Fréchet Mean Computation in Graph Space through Projected Block Gradient Descent

Nicolas Boria, Benjamin Negrevergne and Florian Yger *

CNRS, LAMSADE, PSL-Université Paris Dauphine
Place du Marécal de Lattre de Tassigny, Paris, France

Abstract. A fundamental concept in statistics is the concept of Fréchet sample mean. While its computation is a simple task in Euclidian space, the same does not hold for less structured spaces such as the space of graphs, where concepts of distance or mid-point can be hard to compute. We present some work in progress regarding new distance measures and new algorithms to compute the Fréchet mean in the space of Graphs.

1 Introduction

In a wide range of scientific domains, graphs provide a powerful tool of representation. However, the computation of some fundamental data analysis concepts such as the distance between two elements or the central element of a collection remain challenging tasks in the space of graphs. The main difficulty when dealing with these tasks is the absence of canonical vectorial representations for graphs, in other words, a single graphs can be represented by an exponential number of different adjacency matrices, that correspond to all possible permutations of the set of vertices. Therefore, measuring a distance between two graphs requires first some method of alignment. A popular approach is based on the so called *Graph Edit Distance* (GED), defined as the minimal cost of an edit path between two graphs, where an edit path is a series of edit operations (insertions, deletions and label substitutions) on both edges and nodes. It was shown in [1] that under natural assumptions on the edit costs, there is a bijection between the space of minimal edit paths and the space of node-maps, that assign or *align*, each vertex of the first graph to either a vertex of the second, or to a dummy vertex that corresponds to a deletion operation. Based on such a measure, the computation of a median element, called *Generalized Median Graph* [2] was proposed through genetic algorithms [3], recursive computation of pairwise weighted mean graphs [4] and Block Coordinate Update [5].

In this work, we focus on a different approach, motivated by machine learning applications. While the generalized median graph provides a good representative element of a collection, it does not provide a generative model for it, which is a fundamental element to compute in machine learning, as it allows to sample new elements for the collection. The graph mean -or graph Fréchet mean is very different in nature as it can be considered as a the parameter of a probability distribution on graphs. Indeed, while the median element of a collection is part of the same space as the elements of the collection (*e.g.* the generalized median

---

graph is indeed a graph), the Fréchet mean of a collection is an element of the convex hull of the underlying space. In this sense, the graph Fréchet mean provides a generative model for a collection of graphs.

The paper is organized as follows: section 2 provides some basic definitions on graph matching, and presents our assumptions on graph generative models, section 3 introduces our general model for Fréchet mean computation and evaluation based on Block coordinate update, section 4 presents some empirical evaluations of our methods, through generative model inference and classification. Finally section 5 gives some conclusions and perspectives for future works.

2 Preliminaries

2.1 Problem Definition and Formulations

We first consider a collection $\mathcal{G} = \{G_1, ..., G_N\}$ of connected simple and unlabeled graphs represented by their adjacency matrices $\mathcal{A} = \{A_1, ..., A_N\}$, and suppose w.l.o.g that all matrices in $\mathcal{A}$ have the same size $n = \max |G_i| : G_i \in \mathcal{G}$. Adjacency matrices of graphs with order smaller than n are spanned with 0 lines and columns. For simplicity, in the rest of the paper we do not distinguish graphs and their matrix representations, so that A_i may denote either an adjacency matrix or the graph it represents. Let Π_n denote the space of permutation matrices of size n. Recall that $\Pi_n = \mathcal{P}_n \cap \mathcal{O}_n$ where $\mathcal{P}_n$ and $\mathcal{O}_n$ denote respectively the spaces of bistochastic and orthogonal matrices of size n.

We define the distance between two graphs A and B as follows:

$$d(A, B) = \sqrt{\min_{P \in \Pi_n} \|P^T A P - B\|_{\mathcal{F}}^2} = \sqrt{\min_{P \in \Pi_n} \|AP - PB\|_{\mathcal{F}}^2} \qquad (1)$$

As remarked in [6], the two formulations are equivalent over the set of permutation matrices but are not equivalent when relaxed. Note that this definition of distance is invariant over permutations of A and B. Following this definition of distance, the Fréchet variance of a matrix B with respect to collection $\mathcal{A}$ is $\Psi(B, \mathcal{A}) = \sum_{A_i \in \mathcal{A}} d^2(A_i, B)$.

The Fréchet mean then consists in minimizing the Fréchet variance of B (and simultaneously over P_i). The distance defining in Eq. 1 is formulated as a Quadratic assignment problem. Hence, this sub-problem of the Fréchet mean being computationally already very challenging, we will relax the constraints over P_i. Namely, instead of requiring P_i to be a permutation matrix, we will only require it be a bistochastic matrix, which - as mentionned earlier - results in two different formulations of the problem:

$$\min_{\substack{P_i \in \mathcal{P}_n \\ B \in \mathcal{B}_n}} \sum_{A_i \in \mathcal{A}} \|P_i^T A_i P_i - B\|_{\mathcal{F}}^2 \qquad (2)$$

$$\min_{\substack{P_i \in \mathcal{P}_n \\ B \in \mathcal{B}_n}} \sum_{A_i \in \mathcal{A}} \|A_i P_i - P_i B\|_{\mathcal{F}}^2 \qquad (3)$$
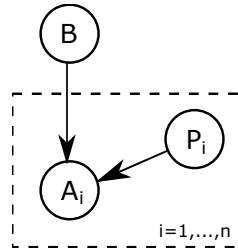
Fig. 1: Generative model implied by the Fréchet mean over graphs.

Where $\mathcal{B}_n$ denotes the convex hull of adjacency matrices of order n. As noted in [6], the relaxed version of the initial problem expressed by Eq. 3 is convex in each P_i (and also in convex in B but not jointly). Hence, in what follows we will denote this formulation as the *biconvex formulation*, while we will denote the one expressed by Eq. 2 as the *base formulation*. Moreover, for a fixed set of bistochastic matrices P_i's, both these formulations admit a closed form for the value of B that optimizes them. For the base formulation, the optimal B is:

$$B^* = \frac{1}{|\mathcal{A}|} \sum_i P_i^T A_i P_i \tag{4}$$

While for the convex formulation, the optimal B is as follows:

$$B^* = \left(\sum_i P^T P \right)^{-1} \sum_i P_i^T A_i P_i \tag{5}$$

It is shown in [6] that the biconvex formulation, while it is easier to solve, will almost certainly yield a suboptimal matching when the resulting bistochastic matrix is projected onto the space of permutation matrices.

2.2 Underlying generative model

The Fréchet mean induces a generative model on the set of graphs where each adjacency matrix A_i would be sampled from an inhomogeneous Erdős–Rényi graph[1] (ERG) shuffled by the permutation P_i. B contains the probability of the ERG model. As illustrated in Figure 1, the variables B and P_i are independent in the model and it suggests the use of an EM-like algorithm that would alternate between the estimation of B and P_i.

3 Block-coordinate descent

Most algorithms proposed in the literature [7] make use of possibly costly graph matching heuristics in order to align each graph of the collection to the current

---

[1]In an inhomogenous ERG, the edge between kl appears with a probability B_{kl}.

mean. In what follows, a separate gradient descent on each P_i will act as our graph matching heuristic, and we denote it by GM*.

In order to avoid the cost induced by the use of GM*, we devise the following Block-coordinate descent (BCD) algorithm for the problem at hand:

Algorithm 1: Algorihtm $\text{BCD}_{k,s}(\mathcal{A})$

Result: a sample Fréchet mean B of collection $\mathcal{A}$
initialize $B \leftarrow A_1$;
initialize P_i's using LSAP based on degrees of vertices ;
while *not converged* **do**
   foreach P_i **do**
      perform k gradient steps on P_i ;
      project P_i onto $\mathcal{P}_n$ using [8] ;
      sample s permutations graph from P_i, and denote
       by P_i^* the one with minimal norm w.r.t B ;
      $P_i \leftarrow P_i^*$
   end
   Update B using either Eq. 4 or Eq. 5 based on the considered model;
end

Moreover, the formulation in 3 being convex in P_i and B advocates the use of a block-coordinate descent (which will converge to a local minimum).

4 Experiments

4.1 Synthetic data models

In order to evaluate the capacity of our methods to infer underlying generative models, we based our experiments on synthetic data, so that the underlying model is known. We developed a stochastic model generator, and tuned it to generate 15 models, based on three shapes : A hamiltonian path (model P), a star (model S), and a sparse union of a hamiltonian path and a star (model PS). For each shape, 5 models with different level of noise σ were generated as represented in Table 1.

4.2 Experimental Setting

We tested 3 different tunings of $\text{BCD}_{k,s}$ along with two algorithms found in the literature. The three versions of $\text{BCD}_{k,s}$ are $\text{BCD}_{2,50}$, $\text{BCD+}_{2,50}$ (where where initial P_i's are generated using GM* instead of LSAP) and $\text{BCD}_{100,50}$ where P_i's are fully optimized at each iteration using GM* (similarily to MMM algorithm proposed in [7]). The two algorithms also found in [7] are DM (Direct Mean) which consists of a single iteration of $\text{BCD}_{100,50}$, IAM (Incremental Arithmetic Mean): Given a random ordering of the graphs, perform an iterative computation of weighted means between the current solution and the next graph in the ordering. Weighted mean is computed using GM* for alignment. Each algorithm was tested using both the base formulation of the distance (M1) and the biconvex one (M2)
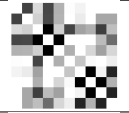
	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
P					
S					
PS					

Table 1: Generative models

4.3 Classification experiments

For each considered value σ, we sampled 30 graphs from each model P, S and PS, which constitute the three classes of our synthetic dataset. A training set of 20 graphs was randomly chosen in each class, and used to compute the mean graph for each class. The remainder of the dataset was classified using a simple Mininum Distance classifier (each graph was classified in the same class as the closest class representative). Each experiment was repeated 20 times, and average classification accuracies on all runs are presented in Table 2. Table 3 provides the time needed to compute Fréchet means averaged on all classes and all runs only for the slowest and fastest algorithm due to space constraints. Typically, $BCD_{2,50}$ was two times faster than the second fastest algorithm (usually IAM), and ten times faster than $BCD_{100,50}$. Moreover, it is interesting to note that among all algorithms, $BCD_{2,50}$ is the only one which running time was not affected by the difficulty of the mean computation, represented by parameter σ. We also note that, although the quality of the classification through $BCD_{2,50}$ means is always outperformed by $BCD_{100,50}$, the difference in accuracy, does not seem significant enough to justify the difference in computation time.

When comparing each model for a fixed choice of solver, we observe that the base formulation seems to perform slightly better than the biconvex formulation. This relates to [6] where a non-convex relaxation was favoured over a convex one for computing a graph matching.

5 Conclusion

In this work, we have proposed a relaxation for the problem of finding the Fréchet mean of a set of unlabelled graphs, and we have shown empirically that aligning graphs to the means in a progressive way (using few gradient steps at a time) proves an efficient method. The extension of this work to labelled graphs and its application to chemoinformatic (as carried out in [5]) is in progress. This work is a first step toward new ways for handling non-Euclidean data [9, 10].

	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
M1-BCD$_{2,50}$	0.970	0.861	0.742	0.605	0.698
M1-BCD+$_{2,50}$	0.975	0.887	0.798	0.587	0.664
M1-BCD$_{100,50}$	0.988	0.877	0.801	0.629	0.7
M1-DM	0.970	0.898	0.774	0.551	0.646
M1-IAM	0.977	0.888	0.722	0.646	0.620
M2–BCD$_{2,50}$	0.929	0.816	0.766	0.601	0.661
M2-BCD+$_{2,50}$	0.977	0.868	0.794	0.6	0.631
M2-BCD$_{100,50}$	0.966	0.848	0.768	0.605	0.672
M2-DM	0.977	0.861	0.761	0.566	0.625
M2-IAM	0.966	0.905	0.764	0.622	0.624

Table 2: Classification accuracies

	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
M1-BCD$_{2,50}$	19.12	21.73	23.94	21.22	18.25
M1-BCD$_{100,50}$	227.2	253.2	317.9	374.0	360.5
M2–BCD$_{2,50}$	20.28	18.83	19.29	21.89	24.71
M2-BCD$_{100,50}$	284.9	272.7	322.7	364.8	345.8

Table 3: Training times

References

[1] Sébastien Bougleux, Luc Brun, Vincenzo Carletti, Pasquale Foggia, Benoit Gaüzere, and Mario Vento. A quadratic assignment formulation of the graph edit distance. 2015.

[2] X. Jiang, A. Munger, and H. Bunke. On median graphs: properties, algorithms, and applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(10):1144–1151, 2001.

[3] M. Ferrer, E. Valveny, and F. Serratosa. Median graphs: A genetic approach based on new theoretical properties. *Pattern Recognition*, 42(9):2003–2012, 2009.

[4] M. Ferrer, E. Valveny, F. Serratosa, K. Riesen, and H. Bunke. Generalized median graph computation by means of graph embedding in vector spaces. *Pattern Recognition*, 43(4):1642–1655, 2010.

[5] Nicolas Boria, Sébastien Bougleux, Benoit Gaüzère, and Luc Brun. Generalized median graph via iterative alternate minimizations. In *GbRPR 2019*, volume 11510 of *LNCS*, pages 99–109, Cham, 2019. Springer.

[6] Vince Lyzinski, Donniell E Fishkind, Marcelo Fiori, Joshua T Vogelstein, Carey E Priebe, and Guillermo Sapiro. Graph matching: Relax at your own risk. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):60–73, 2015.

[7] Brijnesh J Jain. Statistical graph space analysis. *Pattern Recognition*, 60:802–812, 2016.

[8] John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the l 1-ball for learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning*, pages 272–279. ACM, 2008.

[9] Aurélien Bellet, Amaury Habrard, and Marc Sebban. A survey on metric learning for feature vectors and structured data. *arXiv preprint arXiv:1306.6709*, 2013.

[10] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.