Geometry and Statistics, Toulouse 09/2019

Analysis of Populations of Networks:

Graph Spaces and the Computation of Summary Statistics.

Anna Calissano
Mox Modelling Scientific Computing
Dept. Mathematics
Politecnico di Milano
anna.calissano@polimi.it

POLITECNICO
MILANO 1863
DIPARTIMENTO DI MATEMATICA

Team

Aasa Faragen

Simone Vantini

UNIVERSITY OF COPENHAGEN

DTU Technical University of Denmark

MOX

POLITECNICO MILANO 1863

DIPARTIMENTO DI MATEMATICA

Index

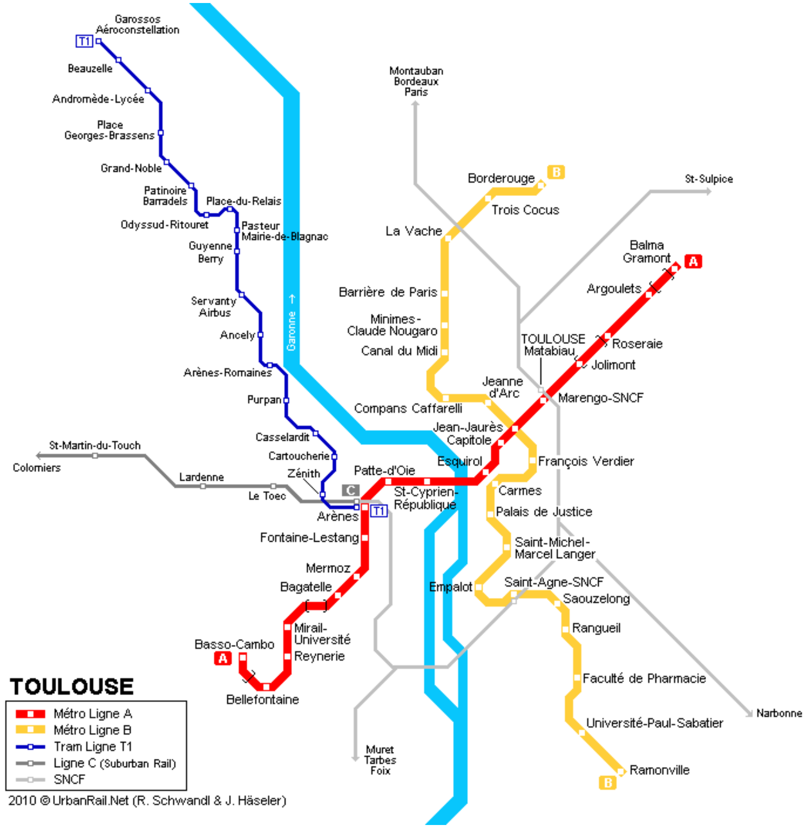
Object Oriented Data Analysis: Population of Networks

Structure Spaces: Graph Space as a particular case

Summary Statistics: Mean and Geodesic PCA

Population of Networks

Network Analysis



Datum:
Nodes and Edges

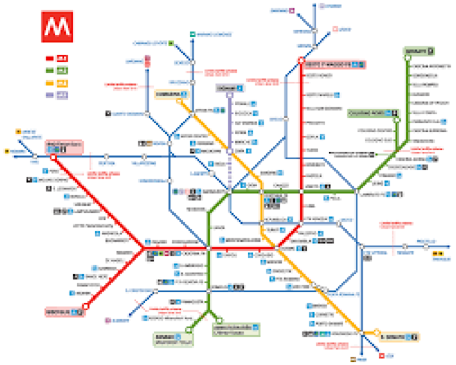
Analysis focusing on:
Features: nodes and edge attributes

Relations: how nodes influence each other

«1° Generation Approach»

Marron, J. Steve, and Andrés M. Alonso. "Overview of object oriented data analysis." *Biometrical Journal* 56.5 (2014): 732-753.

Network Analysis: from one to many

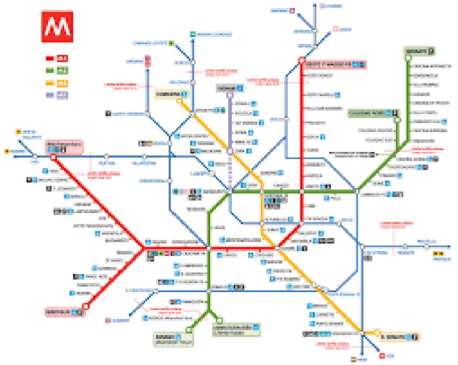


Objects are Networks $X_1, X_2, \ldots$

«2° Generation Approach»

Marron, J. Steve, and Andrés M. Alonso. "Overview of object oriented data analysis." *Biometrical Journal* 56.5 (2014): 732-753.

Network Analysis: from one to many



Many new questions arise:

- How can we describe these networks?
- How can we relate nodes in networks?
- Along which relations/features are different?
- Can we do any statistical analysis and how?

OODA for Networks

OODA for Networks: State of the Art

OODA for trees Wang and Marron (2007), Aydın, et al. (2009), Feragen et al. (2013), Nye, et al. (2017),

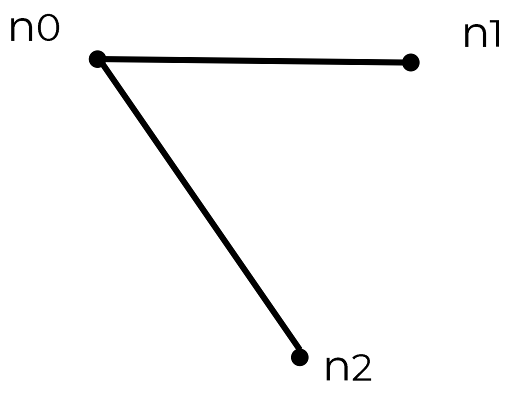
Hypothesis Testing Simpson, et al. (2013), Ginestet, et al. (2017), Lovato, et al. (2017)

Bayesian Generative Models Durante et al. (2017), Durante and Dunson (2018)

Graph Embedding Duvenaud et al. (2015)

Structure Spaces Jain et al. (2009)

Structure Spaces



A-attributed R-structure: $x = (P, R, \alpha)$

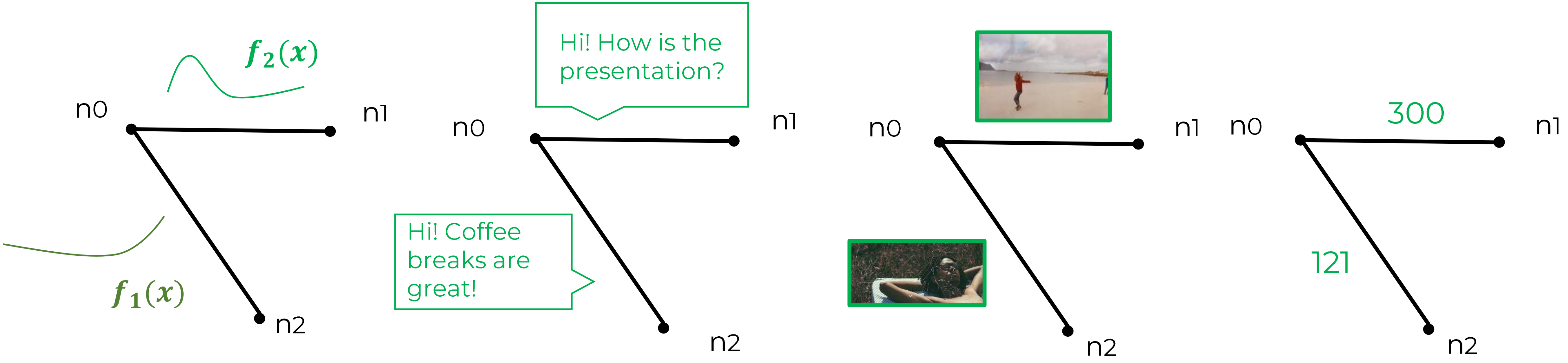
P- set of nodes

$R \in P^r$ - set of relations (edges if r=2).

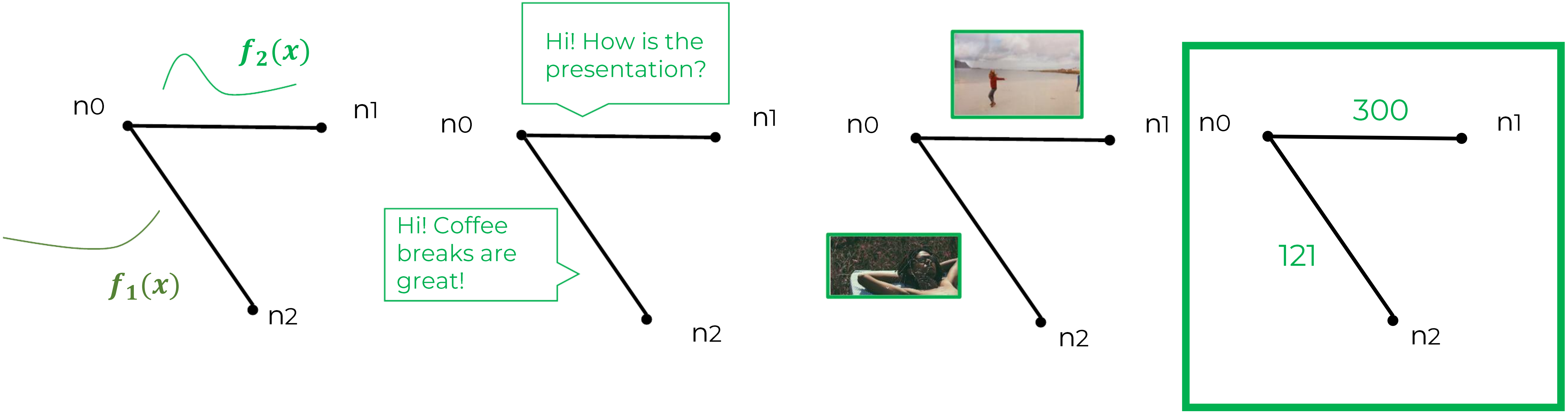
$\alpha: R \rightarrow M$ - function assigning attributes to edges

Jain, Brijnesh J., and Klaus Obermayer. "Structure spaces." *Journal of Machine Learning Research* 10.Nov (2009): 2667-2714.

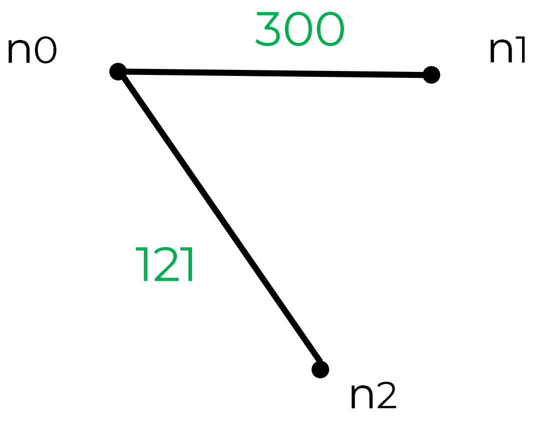
Structure Spaces



Structure Spaces



Graph Spaces



A-attributed R-structure: $\mathrm{x} = (P, R, \alpha)$

$\mathrm{P}=\{0,1,2\}$- set of nodes

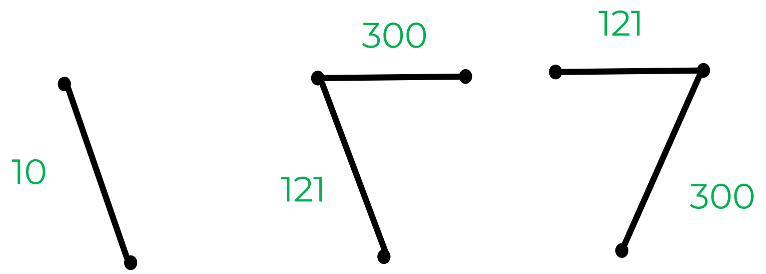
$R \in P^2, R = \{(0,1), (0,2)\}$ - set of relations

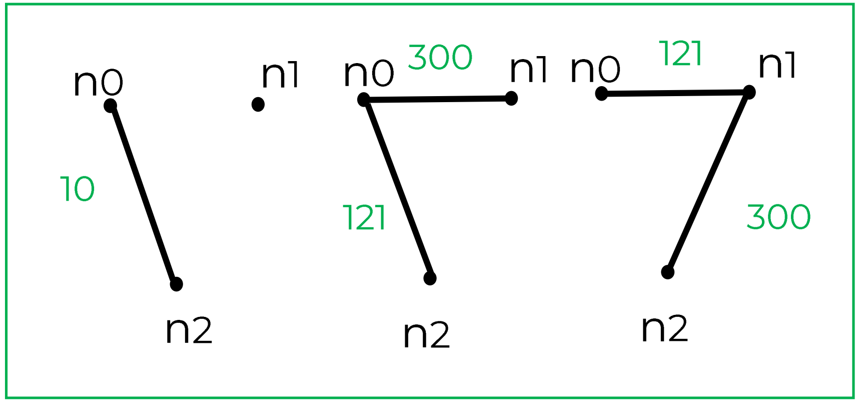
$\alpha\colon \mathrm{R} \rightarrow \mathbb{R}$ - weights

$\alpha\big((0,1)\big) = 300$

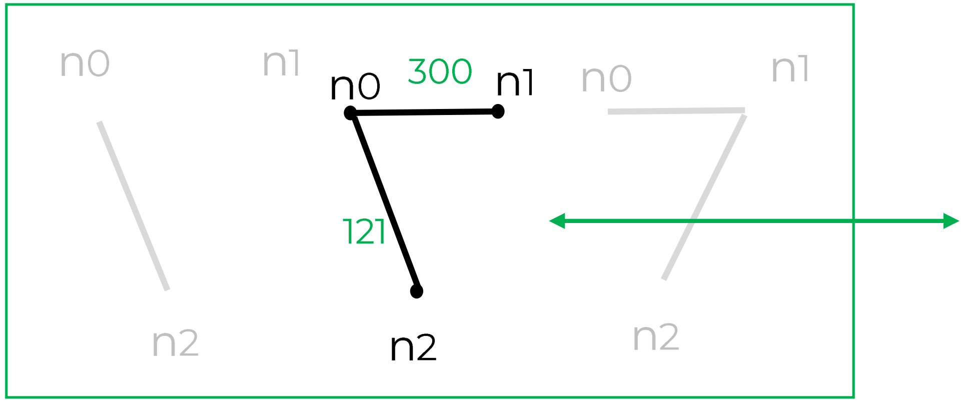
$\alpha\big((0,2)\big) = 121$

Graph Spaces





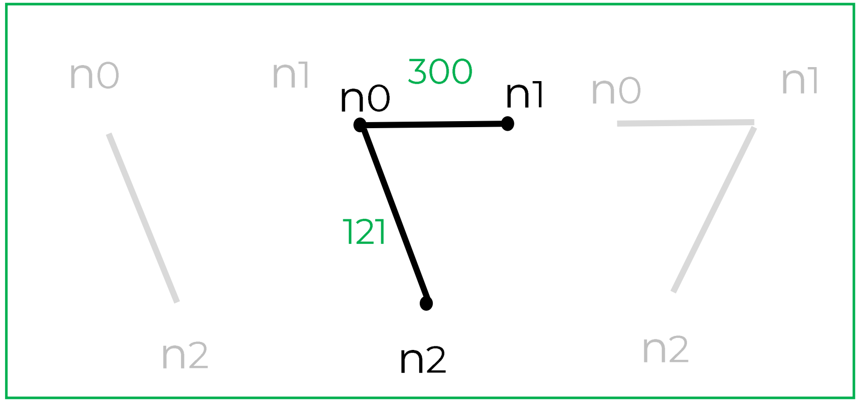
Graph Spaces



n0	300	121
300	n1	0
121	0	n2

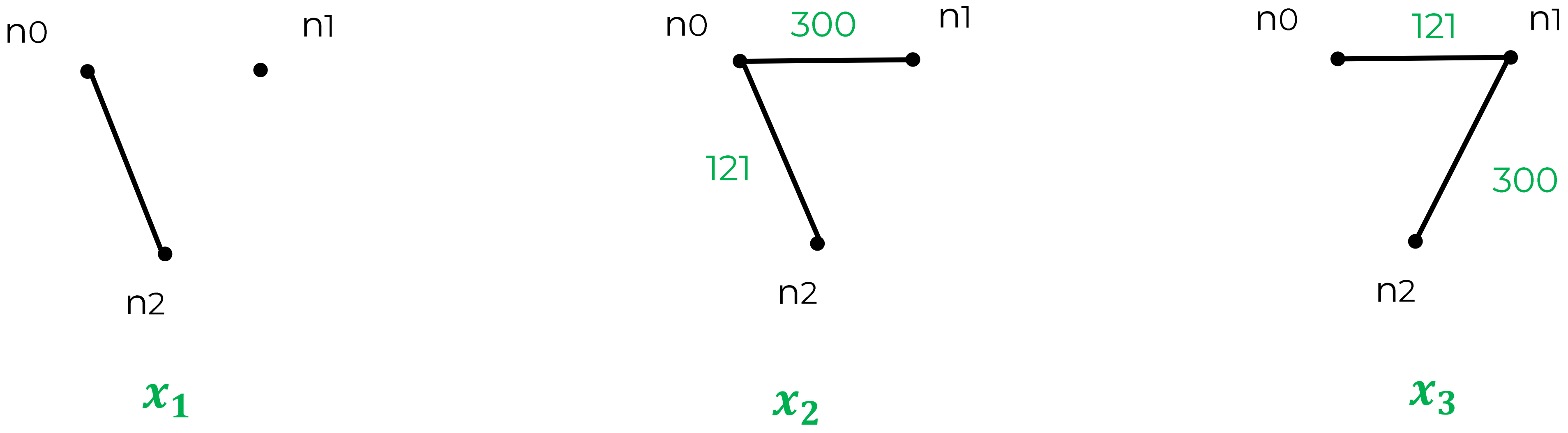
n0	300	121	300	n1	0	121	0	n2

Graph Spaces

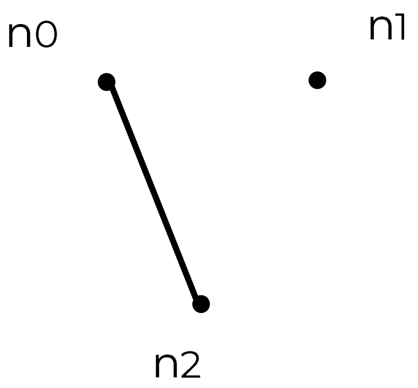


Space: $\mathbf{X} = \mathbb{R}^9$

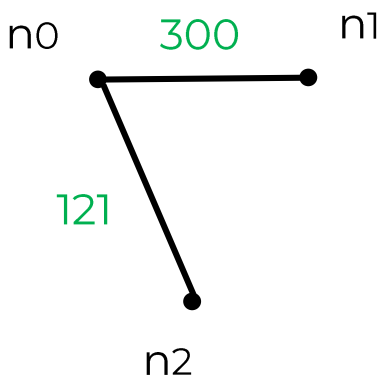
Graph Spaces



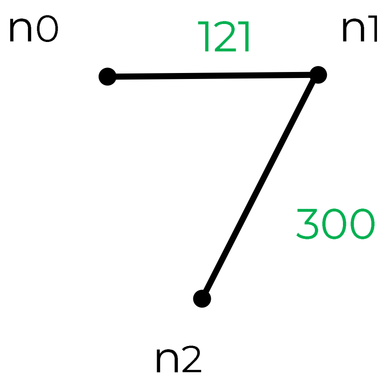
Graph Spaces



x_1



x_2

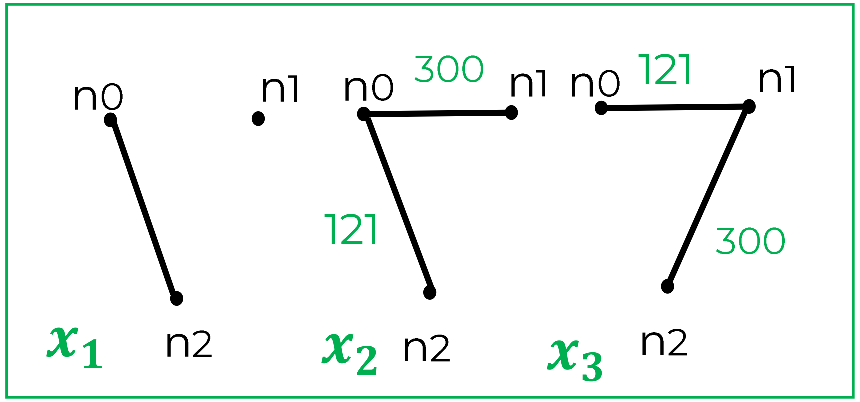


x_3

$x_2 = x_3$

Allowing permutation of nodes

Graph Spaces



$$\lfloor x_1 \rfloor = \{T x_1 : T \in \mathcal{T}\}$$

$$\lfloor x_2 \rfloor = \{T x_2 : T \in \mathcal{T}\}$$

Space: X

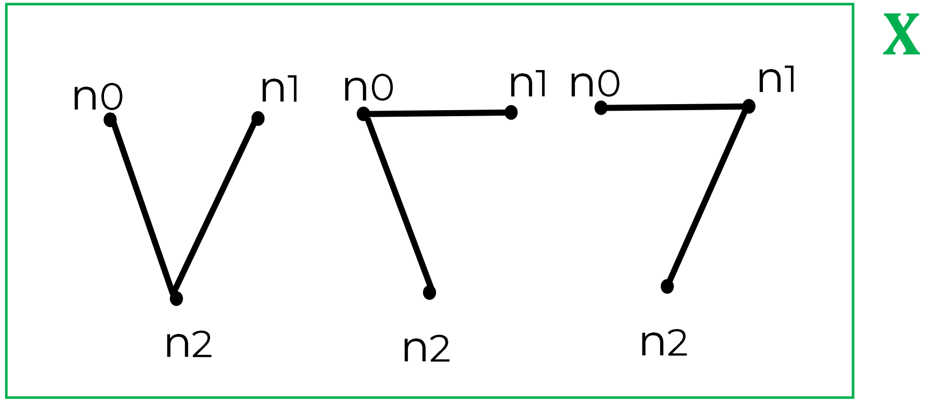
$\mathcal{T}$: **Permutation Action**

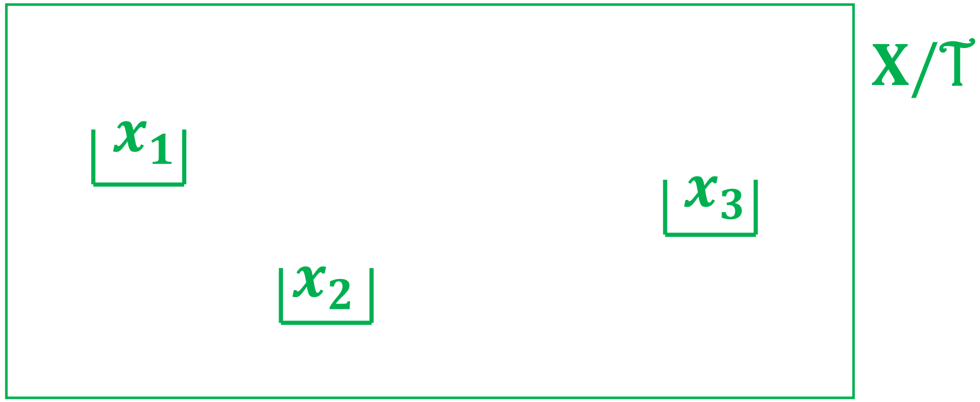
Permuting nodes

Graph Space: $X/\mathcal{T}$

Graph Space Properties

Metric Space

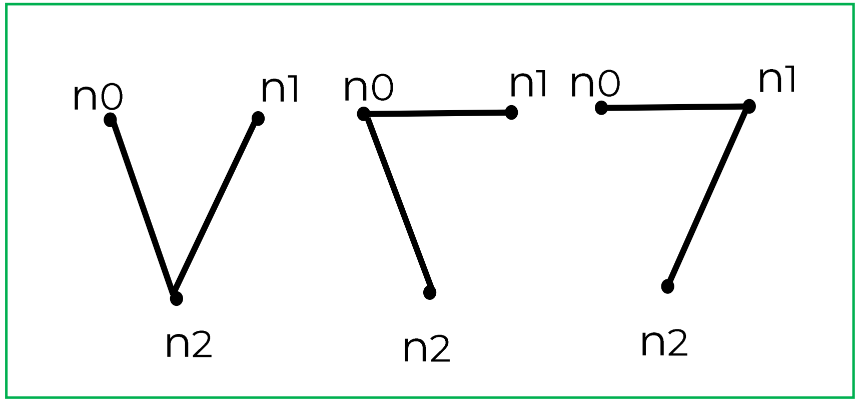




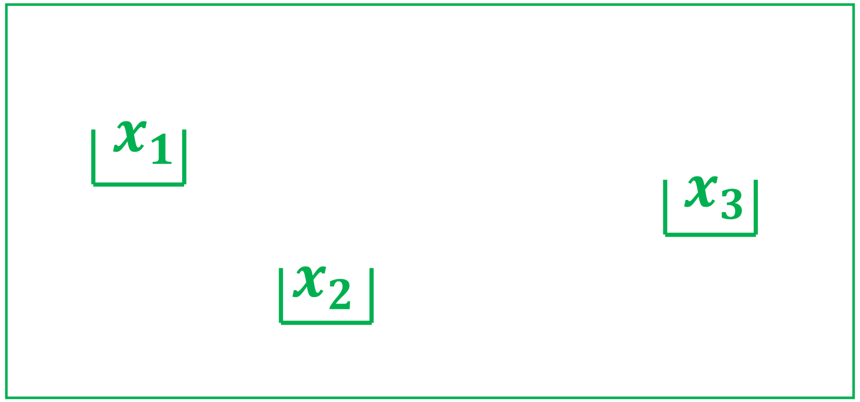
Metric Space
invariant with respect to permutation

Metric Space
Given two equivalent classes, find the permuted elements that have minimum distance

Geodesic Space



$\mathbf{X}$



$\mathbf{X}/\mathcal{T}$

Geodesic Space

Euclidean Space, complete, locally compact.

$\mathcal{T}$: **Finite Group Action**

Geodesic Space

Properties and Consequences

Not a Manifold (Not Free Action):

→ Can’t use all the literature about Manifold statistics

Unbounded:

→ No uniqueness of the geodesic even locally.

Isometric and Finite Dimension Action:

→ Allows to transfer easily computation from X to X/T

Properties and Consequences

Not a Manifold (Non Trivial Action):

→ Can't all the literature about Manifold statistics

Unbounded:

→ No uniqueness of the geodesic even locally.

Isometric and Finite Dimension Action:

→ Allows to transfer easily computation from X to X/T

Properties and Consequences

Not a Manifold (Non Trivial Action):
→ Can’t all the literature about Manifold statistics

Unbounded:
→ No uniqueness of the geodesic even locally.

Isometric and Finite Dimension Action:
→ Allows to transfer computations from X to X/T

Properties and Consequences

Isometric and Finite Dimension Action:

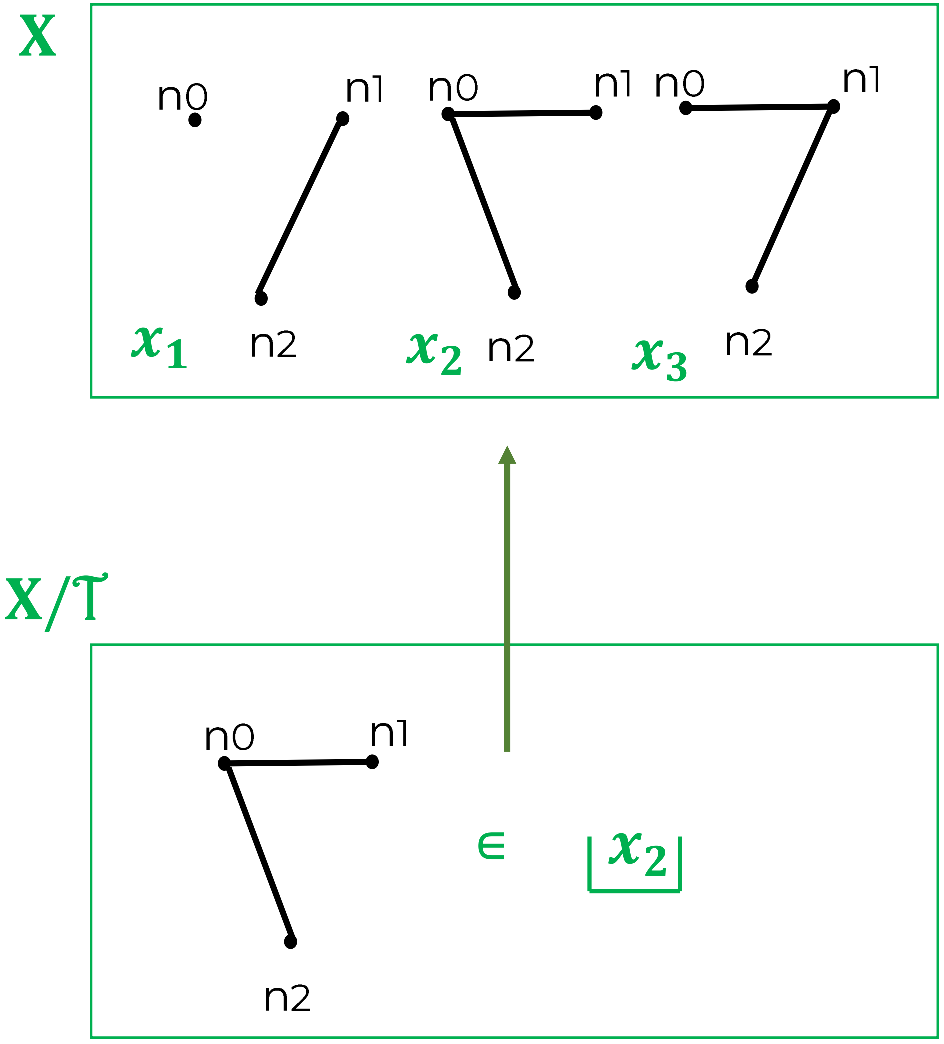
→ Allows to transfer easily computation from X to X/T

↓

Align All and Compute Algorithm to be able to compute statistics on this space.

Align All and Compute

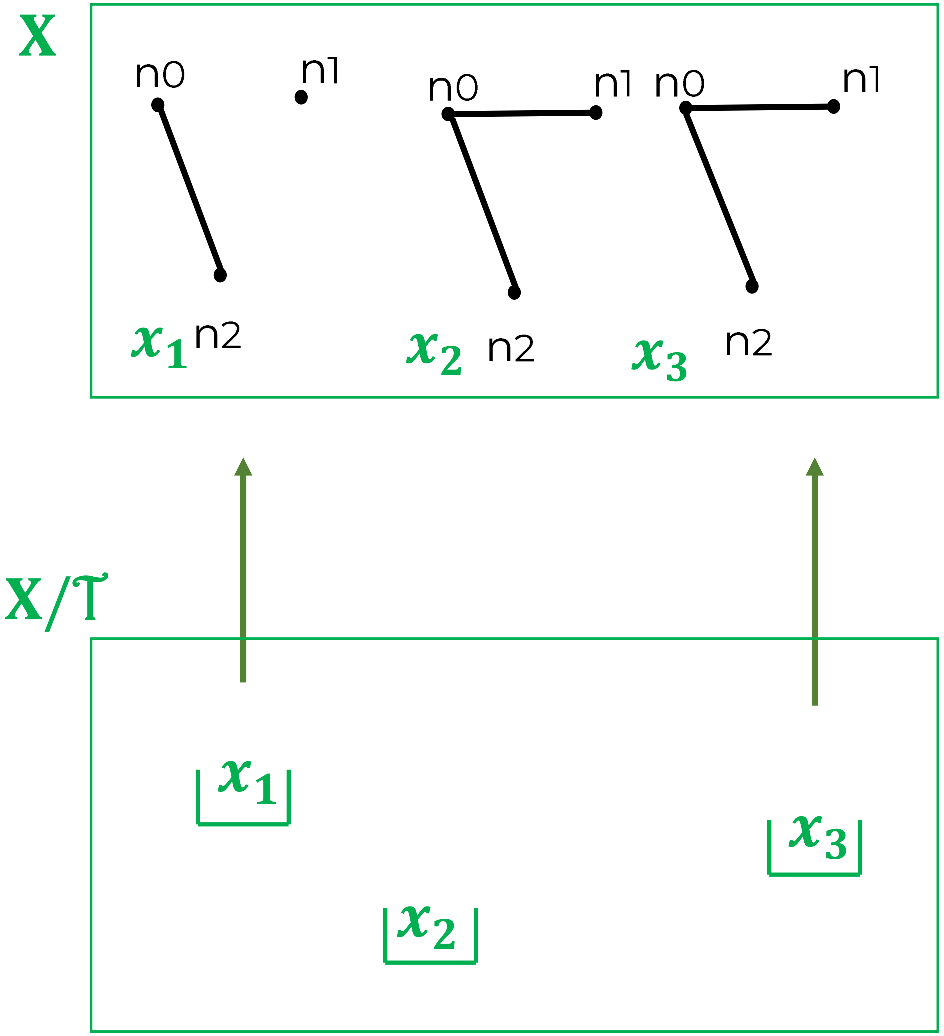
Align All and Compute



Based on the Generalized Procrustes Algorithm:

1) Select a random candidate point $\varkappa \in \lfloor x_2 \rfloor$ in **X/T**

Align All and Compute

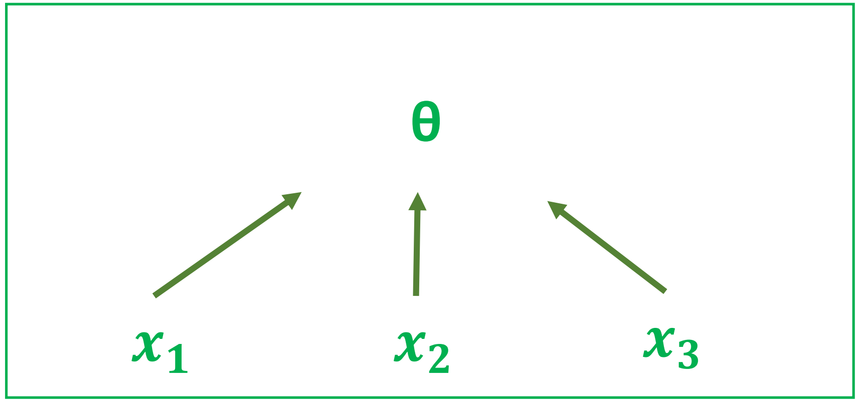


Based on the Generalized Procrustes Algorithm:

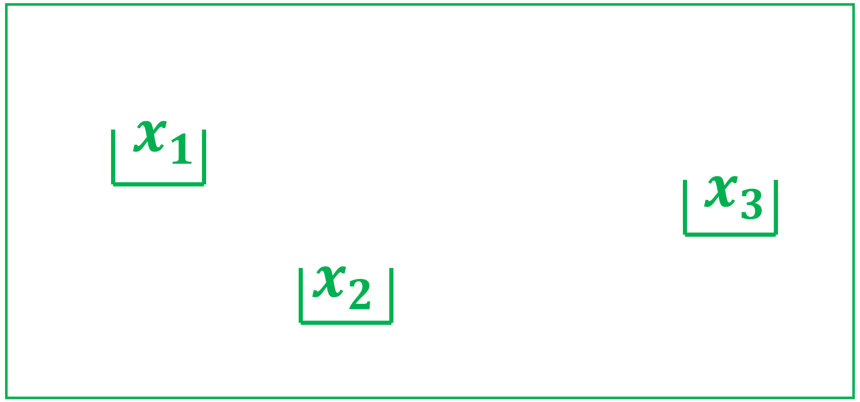
1) Select a random candidate point $\varkappa \in \lfloor x_2 \rfloor$ in $\mathbf{X/T}$
2) Align all the points to $\varkappa$ obtaining $x_1, x_2, \ldots, x_n$ in $\mathbf{X}$

Align All and Compute

X



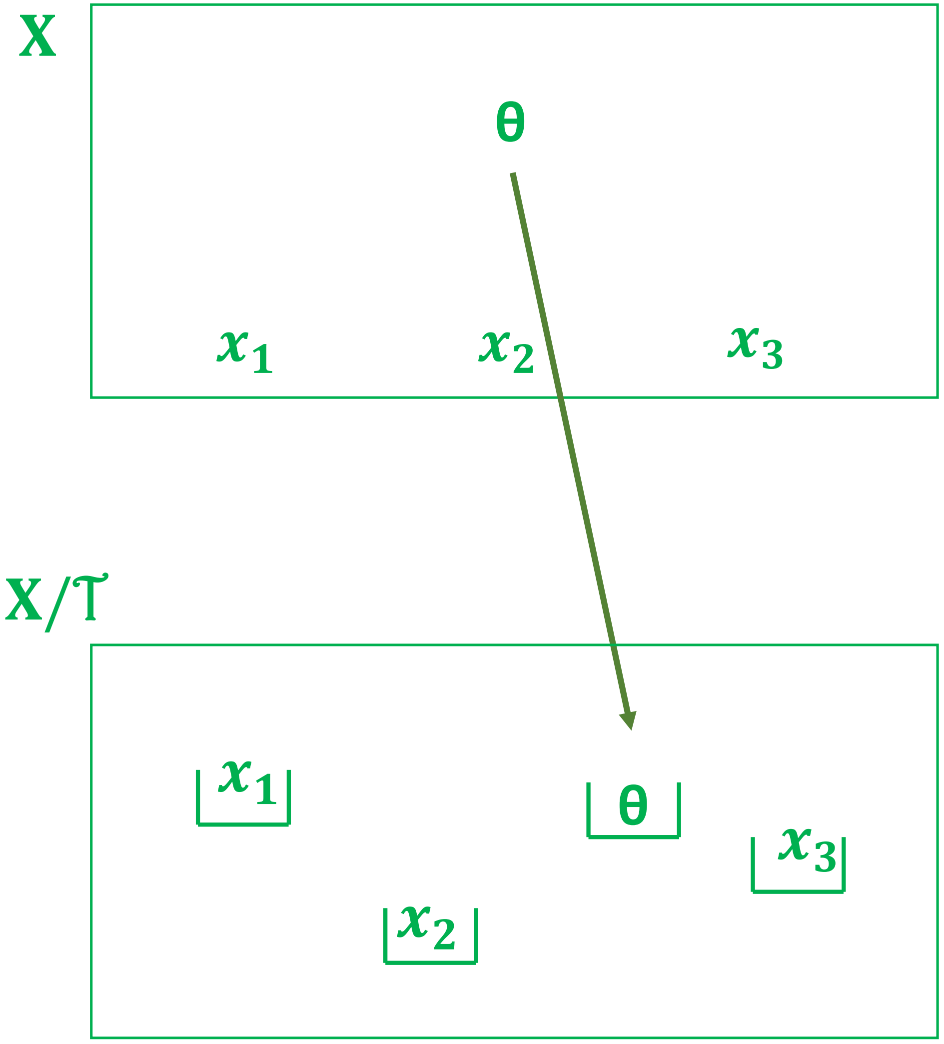
X/T



Based on the Generalized Procrustes Algorithm:

1) Select a random candidate point $\varkappa \in \lfloor x_2 \rfloor$ in **X/T**
2) Align all the points to $\varkappa$ obtaining $x_1, x_2, \ldots, x_n$ in **X**
3) Compute the Statistics θ in X

Align All and Compute



Based on the Generalized Procrustes Algorithm:

1) Select a random candidate point $\tilde{x} \in \lfloor x_2 \rfloor$ in $\mathbf{X/T}$
2) Align all the points to $\tilde{x}$ obtaining $x_1, x_2, \ldots, x_n$ in $\mathbf{X}$
3) Compute the Statistics θ in X
4) Set the $\varkappa = \theta$
5) Do 1 – 4 until the algorithm converge

AAC in Action

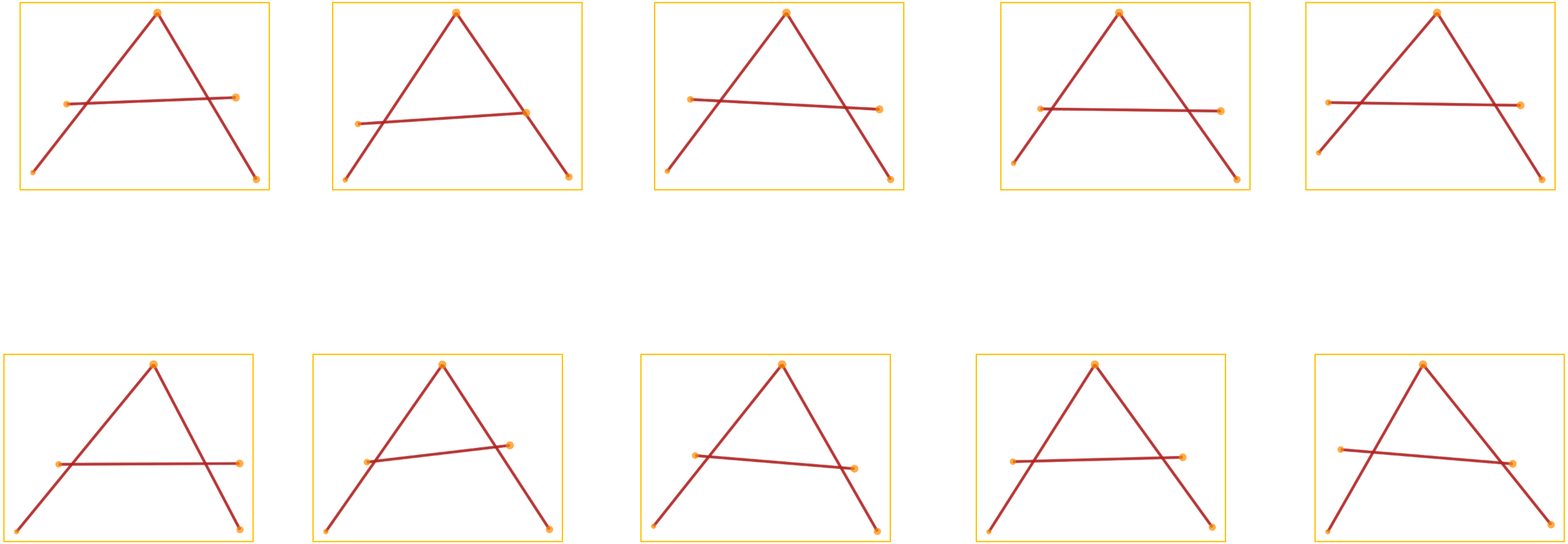
Align All and Compute Algorithm to be able to compute statistics on this space:

→ Fréchet Mean

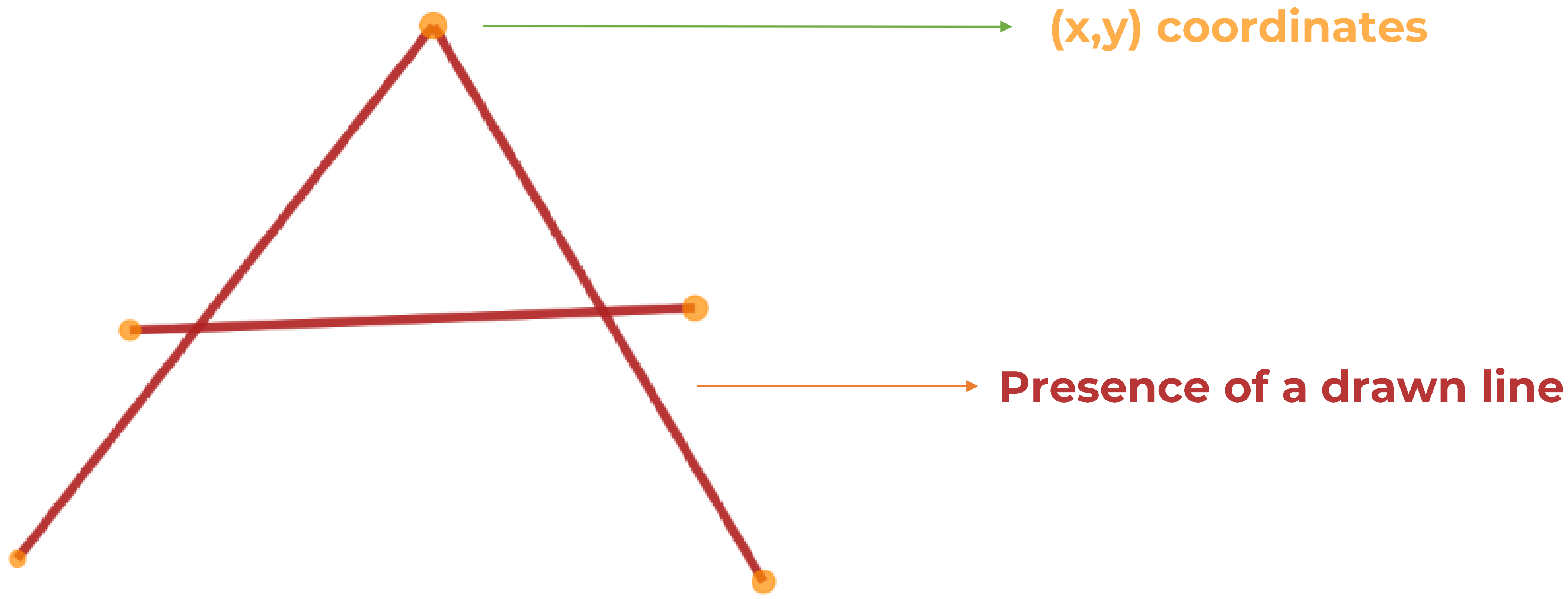
→ Geodesic Principal Components Analysis: following the framework introduced in Huckemann, Hotz, & Munk (2010)

Example

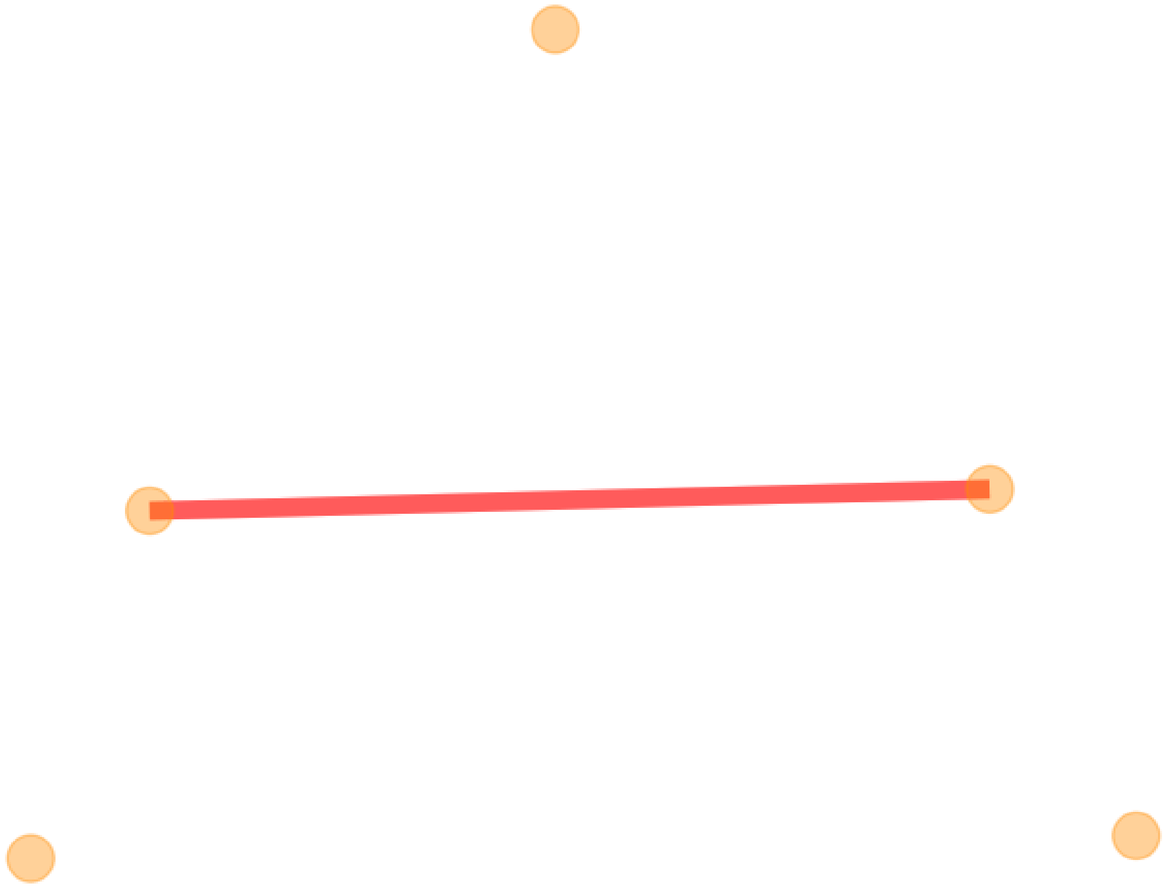
Example: Attributes and Topology



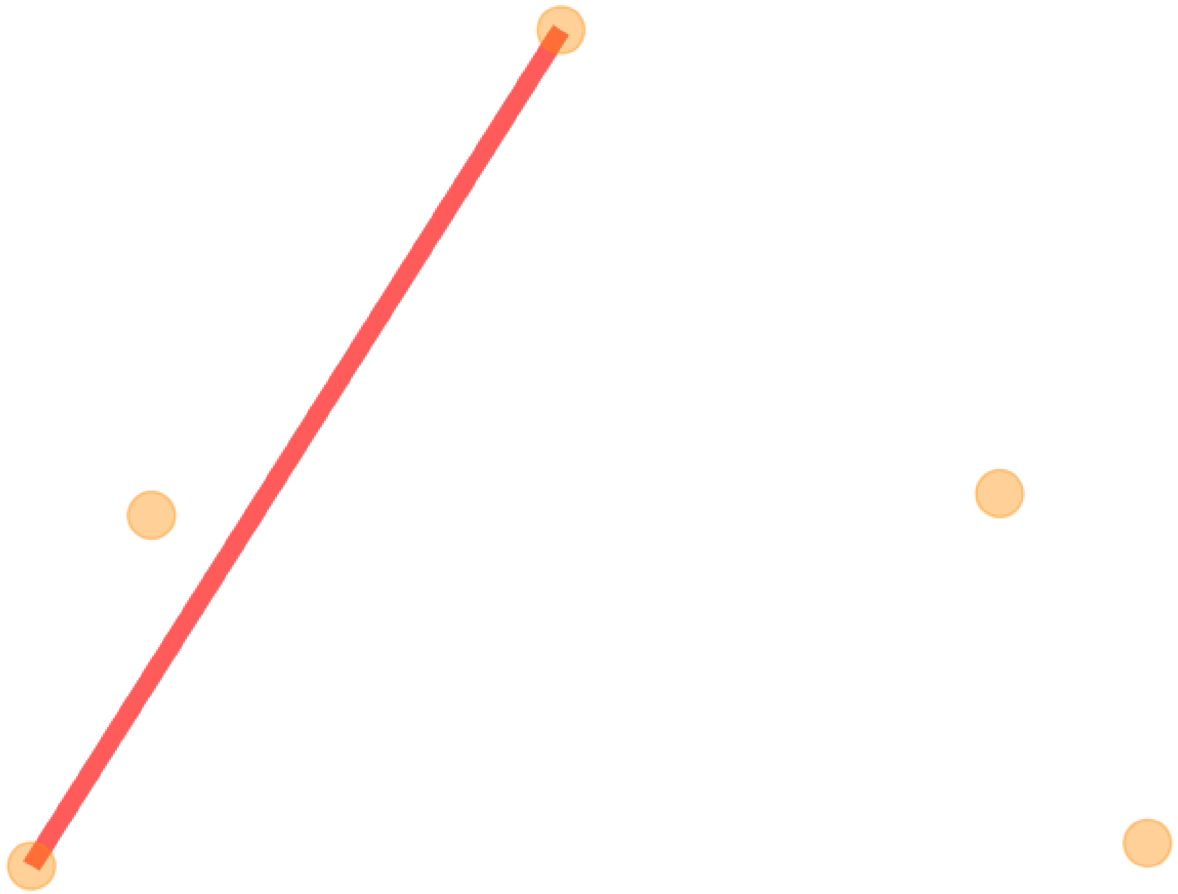
Example: Letters



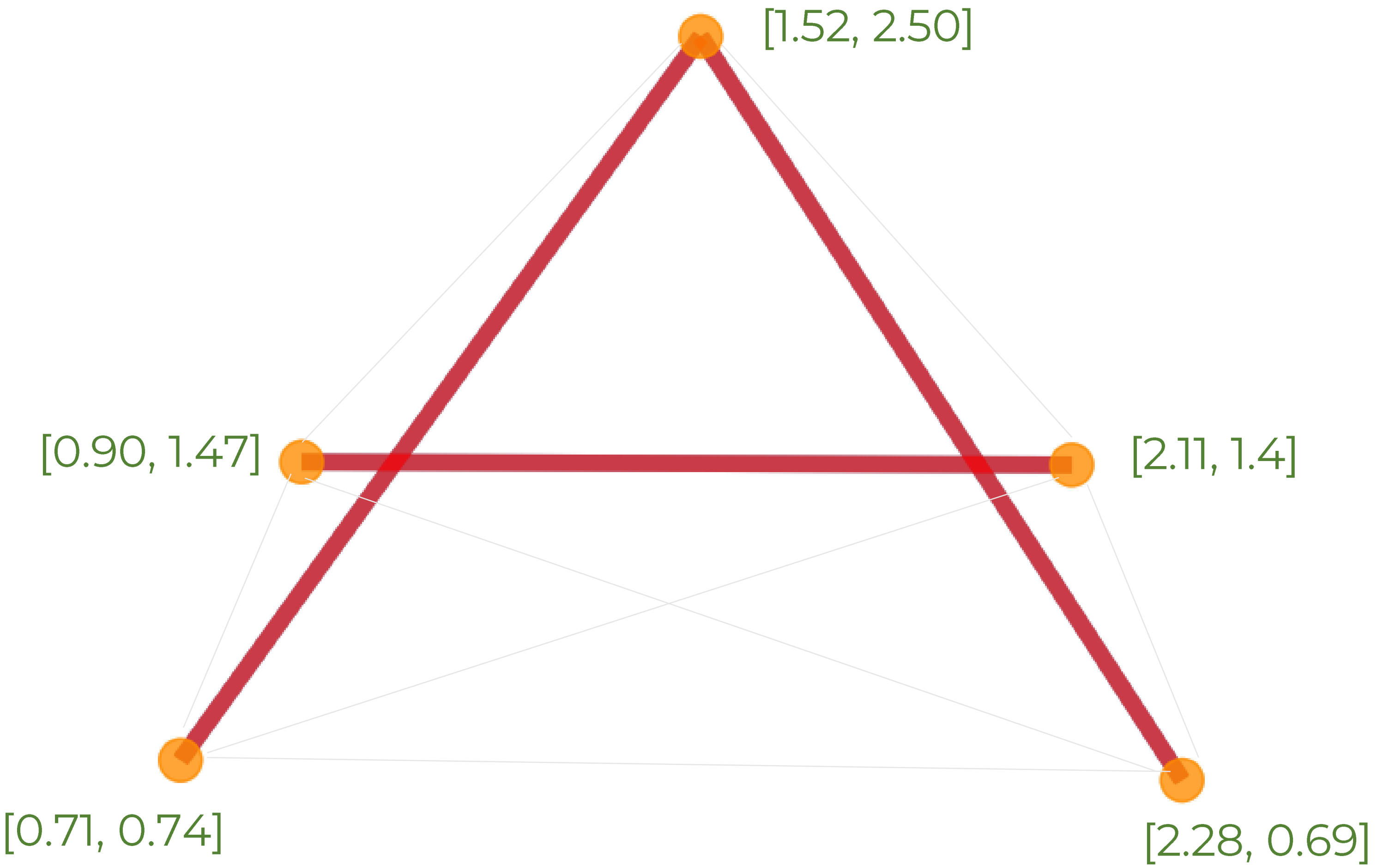
Example: Letters



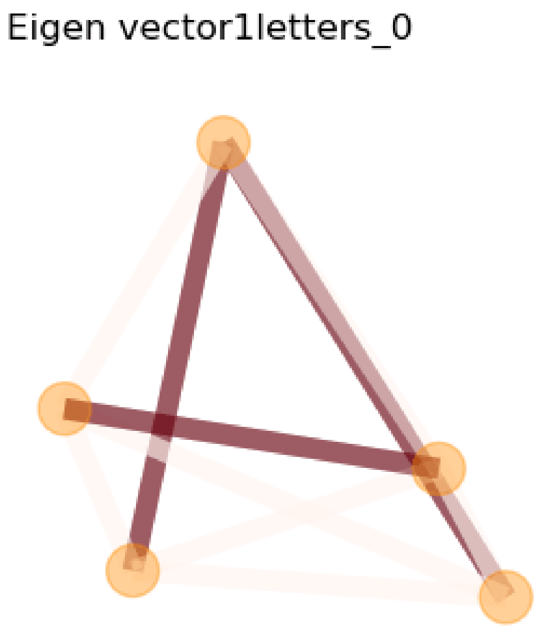
Example: Letters



A mean

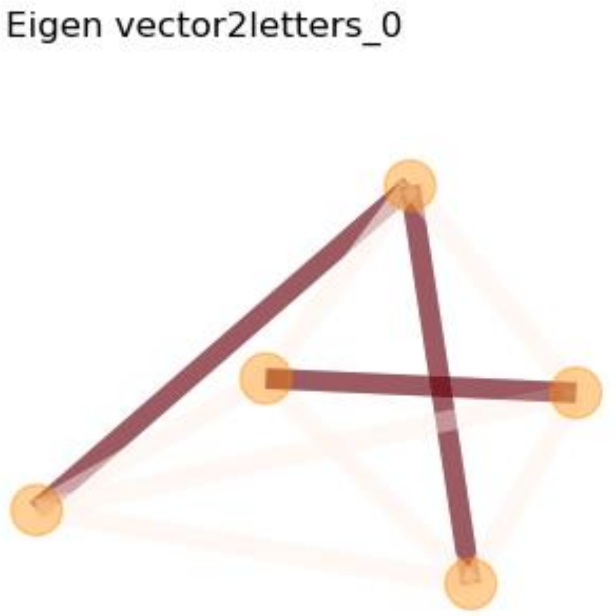


Geodesic Principal Component Analysis



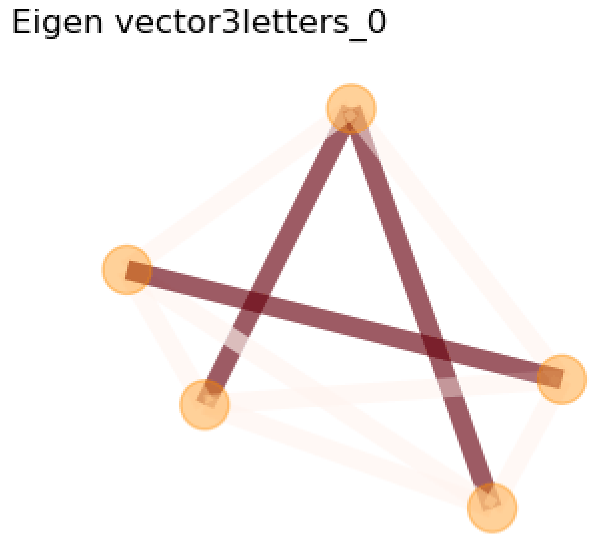
Capturing 18% of the total variance

Geodesic Principal Component Analysis



Capturing 16% of the total variance

Geodesic Principal Component Analysis



Capturing 14% of the total variance

Conclusion and Further Developments

Starting from Strcuture Spaces defined by Jain and Obermayer (2009), we introduced:

⟹ the GPCA for the Graph Space

⟹ AAC Algorithm for computing statistics such as the Fréchet Mean

⟹ Python Package

Next Step:

- Pure Topological Geodesic Principal Component
- Network on Network Regression Model

Analysis of Populations of Networks:

Structure Spaces and the Computation of Summary Statistics.

Anna Calissano
Mox Modelling Scientific Computing
Dept. Mathematics
Politecnico di Milano
anna.calissano@polimi.it

POLITECNICO
MILANO 1863
DIPARTIMENTO DI MATEMATICA

Some References

Aydın, Burcu, et al. "A principal component analysis for trees." *The Annals of Applied Statistics* 3.4 (2009): 1597-1615.

Dumont, J. “Object Oriented Data Analysis in the study of two population of texts” sup. Calissano, Vantini *Thesis KTH-Polimi* (2017)

Durante, Daniele, and David B. Dunson. "Bayesian inference and testing of group differences in brain networks." *Bayesian Analysis* 13.1 (2018): 29-58.

Durante, Daniele, David B. Dunson, and Joshua T. Vogelstein. "Nonparametric Bayes modeling of populations of networks." *Journal of the American Statistical Association* 112.520 (2017): 1516-1530.

Duvenaud, David K., et al. "Convolutional networks on graphs for learning molecular fingerprints." *Advances in neural information processing systems*. 2015.

Feragen, Aasa, et al. "Tree-space statistics and approximations for large-scale analysis of anatomical trees." *International Conference on Information Processing in Medical Imaging*. Springer, Berlin, Heidelberg, 2013.

Fletcher, Thomas. "Geodesic regression on Riemannian manifolds." *Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods for Modelling Biological Shape Variability*. 2011.

Ginestet, Cedric E., et al. "Hypothesis testing for network data in functional neuroimaging." *The Annals of Applied Statistics* 11.2 (2017): 725-750.

Some References

Huckemann, Stephan, Thomas Hotz, and Axel Munk. "Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions." *Statistica Sinica* (2010): 1-58.

Jain, Brijnesh J., and Klaus Obermayer. "Structure spaces." *Journal of Machine Learning Research* 10.Nov (2009): 2667-2714.

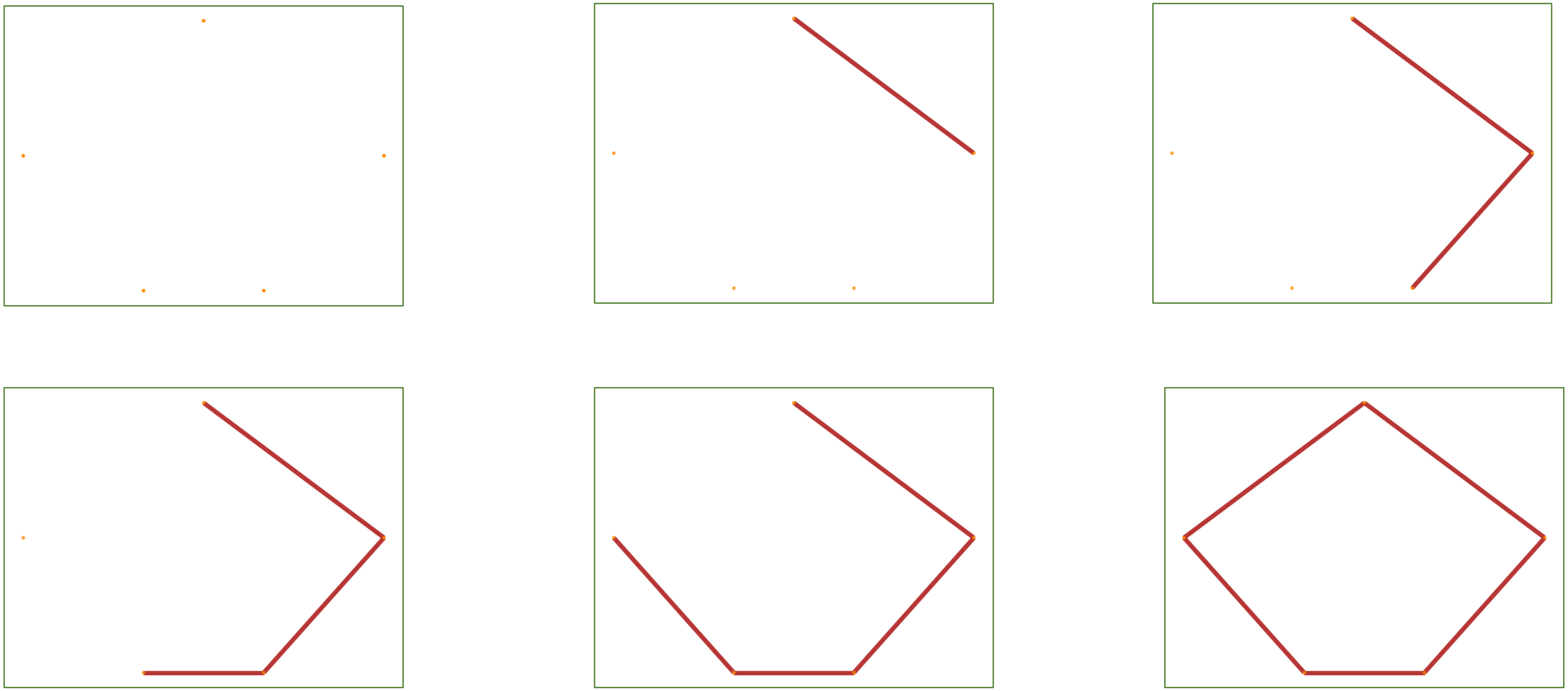
Marron, J. Steve, and Andrés M. Alonso. "Overview of object oriented data analysis." *Biometrical Journal* 56.5 (2014): 732-753.

Nye, Tom MW, et al. "Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees." *Biometrika* 104.4 (2017): 901-922

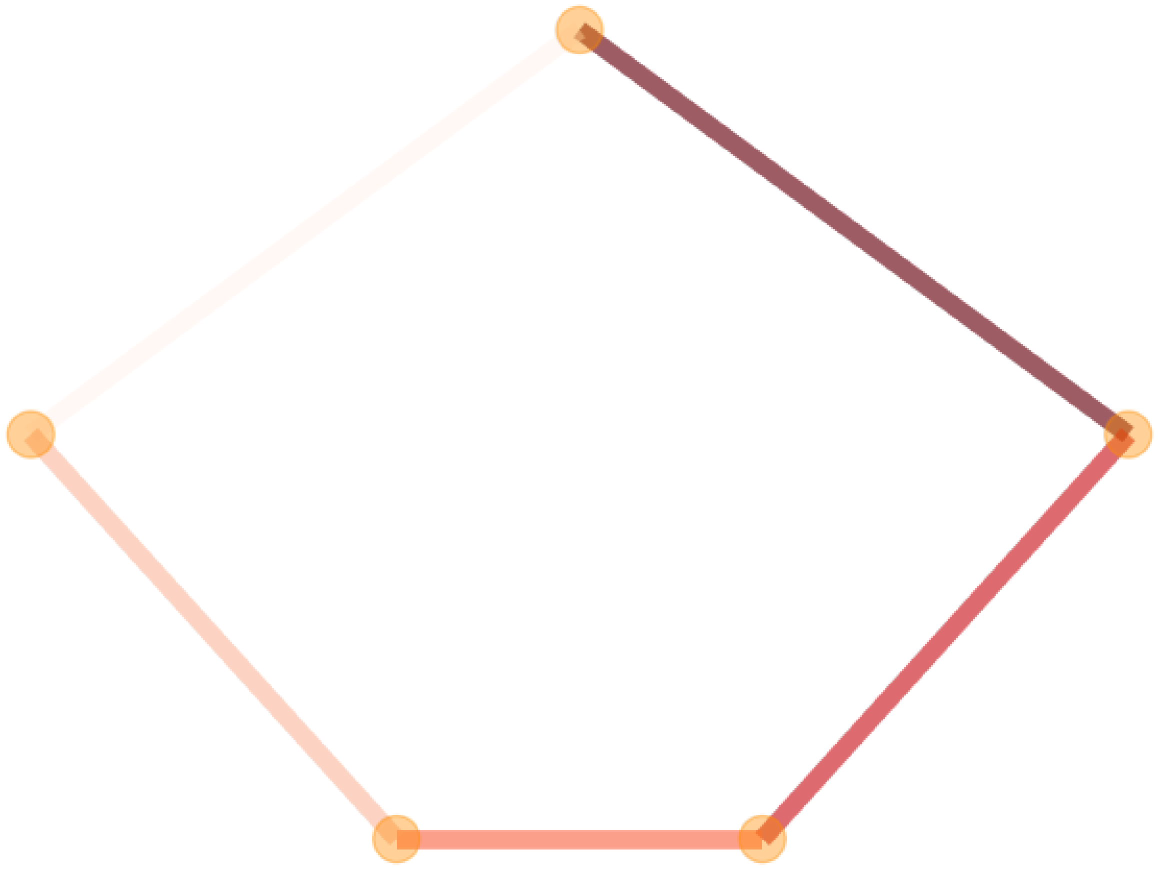
Ramsay, James O., and Bernard W. Silverman. *Applied functional data analysis: methods and case studies*. *Springer,* 2007.

Simpson, Sean L., F. DuBois Bowman, and Paul J. Laurienti. "Analyzing complex functional brain networks: fusing statistics and network science to understand the brain." *Statistics surveys 7* (2013): 1.

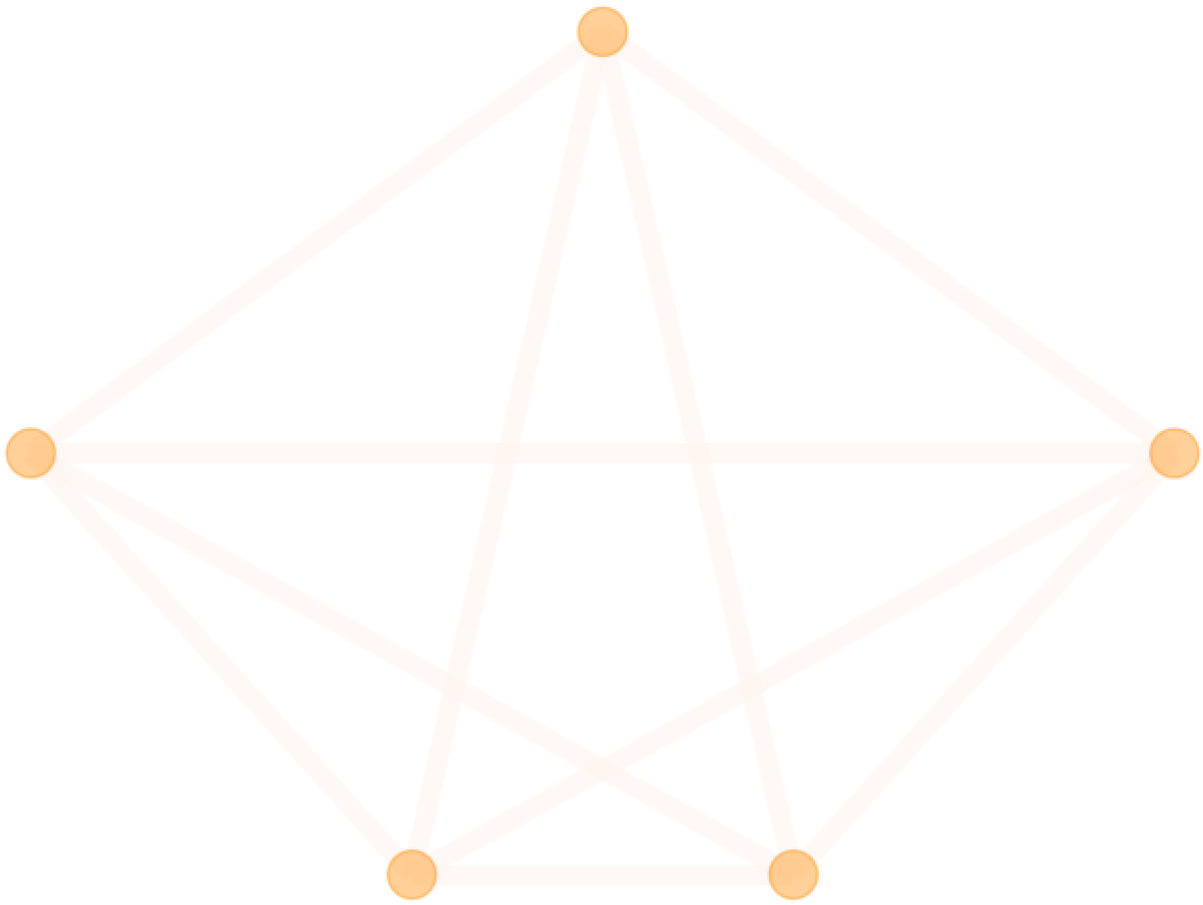
Example 1: Topological Variation



A Mean



Geodesic Principal Component Analysis



Capturing 60% of the total variance