



HAL
open science

Graph-Valued Models for Dimensionality Reduction and Regression

Anna Calissano, Aasa Feragen, Simone Vantini

► **To cite this version:**

Anna Calissano, Aasa Feragen, Simone Vantini. Graph-Valued Models for Dimensionality Reduction and Regression. JSM 2021 - Joint Statistical Meeting, Aug 2021, Online, United States. hal-03518747

HAL Id: hal-03518747

<https://inria.hal.science/hal-03518747>

Submitted on 10 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

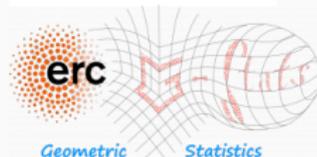
Graph-Valued Models for Dimensionality Reduction and Regression

Anna Calissano

8th August 2021

Politecnico di Milano

Now at: Inria and Université Côte d'Azur



The current work has been developed during my PhD under the supervision of:

Prof. Aasa Feragen DTU
Copenhagen and University of
Copenhagen

Prof. Simone Vantini Politecnico
di Milano



Currently working at INRIA Sophia Antipolis and Université Côte d'Azur within the **ERC Project: Geometry and Statistics** with Xavier Pennec.

Populations of Graphs

Complex data type like networks can be analysed in two approaches [Wang and Marron, 2007]:

first generation approach : the analysis and modelling of a single network datum with a set of nodes and a set of edges.



$$G = (V, E, a)$$

V set of nodes $|V| = n$

E set of edges $|E| = n^2$

$a : E \rightarrow A$

second generation approach: the analysis of a population of network-valued data, i.e. the analysis of not one but a set of networks.



$$X_1, \dots, X_k$$

$$X_i = (V, E_i, a_i)$$

X embedding space

Why Population of Graphs

Networks (or Graphs) are a powerful mathematical representation of reality:

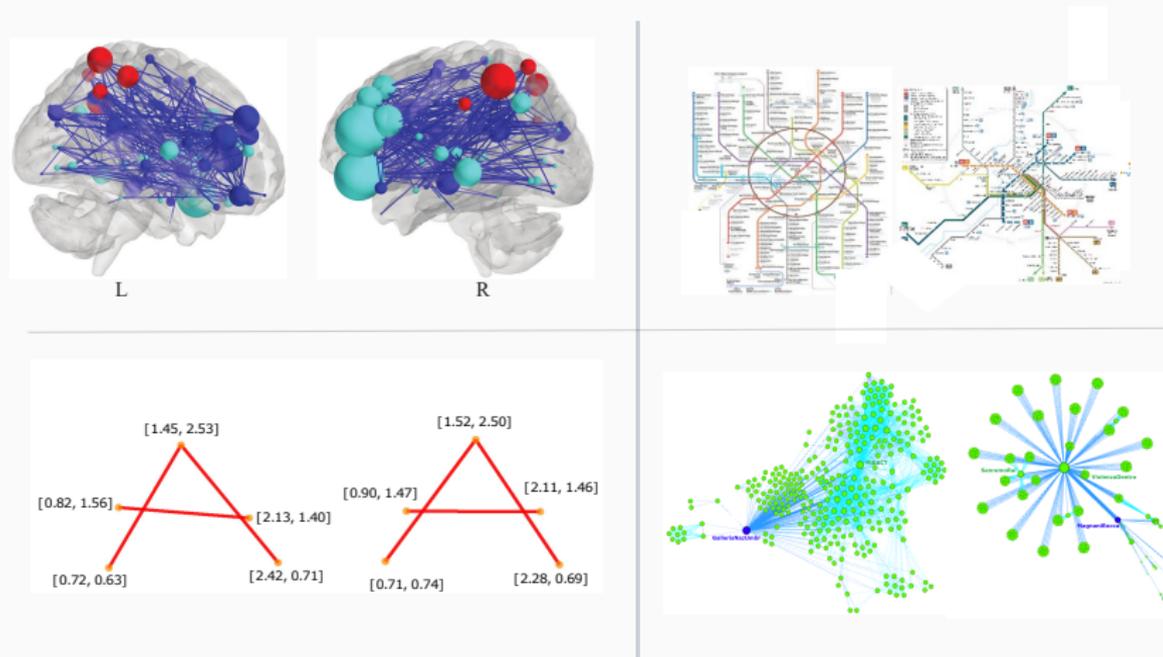
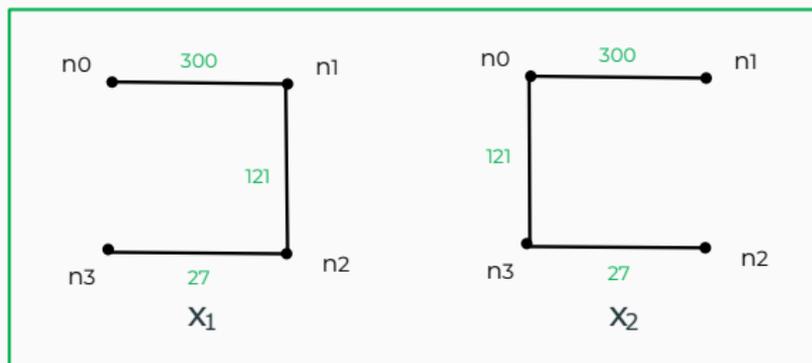


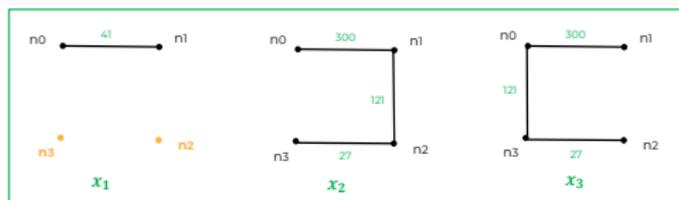
Figure 1: Brain Connectivity Networks in response to different cognitive tasks [Ginestet et al., 2017]; Public Transport Networks in different cities; Letters [Kersting et al., 2016, Riesen and Bunke, 2008]; Twitter Networks of different museums



- **Nodes Labelled Networks:** $x_1 \neq x_2$ if the nodes are labelled
- **Nodes Unlabelled Networks:** $x_1 = x_2$ if the nodes are unlabelled and a matching between the nodes can be defined

Graph Space

HP. Directed, Un-directed; Euclidean scalar or vector attributes; Finite nodes



Adjacency matrices: every G_i is represented as an adjacency matrix or tensor.

Adding zero-nodes: smaller graphs are grown by adding $n - |V|$ null nodes.

Total Space: $X = \mathbb{R}^{n \times n \times d}$, where d is the attributes dimension.

Definition (Graph Space and Metric [Jain and Obermayer, 2009])

Graph Space is a quotient space X/T obtained by applying the permutation action T to the space of adjacency matrices or tensors X . Any metric d_X on X defines a quotient metric on Graph Space X/T

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

A concept of **optimal alignment** between two points follows from the metric.

Statistical Analysis on Graph Space

The statistical analysis of object data is divided in two strategies to develop statistical tools [Marron and Alonso, 2014]:

Extrinsic Statistics: define statistics on an "easier" tangent space - i.e. Manifold Statistics (see for example [Fletcher, 2013, Fletcher and Joshi, 2004, Huckemann et al., 2010, Kendall, 1984, Mallasto and Feragen, 2018, Pennec et al., 2006, Srivastava et al., 2005, Zhang and Fletcher, 2013])

Intrinsic Statistics: defining statistics within the chosen embedding space [Marron and Alonso, 2014, Wang and Marron, 2007]

A geometrical characterization of Graph Space is required to understand the applicability of the first strategy.

Lemma

Graph Space X/T is not a manifold

⇒ Many well-known tools from *manifold statistics* are unfortunately not directly applicable to Graph Space

Lemma

Graph Space is a geodesic space. Geodesics connecting pairs of points are not generally unique in X/T .

⇒ We can define methods using geodesics on the total space and on the quotient space

Theorem (The curvature of Graph Space)

Graph space does not have curvature $\leq \kappa$ in the sense of Alexandrov for any $\kappa \in \mathbb{R}$. For any graph $[x] \in X/T$ and any $R_\kappa > 0$, we can find two graphs $[y_\epsilon], [z_\epsilon] \in B([x], R_\kappa) \subset X/T$ which are connected by two geodesics.

⇒ Curvature affects convergence proofs for common statistical estimation algorithms, which often rely on assumptions of bounded curvature

The geometrical characterization of the Graph Space suggests:

Extrinsic Statistics: it is not applicable due to the non-manifold structure of the space

Intrinsic Statistics: it is a solution but how can we compute statistics in a space with unbounded curvature?

⇒ define intrinsic tools and a general strategy namely **Align All and Compute (AAC)** to estimate these intrinsic statistics on Graph Space [C. et al., 2020a].

The strategy is used to compute the Generalized Geodesic Principal Components [C. et al., 2020a] and the **Generalized Geodesic Regression** model [C. et al., 2020b].

Definition

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

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

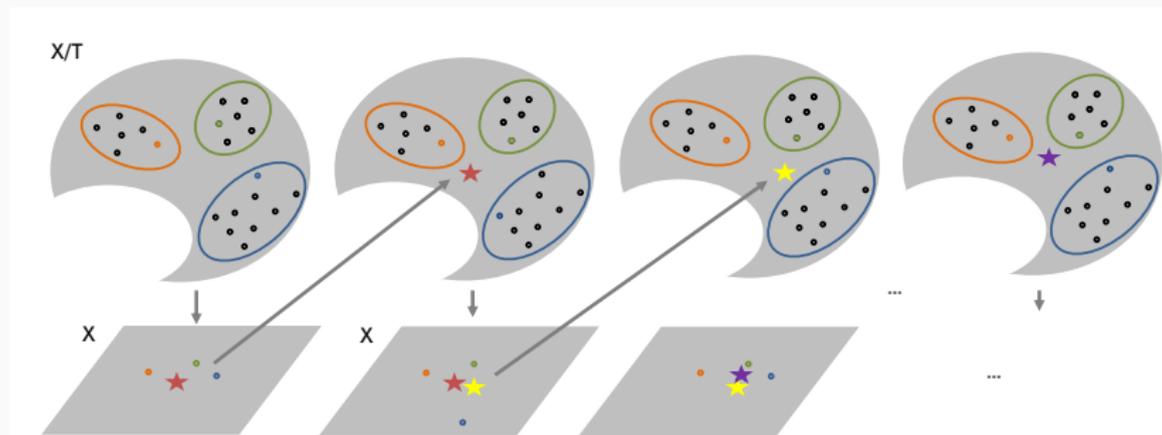
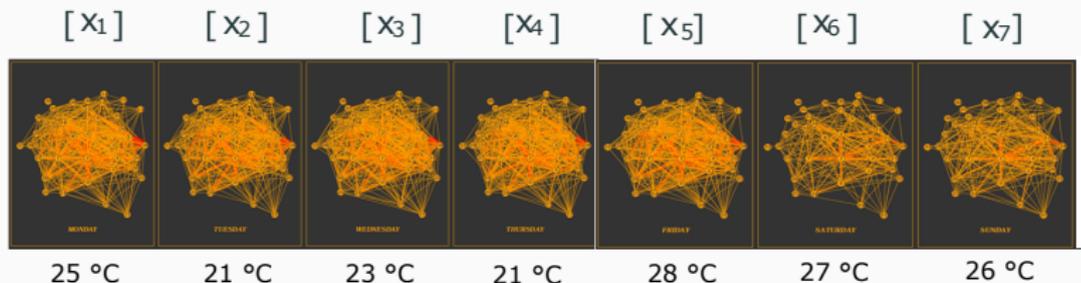


Figure 2: Conceptual visualization of ACC for the estimation of the Fréchet Mean.

Regression Model



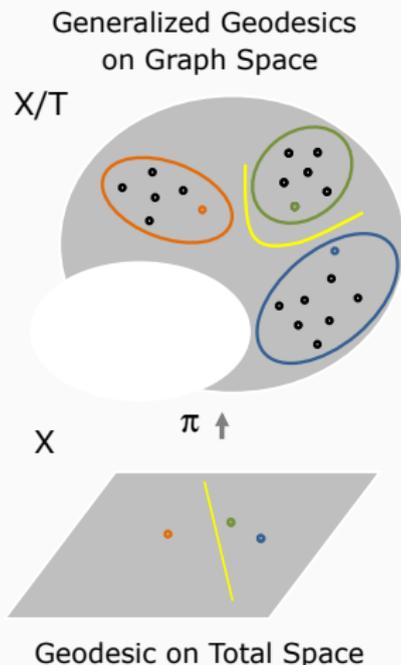
Given a sample $(\underline{s}_1, [x_1]), \dots, (\underline{s}_k, [x_k])$ where $(\underline{s}_i, [x_i]) \in \mathbb{R}^p \times X/T$, we aim to describe the relationship:

$$f: \mathbb{R}^p \rightarrow X/T$$

minimizing:

$$\sum_{i=1}^k d_{X/T}^2([x_i], f(\underline{s}_i)) \quad (2)$$

Building a Regression to model a set of networks in Graph Space from a set of values means to interpolate between graphs, finding an optimal geodesic.



Definition (C. et al. [2020a])

Denote by $\Gamma(X) := \{\gamma : \mathbb{R}^p \rightarrow X\}$ the set of all straight lines ($p = 1$), planes, or hyper-planes ($p > 1$) in X . A curve δ is a *generalized geodesic* on the Graph Space X/T , if it is a projection of a straight line, plane, or hyper-plane on X :

$$\Gamma(X/T) = \{\delta = \pi \circ \gamma : \gamma \in \Gamma(X)\}. \quad (3)$$

where $\pi : X \rightarrow X/T$ is the canonical projection.

Definition

(Family of Generalized Geodesic Regression Models in Graph Space
[C. et al., 2020b])

Consider the regression model

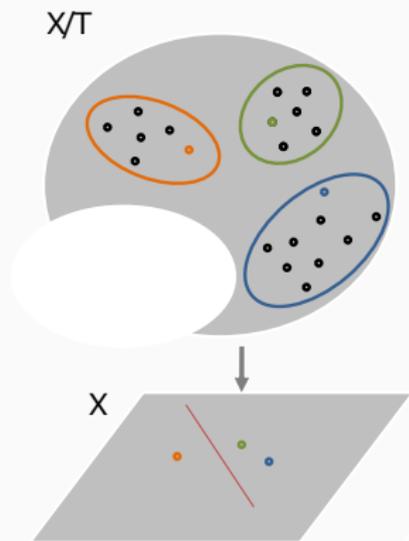
$$f_{\beta}: \mathbb{R}^p \rightarrow X/T, \quad s \mapsto f_{\beta}(\underline{s}) \in X/T$$

given by $f_{\beta}(\underline{s}) = \pi \circ h_{\beta}(\underline{s})$, where $\pi: X \rightarrow X/T$ is the canonical projection from total to quotient space, $h_{\beta}: \mathbb{R}^p \rightarrow X$ is a linear regression model on X of the form

$$[h_{\beta}(\underline{s})]_j = \sum_{m=0}^p s_m \beta_{m,j} \quad (4)$$

for edge- and node-wise coefficients $\beta \in X$.

Denote by $\mathcal{F}(X/T) := \{f_{\beta}: \mathbb{R}^p \rightarrow X/T\}$, the family of such models.



We defined:

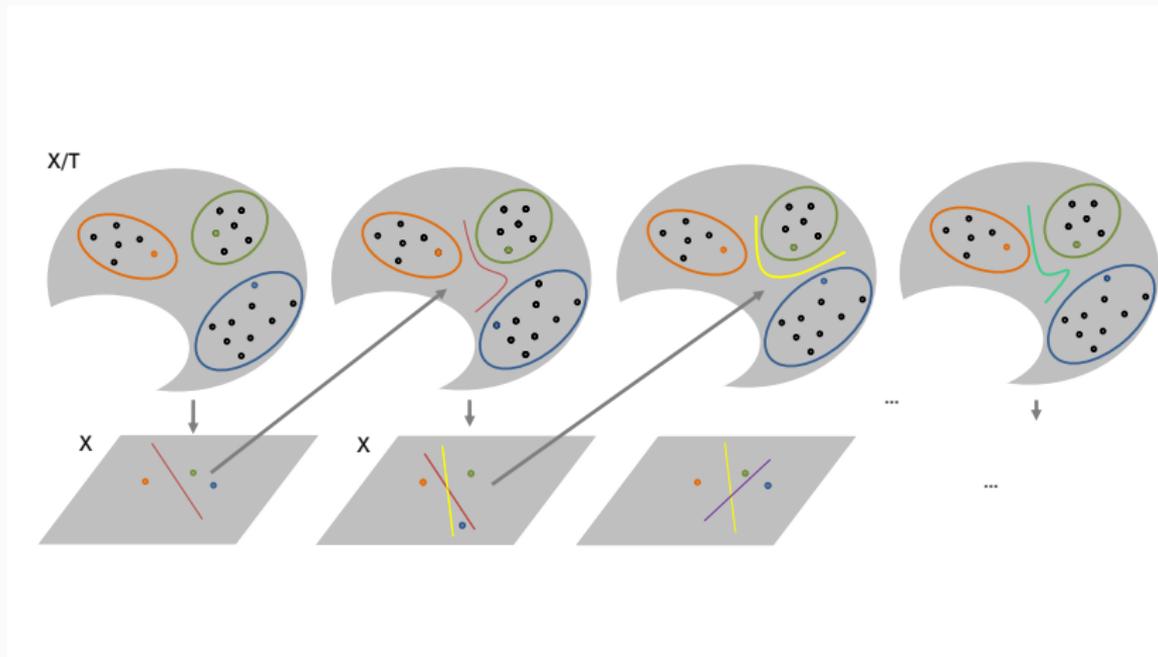
1. $h_{\beta}(s)$ a regression model on total space X
2. $\mathcal{F}(X/T) = \{f_{\beta} \in \Gamma(X/T) : f_{\beta}(s) = \pi \circ h_{\beta}(s)\}$ a regression model on Graph Space X/T

\Rightarrow the tangent space approach would align all the data once and then compute the $h_{\beta}(s)$

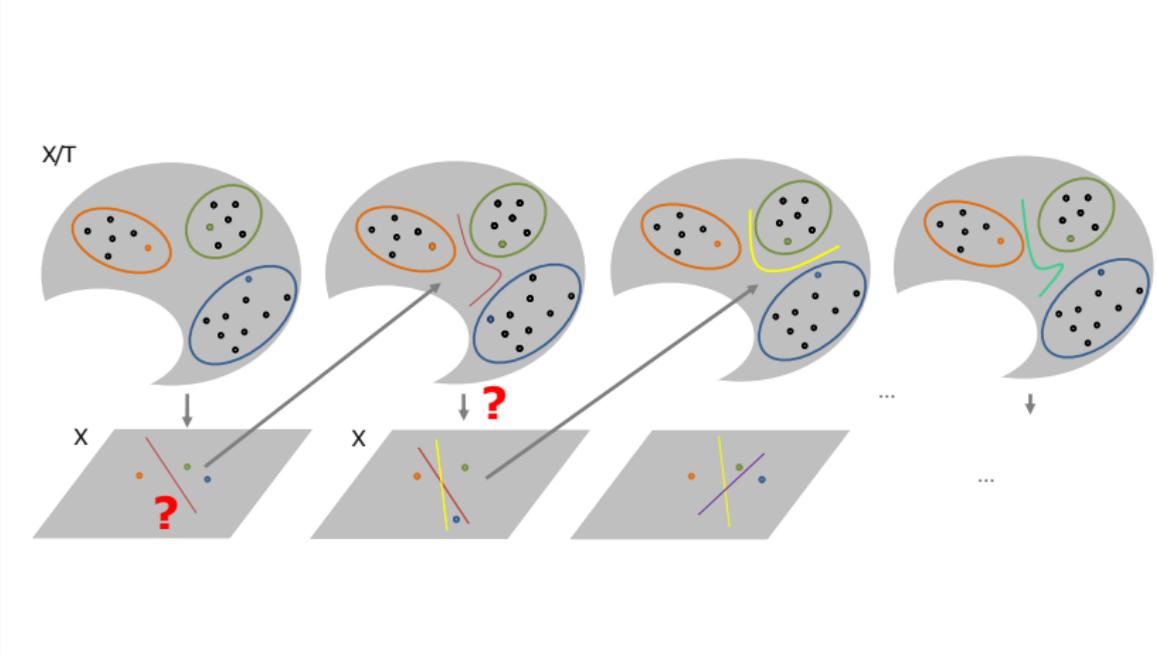
The tangent space is not a good approximation of the graph space even locally due to the non-manifold structure

\Rightarrow Intrinsic Regression Model with **AAC for regression**

AAC for Regression



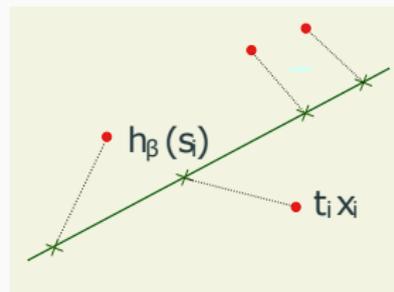
AAC for Regression



Definition**(Regression Alignment [C. et al., 2020b])**

Consider $(\underline{s}_i, [x_i]) \in \mathbb{R}^p \times X/T$, $i = 1, \dots, k$, $t \in T$, and $f_\beta : \mathbb{R}^p \rightarrow X/T$ a generalized geodesic in X/T with representative $h_\beta : \mathbb{R}^p \rightarrow X$. The graph representative $t_i x_i \in X$ is in *optimal regression position* with respect to the regression line f_β on X if

$$d_X(t_i x_i, h_\beta(\underline{s}_i)) = d_{X/T}([x_i], f_\beta(\underline{s}_i)) \quad (5)$$

**Definition (Generalized Geodesic Regression Model [C. et al., 2020b])**

Given a sample $(\underline{s}_1, [x_1]), \dots, (\underline{s}_k, [x_k])$ where $(\underline{s}_i, [x_i]) \in \mathbb{R}^p \times X/T$, $f_\beta \in \mathcal{F}(X/T)$ is a *Generalized Geodesic Regression* on the Graph Space X/T , if it is the projection $f_\beta(\underline{s}) = \pi \circ h_\beta(\underline{s})$ of a linear regression model $h_\beta(\underline{s})$ on X , minimizing

$$\sum_{i=1}^k d_X^2(t_i x_i, h_\beta(\underline{s}_i)) \quad (6)$$

where $t_i x_i$ are in *optimal regression position* as defined above in Definition 5.

Case Studies

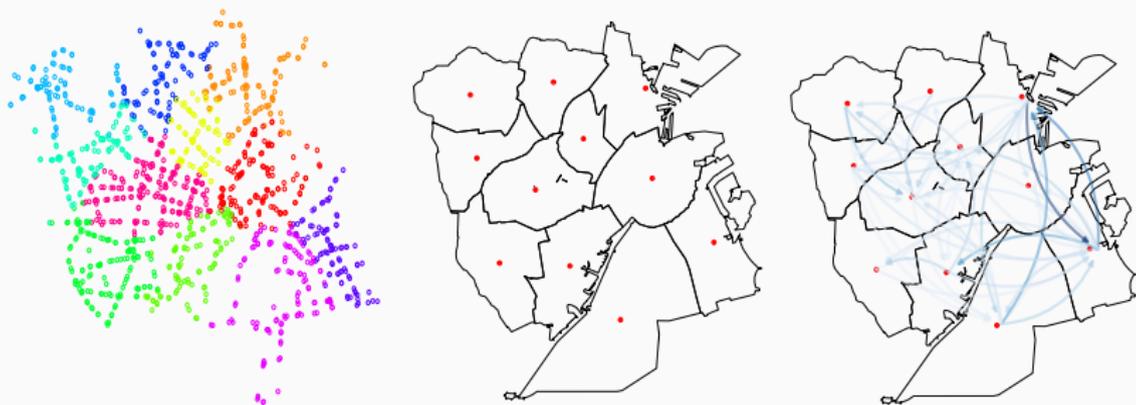
Movia is the bus company of Copenhagen and the Zealand Region. All the users which are not using the app or have year subscription should check in when they drop on and check out when they drop off.



Figure 3: Rejsekord Check in example. A bus and a boat run by Movia.

Data: trips between the 27th of February and the 13th of May, 2020.

$[x_i]$ - the origin destination matrix of fluxes between the areas per day



s_i - two dummy variables describing three different phases:

- *pre lock-down*: 27/02 – 13/03
- *lock-down*: 13/03 – 13/04
- *phase II*: 13/04 – 13/05

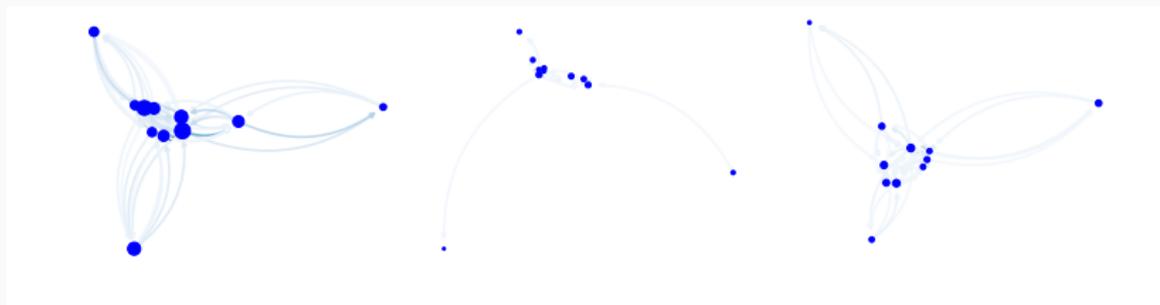
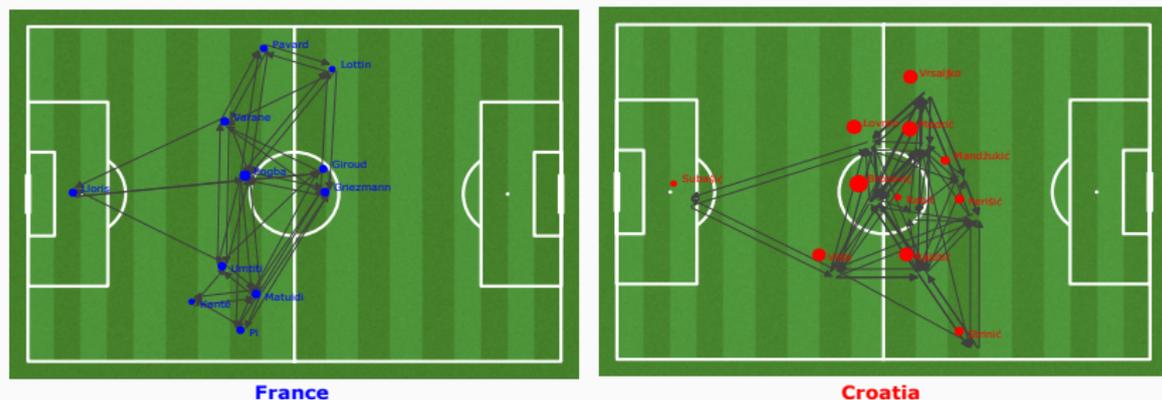


Figure 5: Pre, During, and Post lockdown.

Model: Given the data $(s_1, [x_1]), \dots, (s_k, [x_k])$ we estimated the **three predicted networks** corresponding to the three phases. Notice that this regression model corresponds to an Anova model.

→ **Interesting Result:** the permutation of the most central area *Indre By* becomes interchangeable with residential area during the lockdown.

Players Passing Network: number of passings between players during a match.



Using open source database [Ltd, 2018], we collect all the PPN of each team in each match for the whole Fifa 2018 World Championship, obtaining a set of $(s_i, [x_i]), i = 1, \dots, 128$, where $[x_i]$ is the PPN and s_i is the match result.



0-3 goals: the topology is more diffuse as well as the density is distributed over the nodes.

4-6 goals: the density increases, showing a more defined playing strategy within the team and the topology change

Conclusions and Further Developments

We presented a geometrical framework for the statistical analysis of population of unlabelled and labelled, one or multi-layer, weighed or unweighted networks.

- the geometrical properties of *Graph Space* to define intrinsic statistics
- graph-on-vector regression model based on generalized geodesics [C. et al., 2020a]
- other methodologies are available in the same framework such as Conformal Prediction intervals for the uncertainty quantification [C. et al., 2020d] and geodesic principal components [C. et al., 2020a]

All the presented methods are implemented within the GraphSpace Python Package [C. et al., 2020c]

- ⇒ **zero-node adding problem:** solving the methodological and application issue from zero node adding solution by defining a stratified quotient space embedding strategy;
- ⇒ **binary network modelling:** how to extend the framework to a discrete total space X , to create a pure network-topology descriptors;
- ⇒ **geomstats+graph space:** how to integrate the graphspace python package within the *geomstats* python framework

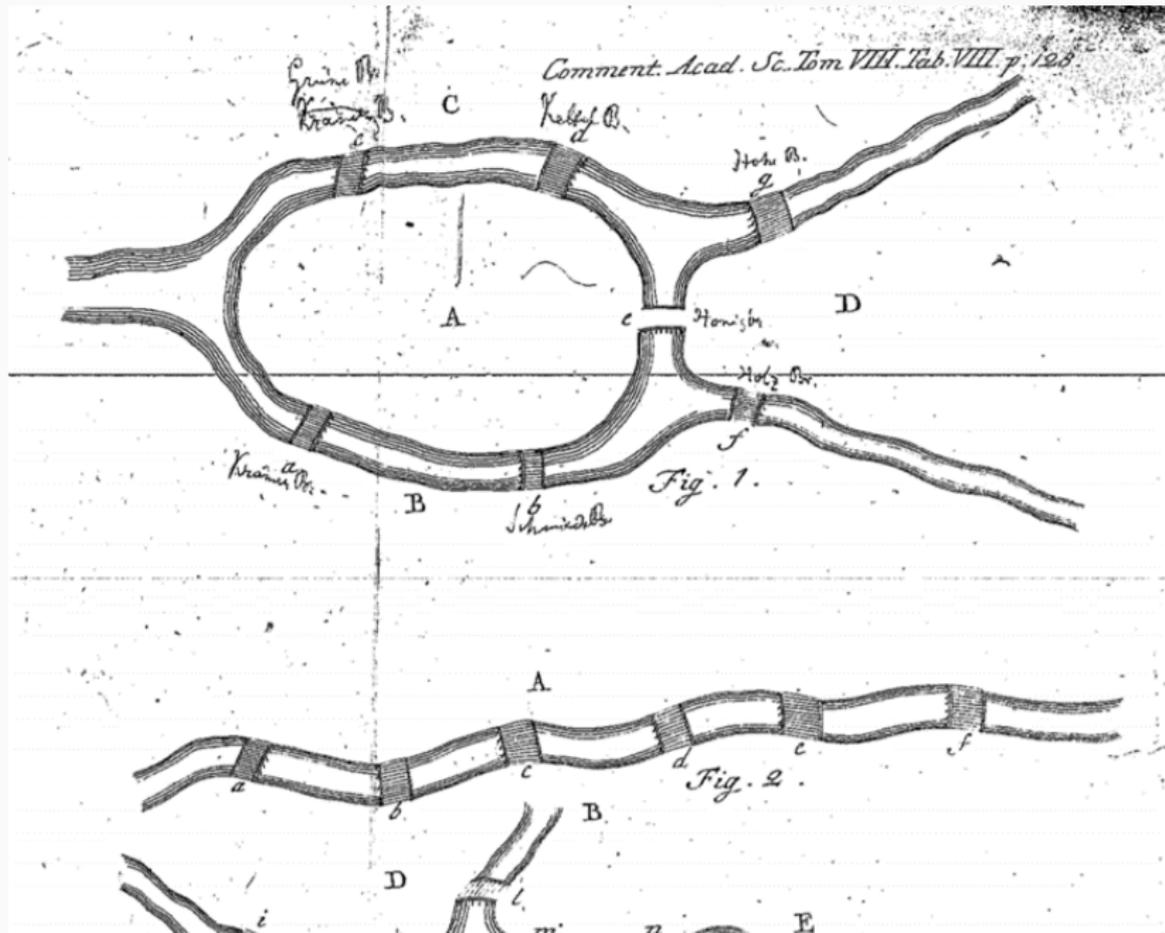
References

- G. E. Bredon. *Introduction to compact transformation groups*, volume 46. Academic press, 1972.
- M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*. Springer, 1999.
- A. C., A. Feragen, and S. Vantini. Populations of unlabeled networks: Graph space geometry and geodesic principal components. *MOX Report*, 2020a.
- A. C., A. Feragen, and S. Vantini. Graph-valued regression: Predicting network from scalar. *In preparation, see:*
<http://www2.compute.dtu.dk/~afhar/publications.php>, 2020b.
- A. C., A. Feragen, and S. Vantini. Graphspace python package.
<https://github.com/annacalissano/GraphSpace.git>, 2020c.
- A. C., G. Zeni, M. Fontana, and S. Vantini. Conformal prediction sets for populations of networks. *In preparation*, 2020d.
- P. T. Fletcher. Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision*, 105(2): 171–185, 2013.
- P. T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer vision and mathematical methods in medical and biomedical image analysis*, pages 87–98. Springer, 2004.

- C. E. Ginestet, J. Li, P. Balachandran, S. Rosenberg, and E. D. Kolaczyk. Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics*, 11(2):725–750, 2017. doi: 10.1214/16-AOAS1015.
- S. Huckemann, T. Hotz, and A. Munk. Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statist. Sinica*, 20(1):1–58, 2010.
- B. J. Jain and K. Obermayer. Structure spaces. *Journal of Machine Learning Research*, 10(Nov):2667–2714, 2009.
- D.G. Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.
- K. Kersting, N. M. Kriege, C. Morris, Mutzel. P., and M. Neumann. Benchmark data sets for graph kernels, 2016. URL <http://graphkernels.cs.tu-dortmund.de>.
- StatsBomb Service Ltd. Statsbomb open data, 2018. URL <https://github.com/statsbomb/open-data>.
- A. Mallasto and A. Feragen. Wrapped Gaussian process regression on Riemannian manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5580–5588, 2018.

- J. S. Marron and A. M. Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753, 2014.
- X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision (IJCV)*, 66(1):41–66, 2006.
- K. Riesen and H. Bunke. lam 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)*, pages 287–297. Springer, 2008.
- A. Srivastava, S. H. Joshi, W. Mio, and X. Liu. Statistical shape analysis: Clustering, learning, and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(4):590–602, 2005.
- H. Wang and S. J. Marron. Object oriented data analysis: Sets of trees. *The Annals of Statistics*, 35(5):1849–1873, 2007.
- M. Zhang and P. T. Fletcher. Probabilistic principal geodesic analysis. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1178–1186. Curran Associates, Inc., 2013.

Thank you!



Appendix

Variability: Geodesic Principal Components

In C. et al. [2020a], we define geodesic principal components to capture the variability within a set of graphs.

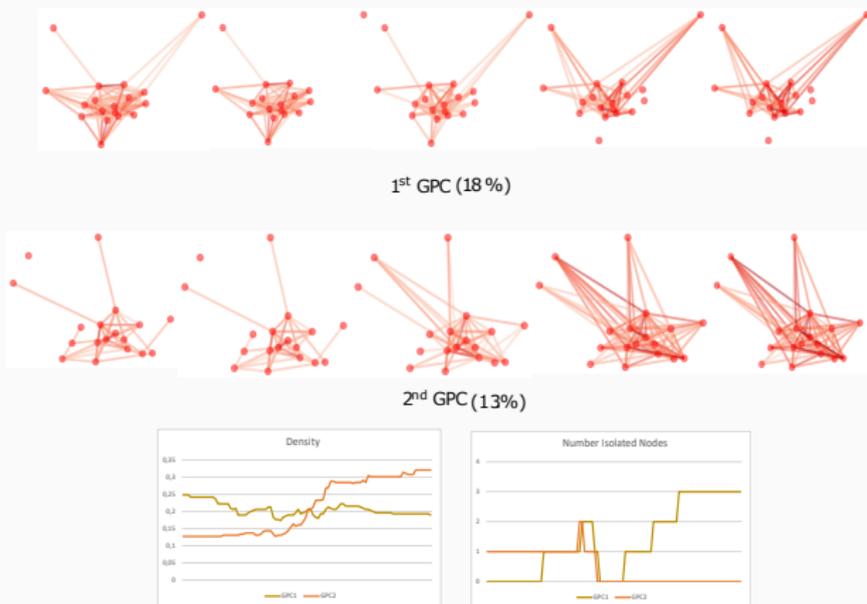


Figure 6: First two geodesic principal components of baboons grooming networks population. Data are sampled along the geodesics and plotted. Indicators are computed to highlight the patterns.

In [C. et al., 2020d], we defined conformal prediction intervals in the Graph Space, building a non parametric strategy to estimate intervals for labelled and unlabelled networks.

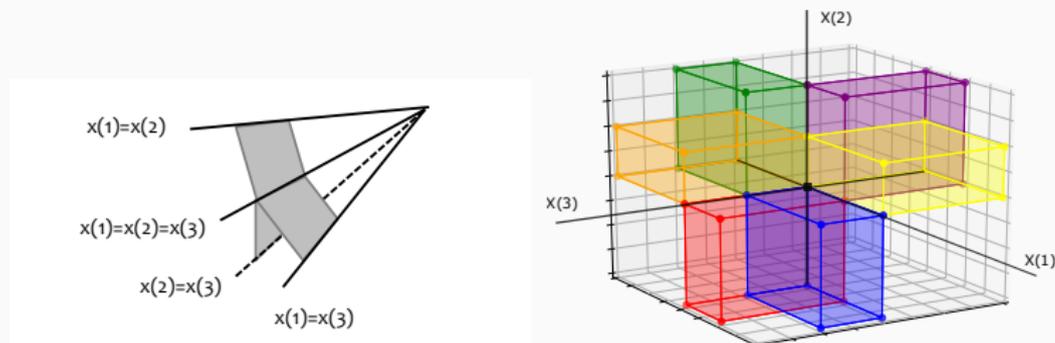


Figure 7: Conceptual visualization of the shape of an interval in the Graph Space and its back-projection π^{-1} in the total space for an un-directed graph with three nodes and no attributes on the nodes.

Geometric Proofs

Lemma ([C. et al., 2020a])

Graph Space X/T is not a manifold

Adding-null nodes model choice: as a permutation $t \in T$ which acts only on the null nodes of a flattened adjacency matrix $x \in X$ does not have any effect on the adjacency matrix, such x will have a larger isotropy subgroup than generic points in X .

\Rightarrow The structure of the isotropy subgroup $T_x = \{t \in T | tx = x\}$ varies for different points $x \in X$ [Bredon, 1972].

Lemma ([C. et al., 2020a])

Graph Space is a geodesic space.

Proof.

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

Lemma ([C. et al., 2020a])

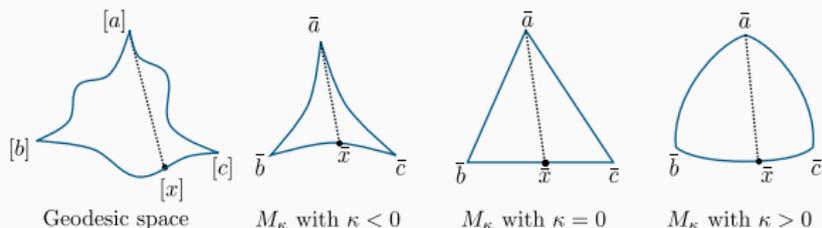
There exist points $y_\varepsilon, z_\varepsilon \in X/T$ which are connected by more than one geodesic.

Proof.

We give an example in the case where node- and edge attributes are scalar. Note that the same example can be adapted to vector valued node- or edge attributes by appending this scalar value with zeros. Let $\varepsilon > 0$, and consider the two graphs y_ε and z_ε shown in the top row on the left hand side of Figure in slide nr. 13. 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. The midpoint of this geodesic is the graph γ_{mid}^n shown on the left hand side of the bottom row. The second geodesic between y_ε and z_ε consists of interpolating the edge attributes with the node matching indicated by the coloring of the nodes. The midpoint of this geodesic is the graph γ_{mid}^e shown on the right hand side of the bottom row. The two points connected by two geodesics are illustrated schematically on the right hand side. \square

Corollary

Geodesics connecting pairs of points are not generally unique in X/T . \square



Definition ($CAT(\kappa)$ space, curvature in the sense of Alexandrov)

Let $(\mathfrak{X}, d_{\mathfrak{X}})$ be a geodesic metric space, and let abc be a geodesic triangle in \mathfrak{X} as described above. Note that any point x from the segment γ_{bc} has a corresponding point \bar{x} on the segment $\bar{\gamma}_{\bar{b}\bar{c}}$ in the comparison triangle, such that $d_{M_{\kappa}}(\bar{x}, \bar{b}) = d_{\mathfrak{X}}(x, b)$. If

$$d_{\mathfrak{X}}(x, a) \leq d_{M_{\kappa}}(\bar{x}, \bar{a}) \quad (7)$$

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. 7. 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 $CAT(\kappa)$.

Regression: Comments and Proofs

Algorithm 1: AAC for Generalized Geodesic Regression [C. et al., 2020b]**Data:** $\{(\underline{s}_1, [x_1]), \dots, (\underline{s}_k, [x_k])\} \in \mathbb{R}^p \times X/T$ **Result:** Generalized Geodesic Regression $f(\underline{s}) \in \Gamma(X/T)$ **Initialize:***Pick rand. $t_i x_i \in [x_i]$ among $\{[x_1], \dots, [x_k]\}$, $t_i \in T$;**Align the data to $t_i x_i$, obtaining $\{t_1 x_1, \dots, t_k x_k\} \in X$ opt. al. wrt $t_i x_i$;**Regression $h(\underline{s}) \in \Gamma(X)$ by solving 4 for $\{(\underline{s}_1, t_1 x_1), \dots, (\underline{s}_k, t_k x_k)\}$ in X ;**Project onto $\Gamma(X/T)$ as $f(\underline{s}) = \pi \circ h(\underline{s})$ and set $\tilde{f}(\underline{s}) = f(\underline{s})$* **while** $\delta > \varepsilon$ **do**Align all the points $\{[x_1], [x_2], \dots, [x_k]\}$ with respect to the generalized geodesic regression $\tilde{f}(\underline{s})$, obtaining a new set of aligned points $t_1 x_1, \dots, t_k x_k \in X$;Perform GGR on $\{(\underline{s}_1, t_1 x_1), \dots, (\underline{s}_k, t_k x_k)\}$ in X obtaining $h(\underline{s}) \in \Gamma(X)$ by solving 6;Project onto $\Gamma(X/T)$ as $f(\underline{s}) = \pi \circ h(\underline{s})$;Compute a step as the distance between the sum of square prediction errors $\delta = \mathcal{D}(\tilde{f}(\underline{s}), f(\underline{s}))$;Align all the observations wrt $f(\underline{s})$, obtaining a set of points $\{t_1 x_1, \dots, t_k x_k\} \in X$ as explained in 5;Set $\tilde{f}(\underline{s}) = f(\underline{s})$.Return $f(\underline{s}) \in \Gamma(X/T)$

Theorem (Convergence [C. et al., 2020b])

Let Graph Space X/T be endowed with a probability measure η which is absolutely continuous with respect to the the push forward of the Lebesgue measure m on X , and let λ be a probability measure absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^P . Let the sample $\{(\underline{s}_1, [x_1]), \dots, (\underline{s}_k, [x_k])\} \subset \mathbb{R}^P \times X/T$ be sampled from $\lambda \times \eta$. Assume that the AAC for Regression (Algorithm 1) fits the regression model f_β defined in Definition 9. Assume moreover that $S \in \mathbb{R}^{k \times P}$ has the first column $[1, \dots, 1]^T$ and full rank. Under these circumstances, we claim that:

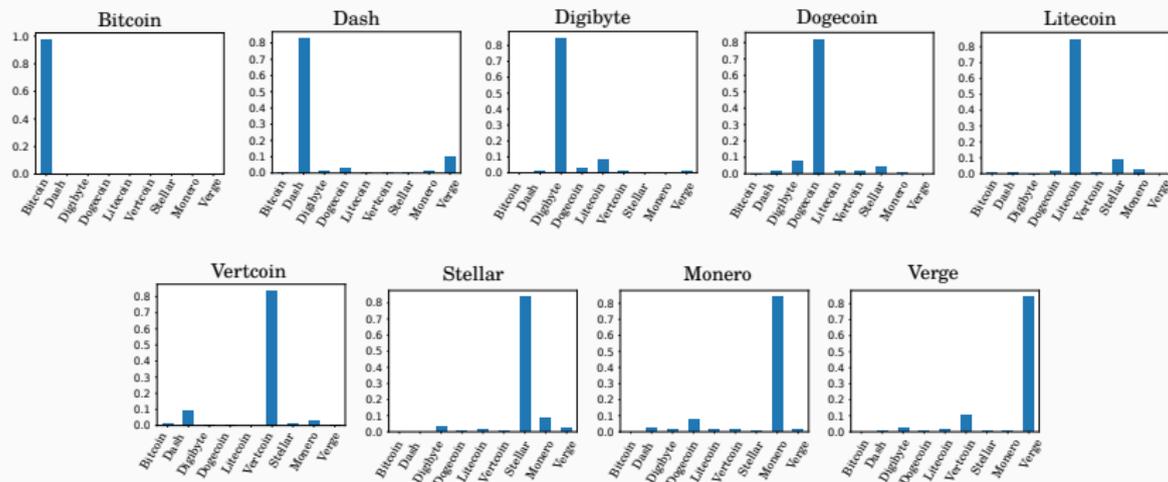
- The AAC algorithm terminates in finite time
- With probability 1, the estimated regression curve f_β returned by the AAC algorithm is a local minimum of the function

$$\beta \mapsto \sum_{i=1}^k d_{X/T}^2([x_i], f_\beta(\underline{s}_i))$$

In other words, the theorem shows that **the alignment procedure** and **the regression estimation procedure** lead to the same minimum.

Cryptocurrency Permutations

The interpretation of the permutations between cryptocurrencies can help understanding the interchangeable relational role.



Is the role of *Indre By* interchangeable over time?
Is there any similar area in terms of mobility?

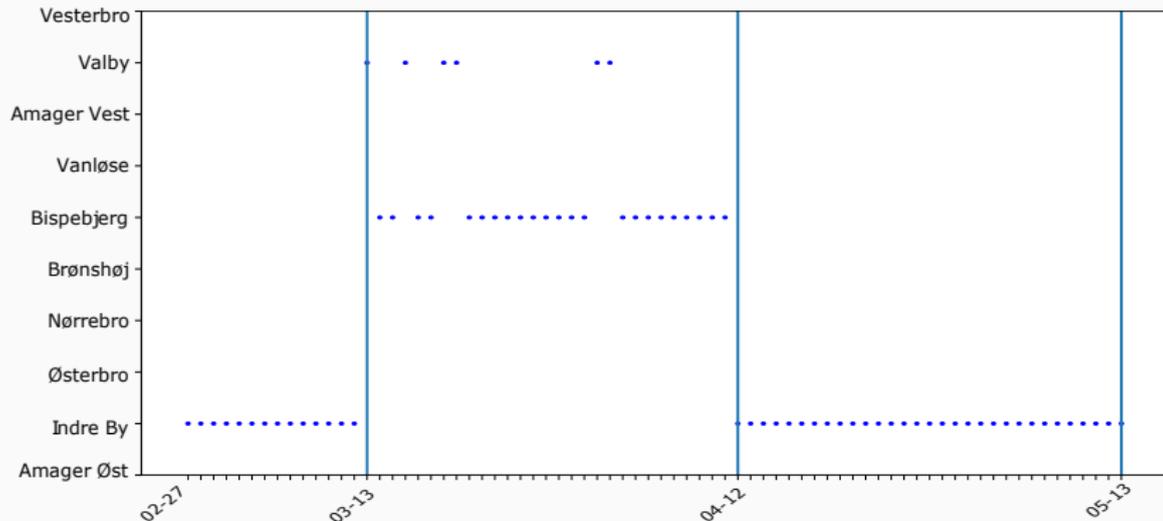


Figure 8: The permutation of the *Indre By* area over time.

The graph-on-vector regression problem generates interesting further developments in both applications and theoretical modelling:

1. Extend the AAC Regression to other regression model in the X space rather than OLS, such as Gaussian Process Regression, Neural Networks, etc.
2. Focus on building regression models for the analysis of time series, taking into account the time dependency
3. Building a graph-on-graph regression model, moving toward equivariant rather than invariant models.

First, we prove convergence in finite time. Algorithm 1 consists of two steps repeated iteratively. Given the observations $\{(s_1, [x_1]), \dots, (s_k, [x_k])\}$ in $\mathbb{R}^p \times X/T$, we shall fit a generalised regression model which is of the form $f(s) = \pi \circ h(s)$ where $h: \mathbb{R}^p \rightarrow X$ is a linear regression model on X . Consider the squared error loss function

$$\sum_{i=1}^k d_X^2(h^{cur}(s_i), x_i^{cur}), \quad (8)$$

where, $h^{cur}(s_i)$ is our current estimation of the regression curve corresponding to input s_i in X , and x_i^{cur} is the current representative in X of the sample network $[x_i]$. Note that both steps of alignment and estimation of the regression cannot increase the loss function. Moreover, if the value of (8) stays fixed two iterations in a row, the algorithm will terminate. Thus, the iterative algorithm will never see the same set of sample-wise alignments twice. As there are only finitely many such alignments, the algorithm is forced to terminate in finite time.

Next, we turn to proving that the estimated regression curve f_{β} is, indeed, a local minimum. We need to show that for some $\epsilon > 0$, $\|\tilde{\beta} - \beta\| < \epsilon$ implies that, with probability 1,

$$\sum_{i=1}^k d_{X/T}^2([x_i], f_{\beta}(s_i)) \leq \sum_{i=1}^k d_{X/T}^2([x_i], f_{\tilde{\beta}}(s_i))$$

We shall rely on the following lemma:

Lemma

Given representatives x_1, \dots, x_k of $[x_1], \dots, [x_k]$ with generalized linear regression model $f(s) = \pi \circ h(s)$ obtained minimizing (8), $h(s): \mathbb{R}^p \rightarrow X$, the following holds with probability 1:

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

$$d_X(h(s_i), x_i) \neq d_X(h(s_i), tx_i),$$

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

This Lemma states that given a regression line and an optimally aligned point to its prediction, with probability 1 this point is unique.

If the lemma holds, we now consider $\tilde{\beta} \in B(\beta, \varepsilon)$. We wish to show that for all $i = 1, \dots, k$ and all $t \in T \setminus T_{x_i}$, we have $d(\tilde{h}(s_i), x_i) < d(\tilde{h}(s_i), tx_i)$, namely that the optimal representative of $[x_i]$ is left unchanged for all $i = 1, \dots, k$, even if we perturb the regression line. Given:

$$\nu = \min\{d_X(h(s_i), tx_i) - d_X(h(s_i), x_i) \mid i = 1, \dots, k, t \in T \setminus T_{x_i}\} > 0.$$

and starting from:

$$d(h(s_i), x_i) \leq d(h(s_i), tx_i) - \nu,$$

we get:

$$d(\tilde{h}(s_i), x_i) < d(\tilde{h}(s_i), tx_i)$$

which proves that even if we perturb the regression line the alignment is the same, so the function $f_\beta = \pi \circ h_\beta$ is a local minimum.

Lemma

Let $\beta \in \mathbb{R}^{p \times J}$ be the parameters of the output of AAC as stated in Theorem 15. Then, with probability 1, $\beta_{j_1} \neq \beta_{j_2} \in \mathbb{R}^p$ for all $j_1 \neq j_2 \in \{1, \dots, k\}$. In particular, $t\beta \neq \beta$ for all $t \in T \setminus \{Id\}$.

The forehead Lemma 17 can be rephrased using the set

$T_{h(s_i)} = \{t \in T : th(s_i) = h(s_i)\}$. Stating that $th(s_i) = h(s_i)$ happens with probability 0 $\forall t \in T \setminus \{Id\}$.

Proof.

Recall from Eq. (??) that

$$\hat{\beta} = (S^T S)^{-1} S^T X.$$

Since, by the assumptions of the theorem, S has full rank, so does

$(S^T S)^{-1} S$. In particular, $(S^T S)^{-1} S$ is a submersion. Thus, if $\beta_{j_1} = \beta_{j_2}$ for some $j_1 \neq j_2$, then the corresponding columns of X belong to the same fiber of $(S^T S)^{-1} S^T$, which happens with probability 0. \square

We show that the following set

$$\mathcal{X}_T = \left\{ \begin{array}{l} ((s_1, [x_1]), \dots, (s_k, [x_k])) \\ \in (\mathbb{R}^p \times X/T)^k \end{array} \middle| \begin{array}{l} d(h(s_i), x_i) = d(h(s_i), tx_i) \\ \text{for some representatives } x_1, \dots, x_k, \\ i = 1, \dots, k \text{ and } t \in T \setminus T_{x_i} \end{array} \right\}$$

has measure $(\lambda \times \eta)_k(\mathcal{X}_T) = 0$. For each element $t \in T$, denote by

$X^t = \{x \in X \mid tx = x\}$ the fixed point set of t . Note that

$(\lambda \times \eta)_k(\mathcal{X}_T) = (\lambda \times m)_k((\text{Id}_{\mathbb{R}^p} \times \pi)^{-1}(\mathcal{X}_T))$, and that

$$(\text{Id}_{\mathbb{R}^p} \times \pi)^{-1}(\mathcal{X}_T) = \bigcup_{i=1}^k \bigcup_{t \in T} \mathcal{X}_{i,t},$$

where

$$\mathcal{X}_{i,t} = \{(s_1, x_1, \dots, s_k, x_k) \in (\mathbb{R}^p \times X) \times \dots \times \underbrace{(\mathbb{R}^p \times X \setminus X^t)}_{i\text{th}} \times \dots \times (\mathbb{R}^p \times X) \mid d_X(h(s_i), x_i) = d_X(h(s_i), tx_i)\}$$

and

$$\mathcal{X}_{i,t} \subset \underbrace{(\mathbb{R}^p \times X) \times \dots \times (\mathbb{R}^p \times X)}_k.$$

To prove that this set has measure zero, we define the corresponding function \mathcal{F}_i :

$$\mathcal{F}_i: \underbrace{X \times \cdots \times X}_k \rightarrow \mathbb{R}, \quad (s_1, x_1, \dots, s_k, x_k) \mapsto d_X^2(h(s_i), x_i) - d_X^2(h(s_i), tx_i)$$

The preimage $\mathcal{F}_i^{-1}(0)$ of the function satisfies

$$\mathcal{F}_i^{-1}(0) \cap (\mathbb{R} \times X) \times \dots \times \underbrace{(\mathbb{R} \times X \setminus X^t)}_{i^{th}} \times \dots \times /(\mathbb{R} \times X) = \mathcal{X}_{i,t}.$$

We show that \mathcal{F}_i is a submersion on

$(\mathbb{R} \times X) \times \dots \times \underbrace{(\mathbb{R} \times X \setminus X^t)}_{i^{th}} \times \dots \times /(\mathbb{R} \times X)$ by showing that it has nonzero

gradient.

Note that

$$\begin{aligned}
 \mathcal{F}_i(s_1, x_1, \dots, s_k, x_k) &= d_X^2(h(s_i), x_i) - d_X^2(h(s_i), tx_i) \\
 &= \|h(s_i) - x_i\|^2 - \|h(s_i) - tx_i\|^2 \\
 &= (h(s_i) - x_i)^T (h(s_i) - x_i) - (h(s_i) - tx_i)^T (h(s_i) - tx_i) \\
 &= 2h(s_i)^T (tx_i - x_i).
 \end{aligned}$$

Given the Lemma 17, it follows that

$$\nabla_{x_i} \mathcal{F}_i(s_1, x_1, \dots, s_k, x_k) = 2h(s_i)^T (t - I) \neq 0,$$

where the last inequality follows from $t \notin T_{h(s_i)}$.

This prove that the set $X_{i,t}$ has measure zero for every i, t so the union has measure zero. This proves Lemma 16, which prove the Theorem 15. \square

Geodesic Principal Components

Definition (Optimal position)

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

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

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

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

Algorithm 2: Optimal Position with respect to a generalized geodesic

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

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

for $s \leftarrow s_{min}$ **to** s_{max} **do**

 Find $t(s) := \arg \min_{t \in T} d_X(tx, \gamma(s))$

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

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

The selected $t^* \in T$ is the geometric projection of the point x on the geodesic γ in the interval selected.

After defining a concept of generalized geodesic, a concept of orthogonality, a concept of optimal alignment with respect to a generalized geodesic:

Definition (Geodesic Principal Components -I [C. et al., 2020a])

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:

$$\delta_1 = \arg \min_{\delta \in \Gamma(X/T)} \sum_{i=1}^k (d_{X/T}^2([x_i], \delta)) \quad (9)$$

- The *second generalized geodesic principal component* $\delta_2 \in \Gamma(X/T)$ minimizes (9) 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 .

Definition (Geodesic Principal Components -II)

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

$$\sum_{i=1}^k (d_{X/T}^2([x_i], [\mu]))^2 \quad (10)$$

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 (9) over all generalized geodesics that meet orthogonally $\delta_1, \dots, \delta_{j-1}$ and cross μ .

Algorithm 3: AAC for the GPCA [C. et al., 2020a]

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

Result: Geodesic Principal Components represented as subspaces of $R^{n \times n}$

Initialization: Select randomly an observed graph and randomly a

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

while *It converges* **do**

 Perform PCA in $R^{n \times n}$;

 Align all the data in optimal position with respect to the first Principal
 component using Algorithm 2

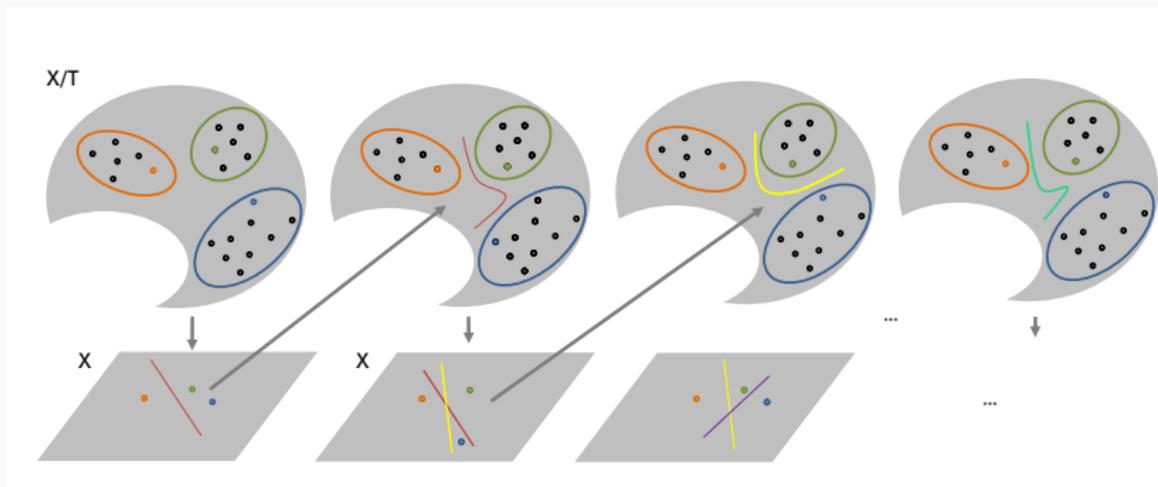
The algorithm converges when the proportion of variance explained is less than a selected threshold.

Theorem

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

$$\sum_{i=1}^k d_{X/T}^2(\delta, [x_i]). \quad (11)$$

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



The convergence of the algorithm is based on the same idea as the one for the Fréchet Mean, because the function to estimate the PCA in the total space X is based on the Euclidean distance, the same distance used to align the points with respect to the generalized geodesics.

Conformal Prediction Intervals

Find an interval $[C]_{k,1-\alpha}$ built on k data with a coverage $1 - \alpha$

$$\mathbb{P}_\eta ([X]_{k+1} \in [C]_{k,1-\alpha}) \geq 1 - \alpha \quad (12)$$

We start by splitting our set of unlabelled graphs $\{[X_1], \dots, [X_k]\}$ in a training set \mathcal{I}_1 and a calibration set \mathcal{I}_2 . $\forall [X_m], m \in \mathcal{I}_2$, one can compute an empirical P-value define exactly as in the labelled case:

$$p_{[X_m]} := \frac{|\{i \in \mathcal{I}_2 : R_i \geq R_m\}|}{|\mathcal{I}_2| + 1} \quad (13)$$

where R is a non-conformity measure. The conformal prediction set:

$$[C]_{k,1-\alpha} := \{[X] \in X/T : p_{[X]} > \alpha\} \quad (14)$$

We define our non-conformity measure R_m to be

$$R_m = \max_{j=1, \dots, p} |t_m X_m(j) - \mathcal{A}(j)| \quad (15)$$

where:

$$t_m = \arg \min_{t \in T} (d_X(t X_m, \mathcal{A})) \quad (16)$$

where \mathcal{A} is a symmetric function of the data.

This non-conformity measure selects the permutation that optimally aligns the two graphs and consequently selects the edge or node that are mostly far apart from each other.

In the case of amplitude modulation, the Equation 15 becomes:

$$R_m = \max_{j=1, \dots, p} \frac{|t_m X_m(j) - \mathcal{A}(j)|}{\hat{s}(j)} \quad (17)$$

Where \hat{s} is an estimator of the variability of the edge or node j after the alignment with respect to the central estimator \mathcal{A} . The whole procedure is summarised in the following Algorithm.

Split Conformal Prediction Parallelotopes for Populations of Unlabelled Graphs with Amplitude Modulation:

Algorithm 3 Split Conformal Prediction Parallelotopes for Populations of Unlabelled Graphs with Amplitude Modulation

- 1: **Require:** Data $[X_i], i = 1, \dots, k,$, type-1 error level $\alpha \in (0, 1)$, Predictive algorithm \mathcal{A} , amplitude modulation algorithm s
 - 2: split randomly $\{1, \dots, k\}$ into two subsets $\mathcal{I}_1, \mathcal{I}_2$
 - 3: $\hat{\mu} = \mathcal{A}(\{[X_l], l \in \mathcal{I}_1\})$, $\hat{s} = \mathcal{S}(\{[X_l], l \in \mathcal{I}_1\})$
 - 4: Find $\{t_1, \dots, t_{|\mathcal{I}_2|}\}$ s.t. $t_m = \arg \min_{t \in T} (d_X(t_m X_m, \hat{\mu}))$
 - 5: $R_m = \max_{j=1, \dots, p} \left(\frac{|(t_m X_m)(j) - \hat{\mu}(j)|}{\hat{s}(j)} \right) = \max_{j=1, \dots, p} \left(\frac{|X_m(\sigma_{t_m}(j)) - \hat{\mu}(j)|}{\hat{s}(j)} \right)$, $m \in \mathcal{I}_2$
 - 6: h is equal to $\lceil (|\mathcal{I}_2| + 1)(1 - \alpha) \rceil$ -th smallest value in the set $\{R_m : m \in \mathcal{I}_2\}$
 - 7: **Ensure:** $\mathcal{C}_{k, 1-\alpha} := \{[X] \in X/T : (tX)(j) \in [\hat{\mu}(j) - h\hat{s}(j), \hat{\mu}(j) + h\hat{s}(j)] \quad \forall j = 1, \dots, p, \quad \forall t \in T\}$
-