Fast Computation of the Barycenter Graph in the Spectral Domain

Olivia Courtney*
University of California, Santa Cruz
ocourtne@ucsc.edu

François G. Meyer*
University of Colorado, Boulder
fmeyer@colorado.edu

Abstract

The barycenter graph (or sample Fréchet mean) of a set of networks is an essential location parameter at the core of almost every machine learning algorithm working on networks. In this paper, we introduce a new algorithm to estimate the graph barycenter which is computationally efficient, has reduced dimensionality, and captures topological information of the training set on various scales. We use the eigenvalues of the adjacency matrix to measure the proximity between graphs, and parameterize the definition of the graph barycenter by approximating complex networks using a stochastic block model (SBM). We provide theoretical justifications for the algorithm's design using analysis from spectral graph theory of the eigenvalues of the SBM. Extensive validation is performed on synthetic networks sampled from true SBMs of various geometries, and our algorithm successfully recovers a Fréchet mean converging to the population Fréchet mean associated with the generative model. Additionally, we test the algorithm's robustness with non-SBM graphs exhibiting community structure, and our results show that the algorithm converges in few iterations, is computationally efficient, and that the parameters describing the output sample Fréchet mean graph can quantify significant topological changes in the real networks.

1 Introduction

Throughout many disciplines, practitioners are interested in the topological information of large sets of graph-valued data. Relevant examples of such topologies include brain regions in functional networks in neuroscience or social groups in connectivity networks. The mean graph is an essential statistic for describing such datasets, its computation is central to a wide range of machine learning algorithms.

In metric spaces, the barycenter graph acts as a "center of mass" for the training set, defined by minimizing the sum of the squared distances to the graphs in the set [1], [2]. We call this the Fréchet mean graph. This central approximation task is non-trivial, as the topological information that the barycenter graph recovers depends critically on the distance used to measure the proximity between graphs [3]. A distance that is useful in various applications captures topological information on many scales and is weakly correlated to the size of the graph. We take a spectral approach where the distance uses the eigenvalues of a matrix such as the adjacency matrix.

Our aim is to develop an algorithm that approximates a barycenter graph which reflects the significant topologies (e.g. community structure) of the ensemble and is computationally efficient for the user in various application areas. Motivated by the stochastic block model's (SBM) universal approximating properties [4] [5] [6], we use SBMs to approximate the ensemble networks. This family of model offers well-understood theoretical properties that help us understand real-world networks with community structure [7] [8].

The edit distance, Hamming distance, and Frobenius norm of adjacency matrices have been widely used to compare graphs of different sizes [9][10][11][12]. These distances are all very sensitive

*Equal contribution.
Code implementation can be found at: https://github.com/ocourtney/frechet_mean.git.

Preprint. Preliminary work.

to local, fine-scale graph structures (e.g. density of triangles), however for large-scale properties like community structure spectral distances are better suited [3]. Several authors have proposed the computation of the sample Fréchet mean using the adjacency spectral distance [6][13][14], and we extend these results by proposing a numerical method for use in various application areas in the form of an algorithm.

The main contributions of this paper are as follows. We propose an algorithm to approximate the barycenter graph which inherits the large-scale topological information from the graphs in the sample. Despite the lack of existing algorithms to compare to, we show that our parametric solution is computationally efficient, independent of the size of the graphs in the sample, and supported by theoretical results we provide. We perform extensive validation on synthetic networks sampled from true SBMs of various geometries, demonstrating the successful recovery of a sample Fréchet mean converging to the population Fréchet mean associated with the generative model. Lastly, experiments with non-SBM graphs exhibiting community structure show that the parameters describing the output sample Fréchet mean graph can quantify significant topological changes in real networks.

The paper is structured as follows. After setting our notational standard in Section 2.1, we formally define the sample Fréchet mean and describe our model in Sections 2.2 and 2.3. In Section 2.4 we walk through the steps to form our optimization problem, and in Section 2.6 we introduce the algorithm. Section 2.7 then provides the mathematical justification of the algorithm. In Section 3 we present our experimental set-up and results for both the synthetic data (Section 3.1) and real-world data (Section 3.2). Finally, we discuss our results in Section 4.

2 Approach

2.1 Notation

We first introduce the notational convention used in this paper as well as some clarifying definitions. Consider an unweighted and undirected graph G with **vertex set** $V = \{1, \dots, n\}$, and where the **edge set** E is a subset of $V \times V$. We denote by $\boldsymbol{A}$ the $n \times n$ adjacency matrix, $a_{ij} = a_{ji} = 1$ if the edge $e = [i, j] \in E$; $a_{ij} = a_{ji} = 0$ otherwise. Let $\boldsymbol{D}$ denote the diagonal degree matrix, $d_{ii} = \sum_{j=1}^{n} a_{ij}$. We are also interested in the **eigenvalues** of $\mathbf{A}$. In the case of an $n \times n$ adjacency matrix, we consider the sequence of eigenvalues ranked in descending order, $\boldsymbol{\lambda} = \lambda_1, \lambda_2, \dots, \lambda_n$.

2.2 The Fréchet Mean

To consider the notion of an average graph, we first look at graph-valued random variables. Consider the set of graphs $\mathcal{G}$ with probability matrix $\mathbb{P}$, where each element is a realization of an unweighted simple graph on vertex set V. The 'average' of $\mathcal{G}$ is computed via the Fréchet mean [15] as follows.

Definition 1 *Given a distance d defined on $\mathcal{G}$, the Fréchet mean is the solution to*

$$\mathbb{E}(G) = \arg\min_{G \in \mathcal{G}} \sum_{H \in \mathcal{G}} d^2(G, H)\mathbb{P}(H).$$

This computation can be understood intuitively as a notion of 'center of mass'. The mean graph is the one which has the minimum distance to all other graphs in the set. Despite its usefulness in metric spaces [12][2], when working with data we rarely have a clearly defined $\mathbb{P}$ matrix to describe the randomness of our set. Instead, in practice we perform the computation using a sample set of graphs and desire that the topological features of interest are inherited by the sample Fréchet mean.

Definition 2 *Given a sample of N graphs $\mathbb{G} = \{G^{(1)}, G^{(2)}, \dots, G^{(N)}\}$ with vertex set $\{1, \dots, n\}$, and a distance d defined on $\mathcal{G}$, the sample Fréchet mean is the solution to*

$$\hat{\mathbb{E}}_N[G] = \arg\min_{G \in \mathcal{G}} \frac{1}{N} \sum_{k=1}^{N} d^2(G, G^{(k)}).$$

The choice of distance d depends on which topological features the sample Fréchet mean shall inherit.

2.3 Distance and Model

The distribution of eigenvalues of a graph contains information about its community structure, making it a powerful tool for our study. We introduce our choice of distance here, and give a more detailed discussion of the spectral theory later in Section 2.7. We choose the adjacency spectral distance as our metric for graph comparison.

Definition 3 *Given graphs G and G' of size n with respective adjacency matrix eigenvalue distributions $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$, the adjacency spectral distance between the two graphs is defined as*

$$d(G, G') = \sqrt{\sum_{j=1}^{n} (\lambda_j - \lambda'_j)^2}.$$

Note that this is simply the ℓ^2 norm of the two spectra. Also note that this is not a true distance, but a pseudo-distance, as it is not injective. In practice, we use only the k dominant eigenvalues as we are interested in the community structure and not in the noise. The choice of k depends largely on the data and will be discussed in detail in Section 2.4.

We present our results in the case where the probability space is formed by **Stochastic Blockmodel** $\text{SBM}(n, \mathbf{P})$ with communities of various sizes. While stochastic block models are only a small subsets of inhomogeneous random graphs, they provide universal approximates to graphs [5, 6, 16]. This approximation property of the stochastic block model holds for various norms: the ℓ^2 difference between the corresponding vectors of eigenvalues [6], or the L^2 difference between the corresponding graph limits, which become functions defined on $[0, 1] \times [0, 1]$ [5].

In the stochastic block model, the vertex set can be partitioned into many non-overlapping sets which we call "communities", $V = C_1 \cup C_2 \cup \ldots \cup C_k$. Each edge $e = (i, j)$ exists independently with probability p_1 if i and j are both in community one (C_1), p_2 if i and j are both in C_2, and so forth. An edge exists with probability q if the nodes are in distinct communities [3]. An SBM graph with n nodes and k communities has a symmetric $k \times k$ matrix $\mathbf{P}$ of edge probabilities. In this paper we assume that the graphs are simple (no self-loops and maximum one edge per vertex pair), undirected, and of the same size. We also assume equal number of communities across graphs.

We choose this model because of its universal approximation properties and its applicability to solving our inverse eigenvalue problem (see Section 2.5). While much work has been done on the recovery of communities and community partitions [17][6], we focus here on the recovery of community density parameters p_i and q.

2.4 Formulation of the Optimization Problem

We begin with Definition 2, which defines the sample Fréchet mean graph $\hat{\mathbb{E}}_N[G]$. The domain of this optimization problem is $\mathcal{G}$, the set of all graphs. The Fréchet mean graph has an associated adjacency matrix, which we will call $\hat{A} = A(\hat{\mathbb{E}}_N[G])$. We will use the adjacency spectral distance as defined as in Definition 3, and obtain

$$\hat{\mathbb{E}}_N[G] = \underset{\hat{\mathbf{A}}}{\operatorname{argmin}} \frac{1}{N} \sum_{k=1}^{N} (\boldsymbol{\lambda}(\hat{\mathbf{A}}) - \boldsymbol{\lambda}(\mathbf{A}^{(k)}))^2. \tag{1}$$

This minimization problem's domain is now $\boldsymbol{\lambda}$ such that it is a valid spectrum. To use this domain in practice, however, one would need to evaluate an inverse eigenvalue problem at each optimization step which is both computationally expensive and difficult to define.

2.5 Solving the Inverse Eigenvalue Problem

Solving the optimization problem given in Equation 1 requires solving an inverse eigenvalue problem. We do not solve this intractable problem directly; instead, we approximate $\hat{\mathbf{A}}$ with a stochastic block model $\text{SBM}(\mathbf{P})$ as described in the previous section. This approximation is possible for graphs of any geometry as the model's parameters can be adjusted accordingly. For example, even a regular graph can be modeled by an SBM with one community (an Erdos Renyi graph in the limit of large n

is regular). We now have the following parametric optimization problem where the eigenvalues are a function of $\mathbf{P}$,

$$\hat{\mathbb{E}}_N[G] = \underset{\mathbf{P}\in[0,1]^k}{\operatorname{argmin}} \underset{\hat{\mathbf{A}}_\mathbf{P}\sim\text{SBM}(\mathbf{P})}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^{N}(\boldsymbol{\lambda}(\hat{\mathbf{A}}_\mathbf{P}) - \boldsymbol{\lambda}(\mathbf{A}^{(k)}))^2. \tag{2}$$

One issue remains now, namely that $\hat{\mathbf{A}}_\mathbf{P}$ is stochastic. Using just a single adjacency matrix realization would produce too large of an error given the stochastic nature of the model, so we use the expected behavior instead. Since we are working in the spectral domain, we can leverage the close relationship between the eigenvalues of $\hat{\mathbf{A}}_\mathbf{P}$ and $\mathbb{E}[\hat{\mathbf{A}}]_\mathbf{P}$. We explore this relationship formally in Section 2.7. Finally, we obtain the following optimization problem.

$$\hat{\mathbb{E}}_N[G] = \underset{\mathbf{P}\in[0,1]^k}{\operatorname{argmin}} F(\mathbf{P}), \tag{3}$$

where function $F(\mathbf{P})$ is the following objective function

$$F(\mathbf{P}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k}(\lambda_j(\mathbb{E}[\hat{\mathbf{A}}]_\mathbf{P}) - \lambda_j(\mathbf{A}^{(i)}))^2. \tag{4}$$

Here $\mathbb{E}[\hat{\mathbf{A}}]_\mathbf{P}$ denotes the expected adjacency matrix generated from input $\mathbf{P}$. Although this may look initially like a simple quadratic optimization problem, it is not. We are still solving a constrained eigenvalue problem which cannot be solved analytically. The change in λ follows from the changed parameters of the expected SBM graph, therefore we optimize over density parameters, not eigenvalues.

The number of communities k can be set with prior knowledge of the data, or approximated via a range of existing community detection methods. Several algorithms have been proposed to estimate this meta-parameter (e.g., [16, 18]). We follow the approach described in [19], and identify the number of communities and their sizes based on the ranking of the empirical degree distribution.

Furthermore, the optimization process provides a solution of density parameters, it does not return a graph object. If a graph is desired, one can generate a single realization of the random graph, or alternatively generate a sample of graphs from $\mathbf{P}$ and then randomly select the approximate mean graph from that sample. Depending on the application area of the algorithm, certain methods will be most useful. This is an area for further research.

2.6 Algorithm

The algorithm's input is a sample of graphs in the form of their adjacency matrices. The eigenvalues are extracted from these matrices, $\boldsymbol{\lambda}(\mathbf{A}^{(1)}), \ldots, \boldsymbol{\lambda}(\mathbf{A}^{(N)})$, and (3) is used to find the parameters ($\mathbf{P}$ matrix) of an approximate mean graph, which the algorithm then outputs. Algorithm 1 provides a description of the proposed algorithm.

At each iteration of the algorithm, the community density matrix $\mathbf{P}$ is updated and a new possible expected mean graph $\mathbb{E}[\mathbf{A}]_\mathbf{P}$ is generated. By comparing eigenvalues, the algorithm finds the matrix $\mathbf{P}$ which generates the graph with the least difference to all other graphs in the set.

Algorithm 1 Approximation of the Sample Fréchet Mean

Require: A set of graphs $G = \{G^{(i)}\}_{i=1}^{N}$.
1: Set k and initialize $\mathbf{P} = \mathbf{P_0}$ via the desired method [**].
2: For $i = 1, \ldots, N$, compute $\boldsymbol{\lambda}^i = [\lambda_1(\mathbf{A}^i), \ldots, \lambda_k(\mathbf{A}^i)]$.
3: **while** The relative change in $\mathbf{P}$ is greater than the tolerance **do**
4: Estimate the gradient of $F(\mathbf{P})$ (given by (4)) via finite differences.
5: Update $\mathbf{P}$ via a projected gradient step.
6: **end while**
7: **return** $\mathbf{P}$.

[**] One can initialize $\mathbf{P}$ randomly or by analysing the degree distribution. See, e.g. [6] for more.

Given a set of N graphs $G^1, \ldots, G^N$, where $G^i = (V, E^i)$ for $1 \geq i \geq N$ the sample Fréchet mean graph is given by

2.6.1 Eigenvalues

The computational cost of a spectral algorithm are often significant, as calculating eigenvalues can be quite computationally expensive. We avoid much of this cost by leveraging the community structure of the model. First, setting the number of communities a priori means we must only compute a very small subset of the graph's eigenvalues. Additionally, the size of this subset does not grow with n. For example, assume we intend to find the average of two graphs $G_1(n = 1200, k = 3)$, and $G_2(n = 1200, k = 3)$. The proposed algorithm only requires the information contained in the dominant three eigenvalues, so instead of computing 2×1200 eigenvalues, we compute 2×3. This is an immense computational difference.

In the case where one knows that all the graphs in the sample have the same number of communities but the exact value of k in unknown, the user can still avoid significant computational cost by setting k as the greatest known upper bound.

Additionally, the algorithm uses a sparse solver whenever possible. This is very effective to efficiently obtain the eigenvalues from the adjacency matrices of the real data, which is often sparse. In an application where the perturbation between graphs is small, one could also implement an iterative method for additional computational efficiency.

It is possible that two graphs in the sample have different numbers of communities, in which case they might have a different number of dominant eigenvalues. When comparing eigenvalues from the bulk with dominant eigenvalues, we are comparing random noise with global topological information. Although one can still make this comparison, it is unclear what the outcome signifies mathematically.

2.6.2 Optimization

Classic gradient descent is used to minimize F over $\mathbf{P}$, therefore a discussion of the function's convexity is warranted [20]. We study numerically the smoothness and convexity of F in Section 3. Implementing stochastic gradient descent or another optimization technique more suited for non-convex functions might be useful in scenarios where F is less smooth.

2.7 Theoretical Justification

Our aim in this section is to provide some theoretical bounds on $\lambda_i(\mathbf{A}) - \lambda_i(\mathbb{E}[\mathbf{A}])$. Since $\mathbf{A}$ is random, $\mathbb{E}[\mathbf{A}]$ is the deterministic matrix which describes its average behavior, and we have

$$\mathbf{A} = \mathbb{E}[\mathbf{A}] + \mathbf{Z}, \tag{5}$$

where $\mathbf{Z}$ captures the random fluctuations around the mean,

$$z_{ij} = \begin{cases} \beta(p_m) & \text{if nodes } i, j \text{ are in community } C_m, \\ \beta(q) & \text{if nodes } i, j \text{ are in distinct communities.} \end{cases} \tag{6}$$

where β is a generalized Bernoulli random variable with zero mean. The eigenvalues of $\mathbf{A}$ can be recovered in the correct order from the eigenvalues of $\mathbb{E}[\mathbf{A}]$ (up to a small perturbation).

Theorem 1 (Weyl-Lidskii) *Let $\boldsymbol{A}$ and $\boldsymbol{Z}$ be two $n \times n$ matrices. Let $\lambda_1(\boldsymbol{A}) \geq \ldots \geq \lambda_n(\boldsymbol{A})$ and $\lambda_1(\boldsymbol{A} + \boldsymbol{Z}) \geq \ldots \geq \lambda_n(\boldsymbol{A} + \boldsymbol{Z})$ denote the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{A} + \boldsymbol{Z}$ respectively. Then,*

$$|\lambda_i(\boldsymbol{A}) - \lambda_i(\boldsymbol{A} + \boldsymbol{Z})| \leq \|\boldsymbol{Z}\|, \forall 1 \leq i \leq n.$$

where $\|\boldsymbol{Z}\| = sup(\lambda(\boldsymbol{Z}))$ is the operator norm.

We can rewrite Theorem 1 as $|\lambda_i(\mathbb{E}[\mathbf{A}] + \mathbf{Z}) - \lambda_i(\mathbb{E}[\mathbf{A}])| \leq \|\mathbf{Z}\|$ for $1 \leq i \leq n$. To understand the above difference, we require a bound on $||\mathbf{Z}||$.

Theorem 2 (Bai-Yin [21]) *Let $\boldsymbol{Z}$ be an $n \times n$ symmetric random matrix with entries such that $\mathbb{E}[z_{ij}] = 0$, $Var[z_{ij}] = 1$, and $\mathbb{E}[z_{ij}^4] < \infty$ for $1 \leq i, j \leq n$. Then almost surely,*

$$\lim_{n \to \infty} \frac{||\boldsymbol{Z}||}{\sqrt{n}} \leq 2.$$

In our model, β in its definition has been shifted to have mean zero and

$$\text{Var}[z_{ij}] = \begin{cases} p_m(1-p_m) & \text{if } i,j \in C_m \\ q(1-q) & \text{otherwise.} \end{cases}$$

Then, in the limit of large n we can say that

$$||\mathbf{Z}|| \leq 2\sqrt{n} \leq 2\sqrt{n} \max_{m\in[1,k]} p_m = 2\sqrt{\Delta},$$

where $\Delta = \max_{i\in[1,n]} \sum_{j=1}^{n} a_{ij}$ is the largest degree of the expected matrix $\mathbb{E}[\mathbf{A}]$. Benaych-Georges et al. show this to be a sharp bound given $\frac{\Delta}{logn} \to \infty$ [22], and many papers have shown this bound to be sharp given $p_m \geq c\frac{logn}{n}$ where c is a constant [23][24][25][26]. Thus, we can conclude that the eigenvalues of $\mathbf{A} - \mathbb{E}[\mathbf{A}]$ (or $\mathbf{Z}$, the random component of $\mathbf{A}$) are almost surely bounded by $2\sqrt{logn}$.

Now that we understand the behavior of the eigenvalues of $\mathbf{A}$, we want to also understand the behavior of the eigenvalues of $\mathbb{E}[\mathbf{A}]$. We begin with a simple example.

Example 1 *Given an SBM(n, $\boldsymbol{P}$) model with two equal-sized communities with density p and cross-community density q, we have*

$$\boldsymbol{P} = \begin{bmatrix} p & q \\ q & p \end{bmatrix}.$$

Then,

$$\lambda_1 = \frac{n(p+q)}{2}, \quad \lambda_2 = \frac{n(p-q)}{2}.$$

Given slightly more complicated SBM(n,$\boldsymbol{P}$) with different densities and

$$\boldsymbol{P} = \begin{bmatrix} p_1 & q \\ q & p_2 \end{bmatrix},$$

we can find that

$$\lambda_1 = \frac{n}{2}\Big(\frac{p_1+p_2}{2}\Big) + \frac{n}{4}\Big[(p_1-p_2)^2 + 4q^2\Big]^{1/2}, \quad \lambda_2 = \frac{n}{2}\Big(\frac{p_1+p_2}{2}\Big) + \frac{n}{2}\Big[(p_1-p_2)^2 + 4q^2\Big]^{1/2}.$$

This illustrates the analytical computation of $\lambda_i(\mathbb{E}[\boldsymbol{A}])$ given very simple geometry.

In our model assumption the graphs have a total of k communities with varying densities, $p_1, \ldots, p_k$. The $\mathbf{P}$ matrix is as follows

$$\mathbf{P} = \begin{bmatrix} p_1 & q & \cdots & \cdots & q \\ q & p_2 & q & \cdots & q \\ \vdots & q & p_3 & & \vdots \\ \vdots & \vdots & & \ddots & q \\ q & q & \cdots & q & p_k \end{bmatrix},$$

and we denote its ith eigenvalue $\mu_i(\mathbf{P})$. Given this, we can obtain the analytical expression

$$\lambda_i(\mathbb{E}[\mathbf{A}]) = \frac{n}{k}\mu_i(\mathbf{P}). \tag{7}$$

However, this alone does not provide any information about the bounds on the eigenvalues of the expected adjacency matrix. We can bound $\mu_i(\mathbf{P})$ using the Kato-Temple theorem as in [27], and obtain that for $q < \min_{i\in[1,k]} \frac{|p_i - q|}{2(k-1)}$, for $1 \leq i \leq k$,

$$|\mu_i - p_i| \leq q.$$

Consequently,

$$\lambda_i(\mathbb{E}[\mathbf{A}]) \approx \frac{n}{k}p_i,$$

and we can conclude that $\lambda_i(\mathbb{E}[\mathbf{A}])$ grows linearly with n. This result motivates the choice in Section 2.4 to optimize over $\mathbf{P}$ as a parameter of $\boldsymbol{\lambda}(\mathbb{E}[\mathbf{A}]_{\mathbf{P}})$ since their linear relationship is evident.

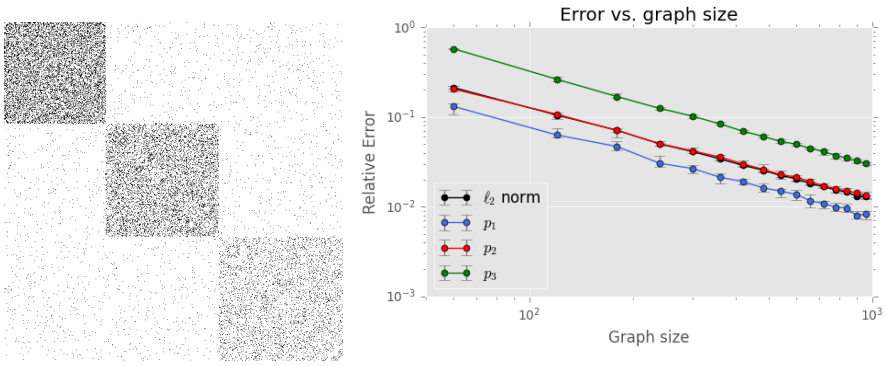


Figure 1: Left: Adjacency matrix of one SBM realization with $n = 512$. Right: Relative error of the algorithm in the limit of large graph size for $N = 30$, $n \in [50, 1000]$, $k = 3$. Community sizes are distinct and relative to n, densities are distinct and constant. The median over 20 trials is plotted.

3 Experiments

3.1 Random graph model

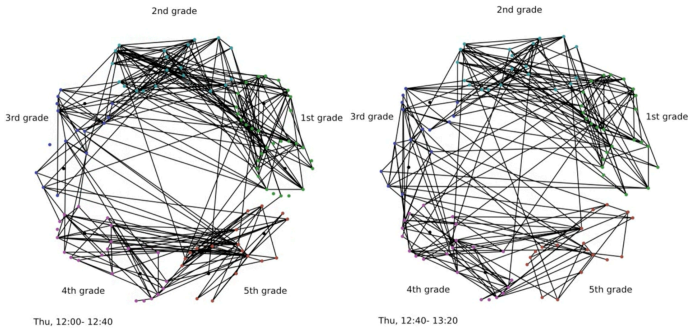
The first set of experiments serve to show that the algorithm converges to the true mean network in the limit of large n (graph size). We generate 50 stochastic block model graphs with $k = 3$ communities for $n \in [50, 1000]$ (see Figure 1-left). The community sizes are distinct and relative to n, and the densities are constant but distinct to one another. We work with SBM graphs to maintain a controlled environment, since these are part of the model. We use three communities because it allows for visual analysis not possible with higher dimensions while maintaining the necessary complexity.

As shown in Figure 1-right, the algorithm converges in the spectral domain. Since the approximated mean graph approaches the true mean graph with large graph size, this shows consistency of the algorithm. We note that there exists a well-known systematic bias introduced by replacing the expected eigenvalues of the adjacency matrix with the eigenvalues of the expected adjacency matrix, which inflates the error here. We bounded this error in the analysis in Section 2.7, so this is not what our experiment is exploring.

When the data exhibits clear community structure following our assumptions, our proposed algorithm is consistent. This is supported by the theoretical analysis, particularly given the result in Equation 7.

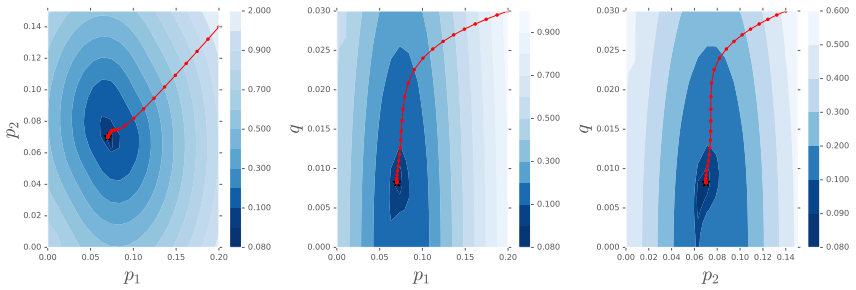
3.2 Real World Networks

Our SBM experiments in Section 3.1 study networks with one isolated feature (community structure), however real networks typically exhibit a combination of features. Therefore, demonstrating the algorithm's functionality on real-world graph data with potential for model error is particularly important to fully show the algorithm's utility.



First Lunch Second Lunch

Figure 2: Snapshots of the contact network during first and second lunch.



(a) (b) (c)

Figure 3: Colormaps of the Fréchet function for sample 1 (first lunch) with axes p_1, p_2, and q. The path of descent is plotted in red.

We study a social contact network between students and teachers of a French primary school recorded over a period of two days in October of 2009 [28] (http://www.sociopatterns.org). Many interesting questions arise with temporal social network data specifically, such as evolving graph structures and change-point detection. Our Fréchet mean algorithm can provide valuable insight into many of these areas.

The primary school face-to-face contacts are recorded every 20 seconds between a total of 242 individuals. The five grades are split into two classes each (ten classes total) that move through instruction, recesses and lunches throughout the day.

From this data we construct the graphs as follows. First, we identify two significant periods in the children's day, namely during first and second lunch. We divide each of these periods into $N = 10$ time intervals that are each four minutes long, and aggregate the two school days. This pre-processing helps ensure that our resulting graphs are well connected. For each time interval we then construct an undirected, unweighted graph of size $n = 242$ where an edge is places if two students had contact within this four minute period (on either day one or day two).

Figure 2 provides a visual representation of the contact networks during these periods. Clearly there is no community structure from the grades, so using the grades to define the stochastic block model approximation would be useless. Instead, we see that students tend to have the most contact with other students in their age group: first, second and third graders play together, and separately fourth and fifth graders play together. Visual inspection of Figure 2 suggests that the communities are more pronounced during first lunch, and less segregated during second lunch. We expect to detect a change in the community densities (contact between younger children and themselves and between older children and themselves) and a change in cross-community density (contact between younger and older children).

We run the Fréchet mean algorithm on both samples. Figure 3 shows the objective function $F(\mathbf{P})$ plotted as a set of three two-dimensional heatmaps with the descent path overlaid for the first lunch networks. The algorithm effectively finds the function minimum with respect to each parameter. We see that the function has an ideal shape for gradient descent algorithms even with real-word data, exhibiting a clear minimum and a smooth contour. See Appendix A for additional results for second lunch.

Each snapshot is taken where the third parameter is held constant at its optimal value. For example, Figure 3(a) shows p_1 vs p_2 for $q \approx 0.008$, which is its argmin value. We see little change in the p_1 and p_2 parameters between the first and second lunch, however q increases from ≈ 0.008 to ≈ 0.010. This change in cross-community density aligns with visual inspection of 2, and indicates how our algorithm may be used for change point detection problems or analysis of a time-series regression.

3.3 Benchmark

Finally, we evaluate the algorithm on the LFR-Benchmark graphs [29]. This benchmark allows for a ground-truth community assignment while still mirroring real complex networks by assigning node degree and community sizes according to heterogeneous distributions. It is widely used to evaluate

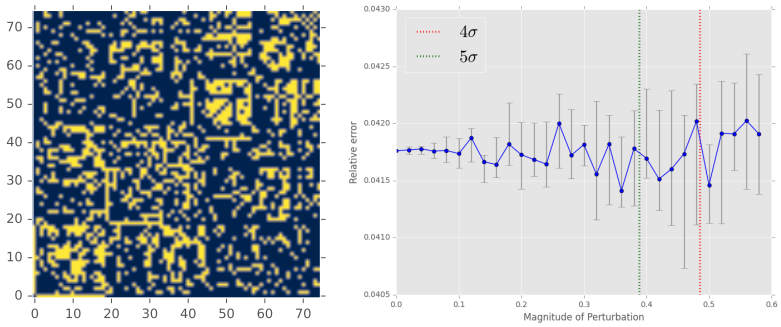


Figure 4: Left: Adjacency matrix of one LFR network with $n = 75$. Right: Relative error of the algorithm as the dominant eigenvalues are perturbed. The median over 20 trials is plotted.

community detection algorithms and offers a more rigorous testing environment than simple SBM synthetic graphs.

We generate one unweighted, undirected LFR network with $n = 75$ and three non-overlapping communities. Then we slightly perturb the eigenvalues of the adjacency matrix of this network to create a 'noisy' sample of 30 graphs. Figure 4 shows the relative error of the recovery of the true mean as the magnitude of the eigenvalue perturbation (i.e. magnitude of the noise) increases. The median of 20 trials is plotted and we see that, on expectation, our algorithm returns the true mean graph. Furthermore, we see that our algorithm acts as an unbiased estimator.

4 Conclusion

We introduce an algorithm that estimates the barycenter graph which inherits large-scale topological features. Our parameterized approach is supported by the presented theory and our experimental results confirm the algorithm's functionality and potential. The primary area of future research is addressing the algorithm's approach to graphs of different sizes using the normalized Laplacian matrix (defined by $\mathcal{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{A})\mathbf{D}^{-1/2}$) instead of $\mathbf{A}$. This allows for spectral comparison between graphs of different sizes. Another current limitation is for graphs with different numbers of communities, for which a more extensive analysis of the separation of eigenvalues would be useful. Higher level considerations such as when the model might fail and adding or subtracting communities are also interesting areas for further investigation.

References

[1] Samir Chowdhury and Facundo Mémoli. The metric space of networks. *arXiv preprint arXiv:1804.02820*, 2018. 1

[2] Eric D. Kolaczyk, Lizhen Lin, Steven Rosenberg, Jackson Walters, and Jie Xu. Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *The Annals of Statistics*, 48(1):514 – 538, 2020. doi: 10.1214/19-AOS1820. URL https://doi.org/10.1214/19-AOS1820. 1, 2

[3] Peter Wills and François G Meyer. Metrics for graph comparison: a practitioner's guide. *Plos one*, 15(2):e0228728, 2020. 1, 2, 3

[4] Edoardo M Airoldi, Thiago B Costa, and Stanley H Chan. Stochastic blockmodel approximation of a graphon: Theory and consistent estimation, 2013. URL https://arxiv.org/abs/1311.1731. 1

[5] Sofia C Olhede and Patrick J Wolfe. Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 111(41):14722–14727, 2014. 1, 3

[6] Daniel Ferguson and François G Meyer. Theoretical analysis and computation of the sample fréchet mean of sets of large graphs for various metrics. *Information and Inference: A Journal of the IMA*, 12(3):1347–1404, 2023. 1, 2, 3, 4

[7] Tom A. B. Snijders. *Statistical Models for Social Networks*, pages 131–153. Annual Review of Sociology. Annual Reviews, 2011. ISBN 978-0-8243-2237-3. doi: 10.1146/annurev.soc.012809.102709. 1

[8] Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018. URL http://jmlr.org/papers/v18/16-480.html. 1

[9] M. Ferrer, E. Valveny, and F. Serratosa. Median graph: A new exact algorithm using a distance based on the maximum common subgraph. *Pattern Recognition Letters*, 30(5):579–588, 2009. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec.2008.12.014. URL https://www.sciencedirect.com/science/article/pii/S0167865509000026. 1

[10] Brijnesh J Jain. Geometry of graph edit distance spaces. *arXiv preprint arXiv:1505.08071*, 2015. 1

[11] Simón Lunagómez, Sofia C Olhede, and Patrick J Wolfe. Modeling network populations via graph distances. *Journal of the American Statistical Association*, 116(536):2023–2040, 2021. 1

[12] Brijnesh J. Jain. On the geometry of graph spaces. *Discrete Applied Mathematics*, 214(C): 126–144, 2016. doi: 10.1016/j.dam.2016.06.027. 1, 2

[13] R. C. Wilson and D. White. Spectral generative models for graphs. In *2007 14th International Conference on Image Analysis and Processing - ICIAP 2007*, pages 35–42, Los Alamitos, CA, USA, sep 2007. IEEE Computer Society. doi: 10.1109/ICIAP.2007.119. URL https://doi.ieeecomputersociety.org/10.1109/ICIAP.2007.119. 2

[14] Miquel Ferrer, Francesc Serratosa, and Alberto Sanfeliu. Synthesis of median spectral graph. In *Pattern Recognition and Image Analysis: Second Iberian Conference, IbPRIA 2005, Estoril, Portugal, June 7-9, 2005, Proceedings, Part II 2*, pages 139–146. Springer, 2005. 2

[15] Maurice Fréchet. 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. URL http://eudml.org/doc/199285. 2

[16] Peter J Bickel and Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50): 21068–21073, 2009. 3, 4

[17] Emmanuel Abbe, Afonso S. Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2016. doi: 10.1109/TIT.2015.2490670. 3

[18] Edo M Airoldi, Thiago B Costa, and Stanley H Chan. Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems*, pages 692–700, 2013. 4

[19] Stanley Chan and Edoardo Airoldi. A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning*, pages 208–216. PMLR, 2014. 4

[20] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 5

[21] Z. D. Bai and Y. Q. Yin. Necessary and Sufficient Conditions for Almost Sure Convergence of the Largest Eigenvalue of a Wigner Matrix. *The Annals of Probability*, 16(4):1729 – 1741, 1988. doi: 10.1214/aop/1176991594. URL https://doi.org/10.1214/aop/1176991594. 5

[22] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Spectral radii of sparse random matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(3):2141 – 2161, 2020. doi: 10.1214/19-AIHP1033. URL https://doi.org/10.1214/19-AIHP1033. 6

[23] Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1), February 2015. ISSN 0090-5364. doi: 10.1214/14-aos1274. URL http://dx.doi.org/10.1214/14-AOS1274. 6

[24] Peter Chin, Anup Rao, and Van Vu. Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Conference on Learning Theory*, pages 391–423. PMLR, 2015. 6

[25] Van Vu. Spectral norm of random matrices. *Combinatorica*, 27:721–736, 11 2007. doi: 10.1007/s00493-007-2190-z. 6

[26] Afonso S. Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4), July 2016. ISSN 0091-1798. doi: 10.1214/15-aop1025. URL http://dx.doi.org/10.1214/15-AOP1025. 6

[27] Arun Kadavankandy. *Spectral analysis of random graphs with application to clustering and sampling*. Theses, COMUE Université Côte d'Azur (2015 - 2019), July 2017. URL https://theses.hal.science/tel-01618579. 6

[28] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems. High-resolution measurements of face-to-face contact patterns in a primary school. *PLOS ONE*, 6(8):1–13, 08 2011. doi: 10.1371/journal.pone.0023176. URL https://doi.org/10.1371/journal.pone.0023176. 8

[29] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E*, 78:046110, Oct 2008. doi: 10.1103/PhysRevE.78.046110. URL https://link.aps.org/doi/10.1103/PhysRevE.78.046110. 8

A Appendix

Figure 5 shows the objective function F plotted as a three-dimensional scatter plot along the p_1, p_2, and p_3 axes. The plot shows that the function indeed has a minimum point at specific density parameters. For simplicity, the cross-community density parameter q was kept constant at the generated value to maintain a three-dimensional visual.

Figure 6 is the same as Figure 3, but for the second lunch time period. We observe the same smooth shape and descent path.

Figure 5: 3D heatmap of the sample Fréchet function for three-community SBM sample.

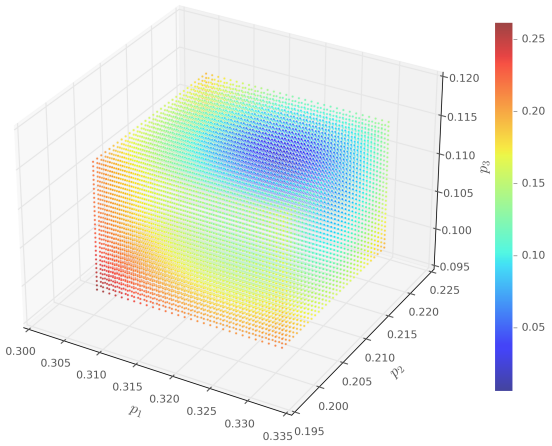
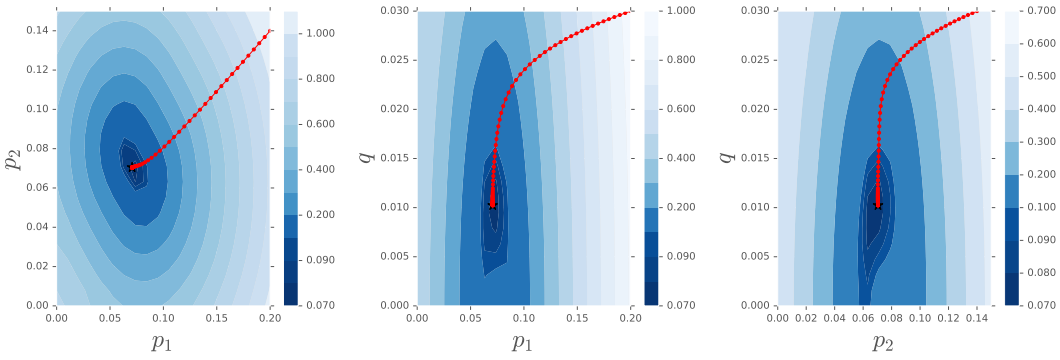


Figure 6: Same as Figure 3 but for sample 2 (second lunch)



(a) (b) (c)