



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

Daniel Ferguson and François G. Meyer^(✉) 

Applied Mathematics, University of Colorado at Boulder, Boulder, CO 80309, USA
fmeyer@colorado.edu
<https://francoismeyer.github.io/>

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

Keywords: Fréchet mean and median graphs · statistical network analysis

1 Introduction

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

We characterize the “average” of a sample of graphs $\{G^{(1)}, \dots, G^{(N)}\}$, which are defined on the same vertex set $\{1, \dots, n\}$, with the sample Fréchet mean and median graphs, [6].

Definition 1. *The sample Fréchet mean graphs are solutions to*

$$\widehat{\mu}_N[G] = \operatorname{argmin}_{G \in \mathcal{G}} \frac{1}{N} \sum_{k=1}^N d^2(G, G^{(k)}), \quad (1)$$

and the sample Fréchet median graphs are solutions to

$$\widehat{\mathbf{m}}_N[G] = \operatorname{argmin}_{G \in \mathcal{G}} \frac{1}{N} \sum_{k=1}^N d(G, G^{(k)}). \quad (2)$$

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

Solutions to the minimization problems (1) and (2) always exist, but the minimizers need not be unique. All our results are stated in terms of any of the elements in the set of minimizers of (1) and (2).

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

The vital role played by the Fréchet mean as a location parameter [9, 10, 13], is exemplified in the works of [1, 14], who have created novel families of random graphs by generating random perturbations around a given Fréchet mean graph.

1.1 Our Main Contributions

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

Because sparse graphs provide prototypical models for real networks, our theoretical analysis is significant since it provides a guarantee that this structural property is preserved when computing a sample mean or median. In a similar vein, the authors in [8] construct a sparse median graph, which provides a more interpretable summary, from a set of graphs that are not necessarily sparse.

Our work answers the question raised by the author in [7]: “does the average of two sparse networks/matrices need to be sparse?” Specifically, we prove the following result: the number of edges of the Fréchet mean or median graphs of a set of graphs is bounded by the sample mean number of edges of the graphs in the sample. We prove this result for the graph Hamming distance, and the spectral adjacency pseudometric, using different arguments.

2 Preliminary and Notations

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

$$\mathcal{S} = \{\mathbf{A} \in \{0, 1\}^{n \times n}; \text{where } a_{ij} = a_{ji}, \text{and } a_{i,i} = 0; 1 \leq i < j \leq n\}. \quad (3)$$

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

$$e(\mathbf{A}) = \sum_{1 \leq i < j \leq n} a_{ij}. \quad (4)$$

We denote by $\boldsymbol{\lambda}(\mathbf{A}) = [\lambda_1(\mathbf{A}) \dots \lambda_n(\mathbf{A})]$, the vector of eigenvalues of \mathbf{A} , with the convention that $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$.

2.1 Distances Between Graphs

In this work, we consider two metrics: the Hamming distance and the spectral adjacency pseudometric. We briefly recall the definitions of these.

Definition 2. Let $G, G' \in \mathcal{G}$ be two unweighted graphs with known vertex correspondence and with adjacency matrix \mathbf{A} and \mathbf{A}' respectively. We define the Hamming distance between G and G' as

$$d_H(\mathbf{A}, \mathbf{A}') \stackrel{\text{def}}{=} \sum_{1 \leq i < j \leq n} |a_{ij} - a'_{ij}| = e(\mathbf{A}) + e(\mathbf{B}) - 2 \sum_{1 \leq i < j \leq n} a_{ij} b_{ij}. \quad (5)$$

The Hamming distance is very sensitive to fine scale fluctuations of the graph connectivity. In contrast, a metric based on the eigenvalues of the adjacency matrix can quantify configurational changes that occur on a graph at many more scales [5, 17].

Definition 3. Let $G, G' \in \mathcal{G}$ with adjacency matrix \mathbf{A} and \mathbf{A}' respectively. We define the adjacency spectral pseudometric as the ℓ_2 norm between the vectors of eigenvalues $\lambda(\mathbf{A})$ and $\lambda(\mathbf{A}')$ of \mathbf{A} and \mathbf{A}' respectively,

$$d_\lambda(\mathbf{A}, \mathbf{A}') = \|\lambda(\mathbf{A}) - \lambda(\mathbf{A}')\|_2. \quad (6)$$

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

3 Main Results

In the following, we consider a set of N unweighted simple labeled graphs, $\{G^{(1)}, \dots, G^{(N)}\}$, with vertex set $\{1, \dots, n\}$. We denote by $\mathbf{A}^{(k)}$ the adjacency matrix of graph $G^{(k)}$. We equip the set \mathcal{G} of all unweighted simple graphs on n nodes with a pseudometric, or a metric, d . The Fréchet mean and median graphs encode two notions of centrality (1) and (2) that minimise the following dispersion function, also called the Fréchet function.

Definition 4. We denote by $\widehat{F}_q(\mathbf{A})$ the sample Fréchet function associated with a sample Fréchet median ($q = 1$) or mean ($q = 2$),

$$\widehat{F}_q(\mathbf{A}) = \frac{1}{N} \sum_{k=1}^N d^q(\mathbf{A}, \mathbf{A}^{(k)}). \quad (7)$$

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

Definition 5. *The sample mean and variance of the number of edges are defined by*

$$\bar{e}_N = \frac{1}{N} \sum_{k=1}^N e(\mathbf{A}^{(k)}), \quad \text{and} \quad \sigma_N^2(e) = \frac{1}{N} \sum_{k=1}^N [e(\mathbf{A}^{(k)})]^2 - [\bar{e}_N]^2. \quad (8)$$

We now turn our attention to the main problem. We consider the following question: if the graphs $G^{(1)}, \dots, G^{(N)}$ all have a similar edge density, can one determine the edge density of the sample Fréchet mean or median graphs? and does that number of edges depend on the choice of metric d in (1) and (2)? We answer both questions in the following theorem.

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

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

$$e_{\hat{\mathbf{m}}} < 2\bar{e}_N + \frac{\sigma_N(e)}{\sqrt{2}}, \quad \text{and} \quad e_{\widehat{\mathbf{m}}} < 2\bar{e}_N, \quad (9)$$

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

$$e_{\hat{\mathbf{m}}} < 9\bar{e}_N, \quad \text{and} \quad e_{\widehat{\mathbf{m}}} < 9\bar{e}_N. \quad (10)$$

Proof. The proof is a direct consequence of Lemmata 6 and 12.

Remark 1. When the graph $G^{(k)}$ are sampled from the inhomogeneous Erdős-Rényi random graph probability space $\mathcal{G}(n, \mathbf{P})$ [3], and if the distance on \mathcal{G} is the Hamming distance, then $\hat{\mathbf{m}}_N[\mathbf{A}] = \widehat{\mathbf{m}}_N[\mathbf{A}]$ with high probability [15]. In this case, a tight bound on $e_{\hat{\mathbf{m}}}$ or $e_{\widehat{\mathbf{m}}}$ in (9) is $2\bar{e}_N$, which – unlike (9) – does not involve $\sigma_N(e)$.

The fact that we overestimate the bound on $e_{\hat{\mathbf{m}}}$ by the addition of the term $\sigma_N(e)/\sqrt{2}$ comes from our technique of proof, which relies on an estimate of the Fréchet function. As explained in Remark 4, our estimate of the Fréchet function is almost tight; it does include the term $\sigma_N(e)$, as it should.

Finally, the following corollary answers the question raised by the author in [7]: “does the average of two sparse networks/matrices need to be sparse?”

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

$$e(\mathbf{A}^{(k)}) = \mathcal{O}(n^2), \quad \text{but} \quad e(\mathbf{A}^{(k)}) = \omega(n). \quad (11)$$

Then the sample Fréchet mean and median graphs – computed according to either the Hamming distance or the adjacency spectral pseudometric – are sparse, as defined by (11).

Proof of Corollary 1. *The corollary is a direct consequence of Theorem 1.*

4 Proofs of the Main Result

We give in the following the proof of Theorem 1. The key observation is that it is relatively easy to derive tight bounds on the number of edges of the sample Fréchet median graph. Inspired by the results in [15] that show that for large classes of random graphs the sample Fréchet median and mean graphs are identical, we prove that the bounds derived for the Fréchet median graphs also hold for the Fréchet mean graphs.

Our analysis begins in Subsect. 4.1 with the sample median graphs computed using the Hamming distance, we then move to the sample mean graphs in Subsect. 4.2. In Subsects. 4.4 and 4.5, we extend these results to the sample mean and median graphs computed with the adjacency spectral pseudometric.

When possible, we use the probability space $\mathcal{G}(n, \mathbf{P})$ of inhomogeneous Erdős-Rényi random graphs [3], equipped with the Hamming distance to test the tightness of our results [15].

4.1 The Median Graphs Computed Using the Hamming Distance

The Hamming distance, by nature, promotes sparsity [5, 17], and we therefore expect that the volumes of the sample Fréchet mean and median graphs computed with this distance be similar to the sample mean number of edges.

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

Lemma 1. *The adjacency matrix $\widehat{\mathbf{m}}_N[\mathbf{A}]$ of a sample median graph $\widehat{\mathbf{m}}_N[G]$ is given by the majority rule,*

$$\left[\widehat{\mathbf{m}}_N[\mathbf{A}] \right]_{ij} = \begin{cases} 0 & \text{if } \sum_{k=1}^N a_{ij}^{(k)} < N/2, \\ 1 & \text{otherwise.} \end{cases} \quad \forall i, j \in \{1, \dots, n\}. \quad (12)$$

Proof of Lemma 1. *The result is classic and we omit the proof, which can be found for instance in [4].*

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

Lemma 2. *Let \bar{e}_N be the sample mean number of edges, given by (8). Then the number of edges of a Fréchet median graph $\widehat{\mathbf{m}}_N[G]$ is bounded by*

$$e_{\widehat{\mathbf{m}}} \leq 2\bar{e}_N. \quad (13)$$

Remark 2. The bound (13) is tight for large N . Indeed, consider a sample of $2N$ graphs, where

$$G^{(k)} = \begin{cases} \text{the complete graph } K_n & \text{if } 1 \leq k \leq N+1, \\ \text{the empty graph} & \text{if } N+2 \leq k \leq 2N. \end{cases} \quad (14)$$

A Fréchet median graph $\widehat{\mathbf{m}}_N[\mathbf{A}]$, given by the majority rule (12) is K_n , and thus $e_{\widehat{\mathbf{m}}} = n(n-1)/2$. On the other hand, the sample mean number of edges is $\bar{e}_N = e_{\widehat{\mathbf{m}}}/2 + e_{\widehat{\mathbf{m}}}/(2N)$. As the sample size N goes to infinity, we have

$$\lim_{N \rightarrow \infty} e_{\widehat{\mathbf{m}}} = 2\bar{e}_N, \quad (15)$$

which proves that the bound (13) is asymptotically tight.

Proof of Lemma 2. Let $\mathcal{E}_{\widehat{\mathbf{m}}} = \{(i, j), i < j, [\widehat{\mathbf{m}}_N[\mathbf{A}]]_{ij} = 1\}$ be the set of edges of $\widehat{\mathbf{m}}_N[\mathbf{G}]$. We have $|\mathcal{E}_{\widehat{\mathbf{m}}}| = e_{\widehat{\mathbf{m}}}$. Now,

$$\sum_{k=1}^N e(\mathbf{A}^{(k)}) = \sum_{1 \leq i < j \leq n} \sum_{k=1}^N a_{ij}^{(k)} = \sum_{i,j \in \mathcal{E}_{\widehat{\mathbf{m}}}} \sum_{k=1}^N a_{ij}^{(k)} + \sum_{i,j \in \mathcal{E}_{\widehat{\mathbf{m}}}^c} \sum_{k=1}^N a_{ij}^{(k)}. \quad (16)$$

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

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

whence we conclude

$$e_{\widehat{\mathbf{m}}} \leq \frac{2}{N} \sum_{k=1}^N e(\mathbf{A}^{(k)}) = 2\bar{e}_N. \quad (17)$$

□

4.2 The Mean Graphs Computed Using the Hamming Distance

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

Lemma 3. Let \mathbf{A} and \mathbf{B} be the adjacency matrices of two unweighted graphs with number of edges $e(\mathbf{A})$ and $e(\mathbf{B})$ respectively. Then

$$|e(\mathbf{A}) - e(\mathbf{B})| \leq d_H(\mathbf{A}, \mathbf{B}). \quad (18)$$

Proof of Lemma 3. The proof is elementary and is skipped.

Next, we derive an upper bound on the deviation of the volume of a Fréchet mean, $e_{\widehat{\mu}}$, away from the sample average volume, \bar{e}_N , given by (8).

Lemma 4. Let $\widehat{\mu}_N[\mathbf{A}]$ be the adjacency matrix of a sample Fréchet mean computed using the Hamming distance, with $e_{\widehat{\mu}}$ edges. Let \bar{e}_N be the sample mean number of edges. Then

$$\left[e_{\widehat{\mu}} - \bar{e}_N \right]^2 \leq \frac{1}{N} \sum_{k=1}^N d_H^2(\widehat{\mu}_N[\mathbf{A}], \mathbf{A}^{(k)}) = \widehat{F}_2(\widehat{\mu}_N[\mathbf{A}]). \quad (19)$$

Remark 3. This bound is not tight. We consider again the probability space of inhomogeneous Erdős-Rényi random graphs equipped with the Hamming distance. In that case, one can show that the population Fréchet mean and median coincide [15], and the adjacency matrix of the population Fréchet mean graph, $\mu[\mathbf{A}]$, is given by the majority rule,

$$[\mu[\mathbf{A}]]_{ij} = \begin{cases} 1 & \text{if } p_{ij} > 1/2, \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

Also, the population Fréchet function, F_2 , evaluated at $\mu[\mathbf{A}]$ is given by [15]

$$F_2(\mu[\mathbf{A}]) = \left[\sum_{1 \leq i < j \leq n} p_{ij} - \sum_{(i,j) \in \mathcal{E}(\mu[\mathbf{A}])} (2p_{ij} - 1) \right]^2 + \sum_{1 \leq i < j \leq n} p_{ij}(1 - p_{ij}), \quad (21)$$

where $\mathcal{E}(\mu[\mathbf{A}])$ is the set of edges of the population Fréchet mean, $\mu[\mathbf{A}]$. We claim that the lower bound on $\hat{F}_2(\hat{\mu}_N[\mathbf{A}])$ in (19),

$$[\bar{e}_N - e_{\hat{\mu}}]^2, \quad (22)$$

can be identified with the first term of $F_2(\mu[\mathbf{A}])$ in (21),

$$\left[\sum_{1 \leq i < j \leq n} p_{ij} - \sum_{(i,j) \in \mathcal{E}(\mu[\mathbf{A}])} (2p_{ij} - 1) \right]^2. \quad (23)$$

Indeed, the first sum inside (23) is the population mean number of edges, $\mathbb{E}[e]$, which matches the sample mean \bar{e}_N in (22). Also, the second sum in (23) is bounded by $e(\mu[\mathbf{A}])$, the number of edges of the population Fréchet mean,

$$0 < \sum_{(i,j) \in \mathcal{E}(\mu[\mathbf{A}])} (2p_{ij} - 1) < \sum_{(i,j) \in \mathcal{E}(\mu[\mathbf{A}])} 1 = e(\mu[\mathbf{A}]). \quad (24)$$

The number of edges $e(\mu[\mathbf{A}])$ matches the sample estimate, $e_{\hat{\mu}}$, in (22). In summary, the first term (23) of the population Fréchet function (21) matches the corresponding sample estimate (22).

However, the second term, $\sum_{1 \leq i < j \leq n} p_{ij}(1 - p_{ij})$ in (21), which accounts for the variance of the $n(n-1)/2$ independent Bernoulli edges, is not present in the lower bound on in $F_2(\mu[\mathbf{A}])$ given by (19), confirming that the lower bound in (19) is missing a variance term, and is therefore not tight.

Proof of Lemma 4. *Because of Lemma 3, we have*

$$|e(\mathbf{A}^{(k)}) - e_{\hat{\mu}}|^2 \leq d_H^2(\hat{\mu}_N[\mathbf{A}], \mathbf{A}^{(k)}). \quad (25)$$

Now, the function

$$x \longmapsto (e_{\hat{\mu}} - x)^2 \quad (26)$$

is strictly convex so,

$$|\bar{e}_N - e_{\hat{\mu}}|^2 = \left| \frac{1}{N} \sum_{k=1}^N e(\mathbf{A}^{(k)}) - e_{\hat{\mu}} \right|^2 < \frac{1}{N} \sum_{k=1}^N \left| e(\mathbf{A}^{(k)}) - e_{\hat{\mu}} \right|^2, \quad (27)$$

and substituting (25) for each k in (27), we get the advertised result. \square

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

Lemma 5. *Let \bar{e}_N and $\sigma_N^2(e)$ be the sample mean and variance of the number of edges (see (8)). Then the Fréchet function $\hat{F}_2(\hat{\mathbf{m}}_N[\mathbf{A}])$ evaluated at a Fréchet median graph is bounded by*

$$\hat{F}_2(\hat{\mathbf{m}}_N[\mathbf{A}]) \leq 2[\bar{e}_N]^2 + \sigma_N^2(e). \quad (28)$$

Remark 4. As explained in Remark 3, when the graphs $G^{(k)}$ are sampled from $\mathcal{G}(n, \mathbf{P})$, then the population Fréchet mean and median graphs coincide, $\mu[G] = \mathbf{m}[G]$. Also, the population Fréchet function $F_2(\mathbf{m}[\mathbf{A}])$ evaluated at a population Fréchet median graph is given by

$$F_2[\mathbf{m}[\mathbf{A}]] = \left[\sum_{1 \leq i < j \leq n} p_{ij} - \sum_{(i,j) \in \mathcal{E}(\mathbf{m}[\mathbf{A}])} (2p_{ij} - 1) \right]^2 + \sum_{1 \leq i < j \leq n} p_{ij}(1 - p_{ij}), \quad (29)$$

where the term $\sum_{(i,j) \in \mathcal{E}(\mathbf{m}[\mathbf{A}])} (2p_{ij} - 1)$ is always positive (since the median graphs are constructed using the majority rule (12)). Therefore, we have

$$F_2[\mathbf{m}[\mathbf{A}]] \leq \left[\sum_{1 \leq i < j \leq n} p_{ij} \right]^2 + \sum_{1 \leq i < j \leq n} p_{ij}(1 - p_{ij}). \quad (30)$$

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

$$F_2(\mathbf{m}[\mathbf{A}]) \leq [\mathbb{E}[e]]^2 + \text{var}[e], \quad (31)$$

where e denotes the number of edges in graphs sampled from $\mathcal{G}(n, \mathbf{P})$. If we replace $\mathbb{E}[e]$ and $\text{var}[e]$ by their respective sample estimates, \bar{e}_N and $\sigma_N^2(e)$, then the bound (28) is only slightly worse (by a factor 2 in front of \bar{e}_N) than the population bound, (31). Interestingly, the variance of the number of edges is present in both expressions.

Proof of Lemma 5. *From (5), one can derive the following expression for the Hamming distance from a Fréchet median graph $\hat{\mathbf{m}}_N[G]$ to a graph $G^{(k)}$,*

$$d_H(\hat{\mathbf{m}}_N[\mathbf{A}], \mathbf{A}^{(k)}) = e_{\hat{\mathbf{m}}} + e(\mathbf{A}^{(k)}) - 2 \sum_{(i,j) \in \mathcal{E}_{\hat{\mathbf{m}}}} a_{ij}^{(k)}, \quad (32)$$

where we recall that $\mathcal{E}_{\widehat{\mathbf{m}}} = \left\{ (i, j), i < j, [\widehat{\mathbf{m}}_N[\mathbf{A}]]_{ij} = 1 \right\}$ is the set of edges of $\widehat{\mathbf{m}}_N[G]$. Taking the square of the Hamming distance given by (32), and summing over all the graphs, yields

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

Expanding all the terms, and using the definition of $\sigma_N^2(e)$ and \bar{e}_N in (8), we get

$$\begin{aligned} \widehat{F}_2(\widehat{\mathbf{m}}_N[\mathbf{A}]) &= [e_{\widehat{\mathbf{m}}}]^2 + 2e_{\widehat{\mathbf{m}}} \bar{e}_N + \sigma_N^2(e) + [\bar{e}_N]^2 + \frac{4}{N} \sum_{k=1}^N \left[\sum_{(i,j) \in \mathcal{E}_{\widehat{\mathbf{m}}}} a_{ij}^{(k)} \right]^2 \\ &\quad - \frac{4}{N} \sum_{k=1}^N e(\mathbf{A}^{(k)}) \left[\sum_{(i,j) \in \mathcal{E}_{\widehat{\mathbf{m}}}} a_{ij}^{(k)} \right] - 4e_{\widehat{\mathbf{m}}} \left[\sum_{(i,j) \in \mathcal{E}_{\widehat{\mathbf{m}}}} \frac{1}{N} \sum_{k=1}^N a_{ij}^{(k)} \right] \\ &= [e_{\widehat{\mathbf{m}}} + \bar{e}_N]^2 + \sigma_N^2(e) + 4 \frac{1}{N} \sum_{k=1}^N \left[\sum_{(i,j) \in \mathcal{E}_{\widehat{\mathbf{m}}}} a_{ij}^{(k)} \right]^2 \\ &\quad - \frac{4}{N} \sum_{k=1}^N e(\mathbf{A}^{(k)}) \left[\sum_{(i,j) \in \mathcal{E}_{\widehat{\mathbf{m}}}} a_{ij}^{(k)} \right] - 4e_{\widehat{\mathbf{m}}} \left[\sum_{(i,j) \in \mathcal{E}_{\widehat{\mathbf{m}}}} \frac{1}{N} \sum_{k=1}^N a_{ij}^{(k)} \right]. \quad (33) \end{aligned}$$

Now, because of the definition of the median graphs (12), we have the following upper bound

$$-4e_{\widehat{\mathbf{m}}} \left[\sum_{(i,j) \in \mathcal{E}_{\widehat{\mathbf{m}}}} \frac{1}{N} \sum_{k=1}^N a_{ij}^{(k)} \right] \leq -2[e_{\widehat{\mathbf{m}}}]^2. \quad (34)$$

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

$$-4 \sum_{k=1}^N e(\mathbf{A}^{(k)}) \sum_{(i,j) \in \mathcal{E}_{\widehat{\mathbf{m}}}} a_{ij}^{(k)} \leq -4 \sum_{k=1}^N \left[\sum_{(i,j) \in \mathcal{E}_{\widehat{\mathbf{m}}}} a_{ij}^{(k)} \right]^2. \quad (35)$$

Finally, after substituting (34) and (35) into (33), we get the bound announced in the lemma,

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

4.3 The Number of Edges of $\widehat{\mathbf{m}}_N[G]$ and $\widehat{\mu}_N[G]$ when $d = d_H$

The following lemma provides the bounds given by Theorem 1 when d is the Hamming distance.

Lemma 6. *Let $\{G^{(1)}, \dots, G^{(N)}\}$ be a sample of unweighted simple labeled graphs with vertex set $\{1, \dots, n\}$. Let $\widehat{\mu}_N[\mathbf{A}]$ be the adjacency matrix of a sample Fréchet mean graph, and $\widehat{\mathbf{m}}_N[\mathbf{A}]$ be the adjacency matrix of a sample Fréchet median graph, computed according to the Hamming distance. Then*

$$e(\widehat{\mu}_N[\mathbf{A}]) < 2\bar{e}_N + \frac{\sigma_N(e)}{\sqrt{2}}, \quad \text{and} \quad e(\widehat{\mathbf{m}}_N[\mathbf{A}]) \leq 2\bar{e}_N. \quad (36)$$

Proof of Lemma 6. *The bound on $e(\widehat{\mathbf{m}}_N[\mathbf{A}])$ is a straightforward consequence of Lemma 4. Indeed, (13) and (8) yield the bound in (36),*

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

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

$$\frac{1}{N} \sum_{k=1}^N d_H^2(\widehat{\mu}_N[\mathbf{A}], \mathbf{A}^{(k)}) \leq \frac{1}{N} \sum_{k=1}^N d_H^2(\widehat{\mathbf{m}}_N[\mathbf{A}], \mathbf{A}^{(k)}). \quad (37)$$

Using (19) as a lower bound and (28) as an upper bound in (37), we get

$$\left[e_{\widehat{\mu}} - \bar{e}_N \right]^2 < 2[\bar{e}_N]^2 + \sigma_N^2(e),$$

and thus

$$|e_{\widehat{\mu}} - \bar{e}_N| \leq \sqrt{2[\bar{e}_N]^2 + \sigma_N^2(e)} \leq \frac{1}{\sqrt{2}} \left\{ \sqrt{2\bar{e}_N} + \sigma_N(e) \right\} = \bar{e}_N + \frac{\sigma_N(e)}{\sqrt{2}}, \quad (38)$$

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

4.4 The Mean Graphs Computed Using the Adjacency Spectral Pseudometric

The technical difficulty in defining the sample Fréchet mean and median graphs according to the adjacency spectral pseudometric stems from the fact that the sample Fréchet function, $\widehat{F}_q(\mathbf{A})$, is defined in the spectral domain, but the domain over which the optimization takes place is the matrix domain. This leads to the definition of the set, Λ , of real spectra that are realizable by adjacency matrices of unweighted graphs (elements of \mathcal{S} , defined by (3)) [11],

$$\Lambda = \{ \boldsymbol{\lambda}(\mathbf{A}) = [\lambda_1(\mathbf{A}) \dots \lambda_n(\mathbf{A})] ; \text{where } \mathbf{A} \in \mathcal{S} \}. \quad (39)$$

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

$\lambda(\mathbf{A}^{(k)})$ be the spectrum of $\mathbf{A}^{(k)}$. The adjacency matrix, $\widehat{\mu}_N[\mathbf{A}]$, of a sample Fréchet mean graph computed according to the adjacency spectral pseudometric, has a vector of eigenvalues, $\lambda(\widehat{\mu}_N[\mathbf{A}]) \in \Lambda$, that satisfies

$$\lambda(\widehat{\mu}_N[\mathbf{A}]) = \operatorname{argmin}_{\lambda \in \Lambda} \sum_{k=1}^N \|\lambda - \lambda(\mathbf{A}^{(k)})\|^2. \quad (40)$$

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

$$\lambda(\widehat{\mathbf{m}}_N[\mathbf{A}]) = \operatorname{argmin}_{\lambda \in \Lambda} \sum_{k=1}^N \|\lambda - \lambda(\mathbf{A}^{(k)})\|. \quad (41)$$

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

Lemma 7. *Let $G \in \mathcal{G}$ with adjacency matrix \mathbf{A} . Let $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$ be the eigenvalues of \mathbf{A} . Then*

$$2e(\mathbf{A}) = \sum_{i=1}^n \lambda_i^2(\mathbf{A}) = \|\lambda(\mathbf{A})\|_2^2. \quad (42)$$

Proof of Lemma 7. *The result is classic; see for instance [2, 16].*

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

Lemma 8. *Let $\widehat{\mathbb{E}}_N[\lambda(\mathbf{A})] = \frac{1}{N} \sum_{k=1}^N \lambda(\mathbf{A}^{(k)})$ be the sample mean spectrum. Then*

$$\frac{1}{2} \left\| \widehat{\mathbb{E}}_N[\lambda(\mathbf{A})] \right\|^2 \leq \bar{e}_N, \quad (43)$$

where \bar{e}_N is the sample mean number of edges, given by (8).

Proof of Lemma 8. *The result is a straightforward consequence of the convexity of the norm combined with (42).*

If Λ were to be a convex set, then the spectrum of a sample Fréchet mean graph would simply be the sample mean spectrum, which would minimize (40). Unfortunately, Λ is not convex [12]. We can nevertheless relate the spectrum of a sample Fréchet mean graph, $\lambda(\widehat{\mu}_N[\mathbf{A}])$, to the mean spectrum $\widehat{\mathbb{E}}_N[\lambda(\mathbf{A})]$. We take a short detour to build some intuition about the geometric position of the spectrum of $\widehat{\mu}_N[\mathbf{A}]$ with respect to $\lambda(\mathbf{A}^{(1)}), \dots, \lambda(\mathbf{A}^{(N)})$.

Warm-Up: The Sample Mean Spectrum. Let $\{G^{(1)}, \dots, G^{(N)}\}$ be a sample of unweighted simple labeled graphs with vertex set $\{1, \dots, n\}$. Let $\mathbf{A}^{(k)}$ be the adjacency matrix of graph $G^{(k)}$, and let $\lambda(\mathbf{A}^{(k)})$ be the spectrum of $\mathbf{A}^{(k)}$.

Lemma 9. Let $\widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\mathbf{A})]$ be the sample mean spectrum. Then $\exists k_0 \in \{1, \dots, N\}$ such that

$$\|\boldsymbol{\lambda}(\mathbf{A}^{(k_0)})\| \leq \|\widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\mathbf{A})]\|. \quad (44)$$

Proof of Lemma 9. A proof by contradiction is elementary.

Using the characterization of a sample Fréchet mean graph, $\widehat{\boldsymbol{\mu}}_N[\mathbf{A}]$, given by (40), we can extend the above lemma to $\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\mathbf{A}])$, and derive the following result.

Lemma 10. Let $\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\mathbf{A}])$ be the spectrum of a sample Fréchet mean graph. Let \bar{e}_N be the sample mean number of edges of the graphs $G^{(1)}, \dots, G^{(N)}$. Then

$$\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\mathbf{A}])\| \leq 3\sqrt{2\bar{e}_N}. \quad (45)$$

Proof of Lemma 10. Because of Lemma 9,

$$\exists k_0 \in \{1, \dots, N\}, \|\boldsymbol{\lambda}(\mathbf{A}^{(k_0)})\| \leq \|\widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\mathbf{A})]\|. \quad (46)$$

Now, because of Lemma 8, (46) implies that

$$\|\boldsymbol{\lambda}(\mathbf{A}^{(k_0)})\| \leq \sqrt{2\bar{e}_N}. \quad (47)$$

Because the vector $\boldsymbol{\lambda}(\mathbf{A}^{(k_0)})$ is in Λ (defined by (39)), we have

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

Expanding the norms squared on both sides yields

$$\begin{aligned} \|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\mathbf{A}])\|^2 - 2\langle \boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\mathbf{A}]), \widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\mathbf{A})] \rangle + \frac{1}{N} \sum_{k=1}^N \|\boldsymbol{\lambda}(\mathbf{A}^{(k)})\|^2 \\ \leq \|\boldsymbol{\lambda}(\mathbf{A}^{(k_0)})\|^2 - 2\langle \boldsymbol{\lambda}(\mathbf{A}^{(k_0)}), \widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\mathbf{A})] \rangle + \frac{1}{N} \sum_{k=1}^N \|\boldsymbol{\lambda}(\mathbf{A}^{(k)})\|^2. \end{aligned} \quad (48)$$

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

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

and therefore

$$\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\mathbf{A}])\| \leq \|\boldsymbol{\lambda}(\mathbf{A}^{(k_0)})\| + 2\|\widehat{\mathbb{E}}_N[\boldsymbol{\lambda}(\mathbf{A})]\|. \quad (49)$$

Finally, using Lemma 8 and (47) in the equation above, we obtain

$$\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\mathbf{A}])\| \leq 3\sqrt{2\bar{e}_N}, \quad (50)$$

which completes the proof of the bound on the spectrum of the Fréchet mean. \square

4.5 The Median Graphs Computed Using the Adjacency Spectral Pseudometric

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

Lemma 11. *Let $\lambda(\widehat{\mathbf{m}}_N[\mathbf{A}])$ be the spectrum of a sample Fréchet median graph. Let \bar{e}_N be the sample mean number of edges of the graphs $G^{(1)}, \dots, G^{(N)}$. Then,*

$$\|\lambda(\widehat{\mathbf{m}}_N[\mathbf{A}])\| \leq 3\sqrt{2\bar{e}_N}. \quad (51)$$

Proof of Lemma 11. *The function Φ ,*

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

is strictly convex, and therefore

$$\Phi(\widehat{\mathbb{E}}_N[\lambda(\mathbf{A})]) = \Phi\left(\frac{1}{N} \sum_{k=1}^N \lambda(\mathbf{A}^{(k)})\right) \leq \frac{1}{N} \sum_{k=1}^N \Phi(\lambda(\mathbf{A}^{(k)})). \quad (52)$$

Now, the right-hand side of (52) is the Fréchet function evaluated at one of its minimizers. Thus $F_1(\lambda(\widehat{\mathbf{m}}_N[\mathbf{A}]))$, is smaller than $F_1(\lambda(\mathbf{A}^{(k_0)}))$, where $\mathbf{A}^{(k_0)}$ is defined in Lemma 9, and (52) becomes

$$\|\lambda(\widehat{\mathbf{m}}_N[\mathbf{A}]) - \widehat{\mathbb{E}}_N[\lambda(\mathbf{A})]\| \leq \frac{1}{N} \sum_{k=1}^N \|\lambda(\mathbf{A}^{(k_0)}) - \lambda(\mathbf{A}^{(k)})\|. \quad (53)$$

Also, because of Lemma 8 and (47), we get

$$\frac{1}{N} \sum_{k=1}^N \|\lambda(\mathbf{A}^{(k_0)}) - \lambda(\mathbf{A}^{(k)})\| \leq \|\lambda(\mathbf{A}^{(k_0)})\| + \sqrt{2\bar{e}_N} \leq 2\sqrt{2\bar{e}_N}. \quad (54)$$

Combining (53) and (54), and using Lemma 8 we conclude that

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

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

4.6 The Number of Edges of $\widehat{\mathbf{m}}_N[\mathbf{G}]$ and $\widehat{\mu}_N[\mathbf{G}]$ when $d = d_\lambda$

The following lemma provides the bounds given by Theorem 1 when d is the spectral adjacency pseudometric.

Lemma 12. *Let $\{G^{(1)}, \dots, G^{(N)}\}$ be a sample of unweighted simple labeled graphs with vertex set $\{1, \dots, n\}$. We consider a sample Fréchet mean, $\widehat{\mu}_N[\mathbf{A}]$, and a sample Fréchet median, $\widehat{\mathbf{m}}_N[\mathbf{A}]$, computed according to the spectral adjacency pseudometric. Then*

$$\max\{e(\widehat{\mu}_N[\mathbf{A}]), e(\widehat{\mathbf{m}}_N[\mathbf{A}])\} \leq 9\bar{e}_N, \quad (55)$$

where \bar{e}_N is the sample mean number of edges given by (8).

Proof of Lemma 12. *We first analyse the case of a sample Fréchet mean graph; a sample Fréchet median graph is handled in the same way. From lemmata 10 and 11, we have*

$$\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\mathbf{A}])\|^2 \leq 18 \bar{e}_N. \quad (56)$$

Now, from (42) we have $e(\widehat{\boldsymbol{\mu}}_N[\mathbf{A}]) = \frac{1}{2}\|\boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}}_N[\mathbf{A}])\|^2$, and therefore

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

which completes the proof of the lemma. \square

References

1. Banks, D., Constantine, G.: Metric models for random graphs. *J. Classif.* **15**(2), 199–223 (1998)
2. Bapat, R.B.: *Graphs and Matrices*. UTX, vol. 27. Springer, London (2010). <https://doi.org/10.1007/978-1-84882-981-7>
3. Bollobás, B., Janson, S., Riordan, O.: The phase transition in inhomogeneous random graphs. *Random Struct. Algorithms* **31**(1), 3–122 (2007)
4. Devroye, L., Györfi, L., Lugosi, G.: *A Probabilistic Theory of Pattern Recognition*, vol. 31. Springer Science & Business Media (2013)
5. Donnat, C., Holmes, S.: Tracking network dynamics: a survey using graph distances. *Ann. Appl. Stat.* **12**(2), 971–1012 (2018)
6. Fréchet, M.: Les espaces abstraits et leur utilité en statistique théorique et même en statistique appliquée. *Journal de la Société Française de Statistique* **88**, 410–421 (1947)
7. Ginestet, C.E., Li, J., Balachandran, P., Rosenberg, S., Kolaczyk, E.D.: Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* **11**(2), 725–750 (2017)
8. Han, F., Han, X., Liu, H., Caffo, B., et al.: Sparse median graphs estimation in a high-dimensional semiparametric model. *Ann. App. Stat.* **10**(3), 1397–1426 (2016)
9. Jain, B.J.: On the geometry of graph spaces. *Discret. Appl. Math.* **214**, 126–144 (2016)
10. Jain, B.J., Obermayer, K.: Learning in Riemannian orbifolds. arXiv preprint [arXiv:1204.4294](https://arxiv.org/abs/1204.4294) (2012)
11. Johnson, C.R., Marijuán, C., Paparella, P., Pisonero, M.: The NIEP. In: *Operator Theory, Operator Algebras, and Matrix Theory*. OTAA, vol. 267, pp. 199–220. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-72449-2_10
12. Knudsen, C., McDonald, J.: A note on the convexity of the realizable set of eigenvalues for nonnegative symmetric matrices. *Electron. J. Linear Algebra* **8**, 110–114 (2001)
13. Kolaczyk, E.D., Lin, L., Rosenberg, S., Walters, J., Xu, J., et al.: Averages of unlabeled networks: geometric characterization and asymptotic behavior. *Ann. Stat.* **48**(1), 514–538 (2020)
14. Lunagómez, S., Olhede, S.C., Wolfe, P.J.: Modeling network populations via graph distances. *J. Am. Stat. Assoc.* **116**(536), 2023–2040 (2021)

15. Meyer, F.G.: The Fréchet mean of inhomogeneous random graphs. In: Benito, R.M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L.M., Sales-Pardo, M. (eds.) COMPLEX NETWORKS 2021. SCI, vol. 1015, pp. 207–219. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-93409-5_18
16. Van Mieghem, P.: Graph Spectra for Complex Networks. Cambridge University Press, Cambridge (2010)
17. Wills, P., Meyer, F.G.: Metrics for graph comparison: a practitioner’s guide. PLOS ONE **15**(2), 1–54 (2020). <https://doi.org/10.1371/journal.pone.0228728>