MOX-Report No. 14/2020

Populations of Unlabeled Networks: Graph Space Geometry and Geodesic Principal Components

Calissano, A.; Feragen, A; Vantini, S.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Populations of Unlabeled Networks: Graph Space Geometry and Geodesic Principal Components

Anna Calissano
MOX- Dept. of Mathematics, Politecnico di Milano
and
Aasa Feragen
DTU Compute, Technical University of Denmark
Dept. of Computer Science, University of Copenhagen
and
Simone Vantini
MOX- Dept. of Mathematics, Politecnico di Milano

February 19, 2020

Abstract

Statistical analysis for populations of networks is widely applicable, but challenging as networks have strongly non-Euclidean behavior. Graph Space is an exhaustive framework for studying populations of networks which are weighted or unweighted, uni- or multi-layer, directed or undirected, labelled or unlabelled. Viewing Graph Space as the quotient of a Euclidean space with respect to a finite group action, we show that it is not a manifold, and that its curvature is unbounded from above. Motivated by these geometric properties, we define geodesic principal components, and we introduce the *Align All and Compute* algorithm, which allows the computation of statistics on Graph Space. The statistics and algorithm are empirically validated on one simulated study and two real datasets, showcasing the framework's potential utility. The whole framework is implemented in a publically available *GraphSpace* Python package.

*Keywords:*Network-valued/Graph-valued data, Computational Geometric Statistics, Quotient Space, Graph Space, Geodesic Principal Components

1 Introduction

In recent years, a soaring attention in the scientific community has been devoted to the analysis of tree- and network data. In many different fields, graph representation of the phenomenon has proven to be very useful. Several examples are available in medicine, social science, chemistry, finance, and many other fields. The scientific literature has so far primarily focused on analyzing networks in a *first generation* (Wang and Marron, 2007) setting, i.e. the analysis and modelling of a single network datum. The *second generation* approach refers to the analysis of a *population* of network-valued data. The need for second generation modelling arises in a series of applied problems, including analysis of brain connectivity (Simpson et al., 2013; Durante et al., 2017), anatomical trees (Feragen et al., 2013; Wang and Marron, 2007), and mobility networks (von Ferber et al., 2009). In these works, the aim is to analyse not just one network, but a set, or population, of networks. If every network has the same nodes, the problem could be re-framed in the analysis of a set of adjacency matrices. However, if the networks describe the same phenomenon, but have different numbers of nodes and different node labels, the mathematical challenges in the analysis become numerous and different.

Populations of tree- or network data have been studied under different perspectives: Nonlinear data spaces such as the Billera-Holmes-Vogtmann space of leaf-labeled trees (Billera et al., 2001), or spaces of unlabeled trees (Feragen et al., 2010, 2013) are examples of non-Euclidean spaces in which tree- and network data can be analysed with geometric tools. Euclidean embeddings form a separate analysis strategy, where trees or networks are embedded in Euclidean space through different approaches such as kernel methods (Shervashidze et al., 2011), convolutional neural networks (Duvenaud et al., 2015), or feature selection algorithms (Bunke and Riesen, 2011). If one is interested in ”simple” tasks such as prediction of class labels, cluster memberships, or a scalar variable, Euclidean Graph Embedding is often a scalable and powerful choice. However,

Euclidean Graph Embedding methods do not ensure that every point in the embedding space is actually a network, and probabilistic models therefore typically assign nonzero probability to points that do not represent networks, causing problems in extending statistical tools to these embedding spaces. Particularly challenging are problems whose answer is a network, such as averages, interpolation, denoising, network-valued regression models, etc. As a response to these limitations, Object Oriented Data Analysis (Marron and Alonso, 2014) aims to model "objects" – in our case, networks – as residing in a space of precisely such objects, ensuring that every point in the embedding space is a meaningful object. However, most existing works in this category have so far focused on tree-valued data rather than networks in general. Among existing models, Ginestet et al. (2017) proposes a model where networks' Laplacian matrices are smoothly injected into a sub-manifold of a Euclidean space; Simpson et al. (2013) and Durante et al. (2017) face the problem of generating and performing test on a population of networks; and Chowdhury and Mémoli (2017) and Chowdhury and Mémoli (2018) study a metric space of networks up to weak isomorphism, which allows the grouping of similar nodes. As a precursor to our work, Jain and Obermayer (2009) introduced an interesting and flexible general space of "Structures", which we call, when restricted to the special case of graphs, "Graph Space". This is a natural space for graphs with different or equal number of nodes, and with labelled or unlabelled nodes. Within Graph Space, networks can also be weighted or unweighted, uni- and multi-layer, directed and undirected, and with different types of attributes on both edges and nodes. The same Graph Space independently appears in two recent preprints, namely Kolaczyk et al. (2017), which studies the behavior of Fréchet mean in Graph Space, as well as in Guo et al. (2019), which proposes a simpler algorithm for principal components, analogous to the tangent space approaches known from manifold statistics (Fletcher and Joshi, 2004). An early version of our approach, which is analogous to intrinsic geodesic principal components from manifold statistics (Huckemann et al., 2010), was presented in the conference proceedings Calissano et al. (2019).

This paper provides a detailed study of the nonlinear nature of the geometry of the Graph Space, showing that even if it is not a manifold, it is a geodesic space with several properties. We prove that its curvature is unbounded from above, which renders standard approaches to nonlinear statistics unpractical, at least in terms of proving their convergence. We propose statistical algorithms that are guaranteed to converge to local optima, and discuss, in particular, the characteristics of the Fréchet Mean computed with this algorithm, compared to other common approaches. We introduce an intrinsic version of geodesic principal components for a population of unlabelled networks, and motivated by our proven geometric properties, propose the *Align All and Compute* (AAC) Algorithm, inspired by Generalized Procrustes Analysis (Gower, 1975).

To demonstrate the potential utility and flexibility of this framework, we illustrate our algorithms using three different datasets, showing our performance of the framework in both capturing topological and attribute variation. These datasets include a simulated dataset of undirected networks with scalar attributes on edges; a real, visually informative dataset of hand written letters (Kersting et al., 2016; Riesen and Bunke, 2008) with vector valued node attributes and scalar edge attributes; and a mobility dataset describing the mobility fluxes of citizens between provinces in Lombardy region (Italy) at different hours and for different modes of transportation.

The paper is organized as follows: In Section 2, we introduce Graph Space, and we describe its geometrical properties in Section 3. Section 4 presents the definition of geodesic principal components (GPCs) for network valued data and the introduction of the Align All and Compute Algorithm to estimate statistics on Graph Space such as Fréchet Mean and GPCs. The previous theoretical results are discussed in the three empirical examples on both simulated and real datasets in Section 5. All the proposed theory is implemented in the python package *GraphSpace* available on GitHub (Calissano et al., 2020).

2 Graph Space

We consider graphs as triples $G = (V, E, a)$ where the node set V has at most n elements, and the edge set $E \subset V^2$ has maximal size n^2. The nodes and edges are attributed with elements of an attribute space A, which in this paper is assumed to be Euclidean, via an attribute map $a\colon E \to A$. Here, the map a allows us to describe attributes on both edges and nodes, as we use self loop edges (diagonal elements) to assign attributes to nodes. From here on, we represent networks mathematically as graphs, and consider these terms equivalent.

In our modelling, we shall represent graphs with fewer nodes than n as having $n - |V|$ additional null nodes, allowing graphs to be represented via fixed-size adjacency matrices. More precisely, a graph with scalar attributes is completely specified by the adjacency matrix of dimension $n \times n$, residing in a space $X = \mathbb{R}^{n^2}$ of flattened adjacency matrices. If the attributes are vectors of dimension d, the graph is represented by a tensor of dimension $n \times n \times d$, residing in a space $X = \mathbb{R}^{n \times n \times d}$.

In many real world applications, populations of graphs describe the same phenomenon in different contexts (e.g. routes of different airline companies, or brain connectivity networks of different patients). Different nodes' labels or order make it challenging to investigate similarities between the topology and attributes of different graphs, and this is often alleviated by explicit or implicit matching of graph nodes. When the graphs are represented as $n \times n$ adjacency matrices, matching two graphs corresponds to finding optimal permutations of their nodes. The group T of node permutations can be represented via permutation matrices, acting on X through matrix multiplication. The binary operation:

$$\cdot\colon T \times X \to X, (T, x) \mapsto Tx$$

thus defines an action of the group T on X. We call the obtained quotient space $X_T = X/T$ *Graph Space*, and each element of X/T is a graph G, represented as an equivalence class $[x] = Tx$ which

contains all the flattened adjacency matrices in X which can be obtained from x by permuting nodes. Note that this Graph Space is a special case of the A-attributed r-structures introduced by Jain and Obermayer (2009), which includes hyper-graphs and more general attribute types. The graphs considered in this paper can be considered A-attributed 2-structures with A Euclidean.

We illustrate the construction of Graph Space with a simple example:

Example 1. Consider the two weighted networks shown in Figure 1. To represent these as points in Graph Space: Add a fictional null node to the first graph; randomly enumerate the nodes; represent them in two weighted, symmetric adjacency matrices as shown in Figure 1. The adjacency matrices can be vectorized as a vector of dimension nine (e.g the first network

$$x_1 = \begin{pmatrix} 0 & 4 & 0 \\ 4 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad x_2 = \begin{pmatrix} 0 & 4 & 0 \\ 4 & 0 & 3 \\ 0 & 3 & 0 \end{pmatrix}$$

Figure 1: Example of two simple weighted undirected networks x_1 and x_2 and their matrix representations.

becomes $[0,4,0,4,0,0,0,0,0]$). The space of flattened adjacency matrices X is thus $\mathbb{R}^9$, and the two networks are represented by two points in $\mathbb{R}^9$. The permutation action can be represented as a binary 9×9 matrix: The permutation reorders the nodes and consequently the edges of the network by permuting the rows and columns of its adjacency matrix, and hence also the positions of the elements in the flattened matrix representation x_i. For example, if the nodes 2 and 3 are permuted in the first network, we obtain the new permuted vector $[0,0,4,0,0,0,4,0,0]$. Each point $[x_i]$ in the quotient space X/T consists of all possible permuted versions of x_i (i.e. permuting the rows and columns of the associated adjacency matrix). Now, the equivalence classes $[x_1]$ and $[x_2]$ are points in X/T, and the maximal size of an equivalence class in X/T is 3!.

Remark 1. In practice, the maximum number of nodes n can limit the possible node matchings, and hence also the geometric relationship, between two graphs. To make sure no such limitations are present, set $n = 2\max(|n_i|)$.

For symmetric adjacency matrices it would suffice to represent them with the upper triangular part. For the sake of generalization we keep the more complete representation.

3 On the geometry of Graph Space

As the Graph Space X/T is the quotient of the total space X of flattened adjacency matrices with respect to the node permutation group T, any metric d_X on X defines a quotient pseudometric

$$d_{X/T}([x_1],[x_2]) = \min_{t\in T} d_X(tx_1, x_2)$$

on X/T which, since the permutation group T is finite, is indeed a metric. Examples of commonly used metrics on X include the l_p metrics for $p \in (0,\infty]$, where $p = 2$ gives the Euclidean distance used in this paper, and $p = 1$ gives the Manhattan distance.

Remark 2. The l_1 distance on X generates the very well known graph edit distance (Sanfeliu and Fu, 1983) on X/T, under the assumption that n is sufficiently large to allow all possible edit paths between the graphs in the dataset. Note that in the graph edit distance, geodesics are not unique, as edit operations can be made in different order with no effect on the total edit cost. Generalizing manifold statistics based on geodesics to graphs based on graph edit distance is therefore futile.

While Graph Space X/T is a metric space, it is not a manifold, even with the Euclidean distance on X. This follows from the fact that the structure of the isotropy subgroup $T_x = \{t \in T | tx = x\}$ varies for different points $x \in X$ (Bredon, 1972). One reason why this happens is that, as explained in the previous section, forcing the networks to all have the same number of

nodes generates networks with a subset of null nodes. As a permutation $t \in T$ which acts only on the null nodes of a flattened adjacency matrix $x \in X$ does not have any effect on the adjacency matrix, such x will have a larger isotropy subgroup than generic points in X. As a consequence, the well-known tools from manifold statistics are unfortunately not applicable to Graph Space.

As is commonly done in *manifold statistics* (Fletcher and Joshi, 2004; Huckemann et al., 2010; Kendall, 1984; Srivastava et al., 2005; Zhang and Fletcher, 2013; Pennec et al., 2006; Fletcher, 2013; Mallasto and Feragen, 2018) as well as in more general nonlinear statistics (Turner et al., 2014; Nye, 2011, 2014; Duncan et al., 2018; Feragen and Nye, 2020; Bačák, 2014; Nye et al., 2017; Miller et al., 2015; Sturm, 2003; Feragen et al., 2013, 2011), we will utilize *geodesics*, or shortest paths, to define and compute statistical properties in Graph Space. In the absence of a manifold structure, we will define and understand geodesics, statistical properties built on geodesics, and their properties, by utilizing geometric constructions from *metric geometry* (Bridson and Haefliger, 1999). To that end, we dedicate this section to surveying necessary concepts from metric geometry and applying them to uncover geometric properties of Graph Space.

3.1 Basic geometric properties of Graph Space: Distances and geodesics

Given a general metric space $(\mathfrak{X}, d_{\mathfrak{X}})$, the *length* of a path $\gamma\colon [0,1] \to \mathfrak{X}$ is given by

$$l(\gamma) = \sup_{x_0=\gamma(0), x_1, \ldots, x_m=\gamma(1)} \sum_{i=1}^{m} d_X(x_{i-1}, x_i), \tag{1}$$

where the supremum is taken over all approximations $x_0 = \gamma(0), x_1 = \gamma(t_1), \ldots, x_{m-1} = \gamma(t_{m-1}), x_m = \gamma(1)$ of γ of some finite length m, where $0 < t_1 < \ldots < t_{m-1} < 1$. Thus, the length of a path can be thought of as the supremum over lengths of all finite approximations of the path.

Given two points $a, b \in \mathfrak{X}$, a *geodesic* from a to b is a path $\gamma\colon [0,1] \to \mathfrak{X}$ such that $\gamma(0) = a$, $\gamma(1) = b$ and $l(\gamma) = d_{\mathfrak{X}}(a,b)$. The metric space $(\mathfrak{X}, d_{\mathfrak{X}})$ is said to be a *length space* if, for every

two points $a, b \in \mathfrak{X}$, we have

$$d_{\mathfrak{X}}(a,b) = \inf\{l(\gamma) \mid \gamma\colon [0,1] \to \mathfrak{X} \text{ s.t. } \gamma(0) = a, \gamma(1) = b\}.$$

That is, the distance between any two points a, b is the infimum over lengths of paths connecting them. Moreover, $\mathfrak{X}$ is a *geodesic space* if every two points $a, b \in \mathfrak{X}$ are connected by a geodesic γ from a to b – that is, there actually exists a path attaining the infimum length.

Lemma 1. *Graph Space is a geodesic space.*

Proof. This result follows from standard properties of metric spaces. As our total space X is Euclidean, it is in particular a length space. Since Graph Space X/T is a metric space, X/T is a length space by (Bridson and Haefliger, 1999, Lemma I.5.20). Moreover, as X/T is the quotient with respect to a finite group, and X is locally compact, the quotient X/T is also locally compact (Bredon, 1972, Theorem I.3.1). Note that any Cauchy sequence $([x_i])_{i\in\mathbb{N}}$ in X/T is the image under π of a Cauchy sequence $(t_i x_i)_{i\in\mathbb{N}}$ in X such that for some $M \in \mathbb{N}$, and for $i, j \geq M$, we have $d_X(t_i x_i, t_j x_j) = d_{X/T}([x_i], [x_j]$. Since X is complete, the sequence $(t_i x_i)_{i\in\mathbb{N}}$ converges to some point $x \in X$, and hence the sequence $([x_i])_{i\in\mathbb{N}}$ converges to $[x] \in X/T$. In other words, X/T is complete. Thus, X/T is a geodesic space by the Hopf-Rinow theorem (Bridson and Haefliger, 1999, Proposition I.3.7). □

Lemma 2. *There exist points $y_\varepsilon, z_\varepsilon \in X/T$ which are connected by more than one geodesic.*

Proof. We give an example in the case where node- and edge attributes are scalar. Note that the same example can be adapted to vector valued node- or edge attributes by appending this scalar value with zeros. Let $\varepsilon > 0$, and consider the two graphs y_ε and z_ε shown in the top row on the left hand side of Figure 2. There are two geodesic paths between these two graphs. The first consists of interpolating the node attributes with the node matching indicated by the planar embedding

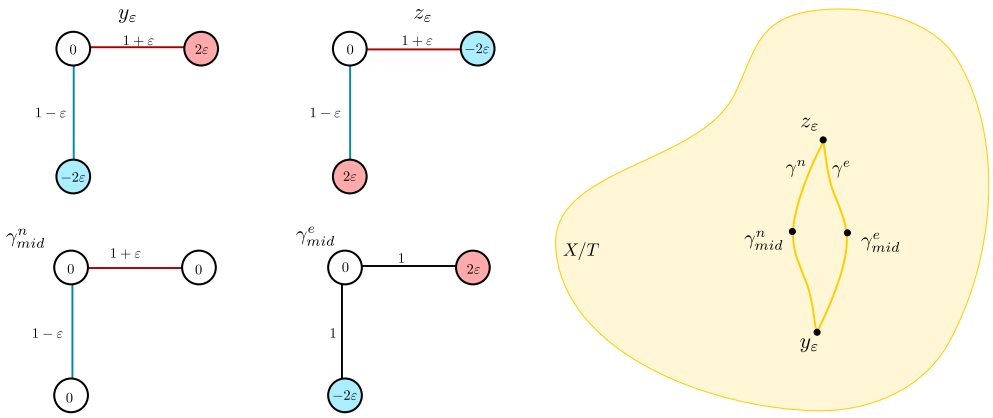


Figure 2: Depending on their node alignment, the two graphs y_ε and z_ε differ either in node attributes or in edge attributes, and there exist two different geodesics γ^n and γ^e connecting the two graphs: One which interpolates node attributes and one which interpolates edge attributes, respectively. These two geodesics are illustrated via their midpoints γ^n_{mid} and γ^e_{mid}.

of the nodes. The midpoint of this geodesic is the graph γ^n_{mid} shown on the left hand side of the bottom row. The second geodesic between y_ε and z_ε consists of interpolating the edge attributes with the node matching indicated by the coloring of the nodes. The midpoint of this geodesic is the graph γ^e_{mid} shown on the right hand side of the bottom row. The two points connected by two geodesics are illustrated schematically on the right hand side. □

Corollary 3. *Geodesics connecting pairs of points are not generally unique in X/T.* □

3.2 The curvature of Graph Space

Curvature affects properties which are important for defining and computing statistical quantities, most notably geodesics (shortest paths) and conditions for their uniqueness. As Graph Space is *not* a manifold, we will utilize more general concepts of curvature from metric geometry (Bridson and Haefliger, 1999).

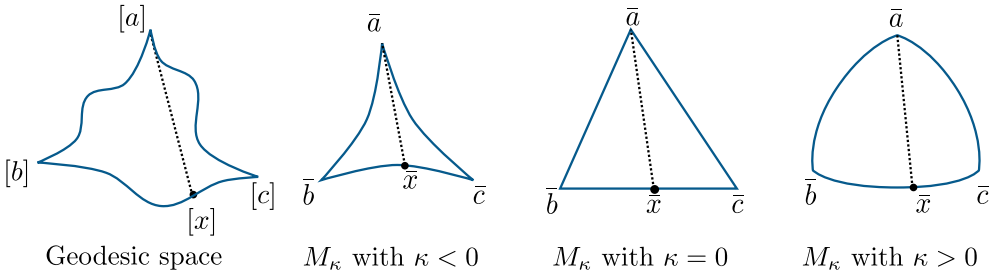


Figure 3: **Left:** A geodesic triangle in $\mathfrak{X}$. **Right:** the corresponding comparison triangles in hyperbolic space $\mathbb{H}^2$, the plane $\mathbb{R}^2$ and the sphere $\mathbb{S}^2$, respectively.

First consider the Graph Space where edges and nodes have real attributes $a\colon E \to \mathbb{R}$. We will show that the curvature of this Graph Space is not bounded from above, which affects our ability to find unique geodesics even locally. In metric geometry, curvature is approached through comparison with *model spaces* M_κ of curvature κ. Different model spaces are used for negative, zero and positive κ. When $\kappa < 0$, the model space is the hyperbolic space of negative curvature κ, namely $M_\kappa = \mathbb{H}_\kappa$. For $\kappa = 0$, the model space is $M_0 = \mathbb{R}^2$, namely the Euclidean plane. Finally, for $\kappa > 0$, the model space is the sphere of curvature κ, namely $M_\kappa = \mathbb{S}^2_\kappa$. An important property of the model spaces of curvature κ is that they are each accompanied by an *diameter* D_κ, such that any two points $a, b \in B(x, D_\kappa)$ for any $x \in M_\kappa$ can be joined by a unique geodesic.

We can compare any given geodesic space $(\mathfrak{X}, d_\mathfrak{X})$ to any one of the model spaces using *comparison triangles*, as follows: A geodesic triangle abc in $\mathfrak{X}$ consists of vertices $a, b, c \in \mathfrak{X}$ joined by geodesic edges γ_{ab}, γ_{bc} and γ_{ac}. We assume that a, b and c are all contained in a ball of perimeter $< 2D_\kappa$. We can then construct a *comparison triangle* $\bar{a}\bar{b}\bar{c}$ in the model space M_κ with vertices $\bar{a}$, $\bar{b}$ and $\bar{c}$ joined by geodesic edges $\bar{\gamma}_{\bar{a}\bar{b}}$, $\bar{\gamma}_{\bar{b}\bar{c}}$ and $\bar{\gamma}_{\bar{a}\bar{c}}$, whose lengths are the same as the lengths of the edges γ_{ab}, γ_{bc} and γ_{ac} in abc. See Figure 3 for an illustration.

Definition 1 ($CAT(\kappa)$ **space, curvature in the sense of Alexandrov).** Let $(\mathfrak{X}, d_\mathfrak{X})$ be a geodesic metric space, and let abc be a geodesic triangle in $\mathfrak{X}$ as described above. Note that any point x

from the segment γ_{bc} has a corresponding point $\bar{x}$ on the segment $\bar{\gamma}_{\bar{b}\bar{c}}$ in the comparison triangle, such that $d_{M_\kappa}(\bar{x},\bar{b}) = d_{\mathfrak{X}}(x,b)$. If

$$d_{\mathfrak{X}}(x,a) \leq d_{M_\kappa}(\bar{x},\bar{a}) \tag{2}$$

for every such x, and similarly for any x on γ_{ab} or γ_{ac}, then the geodesic triangle abc satisfies the *CAT*(κ) *condition*. The metric space $\mathfrak{X}$ is a *CAT*(κ) space if any geodesic triangle abc in $\mathfrak{X}$ of perimeter $< 2D_\kappa$ satisfies the *CAT*(κ) condition given in eq. 2. Geometrically, this means that triangles in $\mathfrak{X}$ are *thinner* than triangles in M_κ. The metric space $\mathfrak{X}$ has *curvature* $\leq \kappa$ *in the sense of Alexandrov* if it is locally *CAT*(κ).

Note that the properties of the model space in relation to its diameter D_κ also transfer to geodesic spaces $\mathfrak{X}$ which are *CAT*(κ) (Bridson and Haefliger, 1999, Proposition II.1.4): Given any $x \in \mathfrak{X}$, any two points $a,b \in B(x,D_\kappa) \subset \mathfrak{X}$ can be joined by a unique geodesic. In particular, $D_\kappa = \infty$ for $\kappa \leq 0$, meaning that in non-positively curved spaces, any two points can be joined by a unique geodesic, regardless of their distance. Moreover, $D_\kappa \geq \frac{\pi}{\sqrt{\kappa}}$ for $\kappa > 0$, meaning that the lower the bound on the (positive) curvature, the larger the radius within which all pairs of points have unique connecting geodesics.

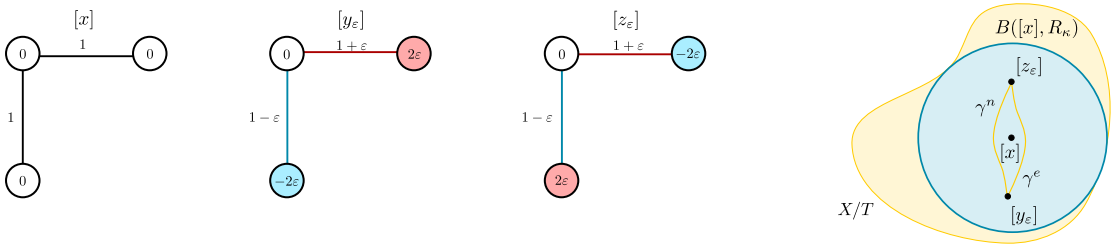


Figure 4: Consider the graph $[x] \in X/T$ shown on the left. A ball about $[x]$ of any radius $R_\kappa > 0$ will always contain two graphs $[y_\varepsilon]$ and $[z_\varepsilon]$, as shown second and third. As we saw in Figure 2, these two points are connected by two different geodesics.

Theorem 4 (The curvature of Graph Space is unbounded from above)**.** *Graph space does not have curvature* $\leq \kappa$ *in the sense of Alexandrov for any* $\kappa \in \mathbb{R}$*. In particular, for any graph* $[x] \in X/T$ *and any* $R_\kappa > 0$*, we can – by assuming a sufficiently high maximal number of nodes* n *– find two graphs* $[y_\varepsilon], [z_\varepsilon] \in B([x], R_\kappa) \subset X/T$ *which are connected by two geodesics.*

Proof. Let $[x], [y_\varepsilon]$ and $[z_\varepsilon]$ be as in Figure 4. Note that for any radius $R_\kappa > 0$, there will exist an $\varepsilon > 0$ such that $[y_\varepsilon], [z_\varepsilon] \in B([x], R_\kappa)$. Moreover, as argued in Lemma 3, there will always be two equally long shortest paths connecting $[y_\varepsilon]$ and $[z_\varepsilon]$. Thus, by (Bridson and Haefliger, 1999, Proposition II.1.4), Graph Space is not locally $CAT(\kappa)$ at $[x]$ for any κ, and thus cannot have curvature $\leq \kappa$ in the sense of Alexandrov for any κ.

To prove the final statement, consider any graph $[x]$ and assume that the maximal number n of nodes considered in a graph is sufficiently high to construct the graphs $[y_\varepsilon]$ and $[z_\varepsilon]$ shown in Figure 5, which can be constructed to both be arbitrarily close to $[x]$. Now, again, there exist two geodesics connecting $[y_\varepsilon]$ and $[z_\varepsilon]$. □

Remark 3. Note that while all of the results and examples above considered the case where node and edge attributes were real valued ($A = \mathbb{R}$), the proofs hold equally well for vector valued node and edge attributes ($A = \mathbb{R}^p$).

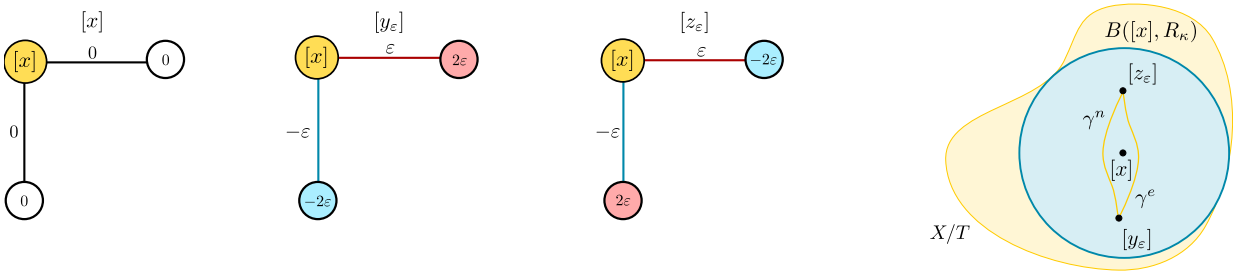


Figure 5: For any graph $[x]$, we can find two graphs $[y_\varepsilon]$ and $[z_\varepsilon]$ arbitrarily close to it, which are connected by more than one geodesic.

4 Statistics

As discussed in the previous section, the curvature of Graph Space is unbounded from above. The total space X is, on the other hand, Euclidean, and we will utilize this to compute statistics in Graph Space. In particular, we propose as a general strategy the *Align All and Compute* (AAC) algorithm, closely related to Generalized Procrustes Analysis (Gower, 1975), which consists of iteratively selecting an *optimal* representative from every equivalence class in Graph Space and computing the wanted statistic with those representatives on the total space. In this section, we first introduce the AAC algorithm in the context of estimating the Fréchet Mean. Next, we extend the concept of geodesic principal components (Huckemann et al., 2010) to Graph Space, and utilize the AAC algorithm to estimate the geodesic principal components.

4.1 The Fréchet Mean

Consider the total space X, the Graph Space X/T obtained by quotienting out the action of the permutation group T on X, and a set of observations $\{[x_1],\dots,[x_k]\} \in X/T$. A basic quantity in nonlinear statistics is the *Fréchet mean*:

Definition 2. The Fréchet mean of a sample $\{[x_1],\dots,[x_k]\} \in X/T$ is given

$$[\bar{x}] = \underset{[x]\in X/T}{\arg\min} \sum_{i=1}^{k} d_{X/T}([x],[x_i])^2. \tag{3}$$

We note that for any geodesic space, the mean of two points is characterized as the midpoint of any geodesic connecting the two points. Thus, as illustrated in Figure 6, we obtain as a direct consequence of Corollary 3 and Theorem 4:

Corollary 5. *Fréchet means are not generally unique in Graph Space X/T. In fact, as shown by Figure 5, for any graph $[x] \in X/T$ and any radius $\varepsilon > 0$, there will be sets of points in $B([x],\varepsilon)$ (e.g. $\{[y_\varepsilon],[z_\varepsilon]\}$) whose Fréchet mean is not unique.*

Even if Graph Space geodesics and means are not generally unique, we might hope for generic uniqueness similar to the results shown for tree spaces in Feragen and Nye (2020). However, even in cases where means are not unique, they are still useful, and their computation an important open problem. In nonlinear statistics, a number of algorithms exist to compute means either approximately (Sturm, 2003; Bačák, 2014; Miller et al., 2015; Afsari et al., 2013; Arnaudon et al., 2013; Arnaudon and Miclo, 2014; Bonnabel, 2013; Hauberg et al., 2015; Turner et al., 2014), or via heuristics (Feragen et al., 2011; Billera et al., 2001; Jain and Obermayer, 2008), whose applicability and efficiency vary with the complexity of the underlying nonlinear data space.

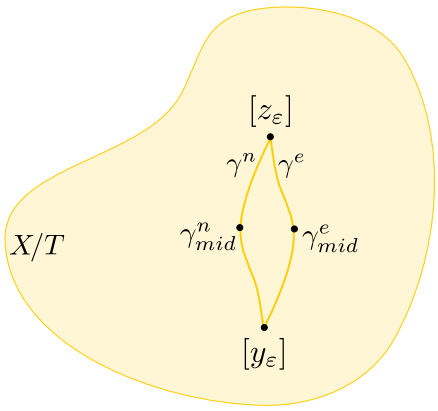


Figure 6: Since geodesics are not necessarily unique in Graph Space X/T, Fréchet means are not necessarily unique either.

A popular strategy for computing Fréchet means in geodesic spaces is the iterative "midpoint" algorithm which obtains an updated mean estimate by stepping $1/k$ along the geodesic from a current mean estimate to a k^{th} random sample from the dataset. In Euclidean space, this computes the mean in finite time when samples are made without replacement. The same "*without* replacement" strategy is applied for trees in Feragen et al. (2011) and for graphs in Jain and Obermayer (2008), but these finite time algorithms do not generally return the mean in tree- or Graph Space, and should be considered heuristics. When running the algorithm *with* replacement, there are a number of scenarios in which it is known to converge towards the Fréchet mean, including non-positively curved spaces (Sturm, 2003; Bačák, 2014; Miller et al., 2015) and certain Riemannian manifolds of bounded curvature (Chakraborty and Vemuri, 2015; Arnaudon et al., 2013; Arnaudon and Miclo, 2014). Note, in particular, that for Riemannian manifolds, this algorithm is a special case of stochastic gradient descent (Bonnabel, 2013). While this algorithm

is easy to generalize to Graph Space, its convergence proofs usually require bounded curvature to have some level of uniqueness for geodesics. This leads us to consider another strategy for computing means, which we call Aligned All and Compute (AAC).

We remind the reader that the metric on Graph Space is given by

$$d_{X/T}([x_1],[x_2]) = \min_{t\in T} d_X(tx_1, x_2),$$

which naturally leads to the concept of *optimal position* (Huckemann et al., 2010) of one point with respect to another.

Definition 3. Given $\tilde{x} \in X$ and $t \in T$, the point $t\tilde{x}$ is in *optimal position* with respect to $x \in X$ if

$$d_X(t\tilde{x}, x) = \min_{t'\in T} d_X(t't\tilde{x}, x) = d_{X/T}([\tilde{x}],[x]).$$

That is, the equivalence class $[\tilde{x}] \in X/T$ contains (at least) one point $t\tilde{x} \in [\tilde{x}]$ which has minimal distance to x, and this point is in optimal position with respect to x.

The distance $d_{X/T}$ between two points in the Graph Space X/T corresponds exactly to the distance in X after posing one point in optimal position with the other. We now define the AAC algorithm, which is based on iteratively first posing observations in optimal position with respect to the current mean estimate, then re-estimating the mean based on the aligned observations, and repeating until convergence. In the case of computing Fréchet means of shapes, the AAC algorithm coincides with generalized Procrustes analysis (Gower, 1975).

This algorithm provides an estimate of a Fréchet Mean which is independent of the order of the data, and with improved convergence properties:

Theorem 6. *AAC for Fréchet Mean (Algorithm 1) converges in finite time. Moreover, assume that Graph Space X/T is endowed with a probability measure η which is absolutely continuous with respect to the pushforward of the Lebesgue measure m on X. In particular, for $A \subset X/T$,*

Algorithm 1: AAC for the Fréchet Mean

Data: $\{[x_1],\dots,[x_k]\} \subset X/T$ observations in X; a threshold $\varepsilon > 0$

Result: An estimate of the Fréchet Mean of $\{[x_1],\dots,[x_k]\}$.

Initialization: Select randomly an observed graph and randomly a representative:

$\tilde{x} = \tilde{x}_i \in [x_i] \in \{[x_1],\dots,[x_k]\}$;

while $s > \varepsilon$ **do**

Put every observation in optimal position with the representative $\tilde{x}$, obtaining an aligned set of representatives $\{\tilde{x_1},\tilde{x_2},\dots,\tilde{x_k}\} \subset X$;

Compute the Fréchet Mean $\bar{x}$ in X of $\{\tilde{x_1},\tilde{x_2},\dots,\tilde{x_k}\}$;

Compute $s = d_X(\tilde{x},\bar{x})$;

Set $\tilde{x} = \bar{x}$;

Return $[\bar{x}]$

we have $\eta(A) = 0$ *if* $m(\pi^{-1}(A)) = 0$. *Let the dataset* $[x_1],\dots,[x_k]$ *is sampled from* η*; now with probability 1, the estimator found by Algorithm 1 is a local minimum of the Fréchet function*

$$\sum_{i=1}^{k} d^2_{X/T}([x],[x_i]). \tag{4}$$

Proof. See Appendix. □

While we show that it theoretically converges in finite time, this might still be a long time, and we thus also add a convergence threshold ε in Algorithm 1.

Note also that our algorithm in practice relies on inexact graph matching, as graph matching is generally NP complete. We thus cannot be sure that our computed means were, in fact, based on completely correct graph matching. However, alternative mean algorithms also rely on graph matching, and are therefore similarly affected.

4.2 Geodesic Principal Component Analysis

In this section we introduce Geodesic principal components (GPCs) in Graph Space, following Huckemann et al. (2010), and propose an AAC algorithm for their computation. Consider the *canonical projection* to the Graph Space X/T:

$$\pi\colon X \to X/T := \{[p] : p \in X\}.$$

Definition 4. Denote by $\Gamma(X)$ the set of all straight lines (geodesics) in X. Following Huckemann et al. (2010), a curve δ is a *generalized geodesic* on the Graph Space X/T, if it is a projection of a straight line on X:

$$\Gamma(X/T) = \{\delta = \pi \circ \gamma\colon \gamma \in \Gamma(X)\}. \tag{5}$$

Just like we defined optimal alignment of sample graph representatives with a fixed graph representative, we shall also utilize optimal alignment of sample graphs with a generalized geodesic:

Definition 5. Consider $[x] \in X/T$, $t \in T$ and δ a generalized geodesic in X/T with representative $\gamma \in \Gamma(X)$. The graph representative $tx \in X$ is in *optimal position* with respect to γ on X if

$$d_X(tx, \gamma) = d_{X/T}([x], \delta).$$

In the Graph Space, the optimal positioning is selecting the best representative $x \in [x]$ (i.e. the "best" node permutation) with respect to a point or a curve, in order to perform computation on the total space X. The optimal positioning of a point with respect to a generalized geodesic is a two step alignment procedure.

Algorithm 2: Algorithm for Optimal Position with respect to a generalized geodesic

Data: A point $x \in [x]$, a straight line $\gamma \in \Gamma(X)$, the domain $[s_{min}, s_{max}]$

Result: $t^* \in T$ such that t^*x is in optimal position wrt γ.

for $s \leftarrow s_{min}$ **to** s_{max} **do**
  Find $t(s) := \arg\min_{t \in T} d_X(tx, \gamma(s))$

Find $s^* = \arg\min_{s \in [s_{min}, s_{max}]} d_X(t(s)x, \gamma(s))$;

Return $t^* = t(s^*)$

The obtained $t^* \in T$ is the permutation such that the point $t^*x \in [x]$ is the closest representative of $[x]$ to the geodesic γ in the interval selected. Since Graph Space is not an inner product space, we consider two generalized geodesics to be orthogonal if they have orthogonal representatives in $\Gamma(X)$. We can define a set of geodesic principal components and a strategy to compute them based on the residuals.

Definition 6. Consider the canonical projection of the Graph Space $\pi\colon X \to X/T$ of X and consider a set $\{[x_1], \ldots, [x_k]\} \subset X/T$ of graphs, $[x] \in X/T$, and $\delta \in \Gamma(X/T)$. The Generalized Principal Components for the set $\{[x_1], \ldots, [x_k]\}$ are defined as:

- The *first geodesic principal component* $\delta_1 \in \Gamma(X/T)$ is the generalized geodesic minimizing the sum of squared residuals:

$$\delta_1 = \underset{\delta \in \Gamma(X/T)}{\arg\min} \sum_{i=1}^{k} (d^2_{X/T}([x_i], \delta)) \tag{6}$$

- The *second geodesic principal component* $\delta_2 \in \Gamma(X/T)$ minimizes (6) over all $\delta \in \Gamma(X/T)$, having at least one point in common with δ_1 and being orthogonal to δ_1 at all points in common with δ_1.

- The point $\mu \in X/T$ is called *Principal Component Mean* if it minimizes

$$\sum_{i=1}^{k} (d^2_{X/T}([x_i], [\mu])^2) \tag{7}$$

where $[\mu]$ only runs over points $\tilde{x}$ in common with δ_1 and δ_2.

- The *j^{th} geodesic principal component* is a $\delta_j \in \Gamma(X/T)$ if it minimizes (6) over all generalized geodesics that meet orthogonally $\delta_1, \ldots, \delta_{j-1}$ and cross μ.

Algorithm 3: AAC to compute the geodesic principal components

Data: $\{[x_1], \ldots, [x_k]\} \in X/T$ observations in X

Result: Geodesic principal components $\delta_1, \ldots, \delta_k \in \Gamma(X/T)$

Initialize:

Select randomly $\tilde{x}_i \in [x_i]$ $\{[x_1], \ldots, [x_k]\}$*;*

Align all the observations to the representative $\tilde{x}_i$*, obtaining a set of points* $\{\tilde{x_1}, \tilde{x_2}, \ldots, \tilde{x_k}\} \in X$ *in optimal position with respect to* $\tilde{x}_i$*;*

Perform PCA on $\{\tilde{x_1}, \tilde{x_2}, \ldots, \tilde{x_k}\}$ *in* X *obtaining* $\gamma_1, \ldots, \gamma_k \in \Gamma(X)$*;*

Project onto $\Gamma(X/T)$ *as* $\delta_i = \pi \circ \gamma_i$*;*

Set $\tilde{\delta}_1 = \delta_1, \ldots, \tilde{\delta}_k = \delta_k$

while $s > \varepsilon$ **do**

Align all the points $\{[x_1], [x_2], \ldots, [x_k]\}$ with respect to the generalized geodesic $\tilde{\delta}$, obtaining a new set of aligned points $\tilde{x_1}, \tilde{x_2}, \ldots, \tilde{x_k} \in X$ using Algorithm 2;

Perform PCA on $\{\tilde{x_1}, \tilde{x_2}, \ldots, \tilde{x_k}\}$ in X obtaining $\gamma_1, \ldots, \gamma_k \in \Gamma(X)$;

Project onto $\Gamma(X/T)$ as $\delta_i = \pi \circ \gamma_i$;

Compute a step distance function $s = f(\tilde{\delta}_i, \delta_i)$;

Set $\tilde{\delta}_1 = \delta_1, \ldots, \tilde{\delta}_k = \delta_k$.

Return $\delta_1, \ldots, \delta_k \in \Gamma(X/T)$

Note that, due to the curvature of the space discussed in Section 3, the Fréchet mean is not ensured to be the same as the Principal Component Mean. Having equipped Graph Space with all

the required definitions, we propose the AAC algorithm to compute the geodesic principal components (see Algorithm 3). A possible choice for the step distance function f is the proportion of variance explained by the first geodesic principal component at the current and the previous step.

Again, the AAC algorithm converges in finite time, and in the case of the first GPC we can show that it converges to a local minimizer of the sum of squared residuals function:

Theorem 7. *AAC for GPCA (Algorithm 3) converges in finite time. Assume that Graph Space X/T is endowed with a probability measure η which is absolutely continuous with respect to the pushforward of the Lebesgue measure m on X, and let the dataset $[x_1], \ldots, [x_k]$ be sampled from η. Now with probability 1, the estimator of the first GPC found by Algorithm 1 is a local minimum of the sum of squared residuals function*

$$\sum_{i=1}^{k} d^2_{X/T}(\delta, [x_i]). \tag{8}$$

where $\delta \in \Gamma(X/T)$.

Proof. See Appendix. □

For the case of the higher GPCs, we do not have a proof of local minimization of (6) and Algorithm 3 should be considered a heuristic.

5 Experiments on real and simulated data

In this section, we illustrate the introduced Graph Space statistics via three case studies emphasizing the framework's flexibility to model different graph features such as directed and undirected edges, or scalar and vector attributes, on both nodes and edges. Each computation of distances and geodesics require graph matching, which is an NP-complete problem with many existing

heuristics for inexact matching (Emmert-Streib et al., 2016; Conte et al., 2004). In the simulations we use the graduate assignment algorithm (Gold and Rangarajan, 1996) for inexact matching. All experiments are conduced using our *GraphSpace* python package, which is available on github (Calissano et al., 2020). Implementation details are listed in the package description. The convergence threshold ε used in the AAC algorithms is set to $\varepsilon = 0.001$ in Algorithm 1 and $\varepsilon = 0.01$ in Algorithm 3.

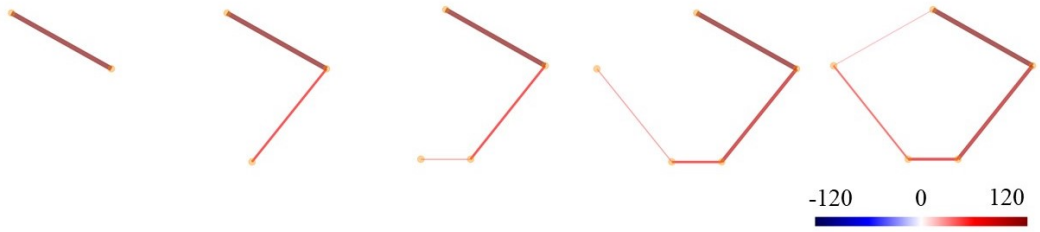


Figure 7: The five networks used to simulate the dataset in Case study 1.

5.1 Case study 1: Undirected Networks with Scalar Edge Attributes

We simulate networks with real-valued node- and edge attributes as follows. Figure 7 shows five networks with edge attributes $\{100, 80, 60, 40, 20\}$ as shown by the darkness of the color, whose equivalence classes in X/T we term $\{[x_1], [x_2], [x_3], [x_4], [x_5]\}$. We generate a synthetic dataset consisting of 50 observations randomly sampled in a stratified fashion by randomly picking one of the five equivalence classes $[x_i], i = 1, 2, 3, 4, 5$, and then we randomly pick an element from this equivalence class by

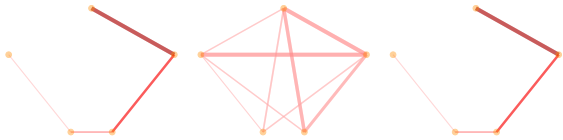


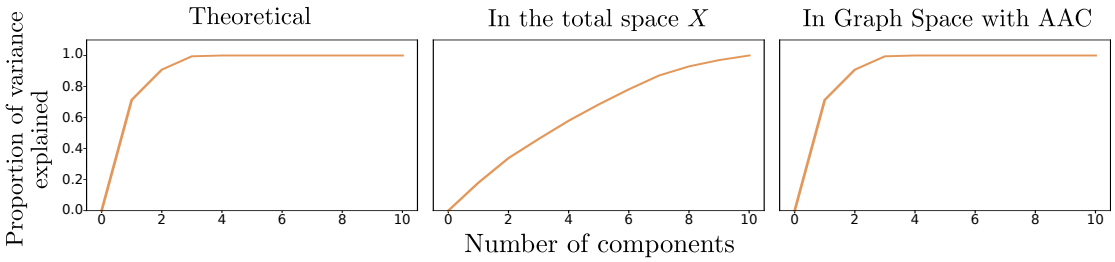
Figure 8: Theoretical Fréchet mean; Fréchet mean in X; Fréchet Mean computed via AAC. The scale is the same as in Figure 7.

randomly permuting the network's nodes. The Fréchet mean and the GPCs are computed using our AAC algorithm, which works on the quotient space X/T. These results are compared to Fréchet means and principal components computed on X, using the initial graph representations, as well as theoretically correct results which can be computed as for this particular synthetic dataset, the alignment can be made by hand. Figure 8 shows the three computed Fréchet means. The mean computed with the AAC algorithm is a good estimate of the theoretical one, both in terms of topological structure and edge weights. The mean obtained in the X space is a complete weighted network shown in Figure 8, capturing neither the topology nor the weights.

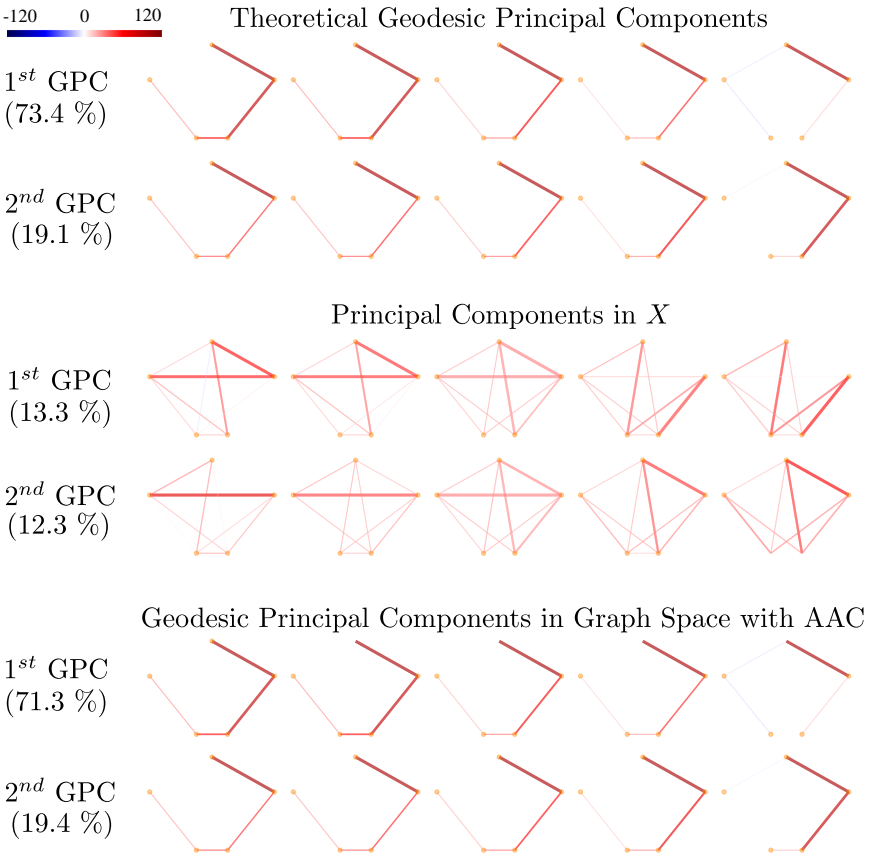
For GPCA, Figure 9(a) shows the cumulative proportion of variance explained as a function of number of GPCs, and Figure 9(b) visualizes the graph variation along the first two principal components. Note that the Graph Space GPCs obtained using AAC capture the same quantitative and visual level of variance as the theoretical GPCs, while this does not hold for the Euclidean version. In particular, the two GPCs explain more then 90% of the original data variance, and the visualization of the 1^{st} principal component shows how it runs from a single edge to the full structure.

5.2 Case study 2: Undirected Networks with Vector Attributes

As an intuitive visual example with real data with vectors attributes, we subsample 20 cases of the letter "A" from the well known hand written letters dataset (Kersting et al., 2016; Riesen and Bunke, 2008). As shown in Figure 10(a), every network has node attributes consisting of the node's x- and y-coordinates, and binary $(0/1)$ edge attributes indicating whether nodes are connected by lines. In Figure 10(a), the Fréchet Mean is shown, underling how the framework is capturing both the topology and the node coordinates. Figure 10(b) plots network variation along the three GPCs. Note in particular how the principal components are capturing the variability in the way the letter A could be written: the variability of the length and the inclination of the

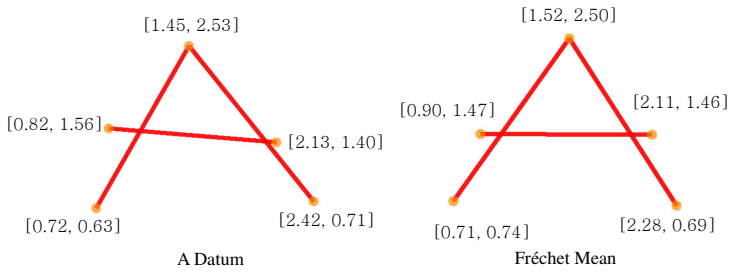


(a) Proportion of variance explained is shown by varying the number of the GPCs.

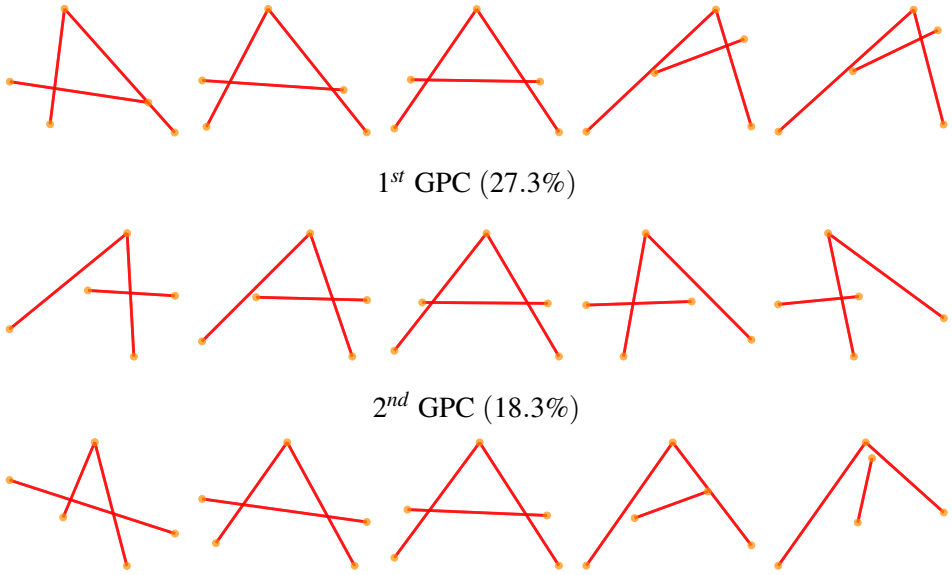


(b) The first two Geodesic Principal Components are shown by plotting the original data projected along the corresponding geodesic (only the $q = 0.1, 0.25, 0.5, 0.75, 0.9$ quantiles are shown).

Figure 9: GPCA analysis for Case study 1.



(a) **Left:** A datum extracted from the A dataset. Every unlabelled node has a bidimensional real valued attribute, while every edge has a $0, 1$ attribute. **Right:** The Fréchet mean.



(b) Visualization of the GPCs. $0.1, 0.25, 0.5, 0.75, 0.9$ quantile of the projected scores are shown for the first three GPCs.

Figure 10: GPCA analysis for Case study 2.

horizontal bar, the angle between the vertical bars, and the reciprocal positions of the bars.

5.3 Case study 3: Directed Networks with Vector Attributes

Our final example uses a mobility dataset from Open Data Regione Lombardia (Region, Region). The dataset consists of origin-destination matrices of the commuting flux of people between the 11 provinces of the Lombardia Region in northern Italy. For every hour of a representative day in 2014, the fluxes were collected counting the number of people travelling by private mobility mode (car), railway system (train or metro), bus public transport system, or bike. This results in a set of 24 multi-layer networks, represented as graphs whose edge attributes are vectors in $\mathbb{R}^4$. Each coordinate represents the flux associated to one of the mobility modes.

Figure 11, top, shows the multi-level networks associated with four different hours (left), as well as the Fréchet Mean on Graph space X/T computed with the AAC algorithm (right). Note that the density of the layers are well represented by the mean. By looking at the permutation of network nodes used to compute the mean and the geodesics, we see that most of the time no permutation is performed (i.e. the node corresponding to a specific province at one hour is matched to the node representing the same province at another hour). This means that even if the province information is not stored in the graph, they are distinguished by their mobility properties. The only permutation happens at 5 p.m., when the town of Brescia is permuted with the town of Bergamo. These two towns are both important commuting satellites of Milan, so their role is interchangeable with respect to the commuting flux in the afternoon. This application shows how this framework does not suffer the possible risk of "over-matching" also in case where the cardinality of the permutation group is pretty high (i.e., even though about 40 million possible node permutations are available, no artificial permutations are introduced by the algorithm). In Figure 11, bottom, we show that the majority of the variability is explained by the first GPC. The first principal component captures 81% of the total variance, and we see that by moving along this

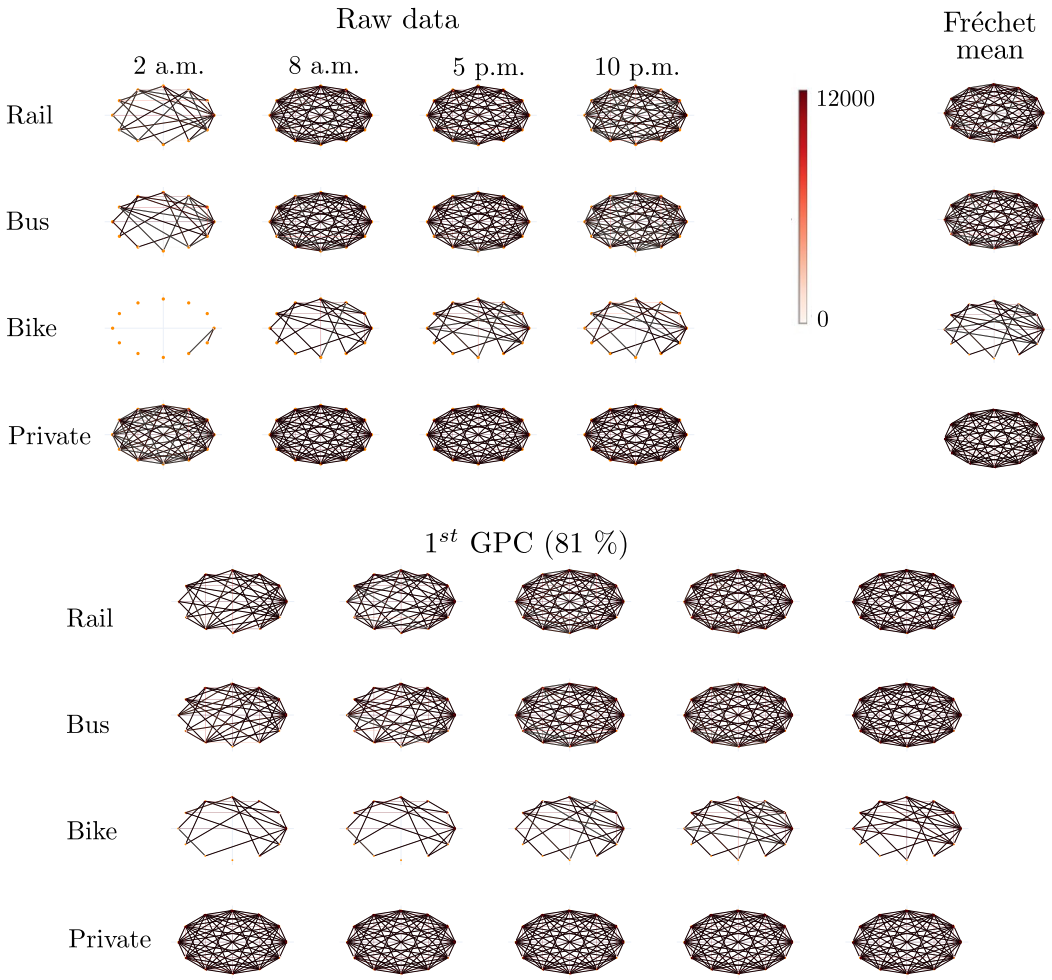


Figure 11: **Top left:** Network of the fluxes between the Lombardia Region provinces at 2 a.m., 8 a.m., 5 p.m. and 10 p.m.. **Top right:** Fréchet Mean in Graph Space X/T for the mobility modes bike, bus, rail and private. **Bottom:** GPCA performed on scaled data. The first GPC is shown, by plotting the quantiles of the original data projected along the corresponding geodesic for each one of the levels ($q = 0.1, 0.25, 0.5, 0.75, 0.9$). The GPC captures the density change along the day for all the transportation modes except from Private car mode.

component, the density of the layers changes, except from the private mode, which is a complete graph at at every hour of the day. As expected, the first GPC captures the main variability of the dataset, which resides in the difference in number of travels along the day, from early-morning and late afternoon peaks to minima at night.

6 Discussion and conclusion

In this work, we study the Graph Space, as a particular case of Jain and Obermayer (2009). As a quotient space, Graph Space allows a meaningful description of a population of networks with both same nodes across the population, i.e. the permutation action is the identity, and different nodes. Moreover, Graph Space allows Euclidean attributes of any dimension on nodes and edges. To perform statistics on the Graph Space, we studied the geometrical properties of the Graph Space such as the metric, the curvature (which we prove to be unbounded from above), and the non uniqueness of the geodesics and the Fréchet Mean. To overcome the problems induced by unbounded curvature, we propose a general algorithm (Align All and Compute Algorithm - AAC) useful to compute statistics on the Graph Space. First, we apply the AAC for the estimation of the Fréchet mean, showing convergence to a local minimum of the variance function in finite time. Second, we introduce the concept of geodesic principal component - GPC - on Graph Space, and perform GPC analysis for populations of network-valued data using the AAC algorithm. GPCA is a useful tool to summarize the complexity of network valued datasets, as well as identifying the major topological and attribute variations. We show the practical utility of the framework on three case studies: A simulated study, showing how AAC for Fréchet Mean and GPCs estimates the expected theoretical results, and two real world examples including both directed and undirected networks, as well as both scalar and vector attributes. The examples emphasize that GPCA is an efficient tool to decompose the complexity and visualize the variability in a population of

networks.

Networks are a powerful and natural mathematical representation of various phenomena. Due to its flexibility, Graph Space can offer interesting practical application in different field such as chemistry, economy, social science, medicine etc. Note that Graph Space is analogous to the classical shape analysis (Kendall, 1984; Dryden and Mardia, 1998), with graph nodes analogous to landmarks. In this sense, Graph Space might also be a potential alternative to shape analysis when the indexing of landmarks is unknown or the number of landmarks varies. On a related note, as our computation of means and principal components is essentially a generalization of generalized Procrustes analysis, our estimators may suffer from similar biases as those shown to exist for shapes (Miolane et al., 2017). As a further development, the analysis of networks with non-Euclidean attributes is a first interesting extension of the current framework. Moreover, the AAC algorithm is a general strategy to compute statistics on the Graph Space, so it can be extended to other statistical tools, such as regression and classification.

Acknowledgements

We thank Brijnesh Jain for sharing his original java implementation of Structure Space computations, and in particular for his detailed explanation of his code and algorithms. The research was supported by Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation, and by the Polisocial Award - Politecnico di Milano, Safari Njema Project.

References

Afsari, B., R. Tron, and R. Vidal (2013). On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM J. Control and Optimization 51*(3), 2230–2260.

Arnaudon, M., F. Barbaresco, and L. Yang (2013). Medians and means in Riemannian geometry: Existence, uniqueness and computation. In *Matrix Information Geometry*, pp. 169–197. Springer.

Arnaudon, M. and L. Miclo (2014). A stochastic algorithm finding generalized means on compact manifolds. *Stochastic Processes and their Applications 124*(10), 3463–3479.

Bačák, M. (2014). Computing medians and means in Hadamard spaces. *SIAM Journal of Optimization 24*(3), 1542–1566.

Billera, L. J., S. P. Holmes, and K. Vogtmann (2001). Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics 27*(4), 733–767.

Bonnabel, S. (2013). Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions on Automatic Control 58*(9), 2217–2229.

Bredon, G. E. (1972). *Introduction to compact transformation groups*, Volume 46. Academic press.

Bridson, M. R. and A. Haefliger (1999). *Metric spaces of non-positive curvature*. Springer.

Bunke, H. and K. Riesen (2011). Improving vector space embedding of graphs through feature selection algorithms. *Pattern Recognition 44*(9), 1928–1940.

Calissano, A., A. Feragen, and S. Vantini (2019). A mathematical framework for population of networks: Comparing public transport of different cities. *Proceedings to SIS 2019*.

Calissano, A., A. Feragen, and S. Vantini (2020). Graphspace python package. *https://github.com/annacalissano/GraphSpace.git*.

Chakraborty, R. and B. C. Vemuri (2015). Recursive Fréchet mean computation on the Grassmannian and its applications to computer vision. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4229–4237.

Chowdhury, S. and F. Mémoli (2017). Distances and isomorphism between networks and the stability of network invariants. *arXiv preprint arXiv:1708.04727*.

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

Conte, D., P. Foggia, C. Sansone, and M. Vento (2004). Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence 18*(03), 265–298.

Dryden, I. and K. Mardia (1998). *Statistical analysis of shape*. Wiley.

Duncan, A., E. Klassen, and A. Srivastava (2018). Statistical shape analysis of simplified neuronal trees. *The Annals of Applied Statistics 12*(3), 1385–1421.

Durante, D., D. B. Dunson, and J. T. Vogelstein (2017). Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association 112*(520), 1516–1530.

Duvenaud, D. K., D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez- Bombarelli, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams (2015). Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems*, pp. 2224–2232.

Emmert-Streib, F., M. Dehmer, and Y. Shi (2016). Fifty years of graph matching, network alignment and network comparison. *Information Sciences 346*, 180–197.

Feragen, A., S. Hauberg, M. Nielsen, and F. Lauze (2011). Means in spaces of tree-like shapes. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 736–746.

Feragen, A., F. Lauze, P. Lo, M. de Bruijne, and M. Nielsen (2010). Geometries on spaces of treelike shapes. In *Asian Conference on Computer Vision*, pp. 160–173. Springer.

Feragen, A. and T. M. W. Nye (2020). Statistics on stratified spaces. In *Riemannian Geometric Statistics in Medical Image Analysis*, pp. 299–342. Elsevier.

Feragen, A., M. Owen, J. Petersen, M. Wille, L. Thomsen, A. Dirksen, and M. de Bruijne (2013). Tree-space statistics and approximations for large-scale analysis of anatomical trees. In *Information Processing in Medical Imaging*, Volume 7917 of *Lecture Notes in Computer Science*, pp. 74–85. Springer.

Fletcher, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision 105*(2), 171–185.

Fletcher, P. T. and S. Joshi (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer vision and mathematical methods in medical and biomedical image analysis*, pp. 87–98. Springer.

Ginestet, C. E., J. Li, P. Balachandran, S. Rosenberg, and E. D. Kolaczyk (2017, 06). Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics 11*(2), 725–750.

Gold, S. and A. Rangarajan (1996). A graduated assignment algorithm for graph matching. *IEEE Transactions on pattern analysis and machine intelligence 18*(4), 377–388.

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika 40*(1), 33–51.

Guo, X., A. Srivastava, and S. Sarkar (2019). A quotient space formulation for statistical analysis of graphical data. *arXiv preprint arXiv:1909.12907*.

Hauberg, S., A. Feragen, R. Enficiaud, and M. J. Black (2015). Scalable robust principal component analysis using Grassmann averages. *IEEE transactions on pattern analysis and machine intelligence 38*(11), 2298–2311.

Huckemann, S., T. Hotz, and A. Munk (2010). Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statist. Sinica 20*(1), 1–58.

Jain, B. and K. Obermayer (2008). On the sample mean of graphs. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 993–1000. IEEE.

Jain, B. J. and K. Obermayer (2009). Structure spaces. *Journal of Machine Learning Research 10*(Nov), 2667–2714.

Kendall, D. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society 16*(2), 81–121.

Kersting, K., N. M. Kriege, C. Morris, P. Mutzel, and M. Neumann (2016). Benchmark data sets for graph kernels.

Kolaczyk, E., L. Lin, S. Rosenberg, J. Xu, and J. Walters (2017). Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *arXiv preprint arXiv:1709.02793*.

Mallasto, A. and A. Feragen (2018). Wrapped Gaussian process regression on Riemannian manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5580–5588.

Marron, J. S. and A. M. Alonso (2014). Overview of object oriented data analysis. *Biometrical Journal 56*(5), 732–753.

Miller, E., M. Owen, and J. S. Provan (2015). Polyhedral computational geometry for averaging metric phylogenetic trees. *Advances in Applied Mathematics 68*, 51 – 91.

Miolane, N., S. Holmes, and X. Pennec (2017). Template shape estimation: correcting an asymptotic bias. *SIAM Journal on Imaging Sciences 10*(2), 808–844.

Nye, T. M. W. (2011). Principal components analysis in the space of phylogenetic trees. *The Annals of Statistics 39*, 2716–2739.

Nye, T. M. W. (2014). An algorithm for constructing principal geodesics in phylogenetic treespace. *IEEE/ACM Trans. Comput. Biology Bioinform. 11*(2), 304–315.

Nye, T. M. W., X. Tang, G. Weyenberg, and R. Yoshida (2017). Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees. *Biometrika 104*, 901–922.

Pennec, X., P. Fillard, and N. Ayache (2006). A Riemannian framework for tensor computing. *International Journal of Computer Vision (IJCV) 66*(1), 41–66.

Region, L. Open Data Regione Lombardia. Accessed: 2019-09-02.

Riesen, K. and H. Bunke (2008). Iam graph database repository for graph based pattern recognition and machine learning. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pp. 287–297. Springer.

Sanfeliu, A. and K.-S. Fu (1983). A distance measure between attributed relational graphs for pattern recognition. *IEEE transactions on systems, man, and cybernetics 3*, 353–362.

Shervashidze, N., P. Schweitzer, E. J. van Leeuwen, K. Mehlhorn, and K. M. Borgwardt (2011). Weisfeiler-Lehman graph kernels. *Journal of Machine Learning Research 12*, 2539–2561.

Simpson, S. L., R. G. Lyday, S. Hayasaka, A. P. Marsh, and P. J. Laurienti (2013). A permutation testing framework to compare groups of brain networks. *Frontiers in computational neuroscience 7*, 171–184.

Srivastava, A., S. H. Joshi, W. Mio, and X. Liu (2005). Statistical shape analysis: Clustering, learning, and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 27*(4), 590–602.

Sturm, K. T. (2003). Probability measures on metric spaces of nonpositive. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces 338*, 357–391.

Turner, K., Y. Mileyko, S. Mukherjee, and J. Harer (2014). Fréchet means for distributions of persistence diagrams. *Disc. Comp. Geom. 52*(1), 44–70.

von Ferber, C., T. Holovatch, Y. Holovatch, and V. Palchykov (2009). Public transport networks: empirical analysis and modeling. *The European Physical Journal B 68*(2), 261–275.

Wang, H. and S. J. Marron (2007). Object oriented data analysis: Sets of trees. *The Annals of Statistics 35*(5), 1849–1873.

Zhang, M. and P. T. Fletcher (2013). Probabilistic principal geodesic analysis. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26*, pp. 1178–1186. Curran Associates, Inc.

Appendix for *Populations of Unlabeled Networks: Graph Space Geometry and Geodesic Principal Components.*
Proof of Theorem 6 and Theorem 7.

Proof of Theorem 6

Proof. First, we prove convergence in finite time. Algorithm 1 consists of two steps repeated iteratively. Consider the function

$$\sum_{i=1}^{k} d_X^2(\mu^{cur}, x_i^{cur}), \tag{9}$$

where, at any point in time, μ^{cur} is the current representative in X of the current estimate of the Fréchet mean, and x_i^{cur} is the current representative (with current optimal node alignment to μ^{cur}) in X of the sample point $[x_i]$. Note that the first step, aligning data points to the current representative of the current mean estimate, cannot increase the value of (9) as an improved alignment would indeed lower the value of (9). Similarly, the second step, which is the re-estimation of the Fréchet mean given the new alignments, also cannot increase the value of (9) as, again, an improved estimate would lower its value. Moreover, if the value of (9) stays fixed two iterations in a row, the algorithm will terminate. Thus, the iterative algorithm will never see the same set of sample-wise alignments twice. As there are only finitely many such sets, the algorithm is forced to terminate in finite time.

Next, we move to convergence to a local minimum. Let $[\mu] \in X/T$ be the estimated mean, let $\mu \in X$ be a representative of it, and let $x_1, \ldots, x_k \in X$ be optimally aligned representatives of the sampled graphs (as in the final step of the AAC algorithm). We will show below that

with probability 1, there exists some $\varepsilon > 0$ such that for any $\mu' \in B_X(\mu, \varepsilon)$, the representatives $x_1, \ldots, x_k$ are also optimally aligned with μ'. In this case, since μ is a local minimizer of (9) within $B(\mu, \varepsilon)$, and $d_{X/T}([\mu'], [x_i]) = d_X(\mu', x_i)$ for all $\mu' \in B_X(\mu, \varepsilon)$, the estimated mean graph $[\mu]$ is a local minimizer of (4).

In order to prove the existence of such an $\varepsilon > 0$, we rely on the following lemma:

Lemma 8. *Given representatives $x_1, \ldots, x_k$ of $[x_1], \ldots, [x_k]$ with mean μ in X, the following holds with probability 1:*

For all $i = 1, \ldots, k$ and for all $t \in T \setminus T_{x_i}$,

$$d(\mu, x_i) \neq d(\mu, tx_i),$$

where T_{x_i} is the stabilizer $T_{x_i} = \{t \in T | tx_i = x_i\}$.

If the lemma holds, then we may define

$$\delta = \min\{d(\mu, tx_i) - d(\mu, x_i) \mid i = 1, \ldots, k,\ t \in T \setminus T_{x_i}\} > 0.$$

We now set $\varepsilon = \frac{\delta}{2}$ and consider $\mu' \in B_X(\mu, \varepsilon)$. We wish to show that for all $i = 1, \ldots, k$ and all $t \in T \setminus T_{x_i}$, we have $d(\mu', x_i) < d(\mu', tx_i)$, namely that the optimal representative of $[x_i]$ is left unchanged for all i.

Note that by the definition of δ, we have

$$d(\mu, x_i) \leq d(\mu, tx_i) - \delta,$$

and by the triangle inequality, we have

$$d(\mu', x_i) \leq \underbrace{d(\mu', \mu)}_{<\varepsilon = \frac{\delta}{2}} + d(\mu, x_i)$$

and

$$d(\mu, tx_i) \leq \underbrace{d(\mu,\mu')}_{<\varepsilon=\frac{\delta}{2}} + d(\mu', tx_i).$$

We compute

$$d(\mu', x_i) \leq \underbrace{d(\mu',\mu)}_{<\frac{\delta}{2}} + d(\mu, x_i) < \frac{\delta}{2} + d(\mu, x_i) \leq \frac{\delta}{2} + d(\mu, tx_i) - \delta < -\frac{\delta}{2} + \frac{\delta}{2} + d(\mu', tx_i) = d(\mu', tx_i),$$

which completes the proof of Theorem 6 under the assumption that Lemma 8 holds. □

Proof of Lemma 8. In order to prove the lemma, we will show that the set

$$\mathscr{X}_T = \left\{ ([x_1], \ldots, [x_k]) \in (X/T)^k \;\middle|\; \begin{array}{c} d(\mu, x_i) = d(\mu, tx_i) \text{ for some representatives } x_1, \ldots, x_k, \\ i = 1, \ldots, k \text{ and } t \in T \setminus T_{x_i} \end{array} \right\}$$

has measure $\eta_k(\mathscr{X}_T) = 0$, where η_k is the product measure induced by η on $\underbrace{X/T \times \ldots \times X/T}_{k}$.

For each element $t \in T$, denote by $X^t = \{x \in X | tx = x\}$ the fixed point set of t. Note that $\eta_k(\mathscr{X}_T) = m_k(\pi^{-1}(\mathscr{X}_T))$, and that

$$\pi^{-1}(\mathscr{X}_T) = \bigcup_{i=1}^{k} \bigcup_{t \in T} \mathscr{X}_{i,t},$$

where

$$\mathscr{X}_{i,t} = \{(x_1, \ldots, x_k) \in X \times \ldots \times \underbrace{X \setminus X^t}_{i^{th}} \times \ldots \times X | d_X(\mu, x_i) = d_X \mu, tx_i)\} \subset \underbrace{X \times \cdots \times X}_{k}.$$

The preimage $f^{-1}(0)$ of the function

$$f \colon \underbrace{X \times \cdots \times X}_{k} \to \mathbb{R}, \qquad (x_1, \ldots, x_k) \mapsto d_X^2(\mu, x_i) - d_X^2(\mu, tx_i)$$

satisfies

$$f^{-1}(0)\cap X\times\ldots\times\underbrace{X\setminus X^t}_{i^{th}}\times\ldots\times X=\mathscr{X}_{i,t}.$$

We show that f is a submersion on $X\times\ldots\times\underbrace{X\setminus X^t}_{i^{th}}\times\ldots\times X$ by showing that it has nonzero gradient. We can rewrite

$$f(x_1,\ldots,x_k)=(\frac{1}{k}\sum_{j=1}^{k}x_j-x_i)^T(\frac{1}{k}\sum_{j=1}^{k}x_j-x_i)-(\frac{1}{k}\sum_{j=1}^{k}x_j-tx_i)^T(\frac{1}{k}\sum_{j=1}^{k}x_j-tx_i)=\frac{2}{k}\sum_{j=1}^{k}\left(x_j^Ttx_i-x_j^Tx_i\right).$$

For $j\neq i$ we obtain

$$\nabla_{x_j}f(x_1,\ldots x_k)=\frac{2}{k}(tx_i-x_i)$$

which is nonzero for $(x_1,\ldots,x_k)\in X\times\ldots\times\underbrace{X\setminus X^t}_{i^{th}}\times\ldots\times X$. It follows that f is a submersion on $X\times\ldots\times\underbrace{X\setminus X^t}_{i^{th}}\times\ldots\times X$. As a result, the set

$$f^{-1}(0)\cap X\times\ldots\times\underbrace{X\setminus X^t}_{i^{th}}\times\ldots\times X=\mathscr{X}_{i,t}$$

has codimension 1 and, in particular,

$$m_k(\mathscr{X}_{i,t})=m_k(f^{-1}(0)\cap X\times\ldots\times\underbrace{X\setminus X^t}_{i^{th}}\times\ldots\times X)=0.$$

But then,

$$\eta_k(\mathscr{X}_k)=m_k(\pi^{-1}(\mathscr{X}_T))=m_k\left(\bigcup_{i=1}^{k}\bigcup_{t\in T}\mathscr{X}_{i,t}\right)\leq\sum_{i=1}^{k}\sum_{t\in T}m_k(\mathscr{X}_{i,t})=0,$$

which proves the lemma. □

Proof of Theorem 7

Proof of Theorem 7. The proof of convergence of AAC for GPCA follows the exact same strategy as for the Fréchet mean in Theorem 6, noting:

- The algorithm converges in finite time because every step either decreases or leaves unchanged the value of the sum of squared residuals for representatives in X:

$$\sum_{i=1}^{k} d^2(x_i, \gamma)$$

  with respect to the first GPC $\delta \in \Gamma(X/T)$ and its representative $\gamma \in \Gamma(X)$.

- For the first GPC, the algorithm converges to a local minimum of (6) following the same argument as above. Here, the ε-neighborhood of the obtained estimate δ has to be considered on the Grassmannian manifold of 1-dimensional subspaces of X.

□

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

13/2020 Pozzi S.; Domanin M.; Forzenigo L.; Votta E.; Zunino P.; Redaelli A.; Vergara C.
A data-driven surrogate model for fluid-structure interaction in carotid arteries with plaque

12/2020 Azzolin, L.; Dede', L.; Gerbi, A.; Quarteroni, A.
Effect of fibre orientation and bulk value on the electromechanical modelling of the human ventricles

08/2020 Antonietti, P. F.; Facciolà, C.; Verani, M.
Polytopic Discontinuous Galerkin methods for the numerical modelling of flow in porous media with networks of intersecting fractures

09/2020 Rea, F.; Ieva, F.,; Pastorino, U.; Apolone, G.; Barni, S.; Merlino, L.; Franchi, M.; Corrao, G.
Number of lung resections performed and long-term mortality rates of patients after lung cancer surgery: evidence from an Italian investigation

10/2020 Bonaventura, L.; Carlini, E.; Calzola, E.; Ferretti, R.
Second order fully semi-Lagrangian discretizations of advection--diffusion--reaction systems

11/2020 Antonietti, P.F.; Facciola', C.; Houston, P.; Mazzieri, I.; Pennes, G.; Verani, M.
High-order discontinuous Galerkin methods on polyhedral grids for geophysical applications: seismic wave propagation and fractured reservoir simulations

07/2020 Fumagalli, A.; Scotti, A.
Reactive flow in fractured porous media

06/2020 Domanin, M.; Piazzoli, G.; Trimarchi, S.; Vergara, C.
Image-based displacements analysis and computational blood dynamics after endovascular aneurysm repair

05/2020 Artioli, E.; Beiraoda Veiga, L.; Verani, M.
An adaptive curved virtual element method for the statistical homogenization of random fibre-reinforced composites

04/2020 Didkovskyi, O.; Azzone, G.; Menafoglio A.; Secchi P.
Social and material vulnerability in the face of seismic hazard: an analysis of the Italian case