

Populations of Unlabelled Networks: Graph Space Geometry and Generalized Geodesic Principal Components

BY ANNA CALISSANO

*MOX- Department of Mathematics, Politecnico di Milano, Via Bonardi 9, 20122 Milano ITA.
Now at Inria centre at Université Côte d'Azur, 2004 Rte des Lucioles, 06902 Valbonne, Francia*
anna.calissano@inria.fr

AASA FERAGEN

*DTU Compute, Technical University of Denmark, 2800 Kongens Lyngby, Denmark and DIKU,
University of Copenhagen, Universitetsparken 1, 2100 København*
afhar@dtu.dk

AND SIMONE VANTINI

MOX- Department of Mathematics, Politecnico di Milano, Via Bonardi 9, 20122 Milano ITA
simone.vantini@polimi.it

SUMMARY

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 unlabelled networks which are weighted or unweighted, uni- or multi-layered, directed or undirected. 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. Within this geometrical framework we define generalized geodesic principal components, and we introduce the Align All and Compute algorithms, all of which allow for the computation of statistics on Graph Space. The statistics and algorithms are compared with existing methods and empirically validated on three real datasets, showcasing the framework potential utility. The whole framework is implemented within the geomstats Python package.

Some key words: Network data, Object data, Graph Space, Computational Geometric Statistics, Quotient Space, Geodesic Principal Components

1. INTRODUCTION

Interest in the analysis of tree- and network-data has soared in recent years due to the utility of representing certain architectures and processes by graphs. The scientific literature has so far primarily focused on analyzing networks in a *first generation* (Wang & 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 & 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 in a particular population

has the same nodes then the problem can be re-framed as the analysis of a set of adjacency matrices; however, if the networks have different numbers of nodes and different node labels then the mathematical challenges multiply. Populations of unlabelled networks are common in real world problems, aiming at describing all those applications where the node matching is non-trivial (e.g. social network of different individuals; trade network of different companies; brain connectivity networks of different patients). Populations of networks also represent one of the major challenges in network-valued data analysis. In fact, with different nodes across networks, a concept of alignment (matching) should be introduced.

Populations of tree- or network data have been studied from different perspectives. A first approach is to embed trees or networks in Euclidean space (e.g., kernel methods by Shervashidze et al. 2011, convolutional neural networks by Duvenaud et al. 2015, feature selection algorithms by Bunke & 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. Particularly challenging are network-valued parameters, such as averages, interpolation, denoising, network-valued regression models, etc. As a response to these limitations, Object Oriented Data Analysis (Marron & 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. Most existing works in this category have so far focused on tree-valued data rather than networks in general (Bille, 2005; Feragen et al., 2010; Garba et al., 2021; Lueg et al., 2021). Among existing models for graphs, Ginestet et al. (2017); Severn et al. (2021) 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 tests on a population of networks; Lunagómez et al. (2020) provide Bayesian modelling for discrete labelled graphs, and Chowdhury & Mémoli (2019) study a metric space of networks up to weak isomorphism, which allows the grouping of similar nodes.

As a precursor to our work, Jain & 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 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 Kolaczyk et al. (2020), which studies the behavior of Fréchet mean in the fundamental domain of Graph Space, as well as in Guo et al. (2021), which proposes a simpler algorithm for principal components, analogous to the tangent space approaches known from manifold statistics (Fletcher & Joshi, 2004).

In Section 2 and 3, we describe Graph Space and the nonlinear nature of its geometry. In Section 4, we introduce an intrinsic version of Geodesic Principal Components for a population of unlabelled networks. We propose the Align All and Compute (AAC) statistical algorithm to compute estimators on the Graph Space, and prove that it converges to a local minimum in finite time (Section 5). We discuss, in particular, the characteristics of the Fréchet Mean and Geodesic Principal Components computed with this algorithm. We compare the proposed method with Tangent-like approaches with two illustrative examples (Section 6). To demonstrate the potential utility and flexibility of this framework we illustrate our algorithms using three different datasets (Section 7). All the proposed theory is implemented in the geometry module GraphSpace and learning module AAC available within geomstats python package (Miolane 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 $a: E \rightarrow A$. 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 $X = \mathbb{R}^{n \times n}$ of 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 type of relational phenomena in different contexts (e.g. routes of different airline companies, or brain connectivity networks of different patients). Differences in the labels or order of nodes 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 by permutation matrices and its action on X in conjugation:

$$\cdot: T \times X \rightarrow X, (t, x) \mapsto t^T x t$$

thus defines an action of the group T on X . We call the obtained quotient space X/T *Graph Space*, and each element of X/T is a graph G , represented as an equivalence class $[x] = \{t^T x t, t \in T\}$ which contains all the adjacency matrices obtained from x by permuting nodes. The *canonical projection* to the Graph Space X/T is defined as:

$$\pi: X \rightarrow X/T, \quad \pi(x) := [x].$$

Graph Space is a special case of the $\mathbb{R}^d$ -attributed 2-structures introduced by Jain & Obermayer (2009). In the implementation the adjacency matrices are converted to flattened vectors without loss of generality. An example of how to build Graph Space is available the Appendix A.

Remark 1. The choice of the maximum number of nodes n will affect downstream analysis. In our experiments, the number of nodes is set to $n = \max(n_i)$. A consequence of this choice is that any node appearing in any graph will implicitly be matched to a node in the maximal size graph. To avoid this, users would need to choose a larger maximal number of nodes.

3. THE GRAPH SPACE IS NOT EUCLIDEAN AND NOT A MANIFOLD

3.1. Graph Space is not a manifold

Graph Space X/T is the quotient of the total space X of adjacency matrices with respect to the node permutation group T . As detailed in Jain & Obermayer (2009), any metric d_X on X defines a quotient pseudometric $d_{X/T}$ on X/T . If d_X is isometric with respect to the group action then it can be expressed as:

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

The resulting quotient metric $d_{X/T}$ is indeed a non-degenerate metric, since the permutation group T is finite. 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.

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 $T_x = \{t \in T \mid t^T x t = x\}$ 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 an 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, many well-known tools from *manifold statistics* (Fletcher & Joshi, 2004; Huckemann et al., 2010; Kendall, 1984; Srivastava et al., 2005; Zhang & Fletcher, 2013; Pennec et al., 2006; Fletcher, 2013; Mallasto & Feragen, 2018) are unfortunately not directly applicable to Graph Space.

However, as is commonly done in both general nonlinear statistics (Turner et al., 2014; Nye, 2011, 2014; Duncan et al., 2018; Feragen & Nye, 2020b; Bacák, 2014; Nye et al., 2017; Miller et al., 2015; Sturm, 2003; Feragen et al., 2013, 2011) and in manifold statistics, 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 and statistical properties built on geodesics by utilizing geometric constructions from *metric geometry* (Bridson & 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.

3.2. Graph Space is a geodesic space

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

$$l(\gamma) = \sup \left\{ \sum_{i=1}^m d_X(x_{i-1}, x_i) \mid x_0 = \gamma(0), x_1, \dots, x_m = \gamma(1) \text{ for some } m \in \mathbb{N} \right\} \quad (1)$$

where the supremum is taken over all approximations $x_0 = \gamma(0), x_1 = \gamma(s_1), \dots, x_{m-1} = \gamma(s_{m-1}), x_m = \gamma(1)$ of γ of arbitrary finite length m , where $0 < s_1 < \dots < s_{m-1} < 1$. 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 $\gamma: [0, 1] \rightarrow \mathfrak{X}$ 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

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

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. Using standard properties of metric spaces we can show:

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

Proof. See Appendix B. $\square$

While any two points can be connected by a geodesic, we illustrate through a series of counterexamples that these geodesics are not always unique. The first counterexample, shown in Figure 1, shows how near a graph $[x]$ with at most $n - 2$ nodes, you can find a pair of points arbitrarily close nearby, that are connected by two geodesics. We formalize this as a lemma:

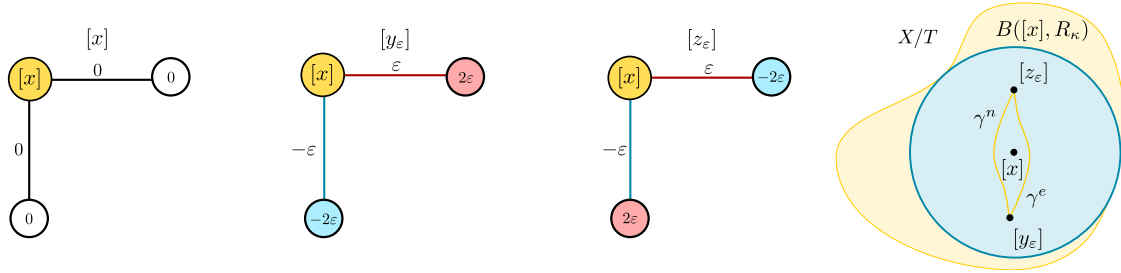


Fig. 1. For any graph $[x]$ with at most $n - 2$ nodes, we can find two graphs $[y_\epsilon]$ and $[z_\epsilon]$ arbitrarily close to it, which are connected by more than one geodesic.

LEMMA 2. *There exist graphs $[x] \in X/T$ such that for any $\epsilon > 0$, there exist points $y_\epsilon, z_\epsilon \in B([x], \epsilon) \subset X/T$ which are connected by more than one geodesic. In particular, the set of such graphs $[x]$ include all graphs $[x]$ with at most $n - 2$ nodes, as well as certain symmetric graphs $[x]$ with up to n nodes.*

Proof. See Appendix B. □

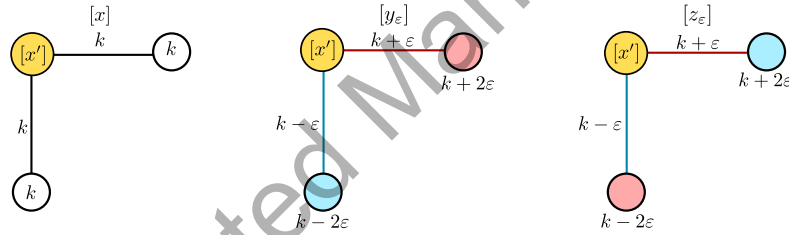


Fig. 2. There exist graphs $[x]$ admitting particular permutation symmetries, such that we can find two graphs $[y_\epsilon]$ and $[z_\epsilon]$ arbitrarily close to $[x]$, which are connected by more than one geodesic. In particular, non-unique geodesics can appear even without having to add zero-attributed nodes.

Note that the first counterexample leading to Lemma 2 can be constructed near any graph $[x]$ with at most $n - 2$ nodes. Are non-unique geodesics caused by using "dummy" nodes with 0 attributes to include graphs with less than n nodes in the graph space? First, note that the graphs $[y_\epsilon]$ and $[z_\epsilon]$ do have n nodes if $[x]$ has $n - 2$ nodes. The geodesic connecting them will pass through nodes with 0 attributes, but if we do not include these points, we would no longer have a geodesic graph space. Second, in Figure 2 a similar example is constructed near graphs $[x]$ with up to n nodes that admit certain symmetries. Here, $[x]$ is constructed from an arbitrary subgraph $[x']$ with at most $n - 2$ nodes. In other words, non-unique geodesics can also be found arbitrarily close to certain graphs with n nodes. Third, the reader might wonder whether these non-uniqueness results are just a function of considering graphs with attributed nodes. Could this be avoided by only allowing attributes on edges? As we show in Figure 3, this is also not the case. Indeed, a completely analogous example can be constructed by building the graph $[x]$ from a subgraph $[x']$ with at most $n - 4$ nodes, using additional edges to play the role taken on by the red and blue nodes in the previous examples. Finally, the reader might ask whether graph space just does not have unique geodesics? We conjecture that within sufficiently strong assumptions, generic pairs of graphs with n nodes may be connected by unique geodesics, in a similar way as

is known for unlabelled trees (Feragen & Nye, 2020a). In this paper, however, we take the route of defining statistics that do not rely on uniqueness of geodesics.

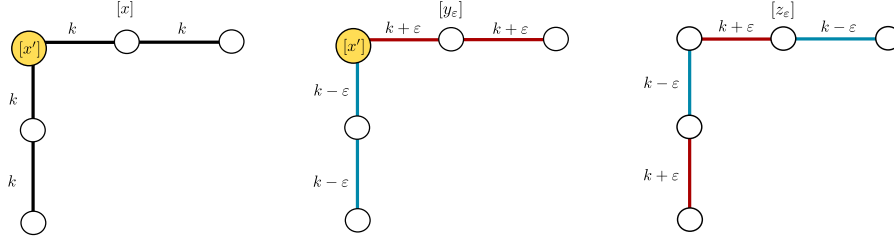


Fig. 3. For graphs without node attributes, there exist graphs $[x]$ such that we can find two graphs $[y_\epsilon]$ and $[z_\epsilon]$ arbitrarily close to $[x]$, by attaching a new edge in place of the weighted node.

3.3. Graph Space curvature

Curvature affects properties which are important for defining and computing statistical quantities. In particular, the convergence proofs for common statistical estimation algorithms (Sturm, 2003; Miller et al., 2015; Chakraborty & Vemuri, 2015) often rely on assumptions of bounded curvature. We show that the curvature of Graph Space is unbounded from above according to generalized concepts of curvature from metric geometry (Bridson & Haefliger, 1999). This will motivate our choice of statistics in the following section.

First consider the Graph Space where edges and nodes have real attributes $a: E \rightarrow \mathbb{R}$. We will show that the curvature of this Graph Space is not bounded from above. 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 $M_\kappa = \mathbb{S}_\kappa^2$. An important property of the model spaces of curvature κ is that they are each accompanied by a *diameter* D_κ , such that any two points $a, b \in B(x, D_\kappa/2) = \{x' \in M_\kappa : d_{M_\kappa}(x', x) < D_\kappa/2\}$ 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 diameter $< 2D_\kappa$. 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 4 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. 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*

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

for every such x , and similarly for any x on γ_{ab} or γ_{ac} , then the geodesic triangle abc satisfies the $CAT(\kappa)$ condition. The metric space $\mathfrak{X}$ is a $CAT(\kappa)$ space if any geodesic triangle abc in $\mathfrak{X}$ of perimeter $< 2D_\kappa$ satisfies the $CAT(\kappa)$ 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 $CAT(\kappa)$.

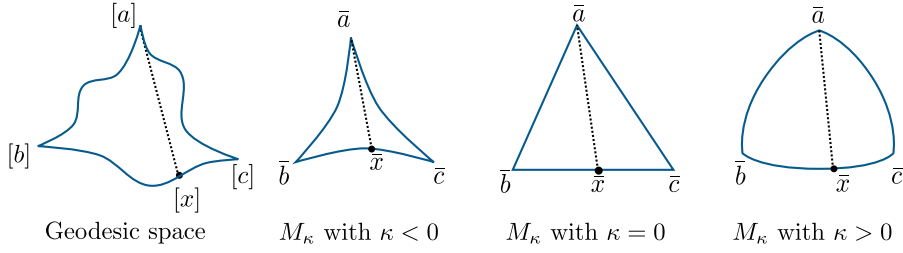


Fig. 4. **Left:** A geodesic triangle in $\mathfrak{X}$. **Right:** the corresponding comparison triangles in hyperbolic space $\mathbb{H}^2$, the plane $\mathbb{R}^2$ and the sphere $\mathbb{S}^2$, respectively.

The properties of the model space in relation to its diameter D_κ also transfer to geodesic spaces $\mathfrak{X}$ which are $CAT(\kappa)$ (Bridson & 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.

THEOREM 1 (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, there exist graphs $[x] \in X/T$ such that for any $R_\kappa > 0$, we can find two graphs $[y_\epsilon], [z_\epsilon] \in B([x], R_\kappa) \subset X/T$ connected by two geodesics.*

Proof. See Appendix B. $\square$

Remark 2. 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 ($A = \mathbb{R}^p$).

4. STATISTICS ON GRAPH SPACE

4.1. The Fréchet Mean

Even though the Graph Space is not a manifold and its curvature is unbounded from above, we can take advantage of its relation to its Euclidean total space X to define and compute intrinsic Graph Space statistics in a well-defined fashion. 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*

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

The Fréchet mean could be a set with more than one element; in this case we denote any one of its elements by $[\bar{x}]$.

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 5, we obtain as a direct consequence of Lemma 2 and Theorem 1:

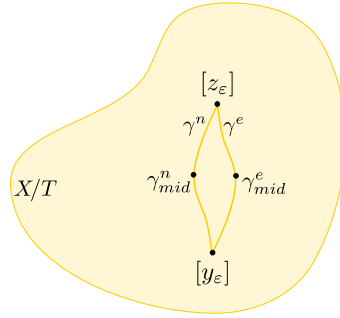


Fig. 5. Since geodesics are not necessarily unique in Graph Space X/T , Fréchet means are not necessarily unique either.

COROLLARY 1. *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.*

Although Fréchet means in Graph Space are not necessarily unique, they are still explanatory and useful – just like on manifolds of positive curvature, where they are also not unique (Huckemann & Eltzner, 2020).

4.2. Geodesic Principal Component Analysis

One approach to defining and computing principal components is to choose node permutations that align each network to a fixed representative of a fixed network, such as the Fréchet mean μ , and performing principal components analysis in X . This is what is done in Guo et al. (2021), and this is analogous to tangent space methods from manifold statistics (Fletcher & Joshi, 2004). However, just like in manifold statistics, the projection to X based on optimal representatives x_i of individual graphs $[x_i]$ with respect to a single graph representative $\mu \in X$, leads to distortion of pairwise distances between observations $d_{X/T}([x_i], [x_j])$, which are not guaranteed to coincide with their distance $d_X(x_i, x_j)$ in the total space X . Thus, in this section, we introduce an intrinsic notion of generalized geodesic principal components (GGPCs) in Graph Space, following Huckemann et al. (2010).

DEFINITION 3. *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 :*

$$\Gamma(X/T) = \{\delta = \pi \circ \gamma : \gamma \in \Gamma(X)\}. \quad (4)$$

Since Graph Space is not an inner product space, we define orthogonality as:

DEFINITION 4. *Two generalized geodesics $\delta_1, \delta_2 \in \Gamma(X/T)$ are orthogonal if they have representatives in $\delta_1 = \pi \circ \gamma_1, \delta_2 = \pi \circ \gamma_2, \gamma_1, \gamma_2 \in \Gamma(X)$ which are orthogonal $\langle \gamma_1, \gamma_2 \rangle_X = 0$.*

In order to bridge computations in Graph Space X/T with computations in the total space X , we introduce a concept of alignment in X . We remind the reader that the metric on Graph Space is

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

which naturally leads to the concept of *optimal position* (Huckemann et al., 2010).

DEFINITION 5 (OPTIMAL POSITION). Given $\tilde{x} \in X$ and $t \in T$, the point $t^T \tilde{x}t$ is in optimal position to $x \in X$ if

$$d_X(t^T \tilde{x}t, x) = d_{X/T}([\tilde{x}], [x]).$$

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

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

The optimal position to a generalized geodesic and its computation is detailed in Algorithm 2. Having concepts of generalized geodesic, optimal position and orthogonality, we now define a set of geodesic principal components:

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

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

$$\delta_1 = \arg \min_{\delta \in \Gamma(X/T)} \sum_{i=1}^k d_{X/T}^2([x_i], \delta) \quad (5)$$

- The second generalized geodesic principal component $\delta_2 \in \Gamma(X/T)$ minimizes (5) 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

$$\sum_{i=1}^k d_{X/T}^2([x_i], [\mu])^2 \quad (6)$$

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

- The j^{th} generalized geodesic principal component is a $\delta_j \in \Gamma(X/T)$ if it minimizes (5) over all generalized geodesics that meet orthogonally $\delta_1, \dots, \delta_{j-1}$ and cross μ .

Remark 3. Since the geodesics are not necessary unique in Graph Space, the Generalized Geodesic Principal Components are non unique as well.

5. ALIGN ALL AND COMPUTE

5.1. AAC for Fréchet Mean

Intrinsic Graph Space statistics can be computed iteratively via a combination of choosing optimal graph representatives in X - not with respect to a single graph but with respect to the wanted statistic - and computing Euclidean statistics in X . Based on this derivation, we propose a general *Align All and Compute* (AAC) strategy for computing Graph Space statistics, which we use to compute both Fréchet means and, in particular, Generalized Geodesic Principal Components.

The distance $d_{X/T}$ between two points in the Graph Space X/T corresponds exactly to the distance in X after bringing one point into optimal position to the other. We now define the AAC

algorithm, which is based on iteratively first bringing the observations into optimal position 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).

Algorithm 1. AAC for the the Fréchet Mean:

Input: $\{[x_1], \dots, [x_k]\} \subset X/T$; a threshold $\varepsilon > 0$
 Initialization: Randomly select $\tilde{x} = \tilde{x}_i \in [x_i] \in \{[x_1], \dots, [x_k]\}$
 While $s > \varepsilon$:
 Obtain $\tilde{x}_i$ optimally aligned wrt $\tilde{x}$ for $i = \{1, \dots, k\}$
 Compute the Fréchet Mean $\bar{x}$ in X of $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\} \in X$
 Compute $s = d_X(\tilde{x}, \bar{x})$
 Set $\tilde{x} = \bar{x}$
 Output: $[\bar{x}]$, an estimate of the Fréchet Mean of $\{[x_1], \dots, [x_k]\} \in X/T$.

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

THEOREM 2. *AAC for Fréchet Mean (Algorithm 1) converges in finite time for $\varepsilon = 0$. 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 $A \subset X/T$, we have $\eta(A) = 0$ if $m(\pi^{-1}(A)) = 0$. Suppose the data set $[x_1], \dots, [x_k]$ is sampled from η ; now with probability 1, the estimator found by Algorithm 1 is a local minimum of the Fréchet function*

$$\sum_{i=1}^k d_{X/T}^2([x], [x_i]). \quad (7)$$

Proof. See Appendix C. □

Even though we show that the algorithm theoretically converges in finite time, we add a convergence threshold ε in the implementation of Algorithm 1. An equally valuable stopping criterion is the difference between the Fréchet function evaluated in two consecutive steps.

All the non invariant models including our algorithm rely on graph matching, which is generally NP complete. In practice, we obtain efficient computation by using faster heuristic algorithms for inexact graph matching such as the Fast Quadratic Assignment by Vogelstein et al. (2015) or Graduate Assignment by Gold & Rangarajan (1996). The graph matching is the computational bottleneck of the procedure.

5.2. AAC for GGPCA

The AAC strategy, due to its iterative alignment with the chosen predictor, helps overcome the non-uniqueness of geodesics. It can readily be extended to other estimators that rely on geodesics such as Generalized Geodesic Principal Components. The alignment with a generalized geodesic (see Definition 5) is performed in a two step alignment procedure:

Algorithm 2. Optimal Position to a generalized geodesic:

Input: A point $x \in [x]$, a straight line $\gamma \in \Gamma(X)$, the domain $[s_{min}, s_{max}]$, k points to sample
 For $\{s_1 = s_{min}, \dots, s_k = s_{max}\}$ sampled from $[s_{min}, s_{max}]$:
 Find $t(s_i) := \arg \min_{t \in T} d_X(t^T x t, \gamma(s_i))$
 Compute $\bar{s} = \arg \min_{s_i \in \{s_1, \dots, s_k\}} d_X(t(s_i)^T x t(s_i), \gamma(s_i))$
 Return: $\bar{t} = t(\bar{s})$
 Output: $\bar{t} \in T$ such that $\bar{t}^T x \bar{t}$ is in optimal position to γ .

The obtained $\bar{t} \in T$ is the permutation such that the point $\bar{t}^T x \bar{t} \in [x]$ is the closest representative of $[x]$ to the geodesic γ in the interval selected - which is $s_i \in [s_{min}, s_{max}]$, discretized into finite steps.

Algorithm 3. AAC to compute the generalized geodesic principal components:

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

Initialization:

Select randomly a candidate $[x_i]$ from $\{[x_1], \dots, [x_k]\}$ and a representative $\tilde{x}_i \in [x_i]$

Align $[x_j]$ wrt $\tilde{x}_i$ obtaining $\tilde{x}_j \in X$ for $j \in \{1, \dots, k\}$

Perform PCA on $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\}$ in X obtaining $\gamma_1, \dots, \gamma_k \in \Gamma(X)$

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

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

While $s > \varepsilon$:

Align $[x_i]$ wrt $\tilde{\delta}_1$ obtaining $\tilde{x}_i \in X$ for $i \in \{1, \dots, k\}$ using Algorithm 2

Perform PCA on $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\}$ in X obtaining $\gamma_1, \dots, \gamma_k \in \Gamma(X)$

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

Compute a step distance function $s = f(\tilde{\delta}_1, \delta_1)$

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

Output: Geodesic principal components $\delta_1, \dots, \delta_k \in \Gamma(X/T)$.

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. A possible choice for the step distance function f is the proportion of variance explained by the first generalized geodesic principal component at the current and the previous step.

THEOREM 3. AAC for GGPCA (Algorithm 3) converges in finite time for $\varepsilon = 0$. 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 $[x_1], \dots, [x_k]$ be sampled from η . Now with probability 1, the estimator of the first GGPC found by Algorithm 1 is a local minimum of the sum of squared residuals function

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

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

Proof. See Appendix D. $\square$

In this paper, we selected as step distance function the proportion of variance explained. For the case of the higher GGPCs, we do not have a proof of local minimization of Equation (5) and Algorithm 3 should be considered a heuristic.

Remark 4. In the case of Fréchet means of shapes, the AAC coincides with Generalized Procrustes Analysis (Gower, 1975). In the statistical literature, a number of algorithms exist to compute means either approximately (Sturm, 2003; Bacák, 2014; Miller et al., 2015; Afsari et al., 2013; Arnaudon et al., 2013; Arnaudon & Miclo, 2014; Bonnabel, 2013; Hauberg et al., 2015; Turner et al., 2014), or via heuristics (Feragen et al., 2011; Billera et al., 2001; Jain & Obermayer, 2008), whose applicability and efficiency vary with the complexity of the underlying nonlinear data space. A particularly 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 & 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; Bacák, 2014; Miller et al., 2015) and certain Riemannian manifolds of bounded curvature (Chakraborty & Vemuri, 2015; Arnaudon et al., 2013; Arnaudon & Miclo, 2014). 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. Moreover, it does not come with an accompanying algorithmic strategy for computing more general statistics such as principal components. This motivates our choice to propose the AAC strategy.

6. COMPARISON WITH EXISTING METHODS

Other strategies have been proposed for the computation of the Fréchet Mean and Principal Components for a set of graphs. In particular, Kolaczyk et al. (2020) focus on the understanding of Graph Space stratified structure and the Fréchet Mean properties when data have the same number of nodes and are restricted to the fundamental domain. Guo et al. (2021) propose an algorithm for the estimation of the Fréchet Mean - similar to the Fréchet Mean AAC - and an algorithm for principal components. Such algorithm applies the Euclidean PCA on the set of graphs aligned with the Fréchet Mean estimation and it recalls a Tangent Space approach (Fletcher, 2011). In the following illustrative examples, we compare the performance of the Tangent PCA method proposed in Guo et al. (2021) and the AAC GGPCA - Algorithm 3. The example is summarized in Figure 6. Consider a weighted directed graph with two nodes - this dimension is driven by the possibility of visualizing the results. The permutation group acts on the nodes, resulting in equivalence classes of dimension 2 if the edges have different values. The edge weights $edge_1$ and $edge_2$ are simulated from two independent Gaussian distributions - Figure 6 (b). In Experiment 1, the edges' distributions have different means and different variances, while in the Experiment 2 the same mean and different variances. As the graphs are considered unlabelled, we randomly choose half of the graphs to be permuted, obtaining the unlabelled datasets in Figure 6 (c). To perform Tangent PCA, firstly the Fréchet Mean is computed to find an optimally aligned dataset and secondly the Euclidean PCA is computed on the aligned dataset - Figure 6 (d). As described in Algorithm 3, AAC GGPCA are computed starting from the unlabelled dataset, without any pre-alignment step - Figure 6 (e). In Experiment 1, the tangent PCA and the AAC GGPCA are equivalent, as all the data points belongs to the fundamental domain of the space (Kolaczyk et al., 2020). In this context, the tangent space approximation happens to be a good approximation of the space. In Experiment 2, the Tangent PCA is capturing the 72.98% of the variability - Figure 6 (d) - while the GGPCA is capturing 95.84% of the data variability. This last example shows how the alignment procedure based on the geodesic alignment rather than a central point alignment is able to capture the complexity of the data structure when the final aim is to fit a geodesic - Figure 6 (e). In real world data, understanding if the data belongs to the fundamental domain is far from trivial (Kolaczyk et al., 2020), thus the AAC GGPCA is a safe model choice, ensuring a correct fit in the whole space and for graphs with different number of nodes.

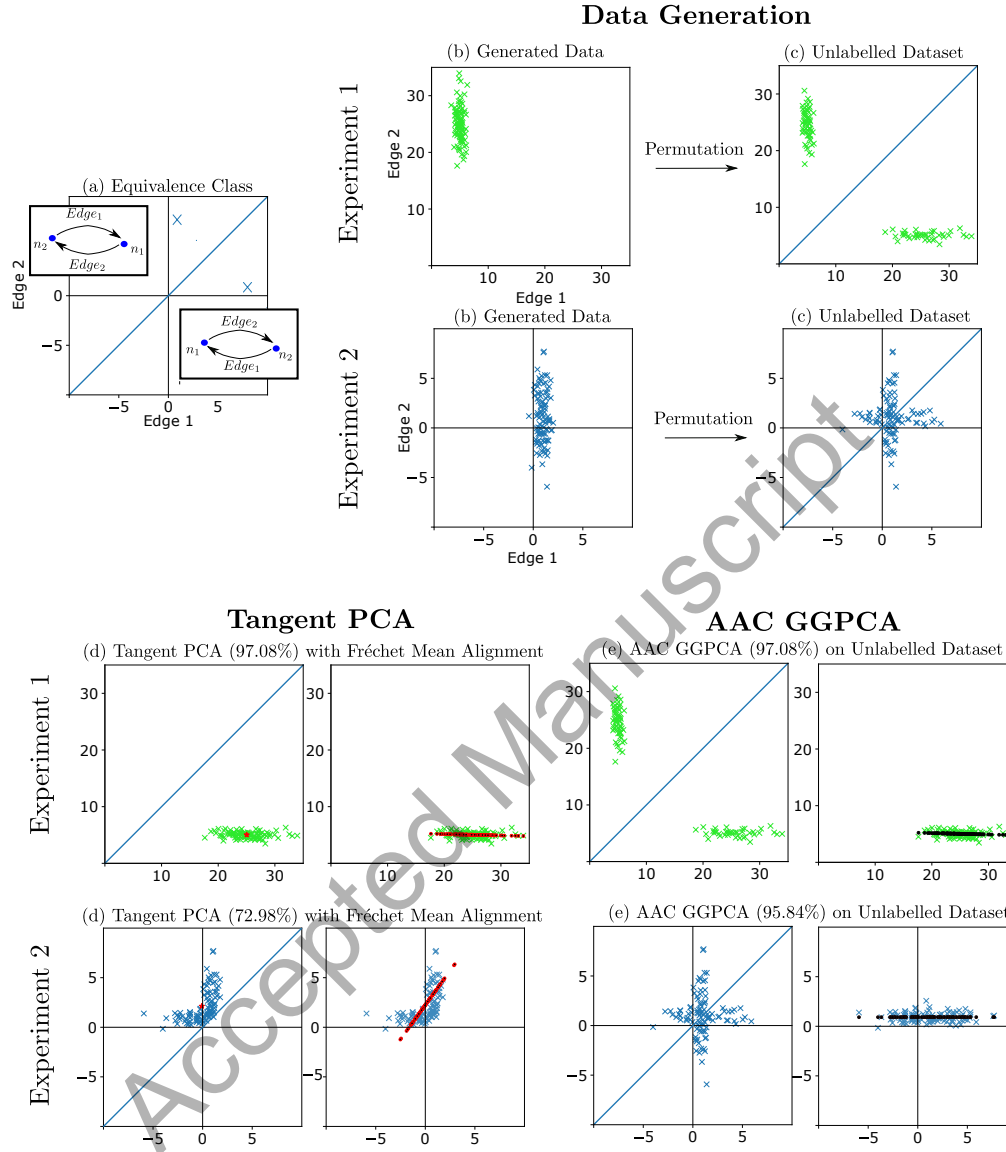


Fig. 6. Comparison between PCA (Guo et al., 2021) and AAC GPCA. FAQ (Vogelstein et al., 2015) matcher is used for both simulations. We sampled 100 data points from: Top Simulation $Edge_1 \sim \mathcal{N}(5, 0.3)$, $Edge_2 \sim \mathcal{N}(25, 7)$; Bottom Simulation $Edge_1 \sim \mathcal{N}(1, 0.3)$, $Edge_2 \sim \mathcal{N}(1, 7)$

7. EXPERIMENTS ON REAL DATA

7.1. Case study 1: Mammals Grooming

This case study focuses on the social interaction of different groups of baboons, collected to observe the effect when an alpha or beta male is knocked out of the community. Social interaction is here defined as grooming interaction between members of the group. The dataset consists of a set of 26 social interaction networks between baboons, where nodes are baboons and edges are the number of grooming interactions between baboons. The networks are naturally unlabelled

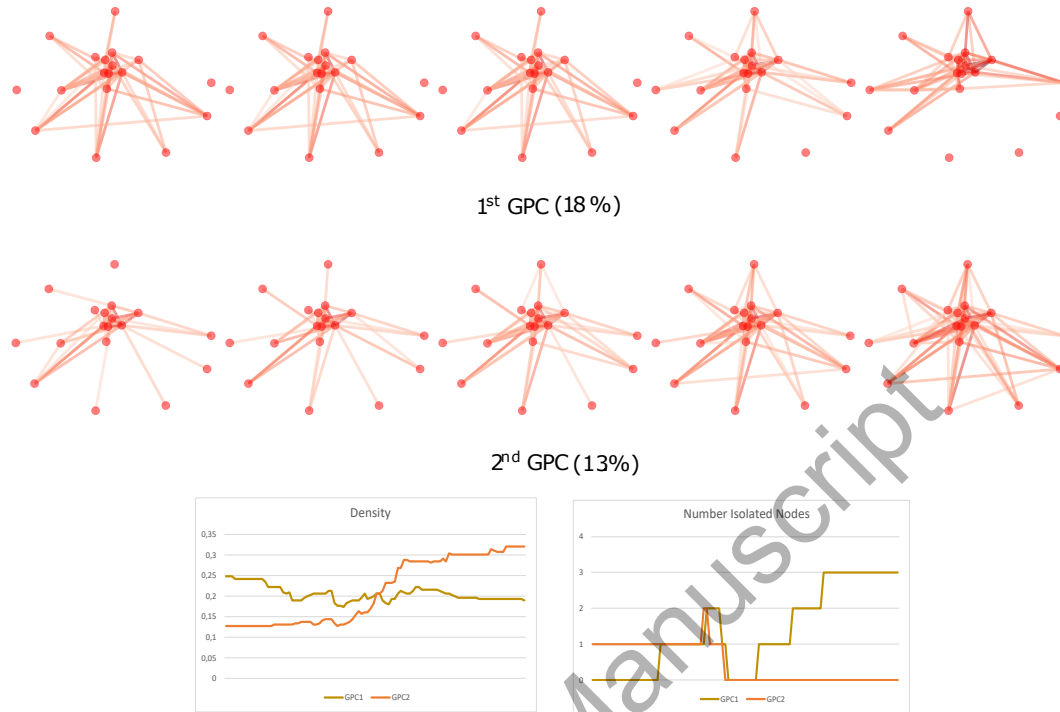


Fig. 7. Data projected along the first two geodesic principal components (0.1, 0.25, 0.5, 0.75, 0.9 percentiles). Graph density along the components as well as the number of isolated nodes along the components are reported. Position of the nodes is computed using the spring layout on the GG-PCA mean.

because the baboons in each network are different, belonging to different groups. Data is available at Rossi & Ahmed (2015), while more details about the case study can be found in Franz et al. (2015). Notice that the original dataset was filtered, keeping only the networks with less than 20 baboons. In Figure 7, we report the first two geodesic principal components along with the density and the number of isolated nodes along the components. Graph density is the number of edges in the graph over the total number of possible edges and an isolated node is a node with degree 0 (i.e. with no connections) (Wasserman & Faust, 1994). The first geodesic principal components reflect the variability in terms of isolated nodes, the second principal component in terms of density. The example is able to show how the GPC analysis in the Graph Space setting is capturing social interaction changes that the original paper (Franz et al., 2015) is not able to capture by just looking at the networks indicators.

7.2. Case study 2: Undirected Networks with Vector Attributes

As an intuitive visual example with real data and associated vectors attributes, we subsample 20 cases of the letter “A” from the well known hand written letters dataset (Kersting et al., 2016; Riesen & Bunke, 2008). As shown in the left panel of Figure 8, every network has node attributes consisting of the node x - and y -coordinates, and binary (0/1) edge attributes indicating whether

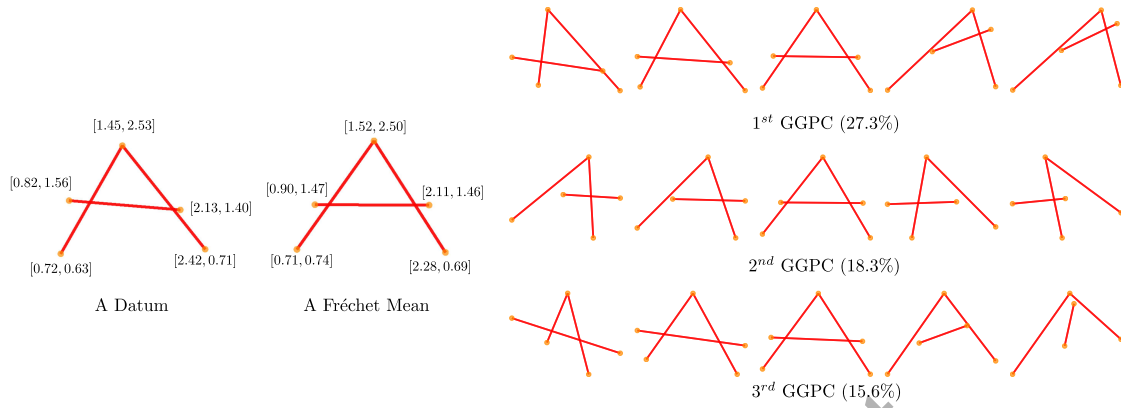


Fig. 8. **Left:** A datum extracted from the A dataset. Every unlabelled node has a bi-dimensional real-valued attribute, while every edge has a 0,1 attribute. The Fréchet mean. **Right:** Visualization of the GGPCs. 0.1, 0.25, 0.5, 0.75, 0.9 quantile of the projected scores are shown for the first three GGPCs.

nodes are connected by lines. In Figure 8, the Fréchet Mean is shown, underlining how the framework is capturing both the topology and the node coordinates. In the right panel of Figure 8 plots network variation along the three GGPCs. 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 horizontal bar, the angle between the vertical bars, and the reciprocal positions of the bars. Notice also that, while the first two GGPCs both capture a trend of slanting letters, they differ in whether the slanting is caused by a translated position in the top node and a rotation of the cross-bar of the letter “A”, or by a rotation in its legs.

7.3. Case study 3: Directed Networks with Vector Attributes

Our final example uses a mobility dataset from Open Data Regione Lombardia (Lombardia Regione, 2019). 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 $\mathbb{R}^4$.

The top panel in Figure 9 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). 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 generalized 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 though the provinces information is not stored in the graph, the provinces 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

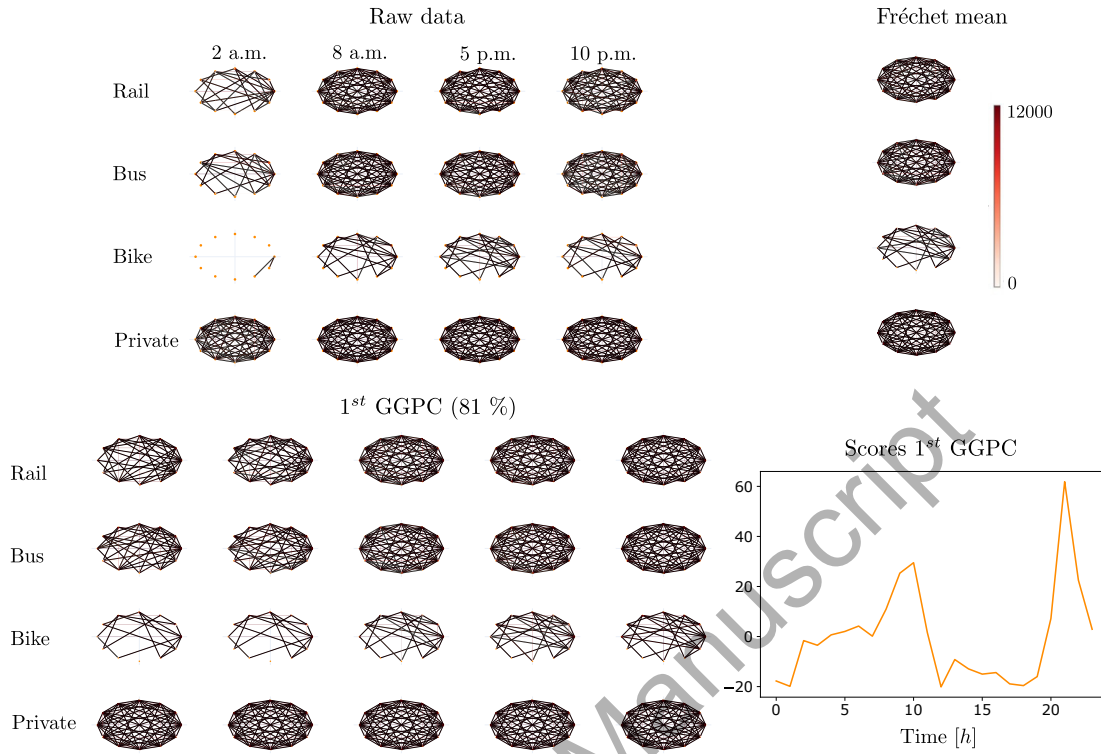


Fig. 9. **Top left:** Network of the fluxes between the Lombardia Region provinces at 2 a.m., 8 a.m., 5 p.m. and 10 p.m.. **Top right:** Fréchet Mean in Graph Space X/T for the mobility modes bike, bus, rail and private. **Bottom Left:** GGPCA performed on scaled data. The first GGPC is shown, by plotting the quantiles of the original data projected along the corresponding generalized geodesic for each one of the levels ($q = 0.1, 0.25, 0.5, 0.75, 0.9$). The GGPC captures the density change along the day for all the transportation modes except from Private car mode. **Bottom Right:** Scores of 1st GGPC plotted along time.

the algorithm). The bottom right panel in Figure 9 we show that the majority of the variability is explained by the first GGPC (81% of the total variance). By moving along this 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 displayed in the bottom right panel in the Figure 9, the scores reflect an expected time pattern, showing the high explanatory power of the GGPCA: two peaks corresponding to the morning and evening activities, one small peak at lunch, and low activity at night.

8. DISCUSSION AND CONCLUSION

The initial aim of this work is to introduce and define a concept of principal components to a population of unlabelled network valued data. In this context, principal component analysis is a useful tool to summarize the complexity of graph valued data and to identify the major topological and attribute variations. The chosen embedding quotient space - Graph Space - allows for a meaningful description of networks weighted, unweighted, directed, undirected, with scalar

and vector attributes. To perform statistics on the Graph Space, we studied the geometrical properties of the Graph Space such as its curvature, which we prove to be unbounded from above, and the corresponding non-uniqueness of its geodesics. Based on these geometrical results, we introduce an intrinsic notion of generalized geodesic principal components (GGPCs). However, computing statistics on Graph Space is not straightforward in practice, due to the problems induced by its unbounded curvature. As a solution, we propose a general algorithm (Align All and Compute Algorithm - AAC) useful to compute statistics on the Graph Space. We compare the AAC with tangent-like approach, showing how the proposed iterative algorithm combined with the suitable alignment procedure reduces the bias of the estimator everywhere in Graph Space, without constraints on the data support. We show the practical utility of the introduced framework on three case studies: three real world examples including both directed and undirected networks, as well as both scalar and vector attributes showing how AAC for Fréchet Mean and GGPCs estimates the expected theoretical results. The examples emphasize that GGPCA is an efficient tool to decompose the complexity and visualize the variability in a population of networks. A first development of the current work could focus on proving consistency and unbiasedness of GPCs within Graph Space, following existing results regarding the Fréchet mean Kolaczyk et al. (2020); Jain & Obermayer (2010b,a) and more general discussions on asymptotic bias and stability within similar quotient spaces (Miolane et al., 2017; Huckemann, 2012). About the uniqueness of the GGPCA, the results of (Kolaczyk et al., 2020) on the fundamental domain can be potentially extended to the geodesics and the GGPCA estimator. In more general terms, AAC is a strategy which can be extended to different estimators, by correctly combining the estimation step with the alignment step (Calissano et al., 2022). A further development in this direction is the extension of the AAC Algorithm to predictive models tailored for graphs.

Accepted Manuscript

A. APPENDIX: EXAMPLE OF GRAPH SPACE CONSTRUCTION

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

Example 1. Consider the two weighted networks shown in Figure 10. To represent these as points in Graph Space, first add a fictional null node to the first graph, then randomly enumerate the nodes, and finally represent them in two weighted, symmetric adjacency matrices as shown in Figure 10. The adjacency matrices can be vectorized as a vector of dimension nine (e.g the first network 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 $3!$. As the graphs are undirected and with only scalar attributes on the edges, we can represent them as points in $\mathbb{R}^2$. The permutation group acting on x_1, x_2 has the effect of inverting the nodes ordering (i.e., the axes). In Figure 10, we show the generalized geodesics connecting the two graphs.

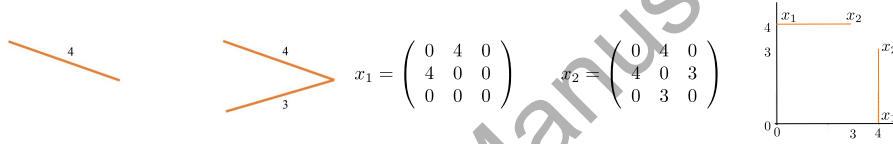


Fig. 10. Example of two simple weighted undirected networks x_1 and x_2 , their matrix representations, and the generalized geodesic connecting the two points.

B. PROOF OF SECTION 3

Proof of Lemma 1. 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 & 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). Any Cauchy sequence $([x_i])_{i \in \mathbb{N}}$ in X/T is the image under π of a Cauchy sequence $(t_i^T x_i t_i)_{i \in \mathbb{N}}$ in X such that for some $M \in \mathbb{N}$, and for $i, j \geq M$, we have $d_X(t_i^T x_i t_i, t_j^T x_j t_j) = d_{X/T}([x_i], [x_j])$. Since X is complete, the sequence $(t_i^T x_i t_i)_{i \in \mathbb{N}}$ 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 & Haefliger, 1999, Proposition I.3.7). $\square$

Proof of Lemma 2. We give an example in the case where node- and edge attributes are scalar. The same example can be adapted to vector valued node- or edge attributes by appending this scalar value with zeros. Consider $\varepsilon > 0$, $c = \sqrt{2}$ and the Euclidean metric on X . Consider the two graphs y_ε and z_ε shown in the top row on the left hand side of Figure 11. The corresponding adjacency matrices are

$$y_\varepsilon = \begin{pmatrix} 0 & 1+\varepsilon & 1-\varepsilon \\ 1+\varepsilon & c\varepsilon & 0 \\ 1-\varepsilon & 0 & -c\varepsilon \end{pmatrix} \quad z_\varepsilon = \begin{pmatrix} 0 & 1+\varepsilon & 1-\varepsilon \\ 1+\varepsilon & -c\varepsilon & 0 \\ 1-\varepsilon & 0 & +c\varepsilon \end{pmatrix}$$

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 of the nodes.

$$\gamma_{mid}^n = \begin{pmatrix} 0 & 1+\varepsilon & 1-\varepsilon \\ 1+\varepsilon & 0 & 0 \\ 1-\varepsilon & 0 & 0 \end{pmatrix} \quad \gamma_{mid}^e = \begin{pmatrix} 0 & 1 & 1 \\ 1 & c\varepsilon & 0 \\ 1 & 0 & -c\varepsilon \end{pmatrix}$$

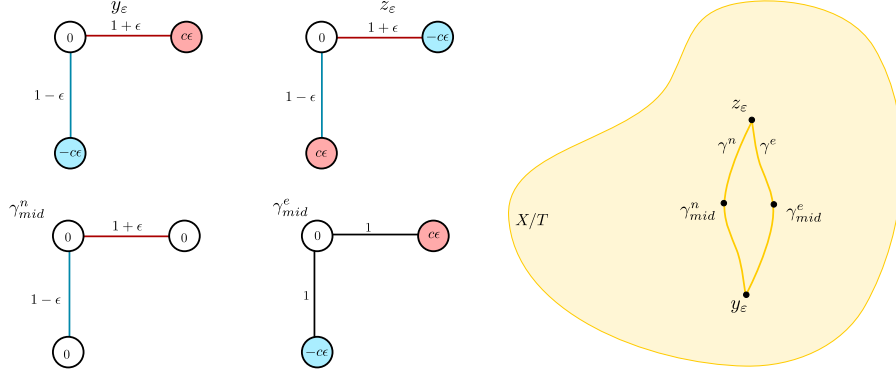


Fig. 11. Depending on their node alignment, the two graphs y_ε and z_ε differ either in node attributes or in edge attributes, and there exist two different geodesics γ^n and γ^e connecting the two graphs: One which interpolates node attributes and one which interpolates edge attributes, respectively. These two geodesics are illustrated via their midpoints γ_{mid}^n and γ_{mid}^e .

The midpoint of this geodesic is the graph γ_{mid}^n shown on the left hand side of the bottom row.

$$d_{X/T}(y_\varepsilon, \gamma_{mid}^n) = d(z_\varepsilon, \gamma_{mid}^n) = 0.2$$

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_{mid}^e shown on the right hand side of the bottom row.

$$d_{X/T}(y_\varepsilon, \gamma_{mid}^e) = d(z_\varepsilon, \gamma_{mid}^e) = 0.2$$

The two points connected by two geodesics are illustrated schematically on the right hand side of Figure 11. The same argument can be built by considering the examples in Figure 1, Figure 2, and Figure 3 $\square$

Proof of Theorem 1. Let $[x]$, $[y_\varepsilon]$ and $[z_\varepsilon]$ be as in Figure 1, Figure 2 or Figure 3. 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 2, there will always be two equally long shortest paths connecting $[y_\varepsilon]$ and $[z_\varepsilon]$. Thus, by (Bridson & 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 .

C. APPENDIX: PROOF OF THEOREM 2

Firstly, we state and prove the following Lemma 3 required to prove Theorem 2.

LEMMA 3. Given representatives $x_1, \dots, x_k$ of $[x_1], \dots, [x_k]$ with mean μ in X , the following holds with probability 1:

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

$$d(\mu, x_i) \neq d(\mu, t^T x_i t),$$

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

Proof of Lemma 3. In order to prove the lemma, we will show that the set

$$\mathcal{X}_T = \left\{ ([x_1], \dots, [x_k]) \in (X/T)^k \mid \begin{array}{l} \text{there exists some } i = 1, \dots, k \text{ and } t \in T \setminus T_{x_i} \\ \text{and some representatives } x_1, \dots, x_k, \\ \text{such that } d(\mu, x_i) = d(\mu, t^T x_i t) \end{array} \right\}$$

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

For each element $t \in T$, denote by $X^t = \{x \in X \mid t^T x = x\}$ the fixed point set of t and define the set
 520 $X_{i \setminus t} = X \times \dots \times \underbrace{X \setminus X^t}_{i^{th}} \times \dots \times X$. Note that $\eta_k(\mathcal{X}_T) = m_k(\pi^{-1}(\mathcal{X}_T))$, and that

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

where

$$\mathcal{X}_{i,t} = \{(x_1, \dots, x_k) \in X_{i \setminus t} \mid d_X(\mu, x_i) = d_X(\mu, t^T x_i t)\} \subset \underbrace{X \times \dots \times X}_k.$$

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

$$f: \underbrace{X \times \dots \times X}_k \rightarrow \mathbb{R}, \quad (x_1, \dots, x_k) \mapsto d_X^2(\mu, x_i) - d_X^2(\mu, t^T x_i t)$$

satisfies

$$f^{-1}(0) \cap X_{i \setminus t} = \mathcal{X}_{i,t}.$$

We show that f is a submersion on $X_{i \setminus t}$ by showing that it has nonzero gradient. We can rewrite

$$\begin{aligned} f(x_1, \dots, x_k) &= \text{tr}\left(\left(\frac{1}{k} \sum_{j=1}^k x_j - x_i\right)^T \left(\frac{1}{k} \sum_{j=1}^k x_j - x_i\right) - \left(\frac{1}{k} \sum_{j=1}^k x_j - t^T x_i t\right)^T \left(\frac{1}{k} \sum_{j=1}^k x_j - t^T x_i t\right)\right) = \\ 525 \quad &= \text{tr}\left(\frac{2}{k} \sum_{j=1}^k \{x_j^T (t^T x_i t) - x_j^T x_i\}\right). \end{aligned}$$

For $j \neq i$ we obtain

$$\nabla_{x_j} f(x_1, \dots, x_k) = \text{tr}\left(\frac{2}{k} (t^T x_i t - x_i)\right)$$

which is nonzero for $(x_1, \dots, x_k) \in X_{i \setminus t}$. It follows that f is a submersion on $X_{i \setminus t}$. As a result, the set

$$f^{-1}(0) \cap X_{i \setminus t} = \mathcal{X}_{i,t}$$

has codimension 1 and, in particular,

$$m_k(\mathcal{X}_{i,t}) = m_k(f^{-1}(0) \cap X_{i \setminus t}) = 0.$$

But then,

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

530 which proves the lemma. $\square$

We can now move to the proof of Theorem 2:

Proof of Theorem 2. First, we prove convergence in finite time. Algorithm 1 consists of two steps repeated iteratively. Consider the function

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

535 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 $[x_i]$. The first step, aligning data points to the current representative of the current mean

estimate, cannot increase the value of (1) as an improved alignment would indeed lower the value of (1). Similarly, the second step, which is the re-estimation of the Fréchet mean given the new alignments, also cannot increase the value of (1) as, again, an improved estimate would lower its value. Moreover, if the value of (1) stays fixed for 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. 540

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 $x_1, \dots, x_k \in X$ 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 $x_1, \dots, x_k$ are also optimally aligned with μ' . In this case, since μ is a local minimizer of (1) 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 (7). 545

As Lemma 3 holds, there exists a $\varepsilon > 0$ and thus we can define:

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

We now set $\varepsilon = \frac{\delta}{2}$ and consider $\mu' \in B_X(\mu, \varepsilon)$. We wish to show that for all $i = 1, \dots, k$ and all $t \in T \setminus T_{x_i}$, we have $d(\mu', x_i) < d(\mu', t^T x_i t)$, namely that the optimal representative of $[x_i]$ is left unchanged for all i . 550

By the definition of δ , we have

$$d(\mu, x_i) \leq d(\mu, t^T x_i t) - \delta,$$

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 555

$$d(\mu, t^T x_i t) \leq \underbrace{d(\mu, \mu')}_{< \varepsilon = \frac{\delta}{2}} + d(\mu', t^T x_i t).$$

We compute

$$\begin{aligned} d(\mu', x_i) &\leq \underbrace{d(\mu', \mu)}_{< \frac{\delta}{2}} + d(\mu, x_i) < \\ &< \frac{\delta}{2} + d(\mu, x_i) \leq \frac{\delta}{2} + d(\mu, t^T x_i t) - \delta < -\frac{\delta}{2} + \frac{\delta}{2} + d(\mu', t^T x_i t) = d(\mu', t^T x_i t), \end{aligned}$$

which completes the proof of Theorem 2 under the assumption that Lemma 3 holds. □

D. APPENDIX : PROOF OF THEOREM 3

Proof Theorem 3. The proof of convergence of AAC for GGPCA follows the exact same strategy as for the Fréchet mean in Theorem 2, noting: 560

- 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 GGPC $\delta \in \Gamma(X/T)$ and its representative $\gamma \in \Gamma(X)$.

- For the first GGPC, the algorithm converges to a local minimum of (5) 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 . $\square$

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., TRON, R. & VIDAL, R. (2013). On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM J. Control and Optimization* **51**, 2230–2260.
- ARNAUDON, M., BARBARESCO, F. & YANG, L. (2013). Medians and means in Riemannian geometry: Existence, uniqueness and computation. In *Matrix Information Geometry*. Springer, pp. 169–197.
- ARNAUDON, M. & MICLO, L. (2014). A stochastic algorithm finding generalized means on compact manifolds. *Stochastic Processes and their Applications* **124**, 3463–3479.
- BACÁK, M. (2014). Computing medians and means in Hadamard spaces. *SIAM Journal of Optimization* **24**, 1542–1566.
- BILLE, P. (2005). A survey on tree edit distance and related problems. *Theoretical computer science* **337**, 217–239.
- BILLERA, L. J., HOLMES, S. P. & VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics* **27**, 733–767.
- BONNABEL, S. (2013). Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions on Automatic Control* **58**, 2217–2229.
- BREDON, G. E. (1972). *Introduction to compact transformation groups*, vol. 46. Academic press.
- BRIDSON, M. R. & HAEFLIGER, A. (1999). *Metric spaces of non-positive curvature*. Springer.
- BUNKE, H. & RIESEN, K. (2011). Improving vector space embedding of graphs through feature selection algorithms. *Pattern Recognition* **44**, 1928–1940.
- CALISSANO, A., FERAGEN, A. & VANTINI, S. (2022). Graph-valued regression: Prediction of unlabelled networks in a non-euclidean graph space. *Journal of Multivariate Analysis* **190**, 104950.
- CHAKRABORTY, R. & VEMURI, B. C. (2015). Recursive Fréchet mean computation on the Grassmannian and its applications to computer vision. In *Proceedings of the IEEE International Conference on Computer Vision*.
- CHOWDHURY, S. & MÉMOLI, F. (2019). The gromov–wasserstein distance between networks and stable network invariants. *Information and Inference: A Journal of the IMA* **8**, 757–787.
- DUNCAN, A., KLASSEN, E. & SRIVASTAVA, A. (2018). Statistical shape analysis of simplified neuronal trees. *The Annals of Applied Statistics* **12**, 1385–1421.
- DURANTE, D., DUNSON, D. B. & VOGELSTEIN, J. T. (2017). Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association* **112**, 1516–1530.
- DUVENAUD, D. K., MACLAURIN, D., AGUILERA-IPARRAGUIRRE, J., GÓMEZ- BOMBARELLI, R., HIRZEL, T., ASPURU-GUZZIK, A. & ADAMS, R. P. (2015). Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems*.
- FERAGEN, A., HAUBERG, S., NIELSEN, M. & LAUZE, F. (2011). Means in spaces of tree-like shapes. In *IEEE International Conference on Computer Vision (ICCV)*.
- FERAGEN, A., LAUZE, F., LO, P., DE BRUIJNE, M. & NIELSEN, M. (2010). Geometries on spaces of treelike shapes. In *Asian Conference on Computer Vision*. Springer.
- FERAGEN, A. & NYE, T. M. W. (2020a). δ -statistics on stratified spaces. In *Riemannian Geometric Statistics in Medical Image Analysis*, X. Pennec, S. Sommer & P. T. Fletcher, eds. Academic Press, pp. 299 – 342.
- FERAGEN, A. & NYE, T. M. W. (2020b). Statistics on stratified spaces. In *Riemannian Geometric Statistics in Medical Image Analysis*. Elsevier, pp. 299–342.
- FERAGEN, A., OWEN, M., PETERSEN, J., WILLE, M., THOMSEN, L., DIRKSEN, A. & DE BRUIJNE, M. (2013). Tree-space statistics and approximations for large-scale analysis of anatomical trees. In *Information Processing in Medical Imaging*, vol. 7917 of *Lecture Notes in Computer Science*. Springer, pp. 74–85.
- FLETCHER, P. T. (2011). Geodesic regression on riemannian manifolds. In *Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods*

- for Modelling Biological Shape Variability.
- FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision* **105**, 171–185. 620
- FLETCHER, P. T. & JOSHI, S. (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer vision and mathematical methods in medical and biomedical image analysis*. Springer, pp. 87–98.
- FRANZ, M., ALTMANN, J. & ALBERTS, S. C. (2015). Knockouts of high-ranking males have limited impact on baboon social networks. *Current zoology* **61**, 107–113.
- GARBA, M. K., NYE, T. M., LUEG, J. & HUCKEMANN, S. F. (2021). Information geometry for phylogenetic trees. *Journal of Mathematical Biology* **82**, 1–39. 625
- GINESTET, C. E., LI, J., BALACHANDRAN, P., ROSENBERG, S. & KOLACZYK, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics* **11**, 725–750.
- GOLD, S. & RANGARAJAN, A. (1996). A graduated assignment algorithm for graph matching. *IEEE Transactions on pattern analysis and machine intelligence* **18**, 377–388. 630
- GOWER, J. C. (1975). Generalized Procrustes analysis. *Psychometrika* **40**, 33–51.
- GUO, X., SRIVASTAVA, A. & SARKAR, S. (2021). A quotient space formulation for generative statistical analysis of graphical data. *Journal of Mathematical Imaging and Vision* **63**, 735–752.
- HAUBERG, S., FERAGEN, A., ENFICIAUD, R. & BLACK, M. J. (2015). Scalable robust principal component analysis using Grassmann averages. *IEEE transactions on pattern analysis and machine intelligence* **38**, 2298–2311. 635
- HUCKEMANN, S. & ELTZNER, B. (2020). Statistical methods generalizing principal component analysis to non-euclidean spaces. *Handbook of variational methods for nonlinear geometric data*, 317–338.
- HUCKEMANN, S., HOTZ, T. & MUNK, A. (2010). Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statist. Sinica* **20**, 1–58.
- HUCKEMANN, S. F. (2012). On the meaning of mean shape: manifold stability, locus and the two sample test. *Annals of the Institute of Statistical Mathematics* **64**, 1227–1259. 640
- JAIN, B. & OBERMAYER, K. (2008). On the sample mean of graphs. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE.
- JAIN, B. & OBERMAYER, K. (2010a). Consistent estimator of median and mean graph. In *2010 20th International Conference on Pattern Recognition*. IEEE. 645
- JAIN, B. J. & OBERMAYER, K. (2009). Structure spaces. *Journal of Machine Learning Research* **10**, 2667–2714.
- JAIN, B. J. & OBERMAYER, K. (2010b). Large sample statistics in the domain of graphs. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer.
- KENDALL, D. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society* **16**, 81–121. 650
- KERSTING, K., KRIEGE, N. M., MORRIS, C., MUTZEL, P. & NEUMANN, M. (2016). Benchmark data sets for graph kernels.
- KOLACZYK, E. D., LIN, L., ROSENBERG, S., WALTERS, J., XU, J. et al. (2020). Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *The Annals of Statistics* **48**, 514–538. 655
- LOMBARDIA REGIONE (2019). Open Data Regione Lombardia. Accessed: 2019-09-02.
- LUEG, J., GARBA, M. K., NYE, T. M. & HUCKEMANN, S. F. (2021). Wald space for phylogenetic trees. In *International Conference on Geometric Science of Information*. Springer.
- LUNAGÓMEZ, S., OLHEDE, S. C. & WOLFE, P. J. (2020). Modeling network populations via graph distances. *Journal of the American Statistical Association*, 1–18. 660
- MALLASTO, A. & FERAGEN, A. (2018). Wrapped Gaussian process regression on Riemannian manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- MARRON, J. S. & ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biometrical Journal* **56**, 732–753.
- MILLER, E., OWEN, M. & PROVAN, J. S. (2015). Polyhedral computational geometry for averaging metric phylogenetic trees. *Advances in Applied Mathematics* **68**, 51 – 91. 665
- MIOLANE, N., GUIGUI, N., LE BRIGANT, A., MATHE, J., HOU, B., THANWERDAS, Y., HEYDER, S., PELTRE, O., KOEP, N., ZAATITI, H. et al. (2020). Geomstats: a python package for riemannian geometry in machine learning. *Journal of Machine Learning Research* **21**, 1–9.
- MIOLANE, N., HOLMES, S. & PENNEC, X. (2017). Template shape estimation: correcting an asymptotic bias. *SIAM Journal on Imaging Sciences* **10**, 808–844. 670
- 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**, 304–315. 675
- NYE, T. M. W., TANG, X., WEYENBERG, G. & YOSHIDA, R. (2017). Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees. *Biometrika* **104**, 901–922.
- PENNEC, X., FILLARD, P. & AYACHE, N. (2006). A Riemannian framework for tensor computing. *International Journal of Computer Vision (IJCV)* **66**, 41–66.

- 680 RIESEN, K. & BUNKE, H. (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)*. Springer.
- ROSSI, R. A. & AHMED, N. K. (2015). The network data repository with interactive graph analytics and visualization. In *AAAI*.
- 685 SEVERN, K. E., DRYDEN, I. L. & PRESTON, S. P. (2021). Non-parametric regression for networks. *Stat* **10**, e373.
- SHERVASHIDZE, N., SCHWEITZER, P., VAN LEEUWEN, E. J., MEHLHORN, K. & BORGWARDT, K. M. (2011). Weisfeiler-Lehman graph kernels. *Journal of Machine Learning Research* **12**, 2539–2561.
- SIMPSON, S. L., LYDAY, R. G., HAYASAKA, S., MARSH, A. P. & LAURIENTI, P. J. (2013). A permutation testing framework to compare groups of brain networks. *Frontiers in computational neuroscience* **7**, 171–184.
- 690 SRIVASTAVA, A., JOSHI, S. H., MIO, W. & LIU, X. (2005). Statistical shape analysis: Clustering, learning, and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* **27**, 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., MILEYKO, Y., MUKHERJEE, S. & HARER, J. (2014). Fréchet means for distributions of persistence diagrams. *Disc. Comp. Geom.* **52**, 44–70.
- 695 VOGELSTEIN, J. T., CONROY, J. M., LYZINSKI, V., PODRAZIK, L. J., KRATZER, S. G., HARLEY, E. T., FISHKIND, D. E., VOGELSTEIN, R. J. & PRIEBE, C. E. (2015). Fast approximate quadratic programming for graph matching. *PLOS one* **10**, e0121002.
- VON FERBER, C., HOLOVATCH, T., HOLOVATCH, Y. & PALCHYKOV, V. (2009). Public transport networks: empirical analysis and modeling. *The European Physical Journal B* **68**, 261–275.
- 700 WANG, H. & MARRON, S. J. (2007). Object oriented data analysis: Sets of trees. *The Annals of Statistics* **35**, 1849–1873.
- WASSERMAN, S. & FAUST, K. (1994). *Social network analysis: Methods and applications*, vol. 8. Cambridge university press.
- 705 ZHANG, M. & FLETCHER, P. T. (2013). Probabilistic principal geodesic analysis. In *Advances in Neural Information Processing Systems* 26, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani & K. Q. Weinberger, eds. Curran Associates, Inc., pp. 1178–1186.