Populations of unlabelled networks

graph space geometry and generalized geodesic principal components

Calissano, Anna; Feragen, Aasa; Vantini, Simone

Published in:
Biometrika

Link to article, DOI:
10.1093/biomet/asad024

Publication date:
2024

Document Version
Peer reviewed version

Biometrika (2023), **00**, 0, *pp.* 1–24 https://doi.org/10.1093/biomet/asad024
Advance Access Publication 4 April 2023

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, 20133 Milano, Italy
anna.calissano@inria.fr

AASA FERAGEN

DTU Compute, Technical University of Denmark, Richard Petersens Plads, 2800 Kongens Lyngby, Denmark
afhar@dtu.dk

and SIMONE VANTINI

MOX Department of Mathematics, Politecnico di Milano, Via Bonardi 9, 20133 Milano, Italy
simone.vantini@polimi.it

Summary

Statistical analysis for populations of networks is widely applicable, but challenging, as networks have strongly non-Euclidean behaviour. Graph space is an exhaustive framework for studying populations of unlabelled networks that are weighted or unweighted, uni- or multilayered, 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 potential utility of the framework. The whole framework is implemented within the `geomstats` Python package.

Some key words: Computational geometric statistic; Geodesic principal component; Graph space; Network data; Object data; Quotient space.

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 analysing 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 (Wang & Marron, 2007; Feragen et al., 2013) 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 reframed 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, including all those applications where the node matching is nontrivial, e.g., social networks of different individuals, trade networks 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) and 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 and so on. 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) and Severn et al. (2021) proposed a model where networks' Laplacian matrices are smoothly injected into a submanifold of a Euclidean space; Simpson et al. (2013) and Durante et al. (2017) faced the problem of generating and performing tests on a population of networks; Lunagómez et al. (2020) provided Bayesian modelling for discrete labelled graphs and Chowdhury & Mémoli (2019) studied 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 numbers of nodes, and with unlabelled nodes. Within graph space, networks can also be weighted or unweighted, uni- or multilayer, directed or undirected, and with different types of attributes on both edges and nodes. The same graph space independently appears in the work of Kolaczyk et al. (2020), who studied the behaviour of the Fréchet mean in the fundamental domain of graph space, as well as in the work of Guo et al. (2021), who proposed a simpler algorithm for principal components, analogous to the tangent space approaches known from manifold statistics (Fletcher & Joshi, 2004).

In this paper we describe graph space and the nonlinear nature of its geometry, and introduce an intrinsic version of geodesic principal components for a population of unlabelled networks. We propose the align-all-and-compute statistical algorithm to compute estimators on the graph space, and prove that it converges to a local minimum in finite time. 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. To demonstrate the potential utility and flexibility of this framework we illustrate our algorithms using three different datasets. All the proposed theory is implemented in the geometry module `GraphSpace` and learning module `AAC` available within the `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 $a\colon E \to A$. Here, the map a allows us to describe attributes on both edges and nodes, as we use self-loop edges to assign attributes to nodes. From here on, we represent networks mathematically as graphs, and consider these terms equivalent.

In our modelling we 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. Group T of node permutations can be represented by permutation matrices and its action on X in conjugation,

$$\cdot\colon T \times X \to X, \qquad (t, x) \mapsto t^{\mathrm{T}} x t,$$

thus defines an action of 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 $[x] = \{t^{\mathrm{T}} x t,\ t \in T\}$ that contains all the adjacency matrices obtained from x by permuting nodes. The canonical projection to graph space X/T is defined as

$$\pi : X \to X/T, \qquad \pi(x) := [x].$$

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 in Appendix 1.

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. 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

$$d_{X/T}([x_1],[x_2]) = \min_{t\in T} d_X(t^{\mathrm{T}}x_1t, x_2).$$

The resulting quotient metric $d_{X/T}$ is indeed a nondegenerate 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 $T_x = \{t \in T \mid t^{\mathrm{T}}xt = x\}$ 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 $t \in T$ that 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 (Kendall, 1984; Fletcher & Joshi, 2004; Srivastava et al., 2005; Pennec et al., 2006; Huckemann et al., 2010; Fletcher, 2013; Zhang & Fletcher, 2013; Mallasto & Feragen, 2018) are unfortunately not directly applicable to graph space.

However, as is commonly the case in both general nonlinear statistics (Sturm, 2003; Feragen et al., 2011, 2013; Nye, 2011, 2014; Bacák, 2014; Turner et al., 2014; Miller et al., 2015; Nye et al., 2017; Duncan et al., 2018; Feragen & Nye, 2020) and in manifold statistics, we utilize geodesics, or shortest paths, to define and compute statistical properties in graph space. In the absence of a manifold structure, we 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 $\gamma\colon [0,1] \to \mathfrak{X}$ is

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

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

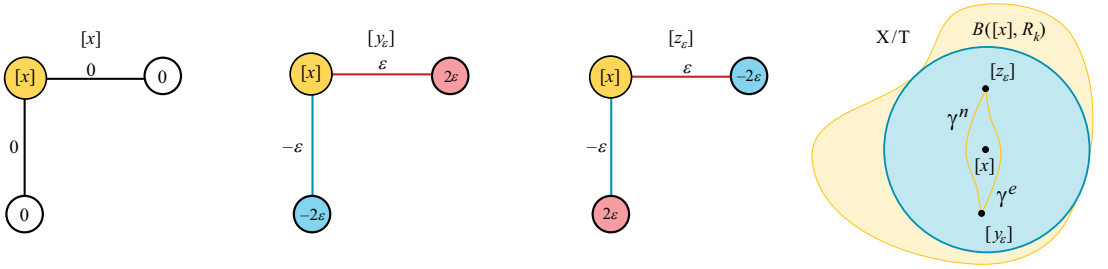


Fig. 1. For any graph $[x]$ with at most $n-2$ nodes, we can find two graphs $[y_\epsilon]$ and $[z_\epsilon]$ arbitrarily close to it, which are connected by more than one geodesic.

space if, for every two points $a, b \in \mathfrak{X}$, we have

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

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 the following lemma.

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

Proof of Lemma 1 can be found in the Appendix along with proofs of all subsequent Lemmas and Theorems.

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 Fig. 1, shows how near a graph $[x]$ with at most $n-2$ nodes you can find a pair of points arbitrarily close to it that are connected by two geodesics. We formalize this as a lemma.

LEMMA 2. *There exist graphs* $[x] \in X/T$ *such that, for any* $\varepsilon > 0$*, there exist points* $y_\varepsilon, z_\varepsilon \in B([x], \varepsilon) \subset X/T$ *that 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.*

The first counterexample leading to Lemma 2 can be constructed near any graph $[x]$ with at most $n-2$ nodes. Are nonunique geodesics caused by using dummy nodes with 0 attributes to include graphs with less than n nodes in graph space? First, note that graphs $[y_\varepsilon]$ and $[z_\varepsilon]$ 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 Fig. 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, nonunique geodesics can also be found arbitrarily close to certain graphs with n nodes. Third, the reader might wonder whether these nonuniqueness 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 Fig. 3, this is also not the case. Indeed, a completely analogous example can be constructed by building graph $[x]$ from a subgraph $[x']$ with at most $n-4$ nodes, using additional edges to play the role taken on by nodes with weights $k-2\varepsilon$ and $k+2\varepsilon$ in the previous examples.

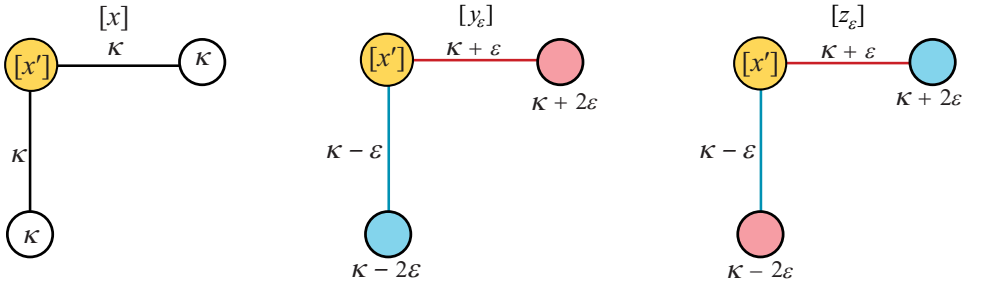


Fig. 2. There exist graphs $[x]$ admitting particular permutation symmetries, such that we can find two graphs $[y_\epsilon]$ and $[z_\epsilon]$ arbitrarily close to $[x]$, which are connected by more than one geodesic. In particular, nonunique geodesics can appear even without having to add zero-attributed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

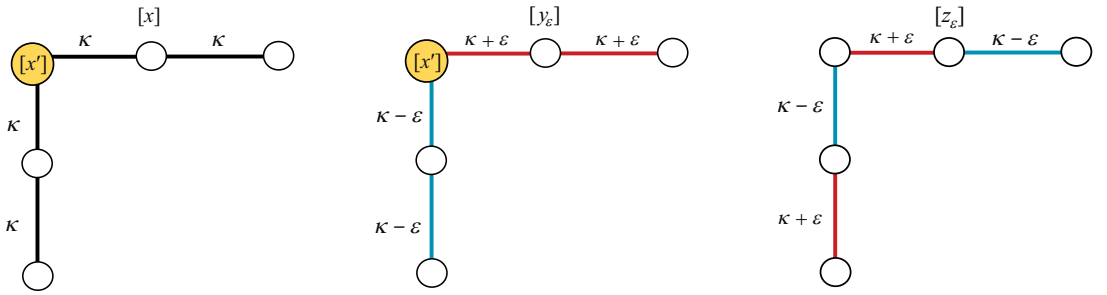


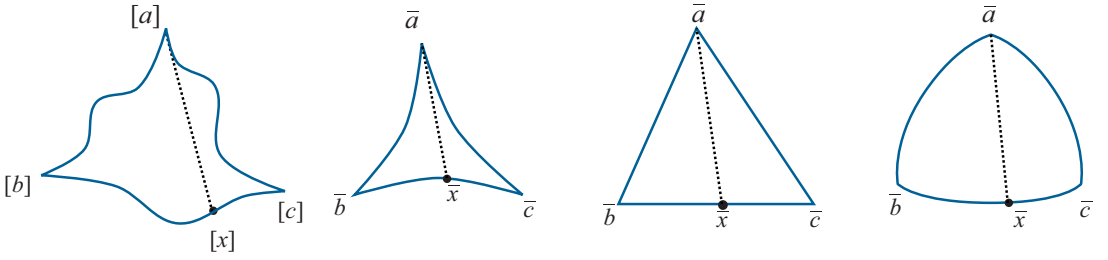
Fig. 3. For graphs without node attributes, there exist graphs $[x]$ such that we can find two graphs $[y_\epsilon]$ and $[z_\epsilon]$ arbitrarily close to $[x]$, by attaching a new edge in place of the weighted node.

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, 2020). In this paper, however, we take the route of defining statistics that do not rely on uniqueness of geodesics.

3.3. *Graph space curvature*

Curvature affects properties that are important for defining and computing statistical quantities. In particular, the convergence proofs for common statistical estimation algorithms (Sturm, 2003; Chakraborty & Vemuri, 2015; Miller et al., 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 $a\colon E \to \mathbb{R}$. We 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, $M_\kappa = \mathbb{S}^2_\kappa$. 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



Geodesic space M_k with $k < 0$ M_k with $k = 0$ M_k with $k > 0$

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

$a, b \in B(x, D_\kappa/2) = \{x' \in M_\kappa : d_{M_\kappa}(x', x) < D_\kappa/2\}$ 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 Fig. 4 for an illustration.

Definition 1 (CAT(κ) 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 segment γ_{bc} has a corresponding point $\bar{x}$ on 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*

$$d_\mathfrak{X}(x, a) \leqslant d_{M_\kappa}(\bar{x}, \bar{a}) \tag{1}$$

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 (1). Geometrically, this means that triangles in $\mathfrak{X}$ are* thinner *than triangles in M_κ. Metric space $\mathfrak{X}$ has* curvature $\leqslant \kappa$ in the sense of Alexandrov *if it is locally* CAT(κ).

The properties of the model space in relation to its diameter D_κ also transfer to geodesic spaces $\mathfrak{X}$ that are CAT(κ) (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 $\kappa \leqslant 0$, meaning that in nonpositively curved spaces, any two points can be joined by a unique geodesic, regardless of their distance. Moreover, $D_\kappa \geqslant \pi/\sqrt{\kappa}$ 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 $\leqslant \kappa$ 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.*

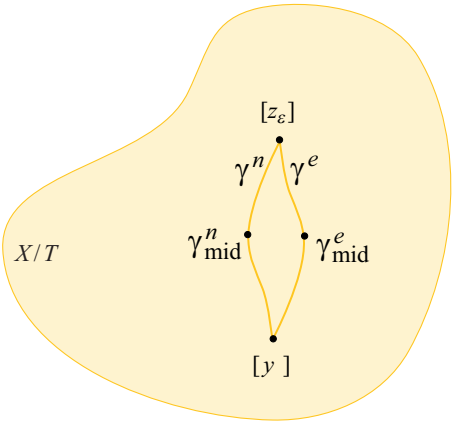


Fig. 5. Since geodesics are not necessarily unique in graph space X/T, Fréchet means are not necessarily unique either.

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

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

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

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

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 Fig. 5, we obtain the following result as a direct consequence of Lemma 2 and Theorem 1.

Corollary 1. *Fréchet means are not generally unique in graph space* X/T. *In fact, as shown by Fig. 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 component analysis in X. This is what was done by 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 in graph space, following Huckemann et al. (2010).

Definition 3. *Denote by $\Gamma(X)$ the set of all straight lines in X. Following Huckemann et al. (2010), a curve δ is a generalized geodesic on graph space X/T if it is a projection of a straight line on X:*

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

Since graph space is not an inner product space, we define orthogonality as follows.

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

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

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 $t^{\mathrm{T}}\tilde{x}t$ is in optimal position to $x \in X$ if*

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

That is, the equivalence class $[\tilde{x}] \in X/T$ contains at least one point $t^{\mathrm{T}}\tilde{x}t \in [\tilde{x}]$ that 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 $t^{\mathrm{T}}xt \in X$ is in optimal position to $\gamma \in \Gamma(X)$ if

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

The optimal position to a generalized geodesic and its computation is detailed in Algorithm 2 below. 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* $\pi : X \to X/T$ *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 geodesic principal components for the set* $\{[x_1], \ldots, [x_k]\}$ *are defined as follows.*

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

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

(ii) The second generalized geodesic principal component $\delta_2 \in \Gamma(X/T)$ *minimizes* (2) *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.

(iii) The point $[\mu] \in X/T$ *is called the principal component mean if it minimizes*

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

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

(iv) The jth generalized geodesic principal component is a $\delta_j \in \Gamma(X/T)$ *if it minimizes* (2) *over all generalized geodesics that meet orthogonally* $\delta_1, \ldots, \delta_{j-1}$ *and cross* μ.

Remark 3. Since the geodesics are not necessarily unique in graph space, the generalized geodesic principal components are nonunique as well.

5. ALIGN-ALL-AND-COMPUTE ALGORITHM

5.1. *Align all and compute for the 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 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 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 align-all-and-compute 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 align-all-and-compute algorithm coincides with generalized Procrustes analysis (Gower, 1975).

Algorithm 1. Align all and compute for the Fréchet mean.

Input: $\{[x_1], \ldots, [x_k]\} \subset X/T$; a threshold $\varepsilon > 0$.
Initialization: Randomly select $\tilde{x} = \tilde{x}_i \in [x_i] \in \{[x_1], \ldots, [x_k]\}$.
While $s > \varepsilon$:

Obtain $\tilde{x}_i$ optimally aligned with respect to $\tilde{x}$ for $i = \{1, \ldots, k\}$.
Compute the Fréchet mean $\bar{x}$ in X of $\{\tilde{x_1}, \tilde{x_2}, \ldots, \tilde{x_k}\} \in X$.
Compute $s = d_X(\tilde{x}, \bar{x})$.
Set $\tilde{x} = \bar{x}$.
Output: $[\bar{x}]$, an estimate of the Fréchet mean of $\{[x_1], \ldots, [x_k]\} \in X/T$.

This algorithm provides an estimate of a Fréchet mean that is independent of the order of the data, and with improved convergence properties.

THEOREM 2. *Algorithm* 1 *converges in finite time for* $\varepsilon = 0$. *Moreover, assume that graph space* X/T *is endowed with a probability measure* η *that 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 that the dataset* $[x_1], \ldots, [x_k]$ *is sampled from* η; *now, with probability* 1, *the estimator found by Algorithm* 1 *is a local minimum of the Fréchet function*

$$\sum_{i=1}^{k} d^2_{X/T}([x], [x_i]). \tag{3}$$

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. *Align all and compute for generalized geodesic principal component analysis*

The align-all-and-compute strategy, due to its iterative alignment with the chosen predictor, helps overcome the nonuniqueness 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 $\{s_1 = s_{\min}, \ldots, s_k = s_{\max}\}$ sampled from $[s_{\min}, s_{\max}]$:
Find $t(s_i) := \arg\min_{t \in T} d_X\{t^{\mathrm{T}} x t, \gamma(s_i)\}$.
Compute $\bar{s} = \arg\min_{s_i \in \{s_1, \ldots, s_k\}} d_X\{t(s_i)^{\mathrm{T}} x t(s_i), \gamma(s_i)\}$.
Return: $\bar{t} = t(\bar{s})$.
Output: $\bar{t} \in T$ such that $\bar{t}^{\mathrm{T}} x \bar{t}$ is in optimal position to γ.

The obtained $\bar{t} \in T$ is the permutation such that the point $\bar{t}^{\mathrm{T}} x \bar{t} \in [x]$ 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. Align all and compute to compute the generalized geodesic principal components.

Input: $\{[x_1], \ldots, [x_k]\} \in X/T$, a threshold $\varepsilon > 0$.
Initialization:
  Select randomly a candidate $[x_i]$ from $\{[x_1], \ldots, [x_k]\}$ and a representative $\tilde{x}_i \in [x_i]$.
  Align $[x_j]$ with respect to $\tilde{x}_i$, obtaining $\tilde{x}_j \in X$ for $j \in \{1, \ldots, k\}$.
  Perform principal component analysis 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 $s > \varepsilon$:
  Align $[x_i]$ with respect to $\tilde{\delta}_1$, obtaining $\tilde{x}_i \in X$ for $i \in \{1, \ldots, k\}$ using Algorithm 2.
  Perform principal component analysis 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 $s = f(\tilde{\delta}_1, \delta_1)$.
  Set $\tilde{\delta}_1 = \delta_1, \ldots, \tilde{\delta}_k = \delta_k$.
Output: geodesic principal components $\delta_1, \ldots, \delta_k \in \Gamma(X/T)$.

Because of the curvature of the space discussed in § 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 previous steps.

THEOREM 3. *Algorithm 3 converges in finite time for $\varepsilon = 0$. Assume that graph space X/T is endowed with a probability measure η that is absolutely continuous with respect to the push-forward 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 generalized geodesic principal component found by Algorithm 1 is a local minimum of the sum of the squared residuals function*

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

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

In this paper, we selected as the step distance function the proportion of variance explained. For the case of the higher generalized geodesic principal components, we do not have a proof of local minimization of (2) and Algorithm 3 should be considered a heuristic.

Remark 4. In the case of Fréchet means of shapes, the align-all-and-compute algorithm coincides with generalized Procrustes analysis (Gower, 1975). In the statistical literature, a number of algorithms exist to compute means either approximately (Sturm, 2003; Afsari et al., 2013; Arnaudon et al., 2013; Bonnabel, 2013; Arnaudon & Miclo, 2014; Bacák, 2014; Turner et al., 2014; Hauberg et al., 2015; Miller et al., 2015) or via heuristics (Billera et al., 2001; Jain & Obermayer, 2008; Feragen et al., 2011), 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 that obtains an updated mean estimate by stepping $1/k$ along the geodesic from a current mean estimate to a kth 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 nonpositively curved spaces (Sturm, 2003; Bacák, 2014; Miller et al., 2015) and certain Riemannian manifolds of bounded curvature (Arnaudon et al., 2013; Arnaudon & Miclo, 2014; Chakraborty & Vemuri, 2015). 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 align-all-and-compute 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) focused 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) proposed an algorithm for the estimation of the Fréchet mean, similar to the Fréchet mean align-all-and-compute algorithm, and an algorithm for principal components. Such an algorithm applies the Euclidean principal component analysis on the set of graphs aligned with the Fréchet mean estimation and it recalls a tangent space approach (Fletcher, 2013). In the following illustrative examples, we compare the performance of the tangent principal component analysis method proposed in Guo et al. (2021) and Algorithm 3. The example is summarized in Fig. 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; see Fig. 6(b). In experiment 1, the edges' distributions have different means and different variances, and in 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 Fig. 6(c). To perform tangent principal component analysis, firstly, the Fréchet mean is computed to find an optimally aligned dataset and, secondly, the Euclidean principal component analysis is computed on the aligned dataset; see Fig. 6(d). As described in Algorithm 3, align-all-and-compute generalized geodesic principal components are computed starting from the unlabelled dataset, without any pre-alignment step; see Fig. 6(e). In experiment 1, the tangent principal component analysis and the align-all-and-compute generalized geodesic principal components are equivalent, as all the data points belong 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 principal component analysis captures 72.98% of the variability, see Fig. 6(d), while the generalized geodesic principal component

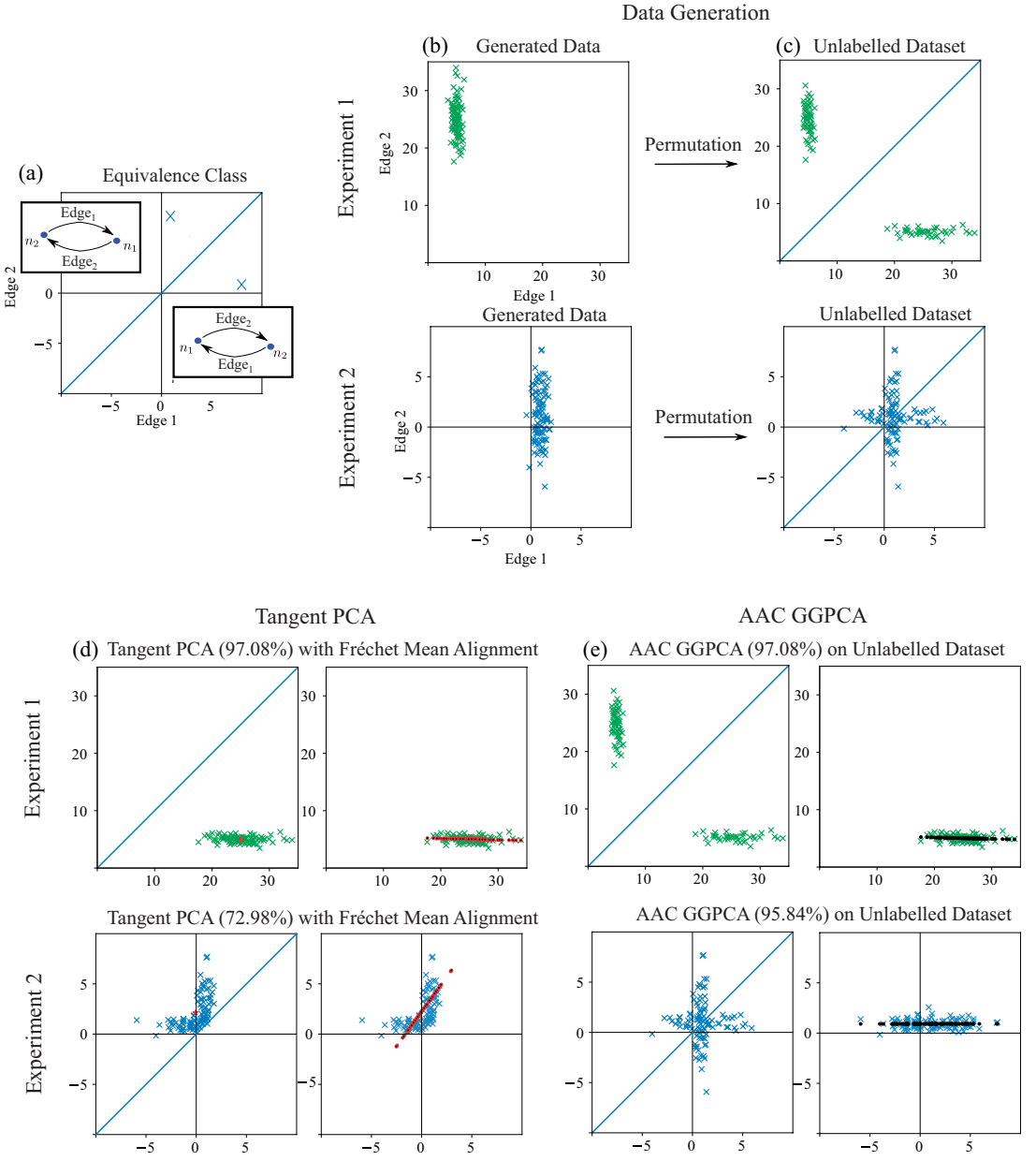


Fig. 6. (a) Visualization of equivalence classes; (b) Generated data; (c) Permuted generated data. (d) Comparison between tangent principal component analysis, PCA (Guo et al., 2021) and (e) align-all-and-compute generalized geodesic principal component analysis, AAC GGPCA. The fast quadratic assignment (Vogelstein et al., 2015) matcher is used for both simulations. We sampled 100 data points from: $Edge_1 \sim \mathcal{N}(5, 0.3)$, $Edge_2 \sim \mathcal{N}(25, 7)$ (Experiment 1); $Edge_1 \sim \mathcal{N}(1, 0.3)$, $Edge_2 \sim \mathcal{N}(1, 7)$ (Experiment 2).

analysis captures 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; see Fig. 6(e). In real-world data, understanding if the data belong to the fundamental domain is far from trivial (Kolaczyk et al., 2020); thus, the align-all-and-compute generalized geodesic principal component analysis is a safe model choice, ensuring a correct fit in the whole space and for graphs with different numbers of nodes.

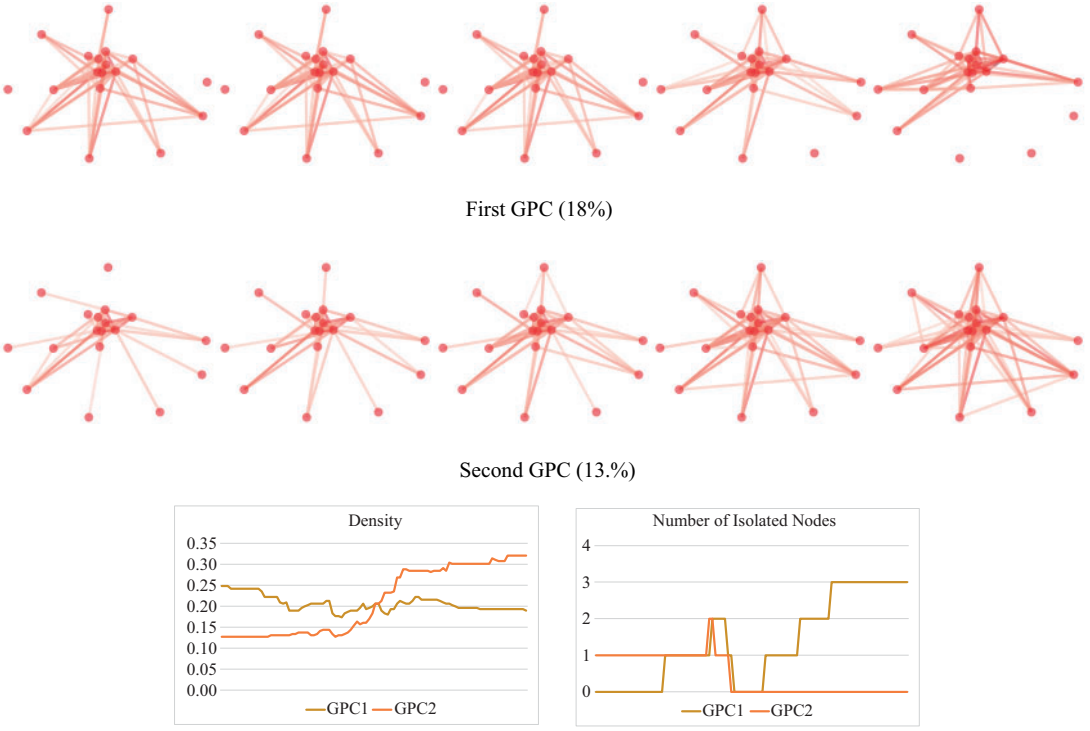


Fig. 7. Data projected along the first two geodesic principal components (0.1, 0.25, 0.5, 0.75, 0.9 percentiles). Graph density along the components as well as the number of isolated nodes along the components are reported. The position of the nodes is computed using the spring layout on the generalized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

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 excluded from the community. Social interaction is defined here 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 because the baboons in each network are different, belonging to different groups. See Rossi & Ahmed (2015) for the data, while more details about the case study can be found in Franz et al. (2015). The original dataset was filtered, keeping only the networks with less than 20 baboons. In Fig. 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, and the second principal component reflects the variability in terms of density. The example is able to show how the generalized principal component 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.

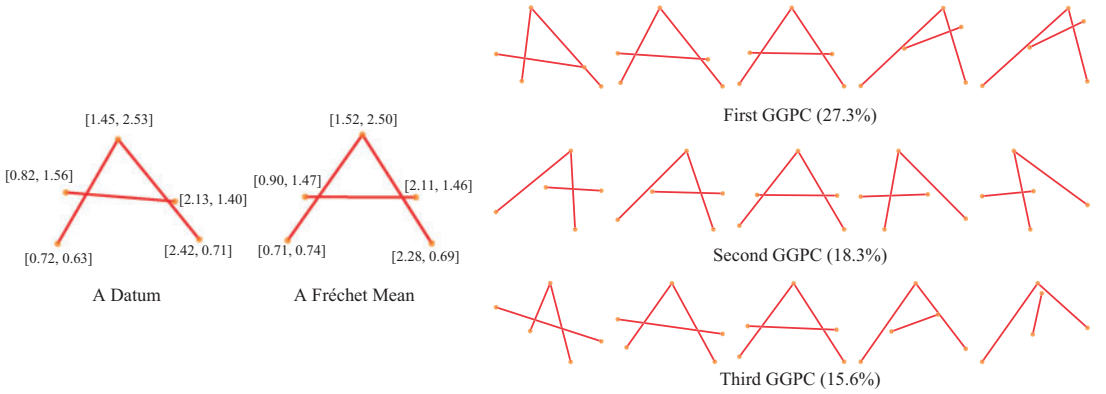


Fig. 8. Left: a datum extracted from the A dataset and the Fréchet mean. Every unlabelled node has a bi-dimensional real-valued attribute, while every edge has a 0, 1 attribute. Right: visualization of the generalized geodesic principal components. We show $0.1, 0.25, 0.5, 0.75, 0.9$ quantiles of the projected scores for the first three generalized geodesic principal components.

7.2. *Case study 2: undirected networks with vector attributes*

As an intuitive visual example with real data and associated vector attributes, we subsample 20 cases of the letter A from the well-known hand written letters dataset (Riesen & Bunke, 2008; Morris et al., 2020). As shown in the left panel of Fig. 8, every network has node attributes consisting of the node x and y coordinates, and binary (0/1) edge attributes indicating whether nodes are connected by lines. In Fig. 8, the Fréchet mean is shown, underlining how the framework is capturing both the topology and the node coordinates. The right panel of Fig. 8 plots network variation along the three generalized geodesic principal components. 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. While the first two generalized geodesic principal components 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 https://dati.lombardia.it/. 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 multilayer networks, represented as graphs whose edge attributes are vectors in $\mathbb{R}^4$.

The top panel of Fig. 9 shows the multilevel networks associated with four different hours on the left, as well as the Fréchet mean on graph space X/T computed with the align-all-and-compute 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

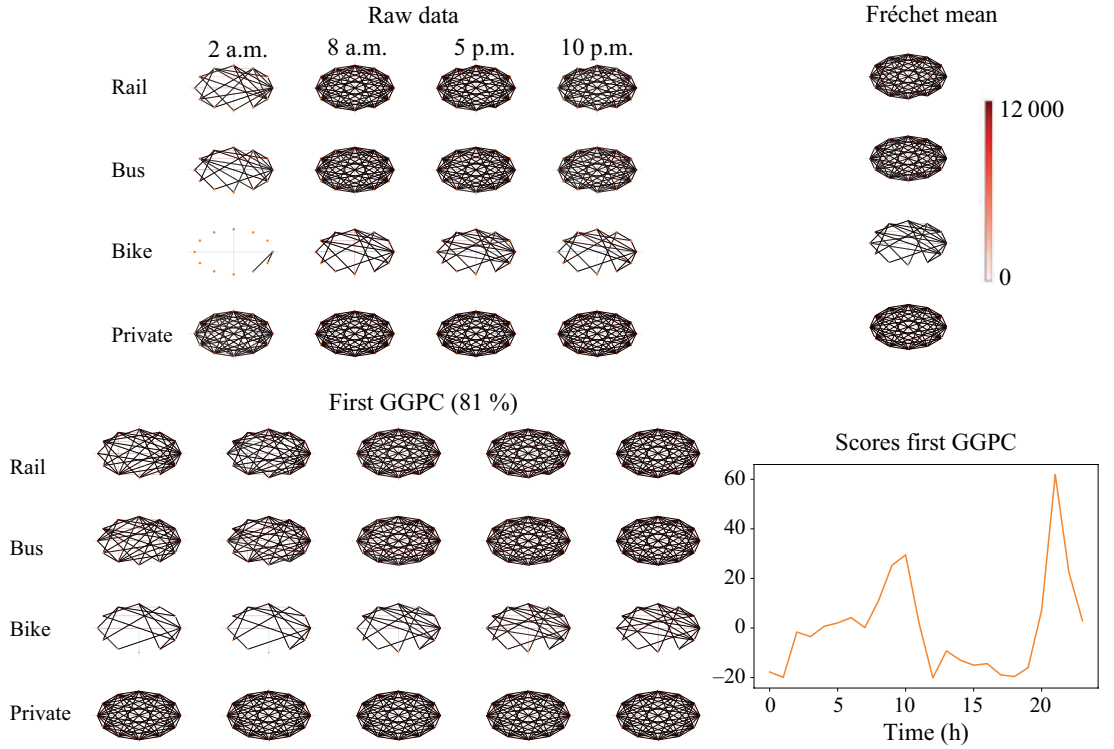


Fig. 9. Top left: network of the fluxes between the Lombardia Region provinces at 2 a.m., 8 a.m., 5 p.m. and 10 p.m. Top right: Fréchet mean in graph space X/T for the four mobility modes: bike, bus, rail and private. Bottom left: generalized geodesic principal component analysis performed on scaled data. The first generalized geodesic principal component 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 generalized geodesic principal component captures the density change along the day for all the transportation modes except for the private car mode. Bottom right: scores for the first generalized geodesic principal component plotted as a function of time.

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 overmatching also in the case where the cardinality of the permutation group is quite high, i.e., even though about 40 million possible node permutations are available, no artificial permutations are introduced by the algorithm. The bottom right panel of Fig. 9 shows that the majority of the variability is explained by the first generalized geodesic principal component, 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 every hour of the day. As displayed in the bottom right panel of Fig. 9, the scores reflect an expected time pattern, showing the high explanatory power of the generalized geodesic principal component analysis: two peaks corresponding to the morning and evening activities, one small peak at lunch and low activity at night.

8. Future research

A first development of the current work could focus on proving consistency and unbiasness of generalized principal components within graph space, following existing results regarding the Fréchet mean (Jain & Obermayer, 2010a,b; Kolaczyk et al., 2020) and more

general discussions on asymptotic bias and stability within similar quotient spaces (Huckemann, 2012; Miolane et al., 2017). Regarding the uniqueness of the generalized geodesic principal component analysis, the results of Kolaczyk et al. (2020) on the fundamental domain can be potentially extended to the geodesics and the generalized geodesic principal component analysis estimator. In more general terms, align all and compute is a strategy that 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 align-all-and-compute algorithm to predictive models tailored for graphs.

ACKNOWLEDGEMENT

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. We sincerely thank the editor and reviewers for providing constructive feedback to improve our work. This work was supported by the Centre for Stochastic Geometry and Advanced Bioimaging, the VKR Centre of Excellence and the Safari Njema Project funded by the Polisocial Award 2018 – Politecnico di Milano. Calissano carried out this work whilst at the Department of Mathematics, Politecnico di Milano. Feragen is also affiliated with the University of Copenhagen.

APPENDIX

Example of graph space construction

Consider the two weighted networks shown in Fig. A1. 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 Fig. A1. 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 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 node ordering, i.e., the axes. In Fig. A1, we show the generalized geodesics connecting the two graphs.

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 Lemma I.5.20 of Bridson & Haefliger (1999). 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 $(t_i^{\mathrm{T}} x_i t_i)_{i\in\mathbb{N}}$ in X such that, for some $M \in \mathbb{N}$ and $i, j \geqslant M$, we have $d_X(t_i^{\mathrm{T}} x_i t, t_j^{\mathrm{T}} x_j t_j) = d_{X/T}([x_i], [x_j])$. Since X is complete, the sequence $(t_i^{\mathrm{T}} x_i t)_{i\in\mathbb{N}}$ converges to some point $x \in X$, and hence the sequence

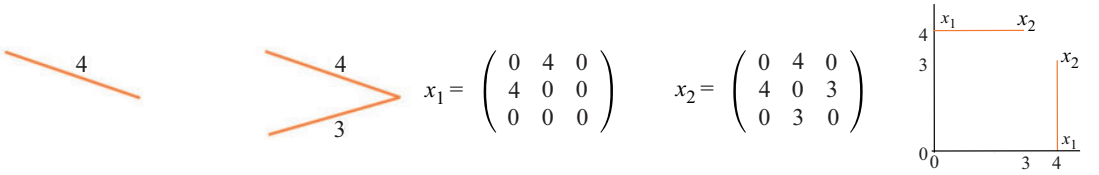


Fig. A1. Example of two simple weighted undirected networks x_1 and x_2, their matrix representations and the generalized geodesic connecting the two points.

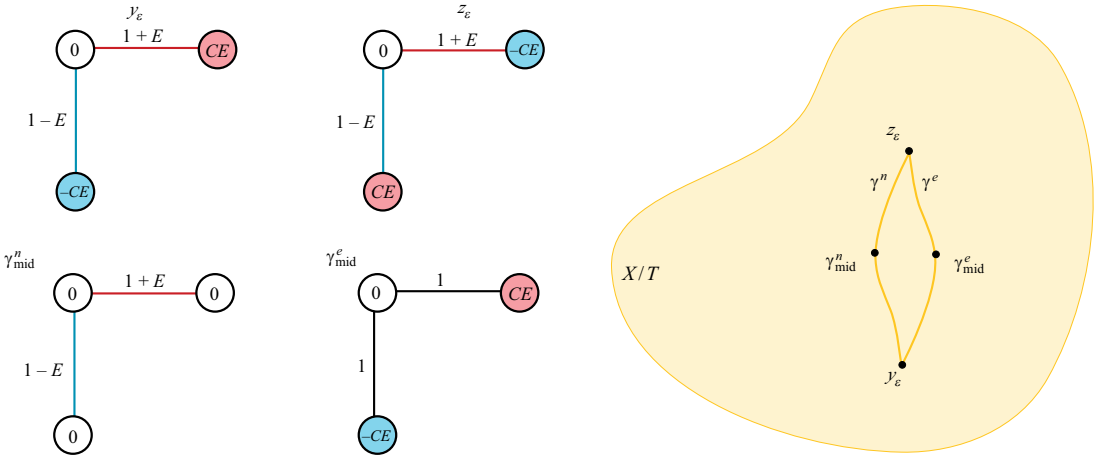


Fig. A2. Depending on their node alignment, the two graphs y_ε and z_ε differ either in node attributes or in edge attributes, and there exist two different geodesics γ^n and γ^e connecting the two graphs: One which interpolates node attributes and one which interpolates edge attributes, respectively. These two geodesics are illustrated via their midpoints γ^n_{mid} and γ^e_{mid}.

$([x_i])_{i\in\mathbb{N}}$ 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).

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 Fig. A2. The corresponding adjacency matrices are

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

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:

$$\gamma^n_{\text{mid}} = \begin{pmatrix} 0 & 1+\varepsilon & 1-\varepsilon \\ 1+\varepsilon & 0 & 0 \\ 1-\varepsilon & 0 & 0 \end{pmatrix}, \qquad \gamma^e_{\text{mid}} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & c\varepsilon & 0 \\ 1 & 0 & -c\varepsilon \end{pmatrix}.$$

The midpoint of this geodesic is the graph γ^n_{mid} shown on the left-hand side of the bottom row,

$$d_{X/T}(y_\varepsilon, \gamma^n_{\text{mid}}) = d(z_\varepsilon, \gamma^n_{\text{mid}}) = 0.2.$$

The second geodesic between y_ε and z_ε consists of interpolating the edge attributes with the node matching indicated by the colouring of the nodes. The midpoint of this geodesic is the graph γ^e_{mid} shown on the right-hand side of the bottom row,

$$d_{X/T}(y_\varepsilon, \gamma^e_{\text{mid}}) = d(z_\varepsilon, \gamma^e_{\text{mid}}) = 0.2.$$

The two points connected by two geodesics are illustrated schematically on the right-hand side of Fig. A2. The same argument can be built by considering the examples in Figs. 1–3. □

Proof of Theorem 1

Let $[x], [y_\varepsilon]$ and $[z_\varepsilon]$ be as in Fig. 1, 2 or 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 Proposition II.1.4 of Bridson & Haefliger (1999), graph space is not locally CAT(κ) at $[x]$ for any κ, and thus cannot have curvature $\leqslant \kappa$ in the sense of Alexandrov for any κ. □

Proof of Theorem 2

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

LEMMA A1. *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 all $t \in T \setminus T_{x_i}$,*

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

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

Proof. In order to prove the lemma, we show that the set

$$\mathcal{X}_T = \{([x_1], \ldots, [x_k]) \in (X/T)^k \mid \text{there exist some } i = 1, \ldots, k \text{ and } t \in T \setminus T_{x_i}$$

and some representatives $x_1, \ldots, x_k$, such that $d(\mu, x_i) = d(\mu, t^{\mathrm{T}} x_i t)\}$ has measure $\eta_k(\mathcal{X}_T) = 0$, where η_k is the product measure induced by η on

$$\underbrace{X/T \times \cdots \times X/T}_{k}.$$

For each element $t \in T$, denote by $X^t = \{x \in X \mid t^{\mathrm{T}} x t = x\}$ the fixed point set of t and define the set

$$X_{i\setminus t} = X \times \cdots \times \underbrace{X \setminus X^t}_{i\text{th}} \times \cdots \times X.$$

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

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

The preimage $f^{-1}(0)$ of the function $f\colon \underbrace{X \times \cdots \times X}_{k} \to \mathbb{R}, \quad (x_1, \ldots, x_k) \mapsto d_X^2(\mu, x_i) - d_X^2(\mu, t^{\mathrm{T}} x_i t)$, 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

$$\begin{aligned} f(x_1, \ldots, x_k) = \operatorname{tr}\Bigg\{ & \left(\frac{1}{k}\sum_{j=1}^{k} x_j - x_i\right)^{\mathrm{T}} \left(\frac{1}{k}\sum_{j=1}^{k} x_j - x_i\right) \\ & - \left(\frac{1}{k}\sum_{j=1}^{k} x_j - t^{\mathrm{T}} x_i t\right)^{\mathrm{T}} \left(\frac{1}{k}\sum_{j=1}^{k} x_j - t^{\mathrm{T}} x_i t\right)\Bigg\} \\ = \operatorname{tr}\left(\frac{2}{k}\sum_{j=1}^{k} \{x_j^{\mathrm{T}}(t^{\mathrm{T}} x_i t) - x_j^{\mathrm{T}} x_i\}\right). \end{aligned}$$

For $j \neq i$, we obtain

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

which is nonzero for $(x_1, \ldots, x_k) \in X_{i\setminus t}$. 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 $\eta_k(\mathcal{X}_k) = m_k\{\pi^{-1}(\mathcal{X}_T)\} = m_k\left(\bigcup_{i=1}^{k}\bigcup_{t\in T}\mathcal{X}_{i,t}\right) \leqslant \sum_{i=1}^{k}\sum_{t\in T} m_k(\mathcal{X}_{i,t}) = 0$, which completes the proof. □

We can now turn to the proof of Theorem 2. First, we prove convergence in finite time. Algorithm 1 consists of two steps repeated iteratively. Consider the function

$$\sum_{i=1}^{k} d_X^2(\mu^{\mathrm{cur}}, x_i^{\mathrm{cur}}), \tag{A1}$$

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 (A1), as an improved alignment would indeed lower the value of (A1). Similarly, the second step, which is the re-estimation of the Fréchet mean given the new alignments, also cannot increase the value of (A1) as, again, an improved estimate would lower its value. Moreover, if the value of (A1) stays fixed for two iterations in a row, the algorithm will terminate. Thus, the iterative algorithm will never see the same set of samplewise 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 align-all-and-compute algorithm. We show below that, with probability 1, there exist 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 (A1) 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 (3).

As Lemma A1 holds, there exists a $\varepsilon > 0$ and thus we can define

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

We now set $\varepsilon = \delta/2$ 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 $d(\mu', x_i) < d(\mu', t^{\mathrm{T}} x_i t)$, namely, that the optimal representative of $[x_i]$ is left unchanged for all i.

By the definition of δ, we have $d(\mu, x_i) \leqslant d(\mu, t^{\mathrm{T}} x_i t) - \delta$, and, by the triangle inequality, we have $d(\mu', x_i) \leqslant \underbrace{d(\mu', \mu)}_{<\epsilon=\delta/2} + d(\mu, x_i)$ and $d(\mu, t^{\mathrm{T}} x_i t) \leqslant \underbrace{d(\mu, \mu')}_{<\epsilon=\delta/2} + d(\mu', t^{\mathrm{T}} x_i t)$. We compute

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

which completes the proof of Theorem 2 under the assumption that Lemma A1 holds.

Proof of Theorem 3

The proof of convergence of align all and compute for generalized geodesic principal component analysis follows the exact same strategy as for the Fréchet mean in Theorem 2, noting the following.

(i) 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 generalized geodesic principal component $\delta \in \Gamma(X/T)$ and its representative $\gamma \in \Gamma(X)$.

(ii) For the first generalized geodesic principal component, the algorithm converges to a local minimum of (2) following the same argument as above. Here, the ε neighbourhood of the obtained estimate δ has to be considered on the Grassmannian manifold of one-dimensional subspaces of X.

REFERENCES

AFSARI, B., TRON, R. & VIDAL, R. (2013). On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM J. Contr. Optimiz.* **51**, 2230–60.

ARNAUDON, M., BARBARESCO, F. & YANG, L. (2013). Medians and means in Riemannian geometry: existence, uniqueness and computation. In *Matrix Information Geometry*, F. Nielsen & R. Bhatia, eds. Berlin: Springer, pp. 169–97.

ARNAUDON, M. & MICLO, L. (2014). A stochastic algorithm finding generalized means on compact manifolds. *Stoch. Proces. Appl.* **124**, 3463–79.

BACÁK, M. (2014). Computing medians and means in Hadamard spaces. *SIAM J. Optimiz.* **24**, 1542–66.

BILLE, P. (2005). A survey on tree edit distance and related problems. *Theor. Comp. Sci.* **337**, 217–39.

BILLERA, L. J., HOLMES, S. P. & VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. Appl. Math.* **27**, 733–67.

Bonnabel, S. (2013). Stochastic gradient descent on Riemannian manifolds. *IEEE Trans. Auto. Contr.* **58**, 2217–29.
Bredon, G. E. (1972). *Introduction to Compact Transformation Groups*, vol. 46. New York: Academic Press.
Bridson, M. R. & Haefliger, A. (1999). *Metric Spaces of Non-Positive Curvature*. Berlin: Springer.
Bunke, H. & Riesen, K. (2011). Improving vector space embedding of graphs through feature selection algorithms. *Pat. Recog.* **44**, 1928–40.
Calissano, A., Feragen, A. & Vantini, S. (2022). Graph-valued regression: prediction of unlabelled networks in a non-Euclidean graph space. *J. Mult. Anal.* **190**, 104950.
Chakraborty, R. & Vemuri, B. C. (2015). Recursive Fréchet mean computation on the Grassmannian and its applications to computer vision. In *2015 IEEE Int. Conf. Comp. Vis.*, pp. 4229–37. Piscataway, NJ: IEEE Press.
Chowdhury, S. & Mémoli, F. (2019). The Gromov–Wasserstein distance between networks and stable network invariants. *Info. Infer.* **8**, 757–87.
Duncan, A., Klassen, E. & Srivastava, A. (2018). Statistical shape analysis of simplified neuronal trees. *Ann. Appl. Statist.* **12**, 1385–421.
Durante, D., Dunson, D. B. & Vogelstein, J. T. (2017). Nonparametric Bayes modeling of populations of networks. *J. Am. Statist. Assoc.* **112**, 1516–30.
Duvenaud, D. K., Maclaurin, D., Aguilera-Iparraguirre, J., Gómez- Bombarelli, R., Hirzel, T., Aspuru-Guzik, A. & Adams, R. P. (2015). Convolutional networks on graphs for learning molecular fingerprints. In *Proc. 28th Int. Conf. Neural Info. Proces. Syst.*, vol. 2, pp. 2224–32. Cambridge, MA: MIT Press.
Feragen, A., Hauberg, S., Nielsen, M. & Lauze, F. (2011). Means in spaces of tree-like shapes. In *2011 IEEE Int. Conf. Comp. Vis.*, pp. 736–46. Piscataway, NJ: IEEE Press.
Feragen, A., Lauze, F., Lo, P., de Bruijne, M. & Nielsen, M. (2010). Geometries on spaces of treelike shapes. In *Computer Vision – ACCV 2010*, R. Kimmel, R. Klette & A. Sugimoto, eds. Berlin: Springer, pp. 160–73.
Feragen, A. & Nye, T. M. W. (2020). Statistics on stratified spaces. In *Riemannian Geometric Statistics in Medical Image Analysis*, X. Pennec, S. Sommer & P. T. Fletcher, eds. New York: Academic Press, 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, J. C. Gee, S. Joshi, K. M. Pohl, W. M. Wells & L. Zöllei, eds. Berlin: Springer, pp. 74–85.
Fletcher, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comp. Vis.* **105**, 171–85.
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*, M. Sonka, I. A. Kakadiaris & J. Kybic, eds. Berlin: Springer, pp. 87–98.
Franz, M., Altmann, J. & Alberts, S. C. (2015). Knockouts of high-ranking males have limited impact on baboon social networks. *Curr. Zool.* **61**, 107–13.
Garba, M. K., Nye, T. M., Lueg, J. & Huckemann, S. F. (2021). Information geometry for phylogenetic trees. *J. Math. Biol.* **82**, 1–39.
Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S. & Kolaczyk, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Statist.* **11**, 725–50.
Gold, S. & Rangarajan, A. (1996). A graduated assignment algorithm for graph matching. *IEEE Trans. Pat. Anal. Mach. Intel.* **18**, 377–88.
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. *J. Math. Imag. Vis.* **63**, 735–52.
Hauberg, S., Feragen, A., Enficiaud, R. & Black, M. J. (2015). Scalable robust principal component analysis using Grassmann averages. *IEEE Trans. Pat. Anal. Mach. Intel.* **38**, 2298–311.
Huckemann, S. F. (2012). On the meaning of mean shape: manifold stability, locus and the two sample test. *Ann. Inst. Statist. Math.* **64**, 1227–59.
Huckemann, S. & Eltzner, B. (2020). Statistical methods generalizing principal component analysis to non-Euclidean spaces. In *Handbook of Variational Methods for Nonlinear Geometric Data*, P. Grohs, M. Holler & A. Weinmann, eds. Cham: Springer, pp. 317–38.
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.
Jain, B. J. & Obermayer, K. (2008). On the sample mean of graphs. In *2008 IEEE Int. Joint Conf. Neural Networks*, pp. 993–1000. Piscataway, NJ: IEEE Press.
Jain, B. J. & Obermayer, K. (2009). Structure spaces. *J. Mach. Learn. Res.* **10**, 2667–714.
Jain, B. J. & Obermayer, K. (2010a). Consistent estimator of median and mean graph. In *2010 20th Int. Conf. Pat. Recog.*, pp. 1032–5. Piscataway, NJ: IEEE Press.

JAIN, B. J. & OBERMAYER, K. (2010b). Large sample statistics in the domain of graphs. In *Structural, Syntactic, and Statistical Pattern Recognition*, E. R. Hancock, R. C. Wilson, T. Windeatt, I. Ulusoy & F. Escolano, eds. Berlin: Springer, pp. 690–7.

KENDALL, D. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. Lond. Math. Soc.* **16**, 81–121.

KOLACZYK, E. D., LIN, L., ROSENBERG, S., WALTERS, J. & XU, J. (2020). Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *Ann. Stat.* **48**, 514–538.

LUEG, J., GARBA, M. K., NYE, T. M. & HUCKEMANN, S. F. (2021). Wald space for phylogenetic trees. In *Geometric Science of Information*, F. Nielsen and F. Barbaresco, eds. Cham: Springer, pp. 710–7.

LUNAGÓMEZ, S., OLHEDE, S. C. & WOLFE, P. J. (2020). Modeling network populations via graph distances. *J. Am. Statist. Assoc.* **116**, 2023–40.

MALLASTO, A. & FERAGEN, A. (2018). Wrapped Gaussian process regression on Riemannian manifolds. In *2018 IEEE/CVF Conf. Comp. Vis. Pat. Recog.*, pp. 5580–8. Piscataway, NJ: IEEE Press.

MARRON, J. S. & ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biomet. J.* **56**, 732–53.

MILLER, E., OWEN, M. & PROVAN, J. S. (2015). Polyhedral computational geometry for averaging metric phylogenetic. *Adv. Appl. Math.* **68**, 51–91.

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. *J. Mach. Learn. Res.* **21**, 1–9.

MIOLANE, N., HOLMES, S. & PENNEC, X. (2017). Template shape estimation: correcting an asymptotic bias. *SIAM J. Imag. Sci.* **10**, 808–44.

MORRIS, C., KRIEGE, N. M., BAUSE, F., KERSTING, K., MUTZEL, P. & NEUMANN, M. (2020). Tudataset: a collection of benchmark datasets for learning with graphs. In *ICML 2020 Workshop on Graph Representation Learning and Beyond (GRL+ 2020)*. www.graphlearning.io.

NYE, T. M. W. (2011). Principal components analysis in the space of phylogenetic trees. *Ann. Statist.* **39**, 2716–39.

NYE, T. M. W. (2014). An algorithm for constructing principal geodesics in phylogenetic treespace. In *IEEE/ACM Trans. Comput. Biology Bioinform.*, vol. 11, pp. 304–15. Piscataway, NJ: IEEE Press.

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**, 901–22.

PENNEC, X., FILLARD, P. & AYACHE, N. (2006). A Riemannian framework for tensor computing. *Int. J. Comp. Vis.* **66**, 41–66.

RIESEN, K. & BUNKE, H. (2008). Iam graph database repository for graph based pattern recognition and machine learning. In *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 287–97. Berlin: Springer.

ROSSI, R. A. & AHMED, N. K. (2015). The network data repository with interactive graph analytics and visualization. In *Proc. 29th AAAI Conf. Artif. Intel.*, pp. 4292–3. AAAI Press.

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. *J. Mach. Learn. Res.* **12**, 2539–61.

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. *Front. Comput. Neurosci.* **7**, 171–84.

SRIVASTAVA, A., JOSHI, S. H., MIO, W. & LIU, X. (2005). Statistical shape analysis: clustering, learning, and testing. *IEEE Trans. Pat. Anal. Mach. Intel.* **27**, 590–602.

STURM, K. T. (2003). Probability measures on metric spaces of nonpositive. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, vol. 338, pp. 357–90. Providence, RI: American Mathematical Society.

TURNER, K., MILEYKO, Y., MUKHERJEE, S. & HARER, J. (2014). Fréchet means for distributions of persistence diagrams. *Disc. Comp. Geom.* **52**, 44–70.

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. *Eur. Phys. J.* B **68**, 261–75.

WANG, H. & MARRON, S. J. (2007). Object oriented data analysis: sets of trees. *Ann. Statist.* **35**, 1849–73.

WASSERMAN, S. & FAUST, K. (1994). *Social Network Analysis: Methods and Applications*, vol. 8. Cambridge: Cambridge University Press.

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, Red Hook, NY: Curran Associates, pp. 1178–86.

[*Received on* 8 *April* 2022. *Editorial decision on* 20 *March* 2023]