frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 08 October 2024
DOI 10.3389/fphy.2024.1455988



OPEN ACCESS

EDITED BY
Víctor M. Eguíluz,
Spanish National Research Council (CSIC), Spain

REVIEWED BY
Renaud Lambiotte,
University of Oxford, United Kingdom
Mingao Yuan,
North Dakota State University, United States

*CORRESPONDENCE
François G. Meyer,
fmeyer@colorado.edu

RECEIVED 27 June 2024
ACCEPTED 21 August 2024
PUBLISHED 08 October 2024

CITATION
Meyer FG (2024) When does the mean network capture the topology of a sample of networks? *Front. Phys.* 12:1455988.
doi: 10.3389/fphy.2024.1455988



# When does the mean network capture the topology of a sample of networks?

François G. Meyer*

Applied Mathematics, University of Colorado at Boulder, Boulder, CO, United States

The notion of Fréchet mean (also known as "barycenter") network is the workhorse of most machine learning algorithms that require the estimation of a "location" parameter to analyse network-valued data. In this context, it is critical that the network barycenter inherits the topological structure of the networks in the training dataset. The metric−which measures the proximity between networks−controls the structural properties of the barycenter. This work is significant because it provides for the first time analytical estimates of the sample Fréchet mean for the stochastic blockmodel, which is at the cutting edge of rigorous probabilistic analysis of random networks. We show that the mean network computed with the Hamming distance is unable to capture the topology of the networks in the training sample, whereas the mean network computed using the effective resistance distance recovers the correct partitions and associated edge density. From a practical standpoint, our work informs the choice of metrics in the context where the sample Fréchet mean network is used to characterize the topology of networks for network-valued machine learning.

KEYWORDS

**network-valued data, network barycenter, network topology, statistical network analysis, Fréchet mean, network distance**

## 1 Introduction

There has been recently a flurry of activity around the design of machine learning algorithms that can analyze "network-valued random variables" (e.g. [1–8], and references therein). A prominent question that is central to many such algorithms is the estimation of the mean of a set of networks. To characterize the mean network we borrow the notion of barycenter from physics, and define the Fréchet mean as the network that minimizes the sum of the squared distances to all the networks in the ensemble. This notion of centrality is well adapted to metric spaces (e.g., [4, 9, 10]), and the Fréchet mean network has become a standard tool for the statistical analysis of network-valued data.

In practice, given a training set of networks, it is important that the topology of the sample Fréchet mean captures the mean topology of the training set. To provide a theoretical answer to this question, we estimate the mean network when the networks are sampled from a stochastic block model. The stochastic block models [11, 12] have great practical importance since they provide tractable models that capture the topology of real networks that exhibit community structure. In addition, the theoretical properties (e.g., degree distribution, eigenvalues distributions, etc.) of this ensemble are well understood. Finally, stochastic block models provide universal approximants to networks and can be used as building blocks to analyse more complex networks [13–15].

In this work, we derive the expression of the sample Fréchet mean of a stochastic block model for two very different distances: the Hamming distance [16] and the effective resistance perturbation distance [17]. The Hamming distance, which counts the number of edges that need to be added or subtracted to align two networks defined on the same vertex set, is very sensitive to fine scale fluctuations of the network connectivity. To detect larger scale changes in connectivity, we use the resistance perturbation distance [17]. This network distance can be tuned to quantify configurational changes that occur on a network at different scales: from the local scale formed by the neighbors of each vertex, to the largest scale that quantifies the connections between clusters, or communities [17]. See ([18–20], and references therein) for recent surveys on network distances.

Our analysis shows that the sample Fréchet mean network computed with the Hamming distance is unable to capture the topology of networks in the sample. In the case of a sparse stochastic block model, the Fréchet mean network is always the empty network. Conversely, the Fréchet mean computed using the effective resistance distance recovers the underlying network topology associated with the generating process: the Fréchet mean discovers the correct partitions and associated edge densities.

## 1.1 Relation to existing work

To the best of our knowledge, we are not aware of any theoretical derivation of the sample Fréchet mean for any of the classic ensemble of random networks. Nevertheless, our work share some strong connections with related research questions.

### 1.1.1 The Fréchet mean network as a location parameter

Several authors have proposed simple models of probability measures defined on spaces of networks, which are parameterized by a location and a scale parameter [5, 21]. These probability measures can be used to assign a likelihood to an observed network by measuring the distance of that network to a central network, which characterizes the location of the distribution. The authors in [5] explore two choices for the distance: the Hamming distance, and a diffusion distance. Our choice of distances is similar to that of [5].

### 1.1.2 Existing metrics for the Fréchet mean network

The concept of Fréchet mean necessitates a choice of metric (or distance) on the probability space of networks. The metric will influence the characteristics that the mean will inherit from the network ensemble. For instance, if the distance is primarily sensitive to large scale features (e.g., community structure or the existence of highly connected "hubs"), then the mean will capture these large scale features, but may not faithfully reproduce the fine scale connectivity (e.g., the degree of a vertex, or the presence of triangles).

One sometimes needs to compare networks of different sizes; the edit distance, which allows for creation and removal of vertices, provides an elegant solution to this problem. When the networks are defined on the same vertex set, the edit distance becomes the Hamming distance [22], which can also be interpreted as the entrywise $\ell_1$ norm between the two adjacency matrices. Replacing the $\ell_1$ norm with the $\ell_2$ norm yields the Frobenius norm, which has also been used to compare networks (modulo an unknown permutation of the vertices–or equivalently by comparing the respective classes in the quotient set induced by the action of the group of permutations [4, 10]). We note that the computation of the sample Fréchet mean network using the Hamming distance is NP-hard (e.g., [23]). For this reason, several alternatives have been proposed (e.g., [3]). Both the Hamming distance and Frobenius norm are very sensitive to the fine scale edge connectivity. To probe a larger range of scales, one can compute the mean network using the eigenvalues and eigenvectors of the respective network adjacency matrices [14, 24, 25].

## 1.2 Content of the paper: our main contributions

Our contributions consists of two results.

### 1.2.1 The network distance is the Hamming distance

We prove that when the probability space is equipped with the Hamming distance, then the sample Fréchet mean network converges in probability to the sample median network (computed using the majority rule), in the limit of large sample size. This result has significant practical consequences. Consider the case where one needs to estimate a "central network" that captures the connectivity structure of a training set of sparse networks. Our work implies that if one uses the Hamming distance, then the sample Fréchet mean will be the empty network.

### 1.2.2 The network distance is the resistance perturbation distance

We prove that when the probability space is equipped with the resistance perturbation distance, then the adjacency matrix of the sample Fréchet mean converges to the sample mean adjacency matrix with high probability, in the limit of large network size. Our theoretical analysis is based on the stochastic block model [12], a model of random networks that exhibit community structure. In practical applications, our work suggests that one should use the effective resistance distance to learn the mean topology of a sample of networks.

## 1.3 Outline of the paper

In Section 2, we describe the stochastic block model, the Hamming and resistance distances that are defined on this probability space. The reader who is already familiar with the network models and distances can skip to Section 3 wherein we detail the main results, along with the proofs of the key results. In Section 4, we discuss the implications of our work. The proofs of some technical lemmata are left aside in Section 5.

# 2 Network ensemble and distances

## 2.1 The network ensemble

Let $\mathcal{G}$ be the set of all simple labeled networks with vertex set $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$, and let $\mathcal{S}$ be the set of $n \times n$ adjacency matrices of networks in $\mathcal{G}$,

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

Because there is a unique correspondence between a network $G = (V, E)$ and its adjacency matrix $\boldsymbol{A}$, we sometimes (by an abuse of the language) refer to an adjacency matrix $\boldsymbol{A}$ as a network. Also, without loss of generality we assume throughout the paper that the network size $n$ is even.

We define the matrix $\boldsymbol{P}$ that encodes the edge density within each community and across communities. $\boldsymbol{P}$ can be written as the Kronecker product of the following two matrices,

$$\boldsymbol{P} = \begin{bmatrix} p & q \\ q & p \end{bmatrix} \otimes \boldsymbol{J}_{n/2} \quad (2)$$

where $\boldsymbol{J}_{n/2}$ is the $n/2 \times n/2$ matrix with all entries equal to 1. We denote by $\mathcal{G}(n,p,q)$, the probability space $\mathcal{S}$ equipped with the probability measure,

$$\forall \boldsymbol{A} \in \mathcal{S},\ \mathbb{P}(\boldsymbol{A}) = \prod_{\substack{1\le i\le n/2 \\ 1\le j\le n/2}} p^{a_{ij}} [1-p]^{1-a_{ij}} \prod_{\substack{1\le i\le n/2 \\ n/2+1\le j\le n}} q^{a_{ij}} [1-q]^{1-a_{ij}}. \quad (3)$$

$\mathcal{G}(n,p,q)$ is referred to as a two-community stochastic blockmodel [12]. One can interpret the stochastic blockmodel as follows: the nodes of a network $G \in \mathcal{G}(n,p,q)$ are partitioned into two communities. The first $n/2$ nodes constitute community $C_1$; the second community, $C_2$, comprises the remaining $n/2$ nodes. Edges in the graph are drawn from independent Bernoulli random variables with the following probability of success: $p$ for edges within each community, and $q$ for the across-community edges.

### 2.2 The Hamming distance between networks

Let $\boldsymbol{A}$ and $\boldsymbol{A}'$ be the adjacency matrices of two unweighted networks defined on the same vertex set. The Hamming distance [16] is defined as follows.

**Definition 1.** The Hamming distance between $\boldsymbol{A}$ and $\boldsymbol{A}'$ is defined as

$$d_H(\boldsymbol{A}, \boldsymbol{A}') = \frac{1}{2}\|\boldsymbol{A} - \boldsymbol{A}'\|_1, \quad (4)$$

where the elementwise $\ell_1$ norm of a matrix $\boldsymbol{A}$ is given by $\|\boldsymbol{A}\|_1 = \sum_{1\le i,j\le n} |a_{ij}|$.

Because the distance $d_H$ is not concerned about the locations of the edges that are different between the two graphs, $d_H(\boldsymbol{A}, \boldsymbol{A}')$ is oblivious to topological differences between $\boldsymbol{A}$ and $\boldsymbol{A}'$. For instance, if $\boldsymbol{A}$ and $\boldsymbol{A}'$ are sampled from $\mathcal{G}(n,p,q)$, then the complete removal of the across-community edges induces the same distance as the removal, or addition, of that same number of edges in either community. In other words, a catastrophic change in the network topology cannot be distinguished from benign fluctuations in the local connectivity within either community. To address the limitation of the Hamming distance we define the resistance distance [17].

### 2.3 The resistance (perturbation) distance between networks

For the sake of completeness, we review the concept of effective resistance (e.g., [26, 27]). Let $\boldsymbol{A}$ denote the adjacency matrix of a network $G = (V, E)$, and let $\boldsymbol{D}$ denote the diagonal degree matrix, $d_{ii} = \sum_{j=1}^{n} a_{ij}$. We consider the combinatorial Laplacian matrix [28] defined by

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}. \quad (5)$$

We denote by $\boldsymbol{L}^\dagger$ the Moore-Penrose pseudoinverse of $\boldsymbol{L}$. Let $i, j$ be two nodes of the network, the effective resistance between $i$ and $j$ is given by

$$R_{ij} = L^\dagger_{ii} + L^\dagger_{jj} - 2L^\dagger_{ij}. \quad (6)$$

Intuitively, $R_{ij}$ depends on the abundance of paths between $i$ and $j$. We have the following lower bound that quantifies the burgeoning of connections around the nodes $i$ and $j$,

$$\frac{1}{d_i} + \frac{1}{d_j} \le R_{ij}, \quad (7)$$

where $d_i$ and $d_j$ are the degrees of nodes $i$ and $j$ respectively. As shown in [29], this lower bound is attained for a large class of graphs (see also Remark 3).

The resistance-perturbation distance (or resistance distance for short) is based on comparing the effective resistances matrices $\boldsymbol{R}$ and $\boldsymbol{R}'$ of $G$ and $G'$ respectively. To simplify the discussion, we only consider networks that are connected with high probability. All the results can be extended to disconnected networks as explained in [17].

**Definition 2.** Let $G = (V, E)$ and $G' = (V, E')$ be two networks defined on the same vertex set $[n]$. Let $R$ and $R'$ denote the effective resistances of $G$ and $G'$ respectively. We define the resistance-perturbation distance [17] to be

$$d_{\text{rp}}(G, G') = \sum_{1\le i<j\le n} \left| R_{ij} - R'_{ij} \right|^2. \quad (8)$$

## 3 Main results

We first review the concept of sample Fréchet mean, and then present the main results. We consider the probability space $(\mathcal{S}, \mathbb{P})$ formed by the adjacency matrices of networks sampled from $\mathcal{G}(n,p,q)$. We equip $\mathcal{S}$ with a distance $d$, which is either the Hamming distance or the resistance distance. Let $\boldsymbol{A}^{(k)}, 1 \le k \le N$ be adjacency matrices sampled independently from $\mathcal{G}(n,p,q)$.

### 3.1 The sample Fréchet mean

The sample Fréchet function evaluated at $\boldsymbol{B} \in \mathcal{S}$ is defined by

$$\hat{F}_2(\boldsymbol{B}) = \frac{1}{N} \sum_{k=1}^{N} d^2\left(\boldsymbol{B}, \boldsymbol{A}^{(k)}\right). \quad (9)$$

The minimization of the Fréchet function $\hat{F}_2(\boldsymbol{B})$ gives rise to the concept of sample Fréchet mean [30], or network barycenter [31].

**Definition 3.** The sample Fréchet mean network is the set of adjacency matrices $\hat{\boldsymbol{\mu}}[\mathbb{P}]$ solutions to

$$\hat{\boldsymbol{\mu}}[\mathbb{P}] = \underset{\boldsymbol{B}\in\mathcal{S}}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^{N} d^2\left(\boldsymbol{B},\boldsymbol{A}^{(k)}\right). \quad (10)$$

Solutions to the minimization problem in Equation 10 always exist, but need not be unique. In Theorem 1 and Theorem 2, we prove that the sample Fréchet mean network of $\mathcal{G}(n,p,q)$ is unique, when $d$ is either the Hamming distance or the resistance distance.

A word on notations is in order here. It is customary to denote by $\boldsymbol{\mu}[\mathbb{P}]$ the *population* Fréchet network of the probability distribution $\mathbb{P}$, (e.g., [31]), since the adjacency matrix $\boldsymbol{\mu}[\mathbb{P}]$ characterizes the location of the probability distribution $\mathbb{P}$. Because we use hats to denote sample (empirical) estimates, we denote by $\hat{\boldsymbol{\mu}}[\mathbb{P}]$ the adjacency matrix of the *sample* Fréchet mean network.

## 3.2 The sample Fréchet mean of $\mathcal{G}(n,p,q)$ computed with the Hamming distance

The following theorem shows that the sample Fréchet mean network converges in probability to the sample Fréchet median network, computed using the majority rule, in the limit of large sample size, $N$.

**Theorem 1.** Let $\hat{\boldsymbol{\mu}}[\mathbb{P}]$ be the sample Fréchet mean network computed using the Hamming distance. Then,

$$\forall\varepsilon>0, \exists N_0, \forall N\geq N_0, \mathbb{P}\left(d_H\left(\hat{\boldsymbol{\mu}}[\mathbb{P}],\hat{\boldsymbol{m}}[\mathbb{P}]\right)<\varepsilon\right)\geq 1-\varepsilon. \quad (11)$$

where $\hat{\boldsymbol{m}}[\mathbb{P}]$ is the adjacency matrix computed using the majority rule,

$$\forall i,j\in[n], \quad \hat{\boldsymbol{m}}[\mathbb{P}]_{ij} = \begin{cases} 1 & if \ \sum_{k=1}^{N} a_{ij}^{(k)} \geq N/2, \\ 0 & otherwise. \end{cases} \quad (12)$$

**Remark 1.** The matrix $\hat{\boldsymbol{m}}[\mathbb{P}]$ is the sample Fréchet median network (e.g., [32], solution to the following minimization problem [21],

$$\hat{\boldsymbol{m}}[\mathbb{P}] = \underset{\boldsymbol{B}\in\mathcal{S}}{\operatorname{argmin}} \hat{F}_1(\boldsymbol{B}), \quad (13)$$

where $\hat{F}_1$ is the Fréchet function associated to the sample Fréchet median, defined by

$$\hat{F}_1(\boldsymbol{B}) = \frac{1}{N}\sum_{k=1}^{N} d_H\left(\boldsymbol{A}^{(k)},\boldsymbol{B}\right). \quad (14)$$

**Remark 2.** The network size $n$ in Theorem 1 is assumed to be constant; the convergence in probability in Theorem 1 happens when the sample size $N\to\infty$. The proof of theorem 1 involves constants that are sublogarithmic functions of $n$ (see $\alpha$ and $\beta$ in the proof of lemma 3 in Section 5.2.)

One could envision a scenario where the network size $n$ would grow with the sample size $N$. In that case, we need $N=\omega(\log n)$ to ensure that lemma 3 provides a useful bound. This is a very weak upper bound on $n$, satisfied for instance for $n=\exp(N^c)$, with $0<c<1$. Finally, theorem 1 holds for any values of the edge densities $p$ and $q$ (whether these depend on $n$ or $N$), as long as they are always distinct from 1/2 (to avoid the instability that occurs when estimating $\hat{\boldsymbol{\mu}}[\mathbb{P}]$; see lemma 4 for details).

Before deriving the proof of theorem 1, we present an extension of the Hamming distance to weighted networks. We remember that the sample Fréchet mean network computed using the Hamming distance has to be an unweighted network, since the Hamming distance is only defined for unweighted networks. This theoretical observation notwithstanding, the proof of theorem 1 becomes much simpler if we introduce an extension of the Hamming distance to weighted networks; in truth, we extend a slightly different formulation of the Hamming distance.

Let $\boldsymbol{A},\boldsymbol{B}\in\mathcal{S}$ be two unweighted adjacency matrices. Because $d_H(\boldsymbol{A},\boldsymbol{B})$ counts the number of (unweighted) edges that are different between the graphs, we have

$$d_H(\boldsymbol{A},\boldsymbol{B}) = \sum_{1\leq i<j\leq n} a_{ij} + \sum_{1\leq i<j\leq n} b_{ij} - 2\sum_{1\leq i<j\leq n} a_{ij}b_{ij}. \quad (15)$$

Now, assume that $\boldsymbol{A}$ and $\boldsymbol{B}$ are two weighted adjacency matrices, with $a_{ij},b_{ij}\in[0,1]$. A natural extension of Equation 15 to matrices with entries in [0,1] is therefore given by

$$\delta(\boldsymbol{A},\boldsymbol{B}) = \sum_{1\leq i<j\leq n} a_{ij} + \sum_{1\leq i<j\leq n} b_{ij} - 2\sum_{1\leq i<j\leq n} a_{ij}b_{ij}. \quad (16)$$

The function $\delta$, defined on the space of weighted adjacency matrices with weights in [0,1], satisfies all the properties of a distance, except for the triangle inequality.

We now present the sample probability matrix $\hat{\boldsymbol{P}}$ and the sample correlation $\hat{\boldsymbol{\rho}}$. Let $\boldsymbol{A}^{(k)}, 1\leq k\leq N$ be adjacency matrices sampled independently from $\mathcal{G}(n,p,q)$. We define

$$\hat{\boldsymbol{P}}_{ij} \stackrel{\text{def}}{=} \hat{\mathbb{E}}\left[a_{ij}\right] \stackrel{\text{def}}{=} \frac{1}{N}\sum_{k=1}^{N} a_{ij}^{(k)}. \quad (17)$$

and

$$\hat{\boldsymbol{\rho}}_{ij,i'j'} \stackrel{\text{def}}{=} \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right] \stackrel{\text{def}}{=} \frac{1}{N}\sum_{k=1}^{N} a_{ij}^{(k)}a_{i'j'}^{(k)}. \quad (18)$$

We can combine the definitions of $\delta$ and $\hat{\boldsymbol{P}}$ to derive the following expression for the Fréchet function $\hat{F}_1$ for the sample median, defined by Equation 14,

$$\hat{F}_1(\boldsymbol{B}) = \delta\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right). \quad (19)$$

The proof of this simple identity is very similar to the proof of lemma 1, and is omitted for brevity. We are now ready to present the proof of theorem 1.

Proof of Theorem 1. The proof relies on the observation (formalized in lemma 1) that the Fréchet function $\hat{F}_2(\boldsymbol{B})$ can be expressed as the sum of a dominant term and a residual. The residual becomes increasingly small in the limit of large sample size (see lemma 3) and can be neglected. We show in lemma 2 that the dominant term is minimum for the sample Fréchet median network $\hat{\boldsymbol{m}}[\mathbb{P}]$ [defined by Equation 12]. We start with the decomposition of $\hat{F}_2(\boldsymbol{B})$ in terms of a dominant term and a residual.

**Lemma 1.** Let $\boldsymbol{B} \in \mathcal{S}$. We denote by $\mathcal{E}(\boldsymbol{B})$ the set of edges of the network with adjacency matrix $\boldsymbol{B}$, we denote by $\bar{\mathcal{E}}(\boldsymbol{B})$ the set of "nonedges." Then

$$\hat{F}_2(\boldsymbol{B}) = \delta^2\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right) - \sum_{\substack{1\le i<j\le n\\ 1\le i'<j'\le n}} \left(\hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'} - \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right]\right) + 4\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \left(\hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'} - \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right]\right). \tag{20}$$

where $\hat{\boldsymbol{P}}$ is defined by Equation 17, and $\hat{\rho}$ is defined by Equation 18.

Proof. The proof of lemma 1 is provided in Section 5.

To call attention to the distinct roles played by the terms in Equation 20, we define the dominant term of $\hat{F}_2(\boldsymbol{B})$,

$$\hat{F}(\boldsymbol{B}) \overset{def}{=} \delta^2\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right) - \sum_{\substack{1\le i<j\le n\\ 1\le i'<j'\le n}} \left[\hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'} - \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right]\right], \tag{21}$$

*and the residual* $\zeta_N$ *is defined by*

$$\zeta_N(\boldsymbol{B}) = 4\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \left(\hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'} - \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right]\right), \tag{22}$$

*so that* $\hat{F}_2(\boldsymbol{B}) = \hat{F}(\boldsymbol{B}) + \zeta_N(\boldsymbol{B})$.

The next step of the proof of theorem 1 involves showing that the sample median network, $\hat{\boldsymbol{m}}[\mathbb{P}]$, [see Equation 12], which is the minimizer of $\hat{F}_1(\boldsymbol{B})$ [see Equation 14], is also the minimizer of $\hat{F}(\boldsymbol{B})$.

**Lemma 2.** $\hat{\boldsymbol{m}}[\mathbb{P}]$ satisfies: $\forall \boldsymbol{B} \in \mathcal{S}, \hat{F}(\hat{\boldsymbol{m}}[\mathbb{P}]) \le \hat{F}(\boldsymbol{B})$.

Proof of lemma 2. We have

$$\hat{F}(\boldsymbol{B}) = \delta^2\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right) - \sum_{\substack{1\le i<j\le n\\ 1\le i'<j'\le n}} \left(\hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'} - \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right]\right) \tag{23}$$

Because $\hat{\boldsymbol{m}}[\mathbb{P}]$ is the minimizer of $\hat{F}_1(\boldsymbol{B}) = \delta(\boldsymbol{B},\hat{\boldsymbol{P}})$ [see Equation 19], $\hat{\boldsymbol{m}}[\mathbb{P}]$ is also the minimizer of $\delta^2(\boldsymbol{B},\hat{\boldsymbol{P}})$. Finally, since $\sum_{\substack{1\le i<j\le n\\ 1\le i'<j'\le n}} (\hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'} - \hat{\mathbb{E}}[\rho_{ij,i'j'}])$ does not depend on $\boldsymbol{B}$, $\hat{\boldsymbol{m}}[\mathbb{P}]$ is the minimizer of $\hat{F}(\boldsymbol{B})$.

*We now turn our attention to the residual and we confirm in the next lemma that* $\zeta_N(\boldsymbol{B}) = \mathcal{O}_P(\frac{1}{\sqrt{N}})$; *to wit* $\zeta_N(\boldsymbol{B})\sqrt{N}$ *is bounded with high probability.*

**Lemma 3.** $\forall \varepsilon > 0, \exists c > 0, \forall\ N \ge 1,$

$$\mathbb{P}\left(\boldsymbol{A}^{(k)} \sim \mathcal{G}(n,p,q); |\zeta_N(\boldsymbol{B})| < \frac{c}{\sqrt{N}}\right) > 1 - \varepsilon. \tag{24}$$

Proof. The proof of lemma 3 is provided in Section 5.2.

The last technical lemma that is needed to complete the proof of theorem 1 is a variance inequality [31] for $\hat{F}$. We assume that the entries of $\boldsymbol{P}$ are uniformly away from 1/2 (this technical condition on $\boldsymbol{P}$ prevents the instability that occurs when estimating $\hat{\boldsymbol{\mu}}[\mathbb{P}]$ for $p_{ij} = 1/2$).

**Lemma 4.** We assume that there exists $\eta > 0$ such that $1 \le i < j \le n, |p_{ij} - 1/2| > \eta$. Then, $\exists \alpha > 0$

$$\forall \boldsymbol{B} \in \mathcal{S}, \quad \alpha\|\boldsymbol{B} - \hat{\boldsymbol{m}}[\mathbb{P}]\|_1^2 \le \left|\hat{F}(\boldsymbol{B}) - \hat{F}(\hat{\boldsymbol{m}}[\mathbb{P}])\right|, \tag{25}$$

with high probability.

Proof. The proof of lemma 4 is provided in Section 5.3.

We are now in position to combine the lemmata and complete the proof of theorem 1.

Let $\hat{\boldsymbol{\mu}}[\mathbb{P}]$ be the sample Fréchet mean network, and let $\hat{\boldsymbol{m}}[\mathbb{P}]$ be the sample Fréchet median network. By definition, $\hat{\boldsymbol{\mu}}[\mathbb{P}]$ is the minimizer of $\hat{F}_2$, and thus

$$\hat{F}(\hat{\boldsymbol{\mu}}[\mathbb{P}]) = \hat{F}_2(\hat{\boldsymbol{\mu}}[\mathbb{P}]) - \zeta_N(\hat{\boldsymbol{\mu}}[\mathbb{P}]) \le \hat{F}_2(\hat{\boldsymbol{m}}[\mathbb{P}]) - \zeta_N(\hat{\boldsymbol{\mu}}[\mathbb{P}]) \tag{26}$$

Now, by definition of $\hat{F}$ in Equation 21, we have

$$\hat{F}_2(\hat{\boldsymbol{m}}[\mathbb{P}]) - \zeta_N(\hat{\boldsymbol{\mu}}[\mathbb{P}]) = \hat{F}(\hat{\boldsymbol{m}}[\mathbb{P}]) + \zeta_N(\hat{\boldsymbol{m}}[\mathbb{P}]) - \zeta_N(\hat{\boldsymbol{\mu}}[\mathbb{P}]), \tag{27}$$

and therefore,

$$0 \le \hat{F}(\hat{\boldsymbol{\mu}}[\mathbb{P}]) - \hat{F}(\hat{\boldsymbol{m}}[\mathbb{P}]) \le \zeta_N(\hat{\boldsymbol{m}}[\mathbb{P}]) - \zeta_N(\hat{\boldsymbol{\mu}}[\mathbb{P}]). \tag{28}$$

This last inequality, combined with Equation 24 proves that $\hat{F}(\hat{\boldsymbol{\mu}}[\mathbb{P}]) - \hat{F}(\hat{\boldsymbol{m}}[\mathbb{P}])$ converges to zero for large $N$. We can say more; using the variance inequality Equation 25, we prove that $d_H(\hat{\boldsymbol{\mu}}[\mathbb{P}], \hat{\boldsymbol{m}}[\mathbb{P}]) = \|\hat{\boldsymbol{\mu}}[\mathbb{P}] - \hat{\boldsymbol{m}}[\mathbb{P}]\|_1$ converges in probability to zero for large $N$.

Let $\varepsilon > 0$, from lemma 4, there exists $\alpha > 0$ such that

$$\mathbb{P}\left(\boldsymbol{A}^{(k)} \sim \mathcal{G}(n,p,q); \quad \alpha\|\hat{\boldsymbol{\mu}}[\mathbb{P}] - \hat{\boldsymbol{m}}[\mathbb{P}]\|_1{}^2 \le \left|\hat{F}(\hat{\boldsymbol{\mu}}[\mathbb{P}]) - \hat{F}(\hat{\boldsymbol{m}}[\mathbb{P}])\right|\right) > 1 - \varepsilon. \tag{29}$$

The term $\zeta_N(\hat{\boldsymbol{m}}[\mathbb{P}]) - \zeta_N(\hat{\boldsymbol{\mu}}[\mathbb{P}])$ is controlled using Lemma 3,

$$\exists C, \forall N \ge 1, \mathbb{P}\left(\forall \mathbb{P} \in \mathcal{S}, \left|\zeta_N(\hat{\boldsymbol{m}}[\mathbb{P}]) - \zeta_N(\hat{\boldsymbol{\mu}}[\mathbb{P}])\right| < \frac{C}{\sqrt{N}}\right) \ge 1 - \varepsilon \tag{30}$$

Combining Equations 28–30 we get

$$\forall N \ge 1, \quad \mathbb{P}\left(\|\hat{\boldsymbol{\mu}}[\mathbb{P}] - \hat{\boldsymbol{m}}[\mathbb{P}]\|_1{}^2 < \frac{C}{\alpha\sqrt{N}}\right) > 1 - \varepsilon. \tag{31}$$

We conclude that $\exists N_1$ such that

$$\forall N \ge N_1, \quad \mathbb{P}\left(\|\hat{\boldsymbol{\mu}}[\mathbb{P}] - \hat{\boldsymbol{m}}[\mathbb{P}]\|_1 < \varepsilon\right) > 1 - \varepsilon, \tag{32}$$

which completes the proof of the theorem.

## 3.3 The sample Fréchet mean of $\mathcal{G}(n,p,q)$ computed with the resistance distance

Here we equip the probability space $(\mathcal{S},\mathbb{P})$ with the resistance metric defined by Equation 8. Let $\boldsymbol{A}^{(k)}, 1 \le k \le N$ be adjacency matrices sampled independently from $\mathcal{G}(n,p,q)$, and let $\boldsymbol{R}^{(k)}$ be their effective resistances. Because the resistance metric relies on the comparison of connectivity at multiple scales, we expect that the sample Fréchet mean network recovers the topology induced by the communities.

In the following, we need to ensure that the effective resistances are always well defined for networks sampled from $\mathcal{G}(n,p,q)$, and we therefore require a very mild condition of the edge density. We assume that $p = \omega(\log n/n)$ and $q = \omega(\log n/n)$. For instance, this condition is satisfied if $p = a_1(\log^{c_1} n)/n$, and $q = a_2(\log^{c_2} n)/n$, with $a_1, a_2 > 0$, $c_1, c_2 > 1$.

The next theorem proves that the sample Fréchet mean converges toward the expected adjacency matrix $\boldsymbol{P}$ (see Section 2) in the limit of large networks.

**Theorem 2.** Let $\hat{\boldsymbol{\mu}}[\mathbb{P}]$ be the sample Fréchet mean computed using the effective resistance distance. Then

$$\hat{\boldsymbol{\mu}}[\mathbb{P}] = \mathbb{E}[\boldsymbol{A}] = \boldsymbol{P}, \tag{33}$$

in the limit of large network size $n$, with high probability.

Proof of theorem 2. The proof combines three elements. We first observe that the effective resistance of the sample Fréchet mean is the sample mean effective resistance.

**Lemma 5.** Let $\hat{\boldsymbol{\mu}}[\mathbb{P}]$ be the sample Fréchet mean computed using the resistance distance. Then

$$\hat{R}_{ij} \stackrel{def}{=} R_{ij}\left(\hat{\boldsymbol{\mu}}[\mathbb{P}]\right) = \frac{1}{N}\sum_{k=1}^{N} R_{ij}^{(k)} \tag{34}$$

Proof of lemma 5. The proof relies on the observation that the Fréchet function in Equation 10, is a quadratic function of $\hat{R}_{ij} = R_{ij}(\hat{\boldsymbol{\mu}}[\mathbb{P}])$. Indeed, we have

$$\frac{1}{N}\sum_{k=1}^{N}\sum_{1\le i<j\le n}\left|\hat{R}_{ij} - R_{ij}^{(k)}\right|^2 = \sum_{1\le i<j\le n}\frac{1}{N}\sum_{k=1}^{N}\left|\hat{R}_{ij} - R_{ij}^{(k)}\right|^2 \tag{35}$$

where we have used the definition of the effective resistance distance given by Equation 8. The minimum of Equation 35 is given by Equation 34.

The second element of the proof of theorem 2 is a concentration result for the effective resistance $R_{ij}$ for networks in $\mathcal{G}(n,p,q)$, when the network size $n$ becomes large. Our technique of proof is different from that of Theorem 1. In Theorem 1, we rely on laws of large numbers (for large sample size $N$) to compute the minimum of the Fréchet function $\hat{F}_2$.

In contrast, the proof of theorem 2 follows a different line of attack, where we replace the law of large number with a concentration result for the effective resistance $R_{ij}$ of $\mathcal{G}(n,p,q)$ for large network size $n$. Our estimates are independent of the sample size $N$; they only become sharper as the graph size $n \to \infty$. Others have derived similar results (e.g., [29, 33–36]).

In the next lemma, we prove that $(1/N)\sum_{k=1}^{N} R_{ij}^{(k)}$ concentrates around $R^{\star}{}_{ij}$ in the limit of large network size $n$.

**Lemma 6.** Let $G = (V,E)$ a graph sampled from $\mathcal{G}(n,p,q)$. Let $i, j$ be two nodes in $V$. Then the effective resistance $R_{ij}$ between $i$ and $j$ is given by

$$R_{ij} = R^{\star}{}_{ij} + o_P\left(\frac{1}{n}\right), \tag{36}$$

where

$$R^{\star}{}_{ij} = \begin{cases} \dfrac{4}{n(p+q)} & \text{if } i \text{ and } j \text{ are in the same community,}\\[2ex] \dfrac{4}{n(p+q)} + \dfrac{(p-q)}{(p+q)}\dfrac{4}{n^2 q} & \text{if } i \text{ and } j \text{ are in different communities.}\end{cases} \tag{37}$$

Before deriving the proof of lemma 6 we make a few remarks to help guide the reader's intuition.

**Remark 3.** We justify Equation 37 with a simple circuit argument. We first analyse the case where $i$ and $j$ belong to the same community, say $C_1$. In this case, we can neglect the other community $C_2$ because of the bottleneck created by the across-community edges. Consequently, $C_1$ is approximately an Erdős-Rényi network wherein the effective resistance $R_{ij}$ concentrates around $4/(n(p+q))$ [29], and we obtain the first term in Equation 37.

On the other hand, when the vertices $i$ and $j$ are in distinct communities, then a simple circuit argument shows that

$$R_{ij} \approx \frac{2}{n(p+q)} + \frac{1}{k} + \frac{2}{n(p+q)}, \tag{38}$$

where $k$ is the number of across-community edges, creating a bottleneck with effective resistance $1/k$ between the two communities [37]; each term $2/(n(p+q))$ accounts for the effective resistance from node $i$ (respectively $j$) to a node incident to an across-community edge. Because the number of across-community edges, $k$, is a binomial random variable, it concentrates around its mean, $qn^2/4$. Finally, $1/k$ is a binomial reciprocal whose mean is given by $4/(qn^2) + ø1/n^3$ [38], and we recover the second term of Equation 37.

Our proof of lemma 6, requires that we introduce another operator on the graph, the normalized Laplacian matrix (e.g., [28]). Let $\boldsymbol{A}$ be the adjacency matrix of a network $(V,E)$, and let $\boldsymbol{D}$ be the diagonal matrix of degrees, $d_i = \sum_{j=1}^{n} a_{ij}$. We normalize $\boldsymbol{A}$ in a symmetric manner, and we define

$$\hat{\boldsymbol{A}} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}, \tag{39}$$

where $\boldsymbol{D}^{-1/2}$ is the diagonal matrix with entries $1/\sqrt{d_i}$. The normalized Laplacian matrix is defined by

$$\mathcal{L} = \boldsymbol{I} - \hat{\boldsymbol{A}}, \tag{40}$$

where $\boldsymbol{I}$ is the identity matrix. $\mathcal{L}$ is positive semi-definite [28], and we will consider its Moore-Penrose pseudoinverse, $\mathcal{L}^{\dagger}$.

Proof of lemma 6. The lemma relies on the characterization of $\boldsymbol{R}$ in terms of $\mathcal{L}^{\dagger}$ [28],

$$R_{ij} = \langle \boldsymbol{u}_i - \boldsymbol{u}_j, \mathcal{L}^{\dagger}\left(\boldsymbol{u}_i - \boldsymbol{u}_j\right)\rangle, \tag{41}$$

where $\boldsymbol{u}_i = (1/\sqrt{d_i})\boldsymbol{e}_i$, and $\boldsymbol{e}_i$ is the $i^{\text{th}}$ vector of the canonical basis. Let $1 = \lambda_1 \ge \lambda_2 \ge \ldots \lambda_n \ge -1$ be the eigenvalues of $\hat{A}$, and let $\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_n$ be the corresponding orthogonal projectors,

$$\hat{\boldsymbol{A}} = \sum_{m=1}^{n} \lambda_m \boldsymbol{\Pi}_m, \tag{42}$$

where $\boldsymbol{\Pi}_1 = \tau^{-1}\boldsymbol{d}^{1/2}\boldsymbol{d}^{1/2^T}$, with $\boldsymbol{d}^{1/2} = [\sqrt{d_1} \;\cdots\; \sqrt{d_n}\,]^T$, and $\tau = \sum_{i=1}^{n} d_i$. Because $\boldsymbol{\Pi}_1$ is also the orthogonal projection on the null space of $\mathcal{L}$, we have

$$\begin{aligned}\mathcal{L}^{\dagger} &= (\mathcal{L} + \boldsymbol{\Pi}_1)^{-1} - \boldsymbol{\Pi}_1 = \left(\boldsymbol{I} - \left(\hat{\boldsymbol{A}} - \boldsymbol{\Pi}_1\right)\right)^{-1} - \boldsymbol{\Pi}_1\\ &= \boldsymbol{I} - \boldsymbol{\Pi}_1 + \frac{\lambda_2}{1-\lambda_2}\boldsymbol{\Pi}_2 + \mathbf{Q},\end{aligned} \tag{43}$$

where

$$\mathbf{Q} = \sum_{m=3}^{n} \frac{\lambda_m}{1-\lambda_m}\boldsymbol{\Pi}_m. \tag{44}$$

Substituting Equation 43 into Equation 41, we get

$$R_{ij} = \langle \boldsymbol{u}_i - \boldsymbol{u}_j, (\boldsymbol{I} - \boldsymbol{\Pi}_1)(\boldsymbol{u}_i - \boldsymbol{u}_j)\rangle + \frac{\lambda_2}{1-\lambda_2}\langle \boldsymbol{u}_i - \boldsymbol{u}_j, \boldsymbol{\Pi}_2(\boldsymbol{u}_i - \boldsymbol{u}_j)\rangle + \langle \boldsymbol{u}_i - \boldsymbol{u}_j, \boldsymbol{Q}(\boldsymbol{u}_i - \boldsymbol{u}_j)\rangle \tag{45}$$

The first (and dominant) term of Equation 45 is

$$\langle \boldsymbol{u}_i - \boldsymbol{u}_j, (\boldsymbol{I} - \boldsymbol{\Pi}_1)(\boldsymbol{u}_i - \boldsymbol{u}_j)\rangle = \langle \boldsymbol{u}_i - \boldsymbol{u}_j, \boldsymbol{u}_i - \boldsymbol{u}_j\rangle = \frac{1}{d_i} + \frac{1}{d_j}. \tag{46}$$

Let us examine the second term of Equation 45. Löwe and Terveer [39] provide the following estimate for $\lambda_2$,

$$\lambda_2 = \frac{p-q}{p+q} + \omega(n), \text{where } \omega(n) = \mathcal{O}\left(\sqrt{\frac{2\log n}{n(p+q)}}\right). \tag{47}$$

The corresponding eigenvector $\boldsymbol{z}$ is given, with probability $(1 - \mathcal{O}1)$, by [40],

$$z_i = \sigma_i \frac{1}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right), \tag{48}$$

where the "sign" vector $\boldsymbol{\sigma}$, which encodes the community, is given by

$$\sigma_i = \begin{cases} 1 & \text{if } 1 \le i \le n/2, \\ -1 & \text{if } n/2 + 1 \le i \le n. \end{cases} \tag{49}$$

We derive from Equation 48 the following approximation to $\langle \boldsymbol{u}_i, \boldsymbol{\Pi}_2 \boldsymbol{u}_j\rangle$,

$$\langle \boldsymbol{u}_i, \boldsymbol{\Pi}_2 \boldsymbol{u}_j\rangle = \boldsymbol{u}_i^T \boldsymbol{z}\boldsymbol{z}^T \boldsymbol{u}_j = \frac{1}{\sqrt{d_i d_j}} z_i z_j = \frac{1}{n\sqrt{d_i d_j}}\sigma_i\sigma_j + o_P\left(\frac{1}{n}\right). \tag{50}$$

We therefore have

$$\langle \boldsymbol{u}_i - \boldsymbol{u}_j, \boldsymbol{\Pi}_2(\boldsymbol{u}_i - \boldsymbol{u}_j)\rangle = \frac{1}{nd_i} + \frac{1}{nd_j} - \frac{2}{n\sqrt{d_i d_j}}\sigma_i\sigma_j + o_P\left(\frac{1}{n}\right). \tag{51}$$

The degree $d_i$ of node $i$ is a binomial random variable, which concentrates around its mean, $p(n/2 - 1) + qn/2 \approx n(p+q)/2$ for large network size $n$. Also, $1/d_i$ is a binomial reciprocal that also concentrates around its mean, which is given by $2/((p+q)n) + ø1/n^2$ [38]. We conclude that in the limit of large network size,

$$\langle \boldsymbol{u}_i - \boldsymbol{u}_j, \boldsymbol{\Pi}_2(\boldsymbol{u}_i - \boldsymbol{u}_j)\rangle = \frac{4}{n^2(p+q)}\left(1 - \sigma_i\sigma_j\right) + o_P\left(\frac{1}{n}\right). \tag{52}$$

Combining Equations 47, 52 yields

$$\langle \boldsymbol{u}_i - \boldsymbol{u}_j, \frac{\lambda_2}{1-\lambda_2}\boldsymbol{\Pi}_2(\boldsymbol{u}_i - \boldsymbol{u}_j)\rangle = \frac{(p-q)}{(p+q)}\frac{4}{n^2 q}\frac{\left(1-\sigma_i\sigma_j\right)}{2} + o_P\left(\frac{1}{n}\right). \tag{53}$$

We note that

$$\frac{(p-q)}{(p+q)}\frac{4}{n^2 q}\frac{\left(1-\sigma_i\sigma_j\right)}{2} = \begin{cases} \dfrac{4}{n(p+q)} & \text{if } i \text{ and } j \text{ are in the same community} \\ \dfrac{4}{n(p+q)} + \dfrac{(p-q)}{(p+q)}\dfrac{4}{n^2 q} & \text{if } i \text{ and } j \text{ are in different communities,} \end{cases} \tag{54}$$

which confirms that $\langle \boldsymbol{u}_i - \boldsymbol{u}_j, \frac{\lambda_2}{1-\lambda_2}\boldsymbol{\Pi}_2(\boldsymbol{u}_i - \boldsymbol{u}_j)\rangle$ provides the correction in Equation 37 created by the bottleneck between the communities. Finally, we show in Section 5.4 that the last term in the expansion of $R_{ij}$ Equation 45 can be neglected,

$$\left|\langle \boldsymbol{u}_i - \boldsymbol{u}_j, \boldsymbol{Q}(\boldsymbol{u}_i - \boldsymbol{u}_j)\rangle\right| \le \left(\frac{1}{d_i} + \frac{1}{d_j}\right)\frac{8\sqrt{2}}{(np)^{3/2}} \quad \text{almost surely.} \tag{55}$$

This concludes the proof the lemma.

**Remark 4.** Lemma 6 can be extended to a stochastic block model of any geometry for which we can derive the analytic expression of the dominant eigenvalues; see (see e.g., [39, 41]) for equal size communities, and (see e.g., [42]) for the more general case of inhomogeneous random networks.

We can apply Lemma 6 to derive an approximation to the sample mean effective resistance.

**Corollary 1.** Let $\boldsymbol{A}^{(k)}, 1 \le k \le N$ be adjacency matrices sampled independently from $\mathcal{G}(n,p,q)$, and let $\boldsymbol{R}^{(k)}, 1 \le k \le N$ be the respective effective resistance matrices. Then

$$\frac{1}{N}\sum_{k=1}^{N} R_{ij}^{(k)} = R^{\star}{}_{ij} + o_P\left(\frac{1}{n}\right), \tag{56}$$

where $R^{\star}{}_{ij}$ is given by Equation 37.

Lastly, the final ingredient of the proof of theorem 2 is Lemma 7 that shows that matrix $\boldsymbol{R}^{\star}$, given by Equation 37, is the effective resistance of the expected adjacency matrix of $(\mathcal{S}, \mathbb{P})$, $\boldsymbol{R}^{\star} = \boldsymbol{R}[\mathbb{E}[\boldsymbol{A}]]$.

**Lemma 7.** Let $\boldsymbol{R}$ be the $n \times n$ effective resistance matrix of a network with adjacency matrix $\boldsymbol{A}$. If

$$\boldsymbol{R} = \frac{4}{n(p+q)}\boldsymbol{J} + \frac{p-q}{p+q}\frac{4}{n^2 q}\boldsymbol{K}, \tag{57}$$

where $\boldsymbol{J} = \boldsymbol{J}_n$, and $\boldsymbol{K}$ is the $n \times n$ matrix associated with the cross-community edges,

$$\boldsymbol{K} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes \boldsymbol{J}_{n/2}. \tag{58}$$

Then $\boldsymbol{A} = \boldsymbol{P}$, where $\boldsymbol{P}$ is given by Equation 2.

Proof of lemma 7. The proof is elementary and relies on the following three identities. First, we recover $\boldsymbol{L}^{\dagger}$, the pseudo-inverse of the combinatorial Laplacian $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, from $\boldsymbol{R}$,

$$\boldsymbol{L}^{\dagger} = -\frac{1}{2}\left[\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right]\boldsymbol{R}\left[\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right]. \tag{59}$$

We can then recover $\boldsymbol{L}$ from $\boldsymbol{L}^{\dagger}$; for every $\alpha \ne 0$, we have

$$\boldsymbol{L} = \left[\boldsymbol{L}^{\dagger} + \frac{\alpha}{n}\boldsymbol{J}\right]^{-1} - \frac{\alpha}{n}\boldsymbol{J}. \tag{60}$$

Finally, $\boldsymbol{A} = -\boldsymbol{L} + \operatorname{diag}(\boldsymbol{L})$.

This concludes the proof of theorem 2.

## 4 Discussion of our results

This paper provides analytical estimates of the sample Fréchet mean network when the sample is generated from a stochastic block model. We derived the expression of the sample Fréchet mean when the probability space $\mathcal{G}(n,p,q)$ is equipped with two very different distances: the Hamming distance and the resistance distance. This work answers the question raised by Lunagómez et al. [5] "what is the "mean" network (rather than how do we estimate the success-probabilities of an inhomogeneous random network), and do we want the "mean" itself to be a network?".

We show that the sample mean network is an unweighted network whose topology is usually very different from the average topology of the sample. Specifically, in the regime of networks where $\min p_{ij} < 1/2$ (e.g., networks with ø$n^2$ but $\omega(n)$ edges), then the sample Fréchet mean is the empty network, and is pointless. In contrast, the resistance distance leads to a sample Fréchet mean that recovers the correct topology induced by the community structure; the edge density of the sample Fréchet mean network is the expected edge density of the random network ensemble. The effective resistance distance is thus able to capture the large scale (community structure) and the mesoscale, which spans scales from the global to the local scales (the degree of a vertex).

This work is significant because it provides for the first time analytical estimates of the sample Fréchet mean for the stochastic blockmodel, which is at the cutting edge of rigorous probabilistic analysis of random networks [12]. The technique of proof that is used to compute the sample Fréchet mean for the Hamming distance can be extended to the large class of inhomogeneous random networks [43]. It should also be possible to extend our computation of the Fréchet mean with the resistance distance to stochastic block models with $K$ communities of arbitrary size, and varying edge density.

From a practical standpoint, our work informs the choice of distance in the context where the sample Fréchet mean network has been used to characterize the topology of networks for network-valued machine learning (e.g., detecting change points in sequences of networks [2, 8], computing Fréchet regression [6], or cluster network datasets [7]). Future work includes the analysis of the sample Fréchet mean when the distance is based on the eigenvalues of the normalized Laplacian Wills and Meyer [20].

## 5 Additional proofs

### 5.1 Proof of lemma 1

We start with a simple result that provides an expression for the Hamming distance squared. Let $\boldsymbol{A},\boldsymbol{B}\in\mathcal{S}$, and let $\mathcal{E}(\boldsymbol{B})$ denote the set of edges of $\boldsymbol{B}$, and $\bar{\mathcal{E}}(\boldsymbol{B})$ denote the set of "nonedges" of $\boldsymbol{B}$. We denote by $|\mathcal{E}(\boldsymbol{B})|$ the number of edges in $\boldsymbol{B}$. Then, the Hamming distance squared is given by

$$d_H^2(\boldsymbol{A},\boldsymbol{B}) = |\mathcal{E}(\boldsymbol{B})|^2 + 2|\mathcal{E}(\boldsymbol{B})|\left[\sum_{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})} a_{ij} - \sum_{[i',j']\in\mathcal{E}(\boldsymbol{B})} a_{i'j'}\right] -4\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} a_{ij}a_{i'j'} + \left[\sum_{1\le i<j\le n} a_{ij}\right]^2 \tag{61}$$

The proof of Equation 61 is elementary, and is omitted for brevity. We now provide the proof of lemma 1.

Proof of lemma 1. Applying Equation 61 for each network $G^{(k)}$, we get

$$\hat{F}_2(\boldsymbol{B}) = |\mathcal{E}(\boldsymbol{B})|^2 + 2|\mathcal{E}(\boldsymbol{B})| \times \left[\sum_{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})} \frac{1}{N}\sum_{k=1}^{N} a_{ij}^{(k)} - \sum_{[i',j']\in\mathcal{E}(\boldsymbol{B})} \frac{1}{N}\sum_{k=1}^{N} a_{i'j'}^{(k)}\right] -4\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \left[\frac{1}{N}\sum_{k=1}^{N} a_{ij}^{(k)}a_{i'j'}^{(k)}\right] + \frac{1}{N}\sum_{k=1}^{N}\left[\sum_{1\le i<j\le n} a_{ij}^{(k)}\right]^2 \tag{62}$$

Using the expressions for the sample mean Equation 17 and correlation Equation 18, and observing that

$$\frac{1}{N}\sum_{k=1}^{N}\left[\sum_{1\le i<j\le n} a_{ij}^{(k)}\right]^2 = \sum_{\substack{1\le i<j\le n\\ 1\le i'<j'\le n}} \frac{1}{N}\sum_{k=1}^{N} a_{ij}^{(k)}a_{i'j'}^{(k)} = \sum_{\substack{1\le i<j\le n\\ 1\le i'<j'\le n}} \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right], \tag{63}$$

we get

$$\hat{F}_2(\boldsymbol{B}) = |\mathcal{E}(\boldsymbol{B})|^2 + 2|\mathcal{E}(\boldsymbol{B})|\left[\sum_{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{ij} - \sum_{[i',j']\in\mathcal{E}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{i',j'}\right] - 4\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right] + \sum_{\substack{1\le i<j\le n\\ 1\le i'<j'\le n}} \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right]. \tag{64}$$

Also, we have

$$|\mathcal{E}(\boldsymbol{B})|^2 + 2|\mathcal{E}(\boldsymbol{B})|\left[\sum_{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{ij} - \sum_{[i',j']\in\mathcal{E}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{i',j'}\right]$$
$$= \left[|\mathcal{E}(\boldsymbol{B})| \quad -2\sum_{(i',j')\in\mathcal{E}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{i',j'}\right]\left[|\mathcal{E}(\boldsymbol{B})| + 2\sum_{(i,j)\in\bar{\mathcal{E}}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{ij}\right] + 4\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'}.$$

Whence

$$|\mathcal{E}(\boldsymbol{B})|^2 + 2|\mathcal{E}(\boldsymbol{B})|\left[\sum_{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{ij} - \sum_{[i',j']\in\mathcal{E}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{i',j'}\right]$$
$$= \left[|\mathcal{E}(\boldsymbol{B})| \quad -2\sum_{(i',j')\in\mathcal{E}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{i',j'}\right] \times \left[|\mathcal{E}(\boldsymbol{B})| - 2\sum_{(i',j')\in\mathcal{E}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{i',j'} + 2\sum_{1\le i<j\le n} \hat{\boldsymbol{P}}_{ij}\right] + 4\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'}$$
$$= \left[|\mathcal{E}(\boldsymbol{B})| \quad -2\sum_{(i',j')\in\mathcal{E}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{i',j'}\right]^2 + 2 \times \sum_{1\le i<j\le n} \hat{\boldsymbol{P}}_{ij}\left[|\mathcal{E}(\boldsymbol{B})| - 2\sum_{(i',j')\in\mathcal{E}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{i',j'}\right] + \left[\sum_{1\le i<j\le n} \hat{\boldsymbol{P}}_{ij}\right]^2 + 4 \times \sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'} - \left[\sum_{1\le i<j\le n} \hat{\boldsymbol{P}}_{ij}\right]^2$$

Completing the square yields

$$|\mathcal{E}(\boldsymbol{B})|^2 + 2|\mathcal{E}(\boldsymbol{B})|\left[\sum_{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{ij} - \sum_{[i',j']\in\mathcal{E}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{i',j'}\right]$$
$$= \left[|\mathcal{E}(\boldsymbol{B})| - 2\sum_{(i',j')\in\mathcal{E}(\boldsymbol{B})} \hat{\boldsymbol{P}}_{i',j'} + \sum_{1\le i<j\le n} \hat{\boldsymbol{P}}_{ij}\right]^2 + 4\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\(i',j')\in\mathcal{E}(\boldsymbol{B})}} \hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'}$$
$$- \left[\sum_{1\le i<j\le n} \hat{\boldsymbol{P}}_{ij}\right]^2$$
$$= \left[\sum_{(i',j')\in\mathcal{E}(\boldsymbol{B})} \left(1 - 2\hat{\boldsymbol{P}}_{i',j'}\right) + \sum_{1\le i<j\le n} \hat{\boldsymbol{P}}_{ij}\right]^2 + 4\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'}$$
$$- \sum_{\substack{1\le i<j\le n\\ 1\le i'<j'\le n}} \hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'}. \tag{65}$$

We can then substitute Equations 63, 65 into Equation 64, and we get the result advertised in the lemma,

$$\hat{F}_2(\boldsymbol{B}) = \left[\sum_{[i,j]\in\mathcal{E}(\boldsymbol{B})} \left(1 - 2\hat{\boldsymbol{P}}_{ij}\right) + \sum_{1\le i<j\le n} \hat{\boldsymbol{P}}_{ij}\right]^2 - \sum_{\substack{1\le i<j\le n\\ 1\le i'<j'\le n}} \hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'}\hat{\boldsymbol{P}}_{i',j'}$$
$$- \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right] + 4\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'} \quad - \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right], \tag{66}$$

where we recognize the first term as $\delta^2(\boldsymbol{B}, \hat{\boldsymbol{P}})$.

## 5.2 Proof of lemma 3

Proof of lemma 3. We recall that the residual $\zeta_N(\boldsymbol{B})$ is a sum of two types of terms,

$$\zeta_N(\boldsymbol{B}) = \sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'} - \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right]. \tag{67}$$

The sample mean $\hat{\boldsymbol{P}}_{ij}$, Equation 17, is the sum of $N$ independent Bernoulli random variables, and it concentrates around its mean $p_{ij}$. The variation of $\hat{\boldsymbol{P}}_{ij}$ around $p_{ij}$ is bounded by Hoeffding inequality,

$$\forall 1\le i<j\le n, \forall N\ge 1, \quad \mathbb{P}\Big(\boldsymbol{A}^{(k)} \sim \mathcal{G}(n,p,q);$$
$$\left|\hat{\boldsymbol{P}}_{ij} - p_{ij}\right| \ge \delta\Big) \le \exp\left(-2N\delta^2\right). \tag{68}$$

Let $\varepsilon > 0$, and let $\alpha \overset{\text{def}}{=} \sqrt{\log\left(n/\sqrt{\varepsilon/2}\right)}$, a union bound yields

$$\forall N\ge 1, \mathbb{P}\left(\boldsymbol{A}^{(k)} \sim \mathcal{G}(n,p,q); \forall 1\le i<j<n, \quad \left|\hat{\boldsymbol{P}}_{ij} - p_{ij}\right| \le \frac{\alpha}{\sqrt{N}}\right)$$
$$> 1 - \varepsilon\Big/4. \tag{69}$$

The sample correlation, $\hat{\rho}_{ij,i'j'}$, Equation 18, is evaluated in Equation 67 for $[i,j] \in \mathcal{E}(\boldsymbol{B})$ and $[i',j'] \in \bar{\mathcal{E}}(\boldsymbol{B})$. In this case, the edges $[i,j]$ and $[i',j']$ are always distinct, thus $a_{ij}^{(k)}$ and $a_{i'j'}^{(k)}$ are independent, and $a_{ij}^{(k)}a_{i'j'}^{(k)}$ is a Bernoulli random variable with parameter $p_{ij}p_{i'j'}$. We conclude that $\hat{\rho}_{ij,i'j'}$ is the sum of $N$ independent Bernoulli random variables, and thus concentrates around its mean, $p_{ij}p_{i'j'}$.

Let $\varepsilon > 0$, and let $\beta \overset{\text{def}}{=} \sqrt{\log\left(n^2/\sqrt{2\varepsilon}\right)}$, Hoeffding inequality and a union bound yield

$$\forall N\ge 1, \quad \mathbb{P}\Big(1\le i<j\le n, 1\le i'<j'\le n,$$
$$[i,j] \ne [i',j'], \left|\hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right] - p_{ij}p_{i'j'}\right| \le \frac{\beta}{\sqrt{N}}\Big) > 1 - \varepsilon\Big/2, \tag{70}$$

Combining Equations 69, 70 yields

$$\forall\varepsilon > 0, \exists\alpha_1,\alpha_2,\beta, \forall N\ge 1, \qquad \forall \qquad 1\le i<j\le n, \forall 1\le i'<j'\le n,$$
$$[i,j] \ne [i',j'], \qquad \left|\hat{\boldsymbol{P}}_{ij} - p_{ij}\right| \le \frac{\alpha_1}{\sqrt{N}}, \quad \left|\hat{\boldsymbol{P}}_{i',j'} - p_{i'j'}\right| \le \frac{1}{\sqrt{N}},$$
$$\text{and} \quad \left|\hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right] - p_{ij}p_{i'j'}\right| \le \frac{\beta}{\sqrt{N}}, \tag{71}$$

with probability $1 - \varepsilon$. Lastly, combining Equations 67, 71, we get the advertised result,

$$\forall\ \varepsilon > 0, \exists\ c > 0, \forall\ N \ge 1,$$
$$\mathbb{P}\left(\left|\sum_{\substack{[i,j]\in\bar{\mathcal{E}}(\boldsymbol{B})\\ [i',j']\in\mathcal{E}(\boldsymbol{B})}} \hat{\boldsymbol{P}}_{ij}\hat{\boldsymbol{P}}_{i',j'} - p_{ij}p_{i'j'} - \hat{\mathbb{E}}\left[\rho_{ij,i'j'}\right] + p_{ij}p_{i'j'}\right| \le \frac{c}{\sqrt{N}}\right)$$
$$= \mathbb{P}\left(|\zeta_N(\boldsymbol{B})| \le \frac{c}{\sqrt{N}}\right) > 1 - \varepsilon. \tag{72}$$

## 5.3 Proof of lemma 4

We first provide some inequalities (the proof of which are omitted) that relate $\delta$ to the matrix norm $\|\ \ \|_1$.

**Lemma 8.** Let $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ be weighted adjacency matrices, with $a_{ij}, b_{ij}, c_{ij} \in [0,1]$. We have

$$\frac{1}{2}\|\boldsymbol{A} - \boldsymbol{B}\|_1 \le \delta(\boldsymbol{A},\boldsymbol{B}), \text{and}\ \frac{1}{2}\|\boldsymbol{A} - \boldsymbol{C}\|_1 \le \delta(\boldsymbol{A},\boldsymbol{B}) + \delta(\boldsymbol{B},\boldsymbol{C}). \tag{73}$$

Proof of lemma 4. Let $\boldsymbol{B}, \hat{\boldsymbol{m}}[\mathbb{P}] \in \mathcal{S}$. From the definition of $\hat{F}$ (see Equation 21) we have

$$\hat{F}(\boldsymbol{B}) - \hat{F}(\hat{\boldsymbol{m}}[\mathbb{P}]) = \delta^2\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right) - \delta^2\left(\hat{\boldsymbol{m}}[\mathbb{P}],\hat{\boldsymbol{P}}\right)$$
$$= \left(\delta\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right) - \delta\left(\hat{\boldsymbol{m}}[\mathbb{P}],\hat{\boldsymbol{P}}\right)\right)$$
$$\times\left(\delta\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right) + \delta\left(\hat{\boldsymbol{m}}[\mathbb{P}],\hat{\boldsymbol{P}}\right)\right). \tag{74}$$

Because of Lemma 8, we have

$$\delta\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right) + \delta\left(\hat{\boldsymbol{m}}[\mathbb{P}],\hat{\boldsymbol{P}}\right) \ge \|\boldsymbol{B} - \hat{\boldsymbol{m}}[\mathbb{P}]\|_1. \tag{75}$$

Also,

$$\delta\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right) - \delta\left(\hat{\boldsymbol{m}}[\mathbb{P}],\hat{\boldsymbol{P}}\right) = \sum_{1\le i<j\le n} b_{ij} - \hat{\boldsymbol{m}}[\mathbb{P}]_{ij} - 2\sum_{1\le i<j\le n} \hat{p}_{ij}\left(b_{ij} - \hat{\boldsymbol{m}}[\mathbb{P}]_{ij}\right)$$
$$= \sum_{1\le i<j\le n} \left(1 - 2\hat{p}_{ij}\right)\left(b_{ij} - \hat{\boldsymbol{m}}[\mathbb{P}]_{ij}\right). \tag{76}$$

The entries of $\hat{\boldsymbol{m}}[\mathbb{P}]$ are equal to 1 only along $\mathcal{E}(\hat{\boldsymbol{m}}[\mathbb{P}])$, and 0 along $\bar{\mathcal{E}}(\hat{\boldsymbol{m}}[\mathbb{P}])$. Therefore,

$$\delta\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right)-\delta\left(\hat{\boldsymbol{m}}[\mathbb{P}],\hat{\boldsymbol{P}}\right)=\sum_{[i,j]\in\bar{\mathcal{E}}(\hat{\boldsymbol{m}}[\mathbb{P}])} b_{ij}\left(1-2\hat{p}_{ij}\right) + \sum_{[i,j]\in\mathcal{E}(\hat{\boldsymbol{m}}[\mathbb{P}])}\left(1-b_{ij}\right)\left(2\hat{p}_{ij}-1\right). \tag{77}$$

Let $\varepsilon>0$, because of the concentration of $\hat{p}_{ij}=\hat{\boldsymbol{P}}_{ij}$ around $p_{ij}$, $\exists N_0$, $\forall N\geq N_0$,

$$\mathbb{P}\left(1\leq i<j\leq n,\quad \left|\hat{\boldsymbol{P}}_{ij}-p_{ij}\right|<\varepsilon/2\right)>1-\varepsilon. \tag{78}$$

We recall that we assume that $|p_{ij}-1/2|>\eta$, $\quad 1\leq i<j\leq n$, and therefore we get that for all $0<\varepsilon<2\eta$,

$$\mathbb{P}\left(1\leq i<j\leq n,\quad \left|2\hat{\boldsymbol{P}}_{ij}-1\right|>2\eta-\varepsilon\right)>1-\varepsilon. \tag{79}$$

Because $\hat{\boldsymbol{m}}[\mathbb{P}]$ is constructed using the majority rule, we have

$$\left|2\hat{\boldsymbol{P}}_{ij}-1\right|=\begin{cases}2\hat{p}_{ij}-1 & \text{if}\quad [i,j]\in\mathcal{E}(\hat{\boldsymbol{m}}[\mathbb{P}]),\\ 1-2\hat{p}_{ij} & \text{if}\quad [i,j]\in\bar{\mathcal{E}}(\hat{\boldsymbol{m}}[\mathbb{P}]).\end{cases} \tag{80}$$

Substituting the expression of $|2\hat{\boldsymbol{P}}_{ij}-1|$ in Equation 79 yields the following lower bounds, with probability $1-\varepsilon$,

$$\begin{cases}2\hat{p}_{ij}-1>2\eta-\varepsilon & \text{if}\quad [i,j]\in\mathcal{E}(\hat{\boldsymbol{m}}[\mathbb{P}]),\\ 1-2\hat{p}_{ij}>2\eta-\varepsilon & \text{if}\quad [i,j]\in\bar{\mathcal{E}}(\hat{\boldsymbol{m}}[\mathbb{P}]).\end{cases} \tag{81}$$

Inserting the inequalities given by Equation 81 into Equation 77 gives the following lower bound that happens with probability $1-\varepsilon$,

$$\delta\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right)-\delta\left(\hat{\boldsymbol{m}}[\mathbb{P}],\hat{\boldsymbol{P}}\right)\geq(2\eta-\varepsilon)\times\left[\sum_{[i,j]\in\mathcal{E}(\hat{\boldsymbol{m}}[\mathbb{P}])}\left(1-b_{ij}\right)+\sum_{[i,j]\in\bar{\mathcal{E}}\left(\hat{\boldsymbol{m}}[\mathbb{P}]_{ij}\right)} b_{ij}\right]. \tag{82}$$

We bring the proof to an end by observing that

$$\begin{aligned}\|\boldsymbol{B}-\hat{\boldsymbol{m}}[\mathbb{P}]\|_1 &= \sum_{1\leq i<j\leq n}\left|b_{ij}-\hat{\boldsymbol{m}}[\mathbb{P}]_{ij}\right|\\ &= \sum_{[i,j]\in\mathcal{E}(\hat{\boldsymbol{m}}[\mathbb{P}])}\left|b_{ij}-1\right|+\sum_{[i,j]\in\bar{\mathcal{E}}(\hat{\boldsymbol{m}}[\mathbb{P}])}\left|b_{ij}\right|\\ &= \sum_{[i,j]\in\mathcal{E}(\hat{\boldsymbol{m}}[\mathbb{P}])}\left(1-b_{ij}\right)+\sum_{[i,j]\in\bar{\mathcal{E}}(\hat{\boldsymbol{m}}[\mathbb{P}])} b_{ij},\end{aligned} \tag{83}$$

whence we conclude that

$$\delta\left(\boldsymbol{B},\hat{\boldsymbol{P}}\right)-\delta\left(\hat{\boldsymbol{m}}[\mathbb{P}],\hat{\boldsymbol{P}}\right)\geq(2\eta-\varepsilon)\|\boldsymbol{B}-\hat{\boldsymbol{m}}[\mathbb{P}]\|_1, \tag{84}$$

with probability $1-\varepsilon$. Finally, combining Equations 75, 84, and letting $\alpha\overset{\text{def}}{=}2\eta-\varepsilon$, we get the inequality advertised in Lemma 4,

$$\alpha\|\boldsymbol{B}-\hat{\boldsymbol{m}}[\mathbb{P}]\|_1^2\leq\left|\hat{F}(\boldsymbol{B})-\hat{F}(\hat{\boldsymbol{m}}[\mathbb{P}])\right|, \tag{85}$$

with probability $1-\varepsilon$.

## 5.4 Proof of Equation 55

We show that $|\langle\boldsymbol{u}_i-\boldsymbol{u}_j,\boldsymbol{Q}(\boldsymbol{u}_i-\boldsymbol{u}_j)\rangle|\leq(\frac{1}{d_i}+\frac{1}{d_j})\frac{8\sqrt{2}}{(np)^{3/2}}$ almost surely.

Proof. Let $i,j\in[n]$. We have.

$$\left|\left\langle\boldsymbol{u}_i-\boldsymbol{u}_j,\boldsymbol{Q}\left(\boldsymbol{u}_i-\boldsymbol{u}_j\right)\right\rangle\right|=\left|\left\langle\boldsymbol{u}_i-\boldsymbol{u}_j,\sum_{m=3}^{n}\frac{\lambda_m}{1-\lambda_m}\boldsymbol{\Pi}_m\left(\boldsymbol{u}_i-\boldsymbol{u}_j\right)\right\rangle\right| \tag{86}$$

$$\begin{aligned}&\leq\left\|\sum_{m=3}^{n}\frac{\lambda_m}{1-\lambda_m}\boldsymbol{\Pi}_m\right\|\|\boldsymbol{u}_i-\boldsymbol{u}_j\|^2\\ &\leq\left\{\frac{1}{d_i}+\frac{1}{d_j}\right\}\left\|\sum_{m=3}^{n}\frac{\lambda_m}{1-\lambda_m}\boldsymbol{\Pi}_m\right\|.\end{aligned} \tag{87}$$

Now

$$\left\|\sum_{m=3}^{n}\frac{\lambda_m}{1-\lambda_m}\boldsymbol{\Pi}_m\right\|^2=\sum_{m=3}^{n}\left|\frac{\lambda_m}{1-\lambda_m}\right|^2\|\boldsymbol{\Pi}_m\|^2\leq\sum_{m=3}^{n}2|\lambda_m|^2\|\boldsymbol{\Pi}_m\|^2\leq2\left[\max_{m=2}^{n}|\lambda_m|\right]^2, \tag{88}$$

because the $\boldsymbol{\Pi}_m$ are orthonormal projectors such that $\sum_{m=1}^{n}\boldsymbol{\Pi}_m=\boldsymbol{I}$. Using the following concentration result (e.g., Theorem 3.6 in [44]),

$$\max_{m=3}^{n}|\lambda_m|\leq\frac{8}{\sqrt{np}}\quad\text{almost surely.} \tag{89}$$

we conclude that

$$\left|\left\langle\boldsymbol{u}_i-\boldsymbol{u}_j,\boldsymbol{Q}\left(\boldsymbol{u}_i-\boldsymbol{u}_j\right)\right\rangle\right|\leq\left(\frac{1}{d_i}+\frac{1}{d_j}\right)\frac{8\sqrt{2}}{(np)^{3/2}}\quad\text{almost surely.} \tag{90}$$

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

FM: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. FM was supported by the National Science Foundation (CCF/CIF 1815971).

## Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

# References


1. Dubey P, Müller H-G. Fréchet change-point detection. *The Ann Stat* (2020) 48: 3312–35. doi:10.1214/19-aos1930

2. Ghoshdastidar D, Gutzeit M, Carpentier A, Von Luxburg U. Two-sample hypothesis testing for inhomogeneous random graphs. *The Ann Stat* (2020) 48: 2208–29. doi:10.1214/19-aos1884

3. Ginestet CE, Li J, Balachandran P, Rosenberg S, Kolaczyk ED. Hypothesis testing for network data in functional neuroimaging. *The Ann Appl Stat* (2017) 11:725–50. doi:10.1214/16-aoas1015

4. Kolaczyk ED, Lin L, Rosenberg S, Walters J, Xu J. Averages of unlabeled networks: geometric characterization and asymptotic behavior. *The Ann Stat* (2020) 48:514–38. doi:10.1214/19-aos1820

5. Lunagómez S, Olhede SC, Wolfe PJ. Modeling network populations via graph distances. *J Am Stat Assoc* (2021) 116:2023–40. doi:10.1080/01621459.2020. 1763803

6. Petersen A, Müller H-G. Fréchet regression for random objects with euclidean predictors. *The Ann Stat* (2019) 47:691–719. doi:10.1214/17-aos1624

7. Xu H. Gromov-Wasserstein factorization models for graph clustering. *Proc AAAI Conf Artif intelligence* (2020) 34(04):6478–85. doi:10.1609/aaai.v34i04.6120

8. Zambon D, Alippi C, Livi L. Change-point methods on a sequence of graphs. *IEEE Trans Signal Process* (2019) 67:6327–41. doi:10.1109/tsp.2019.2953596

9. Chowdhury S, Mémoli F. *The metric space of networks* (2018). arXiv preprint arXiv: 1804.02820.

10. Jain BJ. Statistical graph space analysis. *Pattern Recognition* (2016) 60:802–12. doi:10.1016/j.patcog.2016.06.023

11. Snijders TA. Statistical models for social networks. *Annu Rev Sociol* (2011) 37: 131–53. doi:10.1146/annurev.soc.012809.102709

12. Abbe E. Community detection and stochastic block models: recent developments. *J Machine Learn Res* (2018) 18:1–86.

13. Airoldi EM, Costa TB, Chan SH. Stochastic blockmodel approximation of a graphon: theory and consistent estimation. In: *Advances in neural information processing Systems* (2013). p. 692–700.

14. Ferguson D, Meyer FG. Theoretical analysis and computation of the sample Fréchet mean of sets of large graphs for various metrics. *Inf Inference* (2023) 12: 1347–404. doi:10.1093/imaiai/iaad002

15. Olhede SC, Wolfe PJ. Network histograms and universality of blockmodel approximation. *Proc Natl Acad Sci* (2014) 111:14722–7. doi:10.1073/pnas. 1400374111

16. Mitzenmacher M, Upfal E. *Probability and computing: randomization and probabilistic techniques in algorithms and data analysis*. Cambridge, United Kingdom: Cambridge University Press (2017).

17. Monnig ND, Meyer FG. The resistance perturbation distance: a metric for the analysis of dynamic networks. *Discrete Appl Mathematics* (2018) 236:347–86. doi:10. 1016/j.dam.2017.10.007

18. Akoglu L, Tong H, Koutra D. Graph based anomaly detection and description: a survey. *Data Mining Knowledge Discov* (2015) 29:626–88. doi:10.1007/s10618-014-0365-y

19. Donnat C, Holmes S. Tracking network dynamics: a survey using graph distances. *The Ann Appl Stat* (2018) 12:971–1012. doi:10.1214/18-aoas1176

20. Wills P, Meyer FG. Metrics for graph comparison: a practitioner's guide. *PLoS ONE* (2020) 15(2):e0228728–54. doi:10.1371/journal.pone.0228728

21. Banks D, Carley K. Metric inference for social networks. *J classification* (1994) 11: 121–49. doi:10.1007/bf01201026

22. Han F, Han X, Liu H, Caffo B. Sparse median graphs estimation in a high-dimensional semiparametric model. *The Ann Appl Stat* (2016) 10:1397–426. doi:10. 1214/16-aoas940

23. Chen J, Hermelin D, Sorge M. On computing centroids according to the p-norms of hamming distance vectors. In: *27th annual European symposium on algorithms*, 144 (2019). p. 1–28.28

24. Ferrer M, Serratosa F, Sanfeliu A. Synthesis of median spectral graph. In: *Pattern recognition and image analysis: second iberian conference* (2005). p. 139–46.

25. White D, Wilson RC. Spectral generative models for graphs. In: *14th international conference on image analysis and processing (ICIAP 2007)*. IEEE (2007). p. 35–42.

26. Doyle PG, Snell JL. *Random walks and electric networks (Mathematical Assoc. of America)* (1984).

27. Klein D, Randić M. Resistance distance. *J Math Chem* (1993) 12:81–95. doi:10. 1007/bf01164627

28. Bapat RB. *Graphs and matrices*, 27. Springer (2010).

29. Von Luxburg U, Radl A, Hein M. Hitting and commute times in large random neighborhood graphs. *The J Machine Learn Res* (2014) 15:1751–98.

30. Fréchet M. Les espaces abstraits et leur utilité en statistique théorique et même en statistique appliquée. *J de la Société Française de Statistique* (1947) 88:410–21.

31. Sturm K-T. Probability measures on metric spaces of nonpositive. *Heat Kernels Analysis Manifolds, Graphs, Metric Spaces* (2003) 338:357.

32. Jiang X, Munger A, Bunke H. On median graphs: properties, algorithms, and applications. *IEEE Trans Pattern Anal Machine Intelligence* (2001) 23:1144–51. doi:10. 1109/34.954604

33. Löwe M, Terveer S. A central limit theorem for the mean starting hitting time for a random walk on a random graph. *J Theor Probab* (2023) 36:779–810. doi:10.1007/ s10959-022-01195-9

34. Sylvester J. Random walk hitting times and effective resistance in sparsely connected Erdős-Rényi random graphs. *J Graph Theor* (2021) 96:44–84. doi:10.1002/jgt.22551

35. Ottolini A, Steinerberger S. Concentration of hitting times in erdős-rényi graphs. *J Graph Theor* (2023) 107:245–62. doi:10.1002/jgt.23119

36. Wills P, Meyer FG. Change point detection in a dynamic stochastic blockmodel. In: *Complex networks and their applications VIII* (2020). p. 211–22.

37. Levin DA, Peres Y, Wilmer EL. *Markov chains and mixing times*. American Mathematical Soc. (2009).

38. Rempała G. Asymptotic factorial powers expansions for binomial and negative binomial reciprocals. *Proc Am Math Soc* (2004) 132:261–72. doi:10.1090/s0002-9939-03-07254-x

39. Löwe M, Terveer S. *Hitting times for random walks on the stochastic block model* (2024). arXiv preprint arXiv:2401.07896.

40. Deng S, Ling S, Strohmer T. Strong consistency, graph laplacians, and the stochastic block model. *J Machine Learn Res* (2021) 22:1–44.

41. Avrachenkov K, Cottatellucci L, Kadavankandy A. Spectral properties of random matrices for stochastic block model. In: *International symposium on modeling and optimization in mobile, ad hoc, and wireless networks* (2015). p. 537–44.

42. Chakrabarty A, Chakraborty S, Hazra RS. Eigenvalues outside the bulk of inhomogeneous Erdős–Rényi random graphs. *J Stat Phys* (2020) 181:1746–80. doi:10.1007/s10955-020-02644-7

43. Kovalenko I. Theory of random graphs. *Cybernetics* (1971) 7:575–9. doi:10.1007/ bf01071028

44. Chung F, Lu L, Vu V. The spectra of random graphs with given expected degrees. *Internet Math* (2003) 1:257–75. doi:10.1080/15427951.2004.10129089