

Large Sample Statistics in the Domain of Graphs

Brijnesh J. Jain and Klaus Obermayer

Berlin Institute of Technology, Germany
{jbj, oby}@cs.tu-berlin.de

Abstract. One challenge in bridging the gap between structural and statistical pattern recognition consists in studying combinatorial structures like graphs using probabilistic methods. This contribution presents the structural counterparts of the first and second fundamental theorem in probability, (1) the law of large numbers and (2) the central limit theorem. In addition, we derive characterizations and uniqueness conditions for the mean of graphs. As a special case, we investigate the weighted mean of two graphs. The proposed results establish a sound statistical foundation for unsupervised structural pattern recognition methods.

1 Introduction

Central points such as the median and mean of a finite set of graphs find their applications in central clustering of graphs [5,9,10,11,19], graph quantization [13], frequent substructure mining [18] and multiple alignment of protein structures [14]. Because of their elementary importance, a thorough understanding of central points for a distribution of graphs is necessary in order to statistically justify and algorithmically improve existing unsupervised structural pattern recognition methods. For this reason, first theoretical results on central points in the domain of graphs have been established [4,6,12,15,17]. Compared to vector spaces, however, a fundamental understanding of the graph mean is still missing.

This paper aims at providing new insight to basic properties in large sample statistics of attributed graphs. We restate the strong law of large numbers for distributions on graphs presented in [15]. As novel results, we (1) propose a central limit theorem for distributions on graphs, (2) characterize the mean of graphs, (3) propose sufficient conditions for uniqueness of the mean of graphs, and (4) present properties of the weighted mean of two graphs. In order to derive these results an appropriate approach to represent graphs is necessary. The approach we suggest is to represent graphs as points in some Riemannian orbifold. An orbifold is a quotient of a manifold by a finite group action and therefore generalizes the notion of manifold. Using orbifolds we can derive an intrinsic metric that enables us to adopt integration locally to a Euclidean space.

The proposed approach has the following properties: First, it can be applied to finite combinatorial structures other than graphs like, for example, point patterns, sequences, trees, and hypergraphs can all be embedded isometrically into a Riemannian orbifold. For the sake of concreteness, we restrict our attention

exclusively to the domain of graphs. Second, for graphs consisting of a single vertex with feature vectors as attributes, the proposed learning graph quantization (LGQ) coincides with LVQ.

This paper is organized as follows: Section 2 represents attributed graphs as points in an orbifold. Section 3 derives properties of the graph mean and Section 4 concludes.

2 Representation of Attributed Graphs

In order to do statistical data analysis, we need an appropriate representation of attributed graphs. We suggest to represent graphs as points in some Riemannian orbifold, since orbifolds allow us to apply useful concepts and techniques from differential geometry.

Let \mathbb{E} be a d -dimensional Euclidean space. An *attributed graph* is a triple $X = (V, E, \alpha)$ consisting of a set V of *vertices*, a set $E \subseteq V \times V$ of *edges*, and an *attribute function* $\alpha : V \times V \rightarrow \mathbb{E}$, such that $\alpha(i, j) \neq \mathbf{0}$ for each edge and $\alpha(i, j) = \mathbf{0}$ for each non-edge. Attributes $\alpha(i, i)$ of vertices i may take any value from \mathbb{E} .

For simplifying the mathematical treatment, we assume that all graphs are of order n , where n is chosen to be sufficiently large. Graphs of order less than n , say $m < n$, can be extended to order n by including isolated vertices with attribute zero. For practical issues, it is important to note that limiting the maximum order to some arbitrarily large number n and extending smaller graphs to graphs of order n are purely technical assumptions to simplify mathematics. For pattern recognition problems, these limitations should have no practical impact, because neither the bound n needs to be specified explicitly nor an extension of all graphs to an identical order needs to be performed. When applying the theory, all we actually require is that the graphs are finite.

A graph X is completely specified by its *matrix representation* $\mathbf{X} = (\mathbf{x}_{ij})$ with elements $\mathbf{x}_{ij} = \alpha(i, j)$ for all $1 \leq i, j \leq n$. Let $\mathcal{X} = \mathbb{E}^{n \times n}$ be the Euclidean space of all $(n \times n)$ -matrices with elements from \mathbb{E} and let Γ denote a subgroup of all $(n \times n)$ -permutation matrices. Two matrices $\mathbf{X}, \mathbf{X}' \in \mathcal{X}$ are said to be equivalent, if there is a permutation matrix $P \in \Gamma$ such that $P^T \mathbf{X} P = \mathbf{X}'$. By

$$\mathcal{X}/\Gamma = \{[\mathbf{X}] : \mathbf{X} \in \mathcal{X}\}$$

we denote the quotient set consisting of all equivalence classes $[\mathbf{X}]$.

For notational convenience, we identify \mathcal{X} with \mathbb{E}^N , where $N = n^2$ and consider vector- rather than matrix representations of graphs. By concatenating the columns of a matrix representation \mathbf{X} of a graph X , we obtain a *vector representation* \mathbf{x} of X .

Now we are in the position to take the final step towards representing graphs as points in a Riemannian orbifold. A *Riemannian orbifold of graphs* is a triple $\mathcal{Q} = (\mathcal{X}, \Gamma, \pi)$ consisting of an Euclidean space \mathcal{X} with norm $\|\cdot\|$, a permutation group Γ acting on \mathcal{X} , and an *orbifold chart*

$$\pi : \mathcal{X} \rightarrow \mathcal{X}_{\mathcal{Q}} = \mathcal{X}/\Gamma, \quad \mathbf{x} \mapsto [\mathbf{x}]$$

that projects each vector \mathbf{x} to its orbit $[\mathbf{x}]$. We use capital letters X, Y, Z to denote graphs from $\mathcal{X}_{\mathcal{Q}}$ and write $\mathbf{x} \in X$ if $\pi(\mathbf{x}) = X$. Each vector $\mathbf{x} \in X$ is a *vector representation* of structure X and the set \mathcal{X} of all vector representations is the *representation space* of $\mathcal{X}_{\mathcal{Q}}$.

The *intrinsic metric* of an orbifold $\mathcal{Q} = (\mathcal{X}, \Gamma, \pi)$ of graphs is of the form

$$d(X, X') = \min \{ \|\mathbf{x} - \mathbf{x}'\| : \mathbf{x} \in X, \mathbf{x}' \in X' \}.$$

We call a pair $(\mathbf{x}, \mathbf{x}') \in X \times X'$ with $\|\mathbf{x} - \mathbf{x}'\| = d(X, X')$ an *optimal alignment* of X and X' . By $\mathcal{A}(X, X')$ we denote the set of all optimal alignments of X and X' . Note that the intrinsic metric is not a artificial construction for analytical purposes but rather appears in different guises as a common choice of proximity measure for graphs [2,3,8,20].

3 The Frechet Mean

In this section, we focus on the mean of a distribution on graphs. Unless otherwise stated, proofs of all results are delegated to [16].

3.1 The Frechet Mean Set of Graphs

Since it is unclear how to define the mean of graphs using a weighted sum or an integral of graphs, we present a definition based on the properties of the usual mean as suggested by Frechet [7]. The basic idea of Frechet to define central points in a metric space is essentially the same as for the concept of graph mean and median proposed by [4,6,12,15,17].

Suppose that (\mathcal{Q}, d) is a metric orbifold of graphs with $\mathcal{Q} = (\mathcal{X}, \Gamma, \pi)$. We define the *Frechet function* as

$$F(Y) = \mathbb{E}_X [d(X, Y)^2] = \begin{cases} \int_{\mathcal{C}_{\mathcal{Q}}} d(X, Y)^2 dP_{\mathcal{Q}}(X) & : X \text{ is continuous} \\ \sum_{X \in \mathcal{C}_{\mathcal{Q}}} d(X, Y)^2 P_{\mathcal{Q}}(X) & : X \text{ is discrete} \end{cases},$$

where $P_{\mathcal{Q}}$ is a probability measure on the Borel sigma-field of $\mathcal{X}_{\mathcal{Q}}$ with support on a measurable subset $\mathcal{C}_{\mathcal{Q}}$ in the continuous case and a probability mass function in the discrete case. A *Frechet mean* is any element $M \in \mathcal{C}_{\mathcal{Q}}$ satisfying

$$F(M) = \inf_{Y \in \mathcal{C}_{\mathcal{Q}}} F(Y) < \infty.$$

The *Frechet mean set* \mathcal{F} is the set of all Frechet means.

3.2 Characterization of Frechet Means

For characterizing a Frechet mean, we need the notion of Dirichlet fundamental domain. A *fundamental domain* of Γ in \mathcal{X} is a closed subset $\mathcal{D} \subset \mathcal{X}$ with

$$\mathcal{X} = \bigcup_{\gamma \in \Gamma} \gamma(\mathcal{D})$$

and $\text{int}(\gamma(\mathcal{D})) \cap \text{int}(\gamma'(\mathcal{D})) = \emptyset$ for all $\gamma, \gamma' \in \Gamma$. Thus, the interior of a fundamental domain projects to the entire domain of graphs, where interior points of the fundamental domain are unique vector representations. A *Dirichlet fundamental domain* of $\mathbf{x} \in \mathcal{X}$ is a fundamental domain satisfying

$$\mathbf{x}' \in \mathcal{D}(\mathbf{x}) \Rightarrow \|\mathbf{x} - \mathbf{x}'\| \leq \|\mathbf{x} - \gamma(\mathbf{x}')\| \quad \forall \gamma \in \Gamma.$$

Each Dirichlet fundamental domain $\mathcal{D}(\mathbf{x})$ is a convex polyhedral cone containing at least one vector representation of each graph. Two vector representations in $\mathcal{D}(\mathbf{x})$ projecting to the same graph always lie on the boundary of $\mathcal{D}(\mathbf{x})$. Since the boundary is of Lebesgue measure zero, we can regard the domain of graphs as being geometrically a polyhedral cone.

Theorem 1 shows that any vector representation \mathbf{m} of a Frechet mean M is a population mean of the lifted probability distribution on its Dirichlet fundamental domain $\mathcal{D}(\mathbf{m})$.

Theorem 1 (Representation of a Frechet Mean). *Let $\mathcal{Q} = (\mathcal{X}, \Gamma, \pi)$ be an orbifold of graphs with intrinsic metric d and let $(\mathcal{X}_{\mathcal{Q}}, \Sigma_{\mathcal{Q}}, P_{\mathcal{Q}})$ be a probability space. Suppose that $M \in \mathcal{F}$ is a Frechet mean of $P_{\mathcal{Q}}$. Then any vector representation $\mathbf{m} \in \mathcal{X}$ that projects to M is of the form*

$$\mathbf{m} = \int_{\mathcal{D}(\mathbf{m})} \mathbf{x} \, dP(\mathbf{x}),$$

where $P(\mathbf{x}) = P_{\mathcal{Q}}(\pi(\mathbf{x}))$ on $\mathcal{D}(\mathbf{m})$.

From the definition of the Dirichlet fundamental domain follows that a vector representation \mathbf{m} of a Frechet mean M is the population mean of the distribution on all vector representations \mathbf{x} of X optimally aligned with \mathbf{m} .

3.3 Uniqueness of Frechet Mean

Next, we show under which assumptions the Frechet mean of graphs consists of a singleton. For this, we define the *injectivity radius* of a structure $X \in \mathcal{X}_{\mathcal{Q}}$ by

$$r_X = \min \{ \|\mathbf{x} - \mathbf{x}'\| : \mathbf{x}' \in \text{bd}(\mathcal{D}(\mathbf{x})) \},$$

where $\mathbf{x} \in X$ is a vector representation. The injectivity radius r_X measures the shortest distance from \mathbf{x} to the boundary of its Dirichlet fundamental region $\mathcal{D}(\mathbf{x})$. The injectivity radius r_X is independent of the choice of vector representation ([16], Prop. 5). Thus, r_X is well-defined. The *injectivity angle* of X is

$$\alpha_X = \arcsin \frac{r_X}{l(X)}.$$

The injectivity angle is the smallest angle between a fixed vector representation \mathbf{x} of X and a vector lying on a boundary of $\mathcal{D}(\mathbf{x})$. By definition, the injectivity angle is independent of the choice of vector representation of X .

The Frechet mean consists of a singleton, if the graphs are distributed within a circular right cone with sufficient narrow opening angle.

Theorem 2 (Uniqueness of Frechet Mean). *Let $\mathcal{Q} = (\mathcal{X}, \Gamma, \pi)$ be an orbifold of graphs with intrinsic metric d and let $(\mathcal{X}_{\mathcal{Q}}, \Sigma_{\mathcal{Q}}, P_{\mathcal{Q}})$ be a probability space. Suppose there is a structure $Z \in \mathcal{X}_{\mathcal{Q}}$ such that the support of $P_{\mathcal{Q}}$ is a measurable subset of the open circular cone*

$$\mathcal{C}_{\Gamma} \left(Z, \frac{\alpha Z}{3} \right) = \left\{ X \in \mathcal{X}_{\mathcal{Q}} : \angle(Z, X) < \frac{\alpha Z}{3} \right\}$$

with cone axis in direction of Z . Then the Frechet mean of $P_{\mathcal{Q}}$ is unique.

Suppose that \mathbf{z} projects to Z and $\mathcal{C} \subseteq \mathcal{D}(\mathbf{z})$ projects to $\mathcal{C}_{\Gamma}(Z, \alpha Z/3)$. Then from the proof of Theorem 2 follows that the elements of \mathcal{C} are pairwise optimally aligned. Hence, we may identify open sets of \mathcal{C}_{Γ} with open set of \mathcal{C} and apply any mathematical result that holds locally in a Euclidean space. In particular, we directly obtain as a Corollary the Law of Large Numbers and the Central Limit Theorem.

3.4 A Strong Law of Large Numbers

Since the distribution $P_{\mathcal{Q}}$ is usually unknown and the underlying metric space often lacks sufficient mathematical structure, the Frechet function $F(Y)$ can neither be computed nor be minimized directly. Instead, we estimate a Frechet mean from empirical data. Suppose that $X_1, X_2, \dots, X_N \in \mathcal{X}_{\mathcal{T}}$ is an independent and identically distributed random sample. We replace the Frechet function by the *empirical Frechet function*

$$\hat{F}_N(Y) = \frac{1}{N} \sum_{i=1}^N d(X_i, Y)^2$$

and approximate a Frechet mean by a global minimum of the empirical Frechet function. By $\hat{\mathcal{F}}_N$ we denote the set of *Frechet sample means* consisting of all global minima of $\hat{F}_N(Y)$.

As shown in [15], the strong law of large numbers for a distribution on graphs can be directly derived from [1].

Theorem 3. *Let $\mathcal{Q} = (\mathcal{X}, \Gamma, \pi)$ an orbifold of graphs with intrinsic metric d and let $(\mathcal{X}_{\mathcal{Q}}, \Sigma_{\mathcal{Q}}, P_{\mathcal{Q}})$ be a probability space. Suppose that the Frechet function F of $P_{\mathcal{Q}}$ is finite. Then for any $\varepsilon > 0$, there is a random variable $n(\omega, \varepsilon) \in \mathbb{N}$ and a $P_{\mathcal{Q}}$ -null set $\mathcal{N}(\omega, \varepsilon)$ such that*

$$\hat{\mathcal{F}}_N \subseteq \mathcal{F}_{\varepsilon} = \left\{ X \in \mathcal{X}_{\mathcal{T}} : \min_{M \in \mathcal{F}} d(X, M)^2 < \varepsilon \right\}$$

outside of $\mathcal{N}(\omega, \varepsilon)$ for all $N \geq n(\omega, \varepsilon)$. In particular, if the set $\mathcal{F} = \{\mu\}$ of Frechet means consists of a singleton μ , then every measurable selection, $\hat{\mu}_N$ from $\hat{\mathcal{F}}_N$ is a strongly consistent estimator of μ .

3.5 A Central Limit Theorem

Next, we want to derive a version of the central limit theorem for graphs. For this we introduce the following notations. The expression $X_N = o(Y_N)$ means that $X_N/Y_N \xrightarrow{P_Q} 0$. In particular $X_N = o(Y_N)$ means that $X_N \xrightarrow{P_Q} 0$.

Theorem 4. *Let $\mathcal{Q} = (\mathcal{X}, \Gamma, \pi)$ an orbifold of graphs with intrinsic metric d and let $(\mathcal{X}_Q, \Sigma_Q, P_Q)$ be a probability space. If*

$$\hat{F}_N(\hat{M}_N) \leq \inf_{Y \in \mathcal{X}_Q} F(Y) + o_{P_Q} \left(\frac{1}{N} \right)$$

and

$$\hat{M}_N \xrightarrow{P_Q} M \in \mathcal{F},$$

where M is nonsingular. Then

$$\sqrt{N} \left(\hat{M}_N - M \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{x}_i - \mathbf{m}) - o_{P_Q} (1),$$

where \mathbf{m} projects to M and $(\mathbf{x}_i, \mathbf{m}) \in \mathcal{A}(X_i, M)$ are optimal alignments. In particular, the sequence is $\sqrt{N} \left(\hat{M}_N - M \right)$ is asymptotically normal with mean zero and covariance matrix

$$\Sigma = \int_{\mathcal{D}(\mathbf{m})} (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^\top dP(\mathbf{x}).$$

3.6 The Frechet Mean of Two Structures

As a special case, we consider the Frechet function of the form

$$F(Y) = p \cdot d(X, Y)^2 + (1 - p) \cdot d(X', Y)^2,$$

where $p = P_Q(X)$ and $1 - p = P_Q(X')$ are the probabilities of the structures X and X' , respectively. Regarding p and $1 - p$ as weights rather than probabilities, the Frechet mean set of $F(Y)$ is an elementary component of competitive learning methods for central clustering and graph quantization [13].

Theorem 5. *Let $\mathcal{Q} = (\mathcal{X}, \Gamma, \pi)$ be an orbifold of graphs with intrinsic metric $d(\cdot|\cdot)$. Consider the Frechet function*

$$F(Y) = p \cdot d(X, Y)^2 + (1 - p) \cdot d(X', Y)^2,$$

where $X, X' \in \mathcal{X}_Q$ and $p \in [0, 1]$. Then the following holds:

(P1) Any vector representation \mathbf{m} of a Frechet mean $M \in \mathcal{F}$ is of the form

$$\mathbf{m} = p \cdot \mathbf{x} + (1 - p) \cdot \mathbf{x}',$$

where $(\mathbf{x}, \mathbf{m}) \in \mathcal{A}(X, M)$ and $(\mathbf{x}', \mathbf{m}) \in \mathcal{A}(X', M)$ are optimal alignments.

(P₂) $M \in \mathcal{F}$ is a Frechet mean if and only if any vector representation $\mathbf{m} \in M$ is of the form

$$\mathbf{m} = p \cdot \mathbf{x} + (1 - p) \cdot \mathbf{x}',$$

where $(\mathbf{x}, \mathbf{x}') \in \mathcal{A}(X, X')$ is an optimal alignment.

(P₃) A Frechet mean $M \in \mathcal{F}$ satisfies

$$d(X, M) = p \cdot d(X, X')$$

$$d(X', M) = (1 - p) \cdot d(X, X').$$

(P₄) $|\mathcal{F}| = 1$ with probability one.

Property (P₁) is a direct consequence of Theorem 1 and restated for sake of completeness. Property (P₂) states that the problem of determining an element of the Frechet mean of two structures is equivalent to finding an optimal alignment of X and X' . Properties (P₁) and (P₂) tell us how to construct a weighted mean. Property (P₃) shows that a Frechet mean of two structures is a weighted mean and therefore justifies, for example, the stochastic update rule of central clustering. The last property asserts that the Frechet mean consists of a singleton almost surely. At first this result may seem a useful achievement for a practical setting. A closer look at the proof, however, reveals that for most application problems, the given graphs lie in the set of Lebesgue measure zero for which no statement about uniqueness and non-uniqueness is given.

4 Conclusion

This contribution focused on large sample statistics for distribution on graphs. We derived structural versions of the two key results from probability theory, the law of large numbers and the central limit theory. In addition, we presented a characterization of Frechet means, sufficient conditions for uniqueness of the Frechet mean, and properties of the weighted Frechet mean of two graphs. The key idea to derive the proposed results is based on identifying graphs as points in some Riemannian orbifold. The results generalize corresponding results in Euclidean spaces. In addition, this work establishes a sound statistical basis for unsupervised structural pattern recognition methods such as PCA for structures, central clustering, and graph quantization. Furthermore, we gain new insight into the geometry of the graph domain, which in turn guides us to derive results for statistical and structural pattern analysis of graphs.

References

1. Bhattacharya, R., Bhattacharya, A.: Statistics on Manifolds with Applications to Shape Spaces. In: Perspectives in Mathematical Sciences, ISI, Bangalore (2008)
2. Caetano, T.S., et al.: Learning graph matching. In: ICCV 2007 Conf. Proc., pp. 1–8 (2007)
3. Cour, T., et al.: Balanced graph matching. In: NIPS 2006 Conf. Proc. (2006)

4. Ferrer, M.: Theory and algorithms on the median graph. application to graph-based classification and clustering, PhD Thesis, Univ. Aut'onoma de Barcelona (2007)
5. Ferrer, M., et al.: Graph-Based k-Means Clustering: A Comparison of the Set Median versus the Generalized Median Graph. In: Jiang, X., Petkov, N. (eds.) CAIP 2009. LNCS, vol. 5702, pp. 342–350. Springer, Heidelberg (2009)
6. Ferrer, M., Valveny, E., Serratos, F.: Median graphs: A genetic approach based on new theoretical properties. *Pattern Recognition* 42(9), 2003–2012 (2009)
7. Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré* 10(3), 215–310 (1948)
8. Gold, S., Rangarajan, A.: Graduated Assignment Algorithm for Graph Matching. *IEEE Trans. PAMI* 18, 377–388 (1996)
9. Gold, S., et al.: Learning with preknowledge: clustering with point and graph matching distance measures. *Neural Comp.* 8(4), 787–804 (1996)
10. Günter, S., Bunke, H.: Self-organizing map for clustering in the graph domain. *Pattern Recognition Letters* 23(4), 405–417 (2002)
11. Jain, B., Wyszotzki, F.: Central Clustering of Attributed Graphs. *Machine Learning* 56, 169–207 (2004)
12. Jain, B., Obermayer, K.: On the sample mean of graphs. In: *IJCNN 2008 Conf. Proc.*, pp. 993–1000 (2008)
13. Jain, B., Obermayer, K.: Graph Quantization, arXiv:1001.0921v1 [cs.AI] (2009)
14. Jain, B., et al.: Multiple alignment of contact maps. In: *IJCNN 2009 Conf. Proc.* (2009)
15. Jain, B., Obermayer, K.: Consistent Estimators of Median and Mean Graph. In: *ICPR 2010 Conf. Proc.* (2010)
16. Jain, B., Obermayer, K.: Supplementary material for the paper Large Sample Statistics in the Domain of Graphs (2010), <http://user.cs.tu-berlin.de/~jbj/publication.html>
17. Jiang, X., Munger, A., Bunke, H.: On Median Graphs: Properties, Algorithms, and Applications. *IEEE Trans. PAMI* 23(10), 1144–1151 (2001)
18. Mukherjee, L., et al.: Generalized median graphs and applications. *Journal of Combinatorial Optimization* 17, 21–44 (2009)
19. Schenker, A., et al.: Clustering of web documents using a graph model. In: *Web Document Analysis: Challenges and Opportunities*, pp. 1–16 (2003)
20. Umeyama, S.: An eigendecomposition approach to weighted graph matching problems. *IEEE Trans. PAMI* 10(5), 695–703 (1988)
21. van der Vaart, A.W.: *Asymptotic Statistics*. Cambridge University Press, Cambridge (2000)