}{2} \right)^{*m}(0) \right) + O(n^{m-1}).$$

Therefore, equations (3), (4) and (5) give us:

$$\lim_{n \to \infty} \mathbb{E} \mu_n(P_m) = \left(\frac{F_{\text{in}} + F_{\text{out}}}{2} \right)^{*m}(0) + \left(\frac{F_{\text{in}} - F_{\text{out}}}{2} \right)^{*m}(0).$$

Finally, since $F_{\text{in}}, F_{\text{out}}$ are equal to their Fourier series at 0, and using $\widehat{F * G}(k) = \widehat{F}(k)\widehat{G}(k)$, we have

$$\lim_{n \to \infty} \mathbb{E} \mu_n(P_m) = \sum_{k \in \mathbb{Z}^d} \left(\frac{\widehat{F}_{\text{in}}(k) + \widehat{F}_{\text{out}}(k)}{2} \right)^m + \left(\frac{\widehat{F}_{\text{in}}(k) - \widehat{F}_{\text{out}}(k)}{2} \right)^m = \mu(P_m).$$

(6)

(ii) For each $m \geq 1$, and n fixed, we define

$$Q_m : SGBM(F_{\text{in}}, F_{\text{out}}) \longrightarrow \mathbb{R}$$
$$A \longmapsto \frac{1}{n^{m-1}} \operatorname{Tr} A^m$$

where $SGBM(F_{\text{in}}, F_{\text{out}})$ denotes the set of adjacency matrices of an SGBM random graph with connectivity functions $F_{\text{in}}, F_{\text{out}}$. Note that $Q_m(A) = n\mu_n(P_m)$.

Let $A, \widetilde{A}$ be two adjacency matrices. We denote the Hamming distance by $d_{\text{Ham}}(A, \widetilde{A}) = \sum_{i=1}^n \sum_{j=1}^n 1(A_{ij} \neq \widetilde{A}_{ij})$. Using Lemma 5 in the Appendix, we show that the function Q_m is (m/n)–Lipschitz for the Hamming distance:

$$\left| Q_m(A) - Q_m(\widetilde{A}) \right| \leq \frac{m}{n} d_{\text{Ham}}(A, \widetilde{A}).$$

(7)

Let M_m be the median of Q_m. Talagrand's concentration inequality [14, Proposition 2.1] states that

$$\mathbb{P}(|Q_m - M_m| > t) \leq 4 \exp \left(-\frac{n^2 t^2}{4m^2} \right),$$

(8)

which after integrating over all t gives

$$|n\mathbb{E}\mu_n(P_m) - M_m| \leq \mathbb{E}|Q_m(A) - M_m| \leq \frac{C_m}{n},$$

since $\mathbb{E}X = \int_0^\infty \mathbb{P}(X > t)dt$ for any positive random variable X. The constant C_m is equal to $8m \int_0^\infty e^{-u^2} du$.

Moreover,

$$
\begin{aligned}
n\,|\mu_n(P_m) - \mathbb{E}\mu_n(P_m)| &\leq |n\mu_n(P_m) - M_m| + |M_m - n\mathbb{E}\mu_n(P_m)| \\
&\leq |Q_m - M_m| + \frac{C_m}{n}.
\end{aligned}
$$

Let $s > 0$. Since C_m/n^2 goes to 0 when n goes to infinity, we can pick n large enough such that $s > \frac{C_m}{n^2}$. Thus, using again inequality (8), we have

$$
\begin{aligned}
\mathbb{P}\left(|\mu_n(P_m) - \mathbb{E}\mu_n(P_m)| > s\right) &\leq \mathbb{P}\left(\frac{1}{n}|Q_m - M_m| > s - \frac{C_m}{n^2}\right) \\
&\leq 4\exp\left(-\frac{n^4}{4m^2}\left(s - \frac{C_m}{n^2}\right)^2\right).
\end{aligned}
$$

However, by (6), $\lim_{n\to\infty} \mathbb{E}\mu_n(P_m) = \mu(P_m)$. Hence $\mu_n(P_m)$ converges in probability to $\mu(P_m)$. Let $s_n = \frac{1}{n^\kappa}$ with $\kappa > 0$, and

$$
\epsilon_n = 4\exp\left(-\frac{n^4}{4m^2}\left(s_n - \frac{C_m}{n^2}\right)^2\right).
$$

Since $\sum_{n\in\mathbb{N}} \epsilon_n < \infty$ when $\kappa < 2$, an application of Borel–Cantelli lemma shows that the convergence holds in fact almost surely. This concludes the proof. $\qquad\square$

3.2 Conditions for the Isolation of the Ideal Eigenvalue

As noticed in Remark 1, the ideal eigenvector for clustering is associated with the eigenvalue of the adjacency matrix A closest to $n\frac{\mu_{\text{in}} - \mu_{\text{out}}}{2}$ (recall that $\mu_{\text{in}} = \widehat{F}_{\text{in}}(0)$ and $\mu_{\text{out}} = \widehat{F}_{\text{out}}(0)$). The following proposition shows that this ideal eigenvalue is isolated from other eigenvalues under certain conditions.

Proposition 1 *Consider the adjacency matrix A of an SGBM defined by (1)-(2), and assume that:*

$$
\begin{aligned}
\widehat{F}_{\text{in}}(k) + \widehat{F}_{\text{out}}(k) &\neq \mu_{\text{in}} - \mu_{\text{out}}, \qquad &\forall k \in \mathbb{Z}^d, \qquad &(9) \\
\widehat{F}_{\text{in}}(k) - \widehat{F}_{\text{out}}(k) &\neq \mu_{\text{in}} - \mu_{\text{out}}, \qquad &\forall k \in \mathbb{Z}^d \setminus \{0\}, \qquad &(10)
\end{aligned}
$$

with $\mu_{\text{in}} \neq \mu_{\text{out}}$. Then, the eigenvalue of A closest to $n\frac{\mu_{\text{in}} - \mu_{\text{out}}}{2}$ is of multiplicity one. Moreover, there exists $\epsilon > 0$ such that for large enough n every other eigenvalue is at distance at least ϵn.

Proof Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A. Let $i^* \in \operatorname{argmin}_{i \in [n]} \left| \frac{\lambda_i}{n} - \frac{\mu_{\text{in}} - \mu_{\text{out}}}{2} \right|$. We shall show that there exists $\epsilon > 0$ such that for large enough n, we have for all $i \neq i^*$:

$$\left| \frac{\lambda_i}{n} - \frac{\mu_{\text{in}} - \mu_{\text{out}}}{2} \right| > \epsilon.$$

Due to condition (9), and the fact that

$$\lim_{|k| \to \infty} \left(\widehat{F}_{\text{in}}(k) + \widehat{F}_{\text{out}}(k) \right) = 0,$$

there is some fixed $\epsilon_1 > 0$ such that

$$\min_{k \in \mathbb{Z}^d} \left(\left| \frac{\widehat{F}_{\text{in}}(k) + \widehat{F}_{\text{out}}(k)}{2} - \frac{\mu_{\text{in}} - \mu_{\text{out}}}{2} \right| \right) > \epsilon_1.$$

Similarly, condition (10) ensures the existence of $\epsilon_2 > 0$ such that

$$\min_{k \in \mathbb{Z}^d \setminus \{0\}} \left(\left| \frac{\widehat{F}_{\text{in}}(k) - \widehat{F}_{\text{out}}(k)}{2} - \frac{\mu_{\text{in}} - \mu_{\text{out}}}{2} \right| \right) > \epsilon_2.$$

Denote $\epsilon_3 = \frac{|\mu_{\text{in}} - \mu_{\text{out}}|}{4}$. Let $\epsilon = \min(\epsilon_1, \epsilon_2, \epsilon_3)$, and consider the interval $B = \left[\frac{\mu_{\text{in}} - \mu_{\text{out}}}{2} - \epsilon, \frac{\mu_{\text{in}} - \mu_{\text{out}}}{2} + \epsilon \right]$. By Theorem 1, a.s.,

$$\lim_{n \to \infty} \mu_n(B) = \mu(B) = 1.$$

Therefore, for n large enough the only eigenvalue of A in the interval B is λ_{i^*}. □

The following proposition shows that conditions (9) and (10) of Proposition 1 are almost always verified for a GBM.

Proposition 2 *Consider the d-dimensional GBM model, where $F_{\text{in}}, F_{\text{out}}$ are 1-periodic, and defined on the flat torus $\mathbf{T}^d$ by $F_{\text{in}}(x) = 1(\|x\| \leq r_{\text{in}})$ and $F_{\text{out}}(x) = 1(\|x\| \leq r_{\text{out}})$, with $r_{\text{in}} > r_{\text{out}} > 0$. Denote by $\mathcal{B}_+$ and $\mathcal{B}_-$ the sets of parameters r_{in} and r_{out} defined by negation of conditions (9) and (10):*

$$\mathcal{B}^+ = \left\{ (r_{\text{in}}, r_{\text{out}}) \in \mathbb{R}_+^2 : \widehat{F}_{\text{in}}(k) + \widehat{F}_{\text{out}}(k) = \mu_{\text{in}} - \mu_{\text{out}} \text{ for some } k \in \mathbb{Z}^d \right\}$$

$$\mathcal{B}^- = \left\{ (r_{\text{in}}, r_{\text{out}}) \in \mathbb{R}_+^2 : \widehat{F}_{\text{in}}(k) - \widehat{F}_{\text{out}}(k) = \mu_{\text{in}} - \mu_{\text{out}} \text{ for some } k \in \mathbb{Z}^d \setminus \{0\} \right\}.$$

Then these sets of 'bad' parameters are of zero Lebesgue measure:

$$\operatorname{Leb}\left(\mathcal{B}^+\right) = 0; \quad \text{and} \quad \operatorname{Leb}\left(\mathcal{B}^-\right) = 0.$$

Hence for $\mathcal{B} = \mathcal{B}^+ \cup \mathcal{B}^-$

$$\text{Leb}(\mathcal{B}) = 0.$$

Proof It is clear that

$$\text{Leb}(\mathcal{B}) \leq \text{Leb}\left(\mathcal{B}^+\right) + \text{Leb}\left(\mathcal{B}^-\right).$$

Thus, it is enough to show that $\text{Leb}(\mathcal{B}^+) = 0$ and $\text{Leb}(\mathcal{B}^-) = 0$. We shall establish the first equality, and the second equality can be proved similarly.

By Lemma 3 in the Appendix, the condition (9) for given functions F_{in} and F_{out} is as follows:

$$r_{\text{in}}^d \prod_{j=1}^d \text{sinc}\left(2\pi r_{\text{in}} k_j\right) + r_{\text{out}}^d \prod_{j=1}^d \text{sinc}\left(2\pi r_{\text{out}} k_j\right) =$$
$$= r_{\text{in}}^d - r_{\text{out}}^d \text{ for some } k = (k_1, \ldots, k_d) \in \mathbb{Z}^d.$$

Notice that $\lim_{k_j \to \infty} \text{sinc}(2\pi r_{\text{in}} k_j) = 0$ and $\lim_{k_j \to \infty} \text{sinc}(2\pi r_{\text{out}} k_j) = 0$ while the right-hand side of the above equation is fixed. Therefore, this equation can hold only for k from a finite set $\mathcal{K}$. Let us fix some $k = (k_1, \ldots, k_d) \in \mathcal{K}$ and denote

$$\mathcal{B}_k^+ = \left\{ (r_{\text{in}}, r_{\text{out}}) \in \mathbb{R}_+^2 : r_{\text{in}}^d \prod_{j=1}^d \text{sinc}\left(2\pi r_{\text{in}} k_j\right) + r_{\text{out}}^d \prod_{j=1}^d \text{sinc}\left(2\pi r_{\text{out}} k_j\right) = r_{\text{in}}^d - r_{\text{out}}^d \right\}.$$

Let us now fix r_{in}, and consider the condition defining $\mathcal{B}_k^+$ as an equation on r_{out}. Define the functions

$$f_k(x) = x^d \left(1 + \prod_{j=1}^d \text{sinc}\left(2\pi x k_j\right) \right);$$
$$g_k(x) = x^d \left(1 - \prod_{j=1}^d \text{sinc}\left(2\pi x k_j\right) \right).$$

Then the mentioned equation takes the form

$$f_k(r_{\text{out}}) = g_k(r_{\text{in}}). \tag{11}$$

Consider the function $h_k : \mathbb{C} \to \mathbb{R}$:

$$h_k(z) = z^d \left(1 + \prod_{j=1}^d \text{sinc}\left(2\pi z k_j\right) \right).$$

Clearly, this function coincides with f_k on $\mathbb{R}$. Moreover, it is holomorphic in $\mathbb{C}$, as $\text{sinc}(z)$ is holomorphic (it can be represented by the series $\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n}$), and the product of holomorphic functions is again holomorphic. But then the derivative $h_k'(z)$ is also holomorphic, therefore, it has a countable number of zeros in $\mathbb{C}$. Clearly, $h_k' \equiv f_k'$ on $\mathbb{R}$, which yields that f_k' has a countable number of zeros in $\mathbb{R}$.

Hence, $\mathbb{R}_+$ is divided into a countable number of intervals on which the function $f_k(x)$ is strictly monotone. That is, $\mathbb{R}_+ = \sqcup_{i=0}^{\infty} [a_i(k), b_i(k)]$ where $f_{k,i} = f_k\big|_{[a_i(k), b_i(k)]}$ is strictly monotone. Then the function $f_{k,i}^{-1}(x)$ is correctly defined and, since $f_{k,i}$ is measurable and injective, $f_{k,i}^{-1}$ is measurable as well. Consequently, there is a unique solution $r_{\text{out}} = f_{k,i}^{-1}(g_k(r_{\text{in}}))$ of equation (11) for $r_{\text{in}} \in [\min f_{k,i}; \max f_{k,i}]$. If $r_{\text{in}} \notin [\min f_{k,i}; \max f_{k,i}]$, there is no solution at all.

Therefore, $\mathcal{B}_{k,i}^+ = \left\{ \left(r_{\text{in}}, f_{k,i}^{-1}(g_k(r_{\text{in}})) \right) : r_{\text{in}} \in [\min f_{k,i}; \max f_{k,i}] \right\}$ is the graph of some measurable function in $\mathbb{R}_+^2$. Since such a graph has a zero Lebesgue measure (see e.g., [17, Lemma 5.3]), we have:

$$\text{Leb}(\mathcal{B}_k^+) = \text{Leb}\left(\cup_{i=1}^{\infty} \mathcal{B}_{k,i}^+ \right) = 0.$$

Hence, we can conclude that

$$\text{Leb}(\mathcal{B}^+) = \text{Leb}\left(\bigcup_{k \in \mathcal{K}} \mathcal{B}_k^+ \right) \leq \sum_{k \in \mathcal{K}} \text{Leb}(\mathcal{B}_k^+) = 0.$$

Carrying out similar argumentation for $\mathcal{B}^-$ completes the proof. $\square$

4 Consistency of Higher-Order Spectral Clustering

In this section we show that spectral clustering based on the ideal eigenvector (see Algorithm 1) is weakly consistent for SGBM (Theorem 2). We then show that a simple extra step can in fact achieve strong consistency.

Algorithm 1: Higher-Order Spectral Clustering (HOSC).

Input: Adjacency matrix A, average intra- and inter-cluster edge densities $\mu_{\text{in}}, \mu_{\text{out}}$.
Output: Node labelling $\widetilde{\sigma} \in \{1, 2\}^n$.

Global step:
Let $\widetilde{\lambda}$ be the eigenvalue of A closest to $\lambda_* = \frac{(\mu_{\text{in}} - \mu_{\text{out}})}{2} n$, and $\widetilde{v}$ be the associated eigenvector.
for $i = 1, \ldots, n$ **do**
⌊ If $\widetilde{v}_i > 0$, let $\widetilde{\sigma}_i = 1$; otherwise, let $\widetilde{\sigma}_i = 2$.

Remark 3 The worst case complexity of the eigenvalue factorization is $O\left(n^3\right)$ [5]. However, when the matrix is sufficiently sparse and the eigenvalues are well separated, the empirical complexity can be close to $O(kn)$, where k is the number of required eigenvalues [5]. Moreover, since Algorithm 1 uses only the sign of eigenvector elements, a quite rough accuracy can be sufficient for classification purposes.

4.1 Weak Consistency of Higher-Order Spectral Clustering

Theorem 2 *Let us consider the d-dimensional SGBM with connectivity probability functions F_{in} and F_{out} satisfying conditions (9)-(10). Then Algorithm 1 is weakly consistent. More precisely, Algorithm 1 misclassifies at most $O\left(\log n\right)$ nodes.*

Proof Let us introduce some notations. Recall that $\mu_{\mathrm{in}} = \widehat{F}_{\mathrm{in}}(0)$ and $\mu_{\mathrm{out}} = \widehat{F}_{\mathrm{out}}(0)$. In the limiting spectrum, the ideal eigenvalue for clustering is

$$\lambda_* = \frac{\mu_{\mathrm{in}} - \mu_{\mathrm{out}}}{2} n.$$

We consider the closest eigenvalue of A to λ_*:

$$\widetilde{\lambda} = \underset{\lambda}{\mathrm{argmin}} \left(|\lambda - \lambda_*| \right).$$

Also, let $\widetilde{v}$ be the normalized eigenvector corresponding to $\widetilde{\lambda}$. In this proof, the Euclidean norm $\| \cdot \|_2$ is used.

The plan of the proof is as follows. We consider the vector

$$v_* = \left(\underbrace{1/\sqrt{n}, \ldots, 1/\sqrt{n}}_{n/2}, \underbrace{-1/\sqrt{n}, \ldots, -1/\sqrt{n}}_{n/2} \right)^{\mathrm{T}},$$

where we supposed without loss of generality that the $n/2$ first nodes are in Cluster 1, and the $n/2$ last nodes are in Cluster 2. The vector v_* gives the perfect recovery by the sign of its coordinates. We shall show that with high probability for some constant $C > 0$

$$\|\widetilde{v} - v_*\|_2 \leq C\sqrt{\frac{\log n}{n}}. \tag{12}$$

We say that an event occurs *with high probability* (*w. h. p.*) if its probability goes to 1 as $n \to \infty$. With the bounding (12), we shall then show that at most $o(n)$ of entries of $\widetilde{v}$ have a sign that differs from the sign of the respective entry in v_*; hence $\widetilde{v}$ retrieves almost exact recovery.

In order to establish inequality (12) we shall use the following theorem from [10].

Theorem 3 *Let A be a real symmetric matrix. If $\widetilde{\lambda}$ is the eigenvalue of A closest to $\rho(v) = \frac{v^T A v}{v^T v}$, δ is the separation of ρ from the next closest eigenvalue and $\widetilde{v}$ is the*

eigenvector corresponding to $\widetilde{\lambda}$*, then*

$$|\sin\angle(v,\widetilde{v})| \leq \frac{\|Av - \rho v\|_2}{\|v\|_2 \delta}.$$

First we deal with $\rho(v_*)$. Since v_* is normalized and real-valued (by the symmetry of A), we have

$$\rho(v_*) = v_*^T A v_*.$$

Denote $u = A v_*$. Then, obviously,

$$u_i = \sum_{j=1}^{n} A_{ij}(v_*)_i = \frac{1}{\sqrt{n}} \sum_{j=1}^{n/2} A_{ij} - \frac{1}{\sqrt{n}} \sum_{j=n/2+1}^{n} A_{ij}. \tag{13}$$

It is clear that each entry A_{ij} with $i \neq j$ is a Bernoulli random variable with the probability of success either μ_{in} or μ_{out}. This can be illustrated by the element-wise expectation of the adjacency matrix:

$$\mathbb{E}A = \begin{pmatrix} \mu_{\text{in}} \cdots \mu_{\text{in}} & \mu_{\text{out}} \cdots \mu_{\text{out}} \\ \vdots \ddots \vdots & \vdots \ddots \vdots \\ \mu_{\text{in}} \cdots \mu_{\text{in}} & \mu_{\text{out}} \cdots \mu_{\text{out}} \\ \mu_{\text{out}} \cdots \mu_{\text{out}} & \mu_{\text{in}} \cdots \mu_{\text{in}} \\ \vdots \ddots \vdots & \vdots \ddots \vdots \\ \mu_{\text{out}} \cdots \mu_{\text{out}} & \mu_{\text{in}} \cdots \mu_{\text{in}} \end{pmatrix}.$$

Let us consider the first term in the right-hand side of (13) for $i \leq n/2$. Since A_{ij} are independent for fixed i, it is easy to see that $Y_i := \sum_{j=1}^{n/2} A_{ij} \sim \text{Bin}(n/2 - 1, \mu_{\text{in}})$ with the expectation $\mathbb{E}Y_i = (n/2 - 1)\mu_{\text{in}}$. Then we can use the Chernoff bound to estimate a possible deviation from the mean. For any $0 < t < 1$

$$\mathbb{P}(|Y_i - \mathbb{E}Y_i| > t\mathbb{E}Y_i) \leq e^{-\mathbb{E}Y_i t^2/2}. \tag{14}$$

Let us take $t = \frac{2\sqrt{\log n}}{\sqrt{(n/2-1)\mu_{\text{in}}}}$. Then for large enough n,

$$\mathbb{P}\left(\left|\sum_{j=1}^{n/2} A_{ij} - \mu_{\text{in}}\frac{n}{2}\right| > \sqrt{2\mu_{\text{in}}n\log n}\right) = \mathbb{P}\left(|Y_i - \mathbb{E}Y_i| > \sqrt{2\mu_{\text{in}}n\log n}\right) \leq \frac{1}{n^2}.$$

Similarly,

$$\mathbb{P}\left(\left|\sum_{j=n/2+1}^{n} A_{ij} - \mu_{\text{out}}\frac{n}{2}\right| > \sqrt{2\mu_{\text{out}}n\log n}\right) \leq \frac{1}{n^2}$$

and

$$\mathbb{P}\left(\left|u_i - (\mu_{\text{in}} - \mu_{\text{out}})\frac{\sqrt{n}}{2}\right| > \sqrt{2(\mu_{\text{in}} + \mu_{\text{out}})\log n}\right) \leq \frac{2}{n^2}. \qquad (15)$$

Denote $\gamma_n = \sqrt{2(\mu_{\text{in}} + \mu_{\text{out}})\log n}$ and notice that $\gamma_n = \Theta(\sqrt{\log n})$. By the union bound, we have for large enough n

$$\mathbb{P}\left(\exists i \leq \frac{n}{2} : \left|u_i - (\mu_{\text{in}} - \mu_{\text{out}})\frac{\sqrt{n}}{2}\right| > \gamma_n\right) \leq \frac{n}{2} \cdot \frac{2}{n^2} = \frac{1}{n}. \qquad (16)$$

By the same argumentation,

$$\mathbb{P}\left(\exists i > \frac{n}{2} : \left|u_i - (\mu_{\text{out}} - \mu_{\text{in}})\frac{\sqrt{n}}{2}\right| > \gamma_n\right) \leq \frac{1}{n}. \qquad (17)$$

Now let us calculate $\rho(v_*)$:

$$\rho(v_*) = \sum_{i=1}^{n}(v_*)_i \, u_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n/2} u_i - \frac{1}{\sqrt{n}}\sum_{i=n/2+1}^{n} u_i.$$

We already established that $u_i \sim (\mu_{\text{in}} - \mu_{\text{out}})\frac{\sqrt{n}}{2}$ for $i \leq \frac{n}{2}$ (which means that $\lim \frac{2u_i}{(\mu_{in} - \mu_{out})\sqrt{n}} = 1$ w.h.p.) and, therefore, that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n/2} u_i \sim (\mu_{\text{in}} - \mu_{\text{out}})\frac{n}{4}$. More precisely, by (16),

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n/2} u_i - \frac{(\mu_{\text{in}} - \mu_{\text{out}})n}{4}\right| > \frac{\gamma_n\sqrt{n}}{2}\right) \leq \frac{1}{n}.$$

In the same way, by (17),

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=\frac{n}{2}+1}^{n} u_i - \frac{(\mu_{\text{out}} - \mu_{\text{in}})n}{4}\right| > \frac{\gamma_n\sqrt{n}}{2}\right) \leq \frac{1}{n}.$$

Finally,

$$\mathbb{P}\left(\left|\rho(v_*) - \frac{(\mu_{\text{in}} - \mu_{\text{out}})n}{2}\right| > \gamma_n\sqrt{n}\right) \leq \frac{2}{n}. \qquad (18)$$

Now let us denote $w = Av_* - \rho(v_*)v_* = u - \rho(v_*)v_*$. As we already know, $u_i \sim (\mu_{\text{in}} - \mu_{\text{out}})\frac{\sqrt{n}}{2}$ and $(\rho(v_*)v_*)_i \sim (\mu_{\text{in}} - \mu_{\text{out}})\frac{\sqrt{n}}{2}$ for $i \leq \frac{n}{2}$. Clearly, for $i \leq \frac{n}{2}$

$$|w_i| \leq \left|u_i - \frac{(\mu_{\text{in}} - \mu_{\text{out}})\sqrt{n}}{2}\right| + \left|\frac{(\mu_{\text{in}} - \mu_{\text{out}})\sqrt{n}}{2} - \frac{1}{\sqrt{n}}\rho(v_*)\right|.$$

Then

$$\mathbb{P}\left(|w_i| > \gamma_n\right) \le \mathbb{P}\left(\left|u_i - \frac{(\mu_{\text{in}} - \mu_{\text{out}})\sqrt{n}}{2}\right| > \gamma_n\right) +$$

$$+ \mathbb{P}\left(\left|\frac{(\mu_{\text{in}} - \mu_{\text{out}})\sqrt{n}}{2} - \frac{1}{\sqrt{n}}\rho(v_*)\right| > \gamma_n\right).$$

A similar bound can be derived for the case $i > n/2$. Taking into account that $\rho(v_*)$ does not depend on i, using the union bound and equations (15) and (18), we get that

$$\mathbb{P}\left(\max_i |w_i| > 2\gamma_n\right) \le n \cdot \frac{2}{n^2} + \frac{2}{n} = \frac{4}{n}.$$

One can readily see that $\|w\|_2 \le \sqrt{n \cdot \max_i w_i^2} = \sqrt{n}\max_i |w_i|$. Thus, we finally can bound the Euclidean norm of the vector w:

$$\mathbb{P}\left(\|w\|_2 > 2\gamma_n\sqrt{n}\right) \le \frac{4}{n} \to 0, \quad n \to \infty.$$

Now we can use Theorem 3. According to this result,

$$|\sin \angle(v_*, \widetilde{v})| \le \frac{\|Av_* - v_*\rho(v_*)\|_2}{\|v_*\|_2\delta} = \frac{\|w\|_2}{\delta} \le \frac{2\gamma_n\sqrt{n}}{\delta} \quad w.h.p.,$$

where $\delta = \min_i |\lambda_i(A) - \rho(v_*)|$ over all $\lambda_i \ne \widetilde{\lambda}$. Since we have assumed that (9) and (10) hold, by Proposition 1, $\delta > \epsilon n$. Then, since v_* is normalized, a simple geometric consideration guarantees that

$$\|v_* - \widetilde{v}\|_2 \le \sqrt{2}|\sin \angle(v_*, \widetilde{v})| \le \frac{2\sqrt{2}\gamma_n\sqrt{n}}{\epsilon n} = \frac{2\sqrt{2}\gamma_n}{\epsilon\sqrt{n}} \quad w.h.p. \quad (19)$$

Let us denote *the number of errors* by

$$r = \left|\left\{i \in [n]: \text{sign}\left((v_*)_i\right) \ne \text{sign}\left(\widetilde{v}_i\right)\right\}\right|.$$

If we now remember that the vector v_* consists of $\pm\frac{1}{\sqrt{n}}$, it is clear that for any i with $\text{sign}((v_*)_i) \ne \text{sign}(\widetilde{v}_i)$

$$\left|(v_*)_i - \hat{v}_i\right| > \frac{1}{\sqrt{n}}.$$

The number of such coordinates is r. Therefore,

$$\|v_* - \widetilde{v}\|_2^2 \ge r\left(\frac{1}{\sqrt{n}}\right)^2 = \frac{r}{n}.$$

Then, by (19), the following chain of inequalities holds:

$$\frac{r}{n} \;\leq\; \|v_* - \widetilde{v}\|_2^2 \;\leq\; \frac{8\gamma_n^2}{\epsilon^2 n} \;=\; \frac{16(\mu_{\text{in}} + \mu_{\text{out}})\log n}{\epsilon^2 n} \quad w.\,h.\,p.$$

Hence, with high probability

$$r \;\leq\; \frac{16(\mu_{\text{in}} + \mu_{\text{out}})\log n}{\epsilon^2} = O(\log n), \quad n \to \infty.$$

Thus, the vector $\widetilde{v}$ provides almost exact recovery. This completes the proof. $\qquad\square$

4.2 Strong Consistency of Higher-Order Spectral Clustering with Local Improvement

In order to derive a strong consistency result, we shall add an extra step to Algorithm 1. Given $\widetilde{\sigma}$, the prediction of Algorithm 1, we classify each node to be in the community where it has the most neighbors, according to the labeling $\widetilde{\sigma}$. We summarize this procedure in Algorithm 2, and Theorem 4 states the exact recovery result.

Algorithm 2: Higher-Order Spectral Clustering with Local Improvement (HOSC-LI).

Input: Adjacency matrix A, average intra- and inter-cluster edge densities $\mu_{\text{in}}, \mu_{\text{out}}$.
Output: Node labelling $\widehat{\sigma} \in \{1, 2\}^n$.

Global step:
Let $\widetilde{\sigma}$ be the output of Algorithm 1.
Local improvement:
for $i = 1, \ldots, n$ **do**
$\quad$ Set $\widehat{\sigma}_i := \underset{k \in \{1,2\}}{\operatorname{argmax}} \sum_{j \neq i} \mathbb{1}\left(\widetilde{\sigma}_j = k\right) a_{ij}.$

Remark 4 The local improvement step runs in $O(n d_{\text{max}})$ operations, where d_{max} is the maximum degree of the graph. Albeit the local improvement step being convenient for the theoretical proof, we will see in Section 5 (Figure 3) that in practice Algorithm 1 already works well, often giving 100% accuracy even without local improvement.

Theorem 4 *Let us consider the d-dimensional SGBM defined by* (1)-(2), *and connectivity probability functions* F_{in} *and* F_{out} *satisfying conditions* (9)-(10). *Then Algorithm 2 provides exact recovery for the given SGBM.*

Proof We need to prove that the almost exact recovery of Algorithm 1 (established in Theorem 2) can be transformed into exact recovery by the local improvement step. This step consists in counting neighbours in the obtained communities. For each node

i we count the number of neighbours in both supposed communities determined by the sign of the vector $\widetilde{v}$ coordinate:

$$\widetilde{Z}_1(i) = \sum_{\text{sign}(\widetilde{v}_j)=1} A_{ij};$$

$$\widetilde{Z}_2(i) = \sum_{\text{sign}(\widetilde{v}_j)=-1} A_{ij}.$$

According to Algorithm 2, if $\widetilde{Z}_1(i) > \widetilde{Z}_2(i)$, we assign the label $\widehat{\sigma}_i = 1$ to node i, otherwise we label it as $\widehat{\sigma}_i = 2$. Suppose that some node i is still misclassified after this procedure and our prediction does not coincide with the true label: $\widehat{\sigma}_i \neq \sigma_i$. Let us assume without loss of generality that $\sigma_i = 1$ and, therefore, $\widehat{\sigma}_i = 2$. Then, clearly, $\widetilde{Z}_2(i) > \widetilde{Z}_1(i)$.

Let us denote by $Z_1(i)$ and $Z_2(i)$ degrees of node i in the communities defined by the true labels σ:

$$Z_j(i) = \sum_{\sigma_i=j} A_{ij}, \quad j = 1, 2.$$

Since $\text{sign}(\widetilde{v}_j)$ coincides with the true community partition for all but $C \log n$ nodes (see the end of the proof of Theorem 2),

$$\left| \widetilde{Z}_j(i) - Z_j(i) \right| \leq C \log n, \quad j = 1, 2,$$

which implies that

$$\widetilde{Z}_1(i) \geq Z_1(i) - C \log n;$$
$$\widetilde{Z}_2(i) \leq Z_2(i) + C \log n.$$

Hence, taking into account that $\widetilde{Z}_2(i) > \widetilde{Z}_1(i)$,

$$Z_2(i) + 2C \log n > Z_1(i),$$

which means that the inter-cluster degree of node i is asymptotically not less than its intra-cluster degree (since $Z_j(i) = \Theta(n) \ \ w.h.p.$). Intuitively, this should happen very seldom, and Lemma 4 in the Appendix gives an upper bound on the probability of this event. Thus, by Lemma 4, for large n,

$$\mathbb{P}\left(Z_2(i) + 2C \log n > Z_1(i) \right) = \mathbb{P}\left(Z_1(i) - Z_2(i) < 2C \log n \right) \leq$$
$$\leq \mathbb{P}\left(Z_1(i) - Z_2(i) < \sqrt{2 \left(\mu_{\text{in}} + \mu_{\text{out}} \right) n \log n} \right) \leq \frac{1}{n} \to 0, \quad n \to \infty.$$

Then each node is classified correctly with high probability and Theorem 4 is proved. $\square$

Fig. 1 Accuracy obtained on a 1-dimensional GBM ($n = 2000$, $r_{in} = 0.08$, $r_{out} = 0.02$) using the different eigenvectors of the adjacency matrix. The eigenvector of index k corresponds to the eigenvector associated with the k-th largest eigenvalue of A
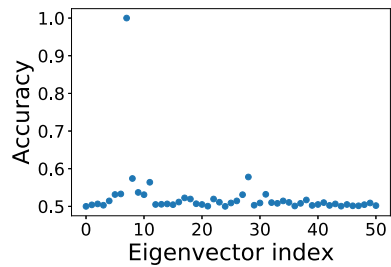
5 Numerical Results

5.1 Higher-Order Spectral Clustering on 1-Dimensional GBM

Let us consider a 1-dimensional GBM, defined in Example 2. We first emphasize two important points of the theoretical study: the ideal eigenvector for clustering is not necesarily the Fiedler vector, and some eigenvectors, including the Fiedler vector, could correspond to geometric configurations.

Figure 1 shows the accuracy (*i.e.,* the ratio of correctly labeled nodes, up to a global permutation of the labels if needed, divided by the total number of nodes) of each eigenvector for a realization of a 1-dimensional GBM. We see that, although the Fiedler vector is not suitable for clustering, there is nonetheless one eigevector that stands out of the crowd.

Then, in Figure 2 we draw the nodes of a GBM according to their respective position. We then show the clusters predicted by some eigenvectors. We see some geometric configurations (Figures 2a and 2c), while the eigenvector leading to the perfect accuracy corresponds to index 4 (Figure 2b).

Figure 3 shows the evolution of the accuracy of Algorithms 1 and 2 when the number of nodes n increases. As expected, the accuracy increases with n. Moreover, we see no significant effect of using the local improvement of Algorithm 2. Thus, we conduct all the rest of numerical experiments with the basic Algorithm 1.

In Figure 4, we illustrate the statement of Proposition 2: for some specific values of the pair (r_{in}, r_{out}), the Conditions (9) and (10) do not hold, and Algorithm 1 is not guaranteed to work. We observe in Figure 4 that these pairs of bad values exactly correspond to the moments when the index of the ideal eigenvector jumps from one value to another.

Finally, we compare in Figure 5 the accuracy of Algorithm 1 with the motif counting algorithms presented in references [8] and [9]. Those algorithms consist in counting the number of common neighbors, and clustering accordingly. We call Motif Counting 1 (resp. Motif Counting 2) the algorithm of reference [8] (resp. of reference [9]). We thank the authors for providing us the code used in their papers. We observed that with present realizations the motif counting algorithms take much more time than HOSC takes. For example on a GBM with $n = 3000$, $r_{in} = 0.08$ and $r_{out} = 0.04$, HOSC takes 8 seconds, while Motif Counting 1 takes 130 seconds and Motif Counting 2 takes 60 seconds on a laptop with 1.90GHz CPU and 15.5 GB memory.
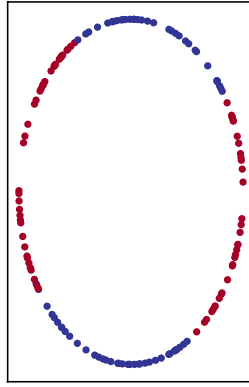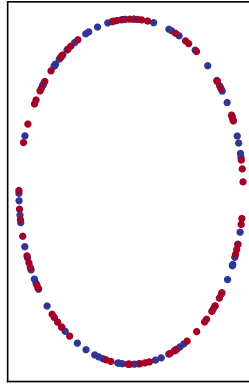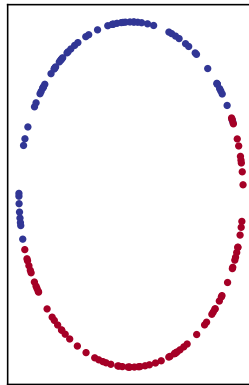
(a) $k = 2$ (b) $k = 4$ (c) $k = 8$

Fig. 2 Example of clustering done using the eigenvector associated to the k-th largest eigenvalue of the adjacency matrix of a 1-dimensional GBM ($n = 150$, $r_{in} = 0.2$, $r_{out} = 0.05$). For clarity edges are not shown, and nodes are positioned on a circle according to their true positions. The Fiedler vector ($k = 2$) is associated with a geometric cut, while the 4-th eigenvector corresponds to the true community labelling and leads to the perfect accuracy. The vector $k = 8$ is associated with yet another geometric cut (Color figure online)

Fig. 3 Accuracy obtained on 1-dimensional GBM as a function of n, when $r_{in} = 0.08$ and $r_{out} = 0.05$, for Algorithm 1 and Algorithm 2. Results are averaged over 100 realisations, and error bars show the standard error
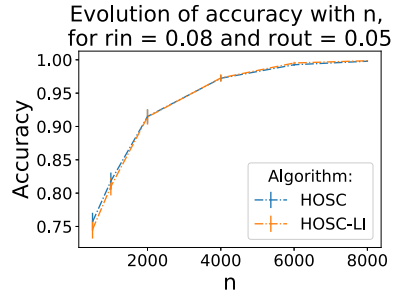

Fig. 4 Evolution of accuracy (blue curve) with respect to r_{in}, for a GBM with $n = 3000$ and $r_{out} = 0.06$. Results are averaged over 5 realisations. By the red curve we show the index of the ideal eigenvector, again with respect to r_{in} (Color figure online)
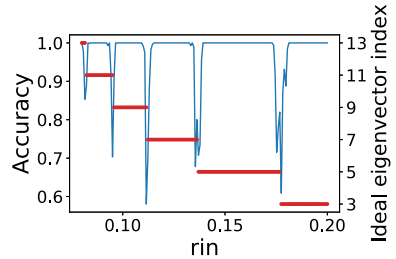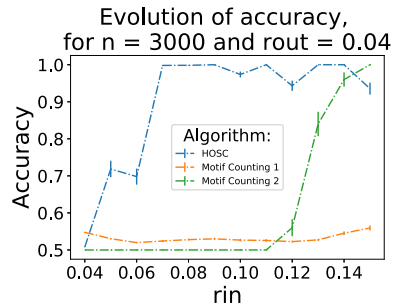

Fig. 5 Accuracy obtained on 1-dimensional GBM for different clustering methods. Motif Counting 1 corresponds to the algorithm described in [8] and Motif Counting 2 to the algorithm described in [9]. Results are averaged over 50 realisations, and error bars show the standard error (Color figure online)

5.2 Waxman Block Model

Let us now consider the Waxman Block Model introduced in Example 3. Recall that $F_{in}(x) = \min(1, q_{in}e^{-s_{in}x})$ and $F_{out}(x) = \min(1, q_{out}e^{-s_{out}x})$, where $q_{in}, q_{out}, s_{in}, s_{out}$ are four positive real numbers. We have the following particular situations:

- if $s_{out} = 0$, then $F_{out}(x) = q_{out}$ and the inter-cluster interactions are independent of the nodes' positions. If $s_{in} = 0$ as well, we recover the SBM;
- if $q_{in} = e^{r_{in}s_{in}}$ and $q_{out} = e^{r_{out}s_{out}}$, then in the limit $s_{in}, s_{out} \gg 1$ we recover the 1-dimensional GBM.

We show in Figure 6 the accuracy of Algorithm 1 on a WBM. In particular, we see that we do not need $\mu_{in} > \mu_{out}$, and we can recover disassociative communities. However, there are obvious dips when q_{in} is close to q_{out} or s_{in} is close to s_{out}. It is clear that if $q_{in} = q_{out}$ on the left-hand side picture or $s_{in} = s_{out}$ on the right-hand side picture, one cannot distinguish two communities in the graph. Thus, for small n, we
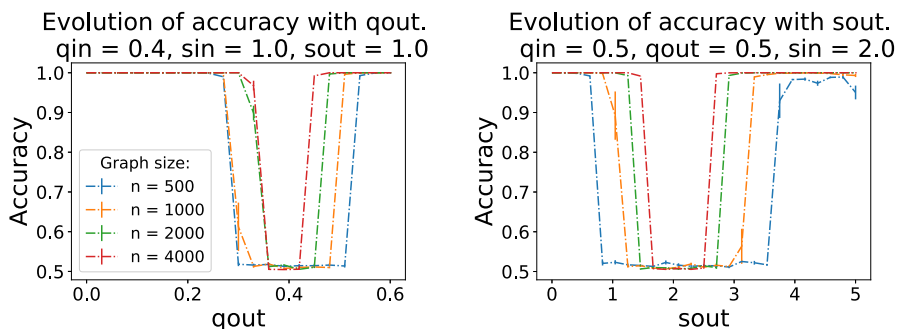
Fig. 6 Accuracy obtained on a 1-dimensional Waxman Block Model. Results are averaged over 10 realizations. Same colors in the two plots correspond to the same graph size

observe some ranges around these 'bad' values where Algorithm 1 fails. As expected, the dips become narrower when n increases.

6 Conclusions and Future Research

In the present paper we devised an effective algorithm for clustering geometric graphs. This algorithm is close in concept to the classical spectral clustering method but it makes use of the eigenvector associated with a higher-order eigenvalue. It provides weak consistency for a wide class of graph models which we call the Soft Geometric Block Model, under some mild conditions on the Fourier transform of F_{in} and F_{out}. A small adjustment of the algorithm leads to strong consistency. Moreover, our method was shown to be effective in numerical simulations even for graphs of modest size.

The problem stated in the current paper might be investigated further in several directions. One of them is a possible study on the SGBM with more than two clusters. The situation here is quite different from the SBM where the spectral clustering algorithm with few eigenvectors associated with the smallest non-zero eigenvalues provides good performance. In the SGBM even the choice of such eigenvectors is not trivial, since the 'optimal' eigenvalue for distinguishing two clusters is often not the smallest one.

Another natural direction of research is the investigation of the sparse regime, since all our theoretical results concern the case of degrees linear in n (this assumption is used for the analysis of the adjacency matrix spectrum and for finding the spectral gap around the 'ideal' eigenvalue $\tilde{\lambda}$). In sparser regimes, there are effective algorithms for some particular cases of the SGBM (e. g., for the GBM), but there is no established threshold when exact recovery is theoretically possible. Unfortunately, the method of the current paper without revision is not appropriate for this situation, and the technique will very likely be much more complicated.

Finally, considering weighted geometric graphs could be an interesting task for applications where clustering of weighted graphs is pertinent. For instance, the functions F_{in} and F_{out} can be considered as weights on the edges in a graph. We believe that most of the results of the paper may be easily transferred to this case.

A Auxiliary Results

A.1 Hamburger Moment Problem for the Limiting Measure

Lemma 1 *Assume that $F : \mathbf{T}^d \to \mathbb{R}$ is such that $F(0)$ is equal to the Fourier series of $F(x)$ evaluated at 0 and $0 \leq F(x) \leq 1$. Consider the measure μ defined by the function F:*

$$\mu = \sum_{k \in \mathbb{Z}^d} \delta_{\widehat{F}(k)}.$$

Then μ is defined uniquely by the sequence of its moments $\{M_n\}_{n=1}^{\infty}$.

Proof It is enough to show that Carleman's condition holds true for μ (see [3]):

$$\sum_{n=1}^{\infty} M_{2n}^{-\frac{1}{2n}} = +\infty. \tag{20}$$

As one can easily see,

$$M_{2n} = \sum_{k \in \mathbb{Z}^d} \widehat{F}^{2n}(k). \tag{21}$$

From the bounds $0 \leq F(x) \leq 1$ it follows that $\widehat{F}(k) \leq 1$. Then it is clear that $\widehat{F}^n(k) \leq \widehat{F}(k)$ for any $n \in \mathbb{N}$. We can write

$$M_{2n} = \sum_{k \in \mathbb{Z}^d} \widehat{F}^{2n}(k) \leq \sum_{k \in \mathbb{Z}^d} \widehat{F}(k) = F(0) \leq 1.$$

Here we used the assumption that $F(0)$ equals its Fourrier series evaluated at 0. Then

$$M_{2n}^{-\frac{1}{2n}} \geq 1,$$

Thus, the series in the right-hand side of (20) is divergent and Carleman's condition is verified. □

A.2 m-times Convolution

Lemma 2 *Let* $m \in \mathbb{N}$ *and* $F_1, \dots, F_m$ *be integrable functions over* $\mathbf{T}^d$. *Then,*

$$F_1 * \cdots * F_m(0) = \int_{(\mathbf{T}^d)^m} \prod_{j=1}^{m} F_j \left(z_j - z_{j+1} \right) \, dz_1 \dots dz_m$$

with the notation $z_{m+1} = z_1$.

Proof With the change of variable $u_i = z_i - z_{i+1}$ for $i = 1, \dots, m-1$, we have

$$\int_{(\mathbf{T}^d)^m} \prod_{j=1}^{m} F_j \left(z_j - z_{j+1} \right) \, dz_1 \dots dz_m$$

$$= \int_{\mathbf{T}^d} dz_1 \int_{(\mathbf{T}^d)^{m-1}} \prod_{i=1}^{m-1} F_i \left(u_i \right) F_m \left(-u_1 - \cdots - u_{m-1} \right) du_1 \dots du_{m-1}$$

We notice that

$$\int_{\mathbf{T}^d} du_{m-1} F_{m-1} \left(u_{m-1} \right) F_m \left(-u_1 - \cdots - u_{m-1} \right)$$

$$= F_{m-1} * F_m \left(-u_1 - \cdots - u_{m-2} \right).$$

Hence,

$$\int_{(\mathbf{T}^d)^{m-1}} \prod_{i=1}^{m-1} F_i \left(u_i \right) F_m \left(-u_1 - \cdots - u_{m-1} \right) du_1 \dots du_{m-1}$$

$$= F_1 * \cdots * F_m(0).$$

Therefore,

$$\int_{(\mathbf{T}^d)^m} \prod_{j=1}^{m} F_j \left(z_j - z_{j+1} \right) \, dz_1 \dots dz_m = \int_{\mathbf{T}^d} dz_1 F_1 * \cdots * F_m(0)$$

$$= F_1 * \cdots * F_m(0),$$

which ends the proof.   □

A.3 Fourier Transform of the Square Wave

Lemma 3 *Let $0 < r < \frac{1}{2}$. Let $F : \mathbb{R}^d \to \mathbb{R}$ be 1-periodic such that $F(x) = 1\,(\|x\| \le r)$ for $x \in \mathbf{T}^d$. Then,*

$$\widehat{F}(k) = 2r^d \prod_{j=1}^{d} \operatorname{sinc}\left(2\pi k_j r\right),$$

where $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$ and

$$\operatorname{sinc}(x) = \begin{cases} 1, & \text{if } x = 0, \\ \frac{\sin x}{x}, & \text{otherwise.} \end{cases}$$

Proof We shall use the set $[-1/2, 1/2]^d$ as a representation of $\mathbf{T}^d$. Let us first notice that for $x \in [-1/2, 1/2]^d$

$$F(x) = 1\,(\|x\| \le r) = 1\left(\max_{1 \le j \le d} |x_j| \le r\right) = \prod_{j=1}^{d} 1\,(|x_j| \le r).$$

Then

$$\widehat{F}(k) = \int_{\left[-\frac{1}{2}, \frac{1}{2}\right]^d} F(x) e^{-2\pi i \langle k, x\rangle} dx$$

$$= \int_{\left[-\frac{1}{2}, \frac{1}{2}\right]^d} \prod_{j=1}^{d} 1(|x_j| \le r) e^{-2\pi i k_j x_j} dx_1 \ldots dx_d$$

$$= \prod_{j=1}^{d} \int_{-1/2}^{1/2} 1(|x_j| \le r) e^{-2\pi i k_j x_j} dx_j.$$

Let us consider some $1 \le j \le d$. If $k_j = 0$, we have $\int_{-1/2}^{1/2} 1(|x_j| \le r) dx_j = \int_{-r}^{r} dx = 2r$. Moreover, for $k_j \ne 0$,

$$\widehat{F}(k) = \int_{-1/2}^{1/2} 1\left(|x_j| \le r\right) e^{-2\pi i k_j x_j} dx_j = \int_{-r}^{r} e^{-2\pi i k_j x_j} dx_j = \frac{e^{-2\pi i k_j r} - e^{2\pi i k_j r}}{-2\pi i k_j} =$$

$$= \frac{\sin(2\pi k_j r)}{\pi k_j} = 2r \frac{\sin\left(2\pi k_j r\right)}{2\pi k_j r} = 2r \operatorname{sinc}\left(2\pi k_j r\right).$$

Hence,

$$\widehat{F}(k) = 2r^d \prod_{j=1}^{d} \operatorname{sinc}(2\pi k_j r),$$

as was needed.   □

A.4 Number of Neighbours in Different Clusters

Lemma 4 *Let us consider the SGBM with connectivity probability functions F_{in} and F_{out} for which $\mu_{\text{in}} = \widehat{F}_{\text{in}}(0) > \widehat{F}_{\text{out}}(0) = \mu_{\text{out}}$. Denote by $Z_{\text{in}}(i)$ (resp., $Z_{\text{out}}(i)$) the 'intra-cluster' (resp., 'inter-cluster') degree of i:*

$$Z_{\text{in}}(i) = \sum_{j\,:\,\sigma_j=\sigma_i} A_{ij};$$

$$Z_{\text{out}}(i) = \sum_{j\,:\,\sigma_j\neq\sigma_i} A_{ij}.$$

Denote $\mathcal{B}_i := \left\{ Z_{\text{in}}(i) - Z_{\text{out}}(i) < \sqrt{2(\mu_{\text{in}} + \mu_{\text{out}})n \log n} \right\}$. Then

$$\mathbb{P}\left(\cup_{i=1}^{n}\mathcal{B}_i\right) \leq \frac{1}{n}.$$

Proof Let us fix $i \in [n]$. Clearly, $Z_{\text{in}}(i) \sim \text{Bin}(\frac{n}{2}-1, \mu_{\text{in}})$ and $Z_{\text{out}}(i) \sim \text{Bin}(\frac{n}{2}, \mu_{\text{out}})$. We again use Chernoff inequality (14). By this bound, for $t = \frac{2\sqrt{\log n}}{\sqrt{(n/2-1)\mu_{\text{in}}}}$ and large enough n

$$\mathbb{P}\left(\left|Z_{\text{in}}(i) - \mu_{\text{in}}\frac{n}{2}\right| > \sqrt{2\mu_{\text{in}}n\log n}\right) \leq \frac{1}{n^2}.$$

By the same reason, Z_{out} is well concentrated around its mean $\mu_{\text{out}}\frac{n}{2}$:

$$\mathbb{P}\left(\left|Z_{\text{out}}(i) - \mu_{\text{out}}\frac{n}{2}\right| > \sqrt{2\mu_{\text{out}}n\log n}\right) \leq \frac{1}{n^2}.$$

Therefore, since $\mu_{\text{in}} > \mu_{\text{out}}$,

$$\mathbb{P}\left(\mathcal{B}_i\right) = \mathbb{P}\left(Z_{\text{in}}(i) - Z_{\text{out}}(i) < \sqrt{2(\mu_{\text{in}} + \mu_{\text{out}})n\log n}\right) \leq \frac{1}{n^2}.$$

By the union bound,

$$\mathbb{P}(\mathcal{B}) \leq n\mathbb{P}\left(\mathcal{B}_1\right) \leq \frac{1}{n},$$

which proves the lemma. □

A.5 Trace Operator is Lipschitz

Lemma 5 *Let $A, \widetilde{A} \in \{0, 1\}^{n \times n}$ be two adjacency matrices, and $m \geq 1$. Then,*

$$\left| \operatorname{Tr} A^m - \operatorname{Tr} \widetilde{A}^m \right| \leq m \, n^{m-2} \mathrm{d}_{\mathrm{Ham}} \left(A, \widetilde{A} \right).$$

Proof Since A and $\widetilde{A}$ are adjacency matrices of graphs without self-loops, we have $\operatorname{Tr} A = 0 = \operatorname{Tr} \widetilde{A}$. Hence $\left| \operatorname{Tr} A - \operatorname{Tr} \widetilde{A} \right| = 0 \leq \frac{1}{n} \mathrm{d}_{\mathrm{Ham}} \left(A, \widetilde{A} \right)$, and the statement holds for $m = 1$.

Let us now consider $m \geq 2$. We have

$$\left| \operatorname{Tr} \left(A^{m-1} \right) - \operatorname{Tr} \left(\widetilde{A}^{m-1} \right) \right| = \left| \sum_{i_1, \ldots, i_m} \left(\prod_{j=1}^{m} A_{i_j i_{j+1}} - \prod_{j=1}^{m} \widetilde{A}_{i_j, i_{j+1}} \right) \right|$$

$$\leq \sum_{i_1, \ldots, i_m} \left| \prod_{j=1}^{m} A_{i_j i_{j+1}} - \prod_{j=1}^{m} \widetilde{A}_{i_j, i_{j+1}} \right|,$$

with the notation $i_{m+1} = i_1$. The quantity $\prod_{j=1}^{m} A_{i_j i_{j+1}}$ is equal to one if $A_{i_j i_{j+1}} = 1$ for all $j = 1, \ldots, m$, and equals zero otherwise. Hence, the difference $\left| \prod_{j=1}^{m} A_{i_j i_{j+1}} - \prod_{j=1}^{m} \widetilde{A}_{i_j, i_{j+1}} \right|$ is non-zero and is equal to one in two scenarii:

- $A_{i_j i_{j+1}} = 1$ for all $j = 1, \ldots, m$, while there is a j' such that $\widetilde{A}_{i_{j'} i_{j'+1}} = 0$,
- there is a j' such that $A_{i_{j'} i_{j'+1}} = 0$ and $\widetilde{A}_{i_j i_{j+1}} = 1$ for all $j = 1, \ldots, m$.

Thus,

$$\left| \prod_{j=1}^{m} A_{i_j i_{j+1}} - \prod_{j=1}^{m} \widetilde{A}_{i_j, i_{j+1}} \right| = 1 \left(\forall j \ A_{i_j i_{j+1}} = 1 \right) 1 \left(\exists j' : \widetilde{A}_{i_{j'} i_{j'+1}} = 0 \right) +$$

$$+ 1 \left(\exists j' : A_{i_{j'} i_{j'+1}} = 0 \right) 1 \left(\forall j \ \widetilde{A}_{i_j i_{j+1}} = 1 \right).$$

But,

$$1 \left(\forall j \ A_{i_j i_{j+1}} = 1 \right) 1 \left(\exists j' : \widetilde{A}_{i_{j'} i_{j'+1}} = 0 \right) \leq \prod_{j=1}^{m} 1 \left(A_{i_j i_{j+1}} = 1 \right) \sum_{j=1}^{m} 1 \left(\widetilde{A}_{i_j i_{j+1}} = 0 \right)$$

$$\leq \sum_{j=1}^{m} 1 \left(A_{i_j i_{j+1}} = 1 \right) 1 \left(\widetilde{A}_{i_j i_{j+1}} = 0 \right).$$

Similarly,

$$1 \left(\exists j' : A_{i_{j'} i_{j'+1}} = 0 \right) 1 \left(\forall j \ \widetilde{A}_{i_j i_{j+1}} = 1 \right) \leq \sum_{j=1}^{m} 1 (A_{i_j i_{j+1}} = 0) 1 (\widetilde{A}_{i_j i_{j+1}} = 1).$$

Therefore,

$$\left| \prod_{j=1}^{m} A_{i_j i_{j+1}} - \prod_{j=1}^{m} \widetilde{A}_{i_j, i_{j+1}} \right| \leq \sum_{j=1}^{m} \mathbb{1}\left(A_{i_j i_{j+1}} \neq \widetilde{A}_{i_j i_{j+1}} \right).$$

This leads to

$$\sum_{i_1,\ldots,i_m} \left| \prod_{j=1}^{m} A_{i_j i_{j+1}} - \prod_{j=1}^{m} \widetilde{A}_{i_j, i_{j+1}} \right| \leq \sum_{i_1,\ldots,i_m} \sum_{j=1}^{m} \mathbb{1}\left(A_{i_j i_{j+1}} \neq \widetilde{A}_{i_j i_{j+1}} \right)$$
$$\leq m \, n^{m-2} \, d_{\text{Ham}}(A, \widetilde{A}),$$

where the last line holds since for all $j = 1, \ldots, m$ and $m \geq 2$

$$\sum_{i_1,\ldots,i_m} \mathbb{1}\left(A_{i_j i_{j+1}} \neq \widetilde{A}_{i_j i_{j+1}} \right) = n^{m-2} \sum_{i_j, i_{j+1}} \mathbb{1}\left(A_{i_j i_{j+1}} \neq \widetilde{A}_{i_j i_{j+1}} \right)$$
$$= n^{m-2} d_{\text{Ham}}\left(A, \widetilde{A} \right).$$

$\square$

References

1. Abbe, E.: Community detection and stochastic block models: recent developments. J. Mach. Learn. Res. **18**(1), 6446–6531 (2017)
2. Abbe, E., Baccelli, F., Sankararaman, A.: Community detection on Euclidean random graphs. arXiv preprint arXiv:1706.09942v5 (2017)
3. Aheizer, N.I., Kemmer, N.: The Classical Moment Problem and Some Related Questions in Analysis. Oliver & Boyd, Edinburgh (1965)
4. Bordenave, C.: Eigenvalues of Euclidean random matrices. Random Struct. Algorithms **33**(4), 515–532 (2008)
5. Demmel, J.W., Marques, O.A., Parlett, B.N., Vömel, C.: Performance and accuracy of Lapack's symmetric tridiagonal eigensolvers. SIAM J. Sci. Comput. **30**(3), 1508–1526 (2008)
6. Fiedler, M.: A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. Czechoslov. Math. J. **25**(4), 619–633 (1975)
7. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010)
8. Galhotra, S., Mazumdar, A., Pal, S., Saha, B.: The geometric block model. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
9. Galhotra, S., Mazumdar, A., Pal, S., Saha, B.: Connectivity of random annulus graphs and the geometric block model. Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019) **145**, 53 (2019)
10. Kahan, W., Parlett, B., Jiang, E.: Residual bounds on approximate eigensystems of nonnormal matrices. SIAM J. Numer. Anal. **19**(3), 470–484 (1982)
11. Lei, J., Rinaldo, A.: Consistency of spectral clustering in stochastic block models. Ann. Stat. **43**(1), 215–237 (2015)
12. Penrose, M.D.: Connectivity of soft random geometric graphs. Ann. Appl. Probab. **26**(2), 986–1028 (2016)
13. Spielman, D.A., Teng, S.H.: Spectral partitioning works: planar graphs and finite element meshes. Linear Algebra Appl. **421**(2), 284–305 (2007)
14. Talagrand, M.: A new look at independence. Ann. Probab. **24**(1), 1–34 (1996)

15. Von Luxburg, U.: A tutorial on spectral clustering. Stat. Comput. **17**(4), 395–416 (2007)
16. Waxman, B.M.: Routing of multipoint connections. IEEE J. Sel. Areas Commun. **6**(9), 1617–1622 (1988)
17. Wheeden, R.L., Zygmund, A.: Measure and Integral: An Introduction to Real Analysis. Chapman & Hall, New York (1977)
18. Zhang, Y., Rohe, K.: Understanding regularized spectral clustering via graph conductance. In: Advances in Neural Information Processing Systems, pp. 10631–10640 (2018)