

Computation of the Sample Fréchet Mean for Sets of Large Graphs with Applications to Regression

Daniel Ferguson*

François G. Meyer†

Applied Mathematics, University of Colorado at Boulder, Boulder CO 80305

Abstract

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

Keywords: Statistical Network Analysis; Fréchet mean.

1 Introduction.

Machine learning almost always requires the estimation of the average of a dataset. Algorithms for clustering, classification, and linear regression all utilize the average value of the data set [17]. When the distance is induced by a norm, then the mean is a simple algebraic operation. If the data lie on a Riemannian manifold, equipped with a metric, then one can extend the notion of mean with the concept of Fréchet mean [27]. In fact the concept of Fréchet mean only requires that a (pseudo)metric between points be defined, and therefore one can consider the Fréchet mean of a set in a pseudometric space [14]. Not surprisingly, many machine learning algorithms, which were developed for Euclidean spaces, can be extended to use the Fréchet mean. The purpose of this paper is to solve the nontrivial problem of determining the sample Fréchet mean for data sets of graphs when the pseudometric is the ℓ_2 distance between the eigenvalues of the adjacency matrices of two graphs.

In this work we consider a set of simple graphs with

n vertices which have an edge density that satisfies,

$$(1.1) \quad n^{-2/3} \ll \rho_n \ll 1.$$

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

Our line of attack involves the following two intermediate results: (1) a graph's largest eigenvalues can be approximated within any precision by the corresponding eigenvalues of a realization of a stochastic block model; (2) given a set of target eigenvalues, one can recover the stochastic block model whose Fréchet mean has a spectrum that best matches these target eigenvalues. We prove various error bounds and convergence results for our algorithm and validate the theory with several experiments.

2 State of the Art.

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

The choice of metric is crucial to the computation of the Fréchet mean, since each metric induces a different mean graph. The Fréchet mean of graphs has been studied in the context where the distance is the edit distance (e.g., [4, 18] and references therein). The edit distance reflects small scale changes in the graphs and therefore the Fréchet mean will be sensitive to the fine structural variations between graphs. Effectively, the Fréchet mean with respect to the edit distance can be interpreted as an average of the fine structures in the observed graphs.

In this paper, we consider that the fine scale, which is defined by the local connectivity at the level of each vertex, may be intrinsically random. The quantification of such random fluctuations is uninformative when comparing graphs. We therefore study a distance that can detect larger scale patterns of connectivity that

*dfe0926@colorado.edu; †fmeyer@colorado.edu

This work was supported by the National Science Foundation, CCF/CIF 1815971

happen at multiple scales (e.g., community structure [1, 22], modularity [16]).

The adjacency spectral distance, which we define as the ℓ_2 norm of the difference between the spectra of the adjacency matrices of the two graphs of interest [31], exhibits good performance when comparing various types of graphs [30], making it a reliable choice for a wide range of problems. Spectral distances also exhibit practical advantages, as they can inherently compare graphs of different sizes and can compare graphs without known vertex correspondence (see e.g., [12] and references therein). The adjacency spectrum in particular is well-understood, and is perhaps the most frequently studied graph spectrum [8, 13].

The eigenvalues of the adjacency matrix carry important topological information about the graph at different structural scales [31]. The spectrum reveals information about large scale features such as community structure [22] or the existence of highly connected “hubs” [13], as well as the smaller scale structure such as local connectivity (i.e. the degree of a vertex) or the ubiquity of substructures such as triangles [11]. In [30], the authors observe that the adjacency spectral pseudo-distance exhibits good performance across a variety of scenarios, making it a reliable choice for a wide range of problems (from the two-sample test problem to the change point detection problem).

A stronger notion of spectral similarity involves the similarity between the eigenspaces of the adjacency matrices of the respective graphs (e.g., [23]). These notions of spectral similarity have regained some interest in the context of graph coarsening (or aggregation) for graph neural networks [6, 26].

Inspired by the conjecture of Vu [29], we propose to use the eigenvalues of the adjacency matrix as “fingerprints” that uniquely (up to trivial permutations and the possible iso-spectral graphs) characterize a graph (see also [20]).

In practice, it is often the case that only the first c eigenvalues are compared, where $c \ll n$. We still refer to such distances as spectral distances but comparison using the first c eigenvalues for small c allows one to focus on the global structure of the graph while ignoring the local structure [22]. We provide in section 6 our recommendation for the choice of c .

Instead of solving the minimization problem associated with the computation of the Fréchet mean in the original set $\mathcal{G}$, the authors in [12] suggest to embed the graphs in Euclidean space, wherein they can trivially find the mean of the set. Because the embedding in [12] is not an isometry, there is no guarantee that the inverse of the average embedded graphs be equal to the Fréchet mean. Furthermore, the inverse embedding may not be

available in closed form. In the case of simple graphs, the Laplacian matrix of the graph uniquely characterizes the graph. The authors in [15] define the mean of a set of graphs using the Fréchet sample mean (computed on the manifold defined by the cone of symmetric positive semi-definite matrices) of the respective Laplacian matrices.

3 Main Contributions.

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

4 Notations.

We denote by $G = (V, E)$ a graph with vertex set $V = \{1, 2, \dots, n\}$ 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 $\mathbf{A}$ is the adjacency matrix of the graph and is defined as

$$(4.2) \quad \mathbf{A}_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E, \\ 0 & \text{else.} \end{cases}$$

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,

$$(4.3) \quad \sigma : \mathbb{M}_{n \times n} \longrightarrow \mathbb{R}^n,$$

$$(4.4) \quad \mathbf{A} \longmapsto \boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_n],$$

where $\lambda_1 \geq \dots \geq \lambda_n$. Because we often consider the c largest eigenvalue of the adjacency matrix $\mathbf{A}$, we define

the mapping to the truncated spectrum as σ_c ,

$$(4.5) \quad \sigma_c : \mathbb{M}_{n \times n} \longrightarrow \mathbb{R}^c,$$

$$(4.6) \quad \mathbf{A} \longmapsto \boldsymbol{\lambda}_c = [\lambda_1, \dots, \lambda_c].$$

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

$$(4.7) \quad d_A(G, G') = \|\sigma(\mathbf{A}) - \sigma(\mathbf{A}')\|_2.$$

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 a similar spectra it can be shown that the graphs have similar global and structural properties [30]. 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 spectra of the respective adjacency matrices,

$$(4.8) \quad d_{A_c}(G, G') = \|\sigma_c(\mathbf{A}) - \sigma_c(\mathbf{A}')\|_2.$$

DEFINITION 4.3. We denote by $\mathcal{G}$ the set of all simple unweighted graphs on n nodes.

4.1 Random Graphs We denote by $\mathcal{M}(\mathcal{G})$ the space of probability measures on $\mathcal{G}$. In this work, when we talk about a measure we always mean a probability measure.

DEFINITION 4.4. We define the set of random graphs distributed according to μ to be the probability space $(\mathcal{G}, \mu)$.

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

This definition allows us to unify various ensemble of random graphs (e.g., Erdős-Rényi, inhomogeneous Erdős-Rényi, Small-World, Barabasi-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 4.5. A probability measure $\mu \in \mathcal{M}(\mathcal{G})$ is called a kernel probability measure if there exist a positive constant ρ_n and a function f ,

$$(4.9) \quad \rho_n f : [0, 1] \times [0, 1] \mapsto [0, 1],$$

such that $f(x, y) = f(1 - y, 1 - x)$, and

$$\forall G \in \mathcal{G}, \text{ with adjacency matrix } \mathbf{A} = (a_{ij}),$$

$$\mu(\{\mathbf{A}\}) = \prod_{1 \leq i < j \leq n} \mathbb{P}(a_{ij}) = \prod_{1 \leq i < j \leq n} \text{Bernoulli}\left(\rho_n f\left(\frac{i}{n}, \frac{j}{n}\right)\right).$$

The function $\rho_n f$ is called a kernel of μ and is denoted $\mu_{\rho_n f}$.

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 .

We note that given the sequence $\left\{\frac{i}{n}\right\}_{i=1}^n$ and the measure μ , the kernel $\rho_n f$ forms an equivalence class of functions, characterized by their values on the grid $\left\{\frac{i}{n}\right\}_{i=1}^n \times \left\{\frac{j}{n}\right\}_{j=1}^n$.

DEFINITION 4.6. We denote by G_μ a random realization of a graph $G \in (\mathcal{G}, \mu)$.

4.1.2 Stochastic Block Models The stochastic block model (see [1]) plays an important role in this work. We review the specific features of this model using the notations that were defined in the previous paragraphs. The key 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 $\mathbf{s} \in \ell_1$ a non-increasing non-negative sequence of relative community sizes with c non-zero entries and $\|\mathbf{s}\| = 1$.

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

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

REMARK 4.3. We allow for infinite vectors with finite number of non-zero entries so that we may introduce new communities smoothly. For example, let $t \in [0, 1]$ and parametrize $\mathbf{s}$ and $\mathbf{p}$ by t as

$$(4.10) \quad \mathbf{s}(t) = [1 - t/2, t/2, 0, \dots]$$

$$(4.11) \quad \rho_n \mathbf{p}(t) = [0.2 + t/2, 0.1 + t/2, 0, \dots].$$

We can parameterize a stochastic block model using one representative of the equivalence class of kernels, f .

We simply consider the function f , which is piecewise constant over the blocks, and is defined by $\rho_n f : [0, 1] \times [0, 1] \rightarrow [0, 1]$

$$(4.12) \quad (x, y) \mapsto \begin{cases} \rho_n p_i & \text{if } \sum_{k=1}^{i-1} s_k \leq x < \sum_{k=1}^i s_k, \\ & \text{and } \sum_{k=1}^{i-1} s_k \leq y < \sum_{k=1}^i s_k, \\ \rho_n q_{ij} & \text{if } \sum_{k=1}^{i-1} s_k \leq x < \sum_{k=1}^i s_k, \\ & \text{and } \sum_{k=1}^{j-1} s_k \leq y < \sum_{k=1}^j s_k. \end{cases}$$

This piecewise constant function is called the canonical kernel of the block model with measure μ (see, e.g. Fig. 1), and we denote it by $f(x, y; \mathbf{p}, \mathbf{Q}, \mathbf{s})$.

Example. Given $\mathbf{s} = [1/2 \ 1/4 \ 1/4 \ 0 \dots]^T$ the values of $f(x, y; \mathbf{p}, \mathbf{Q}, \mathbf{s})$ in the unit square are shown in Figure 1.

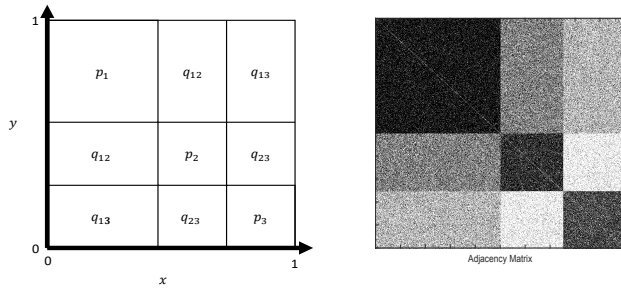


Figure 1: **Left:** Example stochastic block model kernel $f(x, y; \mathbf{p}, \mathbf{Q}, \mathbf{s})$. **Right:** Example adjacency matrix from $\mu_{\rho_n f}$

5 The Fréchet mean and sample Fréchet Mean

We equip the set $\mathcal{G}$ of graphs defined on n vertices (see definition 4.3) with the pseudometric defined by the ℓ_2 norm between the spectra of the respective adjacency matrices, d_{A_c} , (see (4.8)). For any subset of graphs $\{G^{(k)}\}_{k=1}^N \subset \mathcal{G}$, using d_{A_c} , we quantify the spread of the graphs, and we define a notion of centrality, which gives the location of the average graph, according to the empirical measure.

Definition 5.1 (sample Fréchet mean [14]) Let $\{G^{(k)}\}_{1 \leq k \leq N}$ be a set of graphs in $\mathcal{G}$. The sample Fréchet mean is defined by

$$(5.13) \quad \{G_N^* \in \mathcal{G}\} = \operatorname{argmin}_{G \in \mathcal{G}} \frac{1}{N} \sum_{k=1}^N d_{A_c}^2(G, G^{(k)}).$$

Because $\mathcal{G}$ is a finite set, the minimization problem (5.14) always has at least one solution. Throughout this work, we are interested in determining at least one element of the set $\{G_N^* \in \mathcal{G}\}$. Since our results hold for any minimizer of (5.14) (i.e. for any sample

Fréchet mean), to ease the exposition, and without loss of generality, we assume that the sample Fréchet mean is unique. Therefore $\{G^* \in \mathcal{G}\}$ is a singleton and we write the sample Fréchet mean as

$$(5.14) \quad G_N^* = \operatorname{argmin}_{G \in \mathcal{G}} \frac{1}{N} \sum_{k=1}^N d^2(G, G^{(k)}).$$

We note the similarity between equation 5.14 and the definition of the barycenter [27]. Indeed, for different sets of graphs $\{G^{(k)}\}_{k=1}^N \subset \mathcal{G}$ we expect that, for a fixed G , $\frac{1}{N} \sum_{k=1}^N d^2(G, G^{(k)})$ will change, and therefore the sample Fréchet mean G_N^* will move inside $\mathcal{G}$. Here, G_N^* plays the role of the center of mass for the set $\{G^{(k)}\}_{k=1}^N \subset \mathcal{G}$.

The computation of the sample Fréchet mean for sets of large graphs is intractable due to two primary issues. The first is that $|\mathcal{G}| = \mathcal{O}(2^{n^2})$ so any brute force procedure to solve the minimization problem in (5.13) will not compute in reasonable time. Second, the set $\mathcal{G}$ is not ordered so searching the space of graphs in a principled manner is difficult (in contrast to the situation with trees [2]). We suggest solving these issues by first lifting the sample Fréchet mean problem to a larger space and defining an approximation to the lifted problem. Our approach involves searching for the correct parameters of a stochastic block model kernel such that the sample Fréchet mean given $\mu_{\rho_n f}$ approximates the target graph, G_N^* , with respect to d_{A_c} .

6 Approximation of the sample Fréchet mean.

We first state the primary theoretical results (Theorem 6.1 and Corollary 6.1) which form the foundation of our algorithm (see Alg. 1). We additionally state theorems that are necessary results for the implementation of our algorithm.

Let $G \in \mathcal{G}$ with adjacency matrix $\mathbf{A}$ such that $n^{-2/3} \ll \rho_n \ll 1$. Assume that

$$(6.15) \quad 0 \leq \sigma_c(\mathbf{A})$$

and for every $1 \leq i \neq j \leq c$, $\lambda_i \neq \lambda_j$. Our primary theoretical contribution, Theorem 6.1, 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, $\mu_{\rho_n f}$, almost surely with respect to the truncated adjacency spectral pseudo-metric.

Theorem 6.1 (Spectrally similar large graphs) $\forall \epsilon > 0$, $\exists n_1 \in \mathbb{N}$ such that $\forall n > n_1$, $\exists f(x, y; \mathbf{p}, \mathbf{Q}, \mathbf{s})$ a canonical stochastic block model kernel with c communities such that

$$(6.16) \quad \lim_{N \rightarrow \infty} d_{A_c}(G, G_{N, \mu_{\rho_n f}}^*) < \epsilon \quad a.s.$$

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

REMARK 6.1. While we are free to choose the geometry vector $\mathbf{s}$, we make the choice that $s_1 \geq s_i$ for $i = 2, \dots, c$ and $s_i = s_j$ for $i, j = 2, \dots, c$.

The following corollary applies Theorem 6.1 to the sample Fréchet mean of any given data set of graphs, $\{G^{(k)}\}_{k=1}^N$. This corollary forms the basis of our approach to solving (5.13).

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 6.1.

Corollary 6.1 (Approximation of the sample Fréchet mean) $\forall \epsilon > 0, \exists n_1 \in \mathbb{N}$ such that $\forall n > n_1, \exists f(x, y; \mathbf{p}, \mathbf{Q}, \mathbf{s})$ a canonical stochastic block model kernel with c communities such that

$$(6.17) \quad \lim_{\tilde{N} \rightarrow \infty} d_{A_c}(G_N^*, G_{\tilde{N}, \mu_{\rho_n f}}^*) < \epsilon \quad a.s.$$

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

Corollary 6.1 tells us that rather than solving the argmin procedure in equation (5.13) we may instead search for a set of parameters for the stochastic block model kernel and compute the sample Fréchet mean. Most notably, the computation of the sample Fréchet mean given $\mu_{\rho_n f}$ is straightforward due to the following theorem. We show that with high probability, the truncated spectrum of the set mean graph of an iid sample drawn from $\mu_{\rho_n f}$ is near the expected spectrum.

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

$$(6.18) \quad \hat{G}_{\tilde{N}, \mu_{\rho_n f}}^* = \operatorname{argmin}_{\tilde{G} \in \{\tilde{G}^{(k)}\}_{k=1}^{\tilde{N}}} \frac{1}{\tilde{N}} \sum_{k=1}^{\tilde{N}} d_{A_c}^2(\tilde{G}, \tilde{G}^{(k)})$$

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

Theorem 6.2 (High probability convergence of the truncated spectrum of the set mean graph)

$\forall \epsilon > 0,$

$$(6.19) \quad \lim_{n \rightarrow \infty} P(\|\sigma_c(\hat{A}_{\tilde{N}, \mu_{\rho_n f}}^*) - \mathbb{E}[\sigma_c(\mathbf{A}_{\mu_{\rho_n f}})]\|_2 > \epsilon) = 0.$$

We prove Theorem 6.2 in the supplementary material [9]. Our final theorem provides us with a method of estimating the expected eigenvalues of graphs drawn from a stochastic block model which will allow us to

search for the correct canonical stochastic block model. Let $f(x, y; \mathbf{p}, \mathbf{Q}, \mathbf{s})$ be a stochastic block model kernel of a kernel probability measure. Let $G_{\mu_{\rho_n f}}$ be a graph distributed according to $\mu_{\rho_n f}$ with adjacency matrix $\mathbf{A}$. Let $\lambda_i(\mathbf{A})$ denote the i -th largest eigenvalue of $\mathbf{A}$. Define the linear integral operator $L_f : L^2([0, 1]) \mapsto L^2([0, 1])$ as

$$(6.20) \quad L_f(t) = \int_0^1 f(x, y; \mathbf{p}, \mathbf{Q}, \mathbf{s}) t(y) dy$$

with eigenvalues denoted $\lambda_i(L_f)$ and orthonormal eigenfunctions $r_i(x)$. Assume $r_i(x)$ is piecewise continuous with finitely many discontinuities.

Theorem 6.3 (Estimation of the Largest Eigenvalues of Stochastic Block Models) For $i = 1, \dots, c$

$$(6.21) \quad \mathbb{E}[\lambda_i(\mathbf{A})] = \lambda_i(L_f) n \rho_n + \mathcal{O}(\sqrt{\rho_n})$$

The full proof of Theorem 6.3 can be found in [9]. The culmination of these theorems results in the following algorithm.

Algorithm 1 Approximate sample Fréchet mean

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

- 1: Compute the average density $\bar{\rho}_n$ of the graphs in M
 - 2: Estimate c and choose $\mathbf{s}$ (see [9] and Remark 6.1).
 - 3: For each $i = 1, \dots, c$ compute $\bar{\lambda}_i = \frac{1}{N} \sum_{k=1}^N \lambda_i(\mathbf{A}^{(k)})$.
 - 4: Randomly initialize $\mathbf{p}$
 - 5: Initialize $\mathbf{Q} = (q_{ij})$ such that $q_{ij} = q$ for all i, j and $\|f(x, y; \mathbf{p}, \mathbf{Q}, \mathbf{s})\|_1 = 1$
 - 6: **while** Relative change in $\mathbf{p}$ and q is large **do**
 - 7: Estimate the gradient of $\sum_{i=1}^c |n \bar{\rho}_n \lambda_i(L_f) - \bar{\lambda}_i|^2$ via centered differences
 - 8: Update $\mathbf{p}$ via a projected gradient descent step
 - 9: Update q such that $\|f(x, y; \mathbf{p}, \mathbf{Q}, \mathbf{s})\|_1 = 1$
 - 10: **end while**
 - 11: Estimate $G_{\tilde{N}, \mu_{\rho_n f}}^*$ with $\hat{G}_{\tilde{N}, \mu_{\rho_n f}}^*$ (Theorem 6.2).
 - 12: **Return:** Estimate of $G_{\tilde{N}, \mu_{\rho_n f}}^*$.
-

Alg. 1 is founded on the following idea: Any graph G can be expressed as the Fréchet mean of some probability measure. For large graphs, our theory shows we may search for a canonical stochastic block model kernel, f , by aligning the eigenvalues of L_f (Steps 6 - 10), and then estimating the sample Fréchet mean of $\mu_{\rho_n f}$ using the set mean graph (Step 11). A full justification of Alg. 1 appears in [9].

Notably, the runtime of Alg. 1 is bottlenecked by Step 11 which has a complexity of $\mathcal{O}(\tilde{N} n^2 c^3)$. Though, for large graphs, taking $\tilde{N} = 1$ is sufficient and reduces the time complexity to $\mathcal{O}(n^2)$.

7 Experimental Validation.

7.1 Assessment, Validation, and Comparison

A brute force computation of the sample Fréchet mean or median graph based on the adjacency spectral pseudo-distance is unrealistic (it requires about $\Omega(n^2 2^{n^2})$ operations), and we therefore do not provide a ground truth. One may consider comparing the Fréchet mean computed here to a Fréchet mean computed with respect to the edit distance for which several optimization algorithms have been proposed (e.g., [3, 10, 18]). While this comparison may be feasible, it is uninformative as the Fréchet mean with respect to the edit distance need not have any resemblance to the Fréchet mean with respect to d_{Ac} .

All the code and data is 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.

7.2 Choice of the Datasets

Graph valued databases have recently been created and made available publicly [25]. These databases are designed for the evaluation of machine learning algorithms and the mean for each class is not provided (even for the edit distance). Consequently, we believe that computing the Fréchet mean of these graph sets provide little scientific value for the purpose of validating our method. Instead we conduct experiments on synthetic datasets that are generated using ensembles of random graphs.

Ensembles of random graphs capture prototypical features of existing real world networks. Because our theoretical analysis and associated algorithms rely on the stochastic block model graphs as the “atoms” that are used to approximate G_N^* , we expect that our algorithm will perform well when computing the Fréchet mean of graphs generated by stochastic block models. Our experimental investigation is thus concerned with the performance of our approach when the families of graph ensembles exhibit structural features that differ from those of the stochastic block models.

Each data set consists of $N = 50$ graphs on $n = 600$ nodes. We consider three different iid data sets of graphs, M_1, M_2, M_3 , drawn from distributions μ_1, μ_2, μ_3 which have the following high level descriptions.

- μ_1 : Barabasi-Albert
- μ_2 : Small world
- μ_3 : Variable community size stochastic block model

The specific parameters of each μ_i are discussed within

each subsection. Note that μ_1 and μ_2 induce graphs with vastly different topologies than those generated by $\mu_{\rho_n f}$. For each dataset, we determine the canonical stochastic block model whose sample Fréchet mean is close to G_N^* and compute $\hat{G}_{N, \mu_{\rho_n f}}^*$.

7.3 Barabasi-Albert approximate sample Fréchet mean For the Barabasi-Albert ensemble, the initial graph is fully connected on $m_0 = 5$ nodes and $m = 5$ edges were added at each step. In Fig. 2 we reorder the nodes by their degree to get a better visual understanding of the similarities between an observation and $\hat{G}_{N, \mu_{\rho_n f}}^*$.

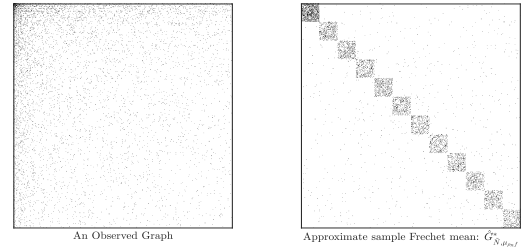


Figure 2: Visualization of a graph in M_1 and the approximate sample Fréchet mean of M_1 , $\hat{G}_{N, \mu_{\rho_n f}}^*$

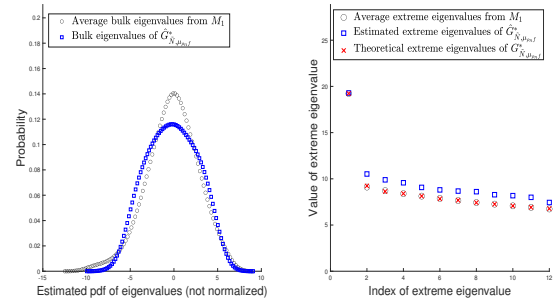


Figure 3: *Left:* The average distribution of bulk eigenvalues from M_1 (black). The distribution of bulk eigenvalues of $\hat{G}_{N, \mu_{\rho_n f}}^*$ (blue). *Right:* The average extreme eigenvalues from M_1 (black). The expected extreme eigenvalues of $\hat{G}_{N, \mu_{\rho_n f}}^*$ (red). The extreme eigenvalues of $\hat{G}_{N, \mu_{\rho_n f}}^*$ (blue).

Fig. 2 is a visual depiction of a graph from M_1 compared to the approximate sample Fréchet mean graph $\hat{G}_{N, \mu_{\rho_n f}}^*$ though we note that there need not be any visual similarity between a graph in M_1 and $\hat{G}_{N, \mu_{\rho_n f}}^*$ since any observation from a distribution μ need not be similar to the mean of μ .

Fig. 3 depicts the alignment of the spectra from the approximate sample Fréchet mean with that of

the average spectra of the graphs from set M_1 . The misalignment in the largest eigenvalues is due to the finite graph approximation.

The theory presented in section 5 only ensures that the c largest eigenvalues can be well approximated by the eigenvalues of the sample mean graph from a stochastic block model. In Fig. 3, we see that the expected eigenvalues (red markers), are near perfect estimates of the average extreme eigenvalues. However there is a notable distance between the extreme eigenvalues of $\hat{G}_{\hat{N}, \mu_{\rho_n f}}^*$ (blue markers), and the expected eigenvalues (red markers). This distance is determined primarily by the size of the graph n and as n increases this distance will decay like $\mathcal{O}(\sqrt{\rho_n})$ per Theorem 6.3.

7.4 Small World approximate sample Fréchet mean The parameters for the Small World ensemble are the number of connected nearest neighbors, $K = 22$, and the probability of rewiring, $\beta = 0.7$.

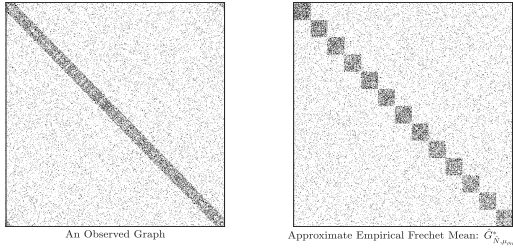


Figure 4: Visualization of a graph in M_2 and the approximate sample Fréchet mean of M_2 , $\hat{G}_{\hat{N}, \mu_{\rho_n f}}^*$

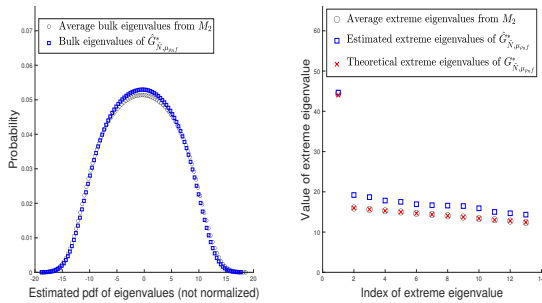


Figure 5: *Left:* The average distribution of bulk eigenvalues from M_2 (black). The distribution of bulk eigenvalues of $\hat{G}_{\hat{N}, \mu_{\rho_n f}}^*$ (blue). *Right:* The average extreme eigenvalues from M_2 (black). The expected extreme eigenvalues of $G_{\hat{N}, \mu_{\rho_n f}}^*$ (red). The extreme eigenvalues of $\hat{G}_{\hat{N}, \mu_{\rho_n f}}^*$ (blue).

Here we see a nice similarity between the adjacency matrices of the two graphs (see Fig. 4). Furthermore

Fig. 5 demonstrates the striking spectral similarity between the two graphs, both in the extreme and bulk eigenvalues.

The alignment of the bulk eigenvalues from the observed set of graphs and the bulk eigenvalues of $\hat{G}_{\hat{N}, \mu_{\rho_n f}}^*$ showcases that the graph structure is entirely determined by the c largest eigenvalues. This has been well known to be true of stochastic block models and in Fig. 5 we see evidence that the Small World ensemble may also be characterized by its largest eigenvalues.

7.5 Variable Community Size approximate sample Fréchet mean The parameters for the stochastic block model in this section are $\mathbf{p} = [0.4, 0.5, 0.6, 0.3, 0.37, 0.65]$, $Q_{ij} = 0.08$, $\mathbf{s} = [\frac{160}{600}, \frac{100}{600}, \frac{60}{600}, \frac{120}{600}, \frac{85}{600}, \frac{75}{600}]$. Fig. 6 visualizes a graph from M_3 and the sample Fréchet mean graph $\hat{G}_{\hat{N}, \mu_{\rho_n f}}^*$.

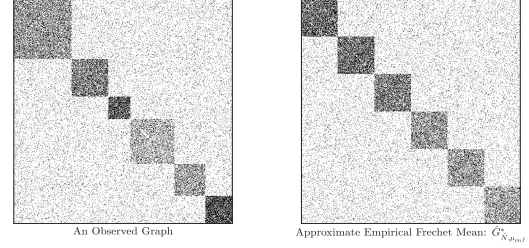


Figure 6: Visualization of a graph in M_3 and the approximate sample Fréchet mean of M_3 , $\hat{G}_{\hat{N}, \mu_{\rho_n f}}^*$

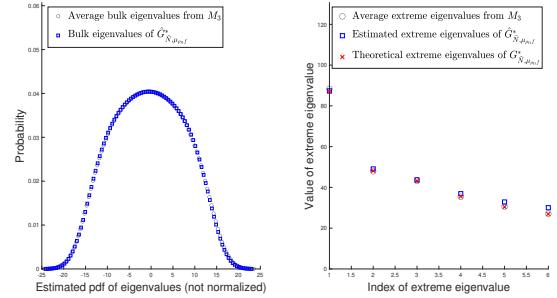


Figure 7: *Left:* The average distribution of bulk eigenvalues from M_3 (black). The distribution of bulk eigenvalues of $\hat{G}_{\hat{N}, \mu_{\rho_n f}}^*$ (blue). *Right:* The average extreme eigenvalues from M_3 (black). The expected extreme eigenvalues of $G_{\hat{N}, \mu_{\rho_n f}}^*$ (red). The extreme eigenvalues of $\hat{G}_{\hat{N}, \mu_{\rho_n f}}^*$ (blue).

In spite of the visual difference between the adjacency matrices of a graph from the set M_3 and the graph $\hat{G}_{\hat{N}, \mu_{\rho_n f}}^*$ respectively (see Fig. 6), we again see a striking similarity between the eigenvalues.

8 Application to Graph Valued Regression

In this section we provide an application of the computation of the sample Fréchet mean: the construction of a regression function in the context where we observe a graph-valued random variable that depends on a real-valued random variable. Our approach is based on the theory developed in [28] where we replace the computation of the sample Fréchet mean with our algorithm. We briefly recall the framework of [28] using our notation. We consider the following scenario. Let $\mu \in \mathcal{M}(\mathcal{G})$, and let T be a random variable with probability density $\mathbb{P}_T(t)$. We consider the random variable formed by the pair G and T , distributed with the joint distribution formed by the product $\mu \times \mathbb{P}_T(t)$. We wish to compute the regression function $\mathbb{E}[G|T=t]$. The authors in [28] propose to compute the following,

$$(8.22) \quad m(t) = \operatorname{argmin}_{G \in \mathcal{G}} \mathbb{E}_{\mu \times \mathbb{P}_T(t)}[s(T, t) d^2(G, G_\mu)],$$

where the expectation in (8.22) is computed jointly over G_μ distributed according to μ , and T , distributed according to $\mathbb{P}_T(t)$, and the bilinear form s is defined by

$$(8.23) \quad s(T, t) = 1 + (T - \mathbb{E}[T]) [\operatorname{var}[T]]^{-1} (t - \mathbb{E}[T]).$$

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

The sample estimate of equation (8.22) is the natural estimate where each unknown term is replaced with the sample alternative as

$$(8.24) \quad \hat{m}(t) = \operatorname{argmin}_{G \in \mathcal{G}} \sum_{i=1}^N s_{i,N}(t) d^2(G, G_i)$$

$$(8.25) \quad s_{i,N}(t) = 1 + (t_i - \bar{T}) \hat{V}(t - \bar{T}).$$

Here we have used $\bar{T}$ and $\hat{V}$ as the sample estimate of the mean and variance of T . The objective in (8.24) can be interpreted as a weighted sample Fréchet mean with weight function $s_{k,N}(t)$. Assume for all t that the graph, $\hat{m}(t)$, satisfies the conditions for Theorem 6.1. This implies the existence of a sequence of stochastic block model kernels depending on t , $\mu_{\rho_n f; t}$, such that, for sufficiently large n ,

$$(8.26) \quad \lim_{N \rightarrow \infty} d_{A_c}(\hat{m}(t), G_{N, \mu_{\rho_n f; t}}^*) < \epsilon \quad a.s.$$

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

8.1 Experimental Validation for Graph Valued Regression We validate the computation of the regression with numerical simulation. We first generate a synthetic data set of graphs by allowing the parameters of the stochastic block model to vary with time. For simplicity we hold the nonzero entries of $\mathbf{Q}$ and $\mathbf{s}$ fixed but allow $\mathbf{p}$ to vary for $t \in [0, 1]$ as

$$(8.27) \quad \rho_n \mathbf{p}(t) = \begin{bmatrix} 0.1 + 0.1t \\ 0.2 + 0.15t \\ 0.35 + 0.2t \end{bmatrix}, \mathbf{s}(t) = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \rho_n q = 0.08.$$

For $T \sim \operatorname{unif}(0, 1)$, the distribution over $\mathcal{G}$ is given as $\mu_{\rho_n f; T}$ where $f(x, y; \mathbf{p}(T), \mathbf{Q}, \mathbf{s})$ is a canonical stochastic block model kernel. For each sample from $\operatorname{unif}(0, 1)$ there is a corresponding sample from the stochastic block model. By construction we know the number of communities in the observed graphs will be constant at $c = 3$ dictating the number of non-zero entries of $\mathbf{p}$ we allow to vary when searching for the stochastic block model kernel in equation (8.26).

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

In an effort of visualization, since we are unable to plot a graph G on the y-axis, we plot in Fig. 8 the largest three eigenvalues of the adjacency matrices of graphs in M (marked by a $\bullet$) and the largest three eigenvalues of $\hat{G}_{N, \mu_{\rho_n f; t'}}^*$ (marked by an $\times$). A vertical line of points in Fig. 8 identifies the largest three eigenvalues of a single graph.

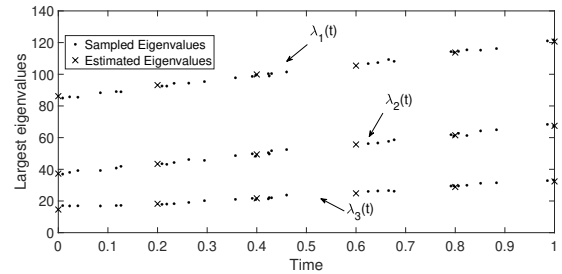


Figure 8: Recovered eigenvalues

The most notable part of Fig. 8 is the construction of a graph that fits the linear regression of each of the largest c eigenvalues simultaneously. To our knowledge, this is the first graph valued linear regression line with respect to a spectral distance.

9 Conclusion.

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

How this approximate sample Fréchet mean is utilized is up to the discretion of the researcher. In section 8 we explore one motivating idea that utilizes the Fréchet mean, termed Fréchet regression in the work in [28]. This is but one example of the utility of the Fréchet mean graph, another interesting application of this graph is to further push the work in [24] which introduces a centered random graph model to capture the variance of a set of observations around a mean graph.

References

- [1] ABBE, E. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research* 18, 1 (2017), 6446–6531.
- [2] BACÁK, M. Computing medians and means in Hadamard spaces *SIAM Journal on Optimization* (2014), pp. 1542–1566.
- [3] BARDAJI, I. AND FERRER, M. AND SANFELIU, A. Computing the barycenter graph by means of the graph edit distance *2010 20th International Conference on Pattern Recognition* (2010)
- [4] BORJA, N., NEGREVERGNE, B., AND YGER, F. Fréchet Mean Computation in Graph Space through Projected Block Gradient Descent. In *ESANN 2020*.
- [5] CHAKRABARTY, A., CHAKRABORTY, A., HAZRA, R. Eigenvalues Outside the Bulk of Inhomogeneous Erdős-Rényi Random Graphs *Jour. of Stat. Physics* (2020)
- [6] CHEN, J., SAAD, Y., AND ZHANG, Z. Graph coarsening: From scientific computing to machine learning arXiv preprint arXiv:2106.11863 (2021).
- [7] FAN, J., FAN, Y., HAN, X., AND LV, J. Asymptotic Theory of Eigenvectors for Random Matrices With Diverging Spikes. *Journal of the American Statistical Association* (2020) pp. 1-14
- [8] FARKAS, I. J. Spectra of “real-world” graphs: Beyond the semicircle law. *Physical Review E* 64, 2 (2001).
- [9] FERGUSON, D. AND MEYER, F. Approximate Fréchet Mean for Data Sets of Sparse Graphs *arXiv preprint arXiv:2105.04062* (2021).
- [10] FERRER, M. AND VALVENY, E. AND SERRATOSA, F. Median graph: A new exact algorithm using a distance based on the maximum common subgraph *Pattern Recognition Letters* 30, 5 (2009), 579–588.
- [11] FARKAS, I. J. Spectra of “real-world” graphs: Beyond the semicircle law. *Physical Review E* 64, 2 (2001).
- [12] FERRER, M., VALVENY, E., SERRATOSA, F., RIESEN, K., AND BUNKE, H. Generalized median graph computation by means of graph embedding in vector spaces. *Pattern Recognition* 43, 4 (2010), 1642–1655.
- [13] FLAXMAN, A., FRIEZE, A., AND FENNER, T. High degree vertices and eigenvalues in the preferential attachment graph. In *Approximation, Randomization, and Combinatorial Optimization. RANDOM 2003 Proceedings* (2003), pp. 264–274.
- [14] FRÉCHET, M. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l’institut Henri Poincaré* 10, 4 (1948), 215–310.
- [15] GINESTET, C. E., LI, J., BALACHANDRAN, P., ROSENBERG, S., AND KOLACZYK, E. D. Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics* 11, 2 (2017), 725–750.
- [16] GIRVAN, M. AND NEWMAN, M. E. J. Community structure in social and biological networks *Proceedings of the National Academy of Sciences*, (2002), 7821-7826
- [17] HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA, 2001.
- [18] JAIN, B. J., AND OBERMAYER, K. Algorithms for the sample mean of graphs. In *Intern. Conf. on Computer Analysis of Images and Patterns* (2009), pp. 351–359.
- [19] JANSON, S. Graphons, cut norm and distance, couplings and rearrangements *NYJM Monographs, State University of New York* (2013).
- [20] JOVANOVIĆ, I. AND STANIĆ, Z. Spectral distances of graphs *Linear algebra and its applications* 436, 5 (2012), 1425–1435.
- [21] LE, C. M., LEVINA, E., AND VERSHYNIN, R. *Concentration of random graphs and application to community detection*. World Scientific, 2018.
- [22] LEE, J. R., GHARAN, S. O., AND TREVISAN, L. Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM* 61, 6 (Dec. 2014), 37:1–37:30.
- [23] LOUKAS, A. Graph Reduction with Spectral and Cut Guarantees. *J. Mach. Learn. Res.*, 20 (2019), 1–42
- [24] LUNAGÓMEZ, S., OLHEDE, S. C., AND WOLFE, P. J. Modeling network populations via graph distances. *Journal of the American Statistical Association* (2020)
- [25] MORRIS, C. KRIEGE, N.M., BAUSE, F., KERSTING, K., MUTZEL, P., AND NEUMANN, M., Tudataset: A collection of benchmark datasets for learning with graphs, *arXiv preprint arXiv:2007.08663* (2020).
- [26] ON, K., KIM, E., KWON, I., YOON, S., AND ZHANG, B. Spectrally Similar Graph Pooling. (2020).
- [27] PENNEC, X. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision* 25, 1 (2006)
- [28] PETERSEN, A., AND MÜLLER, H.-G. Fréchet regression for random objects with euclidean predictors. *Ann. Statist.* 47, 2 (04 2019), 691–719.
- [29] VU, V. *Combinatorial problems in random matrix theory*. Proceedings ICM 4, (2014), 489–508.
- [30] WILLS, P., AND MEYER, F. G. Metrics for graph comparison: A practitioner’s guide. *PLOS ONE* 15, 2 (02 2020), 1–54.
- [31] WILSON, R. C., AND ZHU, P. A study of graph spectra for comparing graphs and trees. *Pattern Recognition* 41, 9 (2008), 2833 – 2841.