Machine Learning on Network-Valued Data: The Spectral Way

by

Daniel Ferguson

B.S., Seattle University, 2017

M.S., University of Colorado Boulder, 2021

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Applied Mathematics

2022

Committee Members:

François G. Meyer, Chair

Stephen Becker

Aaron Clauset

Sean O'Rourke

Juan Restrepo

Ferguson, Daniel (Ph.D., Applied Mathematics)

Machine Learning on Network-Valued Data:

The Spectral Way

Thesis directed by Prof. François G. Meyer

Abstract

This dissertation considers the problem of machine learning given network-valued data, sometimes referred to as graph-valued data. Many machine learning algorithms utilize various statistics to answer a wide variety of questions in a principled manner. When data comes from Euclidean space, determination of the mean and variance is a simple algebraic operation. For graph valued data, simple statistics are non-trivial to compute; they are generalized to metric spaces in [42] and are defined by an optimization procedure over the metric space.

Chapter 1 introduces the overarching questions addressed within this dissertation and the contributions made to the field. In Chapter 2, the computation of the sample mean given a data set of graphs is explored in depth when considering a metric that compares the eigenvalues of the adjacency matrices of two graphs. The work in this chapter is novel, to the best of our knowledge, this chapter introduces the first method to compute the mean of a data set of graphs with respect to spectral information. Within Chapter 3, both the mean and the variance are used to infer the parameters of a distribution on the space of graphs with respect to spectral information. Chapter 4 considers the theoretical properties of the sample mean graph and how those properties relate to the sample data set.

Acknowledgements

Firstly, I thank my advisor François G. Meyer for his support throughout my years at CU Boulder, his patience with this work, and his instruction on how to perform research. I also extend thanks to Stephen Becker, Aaron Clauset, Sean O'Rourke, and Juan Restrepo for serving on my committee.

My time in Boulder has been remarkable, largely thanks to the friends I have made and the people I have met throughout my Ph.D. In particular, my thanks go to Liam Madden, Eli Halperin, and Erik Johnson.

Lastly, I thank my family and my partner, Claire McAteer, for their love and support.

Parts of the work in this dissertation were supported by the National Science Foundation, CCF/CIF 1815971.

Contents

Chapter

List of Figures

Figure

Chapter 1

Introduction

The ubiquity of graphs to represent real-world objects and relationships such as social networks [65], biology networks [52], traffic networks [54] etc. has showcased a need for advancement in algorithms that analyze such objects. Analysis of graphs can be categorized, albeit broadly, into two types: the study of graph patterns and the study of patterns of graphs. The former covers topics such as node classification, degree distribution, link prediction, graph embedding, and subgraph presence. The latter is interested in patterns among sets of graphs. Examples of such patterns could be the presence of communities [1], hubs [2], or the small-world phenomenon [102]; each of which has been observed in various real-world networks.

This dissertation is concerned with the latter of the two categories, the analysis of patterns of graphs. To quantify patterns among sets of graphs, a common approach is to condition the results with respect to a metric, or measure of similarity, defined on the space of graphs. Popular metrics are typically defined as a function of some matrix representation of the graphs themselves, whether it be the adjacency matrix, Laplacian matrix, or normalized Laplacian matrix.

1.1 Preliminary notations and definitions

We denote by $G = (V, E)$ a graph with vertex set $V = \{1, 2, ..., n\}$ and edge set $E \subset V \times V$. For vertices $i, j \in V$ an edge exists between them if the pair $(i, j) \in E$. The size of a graph is called $n = |V|$ and the number of edges is $m = |E|$. The density of a graph is called $\rho_n = \frac{m}{n(n-1)/2}$.

The matrix $\boldsymbol{A}$ is the adjacency matrix of the graph and is defined as

$$a_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E, \\ 0 & \text{else.} \end{cases} \tag{1.1}$$

We define the function σ to be the mapping from the set of $n \times n$ adjacency matrices (square, symmetric matrices with zero entries on the diagonal), $\mathbb{M}_{n\times n}$ to $\mathbb{R}^n$ that assigns to an adjacency matrix the vector of its n sorted eigenvalues,

$$\sigma : \mathbb{M}_{n\times n} \longrightarrow \mathbb{R}^n, \tag{1.2}$$

$$\boldsymbol{A} \longmapsto \boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n], \tag{1.3}$$

where $\lambda_1 \geq \ldots \geq \lambda_n$. Because we often consider the c largest eigenvalue of the adjacency matrix $\boldsymbol{A}$, we define the mapping to the truncated spectrum as σ_c,

$$\sigma_c : \mathbb{M}_{n\times n} \longrightarrow \mathbb{R}^c, \tag{1.4}$$

$$\boldsymbol{A} \longmapsto \boldsymbol{\lambda}_c = [\lambda_1, \ldots, \lambda_c]. \tag{1.5}$$

Definition 1 We define the adjacency spectral pseudometric as the ℓ_2 norm between the spectra of the respective adjacency matrices,

$$d_A(G, G') = ||\sigma(\boldsymbol{A}) - \sigma(\boldsymbol{A'})||_2. \tag{1.6}$$

Definition 2 We define the truncated adjacency spectral pseudometric as the ℓ_2 norm between the largest c eigenvalues of the respective adjacency matrices,

$$d_{A_c}(G, G') = ||\sigma_c(\boldsymbol{A}) - \sigma_c(\boldsymbol{A'})||_2. \tag{1.7}$$

When the graphs have node correspondence, which is always assumed throughout this work, the Hamming distance may be defined as follows.

Definition 3 The Hamming distance is defined as

$$d_H(G, G') = ||\boldsymbol{A} - \boldsymbol{A'}||_1 = \sum_{i,j} |a_{ij} - a'_{ij}|. \tag{1.8}$$

Definition 4 We denote by $\mathcal{G}$ the set of all simple unweighted graphs on n nodes.

1.2 Choice of metric

The distance d_A, [104], exhibits good performance when comparing various types of graphs [103]. Spectral distances also exhibit practical advantages, as they can inherently compare graphs of different sizes and can compare graphs without known vertex correspondence (see, e.g., [34, 35] and references therein).

The adjacency spectrum is well-understood and is perhaps the most frequently studied graph spectrum (e.g., [31, 37]). The eigenvalues of the adjacency matrix carry important topological information about the graph at different structural scales [74, 104]. The spectrum reveals information about large-scale features such as community structure [69] or the existence of highly connected "hubs" [37], as well as the smaller scale structures such as the local connectivity (i.e., the degree of a vertex) or the ubiquity of substructures such as triangles [31]. In [103], the authors observe that the adjacency spectral pseudo-distance exhibits good performance across various scenarios, making it a reliable choice for a wide range of problems (from the two-sample test problem to the change point detection problem).

In practice, it is often the case that d_{A_c}, the truncated adjacency spectral distance is used. We still refer to such distances as spectral distances, and comparison using the largest c eigenvalues for small c allows one to focus on the global structures of the graphs while ignoring the local structures [69].

Another popular metric for graph comparison is the Hamming distance which, when the graphs have node correspondence, is given by Definition 3. Often, graphs in a data set do not have node correspondence, and a lengthy minimization procedure (see [11]) is solved before graph comparisons are made. The Hamming distance, which is sometimes referred to as the edit distance, reflects small-scale changes in the graphs and will be sensitive to the fine structural variations between graphs.

Much of the current work in the field involving machine learning from data sets of graphs is performed with respect to the Hamming distance (see, e.g. [19, 41, 55, 56, 59, 60, 87, 88]).

These references provide rigorous techniques to learn the local connectivity patterns of a data set of graphs. Throughout this dissertation, we consider machine learning questions that regard the global behaviors among sets of graphs rather than local behaviors. The adjacency spectral pseudo-metric is well suited to capture such global properties. With this perspective, our work is complementary to the work on the Hamming distance. We offer new techniques to compute a data set's sample mean and variance when considering the truncated adjacency spectral pseudo-metric.

1.3 Summary statistics for graph valued data sets

A classic approach in machine learning for any data set is to capture the behavior of the data via summary statistics such as the mean and variance. The mean and variance were generalized for data on a Riemannian manifold, equipped with a metric, with the concept of the Fréchet mean, and total Fréchet variance [84]. The concept of Fréchet mean and total Fréchet variance only requires that a (pseudo)metric between points be defined, and therefore one can consider the Fréchet mean and total Fréchet variance in a pseudo-metric space [42]. Not surprisingly, many machine learning algorithms developed for Euclidean spaces can be extended to use the Fréchet mean and total Fréchet variance.

1.3.1 Fréchet mean

Definition 5 (Fréchet mean [42]) The Fréchet mean of the probability measure μ in the pseudometric space $(\mathcal{G}, d_{A_c})$ is the set of graphs G^* whose expected distance squared to $\mathcal{G}$ is minimum,

$$\{G^* \in \mathcal{G}\} = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \, \mathbb{E}_\mu \left[d^2_{A_c}(G, G_\mu) \right], \tag{1.9}$$

where G_μ is a random realization of a graph distributed according to the probability measure μ, and the expectation $\mathbb{E}_\mu \left[d^2_{A_c}(G, G_\mu) \right]$ is computed with respect to the probability measure μ.

For graph-valued data sets, the Fréchet mean is a non-trivial problem to solve because it is defined by a minimization procedure over the space of graphs. The computation of a Fréchet mean

graph, G^*, for large n is intractable due to two primary issues. The first is that $|\mathcal{G}| = \mathcal{O}(2^{n^2})$ so any brute force procedure to solve the minimization problem in (1.9) will not compute in reasonable time. Second, the set $\mathcal{G}$ is not ordered, so searching the space of graphs in a principled manner is difficult (in contrast to the situation with trees [10, 14]).

When the Hamming distance defines the metric, the Fréchet mean has been well studied (e.g., [19, 41, 56, 55, 60] and references therein). Because the Hamming distance detects the local behavior within graphs, the Fréchet mean with respect to the Hamming distance can be interpreted as an average of the local structures in the observed graphs.

Throughout this dissertation, we consider that the fine scale, defined by the local connectivity at the level of each vertex, may be intrinsically random, and the quantification of random fluctuations may be uninformative when comparing graphs. For this reason, we always consider the adjacency spectral pseudo-distance. A primary contribution of this dissertation is a method to approximate the sample Fréchet mean graph when the metric is defined by adjacency spectral pseudo-distance. To the best of our knowledge, this is the first computation of the sample Fréchet mean when considering a spectral distance. We give below the related publications and submissions which constitute Chapter 2 of this dissertation.

Daniel Ferguson and François G. Meyer. *Computation of the Sample Fréchet Mean for Sets of Large Graphs with Applications to Regression* SIAM International Conference on Data Mining, (2022) 379 - 387.

Abstract: To characterize the location (mean, median) of a set of graphs, one needs a notion of centrality adapted to metric spaces since graph sets are not Euclidean spaces. A standard approach is to consider the Fréchet mean. In this work, we equip the set of graphs with the pseudometric defined by the ‘2 norm between the eigenvalues of their respective adjacency matrix. Unlike the edit distance, this pseudometric reveals structural changes at multiple scales and is well adapted to studying various statistical problems for graph-valued data. Using this pseudometric, we describe an algorithm to compute an approximation to the sample Fréchet mean of a set of undirected unweighted graphs with a fixed size.

The above work received the award for best conference paper. A full version of the article has been submitted to *Information and Inference: A Journal of the IMA*. The title and abstract follow below.

Daniel Ferguson and François G. Meyer. *Theoretical analysis and computation of the sample Fréchet mean for sets of large graphs based on spectral information* submitted: Information and Inference: A Journal of the IMA, (2022) 1-38.

Abstract: To characterize the location (mean, median) of a set of graphs, one needs a notion of centrality adapted to metric spaces since graph sets are not Euclidean spaces. A standard approach is to consider the Fréchet mean. In this work, we equip a set of graphs with the pseudometric defined by the ℓ_2 norm between the eigenvalues of their respective adjacency matrix. Unlike the edit distance, this pseudometric reveals structural changes at multiple scales and is well adapted to studying various statistical problems for graph-valued data. Using this pseudometric, we describe an algorithm to compute an approximation to the sample Fréchet mean of a set of undirected unweighted graphs with a fixed size. We provide a rigorous theoretical analysis of our algorithm and give several use cases for the computed mean graph.

1.3.2 Total Fréchet variance

Total Fréchet variance is closely connected to the Fréchet mean by definition.

Definition 6 (Total Fréchet variance [42]) The total Fréchet variance of the probability measure μ in the pseudometric space $(\mathcal{G}, d_{A_c})$ is defined as

$$V^*_{tot} = \mathbb{E}_\mu\big[d^2_{A_c}(G^*, G_\mu)\big]\,, \tag{1.10}$$

which is the evaluation of the Fréchet mean objective at a Fréchet mean graph.

Because a real positive number gives the total Fréchet variance, it is often seen as more tractable when analyzing graph-valued data sets. The total Fréchet variance has been used in the works of [27] to construct hypothesis tests and for change-point detection [28]. When considering the adjacency spectral pseudo-distance, the results of this dissertation show that the sample covariance matrix of the c largest eigenvalues captures similar information to the sample alternative of total Fréchet variance (see Chapter 3).

1.4 Generative models for graph valued data

Generative models for graphs have a long history. Popular ensembles of graphs aim to capture various observed real-world phenomena such as the presence of "hubs" [2], small world phenomenon [102], community structure [51], or specific subgraphs [64]. Popular variations of such ensembles further generalize the structures captured by the ensemble (see for instance the degree corrected stochastic block model [72], inhomogeneous Erdős-Rényi models [12, 22], or exponential random graph models [87]). A new method using neural networks, termed deep generative models, seeks to model the structures of observed graphs without pre-specifying the structures and instead learns the relevant structures in the observed graphs (see [16, 47] for reviews on these models).

While each ensemble captures various structural phenomena, the graph generation process depends upon a set of initial parameters. A data-driven generative model will seek to infer the correct value of these parameters conditioned on a set of observed graphs.

The authors in [88] offer current work in this field wherein they specify a generative process that distributes graphs about the Fréchet mean graph of an observed set. In this work, the Fréchet mean graph is determined with respect to the Hamming distance, but the ideas generalize to any Fréchet mean graph (that is, for any choice of distance assuming one can compute it). A mixture model with respect to the Hamming distance is proposed in [59] where multiple mean graphs are considered as defining the centers of each mixture component, and graphs are sampled within each mixture according to the observed deviations from the mean graph. When considering a Markov random graph model (or its generalization, the exponential random graph model), various procedures have been proposed to determine the parameters, for instance, maximum pseudo-likelihood estimation [95], and Monte Carlo Markov chain maximum [53, 92].

All parameter estimations in the above works compare graphs using the Hamming distance. In Chapter 3, we take the same perspective as in [88] in that the generative model parameters are inferred with respect to the sample Fréchet mean, however, we also consider the sample total Fréchet variance. A notable distinction of the work in this dissertation is that we compute these

quantities using the adjacency spectral pseudo-metric instead of the Hamming distance. Using these summary statistics of a data set of graphs, the proposed generative model captures the statistics of the largest eigenvalues of a sample set of graphs. To the best of our knowledge, this is the first generative model for heterogeneous data sets of graphs with respect to spectral information. A preprint of related work for this chapter is on the arxiv; the title and abstract are given below.

Daniel Ferguson and François G. Meyer. *Probability density estimation for sets of large graphs with respect to spectral information using stochastic block models* (2022) 1-47.

Abstract: For graph-valued data sampled iid from a distribution μ, the sample moments are computed with respect to a choice of metric. In this work, we equip the set of graphs with the pseudo-metric defined by the ℓ_2 norm between the eigenvalues of the respective adjacency matrices. We use this pseudo metric and the respective sample moments of a graph valued data set to infer the parameters of a distribution $\hat{\mu}$ and interpret this distribution as an approximation of μ.

1.5 Relationship between density and the metric

When determining the Fréchet mean with respect to a distance, a preliminary question is whether the density of the Fréchet mean graph is inherited from the population. In Chapter 4, we show that when considering either the Hamming metric or the adjacency spectral pseudo-metric, the sample Fréchet mean graph's density is bounded by the density of the observed graphs in the data set. The related work constitutes the final chapter of this dissertation. The title and abstract for the corresponding publication are presented below.

Daniel Ferguson and François G. Meyer. *On the Number of Edges of the Fréchet Mean and Median Graphs.* International Conference on Network Science, 2022.

Abstract: The availability of large datasets composed of graphs creates an unprecedented need to invent novel tools in statistical learning for graph-valued random variables. To characterize the average of a sample of graphs, one can compute the sample Frechet mean and median graphs. In this paper, we address the following foundational question: does a mean or median graph inherit the structural properties of the graphs in the sample? An important graph property is the edge density; we establish that edge density is a hereditary property, which can be transmitted from a graph sample to its sample Frechet mean or median graphs, irrespective of the method used to estimate the mean or the median. Because of the prominence of the Frechet mean in graph-valued machine learning, this novel theoretical result has some significant practical consequences.

1.6 Random Graphs

An important aspect of this study is the concept of random graphs. Here the concept of a probability measure on the space of graphs, $\mathcal{G}$, is introduced which allows for the unification of several different ensembles of graphs with a singular notation.

Definition 7 We denote by $\mathcal{M}(\mathcal{G})$ the space of probability measures on $\mathcal{G}$.

In this work, we always mean a probability measure when discussing a measure.

Definition 8 We define the set of random graphs distributed according to $\mu \in \mathcal{M}(\mathcal{G})$ as the probability space $(\mathcal{G}, \mu)$.

Remark 1 The σ-field associated with the $(\mathcal{G}, \mu)$ will always be the power set of $\mathcal{G}$.

This definition allows us to unify various ensembles of random graphs (e.g., Erdős-Rényi, inhomogeneous Erdős-Rényi, Small-World, Barabási-Albert, etc.) through the unique concept of a probability space.

1.6.1 Kernel Probability Measures

Here we define an essential class of probability measures for this study.

Definition 9 A probability measure $\mu \in \mathcal{M}(\mathcal{G})$ is called a kernel probability measure if there exists a positive constant ω_n and a function f,

$$f : [0,1] \times [0,1] \mapsto [0,1], \tag{1.11}$$

such that $f(x,y) = f(y,x)$, and

$$\forall G \in \mathcal{G}, \text{with adjacency matrix } \boldsymbol{A} = (a_{ij}),$$
$$\mu(\{\boldsymbol{A}\}) = \prod_{1 \le i < j \le n} \mathbb{P}(a_{ij}) = \prod_{1 \le i < j \le n} \text{Bernoulli}\left(\omega_n f(\frac{i}{n}, \frac{j}{n})\right).$$

The function f is called a kernel of μ and the probability measure is denoted $\mu_{\omega_n f}$.

Remark 2 We refer to these measures as kernel probability measures since the kernels naturally give rise to linear integral operators with kernels f. Furthermore, when the function f satisfies, $||f||_1 = 1$ then ω_n denotes the expected density of graphs sampled according to $\mu_{\omega_n f}$.

Note that given the sequence $\left\{\frac{i}{n}\right\}_{i=1}^{n}$ there exists a an equivalence class of functions f, defined on the grid points $\left\{\frac{i}{n}\right\}_{i=1}^{n} \times \left\{\frac{j}{n}\right\}_{j=1}^{n}$, which identify an equivalent kernel probability measure to $\mu_{\omega_n f}$. Throughout our analysis, we always refer to a piecewise Lipschitz representative function f for the kernel probability measure $\mu_{\omega_n f}$.

Definition 10 We denote by G_μ a random realization of a graph $G \in (\mathcal{G}, \mu)$.

1.6.1.1 Stochastic Block Models

The stochastic block model (see [1]) plays a vital role in this work. We review this model's specific features using the notations defined in the previous paragraphs. The critical aspects of the model are the geometry of the blocks, the within-community edges densities, and the across-community edge densities. An example of the kernel function and associated adjacency matrix from a stochastic block model is given in Fig. 1.6.1.1.

We denote by c the number of communities in the stochastic block model. The **geometry** of the stochastic block model is encoded using the relative sizes of the communities. We denote by $\boldsymbol{s} \in \ell_1$ a non-increasing non-negative sequence of relative community sizes with c non-zero entries and $||\boldsymbol{s}||_1 = 1$.

For the geometry specified by $\boldsymbol{s}$ we define an associated **edge density** vector $\boldsymbol{p} \in \ell_\infty$ such that $0 < p_i$ for $i = 1, ..., c$ and $p_i = 0$ for $i > c$ which describes the within-community edge densities.

Finally, we denote by $\boldsymbol{Q} = (q_{ij})$ an infinite matrix of cross-community edge densities where $q_{i,i} = 0$, $q_{i,j} = q_{j,i}$, and $q_{i,j} = 0$ if $i > c$ or $j > c$.

Remark 3 We allow for infinite vectors with a finite number of non-zero entries so that we may allow for the smooth introduction of new communities within the stochastic block model. For example, let $t \in [0, 1]$ and parametrize $\boldsymbol{s}$ and $\boldsymbol{p}$ by t as

$$\boldsymbol{s}(t) = \begin{bmatrix} 1 - t/2 \\ t/2 \\ 0 \\ \vdots \end{bmatrix} \quad \boldsymbol{p}(t) = \begin{bmatrix} 0.2 + t/2 \\ 0.1 + t/2 \\ 0 \\ \vdots \end{bmatrix}. \tag{1.12}$$

We can parameterize a stochastic block model using one representative of the equivalence class of kernels, f, which we call the canonical stochastic block model kernel.

Definition 11 (Canonical stochastic block model kernel) The function f, which is piecewise constant over the blocks, and is defined by

$f : [0, 1] \times [0, 1] \longrightarrow [0, 1]$

$$(x, y) \longmapsto \begin{cases} p_i & \text{if} \quad \sum_{k=1}^{i-1} s_k \leq x < \sum_{k=1}^{i} s_k, \\ & \text{and} \quad \sum_{k=1}^{i-1} s_k \leq y < \sum_{k=1}^{i} s_k, \\ q_{ij} & \text{if} \quad \sum_{k=1}^{i-1} s_k \leq x < \sum_{k=1}^{i} s_k, \\ & \text{and} \quad \sum_{k=1}^{j-1} s_k \leq y < \sum_{k=1}^{j} s_k \end{cases} \tag{1.13}$$

is called the canonical kernel of the stochastic block model with measure μ (see, e.g. Fig. 1.6.1.1), and we denote it by $f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})$.

Definition 12 (Set of canonical stochastic block model kernels) We denote the set of all canonical stochastic block model kernels f as

$$\mathcal{F} = \{f | f \text{ is a canonical stochastic block model kernel.}\} \tag{1.14}$$

Example 1. Given $\boldsymbol{s} = \begin{bmatrix} 1/2 & 1/4 & 1/4 & 0 \cdots \end{bmatrix}^T$ the values of $f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})$ in the unit square are shown in Fig. 1.6.1.1.

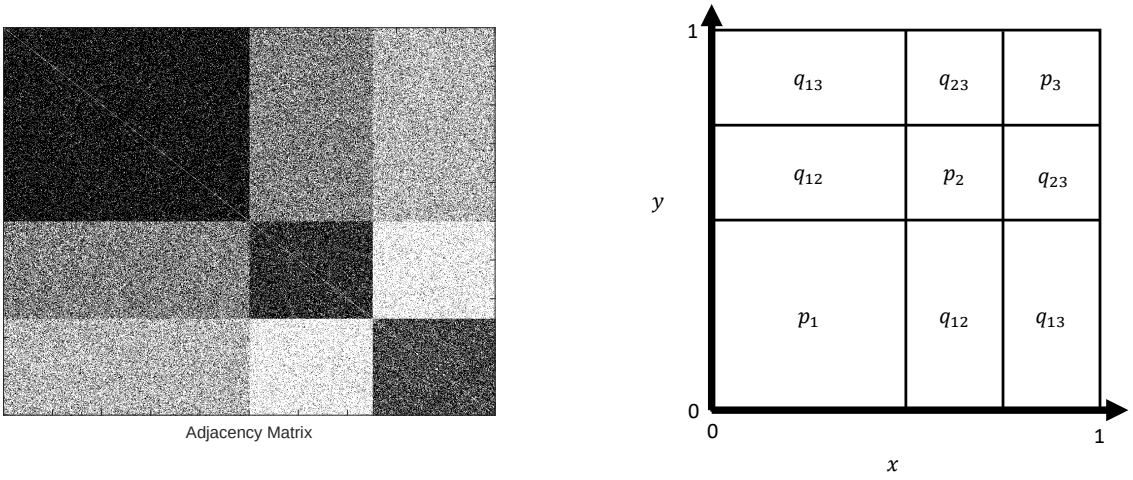

Figure 1.1: (Left) An example adjacency matrix: $\boldsymbol{A}$. (Right) An example $f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})$.

Chapter 2

Theoretical analysis and computation of the sample Fréchet mean for graphs

2.1 Introduction.

Machine learning almost always requires the estimation of the average of a dataset. Algorithms for clustering, classification, and linear regression all utilize the average value [49]. The mean is a simple algebraic operation when a norm induces the distance. This chapter aims to solve the nontrivial problem of determining the Fréchet mean for data sets of graphs when the pseudometric is the ℓ_2 distance between the eigenvalues of the adjacency matrices of two graphs.

In this work, we consider a set of simple graphs with n vertices which have an edge density, ρ_n, that satisfies,

$$\rho_n = \omega(n^{-2/3}) \tag{2.1}$$

We note that the vertex set must be sufficiently large and that the technique introduced in this chapter will perform poorly for sets of small graphs.

Our line of attack involves the following two intermediate results which are shown in this chapter: (1) given a graph, the largest eigenvalues of the adjacency matrix can be approximated with arbitrary precision by the eigenvalues of the mean graph from a stochastic block model; (2) given a set of target eigenvalues; one can recover the stochastic block model whose Fréchet mean has a spectrum that best matches these target eigenvalues. We prove various error bounds and convergence results for our algorithm and validate the theory with several experiments.

2.2 State of the Art.

We consider the set of undirected, unweighted graphs of fixed size n, wherein we define a distance. To characterize the location (mean, median) of the set of graphs, we need a notion of centrality adapted to metric spaces since graph sets are not Euclidean spaces. A standard approach considers the Fréchet mean and the total Fréchet variance.

The choice of metric is crucial to the computation of the Fréchet mean since each metric induces a different mean graph. The Fréchet mean of graphs has been studied in the context where the distance is the edit distance (e.g., [19, 41, 56, 55, 60] and references therein). The edit distance reflects small-scale changes in the graphs; therefore, the Fréchet mean will be sensitive to the fine structural variations between graphs. Effectively, the Fréchet mean with respect to the edit distance can be interpreted as an average of the fine structures in the observed graphs.

In this chapter, we consider that the fine scale, defined by the local connectivity at the level of each vertex, may be intrinsically random, and the quantification of such random fluctuations is uninformative when comparing graphs. We study a distance that can detect large-scale patterns of connectivity that happen at multiple scales (e.g., community structure [1, 69], modularity [44]).

The adjacency spectral distance, which we define as the ℓ_2 norm of the difference between the spectra of the adjacency matrices of the two graphs of interest [104], exhibits good performance when comparing various types of graphs [103], making it a reliable choice for a wide range of problems. Spectral distances also exhibit practical advantages, as they can inherently compare graphs of different sizes and can compare graphs without known vertex correspondence (see, e.g., [34, 35] and references therein).

A more robust notion of spectral similarity involves the similarity between the eigenspaces of the adjacency matrices of the respective graphs (e.g., [24, 61, 71]). These notions of spectral similarity have regained some interest in the context of graph coarsening (or aggregation) for graph neural networks [21, 23, 70, 82].

Inspired by the conjecture of Vu [101, 100], we propose to use the eigenvalues of the adjacency

matrix as "fingerprints" that uniquely (up to trivial permutations and the possible iso-spectral graphs) characterize a graph (see also [63]).

Instead of solving the minimization problem associated with the computation of the Fréchet mean in the original set $\mathcal{G}$, the authors in [35] suggest embedding the graphs in Euclidean space, wherein they can trivially find the mean of the set. Because the embedding in [35] is not an isometry, there is no guarantee that the inverse of the average embedded graphs is equal to the Fréchet mean. Furthermore, the inverse embedding may not be available in closed form. In the case of simple graphs, the Laplacian matrix of the graph uniquely characterizes the graph. The authors in [43] define the mean of a set of graphs using the Fréchet sample mean (computed on the manifold defined by the cone of symmetric positive semi-definite matrices) of the respective Laplacian matrices.

2.3 Main Contributions.

The Fréchet mean graph has become a standard tool for analyzing graph-valued data. In this chapter, we derive a method to compute the sample Fréchet mean when the distance between graphs is computed by comparing the spectra of the adjacency matrices of the respective graphs. We provide a rigorous theoretical analysis of our algorithm that demonstrates that our estimator converges toward the true sample Fréchet mean in the limit of large graph sizes. This is the first computation of the sample Fréchet mean for graphs when considering a spectral distance. This novel theoretical result relies on a combination of two ideas: stochastic block models provide universal approximants in the spectral adjacency pseudometric (see Theorem 1), and the dominant eigenvalues of the adjacency matrix of a stochastic block model can be used to recover the corresponding graph. We use numerical simulations to compare our theoretical analysis to the finite graph size estimates obtained with our algorithm.

2.4 The Fréchet mean and sample Fréchet mean

This section specifies the minimization procedures associated with the Fréchet mean, and sample Fréchet mean problem. We equip the set $\mathcal{G}$ of graphs defined on n vertices (see Definition 4) with the pseudometric defined by the ℓ_2 norm between the spectra of the respective adjacency matrices, d_{A_c}, (see Definition 2). We consider a probability measure $\mu \in \mathcal{M}(\mathcal{G})$ that describes the probability of obtaining a given graph when we sample $\mathcal{G}$ according to μ. Using d_{A_c}, we quantify the spread of the graphs, and we define a notion of centrality, which gives the location of the expected graph according to μ.

Definition 13 (Fréchet mean [42]) The Fréchet mean of the probability measure μ in the pseudometric space $(\mathcal{G}, d_{A_c})$ is the set of graphs G^* whose expected distance squared to the observed graphs is minimum,

$$\{G^* \in \mathcal{G}\} = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \, \mathbb{E}_\mu\big[d^2_{A_c}(G, G_\mu)\big] \,, \tag{2.2}$$

where G_μ is a random realization of a graph distributed according to the probability measure μ, and the expectation $\mathbb{E}_\mu\big[d^2(G, G_\mu)\big]$ is computed with respect to the probability measure μ.

Because $\mathcal{G}$ is a finite set, the minimization problem (2.2) always has at least one solution. Throughout this work, we are interested in determining at least one element of the set $\{G^* \in \mathcal{G}\}$. Because our results hold for any minimizer of (2.2) (i.e., for any Fréchet mean of μ), to ease the exposition, we write that the Fréchet mean is a singleton and we denote the Fréchet mean as

$$G^* = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \, \mathbb{E}_\mu\big[d^2_{A_c}(G, G_\mu)\big] \,. \tag{2.3}$$

We note the similarity between equation (2.3) and the definition of the barycenter [84]. Indeed, as we change μ, we expect that, for a fixed G, $\mathbb{E}_\mu\big[d^2_{A_c}(G, G_\mu)\big]$ will change, and therefore the Fréchet mean, G^*, will move inside $\mathcal{G}$ for different choices of the probability measure μ. Observe that G^* is the center of mass for the mass distribution associated with μ.

In practice, the only information known about a distribution on $\mathcal{G}$ comes from a sample of graphs. Therefore, we need a notion of the sample Fréchet mean, defined by replacing μ with the empirical measure.

Definition 14 (Sample Fréchet mean) Let $\left\{G^{(k)}\right\}$ $1 \leq k \leq N$ be a set of graphs in $\mathcal{G}$. The sample Fréchet mean is defined by

$$G_N^* = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \frac{1}{N} \sum_{k=1}^{N} d_{A_c}^2(G, G^{(k)}). \tag{2.4}$$

Again, our results hold for any minimizer of (2.4), so we write the sample Fréchet mean as a singleton set. The dependence on N is explicitly written here but may be suppressed throughout the chapter when it is obvious.

The computation of G_N^* for sets of large graphs is intractable due to two primary issues. The first is that $|\mathcal{G}| = \mathcal{O}(2^{n^2})$ so any brute force procedure to solve the minimization problem in (2.4) will not compute in reasonable time. Second, the set $\mathcal{G}$ is not ordered, so searching the space of graphs in a principled manner is difficult (in contrast to the situation with trees [10, 14]). We suggest solving these issues by first lifting the sample Fréchet mean problem to $\mathcal{M}(\mathcal{G})$ and defining an approximation to the lifted problem. Our specific approach involves searching for the correct parameters of a canonical stochastic block model kernel, f, such that the sample Fréchet mean given $\mu_{\omega_n f}$ almost surely approximates the target graph, G_N^*, with respect to d_{A_c}.

2.5 Approximation of the sample Fréchet mean.

We first state the primary theoretical results, Theorem 1 and Corollary 1, which constitute the main contributions of this work and which form the foundation of our algorithm (see Alg. 2). This section also states the necessary theorems for implementing our algorithm. We begin with our assumptions. Let $G \in \mathcal{G}$ with adjacency matrix $\boldsymbol{A}$ and assume

1. $\rho_n = \omega_n(n^{-2/3})$.

2. $\lim_{n\to\infty} \rho_n = 0$.

3. $0 \preccurlyeq \sigma_c(\boldsymbol{A})$.

4. For every $1 \leq i \neq j \leq c$, $\lambda_i \neq \lambda_j$.

Assumption 1 is perhaps the most restrictive, stating that the graphs we consider are not too sparse. It has been seen that most real-world networks are considerably sparse and may not satisfy this assumption. It is noteworthy, however, that this assumption is determined by the solution technique we employ when solving the sample Fréchet mean problem. It is a product of the results in [22] which characterizes graphs with the same restriction. The remaining assumptions are relatively mild. Many graphs have a decaying density as n grows, which states that the average degree of a vertex is dominated by n. Furthermore, the largest eigenvalues of graphs are distinct with high probability and positive.

Our primary theoretical contribution, Theorem 1, states that we may approximate well any graph G that satisfies our assumptions by the sample Fréchet mean of an appropriate stochastic block model kernel probability measure, $\mu_{\omega_n f}$, almost surely with respect to the truncated adjacency spectral pseudo-metric.

Theorem 1 (Spectrally similar large graphs) $\forall \epsilon > 0$, $\exists n_1 \in \mathbb{N}$ such that $\forall n > n_1$, $\exists c > 0$, and $\exists f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})$ a canonical stochastic block model kernel with c communities such that

$$\lim_{N \to \infty} d_{A_c}(G, G^*_{N, \mu_{\omega_n f}}) < \epsilon \quad a.s. \tag{2.5}$$

where $G^*_{N, \mu_{\omega_n f}}$ denotes the sample Fréchet mean of $\{G^{(k)}\}_{k=1}^N$, an iid sample distributed according to $\mu_{\omega_n f}$.

Proof of Theorem 1 The proof is in Appendix A.3.

Remark 4 The choice of c for the above theorem is of significant interest as it specifies the number of non-zero entries in the geometry vector $\boldsymbol{s}$ and additionally specifies the number of largest eigenvalues we consider when comparing graphs. Though other methods also exist, we give in Alg. 1 one suggestion for c. It is worth mentioning that under-estimating c may be preferred to over-estimating c since an overestimate of c will compare eigenvalues that capture fine-scale behavior

in the graphs to eigenvalues that determine global structures of the graphs, leading to potentially inaccurate conclusions.

While we are free to choose the entries of the geometry vector $\boldsymbol{s}$, we make the choice that $s_1 \geq s_i$ for $i = 2, ..., c$ and $s_i = s_j$ for $i, j = 2, ..., c$. This choice is not required and any choice of the nonzero entries of the vector $\boldsymbol{s}$ would be suitable (see e.g. subsection 2.6.5).

The following corollary applies Theorem 1 to the sample Fréchet mean of any given data set of graphs, $\{G^{(k)}\}_{k=1}^{N}$. It states that for any given set of graphs whose sample Fréchet mean, G_N^*, satisfies the assumptions of Theorem 1, there exists a canonical stochastic block model kernel defining a probability measure, $\mu_{\omega_n f}$, where the sample Fréchet mean of an iid sample from $\mu_{\omega_n f}$, denoted $G^*_{\tilde{N},\mu_{\omega_n f}}$, is almost surely close to G_N^*. This corollary forms the basis of our approach to solving the sample Fréchet mean problem, equation (2.4).

Let $\{G^{(k)}\}_{k=1}^{N}$ be a set of graphs with sample Fréchet mean G_N^*. Assume G_N^* satisfies the assumptions of Theorem 1.

Corollary 1 (Approximation of the sample Fréchet mean) $\forall \epsilon > 0, \exists n_1 \in \mathbb{N}$ such that $\forall n > n_1$, $\exists f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})$ a canonical stochastic block model kernel with c communities such that

$$\lim_{\tilde{N} \to \infty} d_{A_c}(G_N^*, G^*_{\tilde{N},\mu_{\omega_n f}}) < \epsilon \quad a.s. \tag{2.6}$$

where $G^*_{\tilde{N},\mu_{\omega_n f}}$ denotes the sample Fréchet mean of $\{\tilde{G}^{(k)}\}_{k=1}^{\tilde{N}}$, an iid sample distributed according to $\mu_{\omega_n f}$.

Remark 5 One requirement on G_N^* is that the density, denoted ρ_n^*, satisfies $\rho_n^* = \omega(n^{-2/3})$. The theory in [32] suggests that as long each graph in our sample set $\{G^{(k)}\}_{k=1}^{N}$ satisfies this density condition, then so too does G_N^*.

Theorem 1 and Corollary 1 suggest that the sample Fréchet mean can be approximated well by the sample Fréchet mean of a data set of graphs sampled according to a stochastic block model in the limit of large graph size. This result is purely existential and does not provide an algorithm to construct the stochastic block model probability measure from which the graphs should be sampled.

We now address how to determine the correct canonical stochastic block model kernel f with the following four theorems. The strategy to determine the kernel f begins with the fact that the eigenvalues, $\sigma_c(\boldsymbol{A}_N^*)$, are approximated by the sample mean spectrum, $\frac{1}{N}\sum_{k=1}^N \sigma_c(\boldsymbol{A}^{(k)})$ (Theorem 2). We can therefore find the canonical stochastic block model kernel f where the sample Fréchet mean from $\mu_{\omega_n f}$ has an adjacency matrix whose eigenvalues best approximate the sample mean spectrum, $\frac{1}{N}\sum_{k=1}^N \sigma_c(\boldsymbol{A}^{(k)})$, (Theorem 3) and estimating the eigenvalues of the sample Fréchet mean graph using either Theorems 4 or 5. Once we have estimated the parameters of the canonical stochastic block model kernel f we estimate the sample Fréchet mean given f, denoted $G^*_{\tilde{N},\mu_{\omega_n f}}$, by sampling graphs distributed according to $\mu_{\omega_n f}$ and computing the set mean graph (Theorem 6).

Let $\{G^{(k)}\}_{k=1}^N$ be a set of graphs with sample Fréchet mean G_N^* with adjacency matrix $\boldsymbol{A}_N^*$. Our following theorem shows that the eigenvalues of the adjacency matrix $\boldsymbol{A}_N^*$ are approximated well by the sample mean spectrum.

Theorem 2 $\forall \epsilon > 0$, $\exists n^* \in \mathbb{N}$ such that $\forall n > n^*$,

$$||\sigma_c(\boldsymbol{A}_N^*) - \frac{1}{N}\sum_{k=1}^N \sigma_c(\boldsymbol{A}^{(k)})||_2 < \epsilon. \tag{2.7}$$

Proof of Theorem 2 The proof is in Appendix A.4

Let $\{\tilde{G}\}_{k=1}^{\tilde{N}}$ be an iid sample of graphs distributed according to $\mu_{\omega_n f}$ with sample Fréchet mean $G^*_{\tilde{N},\omega_n f}$. Let $\boldsymbol{A}^*_{\tilde{N},\omega_n f}$ be the adjacency matrix of $G^*_{\tilde{N},\omega_n f}$. Our next two theorems show how we estimate the eigenvalues, $\sigma_c(\boldsymbol{A}^*_{\tilde{N},\omega_n f})$, in terms of the kernel function f. We first show that the expected eigenvalues, $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]$, are almost surely within ϵ of $\sigma_c(\boldsymbol{A}^*_{\tilde{N},\omega_n f})$.

Theorem 3 (The Eigenvalues of the sample Fréchet Mean of Stochastic Block Models) $\forall \epsilon > 0$, $\exists n^* \in \mathbb{N}$ such that for all $n > n^*$,

$$\lim_{\tilde{N}\to\infty} ||\sigma_c(\boldsymbol{A}^*_{\tilde{N},\omega_n f}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 < \epsilon \quad a.s. \tag{2.8}$$

Proof of Theorem 3 The proof is in Appendix A.3.

Since we do not have a closed form expression for $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]$, Theorem 4 and Theorem 5 show that we can estimate the expected eigenvalues, $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]$, in terms of the kernel function f. It should also be noted that the following two theorems are a modification of Theorem 2.4 in [22].

Let $\mu_{\omega_n f} \in \mathcal{M}(\mathcal{G})$ be a kernel probability measure with kernel f. Let L_f be the linear integral operator with the same kernel function, f. Assume L_f has a finite rank of c. Denote the eigenvalues and eigenfunctions of L_f as $\lambda_i(L_f)$ and $r_i(x)$ respectively where for each $i = 1, ..., c$, $r_i(x)$ is assumed to be piecewise Lipschitz with finitely many discontinuities.

Theorem 4 For every $1 \leq i \leq c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] = \lambda_i(\boldsymbol{B}^*) + \mathcal{O}(\sqrt{\omega}_n), \tag{2.9}$$

where

$$\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}. \tag{2.10}$$

and

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b^{*,(1)}_{j,l} = \sqrt{\theta_j \theta_l} n \omega_n \int_0^1 r_j(x) r_l(x) dx = \begin{cases} \theta_j n \omega_n & j = l \\ 0 & j \neq l \end{cases} \tag{2.11}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b^{*,(2)}_{j,l} = \theta_i^{-2} \sqrt{\theta_j \theta_l} \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx. \tag{2.12}$$

Proof of Theorem 4 The proof is in Appendix A.2. This is a modification of Theorem 2.4 from [22].

The above theorem provides a first-order approximation of the expected eigenvalues of stochastic block model graphs in terms of the eigenvalues of the matrix $\boldsymbol{B}^*$. Often these eigenvalues are not explicit in terms of the parameters, $\boldsymbol{p}, \boldsymbol{Q}$, and $\boldsymbol{s}$. A less accurate, but simpler, estimate comes in the form of the next theorem.

Theorem 5 (Estimation of the Largest Eigenvalues of Stochastic Block Models)

For $i = 1, ..., c$

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] = \lambda_i(L_f)n\omega_n + \mathcal{O}(1) \tag{2.13}$$

Proof of Theorem 5 The proof is in Appendix A.2.

Remark 6 Both theorems provide an estimate of the largest eigenvalues of the graphs of interest. When an $\mathcal{O}(1)$ error in the eigenvalues is of little consequence we suggest using the $\mathcal{O}(1)$ estimate from Theorem 5 due it its computational simplicity when implementing Alg. 2.

Recall that our goal is to approximately compute G_N^* by finding a stochastic block model kernel f and determining the sample mean of graphs distributed according to $\mu_{\omega_n f}$. To determine the canonical stochastic block model kernel, we may align the estimates of the expected eigenvalues from either Theorem 4 or Theorem 5, with the sample mean spectrum $\frac{1}{N}\sum_{k=1}^{N}\sigma_c(\boldsymbol{A}^{(k)})$, Theorem 2. Equation (2.14) outlines the objective function to minimize when considering Theorem 4 while equation (2.15) outlines the objective function to minimize when considering the estimate from Theorem 5.

Let $\boldsymbol{B}^*$ be as defined in Theorem 4. Determine f according to the following minimization problem,

$$f^* = \underset{f\in\mathcal{F}}{\operatorname{argmin}} \sum_{i=1}^{c}\left|\lambda_i(\boldsymbol{B}^*) - \frac{1}{N}\sum_{k=1}^{N}\lambda_i(\boldsymbol{A}^{(k)})\right|^2 \tag{2.14}$$

where $\boldsymbol{B}^*$ implicit depends on the function f. To determine f using the estimations from Theorem 5 we solve the following minimization problem,

$$f^* = \underset{f\in\mathcal{F}}{\operatorname{argmin}} \sum_{i=1}^{c}\left|n\omega_n\lambda_i(L_f) - \frac{1}{N}\sum_{k=1}^{N}\lambda_i(\boldsymbol{A}^{(k)})\right|^2. \tag{2.15}$$

Remark 7 We suggest a gradient descent algorithm to determine the correct canonical stochastic block model kernel function f^* where the gradient is computed with respect to the variables $\boldsymbol{p}$ and q which define the kernel function f.

Given the canonical stochastic block model kernel f that solves equation (2.14), Theorem 6 shows a method of estimating the sample Fréchet mean of graphs distributed iid according to $\mu_{\omega_n f}$ by sampling from $\mu_{\omega_n f}$.

Let $\{\tilde{G}^{(k)}\}_{k=1}^{\tilde{N}}$ be a sample of graphs distributed according to $\mu_{\omega_n f}$ where f is the canonical stochastic block model kernel. Define the set mean graph by

$$\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}} = \underset{\tilde{G}\in\{\tilde{G}^{(k)}\}_{k=1}^{\tilde{N}}}{\operatorname{argmin}} \frac{1}{\tilde{N}} \sum_{k=1}^{\tilde{N}} d^2_{A_c}(\tilde{G}, \tilde{G}^{(k)}) \tag{2.16}$$

with adjacency matrix $\hat{\boldsymbol{A}}^*_{\tilde{N},\mu_{\omega_n f}}$.

Theorem 6 (Convergence in probability of the truncated spectrum of the set mean graph)

$\forall \epsilon > 0$,

$$\lim_{n\to\infty} P(||\sigma_c(\hat{\boldsymbol{A}}^*_{\tilde{N},\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 > \epsilon) = 0. \tag{2.17}$$

Proof of Theorem 6 The proof is in Appendix A.5.

Due to Theorem 2.3 in [22], which we restate as Theorem 7 below concerning the convergence in distribution to a multivariate normal of the eigenvalues of adjacency matrices from the stochastic block model, we observe that relatively small size of $\tilde{N}$ is sufficient. Throughout our experiments in Section 2.6, we take $\tilde{N} = 5$, which provides good results for the case of finite graph size.

The practical significance of our theoretical analysis is the invention of an algorithm to approximate the solution to the sample Fréchet mean problem with respect to d_{A_c}, equation (2.4), which we give the pseudo-code for below.

2.5.1 Summary and Algorithm

Given a finite sample of graphs $\{G^{(k)}\}_{k=1}^{N}$, our theory allows us to estimate the sample Fréchet mean graph, G^*_N, by solving an approximate problem in two steps:

(1) Identify the correct canonical stochastic block model kernel f by solving equation (2.14) or (2.15).

(2) Estimate $G^*_{\tilde{N},\mu_{\omega_n f}}$ using Theorem 6 taking $\tilde{N}$ as large as desired.

A notable first step is to estimate c, the number of eigenvalues of G^*_N to consider. We suggest estimating c as per Alg. 1 and use this estimate in Alg. 2.

Algorithm 1 Determine c for the approximate sample Fréchet mean

Require: Set of graphs, $M = \{G^{(k)}\}_{k=1}^N$, and integer K

1: Compute the geometric average spectrum of graphs in M as $\bar{\boldsymbol{\lambda}} = \frac{1}{N}\sum_{k=1}^N \boldsymbol{\lambda}^{(k)}$.

2: Initialize $i = 0$.

3: **Do**

4: $i = i + 1$

5: Initialize $r = \bar{\boldsymbol{\lambda}}(i)$

6: Initialize the semi-circle probability density function (see e.g. [5]), as $s(\lambda; r)$ where r is the radius.

7: Assume $\bar{\boldsymbol{\lambda}}(j) \sim s(\lambda; r)$ for $j = i, ..., n$.

8: Determine the PDF of the K largest order statistics with a sample size $n - i$, $\lambda_{(n-i)}, ..., \lambda_{(n-i-K+1)}$

9: Compute the expected value of the K largest order statistics from the PDF $s(\lambda; r)$ with a sample size of $n - i$.

10: With sample size $n - i$, compute the standard deviation of the K largest order statistics, $\sigma_{n-i}, ..., \sigma_{n-i-K+1}$

11: **While** $|\bar{\lambda}(1 + i) - \mathbb{E}\left[\lambda_{(n-i)}\right]| > \sigma_{n-i} \vee ... \vee |\bar{\lambda}(K + i) - \mathbb{E}\left[\lambda_{(n-i-K)}\right]| > \sigma_{n-i-K+1}$

12: **Return:** $c = i - 1$

Alg. 1 assumes that all but the c largest eigenvalues in the vector $\bar{\boldsymbol{\lambda}}$ follow a bulk distribution given by the semi-circle law (see e.g. [5, 29] and references therein). We determine the edge of the bulk iteratively by assuming the edge is defined by the largest observed eigenvalue and compute whether the next K sequential eigenvalues are within a standard deviation of their expected value. Upon termination, the number of eigenvalues left outside the bulk determines our choice for c. Note

that any estimate of c will suffice, and the algorithm above is a suggestion.

An important discussion as to the choice of c is in order. Alg. 1 determines c for the truncated adjacency spectral pseudo-metric based on the information provided by the arithmetic average of the observed eigenvalues. We then use this estimate for the number of extreme eigenvalues of the adjacency matrix of the sample Fréchet mean graph. We always compare two graphs based on our estimate of c, which is interpreted as the number of extreme eigenvalues of the adjacency matrix of the sample Fréchet mean graph. However, it could very well be possible that there exist graphs $G^{(j)}$ and $G^{(k)}$ in our dataset M that have a very different number of extreme eigenvalues from c. We explore this problem via a simple example.

Take, for instance, the case of an observation from a two community stochastic block model and an observation from a one community stochastic block model. To compare the two graphs, we must choose c in the pseudo-metric d_{A_c}. It is unclear whether the choice for c should be 1 or 2 since the two graphs have differing numbers of extreme eigenvalues.

Note the issue is still not necessarily resolved even when considering the adjacency spectral pseudo-metric d_A (equivalently, $c = n$ for d_{A_c}). While we may be able to compare all the eigenvalues of each graph, it is not obvious what information is provided by comparing the second largest eigenvalue from a two-community stochastic block model graph's adjacency matrix to the second largest eigenvalue of a homogeneous Erdős-Rényi. In a sense, the comparison is between the structural information from one graph and the random noise from the other.

We acknowledge that these issues persist throughout our algorithm and use our estimate of c from Alg. 1 for the pseudo-metric d_{A_c}. It is worth noting that any estimate of c will suffice for an implementation of our algorithm.

Algorithm 2 Approximate sample Fréchet mean

Require: Set of graphs, $M = \{G^{(k)}\}_{k=1}^{N}$

1: Compute the average density $\bar{\rho}_n$ of the graphs in M and set $\omega_n = \bar{\rho}_n$.

2: Approximate c via Alg. 1 and determine $\boldsymbol{s}$ (see Remark 4).

3: For each $i = 1, ..., c$ compute $\bar{\lambda}_i = \frac{1}{N}\sum_{k=1}^{N} \lambda_i(\boldsymbol{A}^{(k)})$.

4: Randomly initialize $\boldsymbol{p}$

5: Initialize $\boldsymbol{Q} = (q_{ij})$ such that $q_{ij} = q$ for all i, j and enforce $||f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})||_1 = 1$.

6: **while** Relative change in $\boldsymbol{p}$ and q is large **do**

7: Estimate the gradient of the objective in equation (2.14) or (2.15) via centered differences.

8: Update $\boldsymbol{p}$ via a projected gradient descent step

9: Update q such that $||f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})||_1 = 1$

10: **end while**

11: Estimate $G^*_{\tilde{N},\mu_{\omega_n f}}$ as $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$ (see Theorem 6)

12: **Return:** $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$.

The founding idea for Alg. 2 is the following: Any graph G can be expressed as the Fréchet mean of some probability measure. For large graphs, our theory shows we may search for a kernel probability measure, $\mu_{\omega_n f}$, by aligning the eigenvalues of L_f (Steps 6 - 10) and then estimating the Fréchet mean of $\mu_{\omega_n f}$ using the set mean graph (Step 11).

Remark 8 Corollary 1 shows the existence of a canonical stochastic block model kernel, f. In our algorithm we choose to seek a kernel f with $||f||_1 = 1$ so that the expected density of graphs distributed according to $\mu_{\omega_n f}$ is $\bar{\rho}_n$.

2.5.2 Computational complexity of Algorithm 2

Step 11, determining $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$, is the most computationally expensive with a complexity of $\mathcal{O}(\tilde{N}n^2c^3)$. This complexity is because we must generate $\tilde{N}$ graphs on n vertices and compute the c largest eigenvalues of each to determine the most central element, $\widehat{G}^*_{\tilde{N},\text{SBM}}$.

Remark 9 It should be noted that for large enough n, taking $\tilde{N} = 1$ provides a sufficient estimate and the number of operations for step 11 is reduced to $\mathcal{O}(n^2)$.

Before providing experimental validation for our algorithm we first would like to confirm that the result of Alg. 2 is near the correct sample Fréchet mean graph. We do so by providing a method to compute a confidence interval for our result. To construct such a confidence interval we utilize Theorem 7 (a restatement of Theorem 2.3 from [22]) which states that the extreme eigenvalues of the adjacency matrices of graphs distributed according to $\mu_{\omega_n f}$ are asymptotically multivariate normal. This theorem allows us to obtain a lower bound on the probability of our confidence set containing the sample Fréchet mean graph. We then determine a confidence set by analyzing the limiting multi-variate Gaussian distribution given in equation (2.18).

Let $\mu_{\omega_n f}$ be a stochastic block model kernel probability measure.

Theorem 7 (Chakrabarty, Chakraborty, and Hazra 2020)

$$\omega_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]) \xrightarrow{d} (Z_i : 1 \leq i \leq c), \tag{2.18}$$

where the right-hand side is a multivariate normal random vector in $\mathbb{R}^c$, with mean zero and

$$Cov(Z_i, Z_j) = 2\int_0^1 \int_0^1 r_i(x) r_i(y) r_j(x) r_j(y) f(x,y) dx dy, \tag{2.19}$$

for all $1 \leq i, j \leq c$.

Proof of Theorem 7 The proof is in [22].

We begin the construction of the confidence set by defining the cumulative distribution functions for the random variables $\omega_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right])$ and $\boldsymbol{Z} = (Z_i : 1 \leq i \leq c)$ respectively as

$$F_n(\boldsymbol{z}) = P_n\left((-\infty, z_1] \times ... \times (-\infty, z_c]\right), \tag{2.20}$$

$$F(\boldsymbol{z}) = P\left((-\infty, z_1] \times ... \times (-\infty, z_c]\right). \tag{2.21}$$

The convergence in distribution of the random vectors is equivalent to the following: $\forall \boldsymbol{z} \in \mathbb{R}^c$,

$$\lim_{n\to\infty} F_n(\boldsymbol{z}) = F(\boldsymbol{z}), \tag{2.22}$$

since $F(\boldsymbol{z})$ is continuous everywhere (see Appendix A.1).

For the construction of a confidence set it is easier to work with the probabilities and random vectors directly and so equation (2.22) takes the form

$$\lim_{n\to\infty} P_n(\omega_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]) \preccurlyeq \boldsymbol{z}) = P(\boldsymbol{Z} \preccurlyeq \boldsymbol{z}). \tag{2.23}$$

With equation (2.23), we can now construct a confidence set for the sample Fréchet mean graph from a stochastic block model probability measure.

Let $M_1 = \{G^{(k)}\}_{k=1}^{N_1}$ be a sample of graphs distributed iid according to some distribution μ. Let $G^*_{N_1}$ be the sample Fréchet mean with adjacency matrix $\boldsymbol{A}^*_{N_1}$.

Lemma 1 Let $\alpha > 0$. Let $\epsilon \in \mathbb{R}$ such that $0 < \epsilon < \alpha$. Take $0 \preccurlyeq \boldsymbol{z}_\alpha \in \mathbb{R}^c$ such that

$$1 - \alpha < P(-\boldsymbol{z}_\alpha \preccurlyeq \boldsymbol{Z} \preccurlyeq \boldsymbol{z}_\alpha) - \epsilon. \tag{2.24}$$

There exists $n^* \in \mathbb{N}$ such that for all $n > n^*$, there exists a stochastic block model kernel function $f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})$ defining the probability measure $\mu_{\omega_n f}$ in Corollary 1. Let $M_2 = \{G^{(k)}\}_{k=1}^{N_2}$ be a sample of graphs distributed according to $\mu_{\omega_n f}$. Let $\widehat{G}^*_{N_2}$ be the set mean graph of the set M_2 (see 2.16) with adjacency matrix $\widehat{\boldsymbol{A}}^*_{N_2}$ such that for $N_2 = 1$

$$1 - \alpha < P_n\left(d_{A_c}\left(G^*_{N_1}, \widehat{G}^*_1\right) < \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right). \tag{2.25}$$

The set $\left\{G \in \mathcal{G} | d_{A_c}\left(G^*_{N_1}, \widehat{G}^*_1\right) < \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right\} \subset \mathcal{G}$ is a $1-\alpha$ confidence set for the sample Fréchet mean graph, $G^*_{N_1}$.

Proof of Lemma 1 We first note that

$$d_{A_c}\left(G^*_{N_1}, \widehat{G}^*_{N_2}\right) = ||\sigma_c(\boldsymbol{A}^*_{N_1}) - \sigma_c(\widehat{\boldsymbol{A}}^*_{N_2})||_2. \tag{2.26}$$

We will from here on be working with the eigenvalues of the adjacency matrices rather than the graphs themselves. Now, due to equation (2.23) we have that there exists an n^* such that for all $n > n^*$

$$1 - \alpha < P(-\boldsymbol{z}_\alpha \preccurlyeq \boldsymbol{Z} \preccurlyeq \boldsymbol{z}_\alpha) - \epsilon \tag{2.27}$$

$$< P_n\left(-\boldsymbol{z}_\alpha \preccurlyeq \omega_n^{-1/2}\left(\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right) \preccurlyeq \boldsymbol{z}_\alpha\right). \tag{2.28}$$

Note that for $N_2 = 1$, the set mean graph, $\widehat{G}_1^*$, is simply a random observation distributed according to $\mu_{\omega_n f}$ and we may replace $\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})$ by $\sigma_c(\widehat{\boldsymbol{A}}_1^*)$ leading to the following inequality,

$$1 - \alpha < P_n\left(-\boldsymbol{z}_\alpha \preccurlyeq \omega_n^{-1/2}\left(\sigma_c(\widehat{\boldsymbol{A}}_1^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right) \preccurlyeq \boldsymbol{z}_\alpha\right). \tag{2.29}$$

By interpolating $\sigma_c(\boldsymbol{A}_{N_1}^*)$ we get the following:

$$1 - \alpha < P_n\left(-\boldsymbol{z}_\alpha \preccurlyeq \omega_n^{-1/2}\left(\sigma_c(\widehat{\boldsymbol{A}}_1^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right) \preccurlyeq \boldsymbol{z}_\alpha\right) \tag{2.30}$$

$$= P_n\left(-\omega_n^{1/2}\boldsymbol{z}_\alpha \preccurlyeq \sigma_c(\widehat{\boldsymbol{A}}_1^*) - \sigma_c(\boldsymbol{A}_{N_1}^*) + \sigma_c(\boldsymbol{A}_{N_1}^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right] \preccurlyeq \omega_n^{1/2}\boldsymbol{z}_\alpha\right) \tag{2.31}$$

$$\leq P_n\left(||\sigma_c(\widehat{\boldsymbol{A}}_1^*) - \sigma_c(\boldsymbol{A}_{N_1}^*) + \sigma_c(\boldsymbol{A}_{N_1}^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \leq \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right) \tag{2.32}$$

where we have taken a norm of the argument in the probability from equation (2.31) resulting in (2.32). Continuing,

$$1 - \alpha \leq P_n\left(||\sigma_c(\widehat{\boldsymbol{A}}_1^*) - \sigma_c(\boldsymbol{A}_{N_1}^*)||_2 + ||\sigma_c(\boldsymbol{A}_{N_1}^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \leq \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right) \tag{2.33}$$

$$\leq P_n\left(||\sigma_c(\widehat{\boldsymbol{A}}_1^*) - \sigma_c(\boldsymbol{A}_{N_1}^*)||_2 \leq \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2 \wedge ||\sigma_c(\boldsymbol{A}_{N_1}^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \leq \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right) \tag{2.34}$$

Let A be the event that $||\sigma_c(\widehat{\boldsymbol{A}}_1^*) - \sigma_c(\boldsymbol{A}_{N_1}^*)||_2 \leq \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2$ and B be the event that $||\sigma_c(\boldsymbol{A}_{N_1}^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \leq \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2$. Then

$$1 - \alpha \leq P_n(A)\, P_n(B|A). \tag{2.35}$$

Dividing by the last term on the right-hand side,

$$\frac{1 - \alpha}{P_n(B|A)} < P_n(A) \tag{2.36}$$

Since $1-\alpha < \frac{1-\alpha}{P_n(B|A)}$ we find that

$$1-\alpha < P_n\left(A\right) \tag{2.37}$$

$$= P_n\left(||\sigma_c(\widehat{\boldsymbol{A}}_1^*) - \sigma_c(\boldsymbol{A}_{N_1}^*)||_2 \leq \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right) \tag{2.38}$$

$$= P_n\left(d_{A_c}(\widehat{G}_1^*, G_{N_1}^*) \leq \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right) \tag{2.39}$$

which is what we aimed to show. □

In practice, we are not always able to control n, the size of the graphs. Since we have the luxury of choosing N_2, one could perform a similar analysis and construct a confidence set for sufficiently large N_2, the number of samples used to construct the set mean graph. We do not explore this case as our regime assumes that n is sufficiently large.

2.6 Experimental Validation.

2.6.1 Assessment, Validation, and Comparison

A brute force computation of the Fréchet mean or median graph based on the adjacency spectral pseudo-distance is unrealistic (it requires about $\Omega\left(n^2 2^{n^2}\right)$ operations for graphs of size n), and we therefore do not provide a ground truth in our experiments (see section 2.6). One may consider comparing the Fréchet mean computed here to a Fréchet mean computed with respect to the edit distance for which several optimization algorithms have been proposed (e.g., [9, 20, 36, 55, 60]). While this comparison may be feasible, it is uninformative as the Fréchet mean with respect to the edit distance need not have any resemblance to the Fréchet mean with respect to d_{A_c}.

All the code and data are provided at `https://github.com/dafe0926/approx_Graph_Frechet_Mean`. To the best of our knowledge, this study provides the first algorithm to compute the sample Fréchet mean for a dataset of graphs when considering a spectral distance, as a consequence, we have no baselines to compare our results with.

2.6.2 Choice of the Datasets

Graph-valued databases have recently been created and made available publicly [86, 83, 77, 76]. These databases are designed for the evaluation of machine learning algorithms (e.g., classification, regression, etc.), and the mean (or median) for each class is not provided (even for the edit distance). Consequently, we believe that computing the Fréchet mean of these graph ensembles provides little scientific value to validate our method.

Instead, we present the results of experiments conducted on synthetic datasets generated using ensembles of random graphs. Ensembles of random graphs capture prototypical features of existing real-world networks. Because our theoretical analysis and associated algorithms rely on the (fixed community size) stochastic block model graphs as the "atoms" that are used to approximate any Fréchet mean, we expect that our algorithm will perform well when computing the Fréchet mean of graphs generated by stochastic block models. Our experimental investigation is therefore concerned with the performance of our approach in scenarios where the families of graph ensembles exhibit structural features that are very different from those of the stochastic block models with fixed community sizes.

We illustrate the theoretical analysis of the previous sections with experimental results using various synthetic datasets of graphs. Each data set consists of $N = 50$ graphs on $n = 600$ nodes. We consider three different iid data sets of graphs, $M_1, ..., M_3$, drawn from distributions $\mu_1, ..., \mu_3$ respectively. The distributions have the following high-level descriptions.

μ_1: Barabási-Albert

μ_2: Watts-Strogatz

μ_3: Variable community size stochastic block model

Note that μ_1 and μ_2 induce graphs with vastly different topologies than those generated by stochastic block models and yet we are still able to provide good approximations of the sample Fréchet mean. We discuss the specific parameters for each distribution when applicable in each subsection. For each dataset, we determine the parameters of the stochastic block model whose sample Fréchet

mean is close to the sample Fréchet mean of each dataset, M_i, and compute $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$.

Throughout our experiments we use the $\mathcal{O}(1)$ estimates from Theorem 5 to estimate the expected eigenvalues of graphs sampled according to a stochastic block model and minimize equation (2.15) rather than equation (2.14).

2.6.3 Barabási-Albert approximate sample Fréchet mean

The probability measure in this section is associated with a Barabási-Albert ensemble. The initial graph is fully connected on $m_0 = 5$ nodes and $m = 5$ edges were added at each step. In Fig. 2.1 we reorder the nodes based on their degree for the Barabási-Albert graph to understand better the similarities between an observed graph and the approximate sample Fréchet mean. Our estimate for the number of communities is $c = 12$ using Alg. 1.

Fig. 2.1 is a visual depiction of a graph from M_1 compared to the approximate sample Fréchet mean graph $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$ though we note that there need not be any visual similarity between a graph in M_1 and $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$ since any observation from a distribution μ need not be similar to the mean of μ.

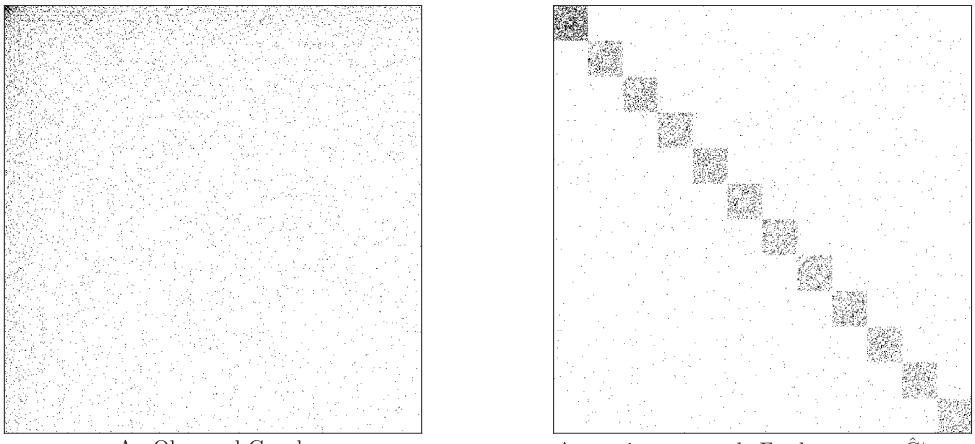

Figure 2.1: Visualization of a graph in M_1 and the approximate sample Fréchet mean of M_1, $\hat{G}^*_{\tilde{N},\mu_{\omega_n f}}$

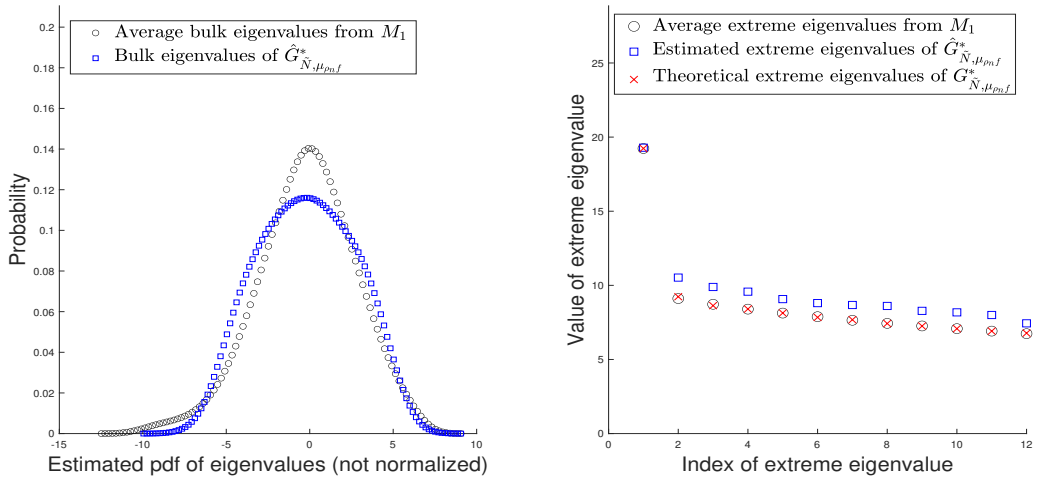

Figure 2.2: **Left**: The average distribution of bulk eigenvalues from M_1 (black). The distribution of bulk eigenvalues of the approximate sample Fréchet mean $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$ (blue). **Right**: The average extreme eigenvalues from M_1 (black). The expected extreme eigenvalues of $G^*_{\tilde{N},\mu_{\omega_n f}}$ from Theorem 5 (red). The extreme eigenvalues of $\widehat{G}_{\tilde{N},\mu^*_{\omega_n f}}$ (blue).

Fig. 2.2 depicts the alignment of the spectra from the approximate sample Fréchet mean with the average spectra of the graphs from set M_1. The misalignment in the largest eigenvalues is due to the finite graph approximation.

The theory presented in Section 2.5 only ensures that the c largest eigenvalues can be well approximated. In Fig. 2.2, we see that the expected eigenvalues, (red markers), are near-perfect estimates of the average extreme eigenvalues. The distinction between the red and blue markers depicts the $\mathcal{O}(1)$ error made when determining the expected eigenvalues of graphs distributed according to a stochastic block model kernel probability measure.

2.6.4 Watts-Strogatz approximate sample Fréchet mean

The parameters for the Watts-Strogatz ensemble (see [102]) are the number of connected nearest neighbors, $K = 22$, and the probability of rewiring, $\beta = 0.7$.

Here we see a nice similarity between the adjacency matrices of the two graphs (see Fig. 2.3). Furthermore, Fig. 2.4 demonstrates the striking spectral similarity between the two graphs in the

extreme and bulk eigenvalues.

The alignment of the bulk eigenvalues from the observed set of graphs and the bulk eigenvalues of $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$ showcases that the graph structure may be entirely determined by the c largest eigenvalues. This has been well known to be true of stochastic block models and in Fig. 2.4 we see evidence that the Watts-Strogatz ensemble might also be characterized by its largest eigenvalues.

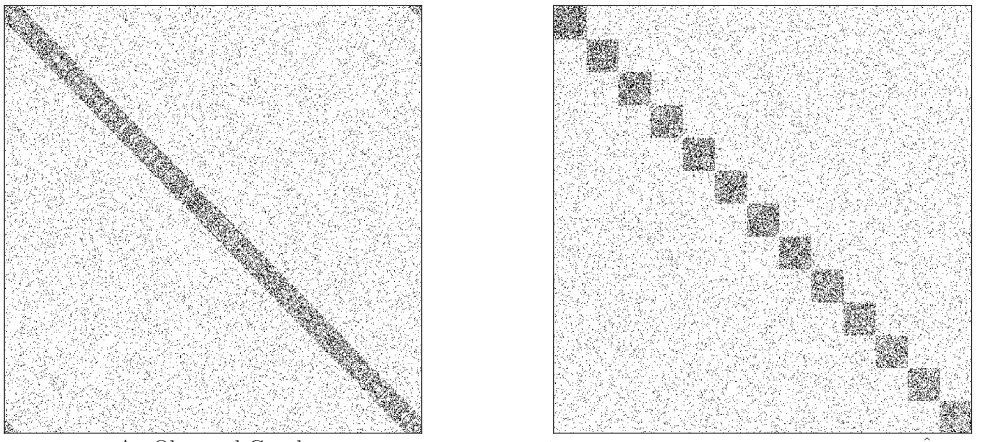

Figure 2.3: Visualization of a graph in M_2 and the approximate sample Fréchet mean of M_2, $\hat{G}^*_{\tilde{N},\mu_{\omega_n f}}$

2.6.5 Variable community size approximate sample Fréchet mean

The probability measure in this section is associated with a variable community sized stochastic block model. The parameters for the stochastic block model are $\boldsymbol{p} = [0.4, 0.5, 0.6, 0.3, 0.37, 0.65, 0, ...]$, $Q_{ij} = 0.08$ for all $i \neq j$, $\boldsymbol{s} = [\frac{160}{600}, \frac{100}{600}, \frac{60}{600}, \frac{120}{600}, \frac{85}{600}, \frac{75}{600}, 0, ...]$. Fig. 2.5 again depicts a visual comparison between a graph from M_3 and the approximate sample Fréchet mean graph $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$.

In spite of the visual difference between the adjacency matrices of a graph from the set M_3 and the graph $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$ respectively (see Fig. 2.5), we again see a striking similarity between the eigenvalues (see Fig. 2.6).

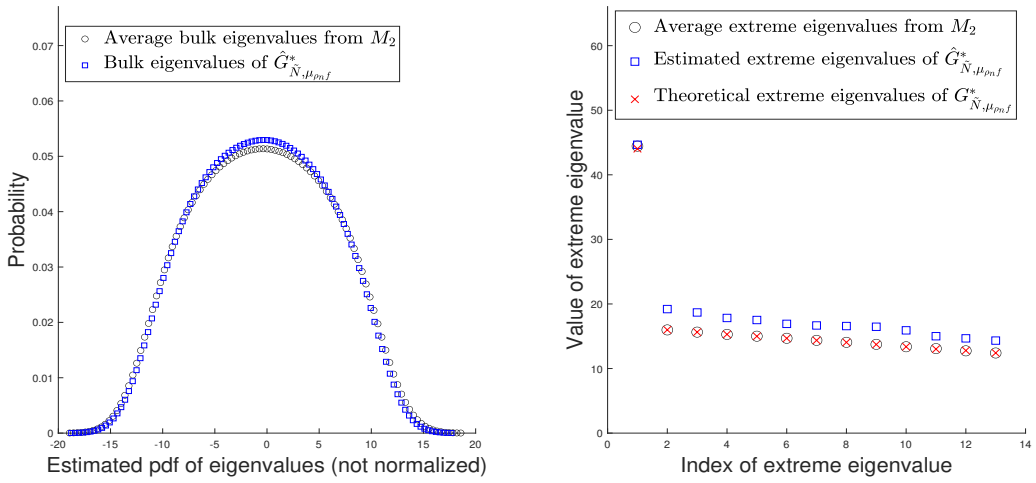

Figure 2.4: **Left**: The average distribution of bulk eigenvalues from M_2 (black). The distribution of bulk eigenvalues of the approximate sample Fréchet mean $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$ (blue). **Right**: The average extreme eigenvalues from M_2 (black). The expected extreme eigenvalues of $G^*_{\tilde{N},\mu_{\omega_n f}}$ (red). The extreme eigenvalues of $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f}}$ (blue).

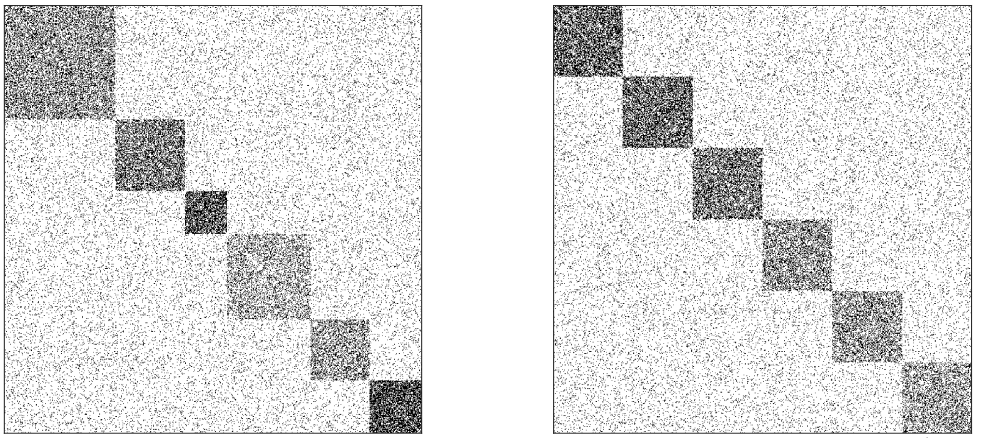

Figure 2.5: Visualization of a graph in M_2 and the approximate sample Fréchet mean of M_2, $\hat{G}^*_{\tilde{N},\mu_{\omega_n f}}$

Note the similarity in the spectra between the graphs despite the obvious difference in the geometry vectors for each stochastic block model. It is for this reason precisely that we are allowed to choose the geometry vector when searching for the canonical stochastic block model kernel that

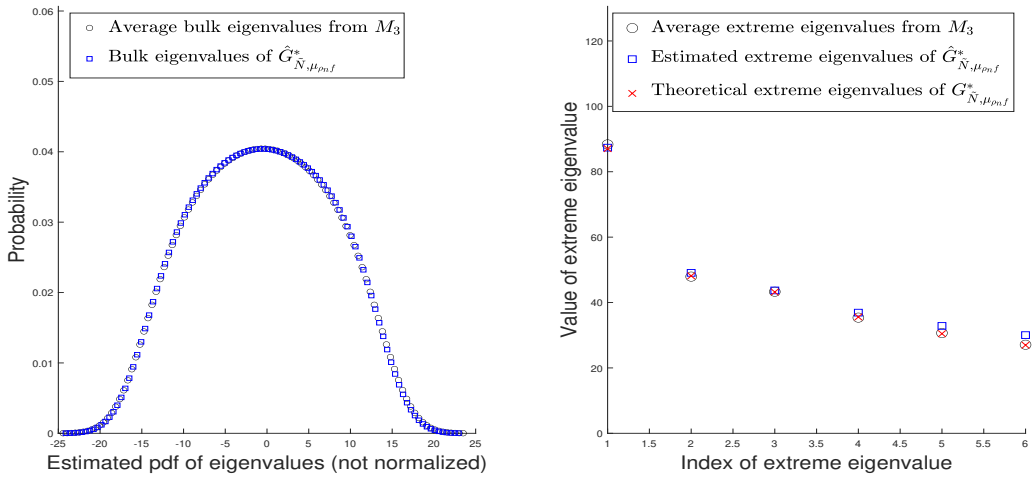

Figure 2.6: **Left**: The average distribution of bulk eigenvalues from M_3 (black). The distribution of bulk eigenvalues of the approximate sample Fréchet mean $\hat{G}^*_{\tilde{N},\mu_{\omega_n f}}$ (blue). **Right**: The average extreme eigenvalues from M_3 (black). The expected extreme eigenvalues of $G^*_{\tilde{N},\mu_{\omega_n f}}$ (red). The extreme eigenvalues of $\hat{G}^*_{\tilde{N},\mu_{\omega_n f}}$ (blue).

solves equation (2.15) as mentioned in remark 4.

2.7 Application to Graph Valued Regression

In this section, we provide an application of the computation of the sample Fréchet mean, the construction of a regression function in the context where we observe a graph-valued random variable that depends on a real-valued random variable. Our approach is based on the theory developed in [85] where we replace the computation of the sample Fréchet mean with our algorithm. Using our notation, we briefly recall the framework of [85]. We consider the following scenario. Let $\mu \in \mathcal{M}(\mathcal{G})$, and let T be a random variable with probability density $\mathbb{P}_T(t)$. We consider the random variable formed by the pair G and T, distributed with the joint distribution formed by the product $\mu \times \mathbb{P}_T(t)$. We wish to compute the regression function

$$\mathbb{E}[G|T=t]. \tag{2.40}$$

The authors in [85] propose to compute the following regression function

$$m(t) = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \, \mathbb{E}_{\mu \times \mathbb{P}_T(t)} \left[s(T,t) d^2(G, G_\mu) \right], \tag{2.41}$$

where the expectation in (2.41) is computed jointly over G_μ distributed according to μ, and T, distributed according to $\mathbb{P}_T(t)$, and the bilinear form $s(T,t)$ is defined by

$$s(T,t) = 1 + (T - \mathbb{E}[T]) \left[\operatorname{var}[T] \right]^{-1} (t - \mathbb{E}[T]). \tag{2.42}$$

The bilinear form, $s(T,t)$, plays the role of a kernel, returning the location of t with respect to the location, $\mathbb{E}[T]$, and scale, $\operatorname{var}[T]$, of T. The regression function $m(t)$ returns a kernel estimate of the linear regression function by summing over all the possible pairs (G_μ, T).

The sample estimate of equation (2.41) is the natural estimate where each unknown term is replaced with the sample alternative as

$$\hat{m}(t) = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \sum_{k=1}^{N} s_{k,N}(t) d^2(G, G^{(k)}) \tag{2.43}$$

$$s_{k,N}(t) = 1 + (t_k - \bar{T}) \hat{V} (t - \bar{T}). \tag{2.44}$$

Here we have used $\bar{T}$ and $\hat{V}$ as the sample estimate of the mean and variance of T. The objective in (2.43) can be interpreted as a weighted sample Fréchet mean with weight function $s_{k,N}(t)$. Assume for all t that the graph, $\hat{m}(t)$, satisfies the conditions for Theorem 1. This implies the existence of a sequence of stochastic block model kernels depending on t, $\mu_{\omega_n f;t}$, such that, for sufficiently large n,

$$\lim_{N \to \infty} d_{A_c}(\hat{m}(t), G^*_{N, \mu_{\omega_n f;t}}) < \epsilon \quad a.s. \tag{2.45}$$

where $G^*_{N, \mu_{\omega_n f;t}}$ denotes the sample Fréchet mean of $\{G_t^{(k)}\}_{k=1}^N$, an iid sample distributed according to $\mu_{\omega_n f;t}$. For each t we may compute $G^*_{N, \mu_{\omega_n f;t}}$ using Alg. 2 as an approximation to the graph $\hat{m}(t)$.

2.7.1 Experimental validation for graph valued regression

We validate the computation of the regression with numerical simulation. We first generate a synthetic data set of graphs by allowing the parameters of the stochastic block model to vary with

time. For simplicity we hold the nonzero entries of $\boldsymbol{Q}$ and $\boldsymbol{s}$ fixed but allow $\boldsymbol{p}$ to vary for $t \in [0,1]$ as

$$\omega_n \boldsymbol{p}(t) = \begin{bmatrix} 0.1+0.1t \\ 0.2+0.15t \\ 0.35+0.2t \\ 0 \\ \vdots \end{bmatrix}, \boldsymbol{s}(t) = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \\ 0 \\ \vdots \end{bmatrix}, \omega_n q = 0.08. \tag{2.46}$$

For $T \sim unif(0,1)$, the distribution over $\mathcal{G}$ is given as $\mu_{\omega_n f;T}$ where $f(x,y;\boldsymbol{p}(T),\boldsymbol{Q},\boldsymbol{s})$ is a canonical stochastic block model kernel. For each sample from $unif(0,1)$ there is a corresponding sample from the stochastic block model. By construction we know the number of communities in the observed graphs will be constant at $c=3$ dictating the number of non-zero entries of $\boldsymbol{p}$ we allow to vary when searching for the stochastic block model kernel in equation (2.45).

We take $n = 600$ and $N = 30$ samples for the sample set $M = \{(t_k, G^{(k)})\}_{k=1}^{30}$ in the experiment and approximate the value of $\hat{m}(t)$ at six different times. For $t' \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ we compute $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f;t'}}$ using Alg. 2.

In an effort of visualization, since we are unable to plot a graph G on the y-axis, we plot in Fig. 2.7 the largest three eigenvalues of the adjacency matrices of graphs in M (marked by a $\bullet$) and the largest three eigenvalues of $\widehat{G}^*_{\tilde{N},\mu_{\omega_n f;t'}}$ (marked by an $\times$). A vertical line of points in Fig. 2.7 identifies the largest three eigenvalues of a single graph. The most significant part of Fig. 2.7 is the construction of a graph that fits the linear regression of each of the largest c eigenvalues simultaneously. To our knowledge, this is the first graph valued linear regression line with respect to a spectral distance.

2.8 Application to K-means clustering.

This section provides another application of the sample Fréchet mean, K-means clustering. We first briefly introduce the clustering problem and the K-means objective (see [89] for details).

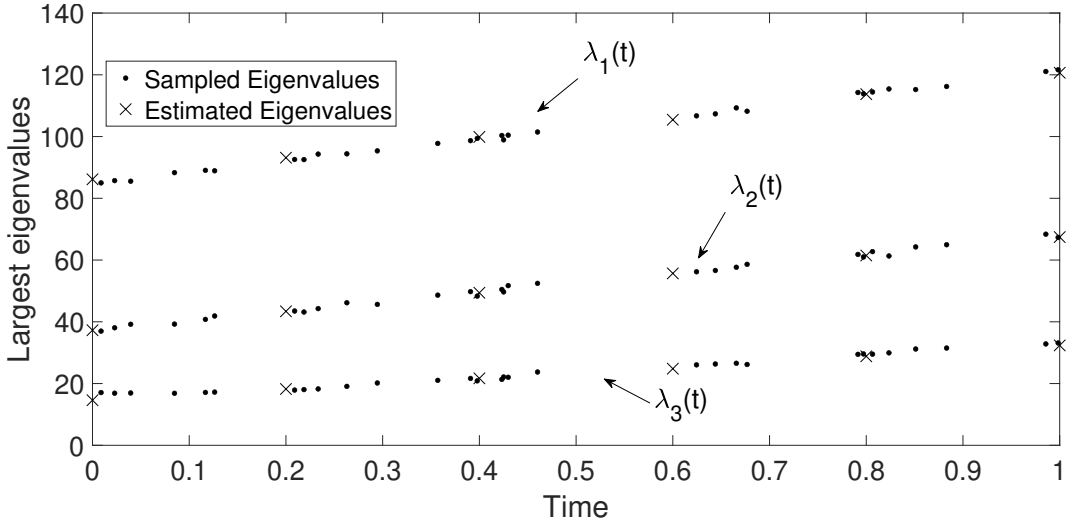

Figure 2.7: Recovered eigenvalues

2.8.1 Setup.

Given a set of graphs $M = \{G^{(k)}\}_{k=1}^{N}$ we seek to partition the data into disjoint sets $\mathcal{M}_1, ..., \mathcal{M}_K$ under the condition that $\cup_{j=1}^{K} \mathcal{M}_j = M$ where each $\mathcal{M}_j$ is represented by its sample Fréchet mean graph

$$G_j^* = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \frac{1}{|M_j|} \sum_{G' \in \mathcal{M}_i} d_{A_c}^2(G, G') \tag{2.47}$$

with adjacency matrix $\boldsymbol{A}_j^*$. We seek to minimize the classic K-means objective function defined as

$$h_0\left((\mathcal{M}_1, ..., \mathcal{M}_K); \{G^{(k)}\}_{k=1}^{N}\right) = \sum_{j=1}^{K} \sum_{G \in \mathcal{M}_j} d_{A_c}^2(G, G_j^*) = \sum_{j=1}^{K} \sum_{G \in \mathcal{M}_j} ||\sigma_c(\boldsymbol{A}) - \sigma_c(\boldsymbol{A}_j^*)||_2^2, \tag{2.48}$$

where $\boldsymbol{A}$ denotes the adjacency matrix of $G \in \mathcal{M}_j$, by determining the optimal disjoint sets $\mathcal{M}_1, ..., \mathcal{M}_K$. Observe that the objective function, h_0, depends on the sets $\mathcal{M}_j$ only through the largest c eigenvalues of $\boldsymbol{A}_j^*$, the adjacency matrix of the representative sample Fréchet mean graph. Theorem 2 states that the eigenvalues of $\boldsymbol{A}_j^*$, $\sigma_c(\boldsymbol{A}_j^*)$, are arbitrarily close to the arithmetic average of the observed set of eigenvalues,

$$\bar{\boldsymbol{\lambda}}^{*,(j)} = \frac{1}{|\mathcal{M}_j|} \sum_{G \in \mathcal{M}_j} \sigma_c(\boldsymbol{A}) \tag{2.49}$$

when the graphs are arbitrarily large. Therefore, to determine the correct disjoint sets $\mathcal{M}_1, ..., \mathcal{M}_K$, we can quickly determine the correct partition of our data by representing the sets $\mathcal{M}_j$ with $\bar{\boldsymbol{\lambda}}^{*,(j)}$, rather than G_j^*, and optimizing the following approximate objective function

$$h_{0,approx}\left((\mathcal{M}_1, ..., \mathcal{M}_K); \{G^{(k)}\}_{k=1}^N\right) = \sum_{j=1}^{K} \sum_{G \in \mathcal{M}_j} ||\sigma_c(\boldsymbol{A}) - \bar{\boldsymbol{\lambda}}^{*,(j)}||_2^2 \tag{2.50}$$

where $\boldsymbol{A}$ is the adjacency matrix of graph $G \in \mathcal{M}_j$. This allows us to avoid the costly computation of determining G_j^* each time the sets $\mathcal{M}_1, ..., \mathcal{M}_K$ are updated in the classic K-means algorithm which we employ with one minor alteration that we discuss in the following section.

2.8.2 Algorithm.

In our implementation of the classic K-means algorithm (see Alg. 3), we consider that when comparing graphs with differing numbers of extremal eigenvalues, there is no obvious choice of c for the pseudo-distance d_{A_c}. The particular difficulty in the K-means algorithm is to compare a graph from $G \in \mathcal{M}_j$ to the sample Fréchet mean graph of the set $\mathcal{M}_r$, given as G_r^*, for $j \neq r$ when the number of extreme eigenvalues from the adjacency matrices of G and G_r^* differ. To be consistent with the work in Section 2.5, whenever we compare a graph G to a mean graph G_j^*, we elect to choose c to be the number of extreme eigenvalues present in the mean graph G_j^*. Therefore, the choice of c is updated every time the sets $\mathcal{M}_1, ..., \mathcal{M}_K$ are updated as this affects the number of extreme eigenvalues in the adjacency matrix of the mean graph G_j^*.

Since this section is merely showcasing an application of our theory, we do not make any attempt to estimate K, the number of clusters in our data set. Below we present our implementation of the classic K-means algorithm. The differences between the standard implementation of K-means and Alg. 3 is due to the metric d_{A_c} and the observation that the choice of c needs to be handled carefully for each cluster.

Algorithm 3 K-means clustering

Require: Set of graphs, $M = \{G^{(k)}\}_{k=1}^{N}$

1: For each $G^{(k)}$ compute and store $\boldsymbol{\lambda}^{(k)} = \sigma(\boldsymbol{A}^{(k)})$

2: Choose K, the number of expected clusters

3: Randomly initialize class assignments, $\mathcal{M}_1, ..., \mathcal{M}_K$

4: **while** Class assignments are changing between iterates **do**

5: For each $\mathcal{M}_j$, estimate c_j as in Alg. 1

6: For each $\mathcal{M}_j$ compute $\bar{\boldsymbol{\lambda}}^{*,(j)} = \frac{1}{|\mathcal{M}_j|} \sum_{G \in \mathcal{M}_j} \boldsymbol{\lambda}(1 : c_j)$

7: For each $\boldsymbol{\lambda}^{(k)}$, and for each j, compute $d_j = \frac{1}{\sqrt{c_j}} \sqrt{\sum_{i=1}^{c_j} (\lambda_i^{(k)} - \bar{\lambda}_i^{*,(j)})^2}$

8: For each $G^{(k)}$ update the class assignment to $j_{new} = \underset{j=1,...,K}{\operatorname{argmin}}\, d_j$

9: **end while**

10: For each $\mathcal{M}_j$ compute G_j^*, the sample Fréchet mean via Alg. 2

11: **Return:** each $\mathcal{M}_j$ and G_j^*

2.8.3 Data and Results

In this section, our data set consists of a mixture of 50 Barabási-Albert (BA) graphs (parameters: $m_0 = 12$ and $m = 12$ edges added at each step), 50 Watts-Strogatz (WS) graphs (parameters: $P = 12$ nearest neighbors with $\beta = 0.4$ probability of rewiring), and 50 stochastic block model (SBM) graphs (parameters: intra-community connection probabilities $\boldsymbol{p} = [0.14, 0.16, 0.1746, 0.2, 0.22, 0.24]$ and constant intercommunity connection probability $q = 0.01$) for a total of 150 graphs. The parameter choices for each ensemble were made such that the number of edges in each graph is relatively consistent. Across 100 different random initializations, Alg. 3 had an accuracy of 0.89 along with the following average confusion matrix. The variance of each element in the average confusion matrix below is denoted by an associated number in parentheses, e.g. $(x.xx)$.

	SBM True	BA True	WS True
SBM Pred.	1 (0.00)	0	0
BA Pred.	0	0.9 (0.0909)	0.19 (0.0909)
WS Pred.	0.23 (0.1789)	0	0.77 (0.1789)

One possible explanation for the misclassification of the Barabási-Albert and Watts-Strogatz graphs comes from the difference in the number of extreme eigenvalues of the adjacency matrices of sample Fréchet means for the respective sets of graphs. For the 50 Barabási-Albert graphs, Alg. 1 estimates that there are $c = 14$ extreme eigenvalues in the adjacency matrix of the sample Fréchet mean whereas for the 50 Watts-Strogatz graphs, Alg. 1 estimates that $c = 34$.

Since the graphs are assigned to a new partition based on a normalized 2-norm between the largest c_j eigenvalues at each step, and because the normalization constant varies when comparing a graph G to the sample Fréchet mean graph from $\mathcal{M}_j$ versus $\mathcal{M}_p$ for $j \neq p$, we find that there exist random initializations such that an empty partition is recovered upon the termination of the K-means algorithm. In this event, all Barabási-Albert graphs are labeled as Watts-Strogatz or vice-versa.

While we acknowledge that fixing c to be constant for each j results in a more numerically stable algorithm, we find that implementing a dynamic choice of c provides a more honest comparison of the graphs in the dataset (see the discussion after Alg. 1). One may still implement a K-means clustering algorithm and recover labels for each graph using a fixed c and estimate the number of extreme eigenvalues of the adjacency matrices of sample Fréchet mean graphs for each recovered partition only at the end of the algorithm, but we have no theoretical justification for this method.

2.9 Conclusion.

In the area of statistical analysis of graph-valued data, determining an average graph is a point of priority among researchers. Throughout this chapter, we have shown that when considering the metric d_{A_c}, it is possible to determine an approximation to the sample Fréchet mean.

How this approximate sample Fréchet mean is utilized is up to the researcher's discretion. In Section 2.7 we explore one motivating idea that utilizes the Fréchet mean, termed Fréchet regression in the work in [85]. This is but one example of the utility of the Fréchet mean graph, another interesting application of this graph is to push further the work in [88] which introduces a centered random graph model to capture the variance of a set of observations around a mean graph.

Chapter 3

Probability density estimation for large graphs

3.1 Introduction.

The process of generating graphs with pre-specified structures observed from real-world networks has led to the advent of various graph ensembles, such as the Barábasi-Albert model, the Watts-Strogatz model, and the stochastic block model. Each ensemble of graphs guarantees that a pre-specified structure exists in the graph with high probability; however, the specifics of each pre-specified structure may vary significantly within the ensemble depending on the initial parameters.

In this work, we consider the problem in graph generation called density estimation. Given a set of graphs with a distribution of structures, we seek to generate new graphs according to this distribution. We utilize the sample Fréchet mean graph and the sample total Fréchet variance of the data to inform the parameters of our generative model. Notably, because our work is always performed with respect to a distance (a requirement to determine the Fréchet mean and variance), we consider the spectral information captured by the adjacency matrix of two graphs to determine their similarity, specifically the ℓ_2 norm between the largest eigenvalues of the adjacency matrices of the observed graphs.

We consider sets of simple graphs with n vertices that have an edge density, ρ_n, that satisfies

$$\rho_n = \omega(n^{-2/3}). \tag{3.1}$$

We also note that the vertex set must be sufficiently large, and the method used in this work will perform poorly for sets of small graphs.

We approach the problem of density estimation by considering the following two ideas: (1) there exists a stochastic block model whose Fréchet mean graph has an adjacency matrix where the largest eigenvalues are arbitrarily close to that of the sample Fréchet mean graph; (2) we may adjust the variance of the recovered stochastic block model by defining a distribution on the parameters. From this perspective, we may align a distribution's mean and variance with the sample mean and sample variance from the data set of graphs.

3.2 State of the Art.

Generative models for graphs have a long history. Popular ensembles of graphs aim to capture various observed real-world phenomenon such as the presence of "hubs" [2], the small-world phenomenon [102], community structure [51], or specific subgraphs [64]. Popular variations of such ensembles further generalize the structures captured by the ensemble (see for instance the degree corrected stochastic block model [72], in-homogeneous Erdős-Rényi models [12, 22], or exponential random graph models [87]). A new method using neural networks, called deep generative models, seeks to model the structures of observed graphs without pre-specifying the structures and instead learns relevant structures in the observed graphs (see [16, 47] for reviews on these models).

While each ensemble captures various structural phenomena, the graph generation process depends on a set of initial parameters. A data-driven generative model will seek to infer the correct value of these parameters conditioned on some set of observations.

Current work in this field is offered by Lunagomez et al. [88], wherein the authors specify a generative process that distributes graphs about the Fréchet mean graph of an observed set. In this work, the Fréchet mean graph is determined with respect to the Hamming distance, but the ideas generalize to any Fréchet mean graph (i.e., for any choice of distance assuming one can compute it). A mixture model with respect to the Hamming distance is proposed in [59], where multiple mean graphs are considered as defining the centers of each mixture component, and graphs are sampled within each mixture according to the observed deviations from the mean graph. When considering a Markov random graph model (or its generalization, the exponential random graph model), various

procedures have been proposed to determine the parameters, e.g., maximum pseudo-likelihood estimation [95] and Monte Carlo Markov chain maximum [53, 92].

All parameter estimations in the above models compare graphs using a Hamming distance. The Hamming distance identifies local changes in the connectivity structure between nodes. At times, the Hamming distance can be used to detect global phenomena in the graphs, i.e., relating the presence of triangles to the density of Erdős-Rényi random graph models; however, most commonly, the Hamming distance measures only the local connectivity.

Taking a similar perspective as Lunagomez et al. [88], we determine the parameters of our model using the sample moments of the observed data. However, as an extension to the work in [88], here we consider both the first and second moments of the data. The choice of metric is crucial to the location and spread of graphs as each metric induces a different mean graph and spread about the mean graph. The Fréchet mean and Fréchet variance have been analyzed with respect to the edit distance (see [19, 41, 56, 55, 60, 88] and references therein); however, we aim to capture the mean and variability of global structures within the data set of graphs.

To this end, we consider a distance that can detect such structural changes (e.g., community structure [1, 69], modularity [44]). The adjacency spectral distance, which we define as the ℓ_2 norm of the difference between the spectra of the adjacency matrices of the two graphs of interest [104] exhibits good performance when comparing various types of graphs [103], making it a reliable choice for a wide range of problems. Spectral distances also offer practical advantages as they can inherently compare graphs of different sizes and graphs without known vertex correspondence (see, e.g., [34, 35] and references therein).

In practice, only the c largest eigenvalues are often considered. We still refer to such distances as spectral distances but comparison using the c largest eigenvalues for small values of c allows one to focus on the global structures of the graphs while omitting the local structures [69]. Of notable importance when considering the distance between the c largest eigenvalues is recent work showcasing how to approximately compute the sample Fréchet mean graph (see prior work in [33]). With access to the sample Fréchet mean graph, a generative model may be centered at this mean

graph.

Further work in the realm of generative modeling when considering the sample moments of the data comes in the form of regression [85] wherein the authors construct a parameterization of a weighted sample Fréchet mean as a generalization of linear regression for metric objects. When considering the second moment of graph-valued data, a theoretical analysis of the Fréchet variance allows the user to construct a test to compare samples of metric-valued objects such as graphs [27].

3.3 Main Contributions.

In this chapter, we introduce a variation on the stochastic block model for graph generation, namely the random-parameter stochastic block model, by allowing the parameters of a stochastic block model graph to vary according to some distribution J where the choice of J at the discretion of the researcher. The need for such a generalization exists since stochastic block models generate graphs in a homogeneous way, meaning there is little variance between graphs sampled according to this model. We showcase that methods that estimate the parameters of a stochastic block model can be used to estimate the moments of the distribution J, which uniquely characterizes the distribution under certain parametric assumptions of J. When J is assumed to be non-parametric, a generalization of kernel density estimation is suggested as a method to infer the distribution J better.

When considering sample sets of graphs, the sample arithmetic mean of the largest eigenvalues of the adjacency matrices of the graphs observed and the corresponding sample covariance is shown to be connected to the sample Fréchet mean graph and total sample Fréchet variance. This indicates that inferring a generative model using the sample mean eigenvalues and sample covariance matrix is equivalent to inferring a model using the sample Fréchet mean and sample total Fréchet variance.

We experimentally verify our results on four different data sets to explore the limitations of the proposed model. We demonstrate that we can recover the parameters of a random-parameter stochastic block model and showcase the impact of the distribution J when considering data not

generated from a random-parameter stochastic block model. We also construct a generative model for real-world data where the graphs in the sample are collected according to face-to-face connections formed at a primary school.

3.4 Random parameter stochastic block models

This section introduces the random parameter stochastic block model to generalize the classic stochastic block model. This generalization is needed when considering density estimation due to the following observation. For a fixed geometry vector $\boldsymbol{s}$, the limiting distribution of the largest c eigenvalues, $\lambda_i(\boldsymbol{A}_{\mu_{\omega_n,\boldsymbol{p},q,\boldsymbol{s}}})$ for $i = 1, ..., c$, have a dependency between the location and its scale. This fact can be seen clearly in [38] where it is shown for an Erdős-Rényi random graph with parameters n and p, that

$$\lambda_1 \xrightarrow{d} N\left((n-2)p + 1, 2p(1-p) + \mathcal{O}(\frac{1}{\sqrt{n}})\right). \tag{3.2}$$

Equation (3.2) shows that a different choice of p affects both the location and scale of the limiting distribution. The same phenomenon can be seen in previous studies [22, 30, 97], where the limiting distribution is derived for inhomogeneous Erdős-Rényi random graphs. The dependency between the location and scale parameters for the largest eigenvalues of stochastic block model graphs is fully expected due to the Bernoulli process that defines the probability of an edge existing in a graph. The implications of these observations suggest that when attempting to model a graph-valued data set, the stochastic block model cannot simultaneously capture notions of location and scale.

By allowing the parameters of a stochastic block model to be random, several degrees of freedom are introduced when considering the distribution of the eigenvalues. Most notably, allowing randomness in the parameter space allows the user to increase the variance of the eigenvalues of the adjacency matrices from a classic stochastic block model while preserving their expected value. We first introduce the random parameter stochastic block model and the distribution of the eigenvalues of graphs sampled in this manner. We then discuss methods by which we estimate the parameters

of this model given a sample set of data (see Alg. 4).

We take a random graph G_μ to be distributed according to a stochastic block model, $\mu_{\omega_n,\boldsymbol{p},q,\boldsymbol{s}}$, with unknown parameter $\boldsymbol{p}$ that is distributed according to some distribution J. Note that by the definition of $f(x,y)$, see Definition 11, the support of J is $[0,1]^c$.

By allowing $\boldsymbol{p}$ to vary according to J, the result is a distribution over $\mathcal{G}$ where the intra-community strengths of the graphs sampled follow a multivariate distribution J. The associated probability measure given some J is denoted by

$$\mu_{\omega_n,J,q,\boldsymbol{s}} \in \mathcal{M}(\mathcal{G}), \tag{3.3}$$

which is similar to the probability measure for a general stochastic block model, except the parameter of intra-community strengths J replaces $\boldsymbol{p}$.

Remark 10 Observe that when $J = \delta(\boldsymbol{p} - \boldsymbol{p}^*)$ for some fixed $\boldsymbol{p}^*$, then

$$\mu_{\omega_n,J,q,\boldsymbol{s}} = \mu_{\omega_n,\boldsymbol{p}^*,q,\boldsymbol{s}} \tag{3.4}$$

which is the typical stochastic block model probability measure with an intra-community probability of connection given by $\boldsymbol{p}^*$.

Because we aim to perform density estimation in $\mathcal{G}$ with respect to d_{A_c} it, is necessary to understand the behavior of the eigenvalues of graphs distributed according to $\mu_{\omega_n,J,q,\boldsymbol{s}}$. By mapping $G_{\mu_{\omega_n,J,q,\boldsymbol{s}}} \mapsto \sigma_c(\boldsymbol{A}_{\mu_{\omega_n,J,q,\boldsymbol{s}}})$, the resulting distribution of the largest c eigenvalues is denoted by H_n and has a probability distribution given by

$$p_{H_n}(\boldsymbol{\lambda}) = \int p_{\mu_{\omega_n,J,q,\boldsymbol{s}}}(\sigma_c(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}) p_J(\boldsymbol{p}) d\boldsymbol{p} \tag{3.5}$$

where the first moment and normalized second moments of H_n are

$$\mathbb{E}_{H_n}[\boldsymbol{\lambda}] = \mathbb{E}_J[\mathbb{E}_\mu[\sigma_c(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}]]\,, \tag{3.6}$$

$$\mathrm{Cov}_{H_n}\left(\frac{\lambda_i}{\sqrt{\omega_n}}, \frac{\lambda_j}{\sqrt{\omega_n}}\right) = \mathrm{Cov}_J(\mathbb{E}_\mu\left[\frac{1}{\sqrt{\omega_n}}\lambda_i(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right], \mathbb{E}_\mu\left[\frac{1}{\sqrt{\omega_n}}\lambda_j(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right]) \tag{3.7}$$

$$+ \mathbb{E}_J\left[\mathrm{Cov}_\mu\left(\frac{1}{\sqrt{\omega_n}}\lambda_i(\boldsymbol{A}), \frac{1}{\sqrt{\omega_n}}\lambda_j(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right)\right]. \tag{3.8}$$

Theorem 8 and Corollary 2 show two methods to determine the quantities

$$\mathbb{E}_\mu[\sigma_c(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}]\,, \tag{3.9}$$

$$\mathrm{Cov}_\mu\left(\frac{1}{\sqrt{\omega_n}}\lambda_i(\boldsymbol{A}), \frac{1}{\sqrt{\omega_n}}\lambda_j(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right), \tag{3.10}$$

from equations (3.6) and (3.8) in terms of the parameters of a stochastic block model.

Let f be a canonical stochastic block model kernel function and let L_f be the associated linear integral operator with eigenfunctions $r_i(x)$ and eigenvalues denoted by $\theta_i = \lambda_i(L_f)$. Assume that $\omega_n = \omega(n^{-2/3})$ and that $\lim_{n\to\infty}\omega_n = 0$. Because μ is taken to be a stochastic block model kernel probability measure with parameters $\omega_n, \boldsymbol{p}$, q, and $\boldsymbol{s}$, we denote the adjacency matrix of a random graph as $\boldsymbol{A}_\mu = \boldsymbol{A}_{\mu_{\omega_n,\boldsymbol{p},q,\boldsymbol{s}}}$ where we have suppressed all the subscripts.

Theorem 8 (Chakrabarty, Chakraborty, Hazra 2020)

$$\left(\omega_n^{-1/2}(\lambda_i(\boldsymbol{A}_\mu) - \mathbb{E}_\mu[\lambda_i(\boldsymbol{A}_\mu)])\right) \xrightarrow{d} (Z_i : 1 \le i \le c), \tag{3.11}$$

where the right-hand side is a multivariate normal random vector in $\mathbb{R}^c$ with zero mean and

$$\mathrm{Cov}(Z_i, Z_j) = 2\int_0^1\int_0^1 r_i(x)r_i(y)r_j(x)r_j(y)f(x,y)dxdy, \tag{3.12}$$

for all $1 \le i, j \le c$. The first order behavior of $\mathbb{E}\left[\lambda_i(\boldsymbol{A}_\mu)\right])$ is given by the following: For every $1 \le i \le c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}})\right] = \lambda_i(\boldsymbol{B}) + \mathcal{O}(\sqrt{\omega}_n + \frac{1}{n\omega_n}), \tag{3.13}$$

where $\boldsymbol{B}$ is a $c \times c$ symmetric deterministic matrix defined by

$$b_{j,l} = \sqrt{\theta_j\theta_l}n\omega_n\boldsymbol{e}_j^T\boldsymbol{e}_l + \theta_i^{-2}\sqrt{\theta_j\theta_l}(n\omega_n)^{-1}\boldsymbol{e}_j^T\mathbb{E}\left[(\boldsymbol{A} - \mathbb{E}\left[\boldsymbol{A}\right])^2\right]\boldsymbol{e}_l + \mathcal{O}(\frac{1}{n\omega_n}), \tag{3.14}$$

and $\boldsymbol{e}_j$ is a vector with entries $\boldsymbol{e}_j(k) = \frac{1}{\sqrt{n}}r_j(\frac{k}{n})$ for $1 \le j \le c$.

Proof of Theorem 8 This is a compilation of Theorems 2.3 and 2.4 from [22].

We offer the following corollary as a minor simplification to the results of [22] which allows for the estimation of $\mathbb{E}\left[\lambda_i(\boldsymbol{A}_\mu)\right]$ and the computation of $\mathrm{Cov}(Z_i, Z_j)$ in a different manner when considering

stochastic block models. First, define the matrices

$$\boldsymbol{M} = \begin{bmatrix} s_1 p_1 & \sqrt{s_1 s_2} q & \dots & \sqrt{s_1 s_c} q \\ \sqrt{s_2 s_1} q & s_2 p_2 & \dots & \sqrt{s_2 s_c} q \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{s_c s_1} q & \sqrt{s_c s_2} q & \dots & s_c p_c \end{bmatrix} \quad \boldsymbol{M}_f = \begin{bmatrix} p_1 & q & \dots & q \\ q & p_2 & \dots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \dots & p_c \end{bmatrix} \tag{3.15}$$

where ν_k and $\boldsymbol{v}_k$ are the eigenvalues and eigenvectors of $\boldsymbol{M}$ respectively.

Corollary 2

$$\left(\omega_n^{-1/2}(\lambda_i(\boldsymbol{A}_\mu) - \mathbb{E}_\mu[\lambda_i(\boldsymbol{A}_\mu)])\right) \xrightarrow{d} (Z_i : 1 \le i \le c), \tag{3.16}$$

where the right-hand side is a multivariate normal random vector in $\mathbb{R}^c$ with zero mean and

$$\mathrm{Cov}(Z_i, Z_j) = 2\left(\boldsymbol{v}_i. * \boldsymbol{v}_j\right)^T \boldsymbol{M}_f \left(\boldsymbol{v}_i. * \boldsymbol{v}_j\right), \tag{3.17}$$

for all $1 \le i, j \le c$ and $.*$ denotes the component-wise product of the vectors.

The first order behavior of $\mathbb{E}\left[\lambda_i(\boldsymbol{A}_\mu)\right])$ is given by the following: For every $1 \le i \le c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_\mu)\right] = \lambda_i(\boldsymbol{B}^*) + \mathcal{O}(\sqrt{\omega_n}) \tag{3.18}$$

where $\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}$, the components of which are given as

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b_{j,l}^{*,(1)} = \begin{cases} \nu_j n \omega_n & \text{if } j = l \\ 0 & \text{if } j \ne l \end{cases} \tag{3.19}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b_{j,l}^{*,(2)} = \nu_i^{-2} \sqrt{\nu_j \nu_l} \sum_{k=1}^{c} \nu_k \sum_{m=1}^{c} \frac{1}{\sqrt{s_m}} \boldsymbol{v}_j(m) \boldsymbol{v}_l(m) \boldsymbol{v}_k(m) \sum_{w=1}^{c} \sqrt{s_w} \boldsymbol{v}_k(w). \tag{3.20}$$

Proof of Corollary 1 We show these results in Appendix B.3. The proof technique observes that the estimation of the expected eigenvalues in Theorem 8 is given in terms of the inner products between discretized eigenfunctions and a discretized operator. Because every term is piecewise Lipschitz, these discretizations converge to their respective limits at a rate $\frac{1}{n}$, which allows one to replace the vectors $\boldsymbol{e}_j$ with the corresponding eigenfunctions. Then, because the eigenfunctions of stochastic block model graphs are piecewise constant, we represent these quantities with the eigenvectors of the matrix $\boldsymbol{M}$.

The prior corollary discusses the convergence in distribution of the random variable

$$\left(\omega_n^{-1/2}(\lambda_i(\boldsymbol{A}_\mu) - \mathbb{E}_\mu[\lambda_i(\boldsymbol{A}_\mu)])\right). \tag{3.21}$$

In general, we do not have convergence of the second moment as a result of convergence in distribution. However, we assume that for a large value of n, the quantity $\mathrm{Cov}(Z_i, Z_j)$ provides a good estimate of the term

$$\mathrm{Cov}_\mu\left(\frac{1}{\sqrt{\omega_n}}\lambda_i(\boldsymbol{A}), \frac{1}{\sqrt{\omega_n}}\lambda_j(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right). \tag{3.22}$$

With this assumption, Corollary 2 provides a relationship between the expected value and covariance of H_n in terms of the parameters of the stochastic block model.

For general stochastic block models, equations (3.17) and (3.18) are not analytic in terms of the parameters of the model even when q is taken to be constant. A consequence is that the estimation of the moments of J becomes non-trivial. We consider a regime where we define $p_{\min} = \min_{i=1,\ldots,c} \boldsymbol{p}$ along with a fixed parameter $\epsilon \ll 1$ and set $q = \epsilon p_{\min}$. Under these conditions, analytic expressions for (3.17) and (3.18) are determined up to $\mathcal{O}(\epsilon^2)$ in the following lemma. The implications of these analytic expressions result in the following lemma to determine the first and second moments of J.

Lemma 2 Let $\boldsymbol{M}$ and $\boldsymbol{M}_f$ be as defined in equation (3.15). Assume $q = \epsilon p_{\min}$, where $p_{\min} = \min_{i=1,\ldots,c} \boldsymbol{p}$ and $\epsilon \ll 1$.

$$\left|\frac{\mathbb{E}\left[\lambda_i(\boldsymbol{A})\right]}{n\omega_n s_i} - p_i\right| = \mathcal{O}\left(\epsilon^2 p_{\min}^2\right) + \mathcal{O}\left(\frac{t_i(\boldsymbol{p})}{n\omega_n s_i}\right), \tag{3.23}$$

$$\left|\mathrm{Cov}(Z_i, Z_j) - \begin{cases} 2p_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}\right| = \mathcal{O}(\epsilon^2 p_{\min}^2), \tag{3.24}$$

where $t_i(\boldsymbol{p})$ is a bounded function of the parameters $\boldsymbol{p}$ and Z is defined in Theorem 8.

Proof of Lemma 2 The proof is in Appendix B.4

When conditioned on an observation $\boldsymbol{p}$, Lemma 2 shows how to determine a relative error in the parameters. Utilizing these results, we now show how to determine the relative error in the first two moments of J with the following lemma.

Lemma 3 Let $\boldsymbol{P}_J$ be an observation from J with components P_i, let $P_{\min} = \min_{i=1,\ldots,c} P_i$ and define $q = \epsilon P_{\min}$ where $\epsilon \ll 1$.

$$\frac{\mathbb{E}_{H_n}[\lambda_i]}{n\omega_n s_i} = \mathbb{E}_J[P_i] + \mathcal{O}(\epsilon^2) + \mathcal{O}\left(\frac{1}{n\omega_n}\right) \tag{3.25}$$

$$\frac{\mathrm{Cov}_{H_n}(\lambda_i, \lambda_j)}{n^2\omega_n^2 s_i s_j} - \begin{cases} \frac{2\mathbb{E}_{H_n}[\lambda_i]}{n^3\omega_n^2 s_i^3} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{3.26}$$

$$= \mathrm{Cov}_J\left(P_i + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}\left(\frac{t_i(\boldsymbol{p})}{n\omega_n}\right), P_j + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}\left(\frac{t_j(\boldsymbol{p})}{n\omega_n}\right)\right) + \mathcal{O}\left(\frac{\epsilon^2}{n^2\omega_n}\right) \tag{3.27}$$

where $t_i(\boldsymbol{p})$ is a bounded function of the parameters for each i.

Proof of Lemma 3 The proof is in Appendix B.5.

When the random parameter stochastic block model has a parameter q that satisfies the assumptions of Lemma 3, the associated probability measure is denoted as $\mu_{\omega_n, J, \epsilon, \boldsymbol{s}}$. Lemma 3 suggests the following algorithm to estimate the probability measure $\mu_{\omega_n, J, \epsilon, \boldsymbol{s}}$ given a sample set of graphs.

Algorithm 4 Estimation of $\mu_{\omega_n,J,\epsilon,\boldsymbol{s}}$ given sample data

Require: Set of graphs, $M = \{G^{(k)}\}_{k=1}^{N}$.

1: Compute the eigenvalues of the adjacency matrix of each graph as $\boldsymbol{\lambda}^{(k)} = \sigma_c(\boldsymbol{A}^{(k)})$.

2: Compute the arithmetic average of the c largest eigenvalues as $\bar{\boldsymbol{\lambda}}$.

3: Compute the sample covariance matrix of the c largest eigenvalues as $\hat{\Sigma}$.

4: Determine the average density of the set of graphs as $\bar{\rho}_n$.

5: Set $\boldsymbol{s}$ such that $s_i = \frac{1}{c}$ for all $1 \leq i \leq c$.

6: Set $\omega_n = C\bar{\rho}_n$. Taking $C = 1$ is typically sufficient, a discussion follows in Remark 11.

7: Determine the first moment of J according to

$$\mathbb{E}_J[P_i] = \frac{\bar{\lambda}_i}{n\omega_n s_i}. \tag{3.28}$$

8: Determine the second moment of J according to

$$\mathrm{Cov}_J(P_i, P_j) = \frac{\hat{\Sigma}_{ij}}{n^2\omega_n^2 s_i s_j} - \begin{cases} \frac{2\bar{\lambda}_i}{n^3\omega_n^2 s_i^3} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{3.29}$$

9: Set

$$\epsilon = \frac{1 - \sum_{i=1}^{c} \mathbb{E}_J[P_i]\, s_i^2}{\min\limits_{i=1,\ldots,c} (\mathbb{E}_J[P_i])(1 - \sum_{i=1}^{c} s_i^2)} \tag{3.30}$$

10: Return: $\omega_n, J, \epsilon, \boldsymbol{s}$

Remark 11 The choice of scaling as $\omega_n = \bar{\rho}_n$ is for simplicity. In practice, taking $\omega_n = C\bar{\rho}_n$, where C is a positive constant, will always yield the same result for the following reason. Observe that for graphs sampled according to a kernel probability measure, the probability of an edge existing in the graph is modeled by

$$\mathbb{P}\left(a_{ij}\right) = \text{Bernoulli}\left(\omega_n f(\frac{i}{n}, \frac{j}{n})\right) = \text{Bernoulli}\left((C\omega_n)\,\frac{f(\frac{i}{n}, \frac{j}{n})}{C}\right). \tag{3.31}$$

Implying that the choice $\tilde{\omega}_n = C\omega_n$ and $\tilde{f}(x,y) = \frac{1}{C} f(x,y)$ for any C such that $\tilde{f}(x,y) \in [0,1]$ defines an equivalent probability measure, $\mu_{\omega_n f} = \mu_{\tilde{\omega}_n \tilde{f}}$.

The choice of ϵ is such that in expectation, the density of graphs sampled from the estimated

probability distribution is equal to the sample density of graphs observed. It is derived by setting $\mathbb{E}_J[||f||_1] = 1$. Note that when taking $\omega_n = C\bar{\rho}_n$, ϵ should be chosen such that $\mathbb{E}_J[||f||_1] = \frac{1}{C}$ so the expected density of the graphs is preserved when this is a desired quantity. The choice of ω_n and ϵ is a suggestion; other quantities may exist to suggest a different method of selecting these model parameters.

The work in [33] shows that for arbitrarily large graphs, when taking s_i as a constant, a disconnected stochastic block model with the correct expected eigenvalues can be determined. We discuss in Remark 13 that in practice, an estimate of $\boldsymbol{s}$ that is data set dependent may perform better.

When fitting a distribution to sample data, aligning the sample moments with the mean and variance of the probability measure is a common practice and is the approach taken when fitting a random parameter stochastic block model. In the following section, it is shown that both $\bar{\boldsymbol{\lambda}}$ and $\hat{\Sigma}$ provide a good estimate of the eigenvalues of the sample Fréchet mean graph and the information contained within the total sample Fréchet variance respectively. Another perspective common to parameter estimation is likelihood maximization which is not explored in this chapter.

3.5 The first and second moments of $\mu \in \mathcal{M}(\mathcal{G})$

The first and second moments of a distribution $\mu \in \mathcal{M}(\mathcal{G})$ respectively characterize the mean and spread of the probability measure. The mean and variance for metric spaces were generalized in [42], respectively called the Fréchet mean and total Fréchet variance along with their sample alternatives. In this section, the Fréchet mean and total Fréchet variance are introduced. It is shown that the arithmetic mean of the eigenvalues of the adjacency matrices of a set of graphs, $\{G^{(k)}\}_{k=1}^N$, and the sample covariance of the eigenvalues are closely related to the sample alternatives of the Fréchet mean and total Fréchet variance when n is large.

3.5.1 The first moment: Fréchet mean

We equip the set of graphs, $\mathcal{G}$, defined on n vertices (see Definition 4) with the pseudometric defined by the ℓ_2 norm between the spectra of the respective adjacency matrices, d_{A_c}, (see (1.7)). We consider a probability measure $\mu \in \mathcal{M}(\mathcal{G})$ that describes the probability of obtaining a given graph when we sample $\mathcal{G}$ according to μ. Using d_{A_c}, we quantify the spread of the graphs, and we define a notion of centrality, which gives the location of the expected graph, according to μ.

Definition 15 (Fréchet mean [42]) The Fréchet mean of the probability measure μ in the pseudometric space $(\mathcal{G}, d_{A_c})$ is the set of graphs G^* whose expected distance squared to $\mathcal{G}$ is the minimum,

$$\{G^* \in \mathcal{G}\} = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \, \mathbb{E}_\mu\big[d^2_{A_c}(G, G_\mu)\big] \,, \tag{3.32}$$

where G_μ is a random realization of a graph distributed according to the probability measure μ, and the expectation $\mathbb{E}_\mu\big[d^2(G, G_\mu)\big]$ is computed with respect to the probability measure μ. The analysis in this section applies to any graph in the Fréchet mean set. We therefore write the Fréchet mean as a singleton set and present the sample Fréchet mean as a unique graph rather than the more general set-valued sample Fréchet mean.

The sample Fréchet mean is a natural extension of the population Fréchet mean defined by replacing the measure μ with the empirical measure.

Definition 16 (Sample Fréchet mean [42]) Let $\big\{G^{(k)}\big\}$ $1 \leq k \leq N$ be a set of graphs in $\mathcal{G}$. The sample Fréchet mean is defined by

$$G^*_N = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \, \frac{1}{N} \sum_{k=1}^{N} d^2_{A_c}(G, G^{(k)}). \tag{3.33}$$

Remark 12 The sample Fréchet mean, when considering the distance d_{A_c}, is discussed at length in [33]. Of particular note, the largest c eigenvalues of the adjacency matrix of the sample Fréchet mean graph are estimated arbitrarily well by the arithmetic average of the largest c eigenvalues from the graphs in the sample, namely Theorem 2 in [33].

3.5.2 The second moment: Fréchet variance

With a notion of mean in hand, the second moment, which captures the variability about the mean, follows naturally.

Definition 17 (Total Fréchet variance [42]) The total Fréchet variance of the probability measure μ in the pseudometric space $(\mathcal{G}, d_{A_c})$ is defined as

$$V^*_{tot} = \mathbb{E}_\mu\big[d^2_{A_c}(G^*, G_\mu)\big]\,, \tag{3.34}$$

which is the evaluation of the Fréchet mean objective at the Fréchet mean graph. Similarly, the sample total Fréchet variance is given by evaluating the sample Fréchet mean objective at the sample Fréchet mean graph.

Definition 18 (Sample total Fréchet variance [42]) Let $\left\{G^{(k)}\right\}$ $1 \leq k \leq N$ be a set of graphs in $\mathcal{G}$. The sample total Fréchet variance is defined as

$$V^*_{N,tot} = \frac{1}{N-1}\sum_{k=1}^{N} d^2_{A_c}(G^*_N, G^{(k)}). \tag{3.35}$$

While the (sample) total Fréchet variance applies for any metric object (all it requires is a distance), the adjacency spectral pseudo-metric, d_{A_c}, allows for a more specific definition of variance. As shown in the following lemma, the covariance matrix of the observed eigenvalues captures identical information to (3.35).

For a dataset of graphs $\{G^{(k)}\}_{k=1}^N$, denote the arithmetic mean of the eigenvalues of the adjacency matrices as

$$\bar{\boldsymbol{\lambda}} = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\lambda}^{(k)} \tag{3.36}$$

where $\boldsymbol{\lambda}^{(k)}$ is the vector of c largest eigenvalues of the adjacency matrix, $\boldsymbol{A}^{(k)}$, of graph $G^{(k)}$. Define the sample covariance matrix as

$$\hat{\Sigma} = \frac{1}{N-1}\sum_{k=1}^{N} (\boldsymbol{\lambda}^{(k)} - \bar{\boldsymbol{\lambda}})(\boldsymbol{\lambda}^{(k)} - \bar{\boldsymbol{\lambda}})^T. \tag{3.37}$$

Lemma 4 Let $\{G^{(k)}\}_{k=1}^{N}$ be a sample of graphs with sample Fréchet mean G_N^* and total sample Fréchet variance $V_{N,tot}^*$. Let $\bar{\boldsymbol{\lambda}}$ denote the arithmetic mean of the largest c eigenvalues, and let $\hat{\Sigma}$ be the sample covariance matrix; then,

$$\lim_{n\to\infty} |V_{N,tot}^* - \sum_{i=1}^{c} \hat{\Sigma}_{ii}| = 0 \tag{3.38}$$

Proof of Lemma 4 The proof is in Appendix B.2.

3.6 Conditions for a feasible distribution J

This section explores the situation in which steps 7 and 8 in Algorithm 4 are solvable independent of any assumptions on J. First, we consider Step 7. Because the support of J is a subset of $[0,1]^c$,

$$\frac{\bar{\lambda}_i}{n\omega_n s_i} \in [0,1] \quad \forall i. \tag{3.39}$$

Remark 13 While it is shown in [33] that the size of the communities is arbitrary when the size of the graphs is arbitrarily large, in practice, this condition suggests that an estimate of s_i, which depends on the data, may perform better. Unless otherwise specified in this chapter, we always take $\boldsymbol{s}$ such that $s_1 \geq s_j$ for all $j = 2, ..., c$ and $s_i = s_j$ for all $2 \leq i, j \leq c$.

Interpreting this condition is straightforward; it states that the largest eigenvalue from a block in the stochastic block model cannot exceed the total number of connections available within that block.

Another condition on the feasibility of the moments of J comes from Step 8. An obvious observation is simply that the variance is bounded below by 0,

$$0 \leq Var_J(P_i) \implies 0 \leq \frac{\hat{\Sigma}_{ii}}{n^2\omega_n^2 s_i^2} - \frac{2\bar{\lambda}_i}{n^3\omega_n^2 s_i^3}. \tag{3.40}$$

This results in the following condition on the relationship between the sample variance and the sample mean eigenvalues;

$$0 \leq \frac{\hat{\Sigma}_{ii}}{\omega_n} - \frac{2\bar{\lambda}_i}{n\omega_n s_i} \iff 0 \leq \hat{\Sigma}_{ii} - \frac{2\bar{\lambda}_i}{ns_i}. \tag{3.41}$$

We have expressed the condition in two forms, the first for interpretability and the second motivated by practical implementation.

We consider that a data set of graphs, $\{G^{(k)}\}_{k=1}^{N}$, is homogeneous when the data set has low total sample Fréchet variance and is heterogeneous when the data set has high total sample Fréchet variance. The condition in equation (3.41) is about the homogeneity of the sample set of graphs. It relates the scaled sample variance $\hat{\Sigma}_{ii}$ to the expected variance inherent to the stochastic block model process. Recall that the expected value of the inherent variance of the stochastic block model when $q = \epsilon p_{\min}$ is estimated to the first order as

$$\mathbb{E}_J[\mathrm{Var}(Z_i)] = \mathbb{E}_J[2P_i] + \mathcal{O}(\epsilon^2) \tag{3.42}$$

where Z is defined as in Corollary 2. By rewriting $\mathbb{E}_J[P_i] = \frac{\bar{\lambda}_i}{n\omega_n s_i}$, it is clear that the condition given in equation (3.41) is equivalent to

$$0 \leq \frac{\hat{\Sigma}_{ii}}{\omega_n} - \mathbb{E}_J[\mathrm{Var}(Z_i)]\,. \tag{3.43}$$

If the sample variance of the eigenvalues, $\hat{\Sigma}_{ii}$, is smaller than $\mathbb{E}_J[\mathrm{Var}(Z_i)]$ then the data set of graphs, $\{G^{(k)}\}_{k=1}^{N}$, is considered to be more homogeneous with respect to the eigenvalues than a set of graphs sampled according to a stochastic block model with parameters $p_i = \frac{\bar{\lambda}_i}{n\omega_n s_i}$ and ω_n.

At times, the term $\frac{2\bar{\lambda}_i}{ns_i}$ will be negligible when compared to $\hat{\Sigma}_{ii}$. This observation, along with equation (3.41), leads to the following three regimes of variance that may be considered when fitting a random parameter stochastic block model to a data set of graphs.

3.6.1 Regimes of variance

Let $\{G^{(k)}\}_{k=1}^{N}$ be a sample set of graphs where $\bar{\boldsymbol{\lambda}}$ and $\hat{\Sigma}$ respectively denote the arithmetic mean of the largest c eigenvalues of the adjacency matrices and the sample covariance of the

eigenvalues about the arithmetic mean. When

$$\hat{\Sigma}_{ii} < \frac{2\bar{\lambda}_i}{ns_i} \tag{3.44}$$

the i-th largest eigenvalue is said to live in the *small variance regime*, and there does not exist a distribution J such that $\mu_{\omega_n,J,q,\boldsymbol{s}}$ captures both the sample mean eigenvalue and the sample variance of the eigenvalues.

The *large variance regime* occurs when the term $\frac{2\bar{\lambda}_i}{ns_i}$ is negligible as compared to $\hat{\Sigma}_{ii}$. For each $i = 1, ..., c$, we determine whether the i-th eigenvalue is in this regime by examining whether

$$\frac{2\bar{\lambda}_i}{ns_i} \ll \hat{\Sigma}_{ii}. \tag{3.45}$$

When the data set of graphs fall within this regime, omitting the contribution to variance from $\frac{2\bar{\lambda}_i}{ns_i}$ as

$$Var_J(P_i) = \frac{\hat{\Sigma}_{ij}}{n^2\omega_n^2 s_i^2} - \frac{2\bar{\lambda}_i}{n^3\omega_n^2 s_i^3} \tag{3.46}$$

$$\approx \frac{\hat{\Sigma}_{ij}}{n^2\omega_n^2 s_i^2} \tag{3.47}$$

provides a reasonable estimate for the variance of J. In this case, equation (3.29) in Step 8 of Alg. 4 simplifies to

$$\mathrm{Cov}(P_i, P_j) = \frac{\hat{\Sigma}_{ij}}{n^2\omega_n^2 s_i s_j} \tag{3.48}$$

The *medium variance regime* occurs when the term $\mathbb{E}_J[2P_i] = \frac{2\bar{\lambda}_i}{ns_i}$ is significant when compared to $\hat{\Sigma}_{ii}$, which is the case when

$$\frac{2\bar{\lambda}_i}{ns_i\hat{\Sigma}_{ii}} = \mathcal{O}(1). \tag{3.49}$$

When this is the case, the distribution J will only add a minor contribution to the total variance of the i-th largest eigenvalue. In this situation, the data may be interpreted as accurately modeled by a classic stochastic block model, indicating no need for a distribution to be defined on the parameters.

3.6.1.1 Summary

When J comes from a two-parameter family of distributions, resolving the first and second moments precisely identifies the distribution J. In some cases, when J has more than two parameters, further moments may need to be considered.

When possible, taking J to be parametric is preferred primarily because of the faster convergence rates when estimating parametric probability density functions. Often it is the case that for large graphs, the number of samples, N, is small; thus, a fast convergence rate with respect to N is desirable. However, when J is non-parametric, we suggest an alternative approach to estimate the probability density from which the graphs $\{G^{(k)}\}_{k=1}^{N}$ were sampled.

3.6.2 Non-parametric density estimation

Assume that J is a non-parametric distribution over the parameters $\boldsymbol{p}$. Let $\boldsymbol{P}_J$ be an observation from J with components P_i, let $P_{\min} = \min_{i=1,\ldots,c} P_i$ and define $q = \epsilon P_{\min}$, where $\epsilon \ll 1$. For a sample set of graphs $\{G^{(k)}\}_{k=1}^{N}$, assume that each $G^{(k)}$ is sampled iid from a random parameter stochastic block model, $\mu_{\omega_n,J,\epsilon,\boldsymbol{s}}$. Rather than determining J via the moments, we suggest an alternative approach: a generalization of kernel density estimation. We define an estimate of $\mu_{\omega_n J,\epsilon,\boldsymbol{s}}$ as

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{N} \mu_{\omega_n^{(k)},J^{(k)},\epsilon^{(k)},\boldsymbol{s}^{(k)}}, \tag{3.50}$$

where the probability measures $\mu_{\omega_n^{(k)},J^{(k)},\epsilon^{(k)},\boldsymbol{s}^{(k)}}$ act analogously to kernels when performing kernel density estimation and the distribution $J^{(k)}$ can be interpreted as the bandwidth parameter which determines the variance of each kernel.

To determine $\mu_{\omega_n^{(k)},J^{(k)},\epsilon^{(k)},\boldsymbol{s}^{(k)}}$ for each k, we suggest the following algorithm, which is closely related to Alg. 4. The difference is in the determination of the second moment of $J^{(k)}$.

Algorithm 5 Determination of $\hat{\mu}$ given sample data

Require: Set of graphs, $M = \{G^{(k)}\}_{k=1}^{N}$.

1: Compute the eigenvalues of the adjacency matrix of each graph as $\boldsymbol{\lambda}^{(k)} = \sigma_c(\boldsymbol{A}^{(k)})$.

2: Define the set $\{\boldsymbol{\lambda}^{(k)}\}_{k=1}^{N} \subset \mathbb{R}^c$.

3: Consider the c largest eigenvalues as a subset of $\mathbb{R}^c$ and perform kernel density estimation. Define

$$\widehat{T}_{\boldsymbol{H}}(\boldsymbol{\lambda}) = \frac{1}{N}\sum_{k=1}^{N} K_{\boldsymbol{H}}^{(k)}(\boldsymbol{\lambda}). \tag{3.51}$$

as the kernel density estimator given $\{\boldsymbol{\lambda}^{(k)}\}_{k=1}^{N} \subset \mathbb{R}^c$ where each kernel $K_{\boldsymbol{H}}^{(k)}(\boldsymbol{\lambda})$ is a probability density function that satisfies

$$\mathbb{E}_{K_{\boldsymbol{H}}^{(k)}}[\boldsymbol{\lambda}] = \boldsymbol{\lambda}^{(k)} \quad \forall k = 1, ..., N \tag{3.52}$$

$$\mathrm{Cov}_{K_{\boldsymbol{H}}^{(k)}}(\lambda_i, \lambda_j) = \boldsymbol{H}_{ij} \quad \forall k = 1, ..., N \tag{3.53}$$

where $\boldsymbol{H}$ is estimated at the choice of the researcher. For an overview of methods see [45].

4: For each k,

5: Compute the eigenvalues of the adjacency matrix of each graph as $\boldsymbol{\lambda}^{(k)} = \sigma_c(\boldsymbol{A}^{(k)})$.

6: Determine the density of each graph as $\rho_n^{(k)}$.

7: Set $\boldsymbol{s}^{(k)}$ such that $s_i^{(k)} = \frac{1}{c}$ for all $1 \leq i \leq c$.

8: Set $\omega_n^{(k)} = \rho_n^{(k)}$.

9: Determine the first moment of $J^{(k)}$ according to

$$\mathbb{E}_{J^{(k)}}[P_i] = \frac{\lambda_i^{(k)}}{n\omega_n^{(k)} s_i^{(k)}}. \tag{3.54}$$

10: Determine the second moment of $J^{(k)}$ according to

$$\mathrm{Cov}_{J^{(k)}}(P_i, P_j) = \frac{\boldsymbol{H}_{ij}}{n^2(\omega_n^{(k)})^2 s_i^{(k)} s_j^{(k)}} - \begin{cases} \frac{2\lambda_i^{(k)}}{n^3(\omega_n^{(k)})^2(s_i^{(k)})^3} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{3.55}$$

11: Set

$$\epsilon^{(k)} = \frac{1 - \sum_{i=1}^{c} \mathbb{E}_{J^{(k)}}[P_i]\,(s_i^{(k)})^2}{\min\limits_{i=1,...,c}(\mathbb{E}_{J^{(k)}}[P_i])(1 - \sum_{i=1}^{c}(s_i^{(k)})^2)} \tag{3.56}$$

12: Return: $\hat{\mu} = \frac{1}{N}\sum_{k=1}^{N} \mu_{\omega_n^{(k)}, \boldsymbol{J}^{(k)}, \epsilon^{(k)}, \boldsymbol{s}^{(k)}}$.

Two new conditions for the feasibility of each $J^{(k)}$ arise for Alg. 5, namely

$$\mathbb{E}_J[P_i] = \frac{\lambda_i^{(k)}}{n\omega_n^{(k)} s_i^{(k)}} \in [0,1], \tag{3.57}$$

$$0 < \boldsymbol{H}_{ii} - 2\omega_n^{(k)}\mathbb{E}_J[P_i] = \boldsymbol{H}_{ii} - \frac{2\lambda_i^{(k)}}{ns_i^{(k)}}. \tag{3.58}$$

Remark 14 In practice, there are two common properties of graph-valued data sets to note. First, for a large value of n, the inherent variance of the largest eigenvalues from a stochastic block model is small. Therefore, for heterogeneous data sets of graphs, we expect a small contribution to the total variance from the inherent variance of the stochastic block model. This suggests that the term $\frac{2\lambda_i^{(k)}}{ns_i^{(k)}}$ is small in practice and that equation (3.58) is likely to be satisfied.

Second, the data is typically coarse for large graphs in that $N \ll n$. Given the coarse data set in $\mathcal{G}$, the expectation is that the data, $\{\boldsymbol{\lambda}^{(k)}\}_{k=1}^N \subset \mathbb{R}^c$, is also coarse. When performing kernel density estimation in $\mathbb{R}^c$, for a coarse data set, the estimation of $\boldsymbol{H}$ will be large; see, for instance, Silverman's rule of thumb when N is small while $\hat{\Sigma}$ is large.

With the expectation that $\boldsymbol{H}_{ii}$ is large and $\frac{2\lambda_i^{(k)}}{ns_i^{(k)}}$ is small, equation (3.58) is expected to be satisfied for a wide variety of graph valued data.

For each i and k where equation (3.58) is not met, we take the i-th marginal distribution of $J^{(k)}$ to be a Dirac delta distribution centered at $\frac{\lambda_i^{(k)}}{ns_i^{(k)}}$ and accept a local over-smoothing[1] of the distribution of the i-th eigenvalue at the k-th data point.

3.7 Simulation studies

Throughout this section, we fit a random parameter stochastic block model to various datasets taking either a parametric or non-parametric approach. We first verify that when the data comes from a random parameter stochastic block model, we recover the model's parameters well. We then

[1] Smoothing is a symptom of bandwidth selection and is relative to the choice of bandwidth selection in $\mathbb{R}^c$. Remark 14 states that the kernel density estimator will over-smooth the i-th eigenvalue at the k-th data point as compared to the method chosen for kernel density estimation in $\mathbb{R}^c$, defined by equation (3.51).

test the model's limits when the data is not sampled according to a random parameter stochastic block model. Within each section, we describe how the data set of graphs was generated and showcase the quality of the estimated probability density.

3.7.1 Recoverability

We first verify that when data is sampled according to a random parameter stochastic block model, Alg. 4 recovers the parameters well. Let $\{G^{(k)}\}_{k=1}^{N}$ be a data set of graphs sampled according to a random parameter stochastic block model, $\mu_{\omega_n J,\epsilon,\boldsymbol{s}}$, where

$$N = 50, \quad n = 1000, \quad \omega_n = 10n^{-1/2}, \quad \epsilon = 0.05, \quad \boldsymbol{s} = [0.5, 0.5, 0, ...]^T, \quad J = \mathcal{U}_{[0.8,0.9]} \times \mathcal{U}_{[0.55,0.6]}. \tag{3.59}$$

Here, $\mathcal{U}_{[0.8,0.9]}$ and $\mathcal{U}_{[0.55,0.6]}$ denote uniform probability measures on the respective intervals. We represent the support of J by determining the centers of each uniform distribution as, $\boldsymbol{p}^* = [0.85, 0.55]$, and capturing the width about the mean values by the vector $\boldsymbol{\delta}^* = [0.1, 0.0.5]$. We denote the estimated parameters by $\hat{\omega}_n, \hat{\boldsymbol{p}}^*$, and $\hat{\boldsymbol{\delta}}^*$.

Remark 15 When determining J we aim to recover the product of $\omega_n \boldsymbol{p}^*$ and $\omega_n \boldsymbol{\delta}^*$, which defines the support of the observations from J when $n = 1000$. If the estimate for $\hat{\omega}_n = \omega_n$, then the direct comparison of $\hat{\boldsymbol{p}}^*$ and $\hat{\boldsymbol{\delta}}^*$ is equivalent, but because we take $\hat{\omega}_n = \bar{\rho}_n$ these comparisons do not provide useful information. This analysis is unique to the case of recoverability; in general, we will not have oracle knowledge of the parameters that generated the data, and we will only report the error in terms of the sample Fréchet mean and sample total Fréchet variance.

Let $\boldsymbol{\lambda}^{(k)} = \sigma_c(\boldsymbol{A}^{(k)})$, where $\boldsymbol{A}^{(k)}$ denotes the adjacency matrix of graph $G^{(k)}$. We compute the sample mean spectrum and sample covariance matrix as

$$\bar{\boldsymbol{\lambda}} = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\lambda}^{(k)}, \quad \hat{\Sigma} = \frac{1}{N-1}\sum_{k=1}^{N} (\boldsymbol{\lambda}^{(k)} - \bar{\boldsymbol{\lambda}})^T (\boldsymbol{\lambda}^{(k)} - \bar{\boldsymbol{\lambda}}). \tag{3.60}$$

The values of the mean eigenvalue and diagonal entries of the covariance are

$$\bar{\boldsymbol{\lambda}} = \begin{bmatrix} 133.9401 \\ \\ 91.1005 \end{bmatrix} \quad \hat{\Sigma}_{11} = 25.3956 \quad \hat{\Sigma}_{22} = 5.5628 \tag{3.61}$$

We assume a constant community size, so $s_i = 0.5$ for each i, and check which regime of variance each eigenvalue falls within by computing $\hat{\Sigma}_{ii} - \frac{2\bar{\lambda}_i}{ns_i}$ for each i:

$$\begin{bmatrix} \hat{\Sigma}_{11} - \frac{2\bar{\lambda}_1}{ns_1} \\ \hat{\Sigma}_{22} - \frac{2\bar{\lambda}_2}{ns_2} \end{bmatrix} = \begin{bmatrix} 24.8599 \\ 5.1984 \end{bmatrix}. \tag{3.62}$$

The magnitude of each term indicates that each eigenvalue is in the large variance regime, and the contribution to variance from $\frac{2\bar{\lambda}_i}{ns_i}$ can be ignored. Determining the parameters according to Alg. 4,

$$\bar{\rho}_n \hat{p}_1^* = 0.2679, \quad \bar{\rho}_n \hat{p}_2^* = 0.1822, \quad \bar{\rho}_n \hat{\delta}_1 = 0.0173, \quad \bar{\rho}_n \hat{\delta}_2 = 0.0079.$$

The absolute relative error for each component of each parameter is

$$\left| \frac{\bar{\rho}_n \hat{p}_1^* - \omega_n p_1^*}{\omega_n p_1^*} \right| = 0.2455\% \quad \left| \frac{\bar{\rho}_n \hat{p}_2^* - \omega_n p_2^*}{\omega_n p_2^*} \right| = 0.0673\%$$
$$\left| \frac{\bar{\rho}_n \hat{\delta}_1 - \omega_n \delta_1}{\omega_n \delta_1} \right| = 9.2371\% \quad \left| \frac{\bar{\rho}_n \hat{\delta}_2 - \omega_n \delta_2}{\omega_n \delta_2} \right| = 0.0958\%$$

The most notable error is in the recovery of ϵ. To compare this term, we consider

$$\left| \frac{\hat{\epsilon} - \epsilon}{\epsilon} \right| = 83.023\% \tag{3.63}$$

which is a significant relative error; however, recall that

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}) | \boldsymbol{P}_J = \boldsymbol{p}\right] = n\omega_n s_i \left(p_i + \mathcal{O}\left(\epsilon^2 p_{\min}^2\right)\right) + \mathcal{O}\left(t_i(\boldsymbol{p})\right). \tag{3.64}$$

The term

$$n\hat{\omega}_n s_1 \hat{\epsilon}^2 \min_{i=1,\ldots,c} \mathbb{E}\left[P_i\right]^2 = n\omega_n s_1 \hat{\epsilon}^2 \hat{p}_2^* = 0.0105 \tag{3.65}$$

suggests that even with a large relative error, the contribution to the largest eigenvalue from this error remains negligible. Evaluating for s_2 shows the error in the second largest eigenvalue, which is similarly small.

In addition to verifying the recovery of the parameters, we also compare the sample Fréchet mean and sample total Fréchet variance of the recovered probability measure. We define a new sample of graphs $\{\tilde{G}^{(k)}\}_{k=1}^{50}$ sampled iid from $\mu_{\hat{\omega}_n,\hat{J},\hat{\epsilon},\boldsymbol{s}}$ and compute the new sample mean eigenvalue and sample variance of the data set as

$$\bar{\boldsymbol{\lambda}}_{new} = \begin{bmatrix} 133.3465 \\ \\ 91.4807 \end{bmatrix} \quad \hat{\Sigma}_{11,new} = 25.1882 \quad \hat{\Sigma}_{22,new} = 6.2949 \tag{3.66}$$

with absolute relative errors of

$$\left|\frac{\bar{\lambda}_{1,new} - \bar{\lambda}_1}{\bar{\lambda}_1}\right| = 0.4432\%, \quad \left|\frac{\bar{\lambda}_{2,new} - \bar{\lambda}_2}{\bar{\lambda}_2}\right| = 0.4173\% \tag{3.67}$$

$$\left|\frac{\hat{\Sigma}_{11,new} - \hat{\Sigma}_{11}}{\hat{\Sigma}_{11}}\right| = 1.3207\%, \quad \left|\frac{\hat{\Sigma}_{22,new} - \hat{\Sigma}_{22}}{\hat{\Sigma}_{22}}\right| = 21.0933\%, \tag{3.68}$$

indicating that the sample Fréchet mean graph and total Fréchet variance are well approximated.

3.7.2 Mixture model estimation via parametric J

We now test the model's flexibility by attempting to fit a random parameter stochastic block model to a mixture model. Let $\{G^{(k)}\}_{k=1}^{N}$ be a set of graphs sampled from a mixture of four stochastic block models where $\boldsymbol{p}$ is allowed to vary between one of four different values. The following parameters are uniform across the mixture of models

$$n = 1000, \quad N = 200, \quad c = 2, \quad q = 0.05, \quad \omega_n = 10n^{-1/2} \tag{3.69}$$

The values of $\boldsymbol{p}$ considered are

$$\boldsymbol{p} \in \left\{ \begin{bmatrix} 0.9 \\ \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.9 \\ \\ 0.3 \end{bmatrix}, \begin{bmatrix} 0.6 \\ \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.6 \\ \\ 0.3 \end{bmatrix} \right\}. \tag{3.70}$$

We generate a data set of graphs, $\{G^{(k)}\}_{k=1}^{N}$, by first assigning equal weights to each possible value of $\boldsymbol{p}$ and randomly selecting a value of $\boldsymbol{p}^{(k)}$ from the possible set. A graph $G^{(k)}$ is then sampled from $\mu_{\omega_n,\boldsymbol{p}^{(k)},q,\boldsymbol{s}}$. As before, we compute $\bar{\boldsymbol{\lambda}}$ and the diagonal of $\hat{\Sigma}$;

$$\bar{\boldsymbol{\lambda}} = \begin{bmatrix} 120.2088 \\ \\ 61.78 \end{bmatrix}, \quad \begin{bmatrix} \hat{\Sigma}_{11} \\ \\ \hat{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} 524.6778 \\ \\ 228.0952 \end{bmatrix} \tag{3.71}$$

which we use to check the regime of variance of the data,

$$\begin{bmatrix} \hat{\Sigma}_{11} - \frac{2\bar{\lambda}_1}{ns_1} \\ \hat{\Sigma}_{22} - \frac{2\bar{\lambda}_2}{ns_2} \end{bmatrix} = \begin{bmatrix} 524.1970 \\ 227.8481 \end{bmatrix}. \tag{3.72}$$

The values indicate that the data resides in the regime of high variance; thus, we may safely ignore the inherent variability of the stochastic block model again.

There are multiple perspectives to consider when constructing a generative model given this data. A researcher could first cluster the data and attempt to reconstruct the mixture, or she may interpret the data as a bi-modal distribution of graphs. We select the latter perspective for experimental purposes and construct a generative model taking J to be a product measure of shifted Beta probability measures. Taking this perspective allows for no consideration of the covariance for J. We first ensure that this assumption is consistent with the data by examining the sample correlation,

$$\frac{\hat{\Sigma}_{1,2}}{\sqrt{\hat{\Sigma}_{1,1}\hat{\Sigma}_{2,2}}} = -0.0188. \tag{3.73}$$

With the correlation being nearly 0, we are assured in our assumption of a negligible covariance contribution, and taking J to be a product between shifted beta probability measures is therefore reasonable.

Each shifted Beta distribution, denoted Beta_i, depends on the four parameters,

a_i: minimum value of range,

b_i: maximum value of range,

α_i: shape parameter,

β_i: shape parameter,

with mean and variance

$$\mathbb{E}_{Beta_i}[P_i] = a_i + (b_i - a_i)\frac{\alpha_i}{\alpha_i + \beta_i}, \tag{3.74}$$

$$Var_{Beta_i}(P_i) = (b_i - a_i)^2 \frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)^2(\alpha_i + \beta_i + 1)}. \tag{3.75}$$

Crude estimates for a_i and b_i are given by

$$a_i = \min_{k=1,\dots,N} \frac{\lambda_i(\boldsymbol{A}^{(k)})}{n\bar{\rho}_n s_i}, \tag{3.76}$$

$$b_i = \max_{k=1,\dots,N} \frac{\lambda_i(\boldsymbol{A}^{(k)})}{n\bar{\rho}_n s_i}. \tag{3.77}$$

where we have taken $\omega_n = \bar{\rho}_n$ as described in Alg. 4. Solving the following set of equalities for α_i and β_i,

$$\mathbb{E}_{Beta_i}[P_i] = a_i + (b_i - a_i)\frac{\alpha_i}{\alpha_i + \beta_i} \tag{3.78}$$

$$Var_{Beta_i}(P_i) = (b_i - a_i)^2 \frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)^2(\alpha_i + \beta_i + 1)} \tag{3.79}$$

where $\mathbb{E}_{Beta_i}[P_i] = \frac{\bar{\lambda}_i}{n\bar{\rho}_n s_i}$ and $Var_{Beta_i}(P_i) = \frac{\hat{\Sigma}_{ii}}{n^2\bar{\rho}_n^2 s_i^2}$ result in the following set of parameters for the two Beta distributions;

$$\boldsymbol{a} = \begin{bmatrix} 1.94656 \\ 0.942685 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2.97084 \\ 1.63105 \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} 0.0996378 \\ 0.111218 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} 0.102785 \\ 0.13049 \end{bmatrix}. \tag{3.80}$$

To compare the goodness of fit between the estimated distribution $\hat{\mu}$ resulting from Alg. 4 we take a new sample of graphs, $\{G_{new}^{(k)}\}_{k=1}^{200}$ from $\mu_{\hat{\omega}_n,\hat{J},\hat{\epsilon},\boldsymbol{s}}$ and compare the sample Fréchet mean and sample total Fréchet variance. Lemma 4 shows that an estimate for the eigenvalues of the sample Fréchet mean for large graphs is given by the arithmetic mean eigenvalue, and an estimate for the sample total Fréchet variance is given by the diagonal of the sample covariance matrix. The sample mean eigenvalue and sample covariance matrix of the new sample are

$$\bar{\boldsymbol{\lambda}}_{new} = \begin{bmatrix} 121.0192 \\ 60.8119 \end{bmatrix}, \quad \begin{bmatrix} \hat{\Sigma}_{11,new} \\ \hat{\Sigma}_{22,new} \end{bmatrix} = \begin{bmatrix} 504.1689 \\ 209.7644 \end{bmatrix}. \tag{3.81}$$

The relative error for each eigenvalue and the diagonal entries of the covariance matrix

$$\left|\frac{\bar{\lambda}_{1,new} - \bar{\lambda}_1}{\bar{\lambda}_1}\right| = 0.6742\%, \quad \left|\frac{\bar{\lambda}_{2,new} - \bar{\lambda}_2}{\bar{\lambda}_2}\right| = 1.567\%, \tag{3.82}$$

$$\left|\frac{\hat{\Sigma}_{11,new} - \hat{\Sigma}_{11}}{\hat{\Sigma}_{11}}\right| = 3.9089\%, \quad \left|\frac{\hat{\Sigma}_{22,new} - \hat{\Sigma}_{22}}{\hat{\Sigma}_{22}}\right| = 8.0365\%. \tag{3.83}$$

While we have confirmed that the statistics of the recovered distribution match both the first and second moments of the original data set, we may further check the goodness of fit for the choice of J by considering an estimate of the probability density of the two largest eigenvalues from each data set, as presented in Fig. 3.1. As is evident in Fig. 3.1, using a random parameter stochastic

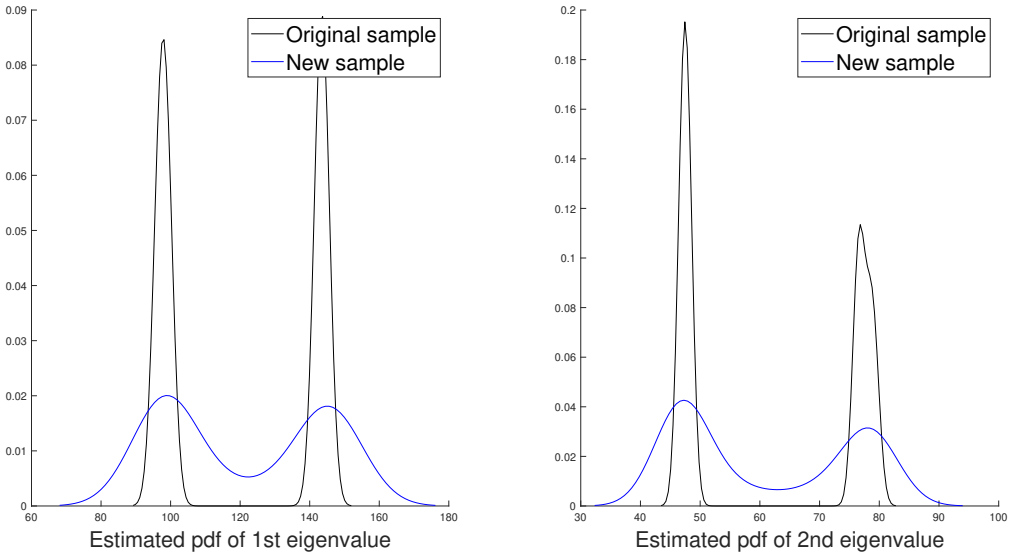

Figure 3.1: A comparison between the estimated probability densities of the two largest eigenvalues from the original data set (black) compared to the recovered data set (blue). Each curve was generated by embedding the observed sets of eigenvalues in $\mathbb{R}^c$ and performing a kernel density estimate.

block model with J given by the product of Beta distributions to fit a mixture model may not always result in an ideal fit of the distributions, even when aligning the first and second moments.

3.7.3 Primary school time-varying networks

The final data set of graphs is a time series of networks collected via RFID tags in a French primary school [40, 93]. These networks exhibit dynamic structural behaviors temporally by merging communities throughout the day in addition to the random fluctuations of the individual interactions. The school is composed of five grades with two classes per grade for a total of 10 different classes and a fixed vertex set of size $n = 242$.

Every time t, a collection of edges is given that corresponds to the face-to-face contacts of the graph, and a new graph is recorded every 20 seconds. We collect all connections within a time window of 45 minutes (2700 seconds) to define a single graph and shift the window by one time step to collect the next graph. The graph $G^{(k)}$ describes all connections made from $20(k-1)$ seconds to $2700 + 20(k-1)$ seconds.

We expect a high degree of correlation from $G^{(k)}$ to $G^{(k+1)}$ and yet, as is displayed in Fig. 3.2, there is notable dynamic behavior in the graphs as observed by the change in the largest eigenvectors. Many perspectives analyze the time series of graphs as a change point problem or anomaly detection. We ask a fundamentally different question: What is the distribution of the graphs during a time interval?

We consider only the initial 2 hours of the school day, 9:00 - 11:00 a.m. This time interval includes the behavior of students both in the classrooms and during the 10:30 - 11:00 a.m. recess and gives a data set of graphs with size $N = 225$, denoted as $M = \{G^{(k)}\}_{k=1}^{N}$.

Informed by the dynamics of the networks, such as the merging of communities during inter-classroom projects and the mixing of classes during recess, we suggest having a dynamic estimate for the geometry vector. Rather than taking $\boldsymbol{s}$ to be constant, as was done in all prior experiments when recovering a random-parameter stochastic block model, we now estimate $\boldsymbol{s}^{(k)}$ for each k. The work in [46, 50, 73, 79] showcases that the sum of the logarithms of the largest eigenvectors can be used to identify the geometry vector $\boldsymbol{s}^{(k)}$ (see Fig. 3.2 and Fig. 3.3). An estimate for the largest K eigenvectors from each $\boldsymbol{A}^{(k)}$ was determined by counting the number of eigenvalues greater than $|\lambda_{242}(\boldsymbol{A}^{(k)})|$, which is interpreted as an estimate of the number of extremal eigenvalues.

The geometry vector $\boldsymbol{s}^{(k)}$ is determined by summing the absolute value of the largest K eigenvectors and analyzing the change points. Each change point is interpreted as a new community within the graph $G^{(k)}$. A robust estimate for change point detection is depicted above, which uses a Bayesian framework (see [107]). However, any state-of-the-art algorithm for change point detection should suffice. When two sequential change points are detected at $index_i$ and $index_{i+1}$ such that $|index_i - index_j| < \lceil \lambda_1^{(k)} \rceil$, we elect to take the average. This is due to the condition on the

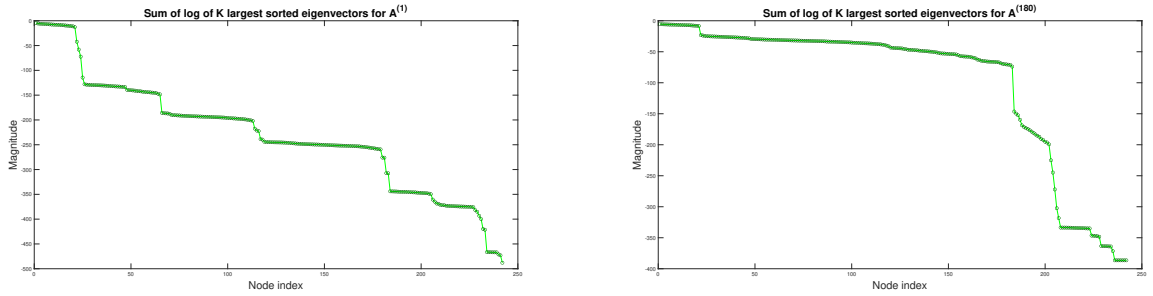

Figure 3.2: (Left) A visualization of the sum of the logarithm of the absolute value of the K largest sorted eigenvectors from $\boldsymbol{A}^{(1)}$. (Right) A visualization of the sum of the logarithm of the absolute value of the K largest sorted eigenvectors from $\boldsymbol{A}^{(180)}$.

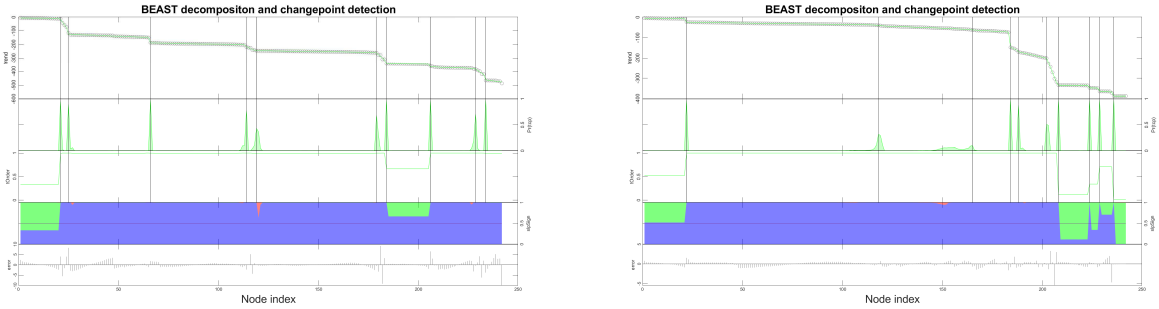

Figure 3.3: An estimation of the change points in the plot in the sum of the largest K eigenvectors for $\boldsymbol{A}^{(1)}$ and $\boldsymbol{A}^{(180)}$. The vertical lines indicate the locations of the changepoints, which are used to infer the communities

geometry equation (3.39), which shows that the number of nodes in a community must be larger than the corresponding eigenvalue.

Detecting the geometry for each graph $G^{(k)}$ leads to a set of geometry vectors $\{\boldsymbol{s}^{(k)}\}_{k=1}^{N}$. Before estimating a distribution of graphs, we now cluster the graphs $G^{(k)}$ such that, within a cluster, the geometry vectors $\boldsymbol{s}^{(k)}$ have the same number of non-zero entries.

For this particular data set, we find four distinct clusters, namely graphs with 4, 5, 6, and 7 communities, respectively. We denote each subset of graphs by $M_c \subset M$, where c denotes the

number of communities. The size of each M_c is given by the following.

$$|M_4| = 20 \tag{3.84}$$

$$|M_5| = 63 \tag{3.85}$$

$$|M_6| = 113 \tag{3.86}$$

$$|M_7| = 29 \tag{3.87}$$

$$M = M_4 \cup M_5 \cup M_6 \cup M_7 \tag{3.88}$$

$$\emptyset = M_i \cap M_j \quad \forall i \neq j \tag{3.89}$$

Here, we analyze only M_5 and M_6, the two clusters with a significant number of graphs, using Alg. 5, where we assume $J^{(k)}$ is a product of uniform probability measures, resulting in two probability measures, $\hat{\mu}_{M_5}$ and $\hat{\mu}_{M_6}$. We then sample $N = 500$ graphs according to $\hat{\mu}_{M_5}$ to determine a new set of graphs $\{G^{(k)}_{new}\}_{k=1}^{57}$. To compare the quality of the estimated probability measure $\hat{\mu}_{M_5}$, we compute the largest 5 eigenvalues for each $G^{(k)} \in M_5$ and each $G^{(k)}_{new} \in \{G^{(k)}_{new}\}_{k=1}^{500}$, perform kernel density estimation for the vectors of the eigenvalues in $\mathbb{R}^5$, and plot the results.

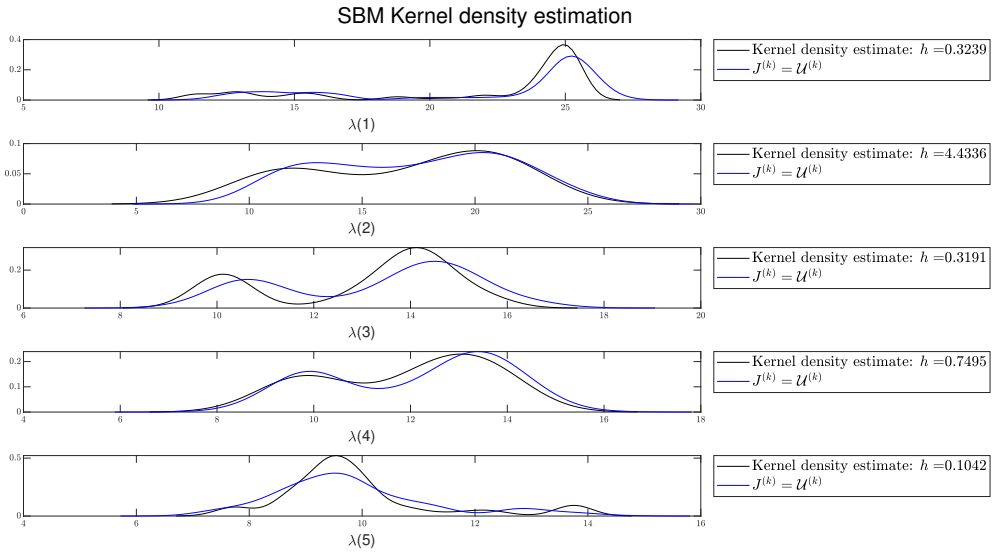

Figure 3.4: (Black) A kernel density estimate of the sample eigenvalues of the adjacency matrices of graphs in M_5. (Blue) A kernel density estimate of the sample eigenvalues of the adjacency matrices of graphs in $\{G_{new}^{(k)}\}_{k=1}^{|M_5|}$ where $G_{new}^{(k)} \sim \hat{\mu}_5$.

For M_6, we perform the same procedure, except we consider the cluster of graphs with 6 non-zero entries in the geometry vectors $\boldsymbol{s}^{(k)}$.

While the black curve need not resemble the true distribution of the eigenvalues well, this is the only baseline with which we may compare our results. Notably, the black curve defines a distribution in the space $\mathbb{R}^c$ but does not determine how to generate graphs with the corresponding eigenvalue distribution. The primary advantage of our method is that the blue curve was generated by first sampling a graph and then computing the eigenvalues of the graph's adjacency matrix. The implication is that the probability measure

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^{N} \mu_{\omega_n^{(k)}, J^{(k)}, \epsilon^{(k)}, \boldsymbol{s}^{(k)}} \tag{3.90}$$

found by Alg. 5 distributes graphs such that the largest eigenvalues follow a distribution that is similar to the black curve.

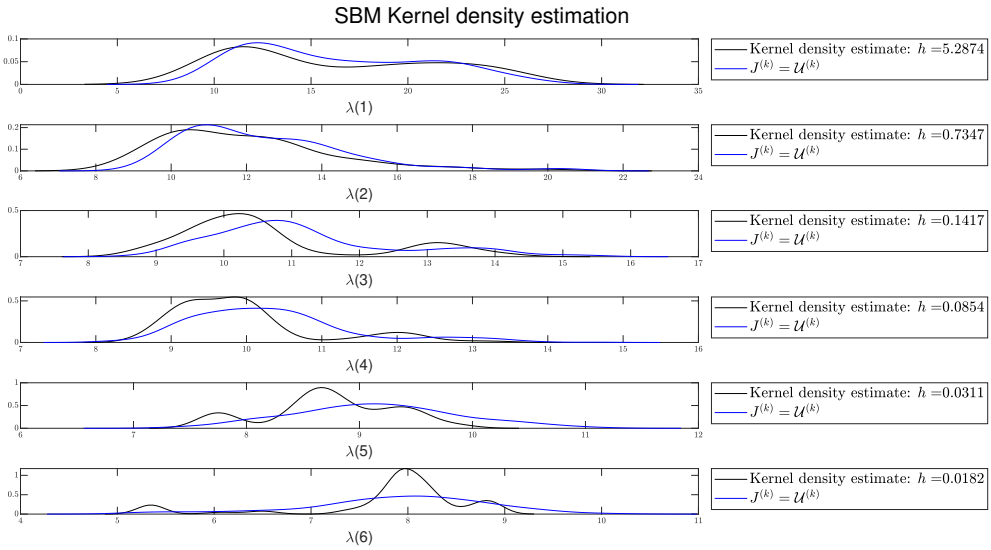

Figure 3.5: (Black) A kernel density estimate of the sample eigenvalues of the adjacency matrices of graphs in M_6. (Blue) A kernel density estimate of the sample eigenvalues of the adjacency matrices of graphs in $\{G_{new}^{(k)}\}_{k=1}^{|M_6|}$ where $G_{new}^{(k)} \sim \hat{\mu}_6$.

The values of h are presented to showcase that when h is small, over-smoothing of the data is expected, which can be seen most clearly in Fig. 3.5 for $\lambda(5)$ and $\lambda(6)$. In contrast, for larger values of H, such as in Fig. 3.5 for $\lambda(1)$, oversmoothing is not expected because the condition given by equation (3.58) is met.

3.8 Conclusions

A preliminary step for the analysis of any graph-valued data set is the choice of metric, or measure of similarity, between graphs. When the distance is chosen with respect to spectral information, such as d_{A_c}, a broad class of generative models can be considered when fitting a data set, called inhomogeneous Erdős-Rényi random graph models. The spectral properties of graphs generated in this way have been well-studied, as evidenced by the work in [12, 22, 30, 97], among others. Two general results are observed for this class of models. First, there is low variability in the largest eigenvalues given a fixed function $f(x, y)$ that defines the inhomogeneous Erdős-Rényi

model. Second, the location and scale of the largest eigenvalues of the adjacency matrices of graphs generated in this way are dependent. It should be noted that here, the term inhomogeneous is used to characterize the edge probabilities of the graphs, and not the data sets of graphs generated in this manner. A sample set of graphs from such a model remains rather homogeneous.

To mitigate the problem of low variance of the largest eigenvalues, when considering the class of stochastic block models, this chapter introduced randomness in the parameter space via the distribution J to better model heterogeneous data sets of graphs, $\{G^{(k)}\}_{k=1}^{N}$. We have shown through Lemma 3 that methods that estimate the relative error in the parameters of stochastic block models can be translated to estimate relative error in the first and second moments of the distribution J. For parametric assumptions of J, the estimation of the first and second moments is typically sufficient to characterize J, but, as evidenced by the real-world data, a finite number of moments of J may not characterize the distribution well. In these situations, a generalization of kernel density estimation for the space of graphs is introduced, and the limitations of such a perspective are explored as a function of sample size. It was shown experimentally that for finite n, we cannot expect to determine the "bandwidth" for the estimator proposed in 3.50 for every sized sample N.

The limitations of the family of stochastic block models as kernels for kernel density estimation point to a fundamental limitation of modeling graphs when a Bernoulli process models the edge probabilities. To specify the limitations of this process, we recall the results from [38], which were presented in equation (3.2). The authors show that for an Erdős-Rényi random graph with parameters n and p, that

$$\lambda_1 \xrightarrow{d} N\left((n-2)p+1, 2p(1-p)+\mathcal{O}(\frac{1}{\sqrt{n}})\right). \tag{3.91}$$

Recall that for an Erdős-Rényi, the entries of the adjacency matrix $\boldsymbol{A}$ are defined by the Bernoulli process

$$a_{ij} \sim \text{Bernoulli}\,(p)\,. \tag{3.92}$$

The expected value and variance for each entry are given by

$$\mathbb{E}\left[a_{ij}\right] = p \tag{3.93}$$

$$\mathrm{Var}(a_{ij}) = p(1-p). \tag{3.94}$$

Rewriting the results of [38],

$$\lambda_1 \xrightarrow{d} N\left((n-2)\mathbb{E}\left[a_{ij}\right] + 1, 2\mathrm{Var}(a_{ij}) + \mathcal{O}(\frac{1}{\sqrt{n}})\right), \tag{3.95}$$

which shows clearly that while the edges of adjacency matrix, $\boldsymbol{A}$, are modeled by a one-parameter family such that the location and scale of each a_{ij} is coupled, then coupling between the location and scale of the largest eigenvalues of a graph is also expected. These results are also seen in the inhomogeneous Erdős-Rényi ensemble in equations (3.12) and (A.2) of Theorem 8 where it can be inferred that any change in the eigenvalues or eigenfunctions of L_f, denoted θ_i and $r_i(x)$ respectively, change the values of equations (3.12) and (A.2). For these reasons, a new model for graphs may be needed when modeling very homogeneous data sets of graphs, a model in which the location and scale of the entries of the adjacency matrix are perhaps independent.

Chapter 4

On the Number of Edges of the Fréchet Mean and Median Graphs

4.1 Introduction

We consider the set $\mathcal{G}$ formed by all undirected unweighted simple labeled graphs with vertex set $\{1, \ldots, n\}$. We equip $\mathcal{G}$ with a metric d to measure the distance between two graphs.

We characterize the "average" of a sample of graphs $\left\{G^{(1)}, \ldots, G^{(N)}\right\}$, which are defined on the same vertex set $\{1, \ldots, n\}$, with the sample Fréchet mean and median graphs, [42].

Definition 19 The sample Fréchet mean graphs are solutions to

$$\widehat{\boldsymbol{\mu}}_N\big[G\big] = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \frac{1}{N} \sum_{k=1}^{N} d^2(G, G^{(k)}), \tag{4.1}$$

and the sample Fréchet median graphs are solutions to

$$\widehat{\boldsymbol{m}}_N\big[G\big] = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \frac{1}{N} \sum_{k=1}^{N} d(G, G^{(k)}). \tag{4.2}$$

Solutions to the minimization problems (4.1) and (4.2) always exist, but the minimizers need not be unique. All our results are stated in terms of any of the elements in the set of minimizers of (4.1) and (4.2).

Because the focus of this work is not the computation of the Fréchet mean or median graphs, but rather a theoretical analysis of the properties that these graphs inherit from the graph sample, we assume that the graphs in the sample are defined on the same vertex set.

The vital role played by the Fréchet mean as a location parameter [57, 56, 67], is exemplified in the works of [7, 88], who have created novel families of random graphs by generating random perturbations around a given Fréchet mean graph.

4.1.1 Our main contributions

We consider a set of N unweighted simple labeled graphs, $\left\{G^{(1)}, \ldots, G^{(N)}\right\}$, with vertex set $\{1, \ldots, n\}$. In this chapter, we address the following foundational question: does a mean or median graph inherit the structural properties of the graphs in the sample? Specifically, we establish that edge density is a hereditary property, which can be transmitted from a graph sample to its sample Fréchet mean or median.

Because sparse graphs provide prototypical models for real networks, our theoretical analysis is significant since it guarantees that this structural property is preserved when computing a sample mean or median. Similarly, the authors in [48] construct a sparse median graph, which provides a more interpretable summary, from a set of graphs that are not necessarily sparse.

Our work answers the question raised by the author in [43]: "does the average of two sparse networks/matrices need to be sparse?" Specifically, we prove the following result: the number of edges of the Fréchet mean or median graphs of a set of graphs is bounded by the sample mean number of edges of the graphs in the sample. We use different arguments to prove this result for the graph Hamming distance and the spectral adjacency pseudometric.

4.2 Preliminary and Notations

We denote by $\mathcal{S}$ the set of $n \times n$ adjacency matrices of graphs in $\mathcal{G}$,

$$\mathcal{S} = \left\{ \boldsymbol{A} \in \{0,1\}^{n \times n}; \text{where } a_{ij} = a_{ji}, \text{and } a_{i,i} = 0;\ 1 \leq i < j \leq n \right\}. \tag{4.3}$$

For a graph $G \in \mathcal{G}$, we denote by $\boldsymbol{A}$ its adjacency matrix, and by $e\big(\boldsymbol{A}\big)$ the number of edges – or **volume** – of G,

$$e\big(\boldsymbol{A}\big) = \sum_{1 \leq i < j \leq n} a_{ij}. \tag{4.4}$$

We denote by $\boldsymbol{\lambda}(\boldsymbol{A}) = \begin{bmatrix} \lambda_1(\boldsymbol{A}) & \ldots & \lambda_n(\boldsymbol{A}) \end{bmatrix}$, the vector of eigenvalues of $\boldsymbol{A}$, with the convention that $\lambda_1(\boldsymbol{A}) \geq \ldots \geq \lambda_n(\boldsymbol{A})$.

4.2.1 Distances between graphs

This work considers two metrics: the Hamming distance and the spectral adjacency pseudometric. We briefly recall the definitions of these.

Definition 20 Let $G, G' \in \mathcal{G}$ be two unweighted graphs with known vertex correspondence and with adjacency matrix $\boldsymbol{A}$ and $\boldsymbol{A'}$ respectively. We define the Hamming distance between G and G' as

$$d_H(\boldsymbol{A}, \boldsymbol{A'}) \stackrel{\text{def}}{=} \sum_{1 \leq i < j \leq n} |a_{ij} - a'_{ij}| = e\big(\boldsymbol{A}\big) + e\big(\boldsymbol{B}\big) - 2 \sum_{1 \leq i < j \leq n} a_{ij} b_{ij}. \tag{4.5}$$

The Hamming distance is very sensitive to fine-scale fluctuations of the graph connectivity. In contrast, a metric based on the eigenvalues of the adjacency matrix can quantify configurational changes that occur on a graph at many more scales [26, 103].

Definition 21 Let $G, G' \in \mathcal{G}$ with adjacency matrix $\boldsymbol{A}$ and $\boldsymbol{A'}$ respectively. We define the adjacency spectral pseudometric as the ℓ_2 norm between the vectors of eigenvalues $\boldsymbol{\lambda}(\boldsymbol{A})$ and $\boldsymbol{\lambda}(\boldsymbol{A'})$ of $\boldsymbol{A}$ and $\boldsymbol{A'}$ respectively,

$$d_\lambda\big(\boldsymbol{A}, \boldsymbol{A'}\big) = ||\boldsymbol{\lambda}(\boldsymbol{A}) - \boldsymbol{\lambda}(\boldsymbol{A'})||_2. \tag{4.6}$$

The pseudometric d_λ satisfies the symmetry and triangle inequality axioms, but not the identity axiom. Instead, d_λ satisfies the reflexivity axiom, $\forall G \in \mathcal{G}$, $d_\lambda(G, G) = 0$. We note that the adjacency spectral pseudometric does not require node correspondence.

4.3 Main Results

In the following, we consider a set of N unweighted simple labeled graphs, $\big\{G^{(1)}, \ldots, G^{(N)}\big\}$, with vertex set $\{1, \ldots, n\}$. We denote by $\boldsymbol{A}^{(k)}$ the adjacency matrix of graph $G^{(k)}$. We equip the set $\mathcal{G}$ of all unweighted simple graphs on n nodes with a pseudometric, or a metric, d. The Fréchet mean and median graphs encode two notions of centrality (4.1) and (4.2) that minimise the following dispersion function, also called the Fréchet function.

Definition 22 We denote by $\widehat{F}_q(\boldsymbol{A})$ the sample Fréchet function associated with a sample Fréchet median ($q = 1$) or mean ($q = 2$),

$$\widehat{F}_q(\boldsymbol{A}) = \frac{1}{N}\sum_{k=1}^{N} d^q(\boldsymbol{A}, \boldsymbol{A}^{(k)}). \tag{4.7}$$

To quantify the connectivity of the graph sample, $\{G^{(1)}, \ldots, G^{(N)}\}$, we define the sample mean and variance of the number of edges.

Definition 23 The sample mean and variance of the number of edges are defined by

$$\overline{e}_N = \frac{1}{N}\sum_{k=1}^{N} e\big(\boldsymbol{A}^{(k)}\big), \quad \text{and} \quad \sigma_N^2(e) = \frac{1}{N}\sum_{k=1}^{N} \big[e\big(\boldsymbol{A}^{(k)}\big)\big]^2 - \big[\overline{e}_N\big]^2. \tag{4.8}$$

We now turn our attention to the main problem. We consider the following question: if the graphs $G^{(1)}, \ldots, G^{(N)}$ all have a similar edge density, can one determine the edge density of the sample Fréchet mean or median graphs? and does that number of edges depend on the choice of metric d in (4.1) and (4.2)? We answer both questions in the following theorem.

Theorem 9 Let $\{G^{(1)}, \ldots, G^{(N)}\}$ be a sample of unweighted simple labeled graphs with vertex set $\{1, \ldots, n\}$. Let $\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]$ be the adjacency matrix of a sample Fréchet mean graph, and let $\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]$ be the adjacency matrix of a sample Fréchet median graph. Let $e_{\widehat{\boldsymbol{\mu}}}$ and $e_{\widehat{\boldsymbol{m}}}$ be the number of edges of $\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]$ and $\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]$ respectively.

If the Fréchet mean and median graphs are computed using the Hamming distance, then

$$e_{\widehat{\boldsymbol{\mu}}} < 2\overline{e}_N + \frac{\sigma_N(e)}{\sqrt{2}}, \quad \text{and} \quad e_{\widehat{\boldsymbol{m}}} < 2\,\overline{e}_N, \tag{4.9}$$

and if the Fréchet mean and median graphs are computed using the adjacency spectral pseudometric, then

$$e_{\widehat{\boldsymbol{\mu}}} < 9\,\overline{e}_N, \quad \text{and} \quad e_{\widehat{\boldsymbol{m}}} < 9\overline{e}_N. \tag{4.10}$$

Proof of Theorem 9 The proof is a direct consequence of lemmata 10 and 16.

Remark 16 When the graph $G^{(k)}$ are sampled from the inhomogeneous Erdős-Rényi random graph probability space $\mathcal{G}\big(n, \boldsymbol{P}\big)$ [15], and if the distance on $\mathcal{G}$ is the Hamming distance, then $\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big] = \widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]$ with high probability [75]. In this case, a tight bound on $e_{\widehat{\boldsymbol{\mu}}}$ or $e_{\widehat{\boldsymbol{m}}}$ in (4.9) is $2\overline{e}_N$, which – unlike (4.9) – does not involve $\sigma_N(e)$.

The fact that we overestimate the bound on $e_{\widehat{\boldsymbol{\mu}}}$ by the addition of the term $\sigma_N(e)/\sqrt{2}$ comes from our technique of proof, which relies on an estimate of the Fréchet function. As explained in Remark 19, our estimate of the Fréchet function is almost tight; it does include the term $\sigma_N(e)$, as it should.

Finally, the following corollary answers the question raised by the author in [43]: "does the average of two sparse networks/matrices need to be sparse?"

Corollary 3 Let $\big\{G^{(1)}, \ldots, G^{(N)}\big\}$ be a sample of unweighted simple labeled graphs with vertex set $\{1, \ldots, n\}$. We assume that the number of edges of each $G^{(k)}$ satisfies

$$e\big(\boldsymbol{A}^{(k)}\big) = \mathcal{O}\left(n^2\right), \text{ but } e\big(\boldsymbol{A}^{(k)}\big) = \omega(n). \tag{4.11}$$

Then the sample Fréchet mean and median graphs – computed according to either the Hamming distance or the adjacency spectral pseudometric – are sparse, as defined by (4.11).

Proof of Corollary 2 The corollary is a direct consequence of theorem 9.

4.4 Proofs of the main result

We give in the following the proof of theorem 9. The key observation is that it is relatively easy to derive tight bounds on the number of edges of the sample Fréchet median graph. Inspired by the results in [75] that show that for large classes of random graphs the sample Fréchet median and mean graphs are identical, we prove that the bounds derived for the Fréchet median graphs also hold for the Fréchet mean graphs.

Our analysis begins in Subsection 4.4.1 with the sample median graphs computed using the Hamming distance, we then move to the sample mean graphs in Subsection 4.4.2. In Subsec-

tions 4.4.4 and 4.4.5, we extend these results to the sample mean and median graphs computed with the adjacency spectral pseudometric.

When possible, we use the probability space $\mathcal{G}\big(n, \boldsymbol{P}\big)$ of inhomogeneous Erdős-Rényi random graphs [15], equipped with the Hamming distance to test the tightness of our results [75].

4.4.1 The median graphs computed using the Hamming Distance

The Hamming distance, by nature, promotes sparsity [26, 103], and we therefore expect that the volumes of the sample Fréchet mean and median graphs computed with this distance are similar to the mean number of edges.

When the distance is the Hamming distance, the sample Fréchet median graphs can in fact be characterized analytically.

Lemma 5 The adjacency matrix $\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]$ of a sample median graph $\widehat{\boldsymbol{m}}_N\big[G\big]$ is given by the majority rule,

$$\Big[\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]\Big]_{ij} = \begin{cases} 0 & \text{if } \sum_{k=1}^{N} a_{ij}^{(k)} < N/2, \\ 1 & \text{otherwise.} \end{cases} \qquad \forall i, j \in \{1, \ldots, n\}\,. \tag{4.12}$$

Proof of Lemma 5 The result is classic and we omit the proof, which can be found for instance, in [25].

In the following lemma, we derive an upper bound on the number of edges of a Fréchet median graph, $e_{\widehat{\boldsymbol{m}}}$.

Lemma 6 Let $\overline{e}_N$ be the sample mean number of edges, given by (4.8). Then the number of edges of a Fréchet median graph $\widehat{\boldsymbol{m}}_N\big[G\big]$ is bounded by

$$e_{\widehat{\boldsymbol{m}}} \leq 2\overline{e}_N. \tag{4.13}$$

Remark 17 The bound (4.13) is tight for large N. Indeed, consider a sample of $2N$ graphs, where

$$G^{(k)} = \begin{cases} \text{the complete graph } K_n & \text{if} \quad 1 \leq k \leq N+1, \\ \text{the empty graph} & \text{if} \quad N+2 \leq k \leq 2N. \end{cases} \tag{4.14}$$

A Fréchet median graph $\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]$, given by the majority rule (4.12) is K_n, and thus $e_{\widehat{\boldsymbol{m}}} = n(n-1)/2$. On the other hand, the sample mean number of edges is $\overline{e}_N = e_{\widehat{\boldsymbol{m}}}/2 + e_{\widehat{\boldsymbol{m}}}/(2N)$. As the sample size N goes to infinity, we have

$$\lim_{N \longrightarrow \infty} e_{\widehat{\boldsymbol{m}}} = 2\overline{e}_N, \tag{4.15}$$

which proves that the bound (4.13) is asymptotically tight.

Proof of Lemma 6 Let $\mathcal{E}_{\widehat{\boldsymbol{m}}} = \big\{(i,j),\ i<j,\ [\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]]_{ij} = 1\big\}$ be the set of edges of $\widehat{\boldsymbol{m}}_N\big[G\big]$. We have $|\mathcal{E}_{\widehat{\boldsymbol{m}}}| = e_{\widehat{\boldsymbol{m}}}$. Now,

$$\sum_{k=1}^{N} e\big(\boldsymbol{A}^{(k)}\big) = \sum_{1 \leq i < j \leq n} \sum_{k=1}^{N} a_{ij}^{(k)} = \sum_{i,j \in \mathcal{E}_{\widehat{\boldsymbol{m}}}} \sum_{k=1}^{N} a_{ij}^{(k)} + \sum_{i,j \in \mathcal{E}_{\widehat{\boldsymbol{m}}}^{c}} \sum_{k=1}^{N} a_{ij}^{(k)}. \tag{4.16}$$

Neglecting the edges (i,j) not in $\mathcal{E}_{\widehat{\boldsymbol{m}}}$, we have

$$\sum_{k=1}^{N} e\big(\boldsymbol{A}^{(k)}\big) \geq \sum_{i,j \in \mathcal{E}_{\widehat{\boldsymbol{m}}}} \sum_{k=1}^{N} a_{ij}^{(k)} > \sum_{i,j \in \mathcal{E}_{\widehat{\boldsymbol{m}}}} \frac{N}{2} = \frac{N}{2} e_{\widehat{\boldsymbol{m}}},$$

whence we conclude

$$e_{\widehat{\boldsymbol{m}}} \leq \frac{2}{N} \sum_{k=1}^{N} e\big(\boldsymbol{A}^{(k)}\big) = 2\overline{e}_N. \tag{4.17}$$

□

4.4.2 The mean graphs computed using the Hamming Distance

First, we recall the following lower bound on the Hamming distance.

Lemma 7 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be the adjacency matrices of two unweighted graphs with number of edges $e\big(\boldsymbol{A}\big)$ and $e\big(\boldsymbol{B}\big)$ respectively. Then

$$\big|e\big(\boldsymbol{A}\big) - e\big(\boldsymbol{B}\big)\big| \leq d_H(\boldsymbol{A}, \boldsymbol{B}). \tag{4.18}$$

Proof of Lemma 7 The proof is elementary and is skipped.

Next, we derive an upper bound on the deviation of the volume of a Fréchet mean, $e_{\widehat{\boldsymbol{\mu}}}$, away from the sample average volume, $\overline{e}_N$, given by (4.8).

Lemma 8 Let $\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]$ be the adjacency matrix of a sample Fréchet mean computed using the Hamming distance, with $e_{\widehat{\boldsymbol{\mu}}}$ edges. Let $\overline{e}_N$ be the sample mean number of edges. Then

$$\Big[e_{\widehat{\boldsymbol{\mu}}} - \overline{e}_N\Big]^2 < \frac{1}{N}\sum_{k=1}^{N} d_H^2(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big], \boldsymbol{A}^{(k)}) = \widehat{F}_2\big(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]\big)\,. \tag{4.19}$$

Remark 18 This bound is not tight. We consider again the probability space of inhomogeneous Erdős-Rényi random graphs equipped with the Hamming distance. In that case, one can show that the population Fréchet mean and median coincide [75], and the adjacency matrix of the population Fréchet mean graph, $\boldsymbol{\mu}\big[\boldsymbol{A}\big]$, is given by the majority rule,

$$\Big[\boldsymbol{\mu}\big[\boldsymbol{A}\big]\Big]_{ij} = \begin{cases} 1 & \text{if} \quad p_{ij} > 1/2, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{4.20}$$

Also, the population Fréchet function, F_2, evaluated at $\boldsymbol{\mu}\big[\boldsymbol{A}\big]$ is given by [75]

$$F_2(\boldsymbol{\mu}\big[\boldsymbol{A}\big]) = \Big[\sum_{1\le i<j\le n} p_{ij} - \sum_{(i,j)\in\mathcal{E}\left(\boldsymbol{\mu}\left[\boldsymbol{A}\right]\right)} (2p_{ij}-1)\Big]^2 + \sum_{1\le i<j\le n} p_{ij}(1-p_{ij}), \tag{4.21}$$

where $\mathcal{E}\big(\boldsymbol{\mu}\big[\boldsymbol{A}\big]\big)$ is the set of edges of the population Fréchet mean, $\boldsymbol{\mu}\big[\boldsymbol{A}\big]$. We claim that the lower bound on $\widehat{F}_2\big(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]\big)$ in (4.19),

$$\big[\overline{e}_N - e_{\widehat{\boldsymbol{\mu}}}\big]^2, \tag{4.22}$$

can be identified with the first term of $F_2(\boldsymbol{\mu}\big[\boldsymbol{A}\big])$ in (4.21),

$$\Big[\sum_{1\le i<j\le n} p_{ij} - \sum_{(i,j)\in\mathcal{E}\left(\boldsymbol{\mu}\left[\boldsymbol{A}\right]\right)} (2p_{ij}-1)\Big]^2. \tag{4.23}$$

Indeed, the first sum inside (4.23) is the population mean number of edges, $\mathbb{E}\,[e]$, which matches the sample mean $\overline{e}_N$ in (4.22). Also, the second sum in (4.23) is bounded by $e\big(\boldsymbol{\mu}\big[\boldsymbol{A}\big]\big)$, the number of edges of the population Fréchet mean,

$$0 < \sum_{(i,j)\in\mathcal{E}\left(\boldsymbol{\mu}\left[\boldsymbol{A}\right]\right)} (2p_{ij}-1) < \sum_{(i,j)\in\mathcal{E}\left(\boldsymbol{\mu}\left[\boldsymbol{A}\right]\right)} 1 = e\big(\boldsymbol{\mu}\big[\boldsymbol{A}\big]\big). \tag{4.24}$$

The number of edges $e\big(\boldsymbol{\mu}\big[\boldsymbol{A}\big]\big)$ matches the sample estimate, $e_{\widehat{\boldsymbol{\mu}}}$, in (4.22). In summary, the first term (4.23) of the population Fréchet function (4.21) matches the corresponding sample estimate (4.22).

However, the second term, $\sum_{1\leq i<j\leq n} p_{ij}(1-p_{ij})$ in (4.21), which accounts for the variance of the $n(n-1)/2$ independent Bernoulli edges, is not present in the lower bound on in $F_2[\boldsymbol{\mu}\big[\boldsymbol{A}\big]]$ given by (4.19), confirming that the lower bound in (4.19) is missing a variance term and is therefore not tight.

Proof of Lemma 8 Because of lemma 7, we have

$$\big|e\big(\boldsymbol{A}^{(k)}\big) - e_{\widehat{\boldsymbol{\mu}}}\big|^2 \leq d_H^2(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big], \boldsymbol{A}^{(k)}). \tag{4.25}$$

Now, the function

$$x \longmapsto \big(e_{\widehat{\boldsymbol{\mu}}} - x\big)^2 \tag{4.26}$$

is strictly convex, so,

$$\big|\overline{e}_N - e_{\widehat{\boldsymbol{\mu}}}\big|^2 = \left|\frac{1}{N}\sum_{k=1}^{N} e\big(\boldsymbol{A}^{(k)}\big) - e_{\widehat{\boldsymbol{\mu}}}\right|^2 < \frac{1}{N}\sum_{k=1}^{N}\Big|e\big(\boldsymbol{A}^{(k)}\big) - e_{\widehat{\boldsymbol{\mu}}}\Big|^2, \tag{4.27}$$

and substituting (4.25) for each k in (4.27), we get the advertised result. □

Finally, we compute an upper bound on the Fréchet function evaluated at a sample Fréchet median graph, $\widehat{F}_2\big(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]\big)$.

Lemma 9 Let $\overline{e}_N$ and $\sigma_N^2(e)$ be the sample mean, and variance of the number of edges (see (4.8)). Then the Fréchet function $\widehat{F}_2\big(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]\big)$ evaluated at a Fréchet median graph is bounded by

$$\widehat{F}_2\big(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]\big) \leq 2\big[\overline{e}_N\big]^2 + \sigma_N^2(e). \tag{4.28}$$

Remark 19 As explained in Remark 18, when the graphs $G^{(k)}$ are sampled from $\mathcal{G}\big(n, \boldsymbol{P}\big)$, then the population Fréchet mean and median graphs coincide, $\boldsymbol{\mu}\big[G\big] = \boldsymbol{m}\big[G\big]$. Also, the population Fréchet function $F_2(\boldsymbol{m}\big[\boldsymbol{A}\big])$ evaluated at a population Fréchet median graph is given by

$$F_2\big[\boldsymbol{m}\big[\boldsymbol{A}\big]\big] = \Big[\sum_{1\leq i<j\leq n} p_{ij} - \sum_{(i,j)\in\mathcal{E}\left(\boldsymbol{m}\left[\boldsymbol{A}\right]\right)} (2p_{ij}-1)\Big]^2 + \sum_{1\leq i<j\leq n} p_{ij}(1-p_{ij}), \tag{4.29}$$

where the term $\sum_{(i,j)\in\mathcal{E}\left(\boldsymbol{m}\left[\boldsymbol{A}\right]\right)}(2p_{ij}-1)$ is always positive (since the median graphs are constructed using the majority rule (4.12)). Therefore, we have

$$F_2\left[\boldsymbol{m}\left[\boldsymbol{A}\right]\right] \leq \Big[\sum_{1\leq i<j\leq n} p_{ij}\Big]^2 + \sum_{1\leq i<j\leq n} p_{ij}(1-p_{ij}). \tag{4.30}$$

The term $\sum_{1\leq i<j\leq n} p_{ij}$ is the expectation of the number of edges, whereas $\sum_{1\leq i<j\leq n} p_{ij}(1-p_{ij})$ is the variance of the number of edges. In summary, we have the following bound on the population Fréchet function,

$$F_2(\boldsymbol{m}\left[\boldsymbol{A}\right]) \leq \left[\mathbb{E}\left[e\right]\right]^2 + \operatorname{var}\left[e\right], \tag{4.31}$$

where e denotes the number of edges in graphs sampled from $\mathcal{G}\big(n, \boldsymbol{P}\big)$. If we replace $\mathbb{E}\left[e\right]$ and $\operatorname{var}\left[e\right]$ by their respective sample estimates, $\overline{e}_N$ and $\sigma_N^2(e)$, then the bound (4.28) is only slightly worse (by a factor 2 in front of $\overline{e}_N$) than the population bound, (4.31). Interestingly, the variance of the number of edges is present in both expressions.

Proof of Lemma 9 From (4.5), one can derive the following expression for the Hamming distance from a Fréchet median graph $\widehat{\boldsymbol{m}}_N\left[G\right]$ to a graph $G^{(k)}$,

$$d_H(\widehat{\boldsymbol{m}}_N\left[\boldsymbol{A}\right], \boldsymbol{A}^{(k)}) = e_{\widehat{\boldsymbol{m}}} + e\big(\boldsymbol{A}^{(k)}\big) - 2\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} a_{ij}^{(k)}, \tag{4.32}$$

where we recall that $\mathcal{E}_{\widehat{\boldsymbol{m}}} = \Big\{(i,j),\ i<j,\ \Big[\widehat{\boldsymbol{m}}_N\left[\boldsymbol{A}\right]\Big]_{ij} = 1\Big\}$ is the set of edges of $\widehat{\boldsymbol{m}}_N\left[G\right]$. Taking the square of the Hamming distance given by (4.32), and summing over all the graphs, yields

$$\begin{aligned}\widehat{F}_2\big(\widehat{\boldsymbol{m}}_N\left[\boldsymbol{A}\right]\big) = \frac{1}{N}\sum_{k=1}^{N}\Bigg\{&\Big[e_{\widehat{\boldsymbol{m}}} + e\big(\boldsymbol{A}^{(k)}\big)\Big]^2 + 4\Big[\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} a_{ij}^{(k)}\Big]^2 \\ &- 4\big(e_{\widehat{\boldsymbol{m}}} + e\big(\boldsymbol{A}^{(k)}\big)\big)\Big[\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} a_{ij}^{(k)}\Big]\Bigg\}.\end{aligned}$$

Expanding all the terms, and using the definition of $\sigma_N^2(e)$ and $\overline{e}_N$ in (4.8), we get

$$\begin{aligned}\widehat{F}_2\big(\widehat{\boldsymbol{m}}_N[\boldsymbol{A}]\big) =& \big[e_{\widehat{\boldsymbol{m}}}\big]^2 + 2e_{\widehat{\boldsymbol{m}}}\,\overline{e}_N + \sigma_N^2(e) + \big[\overline{e}_N\big]^2 + \frac{4}{N}\sum_{k=1}^{N}\Big[\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} a_{ij}^{(k)}\Big]^2 \\ &- \frac{4}{N}\sum_{k=1}^{N} e\big(\boldsymbol{A}^{(k)}\big)\Big[\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} a_{ij}^{(k)}\Big] - 4e_{\widehat{\boldsymbol{m}}}\Big[\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} \frac{1}{N}\sum_{k=1}^{N} a_{ij}^{(k)}\Big] \\ =& \big[e_{\widehat{\boldsymbol{m}}} + \overline{e}_N\big]^2 + \sigma_N^2(e) + 4\frac{1}{N}\sum_{k=1}^{N}\Big[\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} a_{ij}^{(k)}\Big]^2 \\ &- \frac{4}{N}\sum_{k=1}^{N} e\big(\boldsymbol{A}^{(k)}\big)\Big[\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} a_{ij}^{(k)}\Big] - 4e_{\widehat{\boldsymbol{m}}}\Big[\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} \frac{1}{N}\sum_{k=1}^{N} a_{ij}^{(k)}\Big]. \end{aligned} \tag{4.33}$$

Now, because of the definition of the median graphs (4.12), we have the following upper bound

$$-4e_{\widehat{\boldsymbol{m}}}\Big[\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} \frac{1}{N}\sum_{k=1}^{N} a_{ij}^{(k)}\Big] \leq -2\big[e_{\widehat{\boldsymbol{m}}}\big]^2. \tag{4.34}$$

Because $e\big(\boldsymbol{A}^{(k)}\big) \geq \sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} a_{ij}^{(k)}$, we get the following upper bound,

$$-4\sum_{k=1}^{N} e\big(\boldsymbol{A}^{(k)}\big)\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} a_{ij}^{(k)} \leq -4\sum_{k=1}^{N}\Big[\sum_{(i,j)\in\mathcal{E}_{\widehat{\boldsymbol{m}}}} a_{ij}^{(k)}\Big]^2. \tag{4.35}$$

Finally, after substituting (4.34) and (4.35) into (4.33), we get the bound announced in the lemma,

$$\begin{aligned}\widehat{F}_2\big(\widehat{\boldsymbol{m}}_N[\boldsymbol{A}]\big) \leq& \big[e_{\widehat{\boldsymbol{m}}} + \overline{e}_N\big]^2 - 2\big[e_{\widehat{\boldsymbol{m}}}\big]^2 + \sigma_N^2(e) = -\big[e_{\widehat{\boldsymbol{m}}} - \overline{e}_N\big]^2 + 2\big[\overline{e}_N\big]^2 + \sigma_N^2(e) \\ \leq& 2\big[\overline{e}_N\big]^2 + \sigma_N^2(e). \end{aligned}$$

□

4.4.3 The number of edges of the median and mean graph when $d = d_H$

The following lemma provides the bounds given by Theorem 9 when d is the Hamming distance.

Lemma 10 Let $\big\{G^{(1)}, \ldots, G^{(N)}\big\}$ be a sample of unweighted simple labeled graphs with vertex set $\{1, \ldots, n\}$. Let $\widehat{\boldsymbol{\mu}}_N[\boldsymbol{A}]$ be the adjacency matrix of a sample Fréchet mean graph, and $\widehat{\boldsymbol{m}}_N[\boldsymbol{A}]$ be the adjacency matrix of a sample Fréchet median graph, computed according to the Hamming distance. Then

$$e\big(\widehat{\boldsymbol{\mu}}_N[\boldsymbol{A}]\big) < 2\overline{e}_N + \frac{\sigma_N(e)}{\sqrt{2}}, \quad \text{and} \quad e\big(\widehat{\boldsymbol{m}}_N[\boldsymbol{A}]\big) \leq 2\overline{e}_N. \tag{4.36}$$

Proof of Lemma 10 The bound on $e\big(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]\big)$ is a straightforward consequence of lemma 8. Indeed, (4.13) and (4.8) yield the bound in (4.36),

$$e\big(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]\big) \leq \frac{2}{N}\sum_{k=1}^{N} e\big(\boldsymbol{A}^{(k)}\big) \leq 2\overline{e}_N.$$

We now move to $e\big(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]\big)$. We use $\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]$ to derive an upper bound on the Fréchet function computed at $\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]$. By definition of the sample Fréchet mean graphs, we have

$$\frac{1}{N}\sum_{k=1}^{N} d_H^2(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big], \boldsymbol{A}^{(k)}) \leq \frac{1}{N}\sum_{k=1}^{N} d_H^2(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big], \boldsymbol{A}^{(k)}). \tag{4.37}$$

Using (4.19) as a lower bound and (4.28) as an upper bound in (4.37), we get

$$\Big[e_{\widehat{\boldsymbol{\mu}}} - \overline{e}_N\Big]^2 < 2\big[\overline{e}_N\big]^2 + \sigma_N^2(e),$$

and thus

$$\big|e_{\widehat{\boldsymbol{\mu}}} - \overline{e}_N\big| \leq \sqrt{2\big[\overline{e}_N\big]^2 + \sigma_N^2(e)} \leq \frac{1}{\sqrt{2}}\Big\{\sqrt{2}\overline{e}_N + \sigma_N(e)\Big\} = \overline{e}_N + \frac{\sigma_N(e)}{\sqrt{2}}, \tag{4.38}$$

from which we get the advertised bound on $e_{\widehat{\boldsymbol{\mu}}}$. □

4.4.4 The mean graphs computed using the adjacency spectral pseudometric

The technical difficulty in defining the sample Fréchet mean and median graphs according to the adjacency spectral pseudometric stems from the fact that the sample Fréchet function, $\widehat{F}_q(\boldsymbol{A})$, is defined in the spectral domain, but the domain over which the optimization takes place is the matrix domain. This leads to the definition of the set, Λ, of real spectra that are realizable by adjacency matrices of unweighted graphs (elements of $\mathcal{S}$, defined by (4.3)) [62],

$$\Lambda = \left\{\boldsymbol{\lambda}(\boldsymbol{A}) = \Big[\lambda_1(\boldsymbol{A}) \quad \ldots \quad \lambda_n(\boldsymbol{A})\Big]\,; \text{where} \boldsymbol{A} \in \mathcal{S}\right\}. \tag{4.39}$$

Let $\big\{G^{(1)}, \ldots, G^{(N)}\big\}$ be a sample of unweighted simple labeled graphs with vertex set $\{1, \ldots, n\}$. Let $\boldsymbol{A}^{(k)}$ be the adjacency matrix of graph $G^{(k)}$, and let $\boldsymbol{\lambda}(\boldsymbol{A}^{(k)})$ be the spectrum of $\boldsymbol{A}^{(k)}$. The adjacency matrix, $\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]$, of a sample Fréchet mean graph computed according to the adjacency spectral pseudometric, has a vector of eigenvalues, $\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]) \in \Lambda$, that satisfies

$$\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]) = \underset{\boldsymbol{\lambda}\in\Lambda}{\operatorname{argmin}} \sum_{k=1}^{N} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}(\boldsymbol{A}^{(k)})||^2. \tag{4.40}$$

Similarly, the adjacency matrix, $\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]$, of a sample Fréchet median computed according to the adjacency spectral pseudometric, has a vector of eigenvalues, $\boldsymbol{\lambda}(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]) \in \Lambda$, that satisfies

$$\boldsymbol{\lambda}(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]) = \underset{\boldsymbol{\lambda}\in\Lambda}{\operatorname{argmin}} \sum_{k=1}^{N} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}(\boldsymbol{A}^{(k)})||. \tag{4.41}$$

We recall the following result that expresses the number of edges as a function of the ℓ^2 norm of the spectrum of the adjacency matrix.

Lemma 11 Let $G \in \mathcal{G}$ with adjacency matrix $\boldsymbol{A}$. Let $\lambda_1(\boldsymbol{A}) \geq \ldots \geq \lambda_n(\boldsymbol{A})$ be the eigenvalues of $\boldsymbol{A}$. Then

$$2e\big(\boldsymbol{A}\big) = \sum_{i=1}^{n} \lambda_i^2(\boldsymbol{A}) = \|\boldsymbol{\lambda}(\boldsymbol{A})\|_2^2. \tag{4.42}$$

Proof of Lemma 11 The result is classic; see for instance [8, 99].

We derive the following lower bound on the sample mean number of edges.

Lemma 12 Let $\widehat{\mathbb{E}}_N\big[\boldsymbol{\lambda}(\boldsymbol{A})\big] = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\lambda}(\boldsymbol{A}^{(k)})$ be the sample mean spectrum. Then

$$\frac{1}{2}\Big\|\widehat{\mathbb{E}}_N\big[\boldsymbol{\lambda}(\boldsymbol{A})\big]\Big\|^2 \leq \overline{e}_N, \tag{4.43}$$

where $\overline{e}_N$ is the sample mean number of edges, given by (4.8).

Proof of Lemma 12 The result is a straightforward consequence of the convexity of the norm combined with (4.42).

If Λ were to be a convex set, then the spectrum of a sample Fréchet mean graph would be the sample mean spectrum, which would minimize (4.40). Unfortunately, Λ is not convex [66]. We can nevertheless relate the spectrum of a sample Fréchet mean graph, $\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big])$, to the mean spectrum $\widehat{\mathbb{E}}_N\big[\boldsymbol{\lambda}(\boldsymbol{A})\big]$. We take a short detour to build some intuition about the geometric position of the spectrum of $\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]$ with respect to $\boldsymbol{\lambda}(\boldsymbol{A}^{(1)}), \ldots, \boldsymbol{\lambda}(\boldsymbol{A}^{(N)})$.

4.4.4.1 Warm-up: The Sample Mean Spectrum.

Let $\{G^{(1)}, \ldots, G^{(N)}\}$ be a sample of unweighted simple labeled graphs with vertex set $\{1, \ldots, n\}$. Let $\boldsymbol{A}^{(k)}$ be the adjacency matrix of graph $G^{(k)}$, and let $\boldsymbol{\lambda}(\boldsymbol{A}^{(k)})$ be the spectrum of $\boldsymbol{A}^{(k)}$.

Lemma 13 Let $\widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\boldsymbol{A})]$ be the sample mean spectrum. Then $\exists\, k_0 \in \{1, \ldots, N\}$ such that

$$\|\boldsymbol{\lambda}(\boldsymbol{A}^{(k_0)})\| \leq \|\widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\boldsymbol{A})]\|. \tag{4.44}$$

Proof of Lemma 13 A proof by contradiction is elementary.

Using the characterization of a sample Fréchet mean graph, $\widehat{\boldsymbol{\mu}}_N[\boldsymbol{A}]$, given by (4.40), we can extend the above lemma to $\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\boldsymbol{A}])$, and derive the following result.

Lemma 14 Let $\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\boldsymbol{A}])$ be the spectrum of a sample Fréchet mean graph. Let $\overline{e}_N$ be the sample mean number of edges of the graphs $G^{(1)}, \ldots, G^{(N)}$. Then

$$\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\boldsymbol{A}])\| \leq 3\sqrt{2\overline{e}_N}. \tag{4.45}$$

Proof of Lemma 14 Because of lemma 13,

$$\exists\, k_0 \in \{1, \ldots, N\},\ \|\boldsymbol{\lambda}(\boldsymbol{A}^{(k_0)})\| \leq \|\widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\boldsymbol{A})]\|. \tag{4.46}$$

Now, because of lemma 12,(4.46) implies that

$$\|\boldsymbol{\lambda}(\boldsymbol{A}^{(k_0)})\| \leq \sqrt{2\overline{e}_N}. \tag{4.47}$$

Because the vector $\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k_0)}\big)$ is in Λ (defined by (4.39)), we have

$$\frac{1}{N}\sum_{k=1}^{N} \|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\boldsymbol{A}]) - \boldsymbol{\lambda}\big(\boldsymbol{A}^{(k)}\big)\|^2 \leq \frac{1}{N}\sum_{k=1}^{N} \|\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k_0)}\big) - \boldsymbol{\lambda}\big(\boldsymbol{A}^{(k)}\big)\|^2.$$

Expanding the norms squared on both sides yields

$$\begin{aligned}\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\boldsymbol{A}])\|^2 - 2\langle \boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\boldsymbol{A}]), \widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\boldsymbol{A})]\rangle + \frac{1}{N}\sum_{k=1}^{N} \|\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k)}\big)\|^2 \\ \leq \|\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k_0)}\big)\|^2 - 2\langle \boldsymbol{\lambda}\big(\boldsymbol{A}^{(k_0)}\big), \widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\boldsymbol{A})]\rangle + \frac{1}{N}\sum_{k=1}^{N} \|\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k)}\big)\|^2.\end{aligned} \tag{4.48}$$

Subtracting $\frac{1}{N}\sum_{k=1}^{N}\|\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k)}\big)\|^2$ and adding $\big\|\widehat{\mathbb{E}}_N\big[\boldsymbol{\lambda}(\boldsymbol{A})\big]\big\|^2$ on both sides we get

$$\big\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big])-\widehat{\mathbb{E}}_N\big[\boldsymbol{\lambda}(\boldsymbol{A})\big]\big\|^2\leq\big\|\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k_0)}\big)-\widehat{\mathbb{E}}_N\big[\boldsymbol{\lambda}(\boldsymbol{A})\big]\big\|^2,$$

and therefore

$$\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big])\|\leq\|\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k_0)}\big)\|+2\big\|\widehat{\mathbb{E}}_N\big[\boldsymbol{\lambda}(\boldsymbol{A})\big]\big\|. \tag{4.49}$$

Finally, using lemma 12 and (4.47) in the equation above, we obtain

$$\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big])\|\leq 3\sqrt{2\overline{e}_N}, \tag{4.50}$$

which completes the proof of the bound on the spectrum of the Fréchet mean. □

4.4.5 The median graphs computed using the adjacency spectral pseudometric

We finally consider the computation of the median graphs. We have the following bound on the norm of the spectrum of $\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]$.

Lemma 15 Let $\boldsymbol{\lambda}(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big])$ be the spectrum of a sample Fréchet median graph. Let $\overline{e}_N$ be the sample mean number of edges of the graphs $G^{(1)},\ldots,G^{(N)}$. Then,

$$\|\boldsymbol{\lambda}(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big])\|\leq 3\sqrt{2\overline{e}_N}. \tag{4.51}$$

Proof of Lemma 15 The function Φ,

$$\begin{aligned}\Phi:\mathbb{R}^n&\longrightarrow[0,\infty)\\ \boldsymbol{x}&\longmapsto\Phi(\boldsymbol{x})=\big\|\boldsymbol{\lambda}(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big])-\boldsymbol{x}\big\|\end{aligned}$$

is strictly convex, and therefore

$$\Phi\big(\widehat{\mathbb{E}}_N\big[\boldsymbol{\lambda}(\boldsymbol{A})\big]\big)=\Phi\left(\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k)}\big)\right)\leq\frac{1}{N}\sum_{k=1}^{N}\Phi\left(\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k)}\big)\right). \tag{4.52}$$

Now, the right-hand side of (4.52) is the Fréchet function evaluated at one of its minimizers. Thus $F_1(\boldsymbol{\lambda}(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]))$, is smaller than $F_1(\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k_0)}\big))$, where $\boldsymbol{A}^{(k_0)}$ is defined in lemma 13, and (4.52) becomes

$$\|\boldsymbol{\lambda}(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big])-\widehat{\mathbb{E}}_N\big[\boldsymbol{\lambda}(\boldsymbol{A})\big]\|\leq\frac{1}{N}\sum_{k=1}^{N}\|\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k_0)}\big)-\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k)}\big)\|. \tag{4.53}$$

Also, because of lemma 12 and (4.47), we get

$$\frac{1}{N}\sum_{k=1}^{N}\|\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k_0)}\big)-\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k)}\big)\|\leq\|\boldsymbol{\lambda}\big(\boldsymbol{A}^{(k_0)}\big)\|+\sqrt{2\overline{e}_N}\leq 2\sqrt{2\overline{e}_N}. \tag{4.54}$$

Combining (4.53) and (4.54), and using lemma 12 we conclude that

$$\big\|\boldsymbol{\lambda}(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big])\big\|\leq\big\|\widehat{\mathbb{E}}_N\big[\boldsymbol{\lambda}(\boldsymbol{A})\big]\big\|+2\sqrt{2\overline{e}_N}\leq 3\sqrt{2\overline{e}_N}.$$

This completes the proof of the bound on the spectrum of a Fréchet median. □

4.4.6 The number of edges of the median and mean graph when $d = d_\lambda$

The following lemma provides the bounds given by Theorem 9 when d is the spectral adjacency pseudometric.

Lemma 16 Let $\big\{G^{(1)},\ldots,G^{(N)}\big\}$ be a sample of unweighted simple labeled graphs with vertex set $\{1,\ldots,n\}$. We consider a sample Fréchet mean, $\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]$, and a sample Fréchet median, $\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]$, computed according to the spectral adjacency pseudometric. Then

$$\max\big\{e\big(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]\big),e\big(\widehat{\boldsymbol{m}}_N\big[\boldsymbol{A}\big]\big)\big\}\leq 9\,\overline{e}_N, \tag{4.55}$$

where $\overline{e}_N$ is the sample mean number of edges given by (4.8).

Proof of Lemma 16 We first analyse the case of a sample Fréchet mean graph; a sample Fréchet median graph is handled similarly. From lemmata 14 and 15, we have

$$\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big])\|^2\leq 18\,\overline{e}_N. \tag{4.56}$$

Now, from (4.42) we have $e\big(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]\big)=\frac{1}{2}\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big])\|^2$, and therefore

$$e\big(\widehat{\boldsymbol{\mu}}_N\big[\boldsymbol{A}\big]\big)\leq 9\,\overline{e}_N,$$

which completes the proof of the lemma. □

Chapter 5

Conclusions

5.1 Summary

The topic of machine learning is responsible for an abundance of modern research. When the data is not Euclidean, as is the case in this dissertation, the field of machine learning is still in its infancy. Notably, many machine learning algorithms developed for Euclidean data rely on the mean and variance and can be generalized to use the notions of Fréchet mean and Fréchet variance. This indicates that such algorithms can be used for data sets of graphs. However, the Fréchet mean and variance for data sets of graphs are not trivial to compute, as shown throughout Chapter 2. Computing the sample alternatives of the Fréchet mean and variance for data sets of graphs is a preliminary step to implementing many machine learning algorithms. Examples of such algorithms showcased in this dissertation include linear regression, K-means clustering, and kernel density estimation.

Throughout this dissertation, we consider a metric defined on the spectra of the adjacency matrix of a graph. Therefore, this work is complementary to the studies on the computation of the Fréchet mean with respect to the Hamming distance (see, e.g., [19, 41, 56, 55, 60, 88, 87, 59]). To the best of our knowledge, this study provides the first quantification of the Fréchet mean and variance for sets of graphs with respect to a spectral metric, which characterizes the global structures in a graph rather than the local structures. We also provide methods to generate new graphs with respect to the properties measured by the adjacency spectral pseudo-metric (see Chapter 3).

5.2 Choice of metric

The choice of metric for sets of graphs deserves an in-depth discussion. Here the relationship between the metric considered throughout this dissertation and the technique taken to approximate the solution to the sample Fréchet mean problem is discussed. The question of how the method generalizes to other choices of metrics is briefly addressed.

The proposed solution first translates the Fréchet mean problem from the space of graphs, $\mathcal{G}$, to the space of probability measures, $\mathcal{M}(\mathcal{G})$. An approximate problem is then introduced by restricting to an appropriate subset of parametrizable probability measures. Note that the specific subset of probability measures is chosen in conjunction with the choice of metric equipped to $\mathcal{G}$. For different choices of metrics, different subsets of probability measures may be considered to approximate the solution to the Fréchet mean problem.

First, recall the sample Fréchet mean problem, defined for a general metric. Let $\{G^{(k)}\}_{k=1}^{N}$ be a set of graphs with sample Fréchet mean given by

$$G_N^* = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \frac{1}{N} \sum_{k=1}^{N} d^2(G, G^{(k)}). \tag{5.1}$$

To determine the graph G_N^*, two intermediate steps are taken. First, the minimization problem is lifted to the space of probability measures $\mathcal{M}(\mathcal{G})$ where a search for a probability measure with the correct sample Fréchet mean can be performed. This process is done independently of the specific choice of metric. Next, we define an approximate problem in the space of probability measures and solve this approximate problem.

To outline these tasks, we first need to characterize the Fréchet mean of a probability measure defined on $\mathcal{G}$. This has been defined in the introduction, but we recall the definition below due to its relevance for this discussion. Let $\nu \in \mathcal{M}(\mathcal{G})$ be an arbitrary probability measure with Fréchet mean

$$G_\nu^* = \underset{G \in \mathcal{G}}{\operatorname{argmin}} \mathbb{E}_\nu\big[d^2(G, G_\nu)\big] \tag{5.2}$$

where the subscript ν denotes this is the Fréchet mean graph given the probability measure ν.

First, we observe that there exists at least one probability measure ν^* with Fréchet mean $G^*_{\nu^*}$ such that

$$d(G^*_{\nu^*}, G^*_N) = 0. \tag{5.3}$$

This can be shown by considering the measure ν^* defined below. Let $M \subset \mathcal{G}$ be a subset of graphs. Let ν^* be such that

$$\nu^*(M) = \begin{cases} 1 & \text{if } G^*_N \in M \\ 0 & \text{if } G^*_N \notin M. \end{cases} \tag{5.4}$$

The Fréchet mean of ν^* is trivially given by G^*_N, the sample Fréchet mean graph of the data set. Therefore, rather than search for the solution to equation (5.1), which requires searching over the space of graphs, we instead search over the space of probability measures and minimize the following objective,

$$\nu^* = \underset{\nu \in \mathcal{M}(\mathcal{G})}{\operatorname{argmin}}\ d^2\left(G^*_N, \underset{G \in \mathcal{G}}{\operatorname{argmin}}\ \mathbb{E}_\nu\left[d^2\left(G, G_\nu\right)\right]\right) \tag{5.5}$$

$$= \underset{\nu \in \mathcal{M}(\mathcal{G})}{\operatorname{argmin}}\ d^2\left(G^*_N, G^*_\nu\right) \tag{5.6}$$

After solving equation (5.5), the result is a probability measure with the correct Fréchet mean graph. In general, the solution to equation (5.5) is not unique, however, finding one such distribution with the correct Fréchet mean is satisfactory since we can then determine the graph G^*_N. For this reason, we write the solution to equation (5.5) as a unique probability measure. Equation (5.5) accomplishes the first part of the solution technique we employ when solving the sample Fréchet mean problem.

To define an approximate problem, we restrict the space of probability measures to a subset. Let $\mathcal{M}_s(\mathcal{G}) \subset \mathcal{M}(\mathcal{G})$ be any subset of probability measures defined on the space $\mathcal{G}$ and consider the problem

$$\nu^*_s = \underset{\nu_s \in \mathcal{M}_s(\mathcal{G})}{\operatorname{argmin}}\ d^2\left(G^*_N, \underset{G \in \mathcal{G}}{\operatorname{argmin}}\ \mathbb{E}_{\nu_s}\left[d^2\left(G, G_{\nu_s}\right)\right]\right) \tag{5.7}$$

$$= \underset{\nu_s \in \mathcal{M}_s(\mathcal{G})}{\operatorname{argmin}}\ d^2\left(G^*_N, G^*_{\nu_s}\right). \tag{5.8}$$

For certain subsets of probability measures, the following equation is satisfied,

$$d\left(G^*_{\nu^*}, G^*_{\nu^*_s}\right) = 0. \tag{5.9}$$

This accomplishes the second step to determine the sample Fréchet mean graph G^*_N.

When considering equation (5.5), we first note that the choice of metric d has not been specified. Consequently, these steps can be taken regardless of the specific metric, examples of different metric choices in this dissertation include the metrics d_H, d_A, or d_{A_c}. Second, each graph only ever appears as an argument of a metric. Therefore, to minimize the objective, only specific properties of the graphs appearing in equation (5.5) need to be considered. We can make this observation explicit by considering the metric d_{A_c}. For this choice of metric,

$$\nu^* = \underset{\nu \in \mathcal{M}(\mathcal{G})}{\operatorname{argmin}}\, d^2_{A_c}\left(G^*_N, G_\nu\right) \tag{5.10}$$

$$= \underset{\nu \in \mathcal{M}(\mathcal{G})}{\operatorname{argmin}}\, ||\sigma_c(\boldsymbol{A}^*_N) - \sigma_c(\boldsymbol{A}_\nu)||^2_2 \tag{5.11}$$

where $\boldsymbol{A}^*_N$ and $\boldsymbol{A}_\nu$ denote the adjacency matrices of the graphs G^*_N and G_ν respectively. To minimize the objective function, all that is needed is an understanding of the largest c eigenvalues of $\boldsymbol{A}^*_N$ and the largest c eigenvalues of $\boldsymbol{A}_\nu$.

The eigenvalues of $\boldsymbol{A}^*_N$, for large graphs, are characterized by the arithmetic mean of the observed eigenvalues of the adjacency matrices of the graphs in the sample set $\{G^{(k)}\}^N_{k=1}$ (see Theorem 2).

The largest eigenvalues of $\boldsymbol{A}_\nu$ are a function of the measure ν. By defining the subset of probability measures $\mathcal{M}_s(\mathcal{G})$ to be the family of canonical stochastic block model kernel probability measures, the expected value of the largest c eigenvalues of the adjacency matrices of graphs sampled according to such a kernel probability measure are well-understood, as shown in the studies of [4, 12, 22]. Thus, the family of stochastic block models is well-paired to the choice of metric d_{A_c} and the approximate problem, equation (5.7) can be solved using the results of the studies mentioned above.

To prove that the solution to equation (5.7) is arbitrarily close to the graph G^*_N, when

considering the family of stochastic block model kernel probability measures, it is shown that there exists a stochastic block model kernel probability measure which assigns a non-zero probability to a graph whose adjacency matrix has eigenvalues close to the arithmetic mean of the eigenvalues of the adjacency matrices of the graphs in the sample, $\{G^{(k)}\}_{k=1}^{N}$. This results from the convergence in distribution of the largest c eigenvalues of graphs sampled according to a stochastic block model kernel probability measure which were studied in the works of [4, 12, 22].

A natural question is whether the results of this dissertation generalize easily to the spectra of a graph's Laplacian or normalized Laplacian matrix. A first step is to quantify the spectra of the sample Fréchet mean graph's (normalized) Laplacian matrix in a similar manner to Theorem 2. The next step is to understand the expected eigenvalues of the Laplacian (or normalized Laplacian) matrices of graphs sampled according to a stochastic block model kernel probability measure. Currently, the spectra of these matrices are not as well understood as the spectra of the adjacency matrices indicating that the results presented in Chapter 2 do not immediately generalize to metrics defined for the spectra of a graph's (normalized) Laplacian matrix. However, because the (normalized) Laplacian matrix can be written as a deterministic function of the adjacency matrix of a graph, it may be possible to show similar results to those given in Chapter 2 in the future.

5.3 General kernel probability measures

This dissertation focuses on the stochastic block model family of probability measures. An extension of this work may consider different families of probability measures rather than the stochastic block model. The family of stochastic block model kernel probability measures is a specific subset of kernel probability measures. First, it is worth noting that an inhomogeneous Erdős-Rényi random graph model is equivalent to the kernel probability measure. Therefore, the literature on inhomogeneous Erdős-Rényi random graphs provides insight into the behavior of graphs sampled according to a kernel probability measure $\mu_{\omega_n f}$. A general kernel probability measure is defined by a symmetric function $f : [0,1]^2 \mapsto [0,1]$ and a constant ω_n. Furthermore, it is shown in [22] that the largest eigenvalues of graphs sampled according to $\mu_{\omega_n f}$ are characterized

in terms of the eigenvalues and eigenfunctions of a linear integral operator with kernel function f. Let θ_i and $r_i(x)$ denote the eigenvalues and eigenfunctions of the operator L_f defined below by its action on a function $g \in L^2([0,1])$

$$L_f(g) = \int_0^1 f(x,y)g(y)dy = \int_0^1 \sum_{i=1}^{c} \theta_i r_i(x) r_i(y) g(y) dy. \tag{5.12}$$

For stochastic block models, the geometry of the graphs sampled is prescribed by $r_i(x)$, the eigenfunctions of the corresponding integral operator L_f. To generalize this notion, the geometry of inhomogeneous Erdős-Rényi random graphs are characterized by the eigenfunctions $r_i(x)$, but now, the geometry need not specify a block structure, as in the case for the stochastic block model.

These observations imply that a generalization of Theorem 1 may hold for kernel probability measures where the eigenfunctions $r_i(x)$ need not be blockwise constant. The caveat, however, is that after fixing the geometry of the kernel probability measure, the functions $r_i(x)$, the expected density of graphs sampled according to $\mu_{\omega_n f}$ need not be the same as the density of the graph G from Theorem 1. When considering the stochastic block model family of kernels, the expected density of graphs sampled was corrected by the parameter q while the parameter $\boldsymbol{p}$ was chosen such that the largest eigenvalues of the adjacency matrix of the sample Fréchet mean graph given the measure $\mu_{\omega_n f}$ were similar to the largest eigenvalues of the adjacency matrix of G.

It is not immediately obvious how to prespecify the eigenfunctions $r_i(x)$ of a kernel probability measure such that both the eigenvalues and density of the Fréchet mean graph given the probability measure $\mu_{\omega_n f}$ are similar to the eigenvalues and density of the graph G. One possible suggestion is to consider the largest eigenvectors of the adjacency matrix of the graph G as a way to infer the eigenfunctions $r_i(x)$ for the linear integral operator L_f which in turn specifies the kernel of the probability measure.

5.4 Random graph models beyond the Bernoulli process

A classic approach when defining a random graph model is to model the entries of the adjacency matrix of a graph according to some distribution. In this dissertation, the distribution

which characterizes the entries of the adjacency matrix is defined by a Bernoulli random variable,

$$P(a_{ij} = 1) = p_{ij} \tag{5.13}$$

$$P(a_{ij} = 0) = 1 - p_{ij}. \tag{5.14}$$

For a kernel probability measure, the value of p_{ij} is specified by a symmetric function $f(x, y)$ so that

$$P(a_{ij} = 1) = p_{ij} = \omega_n f(\frac{i}{n}, \frac{j}{n}) \tag{5.15}$$

$$P(a_{ij} = 0) = 1 - p_{ij} = 1 - \omega_n f(\frac{i}{n}, \frac{j}{n}). \tag{5.16}$$

It is seen in Chapter 3 that prescribing the distribution of the entries of the adjacency matrices in this way leads to a coupling between the mean and the variance of the largest eigenvalues of the adjacency matrix of a graph sampled according to a kernel probability measure. It is shown that this class of models is essentially a one-parameter family with respect to the metric d_{A_c} since the Fréchet mean graph of the probability measure $\mu_{\omega_n f}$ specifies all information about the kernel probability measure. This is analogous to how the Bernoulli distribution is a one-parameter family and can be characterized entirely by its mean.

This suggests that a different model for the entries of the adjacency matrices may be beneficial when considering generative modeling for graphs. A model where the entries of the adjacency matrix are defined by a two-parameter family, a location parameter, and a scale parameter. Prescribing such a model may allow the mean and variance of the largest eigenvalues of the adjacency matrices of graphs sampled in this way to be adjusted independently.

This concludes a brief summary of the observations made within this dissertation.

"But where have we come? / And where shall we end?"

— *Patrick McHale, Into the Unknown*

Bibliography

[1] E. Abbe. Community detection and stochastic block models: recent developments. Journal of Machine Learning Research, 18(1):6446–6531, 2017.

[2] R. Albert and A Barabási. Statistical mechanices of complex networks. Reviews of Modern Physics, 74(1):47–97, 2002.

[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Gradient Flows in Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics ETH Zürich. Birkhäuser, 2. ed edition. OCLC: 254181287.

[4] Avanti Athreya, Joshua Cape, and Minh Tang. Eigenvalues of Stochastic Blockmodel Graphs and Random Graphs with Low-Rank Edge Probability Matrices. Sankhya A: The Indian Journal of Statistics, 84(1):36–63, June 2022.

[5] Konstantin Avrachenkov, Laura Cottatellucci, and Arun Kadavankandy. Spectral properties of random matrices for stochastic block model. In 2015 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), pages 537–544, 2015.

[6] Luca Baldesi, Carter T Butts, and Athina Markopoulou. Spectral graph forge: Graph generation targeting modularity. In IEEE INFOCOM 2018-IEEE Conference on Computer Communications, pages 1727–1735. IEEE, 2018.

[7] D. Banks and G. Constantine. Metric models for random graphs. Journal of Classification, pages 199–223, 1998.

[8] R.B. Bapat. Graphs and matrices. 27, 2010.

[9] Itziar Bardaji, Miquel Ferrer, and Alberto Sanfeliu. Computing the barycenter graph by means of the graph edit distance. In 2010 20th International Conference on Pattern Recognition, pages 962–965. IEEE, 2010.

[10] Miroslav Bačák. Computing medians and means in hadamard spaces. SIAM Journal on Optimization, 24(3):1542–1566, 2014.

[11] Mohsen Bayati, Margot Gerritsen, David F Gleich, Amin Saberi, and Ying Wang. Algorithms for large, sparse network alignment problems. In 2009 Ninth IEEE International Conference on Data Mining, pages 705–710. IEEE, 2009.

[12] F. Benyach-Georges, C. Bordenave, and A. Knowles. Largest eigenvalues of sparse inhomogeneous erdős–rényi graphs. The Annals of Probability, 47(3):1653–1676, 2019.

[13] Rabindra Nath Bhattacharya and Edward C Waymire. A basic course in probability theory, volume 69. Springer, 2007.

[14] Louis J Billera, Susan P Holmes, and Karen Vogtmann. Geometry of the space of phylogenetic trees. Advances in Applied Mathematics, 27(4):733–767, 2001.

[15] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. Random Structures and Algorithms, pages 3–122, 2007.

[16] A. Bonifati, I. Holubová, A. Prat-Pérez, and S. Sakr. Graph generators: State of the art and open challenges. arXiv preprint arXiv:2001.07906, 2020.

[17] Christian Borgs, Jennifer T Chayes, Henry Cohn, and László Miklós Lovász. Identifiability for graphexes and the weak kernel metric. In Building Bridges II, pages 29–157. Springer, 2019.

[18] Christian Borgs, Jennifer T Chayes, László Lovász, Vera T Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs ii. multiway cuts and statistical physics. Annals of Mathematics, pages 151–219, 2012.

[19] N. Boria, B. Negrevergne, and F. Yger. Fréchet mean computation in graph space through projected block gradient descent. ESANN, pages 705–710, 2020.

[20] Nicolas Boria, Sébastien Bougleux, Benoit Gaüzère, and Luc Brun. Generalized median graph via iterative alternate minimizations. In International Workshop on Graph-Based Representations in Pattern Recognition, pages 99–109. Springer, 2019.

[21] Gecia Bravo Hermsdorff and Lee Gunderson. A unifying framework for spectrum-preserving graph sparsification and coarsening. Advances in Neural Information Processing Systems, 32, 2019.

[22] A. Chakrabarty, A. Chakraborty, and R. Hazra. Eigenvalues outside the bulk of inhomogeneous erdös-rényi random graphs. Journal of Statistical Physics, 2020.

[23] Jie Chen, Yousef Saad, and Zechen Zhang. Graph coarsening: from scientific computing to machine learning. SeMA Journal, 79(1):187–223, 2022.

[24] Chenhui Deng, Zhiqiang Zhao, Yongyu Wang, Zhiru Zhang, and Zhuo Feng. Graphzoom: A multi-level spectral approach for accurate and scalable graph embedding. arXiv preprint arXiv:1910.02370, 2019.

[25] L. Devroye, L. Gyorfi, and G. Lugosi. A probabilistic theory of pattern recognition. Springer Science and Business Media, 2013.

[26] C. Donnat and S. Holmes. Tracking network dynamics: A survey using graph distances. The Annals of Applied Statistics, pages 971–1012, 2018.

[27] P. Dubey and H.G. Müller. Fréchet analysis of variance for random objects. Biometrika, 2017.

[28] Paromita Dubey and Hans-Georg Müller. Fréchet change-point detection. The Annals of Statistics, 48(6):3312–3335, 2020.

[29] László Erdős, Horng-Tzer Yau, and Jun Yin. Bulk universality for generalized wigner matrices. Probability Theory and Related Fields, 154(1):341–407, 2012.

[30] Jianqing Fan, Yingying Fan, Xiao Han, and Jinchi Lv. Asymptotic theory of eigenvectors for random matrices with diverging spikes. Journal of the American Statistical Association, pages 1–14, 2020.

[31] I. J. Farkas. Spectra of "real-world" graphs: Beyond the semicircle law. Physical Review, 64(2), 2001.

[32] Daniel Ferguson and Francois G Meyer. The sample fréchet mean (or median) graph of sparse graphs is sparse. arXiv preprint arXiv:2105.14397, 2021.

[33] Daniel Ferguson and Francois G Meyer. Theoretical analysis and computation of the sample frechet mean for sets of large graphs based on spectral information. arXiv preprint arXiv:2201.05923, 2022.

[34] M. Ferrer, F. Serratosa, and A. Sanfeliu. Synthesis of median spectral graph. Springer Berlin Heidelberg, 3523:139–146, 2005.

[35] M. Ferrer, E. Valveny, F. Serratosa, K. Riesen, and H. Bunke. Generalized median graph computation by means of graph embedding in vector spaces. Pattern Recognition, 43(4):1642–1655, 2010.

[36] Miquel Ferrer, Ernest Valveny, and Francesc Serratosa. Median graph: A new exact algorithm using a distance based on the maximum common subgraph. Pattern Recognition Letters, 30(5):579–588, 2009.

[37] A. Flaxman, A. Frieze, and T. Fenner. High degree vertices and eigenvalues in the preferential attachment graph. Approximation, Randomization, and Combinatorial Optimization., pages 264–274, 2003.

[38] Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. Combinatorica, 1(3):233–241, 1981.

[39] Shuang Gao and Peter E Caines. Spectral representations of graphons in very large network systems control. In 2019 IEEE 58th conference on decision and Control (CDC), pages 5068–5075. IEEE, 2019.

[40] Valerio Gemmetto, Alain Barrat, and Ciro Cattuto. Mitigation of infectious disease at school: targeted class closure vs school closure. BMC infectious diseases, 14(1):1–10, 2014.

[41] C. E. Ginestet. Strong consistency of fréchet sample mean sets for graph-valued random variables. arXiv preprint arXiv:1204.3183, 2012.

[42] C.E. Ginestet, J. Li, P. Balachandran, S. Rosenberg, and E.D. Kolaczyk. Les espaces abstraits et leur utilité en statistique théorique et même en statistique appliquée. Journal de la Société Française de Statistique, pages 410–421, 1948.

[43] C.E. Ginestet, J. Li, P. Balachandran, S. Rosenberg, and E.D. Kolaczyk. Hypothesis testing for network data in functional neuroimaging. The Annals of Applied Statistics, pages 725–750, 2017.

[44] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 60:7821–7826, 2002.

[45] Artur Gramacki. Nonparametric kernel density estimation and its computational aspects, volume 37. Springer, 2018.

[46] Lennart Gulikers, Marc Lelarge, and Laurent Massoulié. A spectral method for community detection in moderately sparse degree-corrected stochastic block models. Advances in Applied Probability, 49(3):686–721, 2017.

[47] X. Guo and L. Zhao. A systematic survey on deep generative models for graph generation. arXiv preprint arXiv:2007.06686, 2012.

[48] F. Han, X. Han, H. Liu, and B. et al. Caffo. Sparse median graphs estimation in a high-dimensional semiparametric model. The Annals of Applied Statistics, pages 1397–1426, 2016.

[49] T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. Springer Series in Statistics, 2001.

[50] V Henson, G Sanders, and J Trask. Extremal eigenpairs of adjacency matrices wear their sleeves near their hearts: Maximum principles and decay rates for resolving community structure. Technical report, Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), 2013.

[51] P. Holland, K. Laskey, and S Leinhardt. Stochastic blockmodels: First steps. Social Networks, 5(2):109–137, 1983.

[52] Wolfgang Huber, Vincent J Carey, Li Long, Seth Falcon, and Robert Gentleman. Graphs in molecular biology. BMC bioinformatics, 8(6):1–14, 2007.

[53] D. Hunter and M. Handcock. Inference in curved exponential family models for networks. Journal of Computational and Graphical Statistics, 15(3):556–583, 2006.

[54] Marios Iliofotou, Prashanth Pappu, Michalis Faloutsos, Michael Mitzenmacher, Sumeet Singh, and George Varghese. Network monitoring using traffic dispersion graphs (tdgs). In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pages 315–320, 2007.

[55] B. J. Jain and K. Obermayer. Algorithms for the sample mean of graphs. International Conference on Computer Analysis of Images and Patterns, pages 351–359, 2009.

[56] B.J. Jain. On the geometry of graph spaces. Discrete Applied Mathematics, pages 126–144, 2016.

[57] B.J. Jain and K. Obermayer. Learning in riemannian orbifolds. arXiv preprint arXiv:1204.4294, 2012.

[58] Svante Janson. Graphons, cut norm and distance, couplings and rearrangements. arXiv preprint arXiv:1009.2376, 2010.

[59] Young. J.G., A. Kirkley, and M.E.J. Newman. Clustering of heterogeneous populations of networks. Phys. Rev. E, 105(1), 2022.

[60] X. Jiang, A. Munger, and H. Bunke. On median graphs: properties, algorithms, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 1144–1151, 2001.

[61] Yu Jin, Andreas Loukas, and Joseph JaJa. Graph coarsening with preserved spectral properties. In International Conference on Artificial Intelligence and Statistics, pages 4452–4462. PMLR, 2020.

[62] C.R. Johnson, C. Marijuán, P. Paparella, and M. Pisonero. Operator Theory, Operator Algebras, and Matrix Theory. Springer, 2018.

[63] Irena Jovanović and Zoran Stanić. Spectral distances of graphs. Linear algebra and its applications, 436(5):1425–1435, 2012.

[64] R. Kindermann and J. Snell. Markov random fields and their applications. American Mathematical Society, 1980.

[65] M Kiranmayi and N Maheswari. A review on privacy preservation of social networks using graphs. Journal of Applied Security Research, 16(2):190–223, 2021.

[66] C. Knudsen and J. McDonald. A note on the convexity of the realizable set of eigenvalues for nonnegative symmetric matrices. The Electronic Journal of Linear Algebra, pages 110–114, 2001.

[67] E.D. Kolaczyk, L. Lin, S. Rosenberg, J. Walters, and J. et al. Xu. Averages of unlabeled networks: Geometric characterization and asymptotic behavior. The Annals of Statistics, pages 514–538, 2020.

[68] Can M Le, Elizaveta Levina, and Roman Vershynin. Concentration of random graphs and application to community detection. In Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018, pages 2925–2943. World Scientific, 2018.

[69] J. R. Lee, S. O. Gharan, and L. Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. J. ACM, pages 37:1–37:30, 2014.

[70] Yujia Li, Oriol Vinyals, Chris Dyer, Razvan Pascanu, and Peter Battaglia. Learning deep generative models of graphs. arXiv preprint arXiv:1803.03324, 2018.

[71] Andreas Loukas. Graph reduction with spectral and cut guarantees. J. Mach. Learn. Res., 20(116):1–42, 2019.

[72] X. Lu and B.K. Szymanski. A regularized stochastic block model for the robust community detection in complex networks. Sci Rep, 9, 2019.

[73] Vince Lyzinski, Daniel L Sussman, Minh Tang, Avanti Athreya, and Carey E Priebe. Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. Electronic journal of statistics, 8(2):2905–2922, 2014.

[74] C. Maas. Computing and interpreting the adjacency spectrum of traffic networks. Journal of Computational and Applied Mathematics, pages 459–466, 1985.

[75] F.G. Meyer. The fréchet mean of inhomogeneous random graphs. Complex Networks and Their Applications, pages 1–12, 2021.

[76] C. Morris, N.M. Kriege, F. Bause, K. Kersting, P. Mutzel, and M. Neumann. Iam graph database repository.

[77] C. Morris, N.M. Kriege, F. Bause, K. Kersting, P. Mutzel, and M. Neumann. Iam graph database repository for graph based pattern recognition and machine learning. arXiv preprint arXiv:2007.08663, 2020.

[78] Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. The Annals of Applied Probability, 26(4):2211–2256, 2016.

[79] Mark EJ Newman. Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23):8577–8582, 2006.

[80] MEJ Newman. Spectral community detection in sparse networks. arXiv preprint arXiv:1308.6494, 2013.

[81] Sofia C Olhede and Patrick J Wolfe. Network histograms and universality of blockmodel approximation. Proceedings of the National Academy of Sciences, 111(41):14722–14727, 2014.

[82] Kyoung-Woon On, Eun-Sol Kim, Il-Jae Kwon, Sangwoong Yoon, and Byoung-Tak Zhang. Spectrally similar graph pooling. 2020.

[83] Research Group on Computer Vision and University of Bern Artificial Intelligence INF. Iam graph database repository. 2019.

[84] X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. Journal of Mathematical Imaging and Vision, 25(1), 2006.

[85] Alexander Petersen and Hans-Georg Müller. Fréchet regression for random objects with euclidean predictors. The Annals of Statistics, 47(2):691–719, 2019.

[86] K. Riesen and H. Bunke. Iam graph database repository for graph based pattern recognition and machine learning. Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR), pages 287–297, 2008.

[87] G. Robins, T. Snijders, P. Wang, M. Handcock, and P. Pattison. Recent developments in exponential random graph (p*) models for social networks. Social Networks, 29(2):192–215, 2007.

[88] Lunagómez S., S.C. Olhede, and P.J. Wolfe. Modeling network populations via graph distances. Journal of the American Statistical Association, pages 1–18, 2020.

[89] S. Shalev-Shwartz and S. Ben-David. Understanding Machine Learning: From Theory to Applications. Cambridge University Press, 2014.

[90] Alana Shine and David Kempe. Generative graph models based on laplacian spectra. In The World Wide Web Conference, pages 1691–1701, 2019.

[91] Abhinav Singh and Mark D Humphries. Finding communities in sparse networks. Scientific reports, 5(1):1–7, 2015.

[92] T. Snijders. Markov chain monte carlo estimation of exponential random graph models. Journal of Social Structure, 3, 2002.

[93] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, et al. High-resolution measurements of face-to-face contact patterns in a primary school. PloS one, 6(8):e23176, 2011.

[94] Gilbert W Stewart. Matrix perturbation theory. 1990.

[95] D. Strauss and M. Ikeda. Pseudolikelihood estimation for social networks. Journal of the American Statistical Association, 85(409):204–212, 1990.

[96] Balázs Szegedy. Limits of kernel operators and the spectral regularity lemma. European Journal of Combinatorics, 32(7):1156–1167, 2011.

[97] Minh Tang. The eigenvalues of stochastic blockmodel graphs. arXiv preprint arXiv:1803.11551, 2018.

[98] Terence Tao. Topics in random matrix theory, volume 132. American Mathematical Soc., 2012.

[99] P. Van-Mieghem. Graph spectra for complex networks. Cambridge University Press, 2010.

[100] V. Vu. Recent progress in combinatorial random matrix theory. Probability Surveys, pages 179–200, 2021.

[101] Van H Vu. Recent progress in combinatorial random matrix theory. Probability Surveys, 18:179–200, 2021.

[102] D Watts and S. Strogatz. Collective dynamics of 'small-world' networks. Nature, 6684(393):440–442, 1998.

[103] P. Wills and F.G. Meyer. Metrics for graph comparison: A practitioner's guide. PLOS ONE, 15:1–54, 2020.

[104] R. C. Wilson and P. Zhu. A study of graph spectra for comparing graphs and trees. Pattern Recognition, 41(9):2833–2841, 2008.

[105] Se-Young Yun and Alexandre Proutiere. Accurate community detection in the stochastic block model via spectral algorithms. arXiv preprint arXiv:1412.7335, 2014.

[106] R. Zhand, X. ND Nadakuditi and M. Newman. Spectra of random graphs with community structure and arbitrary degrees. arXiv preprint arXiv:1310.0046, 2013.

[107] Kaiguang Zhao, Michael A Wulder, Tongxi Hu, Ryan Bright, Qiusheng Wu, Haiming Qin, Yang Li, Elizabeth Toman, Bani Mallick, Xuesong Zhang, et al. Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A bayesian ensemble algorithm. Remote sensing of Environment, 232:111–181, 2019.

[108] Yizhe Zhu. A graphon approach to limiting spectral distributions of wigner-type matrices. Random Structures & Algorithms, 56(1):251–279, 2020.

Appendix A

Supplementary material for Chapter 2

We split the appendix into five sections. Appendix A.1 establishes a few classic results that we refer to in our proofs. Appendix A.2 outlines three methods of estimating the expected value of the largest eigenvalues. The proof of our primary contribution, Theorem 1, is contained in Appendix A.3. Within this appendix we also prove Theorem 3 and Theorem 4 since these are necessary results for our proof of Theorem 1. Appendix A.4 and Appendix A.5 are short appendices in which we prove Theorems 2 and 6 respectively.

A.1 Appendix: Classic Results

Theorem 10 (Weyl-Lidskii)

Let $\boldsymbol{H}$ be a self-adjoint operator on a Hilbert space $\mathcal{H}$. Let $\boldsymbol{A}$ be a bounded operator on $\mathcal{H}$ Let $\sigma(\boldsymbol{H})$ and $\sigma(\boldsymbol{H}+\boldsymbol{A})$ denote the spectra of $\boldsymbol{H}$ and $(\boldsymbol{H}+\boldsymbol{A})$ respectively. Then

$$\sigma(\boldsymbol{H}+\boldsymbol{A}) \subset \{\lambda : dist(\lambda, \sigma(\boldsymbol{H})) \leq ||\boldsymbol{A}||\} \tag{A.1}$$

where $||\boldsymbol{A}||$ denotes the operator norm of $\boldsymbol{A}$.

Proof of Theorem 10 These are standard bounds that can be found in many good books on matrix perturbation theory (e.g., [94]).

Let P_n be probabilities on the Borel σ-field of $\mathbb{R}^c$ and suppose $P_n \to P$ weakly.

Theorem 11 (Finite Dimensional Convergence in Distribution)

Let $F_n(\boldsymbol{x}) := P_n((-\infty, x_1] \times ... \times (-\infty, x_c])$ and $F(\boldsymbol{x}) := P((-\infty, x_1] \times ... \times (-\infty, x_c])$ for any $\boldsymbol{x} \in \mathbb{R}^c$. Then $F_n(\boldsymbol{x}) \to F(\boldsymbol{x})$ as $n \to \infty$ for every point of continuity $\boldsymbol{x}$ of $F(\boldsymbol{x})$.

Proof of Theorem 11 This is a standard equivalence for convergence in distribution found in e.g. [13].

A.2 Approximating expected eigenvalues of stochastic block model graphs

We first introduce a recent theorem from [22] that discusses an estimate of the expected eigenvalues of an inhomogeneous Erdős-Rényi random graph. Let $\mu_{\omega_n f} \in \mathcal{M}(\mathcal{G})$ be a kernel probability measure with kernel f. Let L_f be the linear integral operator with the same kernel function, f. Assume L_f has a finite rank of c. Denote the eigenvalues and eigenfunctions of L_f as θ_i and $r_i(x)$ respectively where for each $i = 1, ..., c$, $r_i(x)$ is assumed to be piecewise Lipschitz with finitely many discontinuities and bounded.

Theorem 12 (Chakrabarty, Chakraborty, Hazra 2020)

For every $1 \leq i \leq c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] = \lambda_i(\boldsymbol{B}) + \mathcal{O}(\sqrt{\omega}_n + \frac{1}{n\omega_n}), \tag{A.2}$$

where $\boldsymbol{B}$ is a $c \times c$ symmetric deterministic matrix defined by

$$b_{j,l} = \sqrt{\theta_j \theta_l} n\omega_n \boldsymbol{e}_j^T \boldsymbol{e}_l + \theta_i^{-2} \sqrt{\theta_j \theta_l} (n\omega_n)^{-1} \boldsymbol{e}_j^T \mathbb{E}\left[(\boldsymbol{A} - \mathbb{E}\left[\boldsymbol{A}\right])^2\right] \boldsymbol{e}_l + \mathcal{O}(\frac{1}{n\omega_n}),$$

and $\boldsymbol{e}_j$ is a vector with entries $\boldsymbol{e}_j(k) = \frac{1}{\sqrt{n}} r_j(\frac{k}{n})$ for $1 \leq j \leq c$.

Proof of Theorem 12 The proof is in [22].

The authors in [22] require that the eigenfunctions be Lipschitz but, as is made clear from their proof, this requirement can be relaxed to include piecewise Lipschitz functions with no adjustments to their proof. The eigenfunctions of integral operators with stochastic block model kernels are therefore within the scope of this theorem. We offer minor simplifications to Theorem [22] in

the regime where $\omega_n \to 0$ and work with the eigenvalues and eigenfunctions of the linear integral operator rather than their finite-dimensional counterparts. We begin with the following lemma which outlines a matrix $\boldsymbol{B}^*$ whose eigenvalues are close to $\boldsymbol{B}$.

Lemma 17 Let $\boldsymbol{B}$ be as defined in Theorem 12. Define the matrices $\boldsymbol{B}^{*,(1)}$ and $\boldsymbol{B}^{*,(2)}$ to have components j, l as

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b^{*,(1)}_{j,l} = \sqrt{\theta_j\theta_l}n\omega_n \int_0^1 r_j(x)r_l(x)dx = \begin{cases} \theta_j n\omega_n & j = l \\ 0 & j \neq l \end{cases} \tag{A.3}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b^{*,(2)}_{j,l} = \theta_i^{-2}\sqrt{\theta_j\theta_l} \int_0^1 r_j(x)r_l(x) \int_0^1 f(x,y)dydx. \tag{A.4}$$

Let $\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}$. For any i,

$$|\lambda_i(\boldsymbol{B}) - \lambda_i(\boldsymbol{B}^*)| = \mathcal{O}(\omega_n). \tag{A.5}$$

Proof of Lemma 17 The proof is straightforward in that we omit all contributions to $b_{j,l}$ that are $\mathcal{O}(\omega_n)$ and, because $\omega_n \to 0$, these contributions are negligible when computing the eigenvalues of $\boldsymbol{B}$ as a consequence of Weyl-Lidksii's theorem on spectral inclusion (Theorem 20 from Appendix A.1). We rely heavily on the fact that because the eigenfunctions are Lipschitz with finitely many discontinuities the right endpoint rule approximation of integrals involving the eigenfunctions converge at a rate $\mathcal{O}(\frac{1}{n})$. With these ideas in mind, we proceed with the computations. We begin by considering the entries of $\boldsymbol{B}$ in two parts, define

$$b^{(1)}_{j,l} = \sqrt{\theta_j\theta_l}n\omega_n \boldsymbol{e}_j^T\boldsymbol{e}_l \tag{A.6}$$

$$b^{(2)}_{j,l} = \theta_i^{-2}\sqrt{\theta_j\theta_l}(n\omega_n)^{-1}\boldsymbol{e}_j^T\mathbb{E}\left[(\boldsymbol{A} - \mathbb{E}\left[\boldsymbol{A}\right])^2\right]\boldsymbol{e}_l \tag{A.7}$$

$$b_{j,l} = b^{(1)}_{j,l} + b^{(2)}_{j,l}. \tag{A.8}$$

For comparison, we also write again the components of $\boldsymbol{B}^{*,(1)}$ and $\boldsymbol{B}^{*,(2)}$

$$b^{*,(1)}_{j,l} = \sqrt{\theta_j\theta_l}n\omega_n \int_0^1 r_j(x)r_l(x)dx \tag{A.9}$$

$$b^{*,(2)}_{j,l} = \theta_i^{-2}\sqrt{\theta_j\theta_l} \int_0^1 r_j(x)r_l(x) \int_0^1 f(x,y)dydx. \tag{A.10}$$

We will show that for every j, l that

$$|b_{j,l}^{(1)} - b_{j,l}^{*,(1)}| = \mathcal{O}(\omega_n) \tag{A.11}$$

$$|b_{j,l}^{(2)} - b_{j,l}^{*,(2)}| = \mathcal{O}(\omega_n). \tag{A.12}$$

Observe that for all j, l we have

$$\int_0^1 r_j(x) r_l(x) dx = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \tag{A.13}$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} e_j(m) e_l(m) \tag{A.14}$$

$$= \lim_{n\to\infty} e_j^T e_l. \tag{A.15}$$

The interpretation is that $e_j^T e_l$ is an approximation to the integral using the right endpoints of the intervals. Denote by

$$R(n) = \left| \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) - \int_0^1 r_j(x) r_l(x) dx \right|, \tag{A.16}$$

the error in the right endpoint approximation. The convergence rate for the right endpoint rule is $\mathcal{O}(\frac{1}{n})$ for functions that are piecewise Lipschitz with finitely many discontinuities. Consequently,

$$n\omega_n R(n) = n\omega_n \mathcal{O}(\frac{1}{n}) = \mathcal{O}(\omega_n). \tag{A.17}$$

This indicates that for every j, l,

$$|b_{j,l}^{(1)} - b_{j,l}^{*,(1)}| = \left| \sqrt{\theta_j \theta_l} n\omega_n \boldsymbol{e}_j^T \boldsymbol{e}_l - \sqrt{\theta_j \theta_l} n\omega_n \int_0^1 r_j(x) r_l(x) dx \right| \tag{A.18}$$

$$= \sqrt{\theta_j \theta_l} n\omega_n \left| \boldsymbol{e}_j^T \boldsymbol{e}_l - \int_0^1 r_j(x) r_l(x) dx \right| \tag{A.19}$$

$$= \sqrt{\theta_j \theta_l} n\omega_n R(n) \tag{A.20}$$

$$= \mathcal{O}(\rho_n). \tag{A.21}$$

We now turn our attention to the second component of $\boldsymbol{B}$ and analyze $b_{j,l}^{(2)}$,

$$b_{j,l}^{(2)} = \theta_i^{-2} \sqrt{\theta_j \theta_l} (n\omega_n)^{-1} \boldsymbol{e}_j^T \mathbb{E}\left[(\boldsymbol{A} - \mathbb{E}\left[\boldsymbol{A}\right])^2 \right] \boldsymbol{e}_l. \tag{A.22}$$

To understand $b_{j,l}^{(2)}$ it is necessary to understand $\mathbb{E}\left[(\boldsymbol{A}-\mathbb{E}\left[\boldsymbol{A}\right])^2\right]$. Let $a_{m,k}=\boldsymbol{A}_{m,k}$, the $(m,k)^{th}$ entry of $\boldsymbol{A}$,

$$\mathbb{E}\left[\left((\boldsymbol{A}-\mathbb{E}\left[\boldsymbol{A}\right])^2\right)_{m,k}\right]=\mathbb{E}\left[\sum_{w=1}^{n}a_{m,w}a_{w,k}-a_{m,w}\mathbb{E}\left[a_{w,k}\right]-\mathbb{E}\left[a_{m,w}\right]a_{w,k}+\mathbb{E}\left[a_{m,w}\right]\mathbb{E}\left[a_{w,k}\right])\right] \tag{A.23}$$

$$=\sum_{w=1}^{n}\mathbb{E}\left[a_{m,w}a_{w,k}\right]-\mathbb{E}\left[a_{m,w}\right]\mathbb{E}\left[a_{w,k}\right]-\mathbb{E}\left[a_{m,w}\right]\mathbb{E}\left[a_{w,k}\right]+\mathbb{E}\left[a_{m,w}\right]\mathbb{E}\left[a_{w,k}\right]) \tag{A.24}$$

$$=\begin{cases}\sum_{w=1}^{n}\mathbb{E}\left[a_{m,w}\right]\left(1-\mathbb{E}\left[a_{m,w}\right]\right) & m=k\\ 0 \quad \text{else.}\end{cases} \tag{A.25}$$

Recall that $\mathbb{E}\left[a_{m,w}\right]=\omega_n f(\frac{m}{n},\frac{w}{n})$. We may compute the term $(n\omega_n)^{-1}\boldsymbol{e}_j^T\mathbb{E}\left[(\boldsymbol{A}-\mathbb{E}\left[\boldsymbol{A}\right])^2\right]\boldsymbol{e}_l$ which appears in the definition of $b_{j,l}^{(2)}$ with this expression for $\mathbb{E}\left[(\boldsymbol{A}-\mathbb{E}\left[\boldsymbol{A}\right])^2\right]$ as

$$(n\omega_n)^{-1}\boldsymbol{e}_j^T\mathbb{E}\left[(\boldsymbol{A}-\mathbb{E}\left[\boldsymbol{A}\right])^2\right]\boldsymbol{e}_l=\sum_{i=1}^{n}\boldsymbol{e}_j(i)\boldsymbol{e}_l(i)\sum_{w=1}^{n}\mathbb{E}\left[a_{i,w}\right]\left(1-\mathbb{E}\left[a_{i,w}\right]\right) \tag{A.26}$$

$$=\sum_{i=1}^{n}\frac{1}{n^2\omega_n}r_j(\frac{i}{n})r_l(\frac{i}{n})\sum_{w=1}^{n}\omega_n f(\frac{i}{n},\frac{w}{n})(1-\omega_n f(\frac{i}{n},\frac{w}{n})) \tag{A.27}$$

$$=\sum_{i=1}^{n}\frac{1}{n}r_j(\frac{i}{n})r_l(\frac{i}{n})\sum_{w=1}^{n}\frac{1}{n}f(\frac{i}{n},\frac{w}{n})(1-\omega_n f(\frac{i}{n},\frac{w}{n})) \tag{A.28}$$

We consider the final term in two separate parts,

$$\sum_{i=1}^{n}\frac{1}{n}r_j(\frac{i}{n})r_l(\frac{i}{n})\sum_{w=1}^{n}\frac{1}{n}f(\frac{i}{n},\frac{w}{n})(1-\omega_n f(\frac{i}{n},\frac{w}{n}))=\sum_{i=1}^{n}\frac{1}{n}r_j(\frac{i}{n})r_l(\frac{i}{n})\sum_{w=1}^{n}\frac{1}{n}f(\frac{i}{n},\frac{w}{n}) \tag{A.29}$$

$$-\omega_n\sum_{i=1}^{n}\frac{1}{n}r_j(\frac{i}{n})r_l(\frac{i}{n})\sum_{w=1}^{n}\frac{1}{n}f(\frac{i}{n},\frac{w}{n})^2 \tag{A.30}$$

We will show that

$$\omega_n\sum_{i=1}^{n}\frac{1}{n}r_j(\frac{i}{n})r_l(\frac{i}{n})\sum_{w=1}^{n}\frac{1}{n}f(\frac{i}{n},\frac{w}{n})^2=\mathcal{O}(\omega_n) \tag{A.31}$$

We then show that

$$\left|\sum_{i=1}^{n}\frac{1}{n}r_j(\frac{i}{n})r_l(\frac{i}{n})\sum_{w=1}^{n}\frac{1}{n}f(\frac{i}{n},\frac{w}{n})-\int_0^1 r_j(x)r_l(x)\int_0^1 f(x,y)dydx\right|=\mathcal{O}(\frac{1}{n}). \tag{A.32}$$

The conclusion is then that

$$|b_{j,l}^{(2)} - b_{j,l}^{*,(2)}| = \mathcal{O}(\omega_n). \tag{A.33}$$

To begin, recall that the expression for $f(\frac{i}{n}, \frac{w}{n})$ in terms of the eigenvalues and eigenfunctions is

$$f(x,y) = \sum_{k=1}^{c} \theta_k r_k(x) r_k(y) \tag{A.34}$$

$$f(x,y)^2 = \left(\sum_{k=1}^{c} \theta_k r_k(x) r_k(y) \right)^2 \tag{A.35}$$

$$= \sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k r_k(x) r_k(y) \theta_m r_m(x) r_m(y). \tag{A.36}$$

We have the following expression which will be used throughout our computations,

$$\sum_{w=1}^{n} \frac{1}{n} f(\frac{i}{n}, \frac{w}{n})^2 = \sum_{w=1}^{n} \frac{1}{n} \sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k r_k(x) r_k(y) \theta_m r_m(x) r_m(y) \tag{A.37}$$

$$= \sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k \theta_m \sum_{w=1}^{n} \frac{1}{n} r_k(x) r_k(y) r_m(x) r_m(y). \tag{A.38}$$

Consider now just (A.30)

$$\omega_n \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} f(\frac{i}{n}, \frac{w}{n})^2 \tag{A.39}$$

$$= \omega_n \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k \theta_m \sum_{w=1}^{n} \frac{1}{n} r_k(\frac{i}{n}) r_k(\frac{w}{n}) r_m(\frac{i}{n}) r_m(\frac{w}{n}) \tag{A.40}$$

$$= \sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k \theta_m \omega_n \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} r_k(\frac{i}{n}) r_k(\frac{w}{n}) r_m(\frac{i}{n}) r_m(\frac{w}{n}) \tag{A.41}$$

We now consider each k, m independently,

$$\theta_k \theta_m \omega_n \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} r_k(\frac{i}{n}) r_k(\frac{w}{n}) r_m(\frac{i}{n}) r_m(\frac{w}{n}) \tag{A.42}$$

$$= \theta_k \theta_m \omega_n \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) r_k(\frac{i}{n}) r_m(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} r_k(\frac{w}{n}) r_m(\frac{w}{n}) \tag{A.43}$$

It is here that we observe that each k, m equation (A.43) can be interpreted as the product of two right endpoint rule approximations to integrals. We have that

$$\int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy = \lim_{n \to \infty} \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) r_k(\frac{i}{n}) r_m(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} r_k(\frac{w}{n}) r_m(\frac{w}{n}) \tag{A.44}$$

We can therefore define the error in the approximation as

$$R(n) = \left| \int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy - \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) r_k(\frac{i}{n}) r_m(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} r_k(\frac{w}{n}) r_m(\frac{w}{n}) \right| \tag{A.45}$$

where, because each function is piecewise Lipschitz, we still have that $R(n) = \mathcal{O}(\frac{1}{n})$. Therefore, for each k, m,

$$\begin{aligned} R(n) &= \theta_k \theta_m \Bigg| \int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy && \text{(A.46)} \\ &\quad - \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) r_k(\frac{i}{n}) r_m(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} r_k(\frac{w}{n}) r_m(\frac{w}{n}) \Bigg| && \text{(A.47)} \\ &= \mathcal{O}(\frac{1}{n}) && \text{(A.48)} \end{aligned}$$

Since taking a finite sum does not impact the order of the error we can conclude that

$$\begin{aligned} &\Bigg| \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} f(\frac{i}{n}, \frac{w}{n})^2 && \text{(A.49)} \\ &\quad - \sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k \theta_m \int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy \Bigg| && \text{(A.50)} \\ &= \mathcal{O}(\frac{1}{n}) && \text{(A.51)} \end{aligned}$$

Observe now the impact of ω_n from equation (A.30),

$$\begin{aligned} &\omega_n \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} f(\frac{i}{n}, \frac{w}{n})^2 && \text{(A.52)} \\ &= \omega_n \left(\sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k \theta_m \omega_n \int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy + \mathcal{O}(\frac{1}{n}) \right). && \text{(A.53)} \end{aligned}$$

Since each $r_j(x)$ is piecewise Lipschitz, we can conclude that the entire term in equation (A.30) is negligible because,

$$\begin{aligned} &\omega_n \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} f(\frac{i}{n}, \frac{w}{n})^2 && \text{(A.54)} \\ &= \omega_n \left(\sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k \theta_m \omega_n \int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy + \mathcal{O}(\frac{1}{n}) \right) && \text{(A.55)} \\ &= \mathcal{O}(\omega_n) + \mathcal{O}(\omega_n \frac{1}{n}). && \text{(A.56)} \end{aligned}$$

Therefore,

$$\sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} f(\frac{i}{n}, \frac{w}{n})(1 - \omega_n f(\frac{i}{n}, \frac{w}{n})) = \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} f(\frac{i}{n}, \frac{w}{n}) + \mathcal{O}(\omega_n) \tag{A.57}$$

By way of the exact same analysis, we can conclude that the right-hand side of (A.29) is

$$\sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} f(\frac{i}{n}, \frac{w}{n}) = \int_0^1 r_j(x) r_l(x) \int_0^1 f(x, y) dy dx + \mathcal{O}(\frac{1}{n}). \tag{A.58}$$

Therefore,

$$\left| \sum_{i=1}^{n} \frac{1}{n} r_j(\frac{i}{n}) r_l(\frac{i}{n}) \sum_{w=1}^{n} \frac{1}{n} f(\frac{i}{n}, \frac{w}{n})(1 - \omega_n f(\frac{i}{n}, \frac{w}{n})) - \int_0^1 r_j(x) r_l(x) \int_0^1 f(x, y) dy dx \right| = \mathcal{O}(\omega_n). \tag{A.59}$$

The conclusion of this analysis shows that

$$\left| b_{j,l}^{(2)} - b_{j,l}^{*,(2)} \right| = \theta_i^{-2} \sqrt{\theta_j \theta_l} \left| (n\omega_n)^{-1} \boldsymbol{e}_j^T \mathbb{E}\left[(\boldsymbol{A} - \mathbb{E}\left[\boldsymbol{A}\right])^2 \right] \boldsymbol{e}_l - \int_0^1 r_j(x) r_l(x) \int_0^1 f(x, y) dy dx \right| \tag{A.60}$$

$$= \mathcal{O}(\omega_n) \tag{A.61}$$

The result puts together equations (A.21) and (A.61) showing that for every j, l

$$|b_{j,l} - \left(b_{j,l}^{*,(1)} + b_{j,l}^{*,(2)} \right)| = |b_{j,l}^{(1)} + b_{j,l}^{(2)} - \left(b_{j,l}^{*,(1)} + b_{j,l}^{*,(2)} \right)| = \mathcal{O}(\omega_n). \tag{A.62}$$

Recall the definition of $\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}$. Define $\boldsymbol{B}^\epsilon$ as

$$\boldsymbol{B}^\epsilon = \boldsymbol{B} - \boldsymbol{B}^*. \tag{A.63}$$

We have just shown that every component of $\boldsymbol{B}^\epsilon$ is $\mathcal{O}(\omega_n)$. We conclude the proof by appealing to Weyl-Lidskii's theorem on the spectral inclusion of the eigenvalues. For this theorem we take

$$\boldsymbol{A} = \boldsymbol{B}^\epsilon \tag{A.64}$$

$$\boldsymbol{H} = \boldsymbol{B}^* \tag{A.65}$$

$$\boldsymbol{H} + \boldsymbol{A} = \boldsymbol{B}^* + \boldsymbol{B}^\epsilon = \boldsymbol{B}^* + \boldsymbol{B} - \boldsymbol{B}^* = \boldsymbol{B} \tag{A.66}$$

where

$$||\boldsymbol{B}^*|| \leq ||\boldsymbol{B}^*||_F = \mathcal{O}(\omega_n), \tag{A.67}$$

and conclude that

$$|\lambda_i(\boldsymbol{B}) - \lambda_i(\boldsymbol{B}^*)| = \mathcal{O}(\omega_n). \tag{A.68}$$

□

The consequence of the above lemma simply shows a method to determine the i^{th} expected eigenvalue of $\boldsymbol{A}_{\mu_{\omega_n f}}$ in terms of the eigenvalues and eigenfunctions of L_f rather than their discretized counterparts. This constitutes the next theorem.

Theorem 13 (Theorem 4 in the main chapter) For every $1 \leq i \leq c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] = \lambda_i(\boldsymbol{B}^*) + \mathcal{O}(\sqrt{\omega_n}), \tag{A.69}$$

where

$$\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}. \tag{A.70}$$

and

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b^{*,(1)}_{j,l} = \sqrt{\theta_j \theta_l} n\omega_n \int_0^1 r_j(x) r_l(x) dx = \begin{cases} \theta_j n\omega_n & j = l \\ 0 & j \neq l \end{cases} \tag{A.71}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b^{*,(2)}_{j,l} = \theta_i^{-2}\sqrt{\theta_j \theta_l} \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx. \tag{A.72}$$

Proof of Theorem 13 The proof is a consequence of Lemma 17 and Theorem 12. Let $\boldsymbol{B}$ be as defined by Theorem 12 and $\boldsymbol{B}^*$ be defined as in Lemma 17. Consider

$$|\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] - \lambda_i(\boldsymbol{B}) + \lambda_i(\boldsymbol{B}) - \lambda_i(\boldsymbol{B}^*)| \leq |\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] - \lambda_i(\boldsymbol{B})| + |\lambda_i(\boldsymbol{B}) - \lambda_i(\boldsymbol{B}^*)| \tag{A.73}$$

$$= \mathcal{O}(\sqrt{\omega_n} + \frac{1}{n\omega_n}) + \mathcal{O}(\omega_n) \tag{A.74}$$

$$= \mathcal{O}(\sqrt{\omega_n}). \tag{A.75}$$

The specific structure of $\boldsymbol{B}^*$ is of note as it is comprised of a diagonal component and a secondary component. In practice, either method can be used depending on the situation. If the eigenfunctions are known then the result from Lemma 17 avoids potentially costly matrix computations for large n. In the following corollary we show the contribution to the i^{th} eigenvalue of $\boldsymbol{B}^*$ in terms of the matrices $\boldsymbol{B}^{*,(1)}$ and $\boldsymbol{B}^{*,(2)}$.

Lemma 18 Let $\boldsymbol{B}^*$ be as defined in Lemma 17 such that

$$\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}. \tag{A.76}$$

Then

$$|\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] - \lambda_i(\boldsymbol{B}^{*,(1)})| = \mathcal{O}(1). \tag{A.77}$$

where

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b_{j,l}^{*,(1)} = \sqrt{\theta_j\theta_l}n\omega_n \int_0^1 r_j(x)r_l(x)dx = \begin{cases} \theta_j n\omega_n & j = l \\ 0 & j \neq l \end{cases} \tag{A.78}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b_{j,l}^{*,(2)} = \theta_i^{-2}\sqrt{\theta_j\theta_l} \int_0^1 r_j(x)r_l(x) \int_0^1 f(x,y)dydx. \tag{A.79}$$

Proof of Lemma 18 The proof is an application of Weyl-Lidksii after we have shown that $||\boldsymbol{B}^{*,(2)}|| = \mathcal{O}(1)$. It is clear by definition that each component of $b_{j,l}^{*,(2)}$ is independent of n, hence

$$b_{j,l}^{*,(2)} = \mathcal{O}(1). \tag{A.80}$$

Because $\boldsymbol{B}^{*,(2)} \in \mathbb{R}^{c\times c}$ where c is fixed and independent of n,

$$||\boldsymbol{B}^{*,(2)}|| \leq ||\boldsymbol{B}^{*,(2)}||_F = \sum_{j=1}^{c}\sum_{l=1}^{c} b_{j,l}^{*,(2)} = \mathcal{O}(1). \tag{A.81}$$

By Weyl-Lidskii, taking

$$\boldsymbol{A} = \boldsymbol{B}^{*,(2)} \tag{A.82}$$

$$\boldsymbol{H} = \boldsymbol{B}^{*,(1)} \tag{A.83}$$

$$\boldsymbol{H} + \boldsymbol{A} = \boldsymbol{B}^* \tag{A.84}$$

we can conclude that

$$|\lambda_i(\boldsymbol{B}^*) - \lambda_i(\boldsymbol{B}^{*,(1)})| = \mathcal{O}(1). \tag{A.85}$$

Since $\lambda_i(\boldsymbol{B}^*)$ is the first order estimate of $\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right]$ we have the final conclusion via the triangle inequality that

$$|\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] - \lambda_i(\boldsymbol{B}^*) + \lambda_i(\boldsymbol{B}^*) - \lambda_i(\boldsymbol{B}^{*,(1)})| \leq |\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] - \lambda_i(\boldsymbol{B}^*)| + |\lambda_i(\boldsymbol{B}^*) - \lambda_i(\boldsymbol{B}^{*,(1)})| \tag{A.86}$$

$$= \mathcal{O}(1) + \mathcal{O}(\sqrt{\omega_n}) = \mathcal{O}(1). \tag{A.87}$$

Theorem 4 is a direct application of the prior lemma.

Theorem 14 (Estimation of the Largest Eigenvalues of Stochastic Block Models)

For $i = 1, ..., c$

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] = \theta_i n\omega_n + \mathcal{O}(1) \tag{A.88}$$

Proof of Theorem 14 The proof is a direct application of the prior Lemma once noting that $\lambda_i(\boldsymbol{B}^{*,(1)}) = n\omega_n\theta_i$ due to the diagonal structure of $\boldsymbol{B}^{*,(1)}$.

While the theorem is rather straightforward it makes the following observation, that $\boldsymbol{B}^{*,(2)}$ is independent of n. Therefore, as n increases, the relative impact of $\boldsymbol{B}^{*,(2)}$ may be of little consequence since

$$\frac{|\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] - n\omega_n\theta_i|}{n\omega_n\theta_i} = \mathcal{O}\left(\frac{1}{n\omega_n}\right). \tag{A.89}$$

A second noteworthy observation is that the $\mathcal{O}(1)$ estimate is independent of the eigenfunctions of L_f and can be computed without determining the eigenfunctions or finite-dimensional approximations of the eigenfunctions as is done in Theorem 12 and Theorem 13. This saves significant computational time, particularly for extremely large graphs.

A.3 Appendix: Proof of Theorem 1

Theorem 1 constitutes the main theoretical contribution of this chapter. Throughout this appendix, we assume that the graph G satisfies assumptions 1-4. The proof involves a few steps which we outline below at a high level.

(1) Given G with adjacency matrix $\boldsymbol{A}$ we compute $\sigma_c(\boldsymbol{A})$ and show that we may construct a canonical stochastic block model kernel, f, where $\boldsymbol{Q} = 0$ with c blocks such that $\lambda_i(L_f) = \frac{\lambda_i(\boldsymbol{A})-1}{n\omega_n}$ where we define $\omega_n = \rho_n$, the density of the observed graph. Notably, the f that we construct will not necessarily satisfy $||f||_1 = 1$ indicating that ρ_n need not denote the expected density of graphs sampled according to $\mu_{\rho_n f}$.

(2) We then show that for each $1 \leq i \leq c$, $\left|\lambda_i(\boldsymbol{B}^*) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}})\right]\right| = \mathcal{O}(\sqrt{\rho_n})$ where $\boldsymbol{B}^*$ is given by Lemma 17. We show that $\lambda_i(\boldsymbol{B}^*) = n\rho_n\lambda_i(L_f) + 1$ due to the block-diagonal structure of $f(x, y)$.

(3) All that is left is to show that for an iid sample of graphs $\{G^{(k)}\}_{k=1}^N$ distributed according to $\mu_{\rho_n f}$, the empirical Fréchet mean, G_N^*, with adjacency matrix $\boldsymbol{A}_N^*$, satisfies

$$||\sigma_c(\boldsymbol{A}_N^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon \quad a.s.$$

The key in this step is to show the existence of a different graph G' with adjacency matrix $\boldsymbol{A}'$ whose eigenvalues are close to $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]$. This graph allows us to bound the distance between the eigenvalues of $\boldsymbol{A}_N^*$ and $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]$.

(4) The final step is to show

$$|\lambda_i(\boldsymbol{A}) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}})\right] + \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}})\right] - \lambda_i(\boldsymbol{A}_N^*)| < \epsilon$$

for each $1 \leq i \leq c$.

We now proceed with our proof.

Step 1: Constructing a stochastic block model kernel

Let $G \in \mathcal{G}$ with adjacency matrix $\boldsymbol{A}$ be such that

$$0 \preccurlyeq \sigma_c(\boldsymbol{A}) \tag{A.90}$$

and for every $1 \leq i \neq j \leq c$, $\lambda_i \neq \lambda_j$.

Lemma 19 Let $\vec{\theta} \in \mathbb{R}^c$ such that $0 \preccurlyeq \vec{\theta}$. Let $\boldsymbol{s}$ be a fixed geometry vector for a canonical stochastic block model kernel with c non-zero entries. There exists a canonical stochastic block model kernel $f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})$ with $\boldsymbol{Q} = 0$ defining the integral operator

$$L_f(t) = \int_0^1 f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s}) t(y) dy \tag{A.91}$$

which satisfies

$$\lambda_i(L_f) = \theta_i \tag{A.92}$$

Proof of Lemma 19 The proof is rather straightforward, we simply construct the equivalent of a block diagonal matrix. The blocks are determined by the geometry vector $\boldsymbol{s}$ which we are free to choose within the constraints that $||\boldsymbol{s}||_1 = 1$, $\boldsymbol{s}$ is non-increasing, non-negative, and has c non-zero entries. For $1 \leq i \leq c$, let $S_i = \sum_{j=1}^{i} s_j$ and define the intervals $\mathcal{I}_i = [S_{i-1}, S_i)$. Note that $S_0 = 0$. Define the function

$$f(x, y) = \begin{cases} \frac{\theta_i}{s_i} & \text{if } (x, y) \in I_i \times I_i \\ 0 & \text{else.} \end{cases} \tag{A.93}$$

Define the linear integral operator $L_f(t) = \int_0^1 f(x, y) t(y) dy$. The eigenfunctions for L_f are

$$r_i(x) = \begin{cases} \frac{1}{\sqrt{s_i}} & \text{if } x \in \mathcal{I}_i \\ 0 & \text{else.} \end{cases} \tag{A.94}$$

which we show in the following computations. We can compute the eigenvalues of L_f as

$$L_f(r_i(x)) = \int_0^1 f(x,y) r_i(y) dy \tag{A.95}$$

$$= \begin{cases} \int_{\mathcal{I}_i} \frac{\theta_i}{s_i} \frac{1}{\sqrt{s_i}} dy & x \in \mathcal{I}_i \\ 0 \quad \text{else} \end{cases} \tag{A.96}$$

$$= \begin{cases} \frac{\theta_i}{s_i} \frac{1}{\sqrt{s_i}} s_i & x \in \mathcal{I}_i \\ 0 \quad \text{else} \end{cases} \tag{A.97}$$

$$= \begin{cases} \theta_i \frac{1}{\sqrt{s_i}} & x \in \mathcal{I}_i \\ 0 \quad \text{else} \end{cases} \tag{A.98}$$

$$= \theta_i r_i(x). \tag{A.99}$$

Next we verify that $||r_i||_2 = 1$.

$$\int_0^1 r_i(x)^2 dx = \int_{\mathcal{I}_i} \frac{1}{s_i} dx \tag{A.100}$$

$$= \frac{s_i}{s_i} \tag{A.101}$$

$$= 1. \tag{A.102}$$

Therefore $r_i(x)$ is an eigenfunction of L_f with eigenvalue θ_i. At this point we note that f is a stochastic block model kernel with $q_{ij} = 0$ for all i, j.

By taking $\boldsymbol{\theta} = \frac{\sigma_c(\boldsymbol{A}) - \mathbf{1}}{n\rho_n}$ in Lemma 19 and setting $\omega_n = \rho_n$ we will have accomplished step 1.

Step 2: Estimating the expected eigenvalues

The estimation of the largest eigenvalues has been discussed at length in Appendix A.2. We will use the estimate of the largest eigenvalues outlined by Theorem 13. Recall that

$$|\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}})\right] - \lambda_i(\boldsymbol{B}^*)| = \mathcal{O}(\sqrt{\rho_n}). \tag{A.103}$$

where

$$(\boldsymbol{B}^*)_{j,l} = b^{*,(1)}_{j,l} + b^{*,(2)}_{j,l} = \sqrt{\theta_j\theta_l}n\rho_n \int_0^1 r_j(x)r_l(x)dx + \theta_i^{-2}\sqrt{\theta_j\theta_l}\int_0^1 r_j(x)r_l(x)\int_0^1 f(x,y)dydx \tag{A.104}$$

when $\omega_n = \rho_n$. For the choice of $f(x,y)$ made in step 1, when $j \neq l$,

$$\forall x, \quad r_j(x)r_l(x)dx = 0. \tag{A.105}$$

Thus $b^{*,(2)}_{j,l} = 0$ for all $j \neq l$ and so $\boldsymbol{B}^*$ is diagonal. The i^{th} eigenvalue is then given as

$$\lambda_i(\boldsymbol{B}^*) = \theta_i n\rho_n + \frac{1}{\theta_i}\int_0^1 r_i(x)r_i(x)\int_0^1 f(x,y)dydx \tag{A.106}$$

Using the definitions of $r_i(x)$ and $f(x,y)$ from equations (A.93) and (A.94),

$$\frac{1}{\theta_i}\int_0^1 r_i(x)r_i(x)\int_0^1 f(x,y)dydx = \frac{1}{\theta_i}\frac{1}{s_i}\int_{I_i}\int_{I_i} f(x,y)dydx \tag{A.107}$$

$$= \frac{1}{\theta_i}\frac{1}{s_i}\frac{\theta_i}{s_i}\int_{I_i}\int_{I_i} dydx \tag{A.108}$$

$$= \frac{1}{s_i^2}s_i^2 = 1 \tag{A.109}$$

since $\int_{I_i} dy = s_i$. This shows that

$$\lambda_i(\boldsymbol{B}^*) = \theta_i n\rho_n + 1 = \lambda_i(\boldsymbol{A}). \tag{A.110}$$

We could also consider the perspective that our adjacency matrix is a series of disconnected Erdős-Rényi random graphs on $\frac{n}{c}$ vertices and arrive at the same conclusion by analyzing each Erdős-Rényi component separately.

Step 3: Eigenvalues of the empirical Fréchet mean adjacency matrix are nearly the expected eigenvalues

Step 3 of our proof is arguably the most interesting. As was stated in the outline, given a stochastic block model kernel probability measure, $\mu_{\rho_n f}$, we show the existence of a graph, G',

whose eigenvalues are close to the $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]$. By showing the existence of this graph, we will be able to provide upper bounds on the distance between the eigenvalues of the adjacency matrix of the sample Fréchet mean graph, $\sigma_c(\boldsymbol{A}_N^*)$ and $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]$.

To prove there exists a graph, G', whose adjacency matrix has eigenvalues close to $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]$, we will show that there exists a positive constant C such that

$$0 < C < P\left(||\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon\right). \tag{A.111}$$

Since the probability measure of the set of graphs that satisfy $||\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon$ is strictly positive, this implies the existence of at least one graph, G', that satisfies the inequality. To prove that the probability is nonzero we reference Theorem 2.3 from [22] on the convergence in distribution of the extreme eigenvalues of inhomogeneous Erdős-Rényi random graphs which we state below.

Theorem 15 (Chakrabarty, Chakraborty, and Hazra 2020)

For every $1 \leq i \leq c$,

$$\rho_n^{-1/2}(\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \xrightarrow{d} (Z_i : 1 \leq i \leq c), \tag{A.112}$$

where the right-hand side is a multivariate normal random vector in $\mathbb{R}^c$, with mean zero and

$$Cov(Z_i, Z_j) = 2\int_0^1 \int_0^1 r_i(x) r_i(y) r_j(x) r_j(y) f(x,y) dx dy, \tag{A.113}$$

for all $1 \leq i, j \leq c$.

Proof of Theorem 15 Theorem 15 is Theorem 2.3 in [22].

We first acknowledge that there exists a very similar theorem to the above in [4] with the difference being the centering of the limiting distribution about the eigenvalues of the expected adjacency matrix, $\lambda_i(\mathbb{E}\left[\boldsymbol{A}_{\mu_{\rho_n f}}\right])$, rather than the expected eigenvalues, $\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}})\right]$.

We consider all the eigenvalues at once by writing $\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right])$ rather than analyzing for each i. Let Z denote the multivariate normal random vector on the right-hand side

in equation (A.112). An equivalent statement to Theorem 15 is then

$$\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \xrightarrow{d} Z. \tag{A.114}$$

One characterization of convergence in distribution for finite-dimensional random variables is pointwise convergence of the cumulative distribution functions (see Theorem 11).

Let P_n denote the probabilities for the sequence of random vectors $\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right])$ and P be the probability for the multivariate Guassian random vector Z. $\forall \boldsymbol{z} \in \mathbb{R}^c$, define the cumulative distribution function of the random variables $\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right])$ and Z respectively as

$$F_n(\boldsymbol{z}) = P_n\left((-\infty, z_1] \times ... \times (-\infty, z_c]\right), \tag{A.115}$$

$$F(\boldsymbol{z}) = P\left((-\infty, z_1] \times ... \times (-\infty, z_c]\right). \tag{A.116}$$

The convergence in distribution of the random vectors is equivalent to the following: $\forall \boldsymbol{z} \in \mathbb{R}^c$,

$$\lim_{n\to\infty} F_n\left(\boldsymbol{z}\right) = F\left(\boldsymbol{z}\right), \tag{A.117}$$

since $F(\boldsymbol{z})$ is continuous everywhere. For our proof, it is easier to work with the probabilities and random vectors directly and so equation (A.117) takes the form

$$\lim_{n\to\infty} P_n(\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq \boldsymbol{z}) = P(Z \preccurlyeq \boldsymbol{z}). \tag{A.118}$$

We are now ready to state and prove our lemma. Let $\mu_{\rho_n f} \in \mathcal{M}(\mathcal{G})$ be a kernel probability measure with kernel f. Let L_f be the linear integral operator with the same kernel f. Assume L_f has a finite rank c and denote the eigenvalues and eigenfunctions of L_f as θ_i and $r_i(x)$ respectively where for each $i = 1, ..., c$, $r_i(x)$ is assumed to be piecewise Lipschitz with finitely many discontinuities.

Lemma 20 $\forall \epsilon > 0$, $\exists n^* \in \mathbb{N}$ where $\forall n > n^*$, $\exists G' \in \mathcal{G}$ with adjacency matrix $\boldsymbol{A}'$ such that

$$||\sigma_c(\boldsymbol{A}') - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon. \tag{A.119}$$

Proof of Lemma 20 Let $\epsilon > 0$. Fix $\boldsymbol{z} \in \mathbb{R}^c$ such that $0 \preccurlyeq \boldsymbol{z}$ and

$$0 < C < P(-\boldsymbol{z} \preccurlyeq Z \preccurlyeq \boldsymbol{z}) - 2\epsilon, \tag{A.120}$$

for some $C > 0$. By equation (A.118), there exists $n_1 \in \mathbb{N}$ where for all $n > n_1$,

$$\left| P_n \left(\rho_n^{-1/2} (\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq \boldsymbol{z} \right) - P\left(Z \preccurlyeq \boldsymbol{z}\right) \right| < \epsilon \tag{A.121}$$

Similarly, there exists $n_2 \in \mathbb{N}$ where for all $n > n_2$,

$$\left| P_n \left(\rho_n^{-1/2} (\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq -\boldsymbol{z} \right) - P\left(Z \preccurlyeq -\boldsymbol{z}\right) \right| < \epsilon \tag{A.122}$$

Furthermore, there exists $n_3 \in \mathbb{N}$ where for all $n > n_3$,

$$||\sqrt{\rho_n}\boldsymbol{z}||_2 < \epsilon. \tag{A.123}$$

Take $n^* = \max(n_1, n_2, n_3)$ and consider the following probability

$$P_n(-\boldsymbol{z} \preccurlyeq \rho_n^{-1/2} (\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq \boldsymbol{z}) = P_n(\rho_n^{-1/2} (\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq \boldsymbol{z}) \tag{A.124}$$

$$- P_n(\rho_n^{-1/2} (\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq -\boldsymbol{z}). \tag{A.125}$$

Now, (A.121) and (A.122) allow us to replace (A.124) and (A.125) with the corresponding expression in terms of $\boldsymbol{z}$. We therefore obtain

$$\left| P_n \left(-\boldsymbol{z} \preccurlyeq \rho_n^{-1/2} (\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq \boldsymbol{z} \right) - P\left(-\boldsymbol{z} \preccurlyeq Z \preccurlyeq \boldsymbol{z}\right) \right| < 2\epsilon, \tag{A.126}$$

from which we obtain

$$P(-\boldsymbol{z} \preccurlyeq Z \preccurlyeq \boldsymbol{z}) - 2\epsilon < P_n(-\rho_n^{1/2}\boldsymbol{z} \preccurlyeq (\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq \rho_n^{1/2}\boldsymbol{z}). \tag{A.127}$$

Since $0 < C < P(-\boldsymbol{z} \preccurlyeq Z \preccurlyeq \boldsymbol{z}) - 2\epsilon$ we have

$$C < P_n(-\rho_n^{1/2}\boldsymbol{z} \preccurlyeq (\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq \rho_n^{1/2}\boldsymbol{z}) \tag{A.128}$$

$$< P_n(0 \leq ||\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 \leq ||\sqrt{\rho_n}\boldsymbol{z}||_2) \tag{A.129}$$

and since n is sufficiently large that $||\rho_n^{1/2}\boldsymbol{z}||_2 < \epsilon$ this implies that the probability measure of the set of graphs that satisfy $||\sigma_c(\boldsymbol{A}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A})\right]||_2 < \epsilon$ is strictly positive. Thus there exists a graph $G' \in \mathcal{G}$ with adjacency matrix $\boldsymbol{A}'$ that satisfies $||\sigma_c(\boldsymbol{A}') - \mathbb{E}\left[\sigma_c(\boldsymbol{A})\right]||_2 < \epsilon$. $\square$

It should also be noted that the constant C can be made arbitrarily close to $1-2\epsilon$ and so the probability of observing a graph such that $||\sigma_c(\boldsymbol{A}') - \mathbb{E}\left[\sigma_c(\boldsymbol{A})\right]||_2 < \epsilon$ can be made arbitrarily large so long as n is also taken to be sufficiently large. This is the theoretical support for Remark 9 though in practice we are not free to choose n, the size of our graphs.

We are now in a position to prove Theorem 3 which we restate for convenience below. Let $\{G^{(k)}\}_{k=1}^N$ be an iid sample of graphs drawn from $\mu_{\rho_n f}$ where f is a canonical stochastic block model kernel. Let $\boldsymbol{A}^{(k)}$ be the adjacency matrix of graph $G^{(k)}$ and let $\boldsymbol{\lambda}^{(k)} = \sigma_c(\boldsymbol{A}^{(k)})$. Let G_N^* be the sample Fréchet mean graph with adjacency matrix $\boldsymbol{A}_N^*$.

Theorem 16 (Theorem 3 in the main chapter) $\forall \epsilon > 0$, $\exists n^* \in \mathbb{N}$ such that for all $n > n^*$,

$$\lim_{N\to\infty} ||\sigma_c(\boldsymbol{A}_N^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon \quad a.s. \tag{A.130}$$

As will become clear in our proof, it will be easier to work with the eigenvalues directly rather than the graphs themselves. To do so we make the following definition.

Definition 24 The set of c largest realizable eigenvalues of adjacency matrices of graphs in $\mathcal{G}$ is defined as

$$\Lambda_n^c = \{\boldsymbol{\lambda} \in \mathbb{R}^c | \exists G \in \mathcal{G} \text{ with adjacency matrix } \boldsymbol{A} \text{ such that } \boldsymbol{\lambda} = \sigma_c(\boldsymbol{A})\}. \tag{A.131}$$

Proof of Theorem 16 We make two observations. First that $\Lambda_n^c \subset \mathbb{R}^c$ and second, equation (A.130) depends only on the eigenvalues of each $\boldsymbol{A}^{(k)}$. These two observations allow us to recast the sample Fréchet mean problem over Λ_n^c and consider the relaxed problem over $\mathbb{R}^c$ whose solution we show to be the optimal eigenvalues of the adjacency matrix of the sample Fréchet mean.

We first recast the problem of computing G_N^* over Λ_n^c as follows. Let $\boldsymbol{\lambda}_N^* = \sigma_c(\boldsymbol{A}_N^*)$. Then

$$\boldsymbol{\lambda}_N^* = \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^N ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}||_2^2. \tag{A.132}$$

We also consider the relaxed version of (A.132) where the solution is in $\mathbb{R}^c$ instead of Λ_n^c. The

relaxed problem is a trivial quadratic optimization problem with a unique solution given by

$$\boldsymbol{\lambda}^*_{N,r} = \underset{\boldsymbol{\lambda}\in\mathbb{R}^c}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^{N} ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}||_2^2. \tag{A.133}$$

$$= \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\lambda}^{(k)} \tag{A.134}$$

which is the classic arithmetic average of the observations $\boldsymbol{\lambda}^{(k)}$. Now, the sample mean, $\boldsymbol{\lambda}^*_{N,r}$, satisfies

$$\frac{1}{N}\sum_{k=1}^{N} ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}||_2^2 = ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2 + \frac{1}{N}\sum_{k=1}^{N} ||\boldsymbol{\lambda}^{(k)}||_2^2 - ||\boldsymbol{\lambda}^*_{N,r}||_2^2. \tag{A.135}$$

Hence, minimizing $||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2$ is equivalent to minimizing $\frac{1}{N}\sum_{k=1}^{N} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^{(k)}||_2^2$ irrespective of the domain over which the function is minimized. This shows that an equivalent formulation of the sample Fréchet mean on the space of realizable eigenvalues is

$$\boldsymbol{\lambda}^*_N = \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^{N} ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}||_2^2 \tag{A.136}$$

$$= \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2. \tag{A.137}$$

Equation (A.137) states that we must find the realizable eigenvalues which are closest to the arithmetic average, the solution to the relaxed problem. Using this formulation of the sample Fréchet mean problem we will show that the eigenvalues of $\boldsymbol{A}^*_N$ converge almost surely to $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]$.

By the strong law of large number, $\forall n$,

$$\lim_{N\to\infty} ||\boldsymbol{\lambda}^*_{N,r} - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]|| = 0 \quad a.s. \tag{A.138}$$

Define

$$\boldsymbol{\lambda}^* = \lim_{N\to\infty} \boldsymbol{\lambda}^*_N \tag{A.139}$$

$$= \lim_{N\to\infty} \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2. \tag{A.140}$$

Since the projection on Λ_n^c is a continuous operator,

$$\lim_{N\to\infty} \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2 = \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} \lim_{N\to\infty} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2. \tag{A.141}$$

Since the norm is continuous,

$$\underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} \lim_{N\to\infty} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}_{N,r}^*||_2^2 = \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} ||\boldsymbol{\lambda} - \lim_{N\to\infty} \boldsymbol{\lambda}_{N,r}^*||_2^2. \tag{A.142}$$

Finally, by (A.138),

$$\underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} ||\boldsymbol{\lambda} - \lim_{N\to\infty} \boldsymbol{\lambda}_{N,r}^*||_2^2 = \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} ||\boldsymbol{\lambda} - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2^2 \quad a.s. \tag{A.143}$$

By Lemma 20, $\forall \epsilon > 0$, there exists $n_1 \in \mathbb{N}$ such that for all $n > n_1$, there exists $G' \in \mathcal{G}$ with adjacency matrix $\boldsymbol{A}'$ such that

$$||\sigma_c(\boldsymbol{A}') - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 \leq \epsilon. \tag{A.144}$$

Because $\sigma_c(\boldsymbol{A}') \in \Lambda_n^c$, the minimizer, $\boldsymbol{\lambda}^*$, in (A.143) satisfies

$$||\boldsymbol{\lambda}^* - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 \leq \epsilon \quad a.s. \tag{A.145}$$

Since $\boldsymbol{\lambda}^* = \lim_{N\to\infty} \boldsymbol{\lambda}_N^* = \lim_{N\to\infty} \sigma_c(\boldsymbol{A}_N^*)$ this concludes our proof. □

Step 4.

All that is left is to compile the results of the prior three steps into a proof for Theorem 1 which we restate below. Assume $G \in \mathcal{G}$ with adjacency matrix $\boldsymbol{A}$ such that

$$0 \preccurlyeq \sigma_c(\boldsymbol{A}) \tag{A.146}$$

and for every $1 \leq i \neq j \leq c$, $\lambda_i \neq \lambda_j$.

Theorem 17 (Theorem 1 from the main chapter) $\forall \epsilon > 0$, $\exists n_1 \in \mathbb{N}$ such that $\forall n > n_1$, $\exists f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})$ a canonical stochastic block model kernel with c communities such that

$$\lim_{N\to\infty} d_{A_c}(G, G_N^*) < \epsilon \quad a.s. \tag{A.147}$$

where G_N^* denotes the empirical Fréchet mean of $\{G^{(k)}\}_{k=1}^N$, an iid sample distributed according to $\mu_{\rho_n f}$.

Proof of Theorem 17 We begin our proof by expanding the left-hand side of equation (A.147).

$$d_{A_c}(G, G_N^*) = ||\sigma_c(\boldsymbol{A}) - \sigma_c(\boldsymbol{A}_N^*)||_2 \tag{A.148}$$

$$= ||\sigma_c(\boldsymbol{A}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right] + \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right] - \sigma_c(\boldsymbol{A}_N^*)||_2. \tag{A.149}$$

Let $\epsilon > 0$. We will show that there exists an $n^* \in \mathbb{N}$ such that for all $n > n^*$, there exists a stochastic block model kernel probability measure $f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s})$ where the following two inequalities hold:

$$||\sigma_c(\boldsymbol{A}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon \tag{A.150}$$

$$||\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right] - \sigma_c(\boldsymbol{A}_N^*)||_2 < \epsilon \quad a.s. \tag{A.151}$$

We begin with (A.150). Define L_f as in Lemma 19 taking $\boldsymbol{\theta} = \frac{\sigma_c(\boldsymbol{A}) - \mathbf{1}}{n\rho_n}$. Then $\lambda_i(\boldsymbol{A}) = n\rho_n\lambda_i(L_f) + 1$. By Theorem 4,

$$\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right] = n\rho_n\boldsymbol{\theta} + \mathbf{1} + \mathcal{O}(\sqrt{\rho_n}). \tag{A.152}$$

Since $\sqrt{\rho_n} \to 0$, there exists an $n_1 \in \mathbb{N}$ such that for all $n > n_1$,

$$||n\rho_n\boldsymbol{\theta} + \mathbf{1} - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon. \tag{A.153}$$

Theorem 3 implies the existence of an $n_2 \in \mathbb{N}$ such that inequality (A.151) holds. Taking $n^* = \max(n_1, n_2)$ implies that both inequalities hold and concludes the proof of Theorem 1. □

A.4 Proof of Theorem 2

The proof of Theorem 2, which we restate below, is a consequence of Lemmas 19 and 20 along with Theorem 4. Let $\{G^{(k)}\}_{k=1}^N$ have sample Fréchet mean G_N^*.

Theorem 18 (Theorem 2 in the main chapter) $\forall \epsilon > 0$, $\exists n^* \in \mathbb{N}$ such that $\forall n > n^*$,

$$\left\| \sigma_c(\boldsymbol{A}_N^*) - \frac{1}{N}\sum_{k=1}^N \sigma_c(\boldsymbol{A}^{(k)}) \right\|_2 < \epsilon. \tag{A.154}$$

Proof of Theorem 18 Let $\boldsymbol{\lambda}^{(k)} = \sigma_c(\boldsymbol{A}^{(k)})$. Define $\bar{\boldsymbol{\lambda}} = \frac{1}{N}\sum_{k=1}^N \boldsymbol{\lambda}^{(k)}$. Let $\bar{\rho}_n$ denote the arithmetic average of the densities of the observed graphs. Take $\boldsymbol{\theta} = \frac{\bar{\boldsymbol{\lambda}} - 1}{n\bar{\rho}_n}$ in Lemma 19 which defines

the disconnected stochastic block model kernel. Let $\epsilon > 0$. By Lemma 20 there exists $n_1 \in \mathbb{N}$ such that for all $n > n_1$, there exists a graph $G' \in \mathcal{G}$ with adjacency matrix $\boldsymbol{A}'$ such that

$$||\sigma_c(\boldsymbol{A}') - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\bar{\rho}_n f}})\right]||_2 < \epsilon. \tag{A.155}$$

By Theorem 4, there exists $n_2 \in \mathbb{N}$ such that for all $n > n_2$,

$$\forall i = 1, ..., c, \left|n\bar{\rho}_n \lambda_i(L_f) + 1 - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\bar{\rho}_n f}})\right]\right| < \epsilon. \tag{A.156}$$

Since $n\bar{\rho}_n \lambda_i(L_f) + 1 = \frac{1}{N}\sum_{k=1}^{N} \lambda_i(\boldsymbol{A}^{(k)})$, there exists a graph G' with adjacency matrix $\boldsymbol{A}'$ that satisfies

$$||\sigma_c(\boldsymbol{A}') - \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\lambda}^{(k)}||_2 < C\epsilon, \tag{A.157}$$

where C is an arbitrary positive constant. We recall, (see (A.137), that the spectrum of the sample Fréchet mean is the solution to

$$\boldsymbol{\lambda}^*_N = \underset{\boldsymbol{\lambda} \in \Lambda_n^c}{\operatorname{argmin}} \; ||\boldsymbol{\lambda} - \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\lambda}^{(k)}||_2^2 \tag{A.158}$$

Now, take $n^* = \max(n_1, n_2)$, then the existence of the graph G' with adjacency matrix $\boldsymbol{A}'$ shows that

$$||\sigma_c(\boldsymbol{A}^*_N) - \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\lambda}^{(k)}||_2^2 < ||\sigma_c(\boldsymbol{A}') - \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\lambda}^{(k)}||_2^2 < \epsilon. \tag{A.159}$$

Thus the adjacency matrix of the sample Fréchet mean must have eigenvalues that are within ϵ of the geometric average. □

A.5 Proof of Theorem 6

Let $\{\tilde{G}^{(k)}\}_{k=1}^{\tilde{N}}$ be a sample of graphs distributed according to $\mu_{\omega_n f}$ where f is the canonical stochastic block model kernel. Define the set mean graph by

$$\widehat{G}^*_{\tilde{N}, \mu_{\omega_n f}} = \underset{\tilde{G} \in \{\tilde{G}^{(k)}\}_{k=1}^{\tilde{N}}}{\operatorname{argmin}} \; \frac{1}{\tilde{N}}\sum_{k=1}^{\tilde{N}} d^2_{A_c}(\tilde{G}, \tilde{G}^{(k)}) \tag{A.160}$$

with adjacency matrix $\hat{\boldsymbol{A}}^*_{N, \mu_{\omega_n f}}$.

Theorem 19 (Theorem 6 from the main chapter) $\forall \epsilon > 0$,

$$\lim_{n\to\infty} P(||\sigma_c(\hat{\boldsymbol{A}}^*_{\tilde{N},\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 > \epsilon) = 0. \tag{A.161}$$

Proof of Theorem 19 We first prove that $\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})$ converges in probability to $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]$ for large graph size. From Theorem 15,

$$\frac{1}{\sqrt{\omega_n}}\left(\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right) \xrightarrow{d} Z \sim N(0,\Sigma) \tag{A.162}$$

where the $c \times c$ covariance matrix Σ is given by (A.113). Let $z > 0$, $\epsilon > 0$, and $\eta > 0$. Because of (A.162), $\exists n_0 \in \mathbb{N}$, $\forall n > n_0$,

$$\left| P\left(\frac{1}{\sqrt{\omega_n}}|\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right]| \leq z\right) - P\left(|z_i| \leq z; i = 1,...,c\right)\right| \leq \frac{\eta}{2}. \tag{A.163}$$

Thus

$$P\left(|z_i| \leq z; i = 1,...,c\right) - \frac{\eta}{2} \leq P\left(\frac{1}{\sqrt{\omega_n}}|\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right]| \leq z\right). \tag{A.164}$$

Now $P(||\boldsymbol{z}|| \leq z) \leq P(|z_i| \leq z; i = 1,...,c)$ and

$$P\left(\frac{1}{\sqrt{\omega_n}}|\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right]| \leq z\right) \leq P\left(||\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \leq \sqrt{c\omega_n}z\right). \tag{A.165}$$

Because $\lim_{z\to\infty} P(||\boldsymbol{z}|| \leq z) = 1$, $\exists z_0 > 0$ such that

$$1 - \frac{\eta}{2} < P(||\boldsymbol{z}|| \leq z_0). \tag{A.166}$$

Also, $\lim_{n\to\infty} \omega_n = 0$ so $\exists n_1 \in \mathbb{N}$ such that $\forall n > n_1$, $\sqrt{\omega_n} < \frac{\epsilon}{z_0\sqrt{c}}$ or $\sqrt{c\omega_n z_0} < \epsilon$. In summary, $\forall \epsilon > 0$, $\forall \eta > 0$, $\exists n_2 = \max(n_0, n_1)$ where

$$1 - \eta < P\left(||\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 < \epsilon\right). \tag{A.167}$$

Equivalently,

$$P\left(||\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \geq \epsilon\right) < \eta. \tag{A.168}$$

In other words, $\forall \epsilon > 0$,

$$\lim_{n\to\infty} P\left(||\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \geq \epsilon\right) = 0. \tag{A.169}$$

We now show that the largest c eigenvalues of the adjacency matrix of the set Fréchet mean graph converges in probability to $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]$. Let $\epsilon > 0$ and let $\eta > 0$. Let $\tilde{N} > 0$ and consider the event

$$\mathcal{E} = \{\boldsymbol{A}^{(1)}, ..., \boldsymbol{A}^{(\tilde{N})}; ||\sigma_c(\widehat{\boldsymbol{A}}_{\tilde{N},\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 > \epsilon\}. \tag{A.170}$$

Note that $\exists k_0 \in \{1, ..., \tilde{N}\}$ where $\widehat{\boldsymbol{A}}_{\tilde{N},\mu_{\omega_n f}} = \boldsymbol{A}^{(k_0)}$ where $\boldsymbol{A}^{(k_0)} \sim \mu_{\omega_n f}$. Now because of (A.169), $\exists n_0 \in \mathbb{N}$ where $\forall n > n_0$, $P(\mathcal{E}) < \eta$. We conclude that $\forall \epsilon > 0$,

$$\lim_{n\to\infty} P\left(||\sigma_c(\widehat{\boldsymbol{A}}_{\tilde{N},\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 > \epsilon\right) = 0. \tag{A.171}$$

Appendix B

Supplementary material for Chapter 3

We divide the appendix into five primary sections. First, we introduce a classic theorem related to our work in Appendix B.1. Appendix B.2 is brief, in which Lemma 4 is proven. We next prove Corollary 2 in Appendix B.3. Appendix B.4 is devoted to the first-order computations of the eigenvalues and eigenvectors of stochastic block model graphs when $q = \epsilon p_{\min}$ where $p_{\min} = \min \boldsymbol{p}$. Appendix B.5 shows the computations for Lemma 3.

B.1 Classical results

Theorem 20 (Weyl-Lidskii)

Let $\boldsymbol{H}$ be a self-adjoint operator on a Hilbert space $\mathcal{H}$. Let $\boldsymbol{A}$ be a bounded operator on $\mathcal{H}$ Let $\sigma(\boldsymbol{H})$ and $\sigma(\boldsymbol{H} + \boldsymbol{A})$ denote the spectra of $\boldsymbol{H}$ and $(\boldsymbol{H} + \boldsymbol{A})$ respectively. Then

$$\sigma(\boldsymbol{H} + \boldsymbol{A}) \subset \{\lambda : dist(\lambda, \sigma(\boldsymbol{H})) \leq ||\boldsymbol{A}||\} \tag{B.1}$$

where $||\boldsymbol{A}||$ denotes the operator norm of $\boldsymbol{A}$.

Proof of Theorem 20 These are standard bounds that can be found in many good books on matrix perturbation theory (e.g., [94]).

B.2 Approximately computing the sample Fréchet variance

This appendix proves Lemma 4, which is a consequence of the following theorem. Let $\{G^{(k)}\}_{k=1}^{N}$ be a sample of graphs with sample Fréchet mean G_N^* with adjacency matrix $\boldsymbol{A}_N^*$. Let $\bar{\boldsymbol{\lambda}}$

denote the arithmetic mean of the c largest eigenvalues.

Theorem 21 $\forall \epsilon > 0$, $\exists n^* \in \mathbb{N}$ such that $\forall n > n^*$,

$$||\sigma_c(\boldsymbol{A}_N^*) - \bar{\boldsymbol{\lambda}}||_2 < \epsilon. \tag{B.2}$$

Proof of Theorem 21 The proof is in [33].

This shows that the c largest eigenvalues of the sample Fréchet mean graph are approximated well by the sample mean eigenvalues when the graphs considered, are sufficiently large. We utilize this fact to compute the value of the sample total Fréchet variance.

Lemma 21 (Lemma 4 from the main chapter) Let $\{G^{(k)}\}_{k=1}^N$ be a sample of graphs with sample Fréchet mean G_N^* and total sample Fréchet variance $V_{N,tot}^*$. Let $\bar{\boldsymbol{\lambda}}$ denote the arithmetic mean of the largest c eigenvalues and $\hat{\Sigma}$ be the sample covariance matrix, then

$$\lim_{n\to\infty} |V_{N,tot}^* - \sum_{i=1}^{c} \hat{\Sigma}_{ii}| = 0 \tag{B.3}$$

Proof of Lemma 21 By Theorem 21, we have $\forall \epsilon > 0$, $\exists n^* \in \mathbb{N}$ such that for all $n > n^*$,

$$||\sigma_c(\boldsymbol{A}_N^*) - \bar{\boldsymbol{\lambda}}||_2 < \epsilon. \tag{B.4}$$

Observe the following,

$$\sum_{i=1}^{c} \hat{\Sigma}_{ii} = \frac{1}{N-1} \sum_{k=1}^{N} (\boldsymbol{\lambda}^{(k)} - \bar{\boldsymbol{\lambda}})^T (\boldsymbol{\lambda}^{(k)} - \bar{\boldsymbol{\lambda}}) \tag{B.5}$$

$$= \frac{1}{N-1} \sum_{k=1}^{N} ||\boldsymbol{\lambda}^{(k)} - \bar{\boldsymbol{\lambda}}||_2^2 \tag{B.6}$$

Let $\boldsymbol{\lambda}_N^* = \sigma_c(\boldsymbol{A}_N^*)$ and consider the sample total Fréchet variance.

$$V_{N,tot}^* = \frac{1}{N-1} \sum_{k=1}^{N} d_{A_c}^2(G_N^*, G^{(k)}) \tag{B.7}$$

$$= \frac{1}{N-1} \sum_{k=1}^{N} ||\sigma_c(\boldsymbol{A}_N^*) - \sigma_c(\boldsymbol{A}^{(k)})||_2^2 \tag{B.8}$$

$$= \frac{1}{N-1} \sum_{k=1}^{N} ||\boldsymbol{\lambda}_N^* - \boldsymbol{\lambda}^{(k)}||_2^2. \tag{B.9}$$

Because the summation is finite, we only need to show that each term in the sums in equations (B.6) and (B.9) are close.

$$||\boldsymbol{\lambda}^{(k)} - \bar{\boldsymbol{\lambda}}||_2 = ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}_N^* + \boldsymbol{\lambda}_N^* - \bar{\boldsymbol{\lambda}}||_2 \tag{B.10}$$

$$= ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}_N^*||_2 + ||\boldsymbol{\lambda}_N^* - \bar{\boldsymbol{\lambda}}||_2 \tag{B.11}$$

$$\leq ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}_N^*||_2 + \epsilon. \tag{B.12}$$

Therefore,

$$|||\boldsymbol{\lambda}^{(k)} - \bar{\boldsymbol{\lambda}}||_2 - ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}_N^*||_2| < |||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}_N^*||_2 + \epsilon - ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}_N^*||_2| \tag{B.13}$$

$$= \epsilon. \tag{B.14}$$

As a consequence, the difference in the summations is

$$\left| V_{N,tot}^* - \sum_{i=1}^{c} \hat{\Sigma}_{ii} \right| \leq \frac{1}{N-1} \sum_{k=1}^{N} \left| ||\boldsymbol{\lambda}_N^* - \boldsymbol{\lambda}^{(k)}||_2^2 - ||\boldsymbol{\lambda}^{(k)} - \bar{\boldsymbol{\lambda}}||_2^2 \right| \tag{B.15}$$

$$\leq \frac{1}{N-1} \sum_{k=1}^{N} |\epsilon| \tag{B.16}$$

$$= \frac{N}{N-1} \epsilon. \tag{B.17}$$

Therefore, the summations are arbitrarily close for sufficiently large graphs, $n > n^*$.

B.3 Proof of Corollary 2

This appendix serves to prove Corollary 2 which is a consequence of Theorem 8. We restate both of these below for convenience. Let f be a canonical stochastic block model kernel function and let L_f be the associated linear integral operator with eigenfunctions $r_i(x)$ and eigenvalues denoted by $\theta_i = \lambda_i(L_f)$. Assume that $n^{-2/3} \ll \omega_n \ll 1$ and that $\lim_{n\to\infty} \omega_n = 0$. Because μ is always taken to be a stochastic block model kernel probability measure with parameters $\omega_n, \boldsymbol{p}$, q, and $\boldsymbol{s}$, we denote the adjacency matrix of a random graph as $\boldsymbol{A}_\mu = \boldsymbol{A}_{\mu_{\omega_n,\boldsymbol{p},q,\boldsymbol{s}}}$, where we have suppressed all the subscripts.

Theorem 22 (Chakrabarty, Chakraborty, Hazra 2020)

$$\left(\omega_n^{-1/2}(\lambda_i(\boldsymbol{A}_\mu) - \mathbb{E}_\mu[\lambda_i(\boldsymbol{A}_\mu)])\right) \xrightarrow{d} (Z_i : 1 \leq i \leq c), \tag{B.18}$$

where the right-hand side is a multivariate normal random vector in $\mathbb{R}^c$ with zero mean and

$$\text{Cov}(Z_i, Z_j) = 2\int_0^1 \int_0^1 r_i(x)r_i(y)r_j(x)r_j(y)f(x,y)dxdy, \tag{B.19}$$

for all $1 \leq i, j \leq c$. The first order behavior of $\mathbb{E}\left[\lambda_i(\boldsymbol{A}_\mu)\right])$ is given by the following: For every $1 \leq i \leq c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_\mu)\right] = \lambda_i(\boldsymbol{B}) + \mathcal{O}(\sqrt{\omega}_n + \frac{1}{n\omega_n}), \tag{B.20}$$

where $\boldsymbol{B}$ is a $c \times c$ symmetric deterministic matrix defined by

$$b_{j,l} = \sqrt{\theta_j\theta_l}n\omega_n \boldsymbol{e}_j^T\boldsymbol{e}_l + \theta_i^{-2}\sqrt{\theta_j\theta_l}(n\omega_n)^{-1}\boldsymbol{e}_j^T\mathbb{E}\left[(\boldsymbol{A}_\mu - \mathbb{E}\left[\boldsymbol{A}_\mu\right])^2\right]\boldsymbol{e}_l + \mathcal{O}(\frac{1}{n\omega_n}), \tag{B.21}$$

and $\boldsymbol{e}_j$ is a vector with entries $\boldsymbol{e}_j(k) = \frac{1}{\sqrt{n}}r_j(\frac{k}{n})$ for $1 \leq j \leq c$.

Proof of Theorem 22 This is a compilation of Theorems 2.3 and 2.4 from [22].

Define the matrices

$$\boldsymbol{M} = \begin{bmatrix} s_1p_1 & \sqrt{s_1s_2}q & \dots & \sqrt{s_1s_c}q \\ \sqrt{s_2s_1}q & s_2p_2 & \dots & \sqrt{s_2s_c}q \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{s_cs_1}q & \sqrt{s_cs_2}q & \dots & s_cp_c \end{bmatrix} \quad \boldsymbol{M}_f = \begin{bmatrix} p_1 & q & \dots & q \\ q & p_2 & \dots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \dots & p_c \end{bmatrix} \tag{B.22}$$

where ν_k and $\boldsymbol{v}_k$ are the eigenvalues and eigenvectors of $\boldsymbol{M}$ respectively.

Corollary 4

$$\left(\omega_n^{-1/2}(\lambda_i(\boldsymbol{A}_\mu) - \mathbb{E}_\mu[\lambda_i(\boldsymbol{A}_\mu)])\right) \xrightarrow{d} (Z_i : 1 \leq i \leq c), \tag{B.23}$$

where the right-hand side is a multivariate normal random vector in $\mathbb{R}^c$ with zero mean and

$$\text{Cov}(Z_i, Z_j) = 2\left(\boldsymbol{v}_k. * \boldsymbol{v}_j\right)^T \boldsymbol{M}_f \left(\boldsymbol{v}_k. * \boldsymbol{v}_j\right), \tag{B.24}$$

for all $1 \leq i, j \leq c$ and $.*$ denotes the component-wise product of the vectors.

The first order behavior of $\mathbb{E}\left[\lambda_i(\boldsymbol{A}_\mu)\right])$ is given by the following: For every $1 \leq i \leq c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_\mu)\right] = \lambda_i(\boldsymbol{B}^*) + \mathcal{O}(\sqrt{\omega_n}) \tag{B.25}$$

where $\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}$ whose components are given as

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b_{j,l}^{*,(1)} = \begin{cases} \nu_j n \omega_n & \text{if } j = l \\ 0 & \text{if } j \neq l \end{cases} \tag{B.26}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b_{j,l}^{*,(2)} = \nu_i^{-2}\sqrt{\nu_j \nu_l}\sum_{k=1}^{c}\nu_k \sum_{m=1}^{c}\frac{1}{\sqrt{s_m}}\boldsymbol{v}_j(m)\boldsymbol{v}_l(m)\boldsymbol{v}_k(m)\sum_{w=1}^{c}\sqrt{s_w}\boldsymbol{v}_k(w). \tag{B.27}$$

Proof of Corollary 3 The proof is a consequence of Theorems 22 and 23 along with Lemmas 22 and 23.

To connect the corollary to Theorem 23, we introduce the following theorem from [33], which shows that the terms of the matrix $\boldsymbol{B}$ defined element-wise by equation (B.21) can be estimated by a similar matrix $\boldsymbol{B}^*$.

Theorem 23 For every $1 \leq i \leq c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] = \lambda_i(\boldsymbol{B}^*) + \mathcal{O}(\sqrt{\omega}_n) \tag{B.28}$$

where $\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}$ whose components are given as

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b_{j,l}^{*,(1)} = \sqrt{\theta_j \theta_l} n \omega_n \int_0^1 r_j(x) r_l(x) dx = \begin{cases} \theta_j n \omega_n & j = l \\ 0 & j \neq l \end{cases} \tag{B.29}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b_{j,l}^{*,(2)} = \theta_i^{-2}\sqrt{\theta_j \theta_l}\int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx. \tag{B.30}$$

Proof of Theorem 23 The proof is in [33]. Notably, this is a minor modification to Theorem 2.4 in [22] when assuming $\omega_n \to 0$. The proof relies on the fact that the terms in equation (B.21) are a discretization of the quantities defined in equations (B.29) and (B.30), and that because the eigenfunctions are piecewise Lipschitz, the discretization converges at the rate $\mathcal{O}(\frac{1}{n})$.

We show the following three equalities within this appendix;

$$\theta_j n\omega_n = \nu_j n\omega_n \tag{B.31}$$

$$\theta_i^{-2}\sqrt{\theta_j\theta_l}\int_0^1 r_j(x)r_l(x)\int_0^1 f(x,y)dydx = \nu_i^{-2}\sqrt{\nu_j\nu_l}\sum_{k=1}^{c}\nu_k\sum_{m=1}^{c}\frac{1}{\sqrt{s_m}}\boldsymbol{v}_j(m)\boldsymbol{v}_l(m)\boldsymbol{v}_k(m)\sum_{w=1}^{c}\sqrt{s_w}\boldsymbol{v}_k(w) \tag{B.32}$$

$$2\int_0^1\int_0^1 r_i(x)r_i(y)r_j(x)r_j(y)f(x,y)dxdy = 2\left(\boldsymbol{v}_k.*\boldsymbol{v}_j\right)^T\boldsymbol{M}_f\left(\boldsymbol{v}_k.*\boldsymbol{v}_j\right). \tag{B.33}$$

The structure of the proof is as follows,

(1) In Lemma 22, we show two quantities: (1) the eigenfunctions of L_f are defined by the components of the vector $\boldsymbol{v}$ and (2) θ_k is given by ν_k.

(2) By relating the eigenfunctions to the components of the vectors, we show that the integrals in equations (B.19) and (B.30) can be written in terms of the vectors $\boldsymbol{v}_k$. Equation (B.29) is a straightforward consequence of part 2 in step 1 because there are effectively no integrals.

Step 1

Lemma 22 Let ν_k and $\boldsymbol{v}_k$ be eigenvalues and eigenvectors of $\boldsymbol{M}$ respectively. Let θ_k and $r_k(x)$ be the eigenvalues and eigenfunctions of L_f respectively. We then have the following,

$$\theta_k = \nu_k \tag{B.34}$$

$$r_k(x_i^*) = \frac{\boldsymbol{v}_k(i)}{\sqrt{s_i}} \quad i = 1, ..., c. \tag{B.35}$$

As shown in Section 3 of [22], $r_k(x)$ is piecewise constant on c different blocks. The values for each block are given by (B.35). We characterize the intervals where each $r_k(x)$ is piecewise constant as

$$S_i = \left[\sum_{j=1}^{i-1}s_j, \sum_{j=1}^{i}s_j\right) \subset [0,1]. \tag{B.36}$$

Proof of Lemma 22 We first define a few notations. Recall that $\boldsymbol{s}$ is the vector of relative community sizes. Observe that $S_i \cap S_j = \emptyset$ for all $i \neq j$ and that $\int_{S_i} dx = s_i$, a fact we use throughout the proof. Additionally, each eigenfunction $r_k(x)$ is piecewise constant on each interval S_i. Define the points

$$x_i^* \in S_i \tag{B.37}$$

as a fixed point in each interval S_i. We first observe that

$$\int_0^1 r_k(x)dx = \sum_{i=1}^{c} \int_{S_i} r_k(x)dx \tag{B.38}$$

$$= \sum_{i=1}^{c} r_k(x_i^*) \int_{S_i} dx \tag{B.39}$$

$$= \sum_{i=1}^{c} r_k(x_i^*)s_i. \tag{B.40}$$

This fact will aid our proof significantly. Define the vector

$$\boldsymbol{r}_k^n = \begin{bmatrix} \sqrt{s_1} r_k(x_1^*) \\ \vdots \\ \sqrt{s_c} r_k(x_c^*) \end{bmatrix}. \tag{B.41}$$

Consider that

$$\langle \boldsymbol{r}_k^n, \boldsymbol{r}_j^n \rangle = \sum_{i=1}^{c} s_i r_k(x_i^*) r_j(x_i^*) \tag{B.42}$$

$$= \sum_{i=1}^{c} r_k(x_i^*) r_j(x_i^*) \int_{S_i} dx \tag{B.43}$$

$$= \sum_{i=1}^{c} \int_{S_i} r_k(x) r_j(x) dx \tag{B.44}$$

$$= \int_0^1 r_k(x) r_j(x) dx \tag{B.45}$$

where we have used the fact that both $r_k(x)$ and $r_j(x)$ are constant on the intervals S_i. This implies that

$$\boldsymbol{r}_k^n \perp \boldsymbol{r}_j^n \quad \forall k \neq j \tag{B.46}$$

$$1 = \langle \boldsymbol{r}_k^n, \boldsymbol{r}_k^n \rangle. \tag{B.47}$$

We now consider

$$\sum_{k=1}^{c} \theta_k \boldsymbol{r}_k^n(i) \boldsymbol{r}_k^n(j) = \sum_{k=1}^{c} \sqrt{s_i}\theta_k r_k(x_i^*)\sqrt{s_j} r_k(x_j^*) \tag{B.48}$$

$$= \sqrt{s_i s_j} \sum_{k=1}^{c} \theta_k r_k(x_i^*)\sqrt{s_j} r_k(x_j^*) \tag{B.49}$$

$$= \sqrt{s_i s_j} f(x_i^*, x_j^*) \tag{B.50}$$

$$= (\boldsymbol{M})_{ij} \tag{B.51}$$

where we have used the definition

$$f(x, y) = \sum_{k=1}^{c} \theta_k r_k(x) r_k(y). \tag{B.52}$$

The conclusion is then

$$\boldsymbol{M} = \sum_{k=1}^{n} \theta_k \boldsymbol{r}_k^n (\boldsymbol{r}_k^n)^T, \tag{B.53}$$

and $\boldsymbol{r}_k^n$ is an eigenvector. Because $\boldsymbol{M}$ is given in terms of the parameters $\boldsymbol{p}, q$, and $\boldsymbol{s}$, we may compute the eigenvalues and eigenfunctions of $\boldsymbol{M}$ as ν_k and $\boldsymbol{v}_k$, where $\boldsymbol{v}_k$ satisfies

$$\boldsymbol{v}_k = \boldsymbol{r}_k^n. \tag{B.54}$$

We can therefore determine the value of each eigenfunction at the points x_i^* as

$$r_k(x_i^*) = \frac{\boldsymbol{v}_k(i)}{\sqrt{s_i}}, \tag{B.55}$$

and the eigenvalues,

$$\theta_k = \nu_k. \tag{B.56}$$

Because the eigenfunctions are piecewise constant and we know one value within each piece, this defines the entire eigenfunction which shows equation (B.34) and equation (B.35).

Step 2

The result of the prior lemma can now be used to calculate the quantities in equations (B.31), (B.32), and (B.33) which will accomplish step 2.

Recall the matrices

$$\boldsymbol{M} = \begin{bmatrix} s_1 p_1 & \sqrt{s_1 s_2} q & \dots & \sqrt{s_1 s_c} q \\ \sqrt{s_2 s_1} q & s_2 p_2 & \dots & \sqrt{s_2 s_c} q \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{s_c s_1} q & \sqrt{s_c s_2} q & \dots & s_c p_c \end{bmatrix} \quad \boldsymbol{M}_f = \begin{bmatrix} p_1 & q & \dots & q \\ q & p_2 & \dots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \dots & p_c \end{bmatrix} \tag{B.57}$$

Lemma 23 Let ν_k, $\boldsymbol{v}_k$ be eigenvalues and eigenvectors of $\boldsymbol{M}$. Then, for every $1 \leq i, j, l \leq c$,

$$\theta_j n \omega_n = \nu_j n \omega_n \tag{B.58}$$

$$\theta_i^{-2} \sqrt{\theta_j \theta_l} \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx = \nu_i^{-2} \sqrt{\nu_j \nu_l} \sum_{k=1}^{c} \nu_k \sum_{m=1}^{c} \frac{1}{\sqrt{s_m}} \boldsymbol{v}_j(m) \boldsymbol{v}_l(m) \boldsymbol{v}_k(m) \sum_{w=1}^{c} \sqrt{s_w} \boldsymbol{v}_k(w) \tag{B.59}$$

$$2 \int_0^1 \int_0^1 r_i(x) r_i(y) r_j(x) r_j(y) f(x,y) dx dy = 2 \left(\boldsymbol{v}_k. * \boldsymbol{v}_j \right)^T \boldsymbol{M}_f \boldsymbol{v}_k. * \boldsymbol{v}_j. \tag{B.60}$$

Proof of Lemma 23 The lemma is shown by using the expressions derived in Lemma 22 along with the summation representation of $f(x,y)$. Equation (B.58) is trivial to show because

$$\theta_j = \nu_j, \tag{B.61}$$

as shown in Lemma 22. We now consider equation (B.59). To show this equality, we show how to compute the integral

$$\int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx \tag{B.62}$$

in terms of the vectors $\boldsymbol{v}_k$. First, we use the summation representation of $f(x,y)$ to simplify as follows

$$\int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx = \int_0^1 r_j(x) r_l(x) \int_0^1 \sum_{k=1}^{c} \theta_k r_k(x) r_k(y) dy dx \tag{B.63}$$

$$= \sum_{k=1}^{c} \theta_k \int_0^1 r_j(x) r_l(x) r_k(x) dx \int_0^1 r_k(y) dy. \tag{B.64}$$

Next, we observe that the eigenfunction is piecewise constant on the same intervals defined by S_i,

$$\sum_{k=1}^{c} \theta_k \int_0^1 r_j(x) r_l(x) r_k(x) dx \int_0^1 r_k(y) dy = \sum_{k=1}^{c} \theta_k \sum_{m=1}^{c} \int_{S_m} r_j(x) r_l(x) r_k(x) dx \sum_{w=1}^{c} \int_{S_w} r_k(y) dy. \tag{B.65}$$

Now, because each eigenfunction is constant on the intervals, we may pull out the constant, which is simply the function values evaluated at a point in the interval.

$$\sum_{k=1}^{c} \theta_k \sum_{m=1}^{c} \int_{S_m} r_j(x) r_l(x) r_k(x) dx \sum_{w=1}^{c} \int_{S_w} r_k(y) dy = \sum_{k=1}^{c} \theta_k \sum_{m=1}^{c} r_j(x_m^*) r_l(x_m^*) r_k(x_m^*) \int_{S_m} dx \sum_{w=1}^{c} r_k(x_w^*) \int_{S_w} dy \tag{B.66}$$

Observe now that the integral $\int_{S_m} dx = s_m$.

$$\sum_{k=1}^{c} \theta_k \sum_{m=1}^{c} r_j(x_m^*) r_l(x_m^*) r_k(x_m^*) \int_{S_m} dx \sum_{w=1}^{c} r_k(x_w^*) \int_{S_w} dy = \sum_{k=1}^{c} \theta_k \sum_{m=1}^{c} s_m r_j(x_m^*) r_l(x_m^*) r_k(x_m^*) \sum_{w=1}^{c} s_w r_k(x_w^*) \tag{B.67}$$

We next utilize equation (B.35)

$$\sum_{k=1}^{c} \theta_k \sum_{m=1}^{c} s_m r_j(x_m^*) r_l(x_m^*) r_k(x_m^*) \sum_{w=1}^{c} s_w r_k(x_w^*) = \sum_{k=1}^{c} \theta_k \sum_{m=1}^{c} s_m \frac{\boldsymbol{v}_j(m)}{\sqrt{s_m}} \frac{\boldsymbol{v}_l(m)}{\sqrt{s_m}} \frac{\boldsymbol{v}_k(m)}{\sqrt{s_m}} \sum_{w=1}^{c} s_w \frac{\boldsymbol{v}_k(w)}{\sqrt{s_w}} \tag{B.68}$$

and we then simplify, which yields the following:

$$\sum_{k=1}^{c} \theta_k \sum_{m=1}^{c} s_m \frac{\boldsymbol{v}_j(m)}{\sqrt{s_m}} \frac{\boldsymbol{v}_l(m)}{\sqrt{s_m}} \frac{\boldsymbol{v}_k(m)}{\sqrt{s_m}} \sum_{w=1}^{c} s_w \frac{\boldsymbol{v}_k(w)}{\sqrt{s_w}} = \sum_{k=1}^{c} \theta_k \sum_{m=1}^{c} \frac{1}{\sqrt{s_m}} \boldsymbol{v}_j(m) \boldsymbol{v}_l(m) \boldsymbol{v}_k(m) \sum_{w=1}^{c} \sqrt{s_w} \boldsymbol{v}_k(w). \tag{B.69}$$

Replacing each $\theta_k = \nu_k$ as shown by equation (B.34) gives the result.

We derive how to express equation (B.19) in terms of the eigenvalues and eigenvectors of $\boldsymbol{M}$ by way of a very similar computation,

$$\begin{aligned}
\mathrm{Cov}(Z_i, Z_j) &= 2 \int_0^1 \int_0^1 r_i(x) r_i(y) r_j(x) r_j(y) f(x,y) dx dy && \text{(B.70)} \\
&= 2 \int_0^1 r_i(x) r_j(x) \int_0^1 \sum_{k=1}^{c} \theta_k r_k(x) r_i(y) r_j(y) r_k(y) dx dy && \text{(B.71)} \\
&= 2 \sum_{m=1}^{c} \int_{S_m} \sum_{w=1}^{c} \int_{S_w} r_i(x) r_j(x) \sum_{k=1}^{c} \theta_k r_k(x) r_i(y) r_j(y) r_k(y) dx dy. && \text{(B.72)}
\end{aligned}$$

We now use the piecewise constant behavior of the functions,

$$\mathrm{Cov}(Z_i, Z_j) = 2\sum_{m=1}^{c}\int_{S_m}\sum_{w=1}^{c}\int_{S_w} r_i(x_m^*)r_j(x_m^*)\sum_{k=1}^{c}\theta_k r_k(x_m^*)r_i(y_w^*)r_j(y_w^*)r_k(y_w^*)dxdy \tag{B.73}$$

$$= 2\sum_{m=1}^{c} s_m \sum_{w=1}^{c} s_w r_i(x_m^*)r_j(x_m^*)\sum_{k=1}^{c}\theta_k r_k(x_m^*)r_i(y_w^*)r_j(y_w^*)r_k(y_w^*) \tag{B.74}$$

where we have used $\int_{S_m} dx = s_m$ and there is no longer any x or y dependence in the equations. Recall now that

$$f(x_m^*, y_w^*) = \sum_{k=1}^{c}\theta_k r_k(x_m^*)r_i(y_w^*). \tag{B.75}$$

Using this and writing $s_w = \sqrt{s_w}\sqrt{s_w}$,

$$\mathrm{Cov}(Z_i, Z_j) = 2\sum_{m=1}^{c}\sum_{w=1}^{c}\sqrt{s_m}r_i(x_m^*)\sqrt{s_m}r_j(x_m^*)f(x_m^*, x_w^*)\sqrt{s_w}r_j(y_w^*)\sqrt{s_w}r_k(y_w^*). \tag{B.76}$$

The right-hand side of the above equation can be written as $\boldsymbol{h}^T\boldsymbol{M}_f\boldsymbol{h}$ where

$$\boldsymbol{h}(m) = \sqrt{s_m}r_i(x_m^*)\sqrt{s_m}r_j(x_m^*). \tag{B.77}$$

Writing in terms of the vector $\boldsymbol{v}_k$ we see that

$$\boldsymbol{h} = \boldsymbol{v}_i . * \boldsymbol{v}_j, \tag{B.78}$$

where $.*$ denotes the component-wise product. Therefore,

$$\mathrm{Cov}(Z_i, Z_j) = 2\left(\boldsymbol{v}_i . * \boldsymbol{v}_j\right)^T \boldsymbol{M}_f \boldsymbol{v}_i . * \boldsymbol{v}_j \tag{B.79}$$

B.4 First-Order Computations

Throughout this appendix, we prove Lemma 2, which is restated below for convenience.

Lemma 24 (Lemma 2 from the main chapter) Let $\boldsymbol{M}$ and $\boldsymbol{M}_f$ be as defined in equation (3.15). Assume $q = \epsilon p_{\min}$ where $p_{\min} = \min_{i=1,\ldots,c} \boldsymbol{p}$ and $\epsilon \ll 1$.

$$\left|\frac{\mathbb{E}\left[\lambda_i(\boldsymbol{A})\right]}{n\omega_n s_i} - p_i\right| = \mathcal{O}\left(\epsilon^2 p_{\min}\right) + \mathcal{O}\left(\frac{t_i(\boldsymbol{p})}{n\omega_n s_i}\right), \tag{B.80}$$

$$\left|\mathrm{Cov}(Z_i, Z_j) - \begin{cases} 2p_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}\right| = \mathcal{O}(\epsilon^2 p_{\min}^2), \tag{B.81}$$

where $t_i(\boldsymbol{p})$ is a bounded function of the parameters $\boldsymbol{p}$ and Z is defined in Theorem 8.

Proof of Lemma 24 The proof is a consequence of Lemmas 25 and 26

The ideas for the proof are related to the quantities in Corollary 2. Recall that

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A})\right] = \lambda_i(\boldsymbol{B}^*) + \mathcal{O}(\sqrt{\omega_n}), \tag{B.82}$$

$$\mathrm{Cov}(Z_i, Z_j) = 2\left(\boldsymbol{v}_i. * \boldsymbol{v}_j\right)^T \boldsymbol{M}_f \boldsymbol{v}_i. * \boldsymbol{v}_j. \tag{B.83}$$

Within this appendix, we always assume that $q = \epsilon p_{\min}$ where $\epsilon \ll 1$ and $p_{\min} = \min \boldsymbol{p}$. Under these assumptions we calculate a first-order expansion for $\lambda_i(\boldsymbol{B}^*)$ and $2\left(\boldsymbol{v}_i. * \boldsymbol{v}_j\right)^T \boldsymbol{M}_f \boldsymbol{v}_i. * \boldsymbol{v}_j.$ to prove Lemma 24. To accomplish these tasks, we split this appendix into three subsections

(1) In B.4.1, we show the first-order behavior of the eigenvectors and eigenvalues of $\boldsymbol{M}$ when $q = \epsilon p_{\min}$.

(2) In B.4.2, we show equation (B.80).

(3) In B.4.3, we show equation (B.81).

B.4.1 Step 1: Preliminaries

Assume that $q = \epsilon p_{\min}$, where $\epsilon \ll 1$ and $p_{\min} = \min \boldsymbol{p}$. Let $\boldsymbol{M}$ be defined as in equation (3.15) with eigenvalues and eigenvectors given by ν_k and $\boldsymbol{v}_k$, respectively. The matrix $\boldsymbol{M}$ may be split into two components,

$$\boldsymbol{M}_0 = \begin{bmatrix} s_1 p_1 & 0 & \dots & 0 \\ 0 & s_2 p_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_c p_c \end{bmatrix} \tag{B.84}$$

$$\epsilon p_{\min} \boldsymbol{M}_1 = \begin{bmatrix} 0 & \sqrt{s_1 s_2} & \dots & \sqrt{s_1 s_c} \\ \sqrt{s_2 s_1} q & 0 & \dots & \sqrt{s_2 s_c} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{s_c s_1} & \sqrt{s_c s_2} & \dots & 0 \end{bmatrix} \tag{B.85}$$

$$\boldsymbol{M} = \boldsymbol{M}_0 + \epsilon p_{\min} \boldsymbol{M}_1 \tag{B.86}$$

Let γ_i and $\boldsymbol{g}_i$ be the eigenvalues and eigenvectors of $\boldsymbol{M}_0$. Clearly,

$$\gamma_i = s_i p_i, \tag{B.87}$$

$$\boldsymbol{g}_i = \boldsymbol{e}_i, \tag{B.88}$$

where $\boldsymbol{e}_i$ denotes the canonical basis vector in $\mathbb{R}^c$. We assume that the eigenvalues of $\boldsymbol{M}_0$, denoted by γ_i, are well separated so that the first order expansion of the eigenvectors is well-behaved. This is an assumption of the model and is the same assumption that the eigenvalues of L_f are well-separated. An expansion for the eigenvectors of $\boldsymbol{M}$, denoted by $\boldsymbol{v}_i$, in terms of the eigenvectors of $\boldsymbol{M}_0$ is then

$$\boldsymbol{v}_i = \boldsymbol{e}_i + \epsilon p_{\min} \sum_{k \neq i}^{c} \frac{\boldsymbol{e}_i \boldsymbol{M}_1 \boldsymbol{e}_k}{\gamma_i - \gamma_k} \boldsymbol{e}_k + \mathcal{O}((\epsilon p_{\min})^2) \tag{B.89}$$

$$= \boldsymbol{e}_i + \epsilon p_{\min} \sum_{k \neq i}^{c} \frac{\sqrt{s_i s_k}}{s_i p_i - s_k p_k} \boldsymbol{e}_k + \mathcal{O}((\epsilon p_{\min})^2). \tag{B.90}$$

This section also determines a first-order expansion for the eigenvalues of $\boldsymbol{M}$, denoted by ν_i.

$$\nu_i = \boldsymbol{v}_i^T \boldsymbol{M} \boldsymbol{v}_i \tag{B.91}$$

$$= \left(\boldsymbol{e}_i + \epsilon p_{\min} \sum_{k \neq i}^{c} \frac{\sqrt{s_i s_k}}{s_i p_i - s_k p_k} \boldsymbol{e}_k \right)^T \boldsymbol{M} \left(\boldsymbol{e}_i + \epsilon p_{\min} \sum_{k \neq i}^{c} \frac{\sqrt{s_i s_k}}{s_i p_i - s_k p_k} \boldsymbol{e}_k \right) + \mathcal{O}((\epsilon p_{\min})^2) \tag{B.92}$$

$$= \boldsymbol{e}_i^T \boldsymbol{M} \boldsymbol{e}_i + \epsilon p_{\min} \sum_{k \neq i}^{c} \frac{\sqrt{s_i s_k}}{s_i p_i - s_k p_k} \boldsymbol{e}_k^T \boldsymbol{M} \boldsymbol{e}_i + \epsilon p_{\min} \boldsymbol{e}_i^T \boldsymbol{M} \sum_{k \neq i}^{c} \frac{\sqrt{s_i s_k}}{s_i p_i - s_k p_k} \boldsymbol{e}_k + \mathcal{O}((\epsilon p_{\min})^2). \tag{B.93}$$

We have the following identity to help simplify:

$$\boldsymbol{e}_i^T \boldsymbol{M} \boldsymbol{e}_k = (\boldsymbol{M})_{ik} = m_{ik} = m_{ki} \tag{B.94}$$

since $\boldsymbol{M}$ is symmetric. Therefore,

$$\nu_i = \boldsymbol{e}_i^T \boldsymbol{M} \boldsymbol{e}_i + \epsilon p_{\min} \sum_{k \neq i}^{c} \frac{\sqrt{s_i s_k}}{s_i p_i - s_k p_k} \boldsymbol{e}_k^T \boldsymbol{M} \boldsymbol{e}_i + \epsilon p_{\min} \boldsymbol{e}_i^T \boldsymbol{M} \sum_{k \neq i}^{c} \frac{\sqrt{s_i s_k}}{s_i p_i - s_k p_k} \boldsymbol{e}_k + \mathcal{O}((\epsilon p_{\min})^2) \quad \text{(B.95)}$$

$$= \boldsymbol{e}_i^T \boldsymbol{M} \boldsymbol{e}_i + 2\epsilon p_{\min} \sum_{k \neq i}^{c} \frac{\sqrt{s_i s_k}}{s_i p_i - s_k p_k} \boldsymbol{e}_k^T \boldsymbol{M} \boldsymbol{e}_i + \mathcal{O}((\epsilon p_{\min})^2) \quad \text{(B.96)}$$

$$= m_{ii} + 2 \sum_{k \neq i}^{c} \epsilon p_{\min} \frac{\sqrt{s_i s_k}}{s_i p_i - s_k p_k} m_{ki} + \mathcal{O}((\epsilon p_{\min})^2) \quad \text{(B.97)}$$

$$= s_i p_i + 2\epsilon p_{\min} \sum_{k \neq i}^{c} \frac{\sqrt{s_i s_k}}{s_i p_i - s_k p_k} \epsilon p_{\min} \sqrt{s_i s_k} + \mathcal{O}((\epsilon p_{\min})^2) \quad \text{(B.98)}$$

$$= s_i p_i + 2(\epsilon p_{\min})^2 \sum_{k \neq i}^{c} \frac{s_i s_k}{s_i p_i - s_k p_k} + \mathcal{O}((\epsilon p_{\min})^2) \quad \text{(B.99)}$$

$$= s_i p_i + \mathcal{O}((\epsilon p_{\min})^2) \quad \text{(B.100)}$$

$$= \gamma_i + \mathcal{O}((\epsilon p_{\min})^2). \quad \text{(B.101)}$$

B.4.2 Step 2: Expected eigenvalues

This section of the appendix computes equation (B.80). We begin this section by recalling the first-order estimate of the expected eigenvalues given by Corollary 2 in terms of the matrix $\boldsymbol{B}^*$.

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_\mu)\right] = \lambda_i(\boldsymbol{B}^*) + \mathcal{O}(\sqrt{\omega_n}), \quad \text{(B.102)}$$

where $\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}$ whose components are given as

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b_{j,l}^{*,(1)} = \begin{cases} \nu_j n \omega_n & j = l \\ 0 & j \neq l \end{cases} \quad \text{(B.103)}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b_{j,l}^{*,(2)} = \nu_i^{-2} \sqrt{\nu_j \nu_l} \sum_{k=1}^{c} \nu_k \sum_{m=1}^{c} \frac{1}{\sqrt{s_m}} \boldsymbol{v}_j(m) \boldsymbol{v}_l(m) \boldsymbol{v}_k(m) \sum_{w=1}^{c} \sqrt{s_w} \boldsymbol{v}_k(w). \quad \text{(B.104)}$$

Observe that the eigenvalues of $\boldsymbol{B}^*$ are determined by the eigenvalues and eigenvectors of the matrix $\boldsymbol{M}$. The results of this subsection are summarized by the following lemma, which shows equation (B.80) and its proof.

Lemma 25 Assume $q = \epsilon p_{\min}$, where $p_{\min} = \min_{i=1,\ldots,c} \boldsymbol{p}$ and $\epsilon \ll 1$ then

$$\left| \frac{\mathbb{E}\left[\lambda_i(\boldsymbol{A})\right]}{n\omega_n s_i} - p_i \right| = \mathcal{O}(\epsilon^2 p_{\min}) + \mathcal{O}\left(\frac{t_i(\boldsymbol{p})}{n\omega_n}\right). \tag{B.105}$$

The proof of the above lemma takes the following intermediate steps. First, we show that

$$\lambda_i(\boldsymbol{B}^*) = \lambda_i(\boldsymbol{B}^{*,(1)}) + \mathcal{O}(t_i(\boldsymbol{p})) \tag{B.106}$$

using Weyl-Lidskii's theorem. Next, we show that

$$\lambda_i(\boldsymbol{B}^{*,(1)}) = n\omega_n \left(s_i p_i + \mathcal{O}(\epsilon^2 p_{\min}^2)\right). \tag{B.107}$$

We conclude the proof using the result that

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_\mu)\right] = \lambda_i(\boldsymbol{B}^*) + \mathcal{O}(\sqrt{\omega_n}). \tag{B.108}$$

We now begin with the proof.

Proof of Lemma 25 As stated, we first show that by Weyl-Lidskii's theorem,

$$\lambda_i(\boldsymbol{B}^*) = \lambda_i(\boldsymbol{B}^{*,(1)}) + \mathcal{O}(t_i(\boldsymbol{p})). \tag{B.109}$$

For the theorem, we take the following quantities.

$$\boldsymbol{A} = \boldsymbol{B}^{*,(2)} \tag{B.110}$$

$$\boldsymbol{H} = \boldsymbol{B}^{*,(1)}. \tag{B.111}$$

Then, $\boldsymbol{B}^* = \boldsymbol{A} + \boldsymbol{H}$. Furthermore, $\boldsymbol{H}$ is self-adjoint because it is diagonal and $\boldsymbol{A}$ is bounded. Since all the eigenvalues are real, we have the following result:

$$|\lambda_i(\boldsymbol{B}^*) - \lambda_i(\boldsymbol{B}^{*,(1)})| \leq ||\boldsymbol{B}^{*,(2)}||. \tag{B.112}$$

Observe that because $\boldsymbol{B}^{*,(2)}$ is independent of n, we may conclude that

$$|\lambda_i(\boldsymbol{B}^*) - \lambda_i(\boldsymbol{B}^{*,(1)})| = \mathcal{O}(t_i(\boldsymbol{p})), \tag{B.113}$$

where $t_i(\boldsymbol{p})$ is a bounded function of the parameters independent of n. This accomplishes the first step of the proof. Next we show that

$$\lambda_i(\boldsymbol{B}^{*,(1)}) = n\omega_n \left(s_i p_i + \mathcal{O}(\epsilon^2 p_{\min}^2)\right). \tag{B.114}$$

This is nearly trivially true. Observe that because $\boldsymbol{B}^{*,(1)}$ is diagonal, the i-th eigenvalue is given as follows,

$$\lambda_i(\boldsymbol{B}^{*,(1)}) = b_{i,i}^{*,(1)} = \nu_i n\omega_n. \tag{B.115}$$

Because of equation (B.101), we have

$$\nu_i n\omega_n = n\omega_n \left(s_i p_i + \mathcal{O}(\epsilon^2 p_{\min}^2)\right). \tag{B.116}$$

To conclude, we have the following set of equalities:

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A})\right] = \lambda_i(\boldsymbol{B}^*) + \mathcal{O}(\sqrt{\omega_n}). \tag{B.117}$$

Equation (B.113) shows that

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A})\right] = \lambda_i(\boldsymbol{B}^{*,(1)}) + \mathcal{O}(t_i(\boldsymbol{p})) \tag{B.118}$$

Replacing $\lambda_i(\boldsymbol{B}^{*,(1)})$ with $n\omega_n \left(s_i p_i + \mathcal{O}(\epsilon^2 p_{\min}^2)\right)$ yields

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A})\right] = n\omega_n \left(s_i p_i + \mathcal{O}(\epsilon^2 p_{\min}^2)\right) + \mathcal{O}(t_i(\boldsymbol{p})) \tag{B.119}$$

and finally, dividing by $n\omega_n s_i$ shows that

$$\frac{\mathbb{E}\left[\lambda_i(\boldsymbol{A})\right]}{n\omega_n s_i} = p_i + \mathcal{O}(\frac{\epsilon^2 p_{\min}^2}{s_i}) + \mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n}) \tag{B.120}$$

Because s_i is a constant with respect to n and ϵ, we simplify this expression as

$$\frac{\mathbb{E}\left[\lambda_i(\boldsymbol{A})\right]}{n\omega_n s_i} = p_i + \mathcal{O}(\epsilon^2 p_{\min}^2) + \mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n}). \tag{B.121}$$

We conclude the proof with the following

$$\left|\frac{\mathbb{E}\left[\lambda_i(\boldsymbol{A})\right]}{n\omega_n s_i} - p_i\right| = \mathcal{O}(\epsilon^2 p_{\min}^2) + \mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n}). \tag{B.122}$$

It is worth noting that the function $t_i(\boldsymbol{p})$ is typically suppressed when reporting the error. In this instance, we make this term explicit because the eventual distribution on the parameters, J, impacts this error term for the random parameter stochastic block model.

The next subsection shows similar calculations except for the term $\text{Cov}(Z_i, Z_j)$ as defined in Corollary 2.

B.4.3 Step 3: Covariance

This section of the appendix is summarized by the following lemma, which shows equation (B.81),

Lemma 26 Assume $q = \epsilon p_{\min}$, where $p_{\min} = \min_{i=1,\dots,c} \boldsymbol{p}$ and $\epsilon \ll 1$ then

$$\left| \text{Cov}(Z_k, Z_l) - \begin{cases} 2p_k & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases} \right| = \mathcal{O}(\epsilon^2 p_{\min}^2), \tag{B.123}$$

where Z is defined in Theorem 8.

Proof of Lemma 26 The proof is split into two subsections. First, we analyze the case of $k \neq l$ and show that all contributions to covariance occur at the second order. Then, we consider the behavior when $k = l$. To begin, we recall a few quantities,

$$\text{Cov}(Z_k, Z_l) = 2\left(\boldsymbol{v}_k. * \boldsymbol{v}_l\right)^T \boldsymbol{M}_f \left(\boldsymbol{v}_k. * \boldsymbol{v}_l\right). \tag{B.124}$$

Recall matrix $\boldsymbol{M}_f$,

$$\boldsymbol{M}_f = \begin{bmatrix} p_1 & \epsilon p_{\min} & \dots & \epsilon p_{\min} \\ \epsilon p_{\min} & p_2 & \dots & \epsilon p_{\min} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon p_{\min} & \epsilon p_{\min} & \dots & p_c \end{bmatrix} \tag{B.125}$$

though it is important to remember that $\boldsymbol{v}_k$ is an eigenvector of $\boldsymbol{M}$ and not $\boldsymbol{M}_f$.

We first compute $\boldsymbol{v}_k. * \boldsymbol{v}_l$ to first order for any choice of k and l.

$$\boldsymbol{v}_k. * \boldsymbol{v}_l = \left(\boldsymbol{e}_k + \sum_{j \neq k}^{c} \frac{\epsilon p_{\min} \sqrt{s_k s_j}}{s_k p_k - s_j p_j} \boldsymbol{e}_j \right) . * \left(\boldsymbol{e}_l + \sum_{r \neq l}^{c} \frac{\epsilon p_{\min} \sqrt{s_l s_r}}{s_l p_l - s_r p_r} \boldsymbol{e}_r \right) + \mathcal{O}((\epsilon p_{\min})^2) \tag{B.126}$$

$$= \boldsymbol{e}_k. * \boldsymbol{e}_l + \sum_{j \neq k}^{c} \frac{\epsilon p_{\min} \sqrt{s_k s_j}}{s_k p_k - s_j p_j} \boldsymbol{e}_j. * \boldsymbol{e}_l + \sum_{r \neq l}^{c} \frac{\epsilon p_{\min} \sqrt{s_l s_r}}{s_l p_l - s_r p_r} \boldsymbol{e}_k. * \boldsymbol{e}_r + \mathcal{O}((\epsilon p_{\min})^2). \tag{B.127}$$

We now show that if $k \neq l$ then the contribution to covariance is only on the order of $\mathcal{O}((\epsilon p_{\min})^2)$. To see this, observe that if $k \neq l$,

$$\boldsymbol{v}_k. * \boldsymbol{v}_l = \sum_{j \neq k}^{c} \frac{\epsilon p_{\min} \sqrt{s_k s_j}}{s_k p_k - s_j p_j} \boldsymbol{e}_j. * \boldsymbol{e}_l + \sum_{r \neq l}^{c} \frac{\epsilon p_{\min} \sqrt{s_l s_r}}{s_l p_l - s_r p_r} \boldsymbol{e}_k. * \boldsymbol{e}_r + \mathcal{O}((\epsilon p_{\min})^2). \tag{B.128}$$

Therefore, the vector

$$\boldsymbol{v}_k. * \boldsymbol{v}_l = \epsilon p_{\min} \boldsymbol{d}^{k,l} + \mathcal{O}((\epsilon p_{\min})^2). \tag{B.129}$$

where $\boldsymbol{d}^{k,l}$ is some vector that depends on k, l. Computing $\mathrm{Cov}(Z_k, Z_l)$ we see

$$\mathrm{Cov}(Z_k, Z_l) = 2 \left(\boldsymbol{v}_k. * \boldsymbol{v}_l \right)^T \boldsymbol{M}_f \left(\boldsymbol{v}_k. * \boldsymbol{v}_l \right) \tag{B.130}$$

$$= 2\epsilon p_{\min} (\boldsymbol{d}^{k,l})^T \boldsymbol{M}_f \left(\epsilon p_{\min} \boldsymbol{d}^{k,l} \right) + \mathcal{O}((\epsilon p_{\min})^2) \tag{B.131}$$

$$= 2(\epsilon p_{\min})^2 (\boldsymbol{d}^{k,l})^T \boldsymbol{M}_f \boldsymbol{d}^{k,l} + \mathcal{O}((\epsilon p_{\min})^2) \tag{B.132}$$

$$= \mathcal{O}((\epsilon p_{\min})^2). \tag{B.133}$$

Next, we show the behavior of $\mathrm{Cov}(Z_k, Z_l)$ for $k = l$. We have

$$\boldsymbol{v}_k. * \boldsymbol{v}_k = \boldsymbol{e}_k. * \boldsymbol{e}_k + \sum_{j \neq k}^{c} \frac{\epsilon p_{\min} \sqrt{s_k s_j}}{s_k p_k - s_j p_j} \boldsymbol{e}_j. * \boldsymbol{e}_k + \sum_{j \neq k}^{c} \frac{\epsilon p_{\min} \sqrt{s_k s_j}}{s_k p_k - s_j p_j} \boldsymbol{e}_k. * \boldsymbol{e}_j + \mathcal{O}((\epsilon p_{\min})^2). \tag{B.134}$$

However, if $j \neq k$, then $\boldsymbol{e}_j. * \boldsymbol{e}_k = \boldsymbol{0}$. Therefore,

$$\boldsymbol{v}_k. * \boldsymbol{v}_k = \boldsymbol{e}_k. * \boldsymbol{e}_k + \mathcal{O}((\epsilon p_{\min})^2) \tag{B.135}$$

$$= \boldsymbol{e}_k + \mathcal{O}((\epsilon p_{\min})^2) \tag{B.136}$$

because $\boldsymbol{e}_k$ is the canonical basis vector. To compute the variance, we need only to compute

$$\mathrm{Cov}(Z_k, Z_k) = 2 \left(\boldsymbol{v}_k. * \boldsymbol{v}_k \right)^T \boldsymbol{M}_f \left(\boldsymbol{v}_k. * \boldsymbol{v}_k \right) \tag{B.137}$$

$$= 2\boldsymbol{e}_k^T \boldsymbol{M}_f \boldsymbol{e}_k + \mathcal{O}((\epsilon p_{\min})^2) \tag{B.138}$$

$$= 2p_k + \mathcal{O}((\epsilon p_{\min})^2). \tag{B.139}$$

The combination of equations (B.133) and (B.139) yields

$$\left| \mathrm{Cov}(Z_k, Z_l) - \begin{cases} 2p_k & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases} \right| = \mathcal{O}(\epsilon^2 p_{\min}^2). \tag{B.140}$$

B.5 Proof of Lemma 3

This appendix proves Lemma 3, which is restated below for convenience.

Lemma 27 (Lemma 3 from the main chapter) Let $\boldsymbol{P}_j$ be an observation from J with components P_i. Let $P_{\min} = \min_{i=1,\ldots,c} P_i$ and define $q = \epsilon P_{\min}$ where $\epsilon \ll 1$.

$$\frac{\mathbb{E}_{H_n}[\lambda_i]}{n\omega_n s_i} = \mathbb{E}_J[P_i] + \mathcal{O}(\epsilon^2) + \mathcal{O}\left(\frac{1}{n\omega_n}\right) \tag{B.141}$$

$$\frac{\mathrm{Cov}_{H_n}(\lambda_i, \lambda_j)}{n^2\omega_n^2 s_i s_j} + \begin{cases} \frac{2\mathbb{E}_{H_n}[\lambda_i]}{n^3\omega_n^2 s_i^3} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{B.142}$$

$$= \mathrm{Cov}_J\left(P_i + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}\left(\frac{t_i(\boldsymbol{p})}{n\omega_n}\right), P_j + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}\left(\frac{t_j(\boldsymbol{p})}{n\omega_n}\right)\right) + \mathcal{O}\left(\frac{\epsilon^2}{n^2\omega_n}\right) \tag{B.143}$$

where $t_i(\boldsymbol{p})$ is a bounded function of the parameters for each i.

Proof of Lemma 27 The proof is a consequence of Propositions 1 and 4 which are direct consequences of Lemma 2.

The proof for the above lemma is given in four parts.

(1) First we show in Proposition 1 that

$$\frac{\mathbb{E}_{H_n}[\lambda_i]}{n\omega_n s_i} = \mathbb{E}_J[P_i] + \mathcal{O}(\epsilon^2) + \mathcal{O}(\frac{1}{n\omega_n}). \tag{B.144}$$

(2) We then turn our attention to the covariance terms. In Proposition 2, we show that

$$\mathbb{E}_J[\mathrm{Cov}(Z_i, Z_j)] = \begin{cases} \mathbb{E}_J[2P_i] + \mathcal{O}(\epsilon^2) & \text{if } i = j \\ \mathcal{O}(\epsilon^2) & \text{if } i \neq j \end{cases}. \tag{B.145}$$

(3) Next, we show in Proposition 3 that

$$\frac{\mathrm{Cov}_J\left(\mathbb{E}_\mu[\lambda_i(\boldsymbol{A})]\,,\mathbb{E}_\mu[\lambda_j(\boldsymbol{A})]\,|\boldsymbol{P}_J=\boldsymbol{p}\right)}{n^2\omega_n^2 s_i s_j}=\mathrm{Cov}_J\left(P_i+\mathcal{O}(\epsilon^2 P_{\min})+\mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n}),P_j+\mathcal{O}(\epsilon^2 P_{\min})+\mathcal{O}(\frac{t_j(\boldsymbol{p})}{n\omega_n})\right). \tag{B.146}$$

(4) After recalling the definition of the first moment of H_n, we have in Proposition 4 that

$$\frac{\mathrm{Cov}_{H_n}(\lambda_i,\lambda_j)}{n^2\omega_n^2 s_i s_j}+\begin{cases}\frac{2\mathbb{E}_{H_n}[\lambda_i]}{n^3\omega_n^2 s_i^3} & \text{if } i=j\\ 0 & \text{if } i\neq j\end{cases} \tag{B.147}$$

$$=\mathrm{Cov}_J\left(P_i+\mathcal{O}(\epsilon^2 P_{\min}^2)+\mathcal{O}\left(\frac{t_i(\boldsymbol{p})}{n\omega_n}\right),P_j+\mathcal{O}(\epsilon^2 P_{\min}^2)+\mathcal{O}\left(\frac{t_j(\boldsymbol{p})}{n\omega_n}\right)\right)+\mathcal{O}\left(\frac{\epsilon^2}{n^2\omega_n}\right) \tag{B.148}$$

where $t_i(\boldsymbol{p})$ is a bounded function of the parameters for each i.

We conclude the proof by combining the results of Step 1 and Step 4.

Step 1:

The following proposition characterizes this step.

Proposition 1

$$\frac{\mathbb{E}_{H_n}[\lambda_i]}{n\omega_n s_i}=\mathbb{E}_J[P_i]+\mathcal{O}(\epsilon^2)+\mathcal{O}(\frac{1}{n\omega_n}) \tag{B.149}$$

Proof of Proposition 1 The proof is a direct consequence of Lemma 2. First, we recall the definition of $\mathbb{E}_{H_n}[\lambda_i]$,

$$\mathbb{E}_{H_n}[\lambda_i]=\mathbb{E}_J[\mathbb{E}_\mu[\lambda_i(\boldsymbol{A})|\boldsymbol{P}_J=\boldsymbol{p}]]\,. \tag{B.150}$$

Lemma 2 shows that

$$\mathbb{E}_\mu[\lambda_i(\boldsymbol{A})|\boldsymbol{P}_J=\boldsymbol{p}]=p_i+\mathcal{O}(\epsilon^2 p_{\min}^2)+\mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n s_i}). \tag{B.151}$$

Substitution yields

$$\mathbb{E}_{H_n}[\lambda_i]=\mathbb{E}_J\left[P_i+\mathcal{O}(\epsilon^2 P_{\min}^2)+\mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n s_i})\right], \tag{B.152}$$

where P_i and $P_{\min}$ denote that these are now random variables. Now, because the support of J is a subset of $[0,1]^c$ and $t_i(\boldsymbol{p})$ is a bounded function of the parameters, we conclude

$$\mathbb{E}_{H_n}[\lambda_i] = \mathbb{E}_J[P_i] + \mathcal{O}(\epsilon^2) + \mathcal{O}(\frac{1}{n\omega_n s_i}). \tag{B.153}$$

Step 2: This step considers $\mathbb{E}_J[\mathrm{Cov}(Z_i, Z_j)]$. It also follows as a result of Lemma 2

Proposition 2 Let Z be as defined in Corollary 2; then,

$$\mathbb{E}_J[\mathrm{Cov}(Z_i, Z_j)] = \begin{cases} \mathbb{E}_J[2P_i] + \mathcal{O}(\epsilon^2) & \text{if } i = j \\ \mathcal{O}(\epsilon^2) & \text{if } i \neq j \end{cases}. \tag{B.154}$$

Proof of Proposition 2 Lemma 2 shows that

$$\mathrm{Cov}(Z_i, Z_j) = \begin{cases} 2p_i + \mathcal{O}(\epsilon^2 p_{\min}^2) & \text{if } i = j \\ \mathcal{O}(\epsilon^2 p_{\min}^2) & \text{if } i \neq j \end{cases}. \tag{B.155}$$

Taking an expectation over J,

$$\mathbb{E}_J[\mathrm{Cov}(Z_i, Z_j)] = \begin{cases} 2\mathbb{E}_J\big[P_i + \mathcal{O}(\epsilon^2 P_{\min}^2)\big] & \text{if } i = j \\ \mathbb{E}_J\big[\mathcal{O}(\epsilon^2 P_{\min}^2)\big] & \text{if } i \neq j \end{cases} \tag{B.156}$$

where P_i and $P_{\min}$ denote random variables. As before, $\mathbb{E}_J\big[\mathcal{O}(\epsilon^2 P_{\min}^2)\big] = \mathcal{O}(\epsilon^2)$ which gives the result

$$\mathbb{E}_J[\mathrm{Cov}(Z_i, Z_j)] = \begin{cases} 2\mathbb{E}_J[P_i] + \mathcal{O}(\epsilon^2) & \text{if } i = j \\ \mathcal{O}(\epsilon^2) & \text{if } i \neq j \end{cases} \tag{B.157}$$

Step 3: This step shows the proper first order scaling of the term $\mathrm{Cov}_J\left(\mathbb{E}_\mu[\lambda_i(\boldsymbol{A})], \mathbb{E}_\mu[\lambda_j(\boldsymbol{A})] \,|\boldsymbol{P}_J = \boldsymbol{p}\right)$. It is another result of Lemma 2.

Proposition 3

$$\frac{\mathrm{Cov}_J\left(\mathbb{E}_\mu[\lambda_i(\boldsymbol{A})], \mathbb{E}_\mu[\lambda_j(\boldsymbol{A})] \,|\boldsymbol{P}_J = \boldsymbol{p}\right)}{n^2\omega_n^2 s_i s_j} = \mathrm{Cov}_J\left(P_i + \mathcal{O}(\epsilon^2 P_{\min}) + \mathcal{O}(\frac{1}{n\omega_n}), P_j + \mathcal{O}(\epsilon^2 P_{\min}) + \mathcal{O}(\frac{1}{n\omega_n})\right). \tag{B.158}$$

Proof of Proposition 3 We have seen that

$$\frac{\mathbb{E}_\mu[\lambda_i(\boldsymbol{A})]}{n\omega_n s_i} = p_i + \mathcal{O}(\epsilon^2 p_{\min}) + \mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n s_i}). \tag{B.159}$$

We use this first-order estimate of the expected eigenvalues to show the result. We begin with

$$\frac{\mathrm{Cov}_J\left(\mathbb{E}_\mu[\lambda_i(\boldsymbol{A})], \mathbb{E}_\mu[\lambda_j(\boldsymbol{A})] \,|\boldsymbol{P}_J = \boldsymbol{p}\right)}{n^2\omega_n^2 s_i s_j} = \mathrm{Cov}_J\left(\frac{\mathbb{E}_\mu[\lambda_i(\boldsymbol{A})]}{n\omega_n s_i}, \frac{\mathbb{E}_\mu[\lambda_j(\boldsymbol{A})]}{n\omega_n s_j}|\boldsymbol{P}_J = \boldsymbol{p}\right). \tag{B.160}$$

We then substitute in equation (B.159) and find

$$\mathrm{Cov}_J\left(\frac{\mathbb{E}_\mu[\lambda_i(\boldsymbol{A})]}{n\omega_n s_i}, \frac{\mathbb{E}_\mu[\lambda_j(\boldsymbol{A})]}{n\omega_n s_j}|\boldsymbol{P}_J = \boldsymbol{p}\right) = \mathrm{Cov}_J\left(P_i + \mathcal{O}(\epsilon^2 P_{\min}) + \mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n s_i}), P_j + \mathcal{O}(\epsilon^2 P_{\min}) + \mathcal{O}(\frac{t_j(\boldsymbol{p})}{n\omega_n s_j})\right) \tag{B.161}$$

which concludes the calculations and the proof.

Step 4: This step puts together the prior three propositions.

Proposition 4 Let $\boldsymbol{P}_J$ be an observation from J with components P_i, let $P_{\min} = \min_{i=1,\ldots,c} P_i$ and define $q = \epsilon P_{\min}$ where $\epsilon \ll 1$. Let Z be as defined in Corollary 2.

$$\frac{\mathrm{Cov}_{H_n}(\lambda_i, \lambda_j)}{n^2\omega_n^2 s_i s_j} + \begin{cases} \frac{2\mathbb{E}_{H_n}[\lambda_i]}{n^3\omega_n^2 s_i^3} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{B.162}$$

$$= \mathrm{Cov}_J\left(P_i + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}\left(\frac{t_i(\boldsymbol{p})}{n\omega_n}\right), P_j + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}\left(\frac{t_j(\boldsymbol{p})}{n\omega_n}\right)\right) + \mathcal{O}\left(\frac{\epsilon^2}{n^2\omega_n}\right) \tag{B.163}$$

where $t_i(\boldsymbol{p})$ is a bounded function of the parameters for each i.

Proof of Proposition 4 We begin with the definition of the scaled covariance term, $\mathrm{Cov}_{H_n}(\frac{\lambda_i}{\sqrt{\omega_n}}, \frac{\lambda_j}{\sqrt{\omega_n}})$.

$$\mathrm{Cov}_{H_n}(\frac{\lambda_i}{\sqrt{\omega_n}}, \frac{\lambda_j}{\sqrt{\omega_n}}) = \mathrm{Cov}_J(\mathbb{E}_\mu\left[\frac{1}{\sqrt{\omega_n}}\lambda_i(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right], \mathbb{E}_\mu\left[\frac{1}{\sqrt{\omega_n}}\lambda_j(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right]) \tag{B.164}$$

$$+ \mathbb{E}_J\left[\mathrm{Cov}_\mu\left(\frac{1}{\sqrt{\omega_n}}\lambda_i(\boldsymbol{A}), \frac{1}{\sqrt{\omega_n}}\lambda_j(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right)\right] \tag{B.165}$$

When n is large we replace $\mathrm{Cov}_\mu\left(\frac{1}{\sqrt{\omega_n}}\lambda_i(\boldsymbol{A}), \frac{1}{\sqrt{\omega_n}}\lambda_j(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right)$ with $\mathrm{Cov}(Z_i, Z_j)$. By Proposition 2, we have

$$\mathrm{Cov}_{H_n}(\frac{\lambda_i}{\sqrt{\omega_n}}, \frac{\lambda_j}{\sqrt{\omega_n}}) = \mathrm{Cov}_J\left(\mathbb{E}_\mu\left[\frac{1}{\sqrt{\omega_n}}\lambda_i(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right], \mathbb{E}_\mu\left[\frac{1}{\sqrt{\omega_n}}\lambda_j(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right]\right) \tag{B.166}$$

$$+ \begin{cases} \mathbb{E}_J[2P_i] + \mathcal{O}(\epsilon^2) & \text{if } i = j \\ \mathcal{O}(\epsilon^2) & \text{if } i \neq j \end{cases}. \tag{B.167}$$

To continue, we divide both sides by $n^2\omega_n s_i s_j$,

$$\frac{\mathrm{Cov}_{H_n}(\frac{\lambda_i}{\sqrt{\omega_n}}, \frac{\lambda_j}{\sqrt{\omega_n}})}{n^2\omega_n s_i s_j} = \frac{\mathrm{Cov}_J(\mathbb{E}_\mu\left[\frac{1}{\sqrt{\omega_n}}\lambda_i(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right], \mathbb{E}_\mu\left[\frac{1}{\sqrt{\omega_n}}\lambda_j(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right])}{n^2\omega_n s_i s_j} \tag{B.168}$$

$$+ \begin{cases} \frac{\mathbb{E}_J[2P_i]+\mathcal{O}(\epsilon^2)}{n^2\omega_n s_i^2} & \text{if } i = j \\ \frac{\mathcal{O}(\epsilon^2)}{n^2\omega_n s_i s_j} & \text{if } i \neq j \end{cases}. \tag{B.169}$$

We rewrite

$$\frac{\mathrm{Cov}_J\left(\mathbb{E}_\mu\left[\frac{\lambda_i(\boldsymbol{A})}{\sqrt{\omega_n}}|\boldsymbol{P}_J = \boldsymbol{p}\right], \mathbb{E}_\mu\left[\frac{\lambda_j(\boldsymbol{A})}{\sqrt{\omega_n}}|\boldsymbol{P}_J = \boldsymbol{p}\right]\right)}{n^2\omega_n s_i s_j} = \mathrm{Cov}_J(\mathbb{E}_\mu\left[\frac{1}{n\omega_n s_i}\lambda_i(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right], \mathbb{E}_\mu\left[\frac{1}{n\omega_n s_i}\lambda_j(\boldsymbol{A})|\boldsymbol{P}_J = \boldsymbol{p}\right]). \tag{B.170}$$

We now use Proposition 3

$$\mathrm{Cov}_J(\mathbb{E}_\mu\left[\frac{\lambda_i(\boldsymbol{A})}{n\omega_n s_i}|\boldsymbol{P}_J = \boldsymbol{p}\right], \mathbb{E}_\mu\left[\frac{\lambda_j(\boldsymbol{A})}{n\omega_n s_j}|\boldsymbol{P}_J = \boldsymbol{p}\right]) = \mathrm{Cov}_J\left(P_i + \mathcal{O}(\epsilon^2 P_{\min}) + \mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n}), P_j + \mathcal{O}(\epsilon^2 P_{\min}) + \mathcal{O}(\frac{t_j(\boldsymbol{p})}{n\omega_n})\right) \tag{B.171}$$

Substituting this in to equation (B.169) and factoring out $\frac{1}{\sqrt{\omega_n}}$ on the left hand side,

$$\frac{\mathrm{Cov}_{H_n}(\lambda_i, \lambda_j)}{n^2\omega_n^2 s_i s_j} = \mathrm{Cov}_J\left(P_i + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}(\frac{1}{n\omega_n}), P_j + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}(\frac{1}{n\omega_n})\right) \tag{B.172}$$

$$+ \begin{cases} \frac{1}{n^2\omega_n s_i^2}\mathbb{E}_J[2P_i] + \mathcal{O}(\frac{\epsilon^2}{n^2\omega_n}) & \text{if } i = j \\ \mathcal{O}(\frac{\epsilon^2}{n^2\omega_n}) & \text{if } i \neq j \end{cases}. \tag{B.173}$$

Recall from Step 1 that

$$\frac{\mathbb{E}_{H_n}[\lambda_i]}{n\omega_n s_i} = \mathbb{E}_J[P_i] + \mathcal{O}(\epsilon^2) + \mathcal{O}(\frac{1}{n\omega_n}). \tag{B.174}$$

Solving this for $\mathbb{E}_J[P_i]$, we have

$$\mathbb{E}_J[P_i] = \frac{\mathbb{E}_{H_n}[\lambda_i]}{n\omega_n s_i} + \mathcal{O}(\epsilon^2) + \mathcal{O}(\frac{1}{n\omega_n}). \tag{B.175}$$

Substituting this expression in to (B.173),

$$\frac{\mathrm{Cov}_{H_n}(\lambda_i, \lambda_j)}{n^2\omega_n^2 s_i s_j} = \mathrm{Cov}_J\left(P_i + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n}), P_j + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}(\frac{t_j(\boldsymbol{p})}{n\omega_n})\right) \tag{B.176}$$

$$+ \begin{cases} \frac{2}{n^2\omega_n s_i^2}\left(\frac{\mathbb{E}_{H_n}[\lambda_i]}{n\omega_n s_i} + \mathcal{O}(\epsilon^2) + \mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n})\right) + \mathcal{O}(\frac{\epsilon^2}{n^2\omega_n}) & \text{if } i = j \\ \mathcal{O}(\frac{\epsilon^2}{n^2\omega_n}) & \text{if } i \neq j \end{cases}. \tag{B.177}$$

Simplifying and collecting all error terms outside the covariance, we have

$$\frac{\mathrm{Cov}_{H_n}(\lambda_i, \lambda_j)}{n^2\omega_n^2 s_i s_j} = \mathrm{Cov}_J\left(P_i + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n}), P_j + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}(\frac{t_j(\boldsymbol{p})}{n\omega_n})\right) \tag{B.178}$$

$$+ \begin{cases} \frac{2\mathbb{E}_{H_n}[\lambda_i]}{n^3\omega_n^2 s_i^3} + \mathcal{O}(\frac{\epsilon^2}{n^2\omega_n}) & \text{if } i = j \\ \mathcal{O}(\frac{\epsilon^2}{n^2\omega_n}) & \text{if } i \neq j \end{cases}. \tag{B.179}$$

And finally, solving for the second moment of J,

$$\frac{\mathrm{Cov}_{H_n}(\lambda_i, \lambda_j)}{n^2\omega_n^2 s_i s_j} - \begin{cases} \frac{2\mathbb{E}_{H_n}[\lambda_i]}{n^3\omega_n^2 s_i^3} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{B.180}$$

$$= \mathrm{Cov}_J\left(P_i + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}(\frac{t_i(\boldsymbol{p})}{n\omega_n}), P_j + \mathcal{O}(\epsilon^2 P_{\min}^2) + \mathcal{O}(\frac{t_j(\boldsymbol{p})}{n\omega_n})\right) + \mathcal{O}(\frac{\epsilon^2}{n^2\omega_n}), \tag{B.181}$$

which concludes the proof.