Information and Inference: A Journal of the IMA (2023) **00**, 1–58
https://doi.org/10.1093/imaiai/iaad002

Theoretical analysis and computation of the sample Fréchet mean of sets of large graphs for various metrics

Daniel Ferguson† and François G. Meyer
Department of Applied Mathematics, University of Colorado, Boulder, Boulder, CO, USA
†Corresponding author. Email: dafe0926@colorado.edu

[Received on 15 March 2022; revised on 24 September 2022; accepted on 17 January 2023]

Abstract

To characterize the location (mean, median) of a set of graphs, one needs a notion of centrality that has been adapted to metric spaces. A standard approach is to consider the Fréchet mean. In practice, computing the Fréchet mean for sets of large graphs presents many computational issues. In this work, we suggest a method that may be used to compute the Fréchet mean for sets of graphs which is metric independent. We show that the technique proposed can be used to determine the Fréchet mean when considering the Hamming distance or a distance defined by the difference between the spectra of the adjacency matrices of the graphs.

Keywords: graph mean; Fréchet mean; statistical network analysis.

AMS Subject Classification: 05C75; 05C80.

1. Introduction

Machine learning algorithms almost always require the estimation of the average of a dataset. Algorithms such as clustering, classification and linear regression all utilize this average value [39]. When notions of addition and multiplication can be defined, the mean is a simple algebraic operation, however, for data in a metric space, an extension of the notion of mean, termed the Fréchet mean, was introduced in [33], which involves an optimization procedure defined over the metric space.

The space of graphs is only one example of a metric space where the notion of Fréchet mean has become a commonplace replacement for the notion of centrality. However, the Fréchet mean for the space of graphs remains difficult to compute and is instead usually approximated by taking the most central element of a given data set (see e.g. [56]). A general critique of the few algorithms that determine the Fréchet mean graph is their inability to be generalized to other choices of metrics. In this work, we introduce a solution technique that can be used to determine the Fréchet mean for the space of graphs, which can be implemented for any choice of metric.

Throughout this work, we consider a set of simple graphs with n vertices. For some of our results, we note that the vertex set must be sufficiently large.

Our approach to determining a solution to the Fréchet mean problem involves the following steps. First, we lift the problem to the space of probability measures and search for a measure with the correct

Fréchet mean. Second, we restrict the space of probability measures to a parametrizable subset and search for the best measure within that subset.

We show that this procedure can be used to determine the Fréchet mean when the metric equipped to the space of graphs is the Hamming distance, which is the most common metric for graph-valued data, and a distance defined by the ℓ_2 difference between the eigenvalues of the adjacency matrices of two graphs. Notably, the high-level steps taken to determine the Fréchet mean graphs for each metric are nearly identical.

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 mean or median of a set of graphs a standard approach is to consider the Fréchet mean. The choice of metric is crucial for the Fréchet mean as different metrics induce a different Fréchet mean set of graphs.

The Fréchet mean has been studied at length when the distance is the Hamming distance (e.g. [19, 36, 40, 41, 43, 58] and references therein). The Hamming distance reflects small-scale changes in the graphs defined by the local connectivity at each the level of each vertex. The average of the local structures in the networks provides useful information when the specific location of an edge in the adjacency matrix is important to the research question at hand. However, because the Hamming distance is defined by the local connectivity, the Fréchet mean graph with respect to this distance need not preserve the observed global properties of graphs in the data set.

A different metric, one which captures the larger scale patterns of connectivity in graphs (e.g. community structure [3, 51], modularity [38]), may also be considered when determining the Fréchet mean. The adjacency spectral distance, which we define as the ℓ_2 norm of the difference between the spectra of the adjacency matrices of two graphs [80], is one such metric that quantifies differences in global connectivity patterns.

The eigenvalues of the adjacency matrix carry important topological information about a graph at a multitude of scales. The largest eigenvalues reveal information about the large-scale features of a network such as community structure [51] and the existence of 'hubs' [32]. While the smaller eigenvalues reveal information about the local features in the graph. Features such as the degree of a vertex or the ubiquity of triangular substructures in the network are examples of features quantified by the smaller eigenvalues of the adjacency matrix [26].

Oftentimes, only the largest eigenvalues of the adjacency matrix are considered when comparing graphs. By choosing to consider only the largest c eigenvalues for some choice of $c < n$, the difference in the global features of the two graphs can be highlighted while the local features of the graphs are ignored.

A primary difficulty when determining the Fréchet mean with respect to the Hamming distance is to have node correspondence between the graphs in the data set. When the graphs do not have node correspondence, a lengthy minimization procedure must be solved to align the graphs before any work can be done to determine the Fréchet mean graph [19]. When the graphs have node correspondence, the solution can be determined analytically, as shown in [58]. One benefit to metrics defined on the spectra of various matrix representations of a graph, in contrast to the Hamming distance, is that they naturally allow for the comparison between graphs defined on different vertex sets.

Further studies on the mean graph appear in [31] where the authors suggest embedding the graphs into Euclidean space where the Fréchet mean can be computed trivially using the algebraic structure of the space. Of note, however, is that the inverse of the embedding may not be available in closed form making it difficult to recover the Fréchet mean graph. The authors in [37] suggest characterizing the mean for weighted graphs by leveraging the information captured in the Laplacian matrices of the graphs and the mean is computed on the manifold defined by the cone of symmetric positive semi-definite matrices.

3. Main contributions

The Fréchet mean (or median) graph has become a standard tool for analyzing graph-valued data. In this paper, we present a solution technique to determine the Fréchet mean (or median) graph, which may be applied for any choice of metric. We showcase the technique by recovering known results when the metric is the Hamming distance (the most widely studied distance when considering the Fréchet mean for graph-valued data) and subsequently show that the same technique can be used when the metric is defined by the difference between the largest spectral values of the respective adjacency matrices of the two graphs.

We provide novel theoretical results about the Fréchet mean graph when the metric is defined as the difference between the spectral values of the adjacency matrices of the graphs. The proofs rely on a combination of ideas: first, the Fréchet mean graph from a stochastic block model ensembles provides a universal approximant in the spectral adjacency pseudo-metric (see Definition 4.7 and Theorem 8.1). Second, the dominant eigenvalues of the adjacency matrices of the sample set of graphs can be used to infer the parameters of a stochastic block model ensemble whose Fréchet mean approximately minimizes the objective value that determines the sample Fréchet mean graph of interest.

We provide experimental results when considering the metric defined on the spectra of the adjacency matrices. We also provide an application of the sample Fréchet mean graph given this metric in the algorithm for K-means clustering. For experiments and use cases of the Fréchet mean with respect to the Hamming distance, see e.g. [19, 36, 40, 41, 43, 58].

In a short and early version of this work [28], we described, without proof, an algorithm to compute an approximation to the Fréchet mean when considering a metric based on the spectral values of the adjacency matrix. This earlier work relied on some theoretical results that were made precise, in their full general context, only recently. These general results are presented in this version of the manuscript in sections 6, 7 and 8.

In comparison to the algorithms presented in the conference paper, the analysis of the optimization problem associated with the computation of the Fréchet mean has been significantly generalized. The optimization problem is recast as a search over a space of probability distributions. This novel and more general perspective can be used to tackle other problems where the solution to an optimization problem is replaced by a set of solutions that can be generated by sampling a distribution. The approach is not limited to graph-valued problems but works in any probability metric space. Conversely, the initial conference paper focused on the algorithmic aspects of the computation of the Fréchet mean with respect to the metric defined by comparing the largest eigenvalues of the adjacency matrices of two graphs, omitting the abstract and more general approach presented in the current manuscript.

Finally, the current manuscript contains a novel and important application of the computation of the Frechet mean to machine learning: the clustering of graph-valued data. For our experiments, we always focus on the metric defined on the spectra of the adjacency matrices rather than the Hamming distance. For experiments and applications of the Fréchet mean with respect to the Hamming distance, we would refer the reader to any of the aforementioned studies.

4. Notations

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

$$\boldsymbol{A}_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E. \\ 0 & \text{else.} \end{cases} \tag{4.1}$$

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:

$$\sigma : \mathbb{M}_{n\times n} \longrightarrow \mathbb{R}^n, \tag{4.2}$$

$$\boldsymbol{A} \longmapsto \boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n], \tag{4.3}$$

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:

$$\sigma_c : \mathbb{M}_{n\times n} \longrightarrow \mathbb{R}^c, \tag{4.4}$$

$$\boldsymbol{A} \longmapsto \boldsymbol{\lambda}_c = [\lambda_1, \ldots, \lambda_c]. \tag{4.5}$$

DEFINITION 4.1. We define the adjacency spectral pseudometric as the ℓ_2 norm between the spectra of the respective adjacency matrices,

$$d_A(G, G') = ||\sigma(\boldsymbol{A}) - \sigma(\boldsymbol{A}')||_2. \tag{4.6}$$

The pseudometric d_A satisfies the symmetry and triangle inequality axioms but not the identity axiom. Instead, d_A satisfies the reflexivity axiom

$$d_A(G, G) = 0, \quad \forall G \in \mathcal{G}.$$

When the adjacency matrices (or Laplacian) of graphs have similar spectra, it can be shown that the graphs have similar global and structural properties [79]. As a natural extension of this spectral metric, sometimes only the largest c eigenvalues are measured where $c \ll n$. We refer to this next metric as a truncation of the adjacency spectral pseudometric.

DEFINITION 4.2. We define the truncated adjacency spectral pseudometric as the ℓ_2 norm between the largest c eigenvalues of the respective adjacency matrices,

$$d_{A_c}(G, G') = ||\sigma_c(\boldsymbol{A}) - \sigma_c(\boldsymbol{A}')||_2. \tag{4.7}$$

Rather than considering the difference between the spectra of the adjacency matrices, the Hamming distance quantifies discrepancies between the existence of an edge at the same location in the adjacency matrix.

DEFINITION 4.3. We define the Hamming distance as follows:

$$d_H(G, G') = \sum_{1 \le i < j \le n} |a_{ij} - a'_{ij}|. \tag{4.8}$$

DEFINITION 4.4. We denote by $\mathscr{G}$ the set of all simple unweighted graphs on n nodes.

4.1 *Random graphs*

We denote by $\mathscr{M}(\mathscr{G})$ the space of probability measures on $\mathscr{G}$. In this work, we always mean a probability measure when discussing a measure.

DEFINITION 4.5. We define the set of random graphs distributed according to μ as the probability space $(\mathscr{G}, \mu)$.

REMARK 4.1. In this paper, the σ-field associated with the $(\mathscr{G}, \mu)$ will always be the power set of $\mathscr{G}$.

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.

4.1.1 *Kernel probability measures.*

Here we define an important class of probability measures for our study.

DEFINITION 4.6. Let ω_n be a positive constant and let f be a function such that

$$f : [0,1] \times [0,1] \mapsto [0,1], \tag{4.9}$$

with $f(x,y) = f(1-y, 1-x)$. The product, $\omega_n f$ defines the kernel of a kernel probability measure, denoted $\mu_{\omega_n f} \in \mathscr{M}(\mathscr{G})$, where

$$\forall G \in \mathscr{G}, \text{with adjacency matrix } \boldsymbol{A}, \tag{4.10}$$

$$\mu_{\omega_n f}(\{\boldsymbol{A}\}) = \prod_{1 \le i < j \le n} \left(\omega_n f\left(\frac{i}{n}, \frac{j}{n}\right)\right)^{a_{ij}} \left(1 - \omega_n f\left(\frac{i}{n}, \frac{j}{n}\right)\right)^{1-a_{ij}}. \tag{4.11}$$

REMARK 4.2. 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 that $||f||_1 = 1$ then the constant ω_n denotes the expected density of graphs sampled according to $\mu_{\omega_n f}$.

We 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 $\{\frac{i}{n}\}_{i=1}^n \times \{\frac{j}{n}\}_{j=1}^n$, which identify an equivalent kernel probability measure $\mu_{\omega_n f}$. Throughout our analysis, we always refer to a piecewise Lipschitz representative function f.

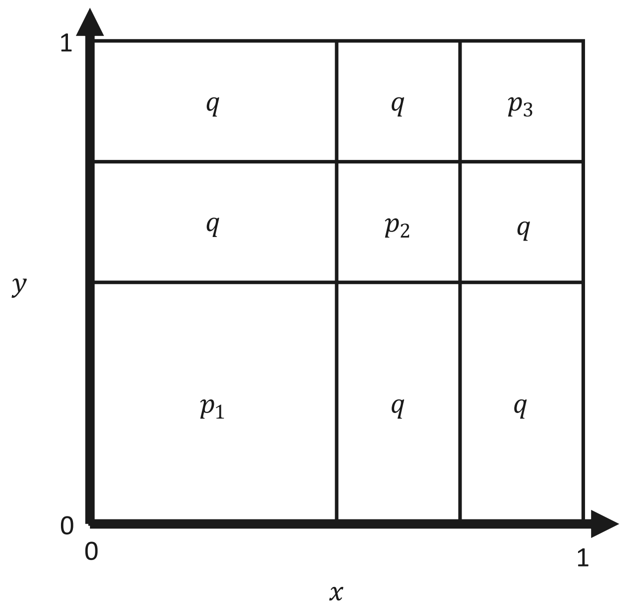

FIG. 1. Example stochastic block model kernel $f(x, y; \boldsymbol{p}, q, \boldsymbol{s})$.

The definition of the kernel $\mu_{\omega_n f}(\{\boldsymbol{A}\})$ in Definition 4.6 is similar to the definition of *graphon* found in [17, 42, 55], or the concept of W-random graphs generated from a graphon W. We follow the approach of [13] and introduce a 'global density' ω_n that quantifies the uniform graph sparsity; see also [1, 14, 16, 46, 47, 81] where the graphon is scaled by a 'target density' function ρ_n.

Our graph model (see Def. 4.6), does not require the introduction of latent variables ξ_i, ξ_j that encode the node location within the domain $[0, 1]^2$. Indeed, we can directly work with the *canonical* representation of the graphon, since the distance that we use to compare graphs, $d_A(G, G')$ (see equation (4.6)), is invariant under the action of any joint (deterministic or stochastic) permutations of the rows and columns of the adjacency matrix.

DEFINITION 4.7. We denote by G_μ a random realization of a graph $G \in (\mathscr{G}, \mu)$.

4.1.2 *Stochastic block models.* The stochastic block model (see [3]) 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.

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.

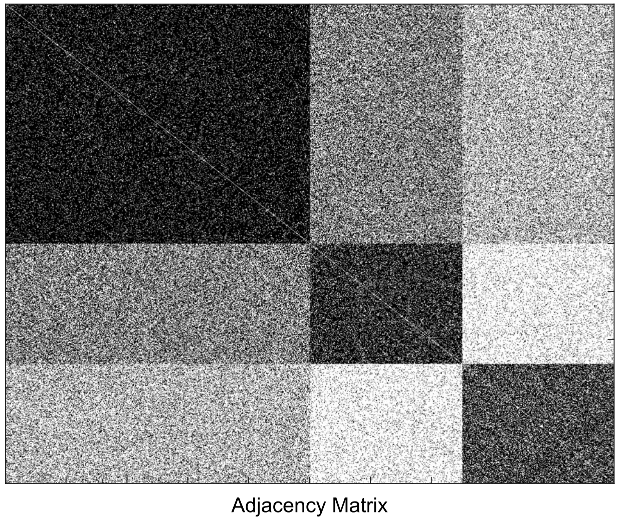

FIG. 2. Example adjacency matrix.

Finally, we denote by q the across-community edge densities. For general stochastic block models, the across-community edge density is allowed to vary between each community. In this work, we always take a constant cross-community edge density.

REMARK 4.3. We allow for infinite vectors with a finite number of non-zero entries so that we may smoothly introduce new communities within the stochastic block model. For example, let $t \in [0, 1]$ and parametrize $\boldsymbol{s}$ and $\boldsymbol{p}$ by t as $\boldsymbol{s}(t) = [1 - t/2, t/2, 0, \vdots]^T$ and $\boldsymbol{p}(t) = [0.2 + t/2, 0.1 + t/2, 0, \ddots$.

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 4.8. (Canonical stochastic block model kernel)The function f, which is piecewise constant over the blocks, and is defined by $f : [0, 1] \times [0, 1] \longrightarrow [0, 1]$

$$f(x, y) = \begin{cases} p_i & \text{if} \quad \sum_{k=1}^{i-1} s_k \le x < \sum_{k=1}^{i} s_k, \\ & \text{and} \quad \sum_{k=1}^{i-1} s_k \le y < \sum_{k=1}^{i} s_k, \\ q & \text{if} \quad \sum_{k=1}^{i-1} s_k \le x < \sum_{k=1}^{i} s_k, \\ & \text{and} \quad \sum_{k=1}^{j-1} s_k \le y < \sum_{k=1}^{j} s_k \end{cases} \tag{4.12}$$

is called the canonical kernel of the stochastic block model with measure denoted by $\mu_{\omega_n f}$ (see, e.g. Fig. 1), and we denote it by $f(x, y; \boldsymbol{p}, q, \boldsymbol{s})$.

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

5. The Fréchet function and sample Fréchet function

We equip the set $\mathscr{G}$ of graphs defined on n vertices (see definition 4.4) with a (pseudo)-metric, d. We consider a probability measure $\mu \in \mathscr{M}(\mathscr{G})$ that describes the probability of obtaining a given graph when we sample $\mathscr{G}$ according to μ and define the Fréchet function.

DEFINITION 5.1. (Fréchet function)We denote by F_r the Fréchet function,

$$F_r(G) = \mathbb{E}_\mu\Big[d^r(G, G_\mu)\Big]. \tag{5.1}$$

Using d, we quantify the spread of the graphs, and we define a notion of centrality by minimizing the Fréchet function for various choices of r, which gives the location of the expected graph according to μ. For $r = 2$ we define the Fréchet mean.

DEFINITION 5.2. (Fréchet mean [33])The Fréchet mean of the probability measure μ in the pseudometric space $(\mathscr{G}, d)$ is the set of graphs $G_2^{*,\mu}$ whose expected distance squared to the observed graphs is minimum,

$$\left\{G_2^{*,\mu} \in \mathscr{G}\right\} = \underset{G\in\mathscr{G}}{\operatorname{argmin}}\, F_2(G) = \underset{G\in\mathscr{G}}{\operatorname{argmin}}\, \mathbb{E}_\mu\Big[d^2(G, G_\mu)\Big], \tag{5.2}$$

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 $\mathscr{G}$ is a finite set, the minimization problem (5.2) always has at least one solution. Throughout this work, we are interested in determining at least one element of the set $\{G_2^{*,\mu} \in \mathscr{G}\}$. Because our results hold for any minimizer of (5.2), i.e., for any Fréchet mean of μ, we work with a single representative graph from the Fréchet mean set so, to ease our exposition, we write the Fréchet mean as follows:

$$G_2^{*,\mu} = \underset{G\in\mathscr{G}}{\operatorname{argmin}}\, \mathbb{E}_\mu\Big[d^2(G, G_\mu)\Big]. \tag{5.3}$$

We note the similarity between equation (5.3) and the definition of the barycenter [64]. Indeed, as we change μ, we expect that, for a fixed G, $\mathbb{E}_\mu\Big[d^2(G, G_\mu)\Big]$ will change, and therefore the Fréchet mean, $G_2^{*,\mu}$, will move inside $\mathscr{G}$ for different choices of the probability measure μ.

Observe that $G_2^{*,\mu}$ is the center of mass for the mass distribution associated with μ. Different centroids (rather than just the mean) can be generalized to metric spaces by considering different choices of r when minimizing the Fréchet function. In particular, taking $r = 1$ defines the Fréchet median graph.

DEFINITION 5.3. (Fréchet median) The Fréchet median of the probability measure μ in the pseudometric space $(\mathscr{G}, d)$ is the set of graphs $G_1^{*,\mu}$ whose expected distance to the observed graphs is minimum,

$$\left\{G_1^{*,\mu} \in \mathscr{G}\right\} = \underset{G\in\mathscr{G}}{\operatorname{argmin}}\, F_1(G) = \underset{G\in\mathscr{G}}{\operatorname{argmin}}\, \mathbb{E}_\mu\Big[d(G, G_\mu)\Big], \tag{5.4}$$

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

Again because our results hold for any graph in the Fréchet median set, we denote the solution to (5.4) as a single graph,

$$G_1^{*,\mu} = \operatorname*{argmin}_{G \in \mathscr{G}} \mathbb{E}_\mu\Big[d(G, G_\mu)\Big]. \tag{5.5}$$

While the mean and median are similar, they capture different notions of centrality. For Euclidean spaces, the differences between the mean and median are well understood. For the space of graphs, the difference depends on the specific distance equipped to $\mathscr{G}$. For example, when $d = d_H$, and for certain classes of probability measures, it is shown in [58] that the (sample) Fréchet mean and median graphs are similar with high probability which is unlike the situation in Euclidean spaces in general.

In practice, the only information known about a distribution on $\mathscr{G}$ comes from a sample of graphs. Therefore, we need a notion of the sample Fréchet function, defined by replacing μ with the empirical measure. Given a set of graphs $\{G^{(k)}\}_{k=1}^N$ we define the sample Fréchet function.

Definition 5.4. (Sample Fréchet function)We denote by $F_{N,r}$ the sample Fréchet function,

$$F_{N,r}(G) = \frac{1}{N}\sum_{k=1}^{N} d^r(G, G^{(k)}). \tag{5.6}$$

The sample Fréchet mean and median are defined by minimizing the sample Fréchet function when $r = 2$ and $r = 1$, respectively.

Definition 5.5. (Sample Fréchet mean)Let $\{G^{(k)}\}$ $1 \le k \le N$ be a set of graphs in $\mathscr{G}$. The sample Fréchet mean is defined by

$$\{G_{N,2}^* \in \mathscr{G}\} = \operatorname*{argmin}_{G \in \mathscr{G}} F_{N,2}(G) = \operatorname*{argmin}_{G \in \mathscr{G}} \frac{1}{N}\sum_{k=1}^{N} d^2(G, G^{(k)}). \tag{5.7}$$

Again, our results hold for any minimizer of (5.7), so we write the sample Fréchet mean as a singleton set,

$$G_{N,2}^* = \operatorname*{argmin}_{G \in \mathscr{G}} \frac{1}{N}\sum_{k=1}^{N} d^2(G, G^{(k)}). \tag{5.8}$$

The dependence on N is explicitly written here but may be suppressed throughout the paper when it is obvious. The sample Fréchet median is specified by setting $r = 1$ and minimizing the sample Fréchet function. Here we introduce the sample Fréchet median as a singleton set for brevity though, in general, it will be set-valued as well.

Definition 5.6. (Sample Fréchet median) Let $\{G^{(k)}\}$ $1 \le k \le N$ be a set of graphs in $\mathscr{G}$. The sample Fréchet median is defined by

$$G_{N,1}^* = \operatorname*{argmin}_{G \in \mathscr{G}} F_{N,1}(G) = \operatorname*{argmin}_{G \in \mathscr{G}} \frac{1}{N}\sum_{k=1}^{N} d(G, G^{(k)}). \tag{5.9}$$

Determining the (sample) Fréchet mean or median is intractable with a direct approach whenever n, the number of vertices in the graph, is reasonably large because $|\mathscr{G}| = \mathscr{O}(2^{n^2})$. A compounding difficulty results from the fact that $\mathscr{G}$ is not ordered, so searching $\mathscr{G}$ in a principled manner is non-trivial (in contrast to the situation with trees [7, 11]). Furthermore, because the geometry of the space $\mathscr{G}$ varies with the choice of metric, developing an algorithm that searches the space $\mathscr{G}$ in a principled manner with respect to one metric need not work well when considering a different metric.

In the following section we propose a technique that can be employed to minimize the (sample) Fréchet function independent of the choice of r or the distance d. We suggest first lifting the problem to the space of probability measures and searching for an approximate solution to the lifted problem by restricting the space of probability measures to a parametrizable subset.

REMARK 5.1. In fact, the technique introduced could be employed to find the minimizer of any graph valued optimization problem, not just the minimizers of the Fréchet function.

6. A solution technique to minimize the sample Fréchet function

Let $\left\{G^{(k)}\right\}_{k=1}^{N}$ be a set of graphs sampled independently according to a probability measure $\nu \in \mathscr{M}(\mathscr{G})$. Let the graph

$$G^*_{N,r} = \underset{G \in \mathscr{G}}{\operatorname{argmin}}\, F_{N,r}(G) = \underset{G \in \mathscr{G}}{\operatorname{argmin}} \sum_{k=1}^{N} d^r(G, G^{(k)}) \tag{6.1}$$

be the minimizer of the sample Fréchet function for any choice of metric d and value of r.

The procedure we suggest to determine $G^*_{N,r}$ takes three fundamental steps. First, we lift the problem to the space of probability measures. Second, we define a suitable parametrizable subset of probability measures to search over. Third, we show that only certain information about the graph $G^*_{N,r}$ is needed to define the solution to the lifted problem allowing for a fast algorithm to be implemented in practice.

The order of the steps presented here is not always the order taken in practice. Specifically, step 3 can be taken at any time.

6.1 *Step 1: Lifting the problem to $\mathscr{M}(\mathscr{G})$*

Let $\mu \in \mathscr{M}(\mathscr{G})$ be a probability measure and define the graph

$$G^{*,\mu}_{r'} = \underset{G \in \mathscr{G}}{\operatorname{argmin}}\, F_{r'}(G) = \underset{G \in \mathscr{G}}{\operatorname{argmin}}\, \mathbb{E}_\mu\Big[d^{r'}(G, G_\mu)\Big], \tag{6.2}$$

where $r' \in \{1, 2\}$. In general, r' can be any natural number but here we are restricting our analysis to only the mean or median graphs.

REMARK 6.1. The value of r taken in equation (6.1) need not be the same value as r' in equation (6.2).

First, we observe that there exists a $\mu^* \in \mathscr{M}(\mathscr{G})$ such that

$$d(G^{*,\mu^*}_{r'}, G^*_{N,r}) = 0. \tag{6.3}$$

This can be shown by considering the measure μ^* defined below. Let $M \subset \mathscr{G}$ be a subset of graphs and define μ^* by the following:

$$\mu^*(M) = \begin{cases} 1 & \text{if} G^*_{N,r} \in M \\ 0 & \text{if} G^*_{N,r} \notin M. \end{cases} \tag{6.4}$$

The graph $G^{*,\mu^*}_{r'}$ of μ^* is trivially given by $G^*_{N,r}$, the minimizer of the sample Fréchet function defined in equation (6.1). The existence of such a probability measure indicates that the following optimization problem

$$\{\mu^* \in \mathscr{M}(\mathscr{G})\} = \underset{\mu \in \mathscr{M}(\mathscr{G})}{\operatorname{argmin}}\ d(G^{*,\mu}_{r'}, G_{N,r^*}) \tag{6.5}$$

has at least one solution. In general, the solution to the above optimization problem is set-valued because there exist many probability measures with the same center of mass. However, because finding any one of these probability measures determines the graph $G^*_{N,r}$, we only need to find at least one solution. Therefore, we represent the solution by a single probability measure and denote the solution as

$$\mu^* = \underset{\mu \in \mathscr{M}(\mathscr{G})}{\operatorname{argmin}}\ d(G^{*,\mu}_{r'}, G^*_{N,r}). \tag{6.6}$$

Equation (6.6) shows how we lift the problem from a search over the space $\mathscr{G}$ to the space $\mathscr{M}(\mathscr{G})$, which constitutes the first step in our approach.

6.2 *Step 2: Defining an approximate problem*

An approximate problem can be easily defined by restricting the searchable space of probability measures to a subset, denoted $\mathscr{M}_s(\mathscr{G}) \subset \mathscr{M}(\mathscr{G})$. A probability measure that solves the restricted problem is denoted by

$$\mu^*_s = \underset{\mu_s \in \mathscr{M}_s(\mathscr{G})}{\operatorname{argmin}}\ d(G^{*,\mu_s}_{r'}, G^*_{N,r}). \tag{6.7}$$

Throughout this manuscript we consider subsets of probability measures that are easily parametrizable. Most often it is the case that we consider the subset of kernel probability measures (Definition 4.6) or, in section 8, a specific class of kernel probability measures known as the stochastic block model kernel probability measures (Definition 4.8).

Whenever $\mu^* \in \mathscr{M}_s(\mathscr{G})$ then $\mu^*_s = \mu^*$ and the solution to the approximate problem is identical to the original problem. It may be the case that $\mu^* \notin \mathscr{M}_s(\mathscr{G})$. In this case, we show that for a specific subset of probability measures and choice of metric, one may still guarantee that the distance

$$d(G^{*,\mu^*_s}_{r'}, G^*_{N,r}) \tag{6.8}$$

is small. We then interpret the graph $G^{*,\mu^*_s}_{r'}$ as an approximation of the sample Fréchet mean graph $G^*_{N,r}$ with respect to the distance d.

6.3 *Step 3: Identifying the necessary information contained within the graph $G^*_{N,r}$ to minimize equation (6.7)*

This step is broken into two smaller parts. First, we observe that to minimize equation (6.7) it is sufficient to have knowledge of the graph $G^*_{N,r}$. However, knowledge of this graph is unknown in general because this is the sample Fréchet mean graph that we are attempting to determine originally. Notably, it is not always necessary to have knowledge of the entire graph $G^*_{N,r}$ to minimize the objective defined in equation (6.6) or the approximate problem defined in (6.7). Instead, it is only necessary to have knowledge of the information contained within the graph that is relevant to the chosen metric d.

The second step within this section discusses briefly general methods to estimate the information contained with the graph $G^*_{N,r}$ that is relevant to the metric d.

The following example shows these two parts of step 3 explicitly when the metric is given by d_A (equation (4.6)).

Example 6.1. Let $d = d_A$ and let $G^{*,\mu_s}_{r'}$ and $G^*_{N,r}$ from equation (6.7) have adjacency matrices denoted by $\boldsymbol{A}^{*,\mu_s}_{r'}$ and $\boldsymbol{A}^*_{N,r}$, respectively. Consider the problem defined by equation (6.7),

$$\mu^*_s = \underset{\mu_s \in \mathscr{M}_s(\mathscr{G})}{\operatorname{argmin}} \; d_A(G^{*,\mu_s}_{r'}, G^*_{N,r}) \tag{6.9}$$

$$= \underset{\mu_s \in \mathscr{M}_s(\mathscr{G})}{\operatorname{argmin}} \; ||\sigma(\boldsymbol{A}^{*,\mu_s}_{r'}) - \sigma(\boldsymbol{A}^*_{N,r})||_2. \tag{6.10}$$

First, in equation (6.10), it is clear that the only information from the graph $G^*_{N,r}$ that is relevant to the problem are the eigenvalues of $\boldsymbol{A}^*_{N,r}$. Therefore, to minimize the objective, it is sufficient to characterize the eigenvalues of the adjacency matrix of the graph $G^*_{N,r}$ rather than the entire graph $G^*_{N,r}$.

Second, the term $\sigma(\boldsymbol{A}^*_{N,r})$ can be estimated in a number of ways. We suggest quantifying the information that is relevant to the metric d_A from each graph $G^{(k)}$ in the data set and summarizing this information via the arithmetic mean. In general, an estimate of $\sigma(\boldsymbol{A}^*_{N,r})$ will be the arithmetic mean of the vectors $\sigma(\boldsymbol{A}^{(k)})$ when $r = 2$ or the median value of the vectors $\sigma(\boldsymbol{A}^{(k)}$ when $r = 1$. For larger values of r a different notion of centrality for the set of vectors $\{\sigma(\boldsymbol{A}^{(k)})\}_{k=1}^N$ may need to be considered as an estimate of $\sigma(\boldsymbol{A}^*_{N,r})$.

Remark 6.2. The relationship between the arithmetic mean of the eigenvalues of the adjacency matrices and the eigenvalues of $\boldsymbol{A}^*_{N,r}$ can be complex. We show in Theorem 8.2 and its proof that the arithmetic mean of the eigenvalues is a good estimate of the information from the graph $G^*_{N,r}$ that is relevant to the metric d_{A_c} (rather than d_A) when the size of the graphs, n, is large.

Similar observations are true of $G^{*,\mu_s}_{r'}$ in that only the eigenvalues of the adjacency matrix of this graph are relevant to the objective value.

Recall now that our goal is to determine the sample Fréchet mean graph $G^*_{N,r}$ (or an approximation of this graph) given the data set of graphs $\left\{G^{(k)}\right\}_{k=1}^N$. Because the probability measure μ^*_s is defined such that the minimizer of the Fréchet function $F_{r'}$ given μ^*_s is (close to) the target sample Fréchet mean graph $G^*_{N,r}$ with respect to d we will be able to either analytically determine the graph $G^{*,\mu^*_s}_{r'}$ (Section 7) or, when an analytic expression for $G^{*,\mu^*_s}_{r'}$ is unknown, we may take an arbitrarily sized sample from μ^*_s and compute an appropriate graph valued statistic given the new sample of graphs (Section 8; Theorem 8.5).

Step 3 is a comment on the objective value for the sample Fréchet mean (or median) problem and does not depend on the underlying space under which the optimization is performed over. Therefore, in certain cases, it is beneficial to perform Step 3 prior to Step 2 because doing so may inform the specific subset of probability measures to restrict to in Step 2.

The following two sections show that the solution technique works well when considering two vastly different metrics with minimal change in the procedure to determine the solution. We first consider the Hamming distance, the metric which has been most widely studied when considering the sample Fréchet mean problem and show that we recover known results using our technique. We then consider the metric d_{A_c} in the case of arbitrarily large graphs. Our results for this metric constitute the bulk of our theoretical contributions which culminate in an algorithm to be implemented in practice. We verify experimentally that the graph recovered by our method, $G_{r'}^{*,\mu_s^*}$, is close to the true sample Fréchet mean graph $G_{N,2}^*$.

7. The sample Fréchet median when $d = d_H$

Here we show that our proposed solution technique recovers known results about the sample Fréchet median graph when the metric equipped to $\mathscr{G}$ is d_H. Because our analysis is about the theoretical properties of the sample Fréchet median graph, we assume that there is a common vertex set for the graphs in the data set and we therefore do not have to solve the graph matching problem.

In this section we determine a precise probability measure $\mu^* \in \mathscr{M}(\mathscr{G})$ whose Fréchet mean is identical to the sample Fréchet median for any given data set of graphs. The specifics of the solution technique are detailed throughout the proof of Lemma 7.1.

Let $\left\{G^{(k)}\right\}_{k=1}^N$ be a set of graphs sampled independently from a distribution $\nu \in \mathscr{M}(\mathscr{G})$. Let $d = d_H$ and denote the median graph as

$$G_{N,1}^* = \operatorname*{argmin}_{G\in\mathscr{G}} \frac{1}{N}\sum_{k=1}^N d_H(G, G^{(k)}) \tag{7.1}$$

with adjacency matrix $\boldsymbol{A}_{N,1}^*$. Define the arithmetic mean of the adjacency matrices as

$$\boldsymbol{A}_{N,a}^* = \frac{1}{N}\sum_{k=1}^N \boldsymbol{A}^{(k)}, \tag{7.2}$$

where $\boldsymbol{A}^{(k)}$ denotes the adjacency matrix of graph $G^{(k)}$. Let $\mu_{\omega_n f}^*$ be a kernel probability measure where the values of $\omega_n f$ on the grid points $\left\{\frac{i}{n}\right\}_{i=1}^n \times \left\{\frac{j}{n}\right\}_{j=1}^n \subset [0,1]^2$ are

$$\omega_n f\left(\frac{i}{n}, \frac{j}{n}\right) = (\boldsymbol{A}_{N,a}^*)_{ij}. \tag{7.3}$$

Because the values of f on the points $[0,1]^2 \backslash \left(\left\{\frac{i}{n}\right\}_{i=1}^n \times \left\{\frac{j}{n}\right\}_{j=1}^n\right)$ bear no impact on the results of this section we leave them unspecified. We summarize the results of this section with the following lemma.

LEMMA 7.1. Let $G_2^{*,\mu^*_{\omega_n f}}$ denote the Fréchet mean of $\mu^*_{\omega_n f}$ with respect to the metric d_H. Then

$$d_H(G_2^{*,\mu^*_{\omega_n f}}, G^*_{N,1}) = 0. \tag{7.4}$$

Proof. The proof for the above lemma is broken into the following steps, which highlight the solution technique. At a very high level, the goal is to determine a probability measure with the correct Fréchet mean or median graph and, as suggested by Example 6.1, the correct probability measure can be determined by quantifying the properties of $G^*_{N,1}$ which are relevant to the metric by analyzing the arithmetic mean of the observed adjacency matrices.

Because the steps outlined in Section 6 do not need to be completed in order, here we elect to begin with step 3 to better understand the information contained within the graph $G^*_{N,1}$ that is pertinent to the metric d_H. We then complete steps 1 and 2 and conclude the proof.

1. Write equation (7.1) over the space of adjacency matrices as

$$\boldsymbol{A}^*_{N,1} = \underset{\boldsymbol{A}\in\{0,1\}^{n\times n}}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^{N}\sum_{1\le i<j\le n}\left|(\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij}\right|. \tag{7.5}$$

Observe that $\left|(\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij}\right| = \left|(\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij}\right|^2$ since $(\boldsymbol{A})_{ij}$ and $(\boldsymbol{A}^{(k)})_{ij}$ are both in the set $\{0,1\}$. Therefore

$$\boldsymbol{A}^*_{N,1} = \underset{\boldsymbol{A}\in\{0,1\}^{n\times n}}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^{N}\sum_{1\le i<j\le n}\left|(\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij}\right|^2. \tag{7.6}$$

2. Relax the problem defined in equation (7.6) to the space of real $n \times n$ matrices

$$\boldsymbol{A}^*_{N,a} = \underset{\boldsymbol{A}\in\mathbb{R}^{n\times n}}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^{N}\sum_{1\le i<j\le n}\left|(\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij}\right|^2. \tag{7.7}$$

The solution to the relaxed problem is simply the arithmetic mean of the adjacency matrices.

3. Rewriting the minimization problems in equations (7.6) and (7.7)

$$\boldsymbol{A}^*_{N,1} = \underset{\boldsymbol{A}\in\{0,1\}^{n\times n}}{\operatorname{argmin}} \frac{1}{N}\sum_{1\le i<j\le n}\sum_{k=1}^{N}\left|(\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij}\right|^2. \tag{7.8}$$

$$\boldsymbol{A}^*_{N,a} = \underset{\boldsymbol{A}\in\mathbb{R}^{n\times n}}{\operatorname{argmin}} \frac{1}{N}\sum_{1\le i<j\le n}\sum_{k=1}^{N}\left|(\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij}\right|^2 \tag{7.9}$$

shows that we may optimize the objective for each i, j independently.

4. For each i, j

$$(\boldsymbol{A}^*_{N,1})_{ij} = \underset{(\boldsymbol{A})_{ij} \in \{0,1\}}{\operatorname{argmin}} \frac{1}{N} \sum_{k=1}^{N} \left| (\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij} \right|^2 \tag{7.10}$$

$$(\boldsymbol{A}^*_{N,a})_{ij} = \underset{(\boldsymbol{A})_{ij} \in \mathbb{R}}{\operatorname{argmin}} \frac{1}{N} \sum_{k=1}^{N} \left| (\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij} \right|^2 \tag{7.11}$$

5. The minimum value of the objective in equation (7.11) provides a lower bound on the objective in equation (7.10),

$$\min_{(\boldsymbol{A})_{ij} \in \mathbb{R}} \frac{1}{N} \sum_{k=1}^{N} \left| (\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij} \right|^2 < \min_{(\boldsymbol{A})_{ij} \in \{0,1\}} \frac{1}{N} \sum_{k=1}^{N} \left| (\boldsymbol{A})_{ij} - (\boldsymbol{A}^{(k)})_{ij} \right|^2 . \tag{7.12}$$

The value of $(\boldsymbol{A}^*_{N,1})_{ij}$, which minimizes the sample Fréchet median objective must be the realizable element of the set $\{0, 1\}$, which least increases the objective value from the point $(\boldsymbol{A}^*_{N,a})_{ij}$.

6. Because the objective in equation (7.11) is quadratic, the closest realizable element to $(\boldsymbol{A}^*_{N,a})_{ij}$ is also the element that least increases the objective value in equation (7.11). We therefore minimize the following objective for each i, j to determine the graph $G^*_{N,1}$,

$$(\boldsymbol{A}^*_{N,1})_{ij} = \underset{(\boldsymbol{A})_{ij} \in \{0,1\}}{\operatorname{argmin}} \left| (\boldsymbol{A})_{ij} - (\boldsymbol{A}^*_{N,a})_{ij} \right|^2 . \tag{7.13}$$

Rather than minimizing for each i, j separately, we may also define the problem for all i, j

$$\boldsymbol{A}^*_{N,1} = \underset{\boldsymbol{A} \in \{0,1\}^{n \times n}}{\operatorname{argmin}} \sum_{1 \le i < j \le n} \left| (\boldsymbol{A})_{ij} - (\boldsymbol{A}^*_{N,a})_{ij} \right|^2 . \tag{7.14}$$

Note, due to the simplicity of the Hamming distance, there are many objectives which may be minimized in this step that yield the same minimizer rather than the objective that is presented in equations (7.13) or (7.14).

A brief aside is in order. At this stage of the proof we have accomplished step 3 of section 6 which discusses the quantification of the information contained with the graph $G^*_{N,1}$ that is pertinent to the objective value in the sample Fréchet mean problem via the arithmetic mean of the adjacency matrices $\boldsymbol{A}^*_{N,a}$.

Furthermore, we acknowledge that the solution to equation (7.13) is analytic and the graph $G^*_{N,1}$ can be determined at this point. The purpose of this manuscript is to showcase a general approach to solving the sample Fréchet mean problem for a variety of distances. The solution at this stage relies on the geometry of the space of graphs induced by the Hamming distance. Rather than rely on this geometry, we instead perform steps 1 and 2 outlined in Section 6 to lift the search space to $\mathscr{M}(\mathscr{G})$, the space of probability measures, where the geometry is not necessarily dependent on the choice of distance equipped to $\mathscr{G}$.

7. Lift the problem defined in equation (7.14) to the space of probability measures,

$$\mu^* = \underset{\mu \in \mathscr{M}(\mathscr{G})}{\operatorname{argmin}} \sum_{1 \le i < j \le n} \left| (\boldsymbol{A}_2^{*,\mu})_{ij} - (\boldsymbol{A}_{N,a}^*)_{ij} \right|^2. \tag{7.15}$$

This accomplishes step 1 of Section 6.

8. Restrict the problem to the space of kernel probability measures, denoted $\mu_{\omega_n f} \in \mathscr{M}_s(\mathscr{G})$,

$$\mu^*_{\omega_n f} = \underset{\mu_{\omega_n f} \in \mathscr{M}_s(\mathscr{G})}{\operatorname{argmin}} \sum_{1 \le i < j \le n} \left| (\boldsymbol{A}_2^{*,\mu_{\omega_n f}})_{ij} - (\boldsymbol{A}_{N,a}^*)_{ij} \right|. \tag{7.16}$$

This accomplishes step 2 from Section 6. Observe that searching the restricted set of kernel probability measures can be done easily independent of the choice of metric equipped to $\mathscr{G}$ by continuously updating the values of $\omega_n f(x,y)$ for each $(x,y) \in [0,1]^2$.
All that remains is a method to characterize the mean of a kernel probability measure with respect to the distance, which for the Hamming distance is analytic.

9. Theorem 1 in [58] states that the Fréchet mean with respect to the Hamming distance of a kernel probability measure, $\mu_{\omega_n f}$, can be determined analytically by thresholding the entries of the expected adjacency matrix, $\mathbb{E}\left[(\boldsymbol{A}_{\mu_{\omega_n f}})_{ij}\right] = \omega_n f\left(\frac{i}{n}, \frac{j}{n}\right)$. This result suggests that there are many kernel probability measures with the same Fréchet mean. Define

$$\omega_n f\left(\frac{i}{n}, \frac{j}{n}\right) = (\boldsymbol{A}_{N,a}^*)_{ij} \tag{7.17}$$

for the kernel probability measure $\mu^*_{\omega_n f}$.

10. Consider now the Fréchet mean of $\mu^*_{\omega_n f}$, denoted $G_2^{*,\mu^*_{\omega_n f}}$ with adjacency matrix $\boldsymbol{A}_2^{*,\mu^*_{\omega_n f}}$. We show that this adjacency matrix minimizes equation (7.16) by showing that for every i,j any change to the graph $G_2^{*,\mu^*_{\omega_n f}}$ either increases the objective value or the objective value remains constant. As a result of Theorem 1 in [58], for every i,j,

$$\left| (\boldsymbol{A}_2^{*,\mu^*_{\omega_n f}})_{ij} - (\boldsymbol{A}_{N,a}^*)_{ij} \right| \in [0, 0.5] \tag{7.18}$$

because the entries of $(\boldsymbol{A}_2^{*,\mu^*_{\omega_n f}})_{ij}$ are the closest realizable element of the set $\{0,1\}$ to the point $(\boldsymbol{A}_{N,a}^*)_{ij}$. In the case that

$$\left| (\boldsymbol{A}_2^{*,\mu^*_{\omega_n f}})_{ij} - (\boldsymbol{A}_{N,a}^*)_{ij} \right| \neq 0.5 \tag{7.19}$$

then changing the entry of $(\boldsymbol{A}_2^{*,\mu^*_{\omega_n f}})_{ij}$ will strictly increase the objective value. In the event that

$$\left| (\boldsymbol{A}_2^{*,\mu^*_{\omega_n f}})_{ij} - (\boldsymbol{A}_{N,a}^*)_{ij} \right| = 0.5 \tag{7.20}$$

then changing the entry of $(\boldsymbol{A}_2^{*,\mu^*_{\omega nf}})_{ij}$ will not increase the objective value and, the conclusion, is that the adjacency matrix defined by $\boldsymbol{A}_2^{*,\mu^*_{\omega nf}}$ minimizes the objective value in equation (7.16). Therefore,

$$d_H(G_2^{*,\mu^*_{\omega nf}}, G^*_{N,1}) = 0 \tag{7.21}$$

and we have determined the sample Fréchet median graph by finding a probability measure with the correct Fréchet mean graph. □

8. The sample Fréchet mean when $d = d_{A_c}$

The focus of this section is to show that the procedure used to determine the sample Fréchet median with respect to the Hamming distance may also be used to determine the sample Fréchet mean with respect to the metric d_{A_c}.

Because the method to compute the sample Fréchet mean graph with respect to d_{A_c} is novel, we also provide the pseudocode for an algorithm to implement in practice to determine the sample Fréchet mean graph. We validate the theory and the proposed algorithm on several data sets in the case of finite graph size and provide an application of the sample Fréchet mean graph, K-means clustering.

Throughout this section, we always work with the mean graph (as opposed to the median graph) so that $r = r' = 2$ for the Fréchet functions considered in this section and we omit the notation in the subscript of the solutions to the sample Fréchet mean problem so that $G^*_{N,2} = G^*_N$.

8.1 *Approximately solving the sample Fréchet mean problem when $d = d_{A_c}$*

Let $\{G^{(k)}\}_{k=1}^N$ be a set of graphs sampled independently from a distribution $\nu \in \mathscr{M}(\mathscr{G})$. Let $d = d_{A_c}$ and denote the sample Fréchet mean graph as

$$G^*_N = \underset{G\in\mathscr{G}}{\text{argmin}}\ \frac{1}{N}\sum_{k=1}^{N} d_{A_c}(G, G^{(k)}). \tag{8.1}$$

Throughout this section, we determine a graph that is arbitrarily close to G^*_N with respect to the metric d_{A_c}.

We begin our analysis with the various assumptions necessary for the theorems presented in this section. Let $G \in \mathscr{G}$ with adjacency matrix $\boldsymbol{A}$. For the graph G assume that

1. $\rho_n = \omega(n^{-2/3})$.
2. $\lim_{n\to\infty} \rho_n = 0$.
3. $0 \preccurlyeq \sigma_c(\boldsymbol{A})$.
4. For every $1 \le i \ne j \le c$, $\lambda_i \ne \lambda_j$.

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. Assumption 1 can be traced back to the state-of-the-art work on the characterization of the limiting distribution of the largest eigenvalues of graphs sampled from inhomogeneous Erdős–Rényi ensembles [21, 25]. We expect that as more general assumptions are considered when quantifying the behavior of the largest eigenvalues of the adjacency matrices of inhomogeneous Erdős–Rényi random graphs then the assumptions and results of this manuscript will similarly generalize.

The remaining assumptions are relatively mild. Many graphs have a decaying density as n grows, which implies that the average degree of a vertex is dominated by n. Furthermore, the largest eigenvalues of graphs are distinct with high probability and positive, which is seen both in practice and for many ensembles of random graphs such as the stochastic block model, the Barabási–Albert ensemble and the Watts–Strogatz ensemble.

8.1.1 *Existence of a stochastic block model kernel probability measure with Fréchet mean close to* G_N^*. Our primary theoretical contribution, Theorem 8.1, states that we may approximate 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, d_{A_c}.

Theorem 8.1 and Corollary 8.1 guarantee that when restricting the problem defined in equation (6.6) to the subset of stochastic block model kernel probability measures, the Fréchet mean graph of the recovered probability measure is close to the target sample Fréchet mean graph G_N^*. These theorems therefore accomplish steps 1 and 2 outlined in section 6 when the metric is d_{A_c}.

THEOREM 8.1. (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

$$\lim_{N\to\infty} d_{A_c}(G, G_N^{*,\mu_{\omega_n f}}) < \epsilon \quad a.s. \tag{8.2}$$

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. The proof is in Appendix C. □

REMARK 8.1. 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 many methods may exist to estimate this quantity, we give in Algorithm 1 one suggestion for the choice of c. It is worth mentioning that under-estimating c may be preferred to over-estimating c because an overestimate of c will compare eigenvalues that capture fine-scale behaviour in the graphs to eigenvalues that determine global structures of the graphs, leading to potentially inaccurate conclusions.

REMARK 8.2. 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 so long as the graphs are sufficiently large. Intuitively, when given a stochastic block model with a set of parameters $\boldsymbol{p}$ and $\boldsymbol{s}$, one can decrease the value of s_i while increasing the value of p_i and maintain the expected i-th eigenvalue of the adjacency matrices of graphs sampled from this ensemble. See e.g. subsection 8.3.3 as an example of two different stochastic block model kernels where the largest eigenvalues of the adjacency matrices are identical.

The following corollary applies Theorem 8.1 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 8.1, 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 (5.7), when $d = d_{A_c}$.

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 8.1.

Corollary 8.1. (Approximation of the sample Fréchet mean)$\forall \epsilon > 0$, $\exists n_1 \in \mathbb{N}$ 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

$$\lim_{\tilde{N}\to\infty} d_{A_c}(G_N^*, G_{\tilde{N}}^{*,\mu_{\omega_n f}}) < \epsilon \quad a.s. \tag{8.3}$$

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 8.3. One requirement on G_N^* is that the density, denoted ρ_n^*, satisfies assumption 1. The theory in [27] 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^*.

Remark 8.4. The Szemerédi regularity lemma [49; 54] states that every dense graph can be partitioned into pairs of random bipartite graphs. The lemma provides a partition of the nodes into equally sized blocks such that the connectivity between the blocks is quasirandom [54]. In its original formulation [74], the number of pairs of bipartite graphs grows exponentially with the inverse of the fraction of the edges not included in the pairs. A weaker version of the regularity lemma yields a manageable number of blocks [34], and in fact proves that the space of step graphons is dense in the space of graphons for the topology induced by the cut-norm. The significance of the result is not that one can approximate any graphon arbitrary well with a step function (after all the result holds for the L^1 and L^2 norms), but rather that the approximation error only depends on the complexity (number of steps) of the step graphon, and not on the complexity of the original graphon. Unfortunately, the regularity lemma is only useful for dense graphs, where $|E| = \mathcal{O}(n^2)$. Some recent results [12, 48, 68] have extended Szemerédi regularity lemma to sparse graphs.

Corollary 8.1 along with Section 6 suggest that the following minimization procedure should be solved to determine the stochastic block model probability measure whose Fréchet mean graph is close to G_N^*,

$$\mu_{\omega_n f}^* = \underset{\mu_{\omega_n f} \in \mathcal{M}_s(\mathcal{G})}{\operatorname{argmin}} \; d_{A_c}^2 \left(G^{*,\mu_{\omega_n f}}, G_N^*\right), \tag{8.4}$$

where $\mathcal{M}_s(\mathcal{G})$ is the subset of stochastic block model kernel probability measures.

Remark 8.5. Equation (8.4) is the identical to equation (6.7) when $d = d_{A_c}$.

Equation (8.4) concludes the first two steps of the solution approach to determine the sample Fréchet mean graph G_N^* as outlined by Section 6. Note, however, that the results of Theorem 8.1 and Corollary 8.1 are purely existential and we have not yet specified a method to implement which will minimize the objective in equation (8.4). To do so we first need to quantify the information from the graph G_N^* that is relevant to the metric d_{A_c} (see Section 8.1.2 and Theorem 8.2). This will conclude step 3 from Section 6. To implement an algorithm in practice that recovers $\mu^*_{\omega_n f}$, we also quantify the information from the graph $G^{*,\mu_{\omega_n f}}$ that is relevant to the metric d_{A_c} (see Section 8.1.3 along with Theorems 8.3 and 8.4).

8.1.2 *Quantifying the properties of G_N^* that are relevant to the metric d_{A_c}.* In a similar manner to Example 6.1, the objective value outlined by equation (8.4) depends only on the largest c eigenvalues of the adjacency matrices of the graphs G_N^* and $G^{*,\mu_{\omega_n f}}$ which is shown explicitly by the following equations,

$$\mu^*_{\omega_n f} = \underset{\mu_{\omega_n f} \in \mathscr{M}_s(\mathscr{G})}{\operatorname{argmin}} \; d^2_{A_c}\left(G^{*,\mu_{\omega_n f}}, G_N^*\right) \tag{8.5}$$

$$= \underset{\mu_{\omega_n f} \in \mathscr{M}_s(\mathscr{G})}{\operatorname{argmin}} \; \left|\left|\sigma_c(\boldsymbol{A}^{*,\mu_{\omega_n f}}) - \sigma_c(\boldsymbol{A}_N^*)\right|\right|_2^2. \tag{8.6}$$

Therefore, to determine the probability measure $\mu^*_{\omega_n f}$, it is only necessary to quantify the largest eigenvalues of $\boldsymbol{A}_N^*$ which is done in the following theorem.

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

THEOREM 8.2. $\forall \epsilon > 0, \exists n^* \in \mathbb{N}$ such that $\forall n > n^*$,

$$\left\| \sigma_c(\boldsymbol{A}_N^*) - \frac{1}{N}\sum_{k=1}^{N} \sigma_c(\boldsymbol{A}^{(k)}) \right\|_2 < \epsilon. \tag{8.7}$$

Proof. The proof is in Appendix D □

REMARK 8.6. Here we again use the arithmetic mean of information in the observed graphs $G^{(k)}$ that is relevant to the metric d_{A_c}. Note the similarity of this theorem to the proof of Lemma 7.1, where the arithmetic mean of the adjacency matrices was utilized to quantify information in the graph G_N^* that was relevant to the metric.

This concludes step 3 of the solution approach that is outlined by Section 6. Because we do not have an analytic expression for the graph $G^{*,\mu_{\omega_n f}}$ for a general stochastic block model kernel probability measure $\mu^*_{\omega_n f}$ it is also necessary to understand the largest c eigenvalues of the adjacency matrix of the graph $G^{*,\mu_{\omega_n f}}$ to minimize equation (8.4), which we analyze in the following section. This step was not done when computing the sample Fréchet mean with respect to the Hamming distance in the prior section where the Fréchet mean of a kernel probability measure was known explicitly.

8.1.3 *Quantifying the properties of $G^{*,\mu_{\omega_n f}}$ that are relevant to the metric d_{A_c}.* 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 $G_{\tilde{N}}^{*,\omega_n f}$. The 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 $\sigma_c(\boldsymbol{A}^{*,\omega_n f}_{\tilde{N}})$.

THEOREM 8.3. (The Eigenvalues of the sample Fréchet mean of stochastic block models)$\forall \epsilon > 0$, $\exists n^* \in \mathbb{N}$ such that for all $n > n^*$,

$$\lim_{\tilde{N}\to\infty} \left\| \sigma_c(\boldsymbol{A}^{*,\omega_n f}_{\tilde{N}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right] \right\|_2 < \epsilon \quad a.s. \tag{8.8}$$

Proof. The proof is in Appendix C. □

Theorem 8.4 then provides an estimate for $\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]$ in terms of the kernel function f since we do not have a closed form expression for this term. It should also be noted that the following theorem is a small modification of Theorem 2.4 in [21].

Let $\mu_{\omega_n f} \in \mathscr{M}(\mathscr{G})$ 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 8.4. For every $1 \le i \le c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] = \lambda_i(\boldsymbol{B}^*) + \mathscr{O}(\sqrt{\omega_n}), \tag{8.9}$$

where

$$\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)} \tag{8.10}$$

and

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b^{*,(1)}_{j,l} = \sqrt{\theta_j\theta_l}\, n\omega_n \int_0^1 r_j(x) r_l(x) dx = \begin{cases} \theta_j n\omega_n & j = l \\ 0 & j \neq l \end{cases} \tag{8.11}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b^{*,(2)}_{j,l} = \theta_i^{-2}\sqrt{\theta_j\theta_l} \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx. \tag{8.12}$$

Proof. The proof is in B. This is a modification of Theorem 2.4 from [21]. □

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 the eigenvalues of $\boldsymbol{B}^*$ are not explicit in terms of the parameters, $\boldsymbol{p}$, q and $\boldsymbol{s}$ and we instead rely on numerical estimates.

8.1.4 *Determining the probabilitiy measure $\mu^*_{\omega_n f}$ from equation (8.4).* The conclusion of the prior steps is that to solve equation (8.4) we may use the estimates provided by Theorems 8.2, 8.3 and 8.4 and minimize the following objective,

$$\mu^*_{\omega_n f} = \underset{\mu_{\omega_n f} \in \mathcal{M}_s(\mathcal{G})}{\operatorname{argmin}} \sum_{i=1}^{c} \left| \lambda_i(\boldsymbol{B}^*) - \frac{1}{N} \sum_{k=1}^{N} \lambda_i(\boldsymbol{A}^{(k)}) \right|^2, \tag{8.13}$$

where $\boldsymbol{B}^*$ is defined as in Theorem 8.4.

Remark 8.7. We have not yet specified the choice for ω_n, a necessary component when defining the stochastic block model kernel probability measure. While several different choices of ω_n will suffice, we elect to choose ω_n such that the expected density of graphs sampled according to the canonical stochastic block model kernel probability measure is equivalent to the average density of the graphs in the data set. In this way, we align not only the largest eigenvalues of the graphs in Corollary 8.1 but also the density of the two graphs.

We suggest a gradient descent algorithm to determine the correct canonical stochastic block model kernel function $f(x, y; \boldsymbol{p}, q, \boldsymbol{s})$ where the gradient is computed with respect to the parameters $\boldsymbol{p}$ and q.

The result of minimizing equation (8.13) is a probability measure $\mu^*_{\omega_n f}$ defined by the values of the kernel function $\omega_n f$ on $\{\frac{i}{n}\}_{i=1}^n \times \{\frac{j}{n}\}_{j=1}^n$. Recall that our goal is to determine a graph that is close to the sample Fréchet mean graph G^*_N defined by equation (8.1). Although Theorem 8.4 provides a good estimate of the largest c eigenvalues of the Fréchet mean graph, $G^{*,\mu^*_{\omega_n f}}$, given the measure $\mu^*_{\omega_n f}$ we have yet to provide a graph that achieves the eigenvalues specified by Theorem 8.4. We address this issue in the next section by taking an arbitrarily sized sample from the recovered probability measure $\mu^*_{\omega_n f}$ and computing an appropriate graph-valued statistic.

8.1.5 *Theoretical analysis of the set mean graph from a sample distributed according to $\mu_{\omega_n f}$.* Given a canonical stochastic block model kernel, Theorem 8.5 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 a canonical stochastic block model kernel. Define the set mean graph by

$$\widehat{G}^{*,\mu_{\omega_n f}}_{\tilde{N}} = \underset{\tilde{G} \in \{\tilde{G}^{(k)}\}_{k=1}^{\tilde{N}}}{\operatorname{argmin}} \frac{1}{\tilde{N}} \sum_{k=1}^{\tilde{N}} d^2_{A_c}(\tilde{G}, \tilde{G}^{(k)}), \tag{8.14}$$

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

Theorem 8.5. (Convergence in probability of the truncated spectrum of the set mean graph) $\forall \epsilon > 0$,

$$\lim_{n \to \infty} P\left(\left\| \sigma_c(\hat{\boldsymbol{A}}^{*,\mu_{\omega_n f}}_{\tilde{N}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) \right] \right\|_2 > \epsilon \right) = 0. \tag{8.15}$$

Proof. The proof is in Appendix E. □

Applying this theorem to the solution of equation (8.13) shows a method by which we can provide a graph, $\widehat{G}^{*,\mu_{\omega_n f}}_{\tilde{N}}$, that is, with high probability, close to the target sample Fréchet mean graph G^*_N, defined in equation (8.1).

Due to Theorem 2.3 in [21] (which we restate as Theorem 8.6 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 8.3, we take $\tilde{N} = 5$, which provides good results for the case of finite graph size.

Our final theoretical contribution is the construction of a confidence interval for the set mean graph about the sample Fréchet mean graph. To construct the confidence set we utilize Theorem 8.6, 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.

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

THEOREM 8.6. (Chakrabarty, Chakraborty, and Hazra 2020)

$$\omega_n^{-1/2}\left(\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right) \xrightarrow{d} (Z_i : 1 \le i \le c), \tag{8.16}$$

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

$$Cov(Z_i, Z_j) = 2\int_0^1\int_0^1 r_i(x)r_i(y)r_j(x)r_j(y)f(x,y)dxdy, \tag{8.17}$$

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

Proof. The proof is in [21]. □

Before introducing the confidence set it is beneficial to introduce a few notations. We define 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 $\boldsymbol{Z} = (Z_i : 1 \le i \le c)$, respectively, as

$$F_n(z) = P_n\left((-\infty, z_1] \times ... \times (-\infty, z_c]\right), \tag{8.18}$$

$$F(z) = P\left((-\infty, z_1] \times ... \times (-\infty, z_c]\right). \tag{8.19}$$

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

$$\lim_{n\to\infty} F_n(z) = F(z), \tag{8.20}$$

since $F(z)$ is continuous everywhere (Appendix A).

For the construction of a confidence set, it is easier to work with the probabilities and random vectors directly and so equation (8.20) takes the form

$$\lim_{n\to\infty} P_n\left(\omega_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right) \preccurlyeq z) = P(\boldsymbol{Z} \preccurlyeq z). \tag{8.21}$$

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 8.1. Let $\alpha > 0$. Let $\epsilon \in \mathbb{R}$ such that $0 < \epsilon < \alpha$. Take $0 \preccurlyeq z_\alpha \in \mathbb{R}^c$ such that

$$1 - \alpha < P(-z_\alpha \preccurlyeq \boldsymbol{Z} \preccurlyeq z_\alpha) - \mathscr{E}. \tag{8.22}$$

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 $\mu_{\omega_n f}$ satisfying the statement of Corollary 8.1 for the graph $G^*_{N_1}$. 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 8.14) with adjacency matrix $\widehat{\boldsymbol{A}}^*_{N_2}$. For $N_2 = 1$

$$1 - \alpha < P_n\left(d_{A_c}\left(G^*_{N_1}, \widehat{G}^*_1\right) < \omega_n^{1/2}||z_\alpha||_2\right). \tag{8.23}$$

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

Proof. We first note that

$$d_{A_c}\left(G^*_{N_1}, \widehat{G}^*_{N_2}\right) = ||\sigma_c(\boldsymbol{A}^*_{N_1}) - \sigma_c(\widehat{\boldsymbol{A}}^*_{N_2})||_2. \tag{8.24}$$

We will from here on be working with the eigenvalues of the adjacency matrices rather than the graphs themselves. Now, due to equation (8.21) we have that there exists an n^* such that for all $n > n^*$

$$1 - \alpha < P(-z_\alpha \preccurlyeq \boldsymbol{Z} \preccurlyeq z_\alpha) - \epsilon \tag{8.25}$$

$$< P_n\left(-z_\alpha \preccurlyeq \omega_n^{-1/2}\left(\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right) \preccurlyeq z_\alpha\right). \tag{8.26}$$

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,

$$1 - \alpha < P_n\left(-z_\alpha \preccurlyeq \omega_n^{-1/2}\left(\sigma_c(\widehat{\boldsymbol{A}}^*_1) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right) \preccurlyeq z_\alpha\right). \tag{8.27}$$

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

$$1 - \alpha < P_n\left(-z_\alpha \preccurlyeq \omega_n^{-1/2}\left(\sigma_c(\widehat{\boldsymbol{A}}^*_1) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right) \preccurlyeq z_\alpha\right) \tag{8.28}$$

$$= P_n\left(-\omega_n^{1/2} z_\alpha \preccurlyeq \sigma_c(\widehat{\boldsymbol{A}}^*_1) - \sigma_c(\boldsymbol{A}^*_{N_1}) + \sigma_c(\boldsymbol{A}^*_{N_1}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right] \preccurlyeq \omega_n^{1/2} z_\alpha\right) \tag{8.29}$$

$$\leq P_n\left(||\sigma_c(\widehat{\boldsymbol{A}}^*_1) - \sigma_c(\boldsymbol{A}^*_{N_1}) + \sigma_c(\boldsymbol{A}^*_{N_1}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \leq \omega_n^{1/2}||z_\alpha||_2\right), \tag{8.30}$$

where we have taken a norm of the argument in the probability from equation (8.29) resulting in (8.30). Continuing,

$$1-\alpha \le P_n\left(||\sigma_c(\widehat{\boldsymbol{A}}_1^*)-\sigma_c(\boldsymbol{A}_{N_1}^*)||_2 + ||\sigma_c(\boldsymbol{A}_{N_1}^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \le \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right) \tag{8.31}$$

$$\le P_n\left(||\sigma_c(\widehat{\boldsymbol{A}}_1^*)-\sigma_c(\boldsymbol{A}_{N_1}^*)||_2 \le \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2 \wedge ||\sigma_c(\boldsymbol{A}_{N_1}^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \le \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right) \tag{8.32}$$

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

$$1-\alpha \le P_n\left(A\right)P_n\left(B|A\right). \tag{8.33}$$

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

$$\frac{1-\alpha}{P_n\left(B|A\right)} < P_n\left(A\right). \tag{8.34}$$

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

$$1-\alpha < P_n\left(A\right) \tag{8.35}$$

$$= P_n\left(||\sigma_c(\widehat{\boldsymbol{A}}_1^*)-\sigma_c(\boldsymbol{A}_{N_1}^*)||_2 \le \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right) \tag{8.36}$$

$$= P_n\left(d_{A_c}(\widehat{G}_1^*, G_{N_1}^*) \le \omega_n^{1/2}||\boldsymbol{z}_\alpha||_2\right), \tag{8.37}$$

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.

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 (8.1), which we give the pseudo-code for below.

8.2 *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 $\mu_{\omega_n f}^*$ by solving equation (8.13).
2. Estimate $G_{\tilde{N},\mu_{\omega_n f}}^*$ using Theorem 8.5 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 Algorithm 1 and use this estimate in Algorithm 2.

Algorithm 1. Determine c for the approximate sample Fréchet mean

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

Compute the arithmetic average spectrum of graphs in M as $\bar{\boldsymbol{\lambda}} = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\lambda}^{(k)}$.
Initialize $i = 0$.
Do
    $i = i + 1$
    Initialize $r = \bar{\boldsymbol{\lambda}}(i)$
    Initialize the semi-circle probability density function (see e.g. [6]), as $s(\lambda; r)$ where r is the radius.
    Assume $\bar{\boldsymbol{\lambda}}(j) \sim s(\lambda; r)$ for $j = i, ..., n$.
    Determine the PDF of the K largest order statistics with a sample size $n - i$, $\lambda_{(n-i)}, ..., \lambda_{(n-i-K+1)}$.
    Compute the expected value of the K largest order statistics from the PDF $s(\lambda; r)$ with a sample size of $n - i$.
    With sample size $n - i$, compute the standard deviation of the K largest order statistics, $\sigma_{n-i}, ..., \sigma_{n-i-K+1}$
While $|\bar{\boldsymbol{\lambda}}(1 + i) - \mathbb{E}\left[\lambda_{(n-i)}\right]| > \sigma_{n-i} \vee ... \vee |\bar{\boldsymbol{\lambda}}(K + i) - \mathbb{E}\left[\lambda_{(n-i-K)}\right]| > \sigma_{n-i-K+1}$
Return: $c = i - 1$

Algorithm 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. [2, 6, 24] and references therein). This assumption need not be true of the graphs in the sample $\{G^{(k)}\}_{k=1}^{N}$ and this algorithm will still provide an estimate of c. 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. Algorithm 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 and, inherently, always compare all graphs using this choice of c for the metric d_{A_c}.

Note, it could very well be possible that there exist graphs $G^{(j)}$ and $G^{(k)}$ in our dataset M that have a 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.

Algorithm 2. Approximate sample Fréchet mean

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

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

Approximate c via Algorithm 1. and determine $\boldsymbol{s}$ (see Remark 8.1).

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

Randomly initialize $\boldsymbol{p}$

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

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

  Estimate the gradient of the objective in equation (8.13) via centered differences.

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

  Update q such that $||f(x, y; \boldsymbol{p}, \boldsymbol{Q}, s)||_1 = 1$. This ensures the expected density of the graphs sampled is given by $\bar{\rho}_n$ as stated in Remark 8.7

end while

Estimate G_N^* as $G_{\tilde{N}}^{*,\mu_{\omega_n f}}$ (see Theorem 8.5)

Return: $G_{\tilde{N}}^{*,\mu_{\omega_n f}}$.

We acknowledge that these issues persist throughout our algorithm and use our estimate of c from Algorithm 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.

Corollary 8.1 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 $\bar{\rho}_n$.

8.3 *Experimental validation*

8.3.1 *Assessment, validation and comparison.* For sets of large graphs, the ground truth for the Fréchet mean problem is unrealistic to compute (it requires about $\Omega\left(n^2 2^{n^2}\right)$ operations for graphs of size n). Instead, we compare the eigenvalues adjacency matrix that results from Algorithm 2 to the arithmetic mean of the eigenvalues of the data set because this provides the optimal spectra of the adjacency matrix (see Theorem 8.2).

One may consider comparing the Fréchet mean computed here to a Fréchet mean computed with respect to the Hamming distance for which several optimization algorithms have been proposed (e.g. [9, 18, 30, 41, 43]). While this comparison may be feasible, it is uninformative as the Fréchet mean with respect to the Hamming 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.

8.3.2 *Choice of the datasets.* Numerous graph-valued databases have recently been made publicly available [60, 61, 66, 67]. For each database, the mean graph has not been provided for any choice of metric. Consequently, we feel that computing the Fréchet mean for these datasets provides little scientific value for validating our method. Instead, we present the results of experiments conducted on synthetic datasets generated using ensembles of random graphs and explore the consequences of differing estimates of c, the number of eigenvalues to consider when comparing graphs using d_{A_c}.

Ensembles of random graphs capture prototypical features of existing real-world networks. Because our theoretical analysis and associated algorithms rely on the 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 different from those of the block models.

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 : Variable community size stochastic block model
μ_2 : Barabási–Albert
μ_3 : Watts-Strogatz

Note that μ_2 and μ_3 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 $\widehat{G}^{*,\mu_{\omega_n f}}_{\tilde{N}}$.

8.3.3 *Variable community size approximate sample Fréchet mean.* The probability measure in this section is associated with a variable community size stochastic block model. The parameters for the model are $\boldsymbol{p} = [0.4, 0.5, 0.6, 0.3, 0.37, 0.65, 0, ...]$, $\boldsymbol{s} = \left[\frac{160}{600}, \frac{100}{600}, \frac{60}{600}, \frac{120}{600}, \frac{85}{600}, \frac{75}{600}, 0, ...\right]$, and $q = 0.08$.

Fig. 3 is a visual depiction of a graph from M_1 as compared to the approximate sample Fréchet mean graphs resulting from Algorithm 2 when considering two different choices of c for the metric d_{A_c}. Here we consider both d_{A_3} and d_{A_6} for comparison purposes.
The metric d_{A_3} emphasizes that the estimate of c need not be exact when implementing Alg 2.

A graph from M_1 is chosen in place of the ground truth for the Fréchet mean for comparison purposes only. While the visual comparison of the graphs is of interest, the primary metric of note is whether the largest c eigenvalues (where c is either 3 or 6) of the graphs are similar to the arithmetic mean of the largest c eigenvalues of the graphs in the sample as guaranteed by Theorem 8.2 which can be seen in Fig 4.

Despite the visual differences between a graph in the sample and the adjacency matrices of the sample Fréchet mean graphs with respect to the metrics d_{A_3} and d_{A_6}, there is a striking similarity between the largest c eigenvalues for both choices of c. Note, when $c = 6$, the alignment between the spectra of the adjacency matrices despite the obvious difference in the geometry vectors that defined the two graphs. This result provides further evidence for the comments made in remark 8.1 that the choice of $\boldsymbol{s}$ is up to the discretion of the research under the assumption that the graphs in question are sufficiently large.

Recall that in our optimization procedure to determine the correct canonical stochastic block model kernel, equation (8.13), the predicted expected eigenvalues from Theorem 8.4 are aligned with the arithmetic mean of the eigenvalues of the graphs in the sample, Theorem 8.2. However, the predicted

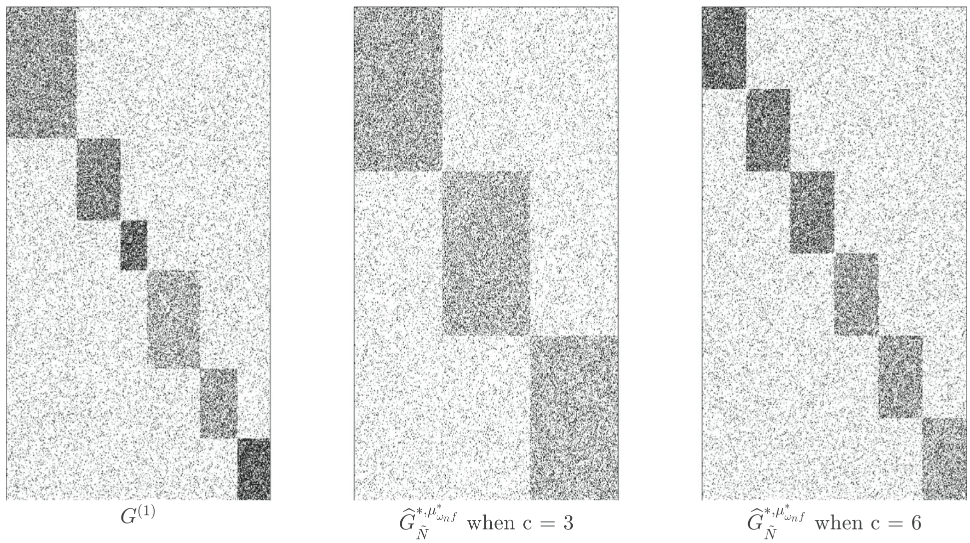

FIG. 3. Visualization of a graph in M_1 (left) and the approximate sample Fréchet mean of M_1, $\widehat{G}^{*,\mu_{\omega n f}}_{\tilde{N}}$, with respect to the metrics d_{A_3} (middle) and d_{A_6} (right).

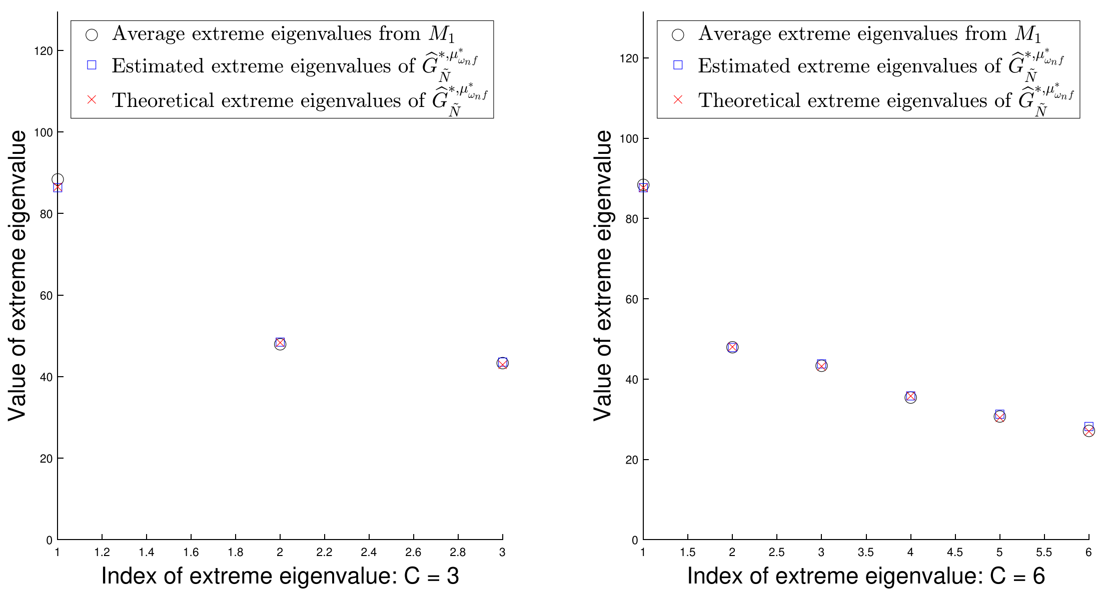

FIG. 4. The average extreme eigenvalues from M_3 (circle). The expected extreme eigenvalues of $\widehat{G}^{*,\mu_{\omega n f}}_{\tilde{N}}$ from Theorem 8.4 (cross). The extreme eigenvalues of $\widehat{G}^{*,\mu_{\omega n f}}_{\tilde{N}}$ (square).

expected eigenvalues only estimate the eigenvalues of the sample Fréchet mean graph of a stochastic block model kernel probability measure in the limit of large graph size. The distinction between the blue markers (the eigenvalues of the resulting graph from Algorithm 2.) and the red markers is due to the finite graph size estimate.

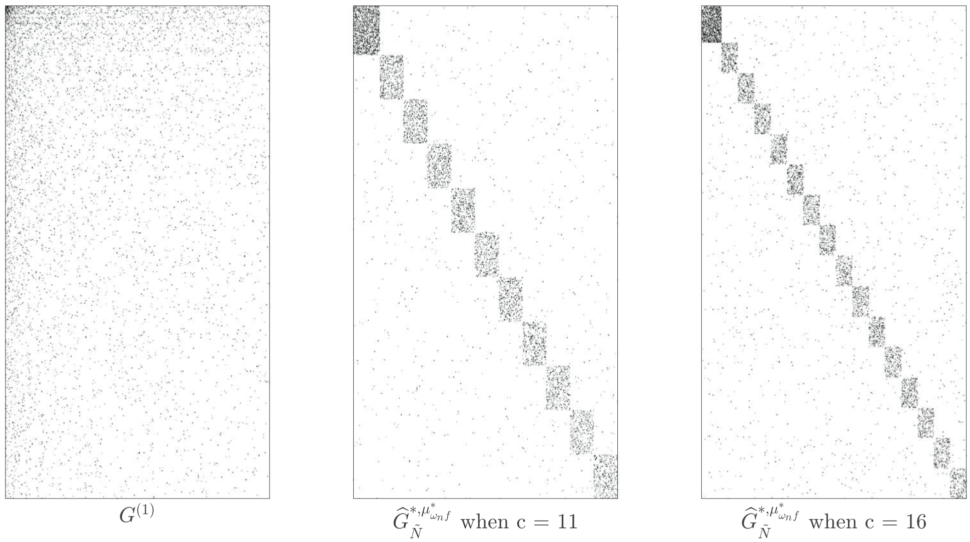

FIG. 5. Visualization of a graph in M_2 (left) and the approximate sample Fréchet mean of M_2, $\widehat{G}^{*,\mu_{\omega nf}}_{\tilde{N}}$, with respect to the metrics $d_{A_{11}}$ (middle) and $d_{A_{16}}$ (right).

8.3.4 *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. 5 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. By choosing $K = 2$ and $K = 4$ for the hyper-parameter in Algorithm 2 we determine two different estimates of c, $c = 11$ and $c = 16$, respectively.

Figure 5 is a visual comparison between a graph in the sample (left) and the approximate sample Fréchet mean graphs that result from Algorithm 2 with respect to the metrics $d_{A_{11}}$ (middle) and $d_{A_{16}}$ (right). We note that there need not be any visual similarity between a graph in the sample set and the sample Fréchet mean of that set since any observation from a distribution need not be similar to the mean of that distribution.

While it is less clear in this case the correct choice of c, as compared with the prior section, we may nonetheless compare the largest c eigenvalues of the graphs adjacency matrices to the arithmetic mean of the largest c eigenvalues of the graphs in the sample (Theorem 8.2).

Furthermore, the estimates in the finite graph size setting are exacerbated by the size of c. We expect that for larger values of c, even larger values of n are needed for the estimate from Theorem 8.4 to be within a pre-determined tolerance. A reason for such an impact of c is that the stochastic block model's expected eigenvalues are determined primarily by the number of vertices within a block, which is inversely proportional to c, the number of communities. For larger c, there are smaller relative community sizes, indicating a larger value of n is needed for the estimates from Theorem 8.4 to be within a certain level of error.

8.3.5 *Watts–Strogatz approximate sample Fréchet mean.* The parameters for the Watts–Strogatz ensemble ([78]) are the number of connected nearest neighbors, $K = 22$ and the probability of rewiring,

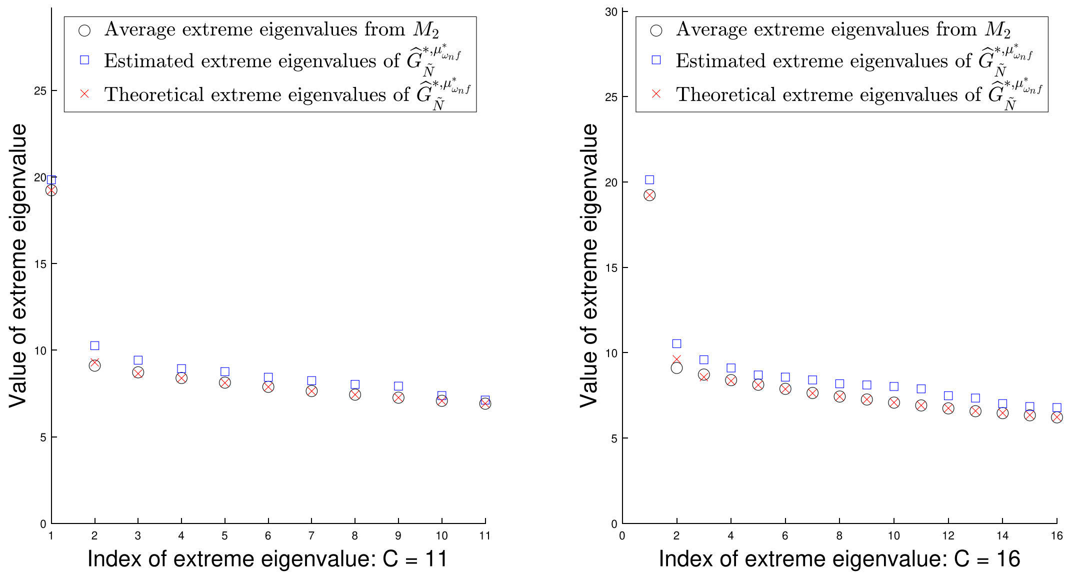

FIG. 6. The average extreme eigenvalues from M_2 (circle). The expected extreme eigenvalues of $\widehat{G}^{*,\mu_{\omega n f}}_{\tilde{N}}$ from Theorem 8.4 (cross). The extreme eigenvalues of$\widehat{G}^{*,\mu_{\omega n f}}_{\tilde{N}}$ (square).

$\beta = 0.7$. Here we again take $K = 2$ and $K = 4$ resulting in two different estimates of c from Algorithm 1, $c = 10$ and $c = 13$, respectively.

Here we see a nice similarity between the adjacency matrices of a graph in M_3 (left) and the approximate sample Fréchet mean graphs with respect to $d_{A_{10}}$ and $d_{A_{13}}$ (Fig. 7).

Visually, $c = 13$ provides a better comparison to a graph in the sample data set; however, this preference is dependent on a particular ordering of the vertices in the adjacency matrix, which, in general, should not impact the choice of c. In the following figure it is shown that for either choice of c, the largest eigenvalues of the approximate sample Fréchet mean graphs are close to the arithmetic mean of the largest eigenvalues of the graphs in the sample.

8.4 *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 [69] for details).

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

$$G_j^* = \underset{G \in \mathscr{G}}{\operatorname{argmin}} \frac{1}{|M_j|} \sum_{G' \in \mathscr{M}_i} d^2_{A_c}(G, G') \tag{8.38}$$

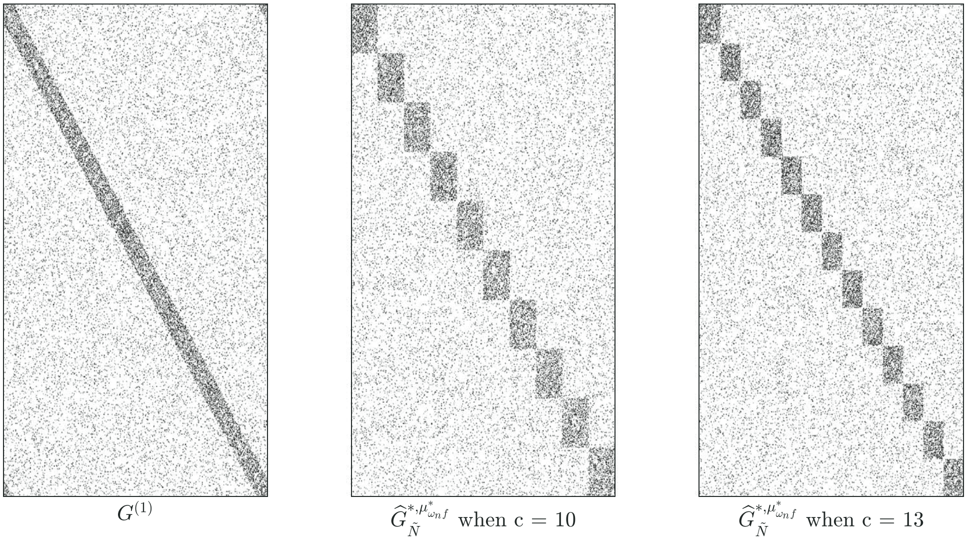

FIG. 7. Visualization of a graph in M_3 (left) and the approximate sample Fréchet mean of M_3, $\widehat{G}^{*,\mu_{\omega nf}}_{\tilde{N}}$, with respect to the metrics $d_{A_{10}}$ (middle) and $d_{A_{13}}$ (right).

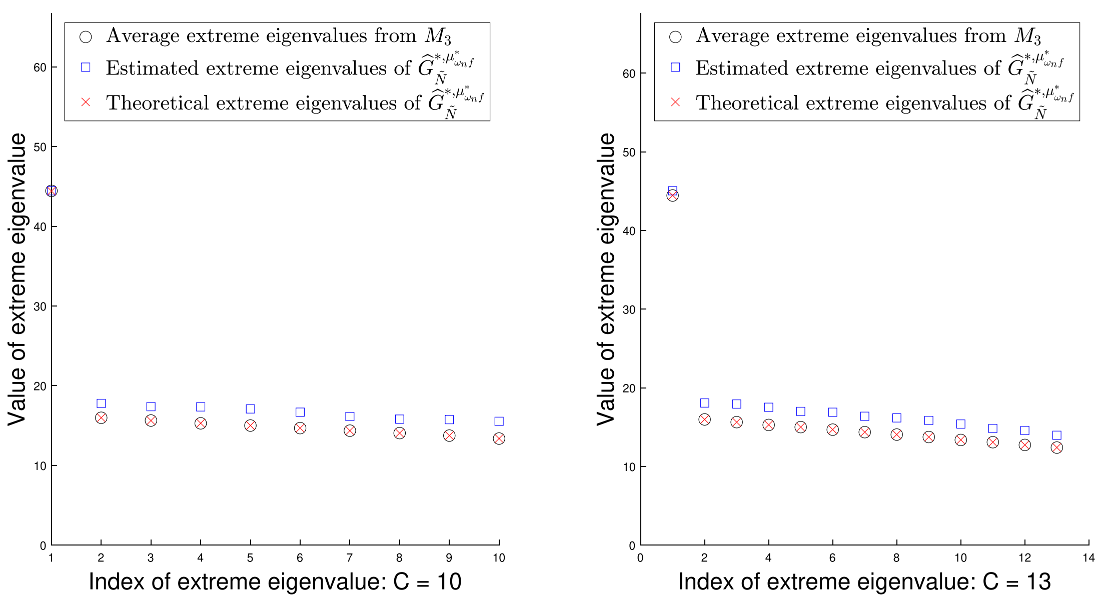

FIG. 8. The average extreme eigenvalues from M_3 (circle). The expected extreme eigenvalues of $\widehat{G}^{*,\mu_{\omega nf}}_{\tilde{N}}$ from Theorem 8.4 (cross). The extreme eigenvalues of $\widehat{G}^{*,\mu_{\omega nf}}_{\tilde{N}}$ (square).

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

$$h_0\left((\mathscr{M}_1,...,\mathscr{M}_K)\,;\{G^{(k)}\}:{}_{k=1}^{N}\right)=\sum_{j=1}^{K}\sum_{G\in\mathscr{M}_j} d^2_{A_c}(G,G_j^*)=\sum_{j=1}^{K}\sum_{G\in\mathscr{M}_j}||\sigma_c(\boldsymbol{A})-\sigma_c(\boldsymbol{A}_j^*)||_2^2, \quad (8.39)$$

where $\boldsymbol{A}$ denotes the adjacency matrix of $G\in\mathscr{M}_j$, by determining the optimal disjoint sets $\mathscr{M}_1,...,\mathscr{M}_K$. Observe that the objective function, h_0, depends on the sets $\mathscr{M}_j$ only through the largest c eigenvalues of $\boldsymbol{A}_j^*$, the adjacency matrix of the representative sample Fréchet mean graph. Theorem 8.2 states that the eigenvalues of $\boldsymbol{A}_j^*$, $\sigma_c(\boldsymbol{A}_j^*)$, are arbitrarily close to the arithmetic average of the largest eigenvalues of the adjacency matrices in the sample set,

$$\bar{\boldsymbol{\lambda}}^{*,(j)}=\frac{1}{|\mathscr{M}_j|}\sum_{G\in\mathscr{M}_j}\sigma_c(\boldsymbol{A}) \quad (8.40)$$

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

$$h_{0,approx}\left((\mathscr{M}_1,...,\mathscr{M}_K)\,;\{G^{(k)}\}:{}_{k=1}^{N}\right)=\sum_{j=1}^{K}\sum_{G\in\mathscr{M}_j}\left\|\sigma_c(\boldsymbol{A})-\bar{\boldsymbol{\lambda}}^{*,(j)}\right\|_2^2, \quad (8.41)$$

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

8.5 *Algorithm*

In our implementation of the classic K-means algorithm, 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 $G\in\mathscr{M}_j$ to the sample Fréchet mean graph of the set $\mathscr{M}_r$, 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 8, 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 $\mathscr{M}_1,...,\mathscr{M}_K$ are updated as this affects the number of extreme eigenvalues in the adjacency matrix of 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 Algorithm 3 is due to the metric d_{A_c} and the observation that the choice of c must be handled carefully for each cluster.

8.5.1 *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)

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, $\mathscr{M}_1, ..., \mathscr{M}_K$
4: **while** Class assignments are changing between iterates **do**
5: For each $\mathscr{M}_j$, estimate c_j as in Algorithm 1
6: For each $\mathscr{M}_j$ compute $\bar{\boldsymbol{\lambda}}^{*,(j)} = \frac{1}{|\mathscr{M}_j|} \sum_{G \in \mathscr{M}_j} \boldsymbol{\lambda}(1 : c_j)$
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 $\mathscr{M}_j$ compute G_j^*, the sample Fréchet mean via Algorithm 2
11: **Return:** each $\mathscr{M}_j$ and G_j^*

graphs (parameters: $P = 12$ nearest neighbors with $\beta = 0.4$ probability of rewiring), and 50 stochastic block model (SBM) graphs (parameters: $\boldsymbol{p} = [0.14, 0.16, 0.1746, 0.2, 0.22, 0.24, 0, ...]$, $q = 0.01$, and $\boldsymbol{s} = [\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, 0, ...]$) 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, Algorithm 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	SW True
SBM Pred.	1(0.00)	0	0
BA Pred.	0	0.9(0.0909)	0.19(0.0909)
SW Pred.	0.23(0.1789)	0	0.77(0.1789)

One possible explanation for the misclassification of the Barabasi–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 Barabasi–Albert graphs, Algorithm 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, Algorithm 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 $\mathscr{M}_j$ versus $\mathscr{M}_p$ 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 Barabasi–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 Algorithm 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.

9. Conclusion

The modern developments of statistical graph analysis typically involve the determination of the average of a sample data set. Throughout this manuscript, we have proposed a general solution technique that can be used to determine the sample Fréchet mean graph with respect to many choices of metrics. We have demonstrated the approach for the Hamming distance as well as the pseudometric d_{A_c} for several choices of c.

Several applications of the sample Fréchet mean exist, in fact as many applications for the mean graph exist as applications for the mean of real-valued data. Here we have shown one application, which is the principled implementation of the K-means algorithm with respect to the metric d_{A_c}. Furthermore, we have shown how to determine the mean graph within a cluster upon the termination of the K-means algorithm. In [65] it is shown that a generalization of linear regression can be defined using the Fréchet mean and sample Fréchet mean indicating that the algorithms presented here can be used for implementations of linear regression algorithms in the future. We may also utilize these results to further push the concepts introduced in [56], which introduces a method to sample new graphs around the sample Fréchet mean graph but, in their experiments, the set mean graph was used as an estimate of the sample Fréchet mean instead.

Data availability statement

The data used in this manuscript is available at the repository https://github.com/dafe0926/approx_Graph_Frechet_Mean.

Funding

National Science Foundation (CCF/CIF 1815971).

References

1. Abbe, E., Li, S. & Sly, A. (2020) Learning sparse graphons and the generalized kesten-stigum threshold. *arXiv preprint arXiv:2006.07695*.
2. Abbe, E., Bandeira, A. S. & Hall, G. (2016) Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory*, 471–487.
3. Abbe, E. (2017) Community detection and stochastic block models: recent developments. *J. Mach. Learn. Res.* **18**, 6446–6531.
4. Ambrosio, L., Gigli, N. & Savare, G. (2008) Gradient flows: In Metric Spaces and in the Space of Probability Measures. Basel, Switzerland: Birkhauser Basel.
5. Athreya, A., Cape, J. & Tang, M. (2021) Eigenvalues of stochastic Blockmodel graphs and random graphs with low-rank edge probability matrices. *Indian J. Stat*. Springer. pp. 36–63. https://econpapers.repec.org/article/sprsankha/v_3a84_3ay_3a2022_3ai_3a1_3ad_3a10.1007_5fs13171-021-00268-x.htm.
6. Avrachenkov, K., Cottatellucci, L., and Kadavankandy, A. (2015) 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)*, pp. 537–544.
7. Bacák, M. (2014) Computing medians and means in Hadamard spaces. *SIAM J. Optim.*, **24**, 1542–1566.
8. Baldesi, L., Markopoulou, A. & Butts, C. T. (2018) Spectral graph forge: graph generation targeting modularity. *CoRR abs/1801.01715*.
9. Bardaji, Itziar and Ferrer, Miquel and Sanfeliu, Alberto (2010) Computing the barycenter graph by means of the graph edit distance. *2010 20th International Conference on Pattern Recognition*, Istan-

bul, Turkey, 962–965. https://www.researchgate.net/publication/333675002_Generalized_Median_Graph_via_Iterative_Alternate_Minimizations/citation/downlo.

10. Bhattacharya, R. & Waymire, E. (2016) *A Basic Course in Probability Theory*. New York, NY: Springer.
11. Billera, L. J., Holmes, S. P. & Karen, V. (2001) Geometry of the space of phylogenetic trees advances in applied mathematics. Elsevier, 733–767. https://www.sciencedirect.com/science/article/pii/S0196885801907596.
12. Bodwin, G. & Vempala, S. (2022) A unified view of graph regularity via matrix decompositions. *Random Structures and Algorithms*, **61**, 62–83.
13. Bollobás, B. & Riordan, O. (2009) *Metrics for sparse graphs London Mathematical Society Lecture Note Series*. Cambridge: Cambridge University Press, pp. 211–288.
14. Borgs, C., Chayes, J. T., Cohn, H. & Ganguly, S. (2021) Consistent nonparametric estimation for heavy-tailed sparse graphs. *Ann. Stat.*, **49**, 1904–1930.
15. Borgs, C., Chayes, J. T., Cohn, H. & Lovász, L. M. (2019) Identifiability for graphexes and the weak kernel metric. *Building Bridges II*. Berlin, Heidelberg: Springer, pp. 29–157.
16. Borgs, C., Chayes, J. T., Cohn, H. & Zhao, Y. (2018) An L^p theory of sparse graph convergence II: LD convergence, quotients, and right convergence. *Ann. Probab.*, **46**, 337–396.
17. Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., and Vesztergombi, K. (2012) Convergent sequences of dense graphs ii. Multiway cuts and statistical physics. *Ann. Math.* **176**, 151–219.
18. Boria, Nicolas and Bougleux, Sébastien and Gaüzère, Benoit and Brun, Luc (2019) Generalized median graph via iterative alternate minimizations *International Workshop on Graph-Based Representations in Pattern Recognition*, 99–109. https://www.researchgate.net/publication/333675002_Generalized_Median_Graph_via_Iterative_Alternate_Minimizations/citation/download.
19. Boria, N., Negrevergne, B. & Yger, F. (2020) Fréchet Mean Computation in Graph Space through Projected Block Gradient Descent. *ESANN 2020*. Bruges, France.
20. Bravo-Hermsdorff, Gecia and Gunderson, Lee MA unifying framework for spectrum-preserving graph sparsification and coarsening *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (2019), Curran Associates Inc. pp. 7736–7747.
21. Chakrabarty, A., Chakraborty, A. & Hazra, R. (2020) Eigenvalues outside the bulk of inhomogeneous Erdös-Rényi random graphs. *J. Statist. Phys.* **181**, 1746–1780.
22. Chen, J., Saad, Y. & Zhang, Z. (2021) Graph coarsening: from scientific computing to machine learning. *arXiv preprint arXiv*, **2106**, 11863.
23. Deng, C., Zhao, Z., Wang, Y., Zhang, Z., and Feng, Z. (2020) GraphZoom: a multi-level spectral approach for accurate and scalable graph embedding *The International Conference on Learning Representations (ICLR)* OpenReview.net.
24. Erdős, L., Yau, HT. and Yin, J. (2012) Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154**, 341–407.
25. Fan, J., Fan, Y., Han, X. & Lv, J. (2019) *Asymptotic theory of eigenvectors for large random matrices. arXiv preprint arXiv:1902.06846*.
26. Farkas, I. J. (2001) Spectra of "real-world" graphs: beyond the semicircle law. *Phys. Rev. E (3)*, **64**, 2.
27. Ferguson, D. & Meyer, F. G. (2021) The sample Fréchet mean (or median) graph of sparse graphs is sparse. *arXiv preprint arXiv:2105.14397*.
28. Daniel Ferguson and François G. Meyer (2022) Computation of the Sample Fréchet Mean for Sets of Large Graphs with Applications to Regression *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, pp. 379–387. https://epubs.siam.org/doi/abs/10.1137/1.9781611977172.43.
29. Ferrer, M., Serratosa, F. & Sanfeliu, A. (2005) *Synthesis of Median Spectral Graph*, vol. **3523**. Berlin, Heidelberg: Springer, pp. 139–146.
30. Ferrer, Miquel and Valveny, Ernest and Serratosa, Francesc (2009) Median graph: a new exact algorithm using a distance based on the maximum common subgraph *Pattern Recognit. Lett.* **30**, 579–588.
31. Ferrer, M., Valveny, E., Serratosa, F., Riesen, K., and Bunke, H. (2010) Generalized median graph computation by means of graph embedding in vector spaces. *Pattern Recognit.* **43**, 1642–1655.

32. Flaxman, A., Frieze, A. & Fenner, T. (2003) High degree vertices and eigenvalues in the preferential attachment graph. *Approximation, Randomization, and Combinatorial Optimization*. Berlin Heidelberg: Springer, pp. 264–274.
33. Fréchet, M. (1948) Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré. Analyse Non Linéaire* **10**, 215–310.
34. Frieze, A. & Kannan, R. (1999) *Quick Approximation to Matrices and Applications*, vol. **19**. Combinatorica, pp. 175–220. https://link.springer.com/article/10.1007/s004930050052#citeas.
35. Gao, S., and Caines, P. E. (2019) Spectral representations of graphons in very large network systems control. *2019 IEEE 58th Conference on Decision and Control (CDC)* Nice, France. https://ieeexplore.ieee.org/document/9030220.
36. Ginestet, C. E. (2012) Strong consistency of fréchet sample mean sets for graph-valued random variables. *arXiv preprint arXiv:1204.3183*.
37. Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S., and Kolaczyk, E. D. (2017) Hypothesis testing for network data in functional neuroimaging. *Ann. Appl Stat.* **11**, 725–750.
38. Girvan, M. & Newman, M. E. J. (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, **99**, 7821–7826. https://www.pnas.org/doi/epdf/10.1073/pnas.122653799.
39. Hastie, T., Tibshirani, R. & Friedman, J. (2001) *The Elements of Statistical Learning. Springer Series in Statistics*. New York, NY, USA: Springer.
40. Jain, B. J. (2016) Statistical graph space analysis. *Pattern Recognit.*, **60**, 802–812.
41. Jain, B. J., and Obermayer, K. (2009) Algorithms for the sample mean of graphs. In *International Conference on Computer Analysis of Images and Patterns*, Berlin, Heidelberg: Springer, pp. 351–359.
42. Janson, S. (2013) *Graphons, Cut Norm and Distance, Couplings and Rearrangements NYJM Monographs*. University at Albany, Albany, NY: State University of New York.
43. Jiang, X., Munger, A., and Bunke, H. (2001) On median graphs: properties, algorithms, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 1144–1151.
44. Jin, Y., Loukas, A., and JaJa, J. (2020) Graph coarsening with preserved spectral properties. *International Conference on Artificial Intelligence and Statistics*, PMLR **108**, 4452–4462. https://proceedings.mlr.press/v108/jin20a.html.
45. Jovanović, Irena and Stanić. (2012) Zoran spectral distances of graphs *Linear Algebra Appl.* **436**, 1425–1435.
46. Klopp, O., Tsybakov, A. B. & Verzelen, N. (2017) Oracle inequalities for network models and sparse graphon estimation. *Ann. Stat.*, **45**, 316–354.
47. Klopp, O. & Verzelen, N. (2019) Optimal graphon estimation in cut distance. *Probab. Theory Related Fields*, **174**, 1033–1090.
48. Kohayakawa, Y. (1997) *Szemerédi's regularity lemma for sparse graphs. Foundations of computational mathematics*. Berlin, Heidelberg: Springer. pp. 216–230.
49. Komlós, J., Shokoufandeh, A., Simonovits, M. & Szemerédi, E. (2000) The regularity lemma and its applications in graph theory summer school on theoretical aspects of computer science. 84–112.
50. Le, C. M., Levina, E. & Vershynin, R. (2018) Concentration of random graphs and application to community detection. *Proceedings of the International Congress of Mathematicians (ICM 2018)*. pp. 2925–2943. https://www.worldscientific.com/doi/10.1142/9789813272880_0166.
51. Lee, J. R., Gharan, S. O. & Trevisan, L. (2014) Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM*, **61**, 1, 37.
52. Li, Y., Vinyals, O., Dyer, C., Pascanu, R. & Battaglia, P. (2018) Learning deep generative models of graphs. *arXiv preprint arXiv:1803.03324*.
53. Loukas, Andreas (2019) Graph reduction with spectral and cut guarantees. *J. Mach. Learn. Res.*, 20 **116**, 42.
54. Lovász, L. (2012) Large networks and graph limits. *American Mathematical Soc*. **60**. https://scholar.google.com/scholar?q=Large+networks+and+graph+limits.+AMS&hl=en&as_sdt=0&as_vis=1&oi=scholart.

55. Lovász, L. & Szegedy, B. (2006) Limits of dense graph sequences. *J. Combinatorial Theory*, **96**, 933–957.
56. Lunagómez, S., Olhede, S. C. & Wolfe, P. J. (2020) Modeling network populations via graph distances. *J. Amer. Statist. Assoc.*, **116**, 1–18.
57. Maas, C. (1985) Computing and interpreting the adjacency spectrum of traffic networks. *J. Comput. Appl. Math.*, **12**, 459–466.
58. Meyer, F. G. Sharp threshold for the Fréchet mean (or median) of inhomogeneous Erdős–Rényi random graphs. *arXiv preprint arXiv:2201.11954*.
59. Mossel, Elchanan, Neeman, Joe, and Sly, Allan (2016) Belief propagation, robust reconstruction and optimal recovery of block models *Ann. Appl. Probab.* **26**, 2211–2256.
60. Morris, C., Kriege, N. M., Bause, F., Kersting, K., Mutzel, P. & Neumann, M. (2020) Tudataset: a collection of benchmark datasets for learning with graphs. *arXiv preprint arXiv:2007.08663*.
61. Morris, C.Kriege, N.M, Bause, F., Kersting, K., Mutzel, P., and Neumann, M., TUDatasets: a collection of benchmark datasets for graph classification and regression, https://chrsmrrs.github.io/datasets/.
62. Newman, M. E. J. (2013) Spectral community detection in sparse networks. *arXiv preprint arXiv:1308.6494*.
63. Olhede, S. C., and Wolfe, P. J. (2014) Network histograms and universality of blockmodel approximation. *Proc. Natl Acad Sci* **111**, 14722–14727.
64. Pennec, X. (2006) Intrinsic statistics on riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vision* **25**, 127–154.
65. Petersen, A., and Müller, H.-G. (2019) Fréchet regression for random objects with euclidean predictors. *Ann. Statist.* **47**, 691–719.
66. Research Group on Computer Vision and Artificial Intelligence INF (2019) University of Bern. *IAM Graph Database Repository*. https://fki.tic.heia-fr.ch/databases/iam-graph-database.
67. Riesen, K. and Bunke, H. (2008) 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)*, 287–297.
68. Scott, A. (2011) Szemerédi's regularity lemma for matrices and sparse graphs. *Combin. Probab. Comput.*, Cambridge University Press, **20**, 455–466. https://www.cambridge.org/core/journals/combinatorics-probability-and-computing/article/abs/szemeredis-regularity-lemma-for-matrices-and-sparse-graphs/476D84CD7176709E9559FEB36EB8BA62.
69. Shalev-Shwartz, S. & Ben-David, S. (2014) *Understanding Machine Learning: From Theory to Applications*. Cambridge: Cambridge University Press.
70. Shine, A., and Kempe, D. (2019) Generative graph models based on Laplacian spectra. *WWW '19: The World Wide Web Conference* NY, United States: Association for Computing Machinery New York, 1691–1701.
71. Singh, A. & Humphries, M. (2015) Finding communities in sparse networks. *Sci Rep*, **5** 8828.
72. Stewart, G. & Sun, J. (1990) *Matrix perturbation Theory*. Academic Press.
73. Szegedy, B. (2011) Limits of kernel operators and the spectral regularity lemma. *Eur. J. Combin.* **32**, 1156–1167. Homomorphisms and Limits.
74. Szemerédi, E. (1975) *Regular partitions of graphs Tech*. Report CS-TR-75-489. Stanford, CA: Stanford University, Department of Computer Science.
75. Tao, T. (2012) Topics in random matrix theory. Providence, RI: Graduate Studies in Mathematics; v. 132. American Mathematical Society.
76. Vu, V. (2014) Combinatorial problems in random matrix theory. *Proceedings ICM* **4**, 489–508.
77. Vu, V. (2021) Recent progress in combinatorial random matrix theory. *Probab Surv*, **18**, 179–200.
78. Watts, D. & Strogatz, S. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442.
79. Wills, P. & Meyer, F. G. (2020) Metrics for graph comparison: a practitioner's guide. *PLoS One*, **15**, 1–54.
80. Wilson, R. C., and Zhu, P. (2008) A study of graph spectra for comparing graphs and trees. *Pattern Recognit.* **41**, 2833–2841.
81. Xu, J (2018) Rates of convergence of spectral methods for graphon estimation *International Conference on Machine Learning*, PMLR pp. 5433–5442. https://proceedings.mlr.press/v80/xu18a.html.

82. Yun, S. & Proutiere, A. Accurate community detection in the stochastic block model via spectral algorithms. *arXiv preprint arXiv:1412.7335*.
83. Zhand, X., Nadakuditi, R. & Newman, M. Spectra of random graphs with community structure and arbitrary degrees. *arXiv preprint arXiv:1310.0046*.
84. Zhu, Y. (2019) A graphon approach to limiting spectral distributions of Wigner-type matrices. *Random Struct Algorithms*, **56**. https://www.researchgate.net/publication/336784977_A_graphon_approach_to_limiting_spectral_distributions_of_Wigner-type_matrices.

Appendix

We split the appendix into five sections. Appendix A establishes a few classic results that we refer to in our proofs. Appendix B outlines three methods of estimating the expected value of the largest eigenvalues. The proof of our primary contribution, Theorem 8.1, is contained in Appendix C. Within this appendix we also prove Theorem 8.3 and Theorem 8.4 since these are necessary results for our proof of Theorem 8.1. Appendix D and Appendix E are short appendices in which we prove Theorems 8.2 and 8.5, respectively.

A. Classic results

Theorem A.1. (Weyl–Lidskii) Let $\boldsymbol{H}$ be a self-adjoint operator on a Hilbert space $\mathscr{H}$. Let $\boldsymbol{A}$ be a bounded operator on $\mathscr{H}$ Let $\sigma(\boldsymbol{H})$ and $\sigma(\boldsymbol{H}+\boldsymbol{A})$ denote the spectra of $\boldsymbol{H}$ and $(\boldsymbol{H}+\boldsymbol{A})$, respectively. Then

$$\sigma(\boldsymbol{H}+\boldsymbol{A}) \subset \{\lambda : dist(\lambda, \sigma(\boldsymbol{H})) \leq ||\boldsymbol{A}||\} \tag{A1}$$

where $||\boldsymbol{A}||$ denotes the operator norm of $\boldsymbol{A}$.

Proof. These are standard bounds that can be found in many good books on matrix perturbation theory (e.g. [72]). □

Let P_n be probabilities on the Borel σ-field of $\mathbb{R}^c$ and suppose $P_n \to P$ weakly.

Theorem A.2. (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. This is a standard equivalence for convergence in distribution found in e.g. [10]. □

B. Approximating expected eigenvalues of stochastic block model graphs

We first introduce a recent theorem from [21] that discusses an estimate of the expected eigenvalues of an inhomogeneous Erdős–Rényi random graph. Let $\mu_{\omega_n f} \in \mathscr{M}(\mathscr{G})$ 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 B.1. (Chakrabarty, Chakraborty, Hazra 2020) For every $1 \leq i \leq c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] = \lambda_i(\boldsymbol{B}) + \mathscr{O}\left(\sqrt{\omega}_n + \frac{1}{n\omega_n}\right), \tag{B1}$$

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

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

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

Proof. The proof is in [21]. □

The authors in [21] 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 [21] 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 B.1. Let $\boldsymbol{B}$ be as defined in Theorem B.1. Define the matrices $\boldsymbol{B}^{*,(1)}$ and $\boldsymbol{B}^{*,(2)}$ to have components j, l as

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b_{j,l}^{*,(1)} = \sqrt{\theta_j\theta_l} n\omega_n \int_0^1 r_j(x) r_l(x) dx = \begin{cases} \theta_j n\omega_n & j = l \\ 0 & j \neq l \end{cases} \tag{B2}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b_{j,l}^{*,(2)} = \theta_i^{-2}\sqrt{\theta_j\theta_l} \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx. \tag{B3}$$

Let $\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}$. For any i,

$$|\lambda_i(\boldsymbol{B}) - \lambda_i(\boldsymbol{B}^*)| = \mathscr{O}(\omega_n). \tag{B4}$$

Proof. The proof is straightforward in that we omit all contributions to $b_{j,l}$ that are $\mathscr{O}(\omega_n)$ 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 A.1 from A). 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 $\mathscr{O}(\frac{1}{n})$. With these ideas in mind, we proceed with the computations. We begin by considering the entries of $\boldsymbol{B}$ in two parts, define

$$b_{j,l}^{(1)} = \sqrt{\theta_j\theta_l} n\omega_n \boldsymbol{e}_j^T \boldsymbol{e}_l \tag{B5}$$

$$b_{j,l}^{(2)} = \theta_i^{-2}\sqrt{\theta_j\theta_l}(n\omega_n)^{-1}\boldsymbol{e}_j^T \mathbb{E}\left[(\boldsymbol{A} - \mathbb{E}\left[\boldsymbol{A}\right])^2\right]\boldsymbol{e}_l \tag{B6}$$

$$b_{j,l} = b_{j,l}^{(1)} + b_{j,l}^{(2)}. \tag{B7}$$

For comparison, we also write again the components of $\boldsymbol{B}^{*,(1)}$ and $\boldsymbol{B}^{*,(2)}$

$$b_{j,l}^{*,(1)} = \sqrt{\theta_j\theta_l}n\omega_n \int_0^1 r_j(x)r_l(x)dx \tag{B8}$$

$$b_{j,l}^{*,(2)} = \theta_i^{-2}\sqrt{\theta_j\theta_l}\int_0^1 r_j(x)r_l(x)\int_0^1 f(x,y)dydx. \tag{B9}$$

We will show that for every j,l that

$$|b_{j,l}^{(1)} - b_{j,l}^{*,(1)}| = \mathscr{O}(\omega_n) \tag{B10}$$

$$|b_{j,l}^{(2)} - b_{j,l}^{*,(2)}| = \mathscr{O}(\omega_n). \tag{B11}$$

Observe that for all j,l we have

$$\int_0^1 r_j(x)r_l(x)dx = \lim_{n\to\infty}\sum_{i=1}^n \frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right) \tag{B12}$$

$$= \lim_{n\to\infty}\sum_{i=1}^n e_j(m)e_l(m) \tag{B13}$$

$$= \lim_{n\to\infty} e_j^T e_l. \tag{B14}$$

The interpretation is that $e_j^T e_l$ is an approximation to the integral using the right endpoints of the intervals. Denote by

$$R(n) = \left|\sum_{i=1}^n \frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right) - \int_0^1 r_j(x)r_l(x)dx\right|, \tag{B15}$$

the error in the right endpoint approximation. The convergence rate for the right endpoint rule is $\mathscr{O}(\frac{1}{n})$ for functions that are piecewise Lipschitz with finitely many discontinuities. Consequently,

$$n\omega_n R(n) = n\omega_n \mathscr{O}\left(\frac{1}{n}\right) = \mathscr{O}(\omega_n). \tag{B16}$$

This indicates that for every j,l,

$$|b_{j,l}^{(1)} - b_{j,l}^{*,(1)}| = \left|\sqrt{\theta_j\theta_l}n\omega_n \boldsymbol{e}_j^T\boldsymbol{e}_l - \sqrt{\theta_j\theta_l}n\omega_n\int_0^1 r_j(x)r_l(x)dx\right| \tag{B17}$$

$$= \sqrt{\theta_j\theta_l}n\omega_n\left|\boldsymbol{e}_j^T\boldsymbol{e}_l - \int_0^1 r_j(x)r_l(x)dx\right| \tag{B18}$$

$$= \sqrt{\theta_j\theta_l}n\omega_n R(n) \tag{B19}$$

$$= \mathscr{O}(\rho_n). \tag{B20}$$

We now turn our attention to the second component of $\boldsymbol{B}$ and analyze $b^{(2)}_{j,l}$,

$$b^{(2)}_{j,l} = \theta_i^{-2}\sqrt{\theta_j\theta_l}(n\omega_n)^{-1}\boldsymbol{e}_j^T\mathbb{E}\left[(\boldsymbol{A}-\mathbb{E}[\boldsymbol{A}])^2\right]\boldsymbol{e}_l. \tag{B21}$$

To understand $b^{(2)}_{j,l}$ it is necessary to understand $\mathbb{E}\left[(\boldsymbol{A}-\mathbb{E}[\boldsymbol{A}])^2\right]$. Let $a_{m,k} = \boldsymbol{A}_{m,k}$, the $(m,k)^{th}$ entry of $\boldsymbol{A}$,

$$\mathbb{E}\left[\left((\boldsymbol{A}-\mathbb{E}[\boldsymbol{A}])^2\right)_{m,k}\right] = \mathbb{E}\left[\sum_{w=1}^{n} a_{m,w}a_{w,k} - a_{m,w}\mathbb{E}\left[a_{w,k}\right] - \mathbb{E}\left[a_{m,w}\right]a_{w,k} + \mathbb{E}\left[a_{m,w}\right]\mathbb{E}\left[a_{w,k}\right])\right] \tag{B22}$$

$$= \sum_{w=1}^{n}\mathbb{E}\left[a_{m,w}a_{w,k}\right] - \mathbb{E}\left[a_{m,w}\right]\mathbb{E}\left[a_{w,k}\right] - \mathbb{E}\left[a_{m,w}\right]\mathbb{E}\left[a_{w,k}\right] + \mathbb{E}\left[a_{m,w}\right]\mathbb{E}\left[a_{w,k}\right]) \tag{B23}$$

$$= \begin{cases}\sum_{w=1}^{n}\mathbb{E}\left[a_{m,w}\right](1-\mathbb{E}\left[a_{m,w}\right]) & m=k\\ 0 \quad \text{else.}\end{cases} \tag{B24}$$

Recall that $\mathbb{E}\left[a_{m,w}\right] = \omega_n f(\frac{m}{n},\frac{w}{n})$. We may compute the term $(n\omega_n)^{-1}\boldsymbol{e}_j^T\mathbb{E}\left[(\boldsymbol{A}-\mathbb{E}[\boldsymbol{A}])^2\right]\boldsymbol{e}_l$ which appears in the definition of $b^{(2)}_{j,l}$ with this expression for $\mathbb{E}\left[(\boldsymbol{A}-\mathbb{E}[\boldsymbol{A}])^2\right]$ as

$$(n\omega_n)^{-1}\boldsymbol{e}_j^T\mathbb{E}\left[(\boldsymbol{A}-\mathbb{E}[\boldsymbol{A}])^2\right]\boldsymbol{e}_l = \sum_{i=1}^{n}\boldsymbol{e}_j(i)\boldsymbol{e}_l(i)\sum_{w=1}^{n}\mathbb{E}\left[a_{i,w}\right](1-\mathbb{E}\left[a_{i,w}\right]) \tag{B25}$$

$$= \sum_{i=1}^{n}\frac{1}{n^2\omega_n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{w=1}^{n}\omega_n f\left(\frac{i}{n},\frac{w}{n}\right)\left(1-\omega_n f\left(\frac{i}{n},\frac{w}{n}\right)\right) \tag{B26}$$

$$= \sum_{i=1}^{n}\frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{w=1}^{n}\frac{1}{n}f\left(\frac{i}{n},\frac{w}{n}\right)\left(1-\omega_n f\left(\frac{i}{n},\frac{w}{n}\right)\right). \tag{B27}$$

We consider the final term in two separate parts,

$$\sum_{i=1}^{n}\frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{w=1}^{n}\frac{1}{n}f\left(\frac{i}{n},\frac{w}{n}\right)\left(1-\omega_n f\left(\frac{i}{n},\frac{w}{n}\right)\right) = \sum_{i=1}^{n}\frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{w=1}^{n}\frac{1}{n}f\left(\frac{i}{n},\frac{w}{n}\right) \tag{B28}$$

$$-\omega_n\sum_{i=1}^{n}\frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{w=1}^{n}\frac{1}{n}f\left(\frac{i}{n},\frac{w}{n}\right)^2 \tag{B29}$$

We will show that

$$\omega_n\sum_{i=1}^{n}\frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{w=1}^{n}\frac{1}{n}f\left(\frac{i}{n},\frac{w}{n}\right)^2 = \mathcal{O}(\omega_n). \tag{B30}$$

We then show that

$$\left|\sum_{i=1}^{n}\frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{w=1}^{n}\frac{1}{n}f\left(\frac{i}{n},\frac{w}{n}\right)-\int_0^1 r_j(x)r_l(x)\int_0^1 f(x,y)dydx\right|=\mathscr{O}\left(\frac{1}{n}\right). \tag{B31}$$

The conclusion is then that

$$\left|b_{j,l}^{(2)}-b_{j,l}^{*,(2)}\right|=\mathscr{O}(\omega_n). \tag{B32}$$

To begin, recall that the expression for $f(\frac{i}{n},\frac{w}{n})$ in terms of the eigenvalues and eigenfunctions is

$$f(x,y)=\sum_{k=1}^{c}\theta_k r_k(x)r_k(y) \tag{B33}$$

$$f(x,y)^2=\left(\sum_{k=1}^{c}\theta_k r_k(x)r_k(y)\right)^2 \tag{B34}$$

$$=\sum_{k=1}^{c}\sum_{m=1}^{c}\theta_k r_k(x)r_k(y)\theta_m r_m(x)r_m(y). \tag{B35}$$

We have the following expression which will be used throughout our computations,

$$\sum_{w=1}^{n}\frac{1}{n}f\left(\frac{i}{n},\frac{w}{n}\right)^2=\sum_{w=1}^{n}\frac{1}{n}\sum_{k=1}^{c}\sum_{m=1}^{c}\theta_k r_k(x)r_k(y)\theta_m r_m(x)r_m(y) \tag{B36}$$

$$=\sum_{k=1}^{c}\sum_{m=1}^{c}\theta_k\theta_m\sum_{w=1}^{n}\frac{1}{n}r_k(x)r_k(y)r_m(x)r_m(y). \tag{B37}$$

Consider now just (B29)

$$\omega_n\sum_{i=1}^{n}\frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{w=1}^{n}\frac{1}{n}f\left(\frac{i}{n},\frac{w}{n}\right)^2 \tag{B38}$$

$$=\omega_n\sum_{i=1}^{n}\frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{k=1}^{c}\sum_{m=1}^{c}\theta_k\theta_m\sum_{w=1}^{n}\frac{1}{n}r_k\left(\frac{i}{n}\right)r_k\left(\frac{w}{n}\right)r_m\left(\frac{i}{n}\right)r_m\left(\frac{w}{n}\right) \tag{B39}$$

$$=\sum_{k=1}^{c}\sum_{m=1}^{c}\theta_k\theta_m\omega_n\sum_{i=1}^{n}\frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{w=1}^{n}\frac{1}{n}r_k\left(\frac{i}{n}\right)r_k\left(\frac{w}{n}\right)r_m\left(\frac{i}{n}\right)r_m\left(\frac{w}{n}\right) \tag{B40}$$

We now consider each k,m independently,

$$\theta_k\theta_m\omega_n\sum_{i=1}^{n}\frac{1}{n}r_j\left(\frac{i}{n}\right)r_l\left(\frac{i}{n}\right)\sum_{w=1}^{n}\frac{1}{n}r_k\left(\frac{i}{n}\right)r_k\left(\frac{w}{n}\right)r_m\left(\frac{i}{n}\right)r_m\left(\frac{w}{n}\right) \tag{B41}$$

$$= \theta_k \theta_m \omega_n \sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) r_k\left(\frac{i}{n}\right) r_m\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} r_k\left(\frac{w}{n}\right) r_m\left(\frac{w}{n}\right). \tag{B42}$$

It is here that we observe that each k, m equation (B42) can be interpreted as the product of two right endpoint rule approximations to integrals. We have that

$$\int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) r_k\left(\frac{i}{n}\right) r_m\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} r_k\left(\frac{w}{n}\right) r_m\left(\frac{w}{n}\right) \tag{B43}$$

We can therefore define the error in the approximation as

$$R(n) = \Bigg| \int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy$$

$$- \sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) r_k\left(\frac{i}{n}\right) r_m\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} r_k\left(\frac{w}{n}\right) r_m\left(\frac{w}{n}\right) \Bigg| \tag{B44}$$

where, because each function is piecewise Lipschitz, we still have that $R(n) = \mathscr{O}\left(\frac{1}{n}\right)$. Therefore, for each k, m,

$$R(n) = \theta_k \theta_m \Bigg| \int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy \tag{B45}$$

$$- \sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) r_k\left(\frac{i}{n}\right) r_m\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} r_k\left(\frac{w}{n}\right) r_m\left(\frac{w}{n}\right) \Bigg| \tag{B46}$$

$$= \mathscr{O}\left(\frac{1}{n}\right). \tag{B47}$$

Since taking a finite sum does not impact the order of the error we can conclude that

$$\Bigg| \sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} f\left(\frac{i}{n}, \frac{w}{n}\right)^2 \tag{B48}$$

$$- \sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k \theta_m \int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy \Bigg| \tag{B49}$$

$$= \mathscr{O}\left(\frac{1}{n}\right). \tag{B50}$$

Observe now the impact of ω_n from equation (B29),

$$\omega_n \sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} f\left(\frac{i}{n}, \frac{w}{n}\right)^2 \tag{B51}$$

$$= \omega_n \left(\sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k \theta_m \omega_n \int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy + \mathscr{O}\left(\frac{1}{n}\right)\right). \tag{B52}$$

Since each $r_j(x)$ is piecewise Lipschitz, we can conclude that the entire term in equation (B29) is negligible because,

$$\omega_n \sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} f\left(\frac{i}{n}, \frac{w}{n}\right)^2 \tag{B53}$$

$$= \omega_n \left(\sum_{k=1}^{c} \sum_{m=1}^{c} \theta_k \theta_m \omega_n \int_0^1 r_j(x) r_l(x) r_k(x) r_m(x) dx \int_0^1 r_k(y) r_m(y) dy + \mathscr{O}\left(\frac{1}{n}\right)\right) \tag{B54}$$

$$= \mathscr{O}(\omega_n) + \mathscr{O}\left(\omega_n \frac{1}{n}\right). \tag{B55}$$

Therefore,

$$\sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} f\left(\frac{i}{n}, \frac{w}{n}\right) \left(1 - \omega_n f\left(\frac{i}{n}, \frac{w}{n}\right)\right) = \sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} f\left(\frac{i}{n}, \frac{w}{n}\right) + \mathscr{O}(\omega_n). \tag{B56}$$

By way of the exact same analysis, we can conclude that the right-hand side of (B28) is

$$\sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} f\left(\frac{i}{n}, \frac{w}{n}\right) = \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx + \mathscr{O}\left(\frac{1}{n}\right). \tag{B57}$$

Therefore,

$$\left| \sum_{i=1}^{n} \frac{1}{n} r_j\left(\frac{i}{n}\right) r_l\left(\frac{i}{n}\right) \sum_{w=1}^{n} \frac{1}{n} f\left(\frac{i}{n}, \frac{w}{n}\right) \left(1 - \omega_n f\left(\frac{i}{n}, \frac{w}{n}\right)\right) - \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx \right| = \mathscr{O}(\omega_n). \tag{B58}$$

The conclusion of this analysis shows that

$$\left| b_{j,l}^{(2)} - b_{j,l}^{*,(2)} \right| = \theta_i^{-2} \sqrt{\theta_j \theta_l} \left| (n\omega_n)^{-1} \boldsymbol{e}_j^T \mathbb{E}\left[(\boldsymbol{A} - \mathbb{E}[\boldsymbol{A}])^2 \right] \boldsymbol{e}_l - \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx \right| \tag{B59}$$

$$= \mathscr{O}(\omega_n). \tag{B60}$$

The result puts together equations (B20) and (B60) showing that for every j, l

$$|b_{j,l} - \left(b_{j,l}^{*,(1)} + b_{j,l}^{*,(2)}\right)| = |b_{j,l}^{(1)} + b_{j,l}^{(2)} - \left(b_{j,l}^{*,(1)} + b_{j,l}^{*,(2)}\right)| = \mathscr{O}(\omega_n). \tag{B61}$$

Recall the definition of $\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}$. Define $\boldsymbol{B}^\epsilon$ as

$$\boldsymbol{B}^\epsilon = \boldsymbol{B} - \boldsymbol{B}^*. \tag{B62}$$

We have just shown that every component of $\boldsymbol{B}^\epsilon$ is $\mathscr{O}(\omega_n)$. We conclude the proof by appealing to Weyl–Lidskii's theorem on the spectral inclusion of the eigenvalues. For this theorem, we take

$$\boldsymbol{A} = \boldsymbol{B}^\epsilon \tag{B63}$$

$$\boldsymbol{H} = \boldsymbol{B}^* \tag{B64}$$

$$\boldsymbol{H} + \boldsymbol{A} = \boldsymbol{B}^* + \boldsymbol{B}^\epsilon = \boldsymbol{B}^* + \boldsymbol{B} - \boldsymbol{B}^* = \boldsymbol{B} \tag{B65}$$

where

$$||\boldsymbol{B}^*|| \leq ||\boldsymbol{B}^*||_F = \mathscr{O}(\omega_n), \tag{B66}$$

and conclude that

$$|\lambda_i(\boldsymbol{B}) - \lambda_i(\boldsymbol{B}^*)| = \mathscr{O}(\omega_n). \tag{B67}$$

□

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 B.2. (Theorem 8.4 in the main document) For every $1 \leq i \leq c$,

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] = \lambda_i(\boldsymbol{B}^*) + \mathscr{O}(\sqrt{\omega_n}), \tag{B68}$$

where

$$\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}. \tag{B69}$$

and

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b^{*,(1)}_{j,l} = \sqrt{\theta_j\theta_l}\, n\omega_n \int_0^1 r_j(x) r_l(x) dx = \begin{cases} \theta_j n\omega_n & j = l \\ 0 & j \neq l \end{cases} \tag{B70}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b^{*,(2)}_{j,l} = \theta_i^{-2}\sqrt{\theta_j\theta_l} \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y)\, dy dx. \tag{B71}$$

Proof. The proof is a consequence of Lemma B.1 and Theorem B.1. Let $\boldsymbol{B}$ be as defined by Theorem B.1 and $\boldsymbol{B}^*$ be defined as in Lemma B.1. Consider

$$|\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] - \lambda_i(\boldsymbol{B}) + \lambda_i(\boldsymbol{B}) - \lambda_i(\boldsymbol{B}^*)| \leq |\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] - \lambda_i(\boldsymbol{B})| + |\lambda_i(\boldsymbol{B}) - \lambda_i(\boldsymbol{B}^*)| \tag{B72}$$

$$= \mathscr{O}\left(\sqrt{\omega_n} + \frac{1}{n\omega_n}\right) + \mathscr{O}(\omega_n) \tag{B73}$$

$$= \mathscr{O}(\sqrt{\omega_n}). \tag{B74}$$

□

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 B.1 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 B.2. Let $\boldsymbol{B}^*$ be as defined in Lemma B.1 such that

$$\boldsymbol{B}^* = \boldsymbol{B}^{*,(1)} + \boldsymbol{B}^{*,(2)}. \tag{B75}$$

Then

$$|\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega n f}})\right] - \lambda_i(\boldsymbol{B}^{*,(1)})| = \mathcal{O}(1). \tag{B76}$$

where

$$\left(\boldsymbol{B}^{*,(1)}\right)_{j,l} = b_{j,l}^{*,(1)} = \sqrt{\theta_j\theta_l}\, n\omega_n \int_0^1 r_j(x) r_l(x) dx = \begin{cases} \theta_j n\omega_n & j = l \\ 0 & j \neq l \end{cases} \tag{B77}$$

$$\left(\boldsymbol{B}^{*,(2)}\right)_{j,l} = b_{j,l}^{*,(2)} = \theta_i^{-2}\sqrt{\theta_j\theta_l} \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y)\,dy\,dx. \tag{B78}$$

Proof. 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,

$$b_{j,l}^{*,(2)} = \mathcal{O}(1). \tag{B79}$$

Because $\boldsymbol{B}^{*,(2)} \in \mathbb{R}^{c\times c}$ where c is fixed and independent of n,

$$||\boldsymbol{B}^{*,(2)}|| \leq ||\boldsymbol{B}^{*,(2)}||_F = \sum_{j=1}^{c}\sum_{l=1}^{c} b_{j,l}^{*,(2)} = \mathcal{O}(1). \tag{B80}$$

By Weyl–Lidskii, taking

$$\boldsymbol{A} = \boldsymbol{B}^{*,(2)} \tag{B81}$$

$$\boldsymbol{H} = \boldsymbol{B}^{*,(1)} \tag{B82}$$

$$\boldsymbol{H} + \boldsymbol{A} = \boldsymbol{B}^* \tag{B83}$$

we can conclude that

$$|\lambda_i(\boldsymbol{B}^*) - \lambda_i(\boldsymbol{B}^{*,(1)})| = \mathcal{O}(1). \tag{B84}$$

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

$$|\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega n f}})\right] - \lambda_i(\boldsymbol{B}^*) + \lambda_i(\boldsymbol{B}^*) - \lambda_i(\boldsymbol{B}^{*,(1)})| \leq |\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega n f}})\right] - \lambda_i(\boldsymbol{B}^*)| + |\lambda_i(\boldsymbol{B}^*) - \lambda_i(\boldsymbol{B}^{*,(1)})| \tag{B85}$$

$$= \mathscr{O}(1) + \mathscr{O}(\sqrt{\omega_n}) = \mathscr{O}(1). \tag{B86}$$

□

Theorem 8.4 is a direct application of the prior lemma.

THEOREM B.3. (Estimation of the largest eigenvalues of Stochastic Block Models) For $i = 1, ..., c$

$$\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] = \theta_i n\omega_n + \mathscr{O}(1). \tag{B87}$$

Proof. 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

$$\frac{|\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right] - n\omega_n\theta_i|}{n\omega_n\theta_i} = \mathscr{O}\left(\frac{1}{n\omega_n}\right). \tag{B88}$$

A second noteworthy observation is that the $\mathscr{O}(1)$ 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 B.1 and Theorem B.2. This saves significant computational time, particularly for extremely large graphs.

C. Proof of Theorem 8.1

Theorem 8.1 constitutes the main theoretical contribution of this chapter. 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 $1 \le i \le c$, $\left|\lambda_i(\boldsymbol{B}^*) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}})\right]\right| = \mathscr{O}(\sqrt{\rho_n})$ where $\boldsymbol{B}^*$ is given by Lemma B.1. 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

$$|\lambda_i(\boldsymbol{A}) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho n f}})\right] + \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho n f}})\right] - \lambda_i(\boldsymbol{A}_N^*)|$$

$$= |\lambda_i(\boldsymbol{B}^*) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho n f}})\right] + \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho n f}})\right] - \lambda_i(\boldsymbol{A}_N^*)| < \epsilon$$

for each $1 \leq i \leq c$.

We now proceed with our proof.

Step 1: Constructing a stochastic block model kernel

Let $G \in \mathscr{G}$ with adjacency matrix $\boldsymbol{A}$ be such that

$$0 \preccurlyeq \sigma_c(\boldsymbol{A}) \tag{C1}$$

and for every $1 \leq i \neq j \leq c$, $\lambda_i \neq \lambda_j$.

LEMMA C.1. 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

$$L_f(t) = \int_0^1 f(x, y; \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{s}) t(y) dy, \tag{C2}$$

which satisfies

$$\lambda_i(L_f) = \theta_i \tag{C3}$$

Proof. 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 $\mathscr{I}_i = [S_{i-1}, S_i)$. Note that $S_0 = 0$. Define the function

$$f(x, y) = \begin{cases} \frac{\theta_i}{s_i} & \text{if}(x, y) \in I_i \times I_i \\ 0 & \text{else.} \end{cases} \tag{C4}$$

Define the linear integral operator $L_f(t) = \int_0^1 f(x, y) t(y) dy$. The eigenfunctions for L_f are

$$r_i(x) = \begin{cases} \frac{1}{\sqrt{s_i}} & \text{if} x \in \mathscr{I}_i \\ 0 & \text{else,} \end{cases} \tag{C5}$$

which we show in the following computations. We can compute the eigenvalues of L_f as

$$L_f(r_i(x)) = \int_0^1 f(x, y) r_i(y) dy \tag{C6}$$

$$= \begin{cases} \int_{\mathscr{I}_i} \frac{\theta_i}{s_i} \frac{1}{\sqrt{s_i}} dy & x \in \mathscr{I}_i \\ 0 & \text{else} \end{cases} \tag{C7}$$

$$= \begin{cases} \frac{\theta_i}{s_i} \frac{1}{\sqrt{s_i}} s_i & x \in \mathscr{I}_i \\ 0 & \text{else} \end{cases} \tag{C8}$$

$$= \begin{cases} \theta_i \frac{1}{\sqrt{s_i}} & x \in \mathscr{I}_i \\ 0 & \text{else} \end{cases} \tag{C9}$$

$$= \theta_i r_i(x). \tag{C10}$$

Next we verify that $||r_i||_2 = 1$.

$$\int_0^1 r_i(x)^2 dx = \int_{\mathscr{I}_i} \frac{1}{s_i} dx \tag{C11}$$

$$= \frac{s_i}{s_i} \tag{C12}$$

$$= 1. \tag{C13}$$

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 $\boldsymbol{\theta} = \frac{\sigma_c(\boldsymbol{A})-\boldsymbol{1}}{n\rho_n}$ in Lemma C.1 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 B. We will use the estimate of the largest eigenvalues outlined by Theorem B.2. Recall that

$$|\mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}})\right] - \lambda_i(\boldsymbol{B}^*)| = \mathscr{O}(\sqrt{\rho_n}). \tag{C14}$$

where

$$(\boldsymbol{B}^*)_{j,l} = b_{j,l}^{*,(1)} + b_{j,l}^{*,(2)} = \sqrt{\theta_j \theta_l} n\rho_n \int_0^1 r_j(x) r_l(x) dx + \theta_i^{-2} \sqrt{\theta_j \theta_l} \int_0^1 r_j(x) r_l(x) \int_0^1 f(x,y) dy dx \tag{C15}$$

when $\omega_n = \rho_n$. For the choice of $f(x, y)$ made in step 1, when $j \neq l$,

$$\forall x, \quad r_j(x) r_l(x) dx = 0. \tag{C16}$$

Thus, $b_{j,l}^{*,(2)} = 0$ for all $j \neq l$ and so $\boldsymbol{B}^*$ is diagonal. The i^{th} eigenvalue is then given as

$$\lambda_i(\boldsymbol{B}^*) = \theta_i n\rho_n + \frac{1}{\theta_i} \int_0^1 r_i(x) r_i(x) \int_0^1 f(x,y) dy dx. \tag{C17}$$

Using the definitions of $r_i(x)$ and $f(x, y)$ from equations (C4) and (C5),

$$\frac{1}{\theta_i} \int_0^1 r_i(x) r_i(x) \int_0^1 f(x,y) dy dx = \frac{1}{\theta_i} \frac{1}{s_i} \int_{I_i} \int_{I_i} f(x,y) dy dx \tag{C18}$$

$$= \frac{1}{\theta_i} \frac{1}{s_i} \frac{\theta_i}{s_i} \int_{I_i} \int_{I_i} dy dx \tag{C19}$$

$$= \frac{1}{s_i^2} s_i^2 = 1 \tag{C20}$$

since $\int_{I_i} dy = s_i$. This shows that

$$\lambda_i(\boldsymbol{B}^*) = \theta_i n\rho_n + 1 = \lambda_i(\boldsymbol{A}). \tag{C21}$$

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

$$0 < C < P\left(||\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon\right). \tag{C22}$$

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 [21] on the convergence in distribution of the extreme eigenvalues of inhomogeneous Erdős–Rényi random graphs, which we state below.

Theorem C.1. (Chakrabarty, Chakraborty, and Hazra 2020) For every $1 \le i \le c$,

$$\rho_n^{-1/2}(\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \xrightarrow{d} (Z_i : 1 \le i \le c), \tag{C23}$$

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

$$Cov(Z_i, Z_j) = 2\int_0^1 \int_0^1 r_i(x)r_i(y)r_j(x)r_j(y)f(x,y)dxdy, \tag{C24}$$

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

Proof. Theorem C.1 is Theorem 2.3 in [21]. □

We first acknowledge that there exists a very similar theorem to the above in [5] 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 (C23). An equivalent statement to Theorem C.1 is then

$$\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \xrightarrow{d} Z. \tag{C25}$$

One characterization of convergence in distribution for finite-dimensional random variables is pointwise convergence of the cumulative distribution functions (see Theorem A.2).

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

$$F_n(\boldsymbol{z}) = P_n\left((-\infty, z_1] \times ... \times (-\infty, z_c]\right), \tag{C26}$$

$$F(\boldsymbol{z}) = P\left((-\infty, z_1] \times ... \times (-\infty, z_c]\right). \tag{C27}$$

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

$$\lim_{n\to\infty} F_n(\boldsymbol{z}) = F(\boldsymbol{z}), \tag{C28}$$

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 (C28) takes the form

$$\lim_{n\to\infty} P_n(\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}})\right]) \preccurlyeq \boldsymbol{z}) = P(Z \preccurlyeq \boldsymbol{z}). \tag{C29}$$

We are now ready to state and prove our lemma. Let $\mu_{\rho_n f} \in \mathscr{M}(\mathscr{G})$ 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 C.2. $\forall \epsilon > 0$, $\exists n^* \in \mathbb{N}$ where $\forall n > n^*$, $\exists G' \in \mathscr{G}$ with adjacency matrix $\boldsymbol{A}'$ such that

$$||\sigma_c(\boldsymbol{A}') - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}})\right]||_2 < \epsilon. \tag{C30}$$

Proof. Let $\epsilon > 0$. Fix $\boldsymbol{z} \in \mathbb{R}^c$ such that $0 \preccurlyeq \boldsymbol{z}$ and

$$0 < C < P(-\boldsymbol{z} \preccurlyeq Z \preccurlyeq \boldsymbol{z}) - 2\epsilon, \tag{C31}$$

for some $C > 0$. By equation (C29), there exists $n_1 \in \mathbb{N}$ where for all $n > n_1$,

$$\left|P_n\left(\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}})\right]) \preccurlyeq \boldsymbol{z}\right) - P\left(Z \preccurlyeq \boldsymbol{z}\right)\right| < \epsilon \tag{C32}$$

Similarly, there exists $n_2 \in \mathbb{N}$ where for all $n > n_2$,

$$\left|P_n\left(\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}})\right]) \preccurlyeq -\boldsymbol{z}\right) - P\left(Z \preccurlyeq -\boldsymbol{z}\right)\right| < \epsilon. \tag{C33}$$

Furthermore, there exists $n_3 \in \mathbb{N}$ where for all $n > n_3$,

$$||\sqrt{\rho_n}\boldsymbol{z}||_2 < \epsilon. \tag{C34}$$

Take $n^* = \max(n_1, n_2, n_3)$ and consider the following probability

$$P_n(-\boldsymbol{z} \preccurlyeq \rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}})\right]) \preccurlyeq \boldsymbol{z}) = P_n(\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}})\right]) \preccurlyeq \boldsymbol{z}) \tag{C35}$$

$$- P_n(\rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho n f}})\right]) \preccurlyeq -\boldsymbol{z}). \tag{C36}$$

Now, (C32) and (C33) allow us to replace (C35) and (C36) with the corresponding expression in terms of z. We therefore obtain

$$\left|P_n\left(-z \preccurlyeq \rho_n^{-1/2}(\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq z\right) - P\left(-z \preccurlyeq Z \preccurlyeq z\right)\right| < 2\epsilon, \tag{C37}$$

from which we obtain

$$P(-z \preccurlyeq Z \preccurlyeq z) - 2\epsilon < P_n(-\rho_n^{1/2} z \preccurlyeq (\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq \rho_n^{1/2} z). \tag{C38}$$

Since $0 < C < P(-z \preccurlyeq Z \preccurlyeq z) - 2\epsilon$ we have

$$C < P_n(-\rho_n^{1/2} z \preccurlyeq (\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]) \preccurlyeq \rho_n^{1/2} z) \tag{C39}$$

$$< P_n(0 \leq ||\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 \leq ||\sqrt{\rho_n} z||_2) \tag{C40}$$

and since n is sufficiently large that $||\rho_n^{1/2} z||_2 < \epsilon$ 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 $G' \in \mathscr{G}$ with adjacency matrix $\boldsymbol{A}'$ that satisfies $||\sigma_c(\boldsymbol{A}') - \mathbb{E}\left[\sigma_c(\boldsymbol{A})\right]||_2 < \epsilon$. □

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.

We are now in a position to prove Theorem 8.3, 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 C.2. (Theorem 8.3 in the main chapter) $\forall\epsilon > 0$, $\exists n^* \in \mathbb{N}$ such that for all $n > n^*$,

$$\lim_{N\to\infty} ||\sigma_c(\boldsymbol{A}_N^*) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon \quad a.s. \tag{C41}$$

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 C.3. (The set of c largest realizable eigenvalues of adjacency matrices of graphs in $\mathscr{G}$)

$$\Lambda_n^c = \left\{\boldsymbol{\lambda} \in \mathbb{R}^c | \exists G \in \mathscr{G} \text{ with adjacency matrix } \boldsymbol{A} \text{ such that } \boldsymbol{\lambda} = \sigma_c(\boldsymbol{A})\right\}. \tag{C42}$$

Proof. We make two observations. First that $\Lambda_n^c \subset \mathbb{R}^c$ and second, equation (C41) depends only on the eigenvalues of $\boldsymbol{A}^{(k)}$. 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

$$\boldsymbol{\lambda}_N^* = \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^N ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}||_2^2. \tag{C43}$$

We also consider the relaxed version of (C43) 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

$$\boldsymbol{\lambda}^*_{N,r} = \operatorname*{argmin}_{\boldsymbol{\lambda}\in\mathbb{R}^c} \frac{1}{N}\sum_{k=1}^{N} ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}||_2^2. \tag{C44}$$

$$= \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{\lambda}^{(k)} \tag{C45}$$

which is the classic arithmetic average of the observations $\boldsymbol{\lambda}^{(k)}$. Now, the sample mean, $\boldsymbol{\lambda}^*_{N,r}$, satisfies

$$\frac{1}{N}\sum_{k=1}^{N} ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}||_2^2 = ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2 + \frac{1}{N}\sum_{k=1}^{N} ||\boldsymbol{\lambda}^{(k)}||_2^2 - ||\boldsymbol{\lambda}^*_{N,r}||_2^2. \tag{C46}$$

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

$$\boldsymbol{\lambda}^*_{N} = \operatorname*{argmin}_{\boldsymbol{\lambda}\in\Lambda_n^c} \frac{1}{N}\sum_{k=1}^{N} ||\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}||_2^2 \tag{C47}$$

$$= \operatorname*{argmin}_{\boldsymbol{\lambda}\in\Lambda_n^c} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2. \tag{C48}$$

Equation (C48) states that we must find the realizable eigenvalues that 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, $\forall n$,

$$\lim_{N\to\infty} ||\boldsymbol{\lambda}^*_{N,r} - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]|| = 0 \quad a.s. \tag{C49}$$

Define

$$\boldsymbol{\lambda}^* = \lim_{N\to\infty} \boldsymbol{\lambda}^*_N \tag{C50}$$

$$= \lim_{N\to\infty} \operatorname*{argmin}_{\boldsymbol{\lambda}\in\Lambda_n^c} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2. \tag{C51}$$

Since the projection on Λ_n^c is a continuous operator,

$$\lim_{N\to\infty} \operatorname*{argmin}_{\boldsymbol{\lambda}\in\Lambda_n^c} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2 = \operatorname*{argmin}_{\boldsymbol{\lambda}\in\Lambda_n^c} \lim_{N\to\infty} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2. \tag{C52}$$

Since the norm is continuous,

$$\operatorname*{argmin}_{\boldsymbol{\lambda}\in\Lambda_n^c} \lim_{N\to\infty} ||\boldsymbol{\lambda} - \boldsymbol{\lambda}^*_{N,r}||_2^2 = \operatorname*{argmin}_{\boldsymbol{\lambda}\in\Lambda_n^c} ||\boldsymbol{\lambda} - \lim_{N\to\infty} \boldsymbol{\lambda}^*_{N,r}||_2^2. \tag{C53}$$

Finally, by (C49),

$$\underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}}\, ||\boldsymbol{\lambda} - \lim_{N\to\infty} \boldsymbol{\lambda}^*_{N,r}||_2^2 = \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}}\, ||\boldsymbol{\lambda} - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2^2 \quad a.s. \tag{C54}$$

By Lemma C.2, $\forall\epsilon > 0$, 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

$$||\sigma_c(\boldsymbol{A}') - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 \le \epsilon. \tag{C55}$$

Because $\sigma_c(\boldsymbol{A}') \in \Lambda_n^c$, the minimizer, $\boldsymbol{\lambda}^*$, in (C54) satisfies

$$||\boldsymbol{\lambda}^* - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 \le \epsilon \quad a.s. \tag{C56}$$

Since $\boldsymbol{\lambda}^* = \lim_{N\to\infty} \boldsymbol{\lambda}^*_N = \lim_{N\to\infty} \sigma_c(\boldsymbol{A}^*_N)$ this concludes our proof. $\square$

Step 4.

All that is left is to compile the results of the prior three steps into a proof for Theorem 8.1, which we restate below. Assume $G \in \mathcal{G}$ with adjacency matrix $\boldsymbol{A}$ that satisfies the following:

1. $\rho_n = \omega(n^{-2/3})$.
2. $\lim_{n\to\infty} \rho_n = 0$.
3. $0 \preccurlyeq \sigma_c(\boldsymbol{A})$.
4. For every $1 \le i \ne j \le c$, $\lambda_i \ne \lambda_j$.

Theorem C.4. (Theorem 8.1 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

$$\lim_{N\to\infty} d_{A_c}(G, G^*_N) < \epsilon \quad a.s. \tag{C57}$$

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. We begin our proof by expanding the left-hand side of equation (C57).

$$d_{A_c}(G, G^*_N) = ||\sigma_c(\boldsymbol{A}) - \sigma_c(\boldsymbol{A}^*_N)||_2 \tag{C58}$$

$$= ||\sigma_c(\boldsymbol{A}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right] + \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right] - \sigma_c(\boldsymbol{A}^*_N)||_2. \tag{C59}$$

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:

$$||\sigma_c(\boldsymbol{A}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon \tag{C60}$$

$$||\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right] - \sigma_c(\boldsymbol{A}^*_N)||_2 < \epsilon \quad a.s. \tag{C61}$$

We begin with (C60). Define L_f as in Lemma C.1 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 8.4,

$$\mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right] = n\rho_n\boldsymbol{\theta} + \mathbf{1} + \mathcal{O}(\sqrt{\rho_n}). \tag{C62}$$

Since $\sqrt{\rho_n} \to 0$, there exists an $n_1 \in \mathbb{N}$ such that for all $n > n_1$,

$$||n\rho_n\boldsymbol{\theta} + \mathbf{1} - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\rho_n f}})\right]||_2 < \epsilon. \tag{C63}$$

Theorem 8.3 implies the existence of an $n_2 \in \mathbb{N}$ such that inequality (C61) holds. Taking $n^* = \max(n_1, n_2)$ implies that both inequalities hold and concludes the proof of Theorem 8.1. □

D. Proof of Theorem 8.2

The proof of Theorem 8.2, which we restate below, is a consequence of Lemmas C.1 and C.2 along with Theorem 8.4. Let $\{G^{(k)}\}_{k=1}^N$ have sample Fréchet mean G_N^*.

Theorem D.1. (Theorem 8.2 in the main chapter) $\forall\epsilon > 0$, $\exists n^* \in \mathbb{N}$ such that $\forall n > n^*$,

$$\left\|\sigma_c(\boldsymbol{A}_N^*) - \frac{1}{N}\sum_{k=1}^N \sigma_c(\boldsymbol{A}^{(k)})\right\|_2 < \epsilon. \tag{D1}$$

Proof. 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 C1 that defines the disconnected stochastic block model kernel. Let $\epsilon > 0$. By Lemma C2 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

$$||\sigma_c(\boldsymbol{A}') - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\bar{\rho}_n f}})\right]||_2 < \epsilon. \tag{D2}$$

By Theorem 8.4, there exists $n_2 \in \mathbb{N}$ such that for all $n > n_2$,

$$\forall i = 1, ..., c, \left|n\bar{\rho}_n\lambda_i(L_f) + 1 - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\bar{\rho}_n f}})\right]\right| < \epsilon. \tag{D3}$$

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

$$\left\|\sigma_c(\boldsymbol{A}') - \frac{1}{N}\sum_{k=1}^N \boldsymbol{\lambda}^{(k)}\right\|_2 < C\epsilon, \tag{D4}$$

where C is an arbitrary positive constant. We recall, (see (C48), that the spectrum of the sample Fréchet mean is the solution to

$$\boldsymbol{\lambda}_N^* = \underset{\boldsymbol{\lambda}\in\Lambda_n^c}{\operatorname{argmin}} \left\|\boldsymbol{\lambda} - \frac{1}{N}\sum_{k=1}^N \boldsymbol{\lambda}^{(k)}\right\|_2^2. \tag{D5}$$

Now, take $n^* = \max(n_1, n_2)$, then the existence of the graph G' with adjacency matrix $\boldsymbol{A}'$ shows that

$$\left\|\sigma_c(\boldsymbol{A}_N^*) - \frac{1}{N}\sum_{k=1}^N \boldsymbol{\lambda}^{(k)}\right\|_2^2 < \left\|\sigma_c(\boldsymbol{A}') - \frac{1}{N}\sum_{k=1}^N \boldsymbol{\lambda}^{(k)}\right\|_2^2 < \epsilon. \tag{D6}$$

Thus, the adjacency matrix of the sample Fréchet mean must have eigenvalues that are within ϵ of the geometric average. □

E. Proof of Theorem 8.5

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

$$\widehat{G}_{\tilde{N}}^{*,\mu_{\omega_n f}} = \underset{\tilde{G}\in\{\tilde{G}^{(k)}\}_{k=1}^{\tilde{N}}}{\operatorname{argmin}} \frac{1}{\tilde{N}} \sum_{k=1}^{\tilde{N}} d_{A_c}^2(\tilde{G}, \tilde{G}^{(k)}) \tag{E1}$$

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

Theorem E.1. $\forall \epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left\|\sigma_c(\hat{\boldsymbol{A}}_{\tilde{N}}^{*,\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right\|_2 > \epsilon\right) = 0. \tag{E2}$$

Proof. 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 C.1,

$$\frac{1}{\sqrt{\omega_n}}\left(\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right) \xrightarrow{d} Z \sim N(0, \Sigma), \tag{E3}$$

where the $c \times c$ covariance matrix Σ is given by (C24). Let $z > 0$, $\epsilon > 0$, and $\eta > 0$. Because of (E3), $\exists n_0 \in \mathbb{N}$, $\forall n > n_0$,

$$\left| P\left(\frac{1}{\sqrt{\omega_n}}|\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right]| \le z\right) - P\left(|z_i| \le z; i = 1, ..., c\right)\right| \le \frac{\eta}{2}. \tag{E4}$$

Thus,

$$P\left(|z_i| \le z; i = 1, ..., c\right) - \frac{\eta}{2} \le P\left(\frac{1}{\sqrt{\omega_n}}|\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right]| \le z\right). \tag{E5}$$

Now $P(||\boldsymbol{z}|| \le z) \le P(|z_i| \le z; i = 1, ..., c)$ and

$$P\left(\frac{1}{\sqrt{\omega_n}}|\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\lambda_i(\boldsymbol{A}_{\mu_{\omega_n f}})\right]| \le z\right) \le P\left(||\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]||_2 \le \sqrt{c\omega_n} z\right). \tag{E6}$$

Because $\lim_{z\to\infty} P(||\boldsymbol{z}|| \le z) = 1$, $\exists z_0 > 0$ such that

$$1 - \frac{\eta}{2} < P(||\boldsymbol{z}|| \le z_0). \tag{E7}$$

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

$$1 - \eta < P\left(\left\|\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega_n f}})\right]\right\|_2 < \epsilon\right). \tag{E8}$$

Equivalently,

$$P\left(\left\|\sigma_c(\boldsymbol{A}_{\mu_{\omega n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega n f}})\right]\right\|_2 \geq \epsilon\right) < \eta. \tag{E9}$$

In other words, $\forall \epsilon > 0$,

$$\lim_{n \to \infty} P\left(\left\|\sigma_c(\boldsymbol{A}_{\mu_{\omega n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega n f}})\right]\right\|_2 \geq \epsilon\right) = 0. \tag{E10}$$

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

$$\mathscr{E} = \{\boldsymbol{A}^{(1)}, ..., \boldsymbol{A}^{(\tilde{N})}; \left\|\sigma_c(\widehat{\boldsymbol{A}}_{\tilde{N}}^{*,\mu_{\omega n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega n f}})\right]\right\|_2 > \epsilon\}. \tag{E11}$$

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 (E10), $\exists n_0 \in \mathbb{N}$ where $\forall n > n_0$, $P(\mathscr{E}) < \eta$. We conclude that $\forall \epsilon > 0$,

$$\lim_{n \to \infty} P\left(\left\|\sigma_c(\widehat{\boldsymbol{A}}_{\tilde{N}}^{*,\mu_{\omega n f}}) - \mathbb{E}\left[\sigma_c(\boldsymbol{A}_{\mu_{\omega n f}})\right]\right\|_2 > \epsilon\right) = 0. \tag{E12}$$

□