Object Oriented Data Analysis of Population of Networks

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

POLITECNICO
MILANO 1863

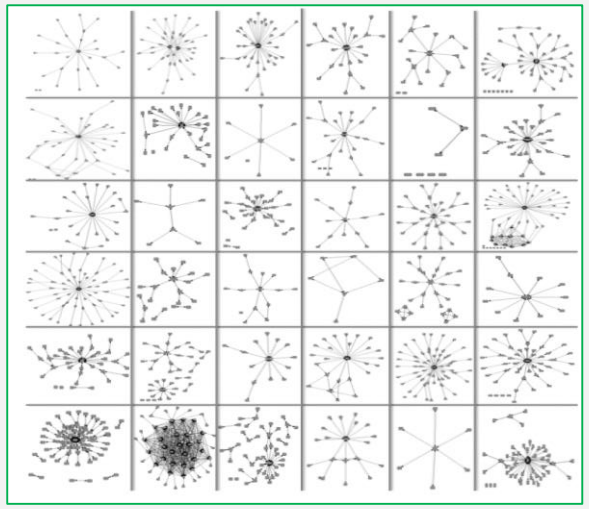
DIPARTIMENTO DI MATEMATICA

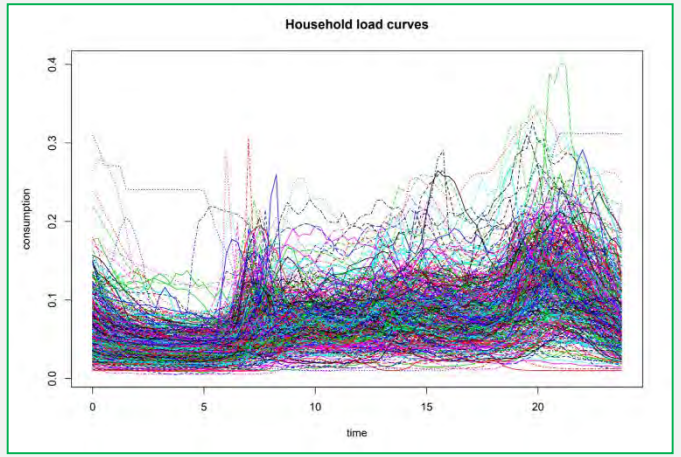
Index

- Complex Data: what is it and which mathematical tools we need
- Graph Space: a good candidate to analyse set of Networks
- Statistic on Graph Space
- Example 1: Simulated Dataset
- Example 2: Airlines Dataset

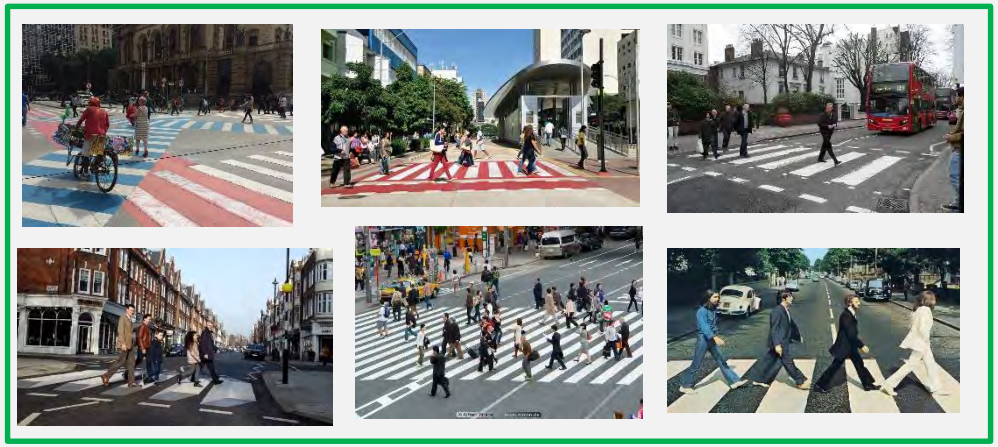
Complex Data

Complex Data

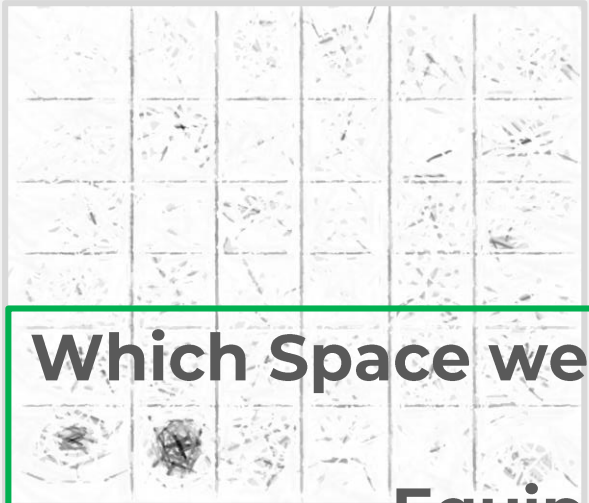


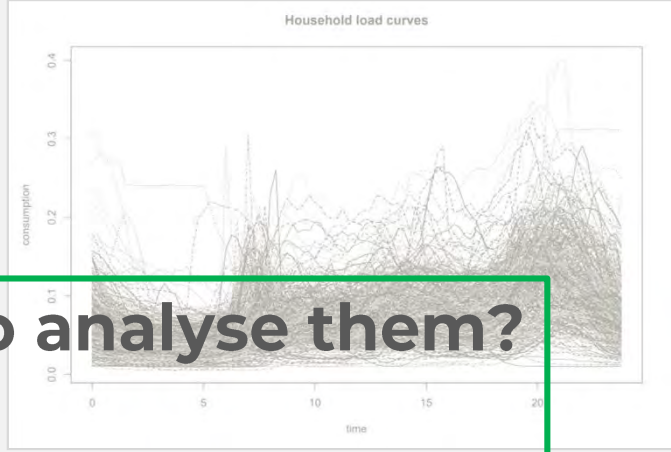


text
RT @simopieranni: I cinesi non sembrano molto #Expottimisti Problem-plagued Milan Expo battles a long list of headaches http://t.co/HlnhtRXâ€¦
RT @eu_echo: Follow us and @EUExpo2015 next week as we focus on EU #foodassistance in lead-up to #Expo2015: http://t.co/IxrxPtBst9 http://tâ€¦
RT @4expo: Five questions for IFAD. An original technique for growing rice, that is spreading #Expo2015 http://t.co/5WRxlAOQDZ http://t.co/â€¦
RT @Expo2015Milano: Shortly press conference to present @OVSofficial and @excelsiormilano as #Expo2015's Official Retailers
RT @MatchforExpo: A Champion on and off the pitch: Hernan #Crespo takes part to the #MatchForExpo as well! #Expo2015 http://t.co/5XM7gT4LIf
RT @PUrquiola: Readings before #Expo2015 #we4expo @Expo2015Milano @WEWomenForExpo http://t.co/HczeEZxNUa
RT @EUExpo2015: Why is the #EU at #Expo2015? Discover our pavilion http://t.co/CgvZjSyJi6 & plan a visit to our pavilion: http://t.co/34y9Jâ€¦
RT @israelexpo2015: Todat is #WorldHealthDay ! @Expo2015Milano @UN #Safefood #Expo2015 #Israel #expottimisti @WEWomenForExpo #foodsafety htâ€¦
#mostraMI #folli50.0 #expo2015 #fuorisalone2015 #eventi #bracco #instart #instapicture @ Bracco Imaging https://t.co/1CjQYGiUVF
Discover the beautiful Milan... #Expo2015 #Milano #promotion http://t.co/9HRCFQtMpG
RT @AmbTelAviv: Chef @craccocarlo on Yedioth Ahronot after #ItalyFoodDesign on #Israel #kosher cuisine & #Expo2015 http://t.co/H78PLaCFO2
"Osteria" with view just launched new Ristorante Terrazza @Triennale #chef @StefanoCerveri #Expo2015 #food #tradition http://t.co/LQ6Kvv7bRG
RT @EUExpo2015: With whom do we share our food every day? Discover #WeEatTogether: http://t.co/itrVistnq4 @WEWomenForExpo #Expo2015 http://â€¦
RT @FraOrlandi_: Expo quando arrivi?! Can't wait! #new #experience #waiting #Expo2015
RT @RaiExpo: Jamie Oliver @jamieoliver: teach your child about food #CountdownToExpo2015/17 #Expo2015
Wal-Mart support of gay rights turns on business: CHICAGO (Reuters) - The chief executive... http://t.co/HB8PBm9Sj4 #expo #expo2015 #biz
EFSAExpo2015 | Shaping the Future of Food Safety, Together #EXPO2015 #Expottimisti #Milan #EFSA #Food http://t.co/18JbXdafKb
#ExpoMilano2015 all in BOTOX MATINEE the friendly after party every sunday morning at Q21 Milan
Follow the adventures of #Students4Expo! Our markoshughes blogs about ESNRomania. ESNITALIA #EUExpo2015 #Expo2015 http://t.co/BUYmjAxC2a
Exxon rejects Beaumont refinery union proposal; offers six-year pact: HOUSTON (Reuters) -... http://t.co/u2O2OWeCFC #expo #expo2015 #biz
Farming First | A global coalition for sustainable agricultural development #EUExpo2015 #Expo2015 http://t.co/WERbYx5Wxi
#HappyEaster #Expo2015 Raisins nutrition facts and health benefits.. #foodie #foodie Dried fruit: http://t.co/hOlqcJjXoJ



Complex Data





Which Space we should choose to analyse them?

Equipped with which tools?

text
RT @simopieranni: I cinesi non sembrano molto #Expottimisti Problem-plagued Milan Expo battles a long list of headaches http://t.co/HlnhtRXâ€¦
RT @eu_echo: Follow us and @EUExpo2015 next week as we focus on EU #foodassistance in lead-up to #Expo2015: http://t.co/IxrxPtBst9 http://tâ€¦
RT @4expo: Five questions for IFAD. An original technique for growing rice, that is spreading #Expo2015 http://t.co/5WRxIAOQDZ http://t.co/aâ€¦
RT @Expo2015Milano: Shortly press conference to present @OVSofficial and @excelsiormilano as #Expo2015's Official Retailers
RT @MatchforExpo: A Champion on and off the pitch: Hernan #Crespo takes part to the #MatchForExpo as well! #Expo2015 http://t.co/5XM7gT4LIf
RT @PUrquiola: Readings before #Expo2015 #we4expo @Expo2015Milano @WEWomenForExpo http://t.co/HczeEZxNUa
RT @EUExpo2015: Why is the #EU at #Expo2015? Discover our pavilion http://t.co/CgvZjSyJi6 & plan a visit to our pavilion: http://t.co/34y9Jâ€¦
RT @israelexpo2015: Todat is #WorldHealthDay ! @Expo2015Milano @UN #Safefood #Expo2015 #Israel #expottimisti @WEWomenForExpo #foodsafety htâ€¦
#mostraMI #folli50.0 #expo2015 #fuorisalone2015 #eventi #bracco #instart #instapicture @ Bracco Imaging https://t.co/1CjQYGiUVF
Discover the beautiful Milan... #Expo2015 #Milano #promotion http://t.co/9HRCFQtMpG
RT @AmbTelAviv: Chef @craccocarlo on Yedioth Ahronot after #ItalyFoodDesign on #Israel #kosher cuisine & #Expo2015 http://t.co/H78PLaCFO2
"Osteria" with view just launched new Ristorante Terrazza @Triennale #chef @StefanoCerveri #Expo2015 #food #tradition http://t.co/LQ6Kvv7bRG
RT @EUExpo2015: With whom do we share our food every day? Discover #WeEatTogether: http://t.co/itrVistnq4 @WEWomenForExpo #Expo2015 http://â€¦
RT @FraOrlandi_: Expo quando arrivi?! Can't wait! #new #experience #waiting #Expo2015
RT @RaiExpo: Jamie Oliver @jamieoliver: teach your child about food #CountdownToExpo2015/17 #Expo2015
Wal-Mart support of gay rights turns on business: CHICAGO (Reuters) - The chief executive... http://t.co/HB8PBm9Sj4 #expo #expo2015 #biz
EFSAExpo2015 | Shaping the Future of Food Safety, Together #EXPO2015 #Expottimisti #Milan #EFSA #Food http://t.co/18JbXdafKb
#ExpoMilano2015 all in BOTOX MATINEE the friendly after party every sunday morning at Q21 Milan
Follow the adventures of #Students4Expo! Our markoshughes blogs about ESNRomania. ESNITALIA #EUExpo2015 #Expo2015 http://t.co/BUYmjAxC2a
Exxon rejects Beaumont refinery union proposal; offers six-year pact: HOUSTON (Reuters) -... http://t.co/u2O2OWeCFC #expo #expo2015 #biz
Farming First | A global coalition for sustainable agricultural development #EUExpo2015 #Expo2015 http://t.co/WERbYx5Wxi
#HappyEaster #Expo2015 Raisins nutrition facts and health benefits.. #foodie #foodie Dried fruit: http://t.co/hOlqcIJXoJ

Which ingredients to make statistics?

Metric Space	→	Cluster Analysis
Geodesics Space	→	Regression, Principal Components Analysis
Measure Space	→	Probability Distribution, Testing

Which ingredients to make statistics?

Euclidean Vectors (finite and infinite dimension)

- Metric Space → Cluster Analysis
- Geodesics Space → Regression, Principal Components Analysis
- Measure Space → Probability, Distribution, Testing

Which ingredients to make statistics?

Mildly non Euclidean
(functions, covariance matrices, ...)

- Metric Space → Cluster Analysis
- Geodesics Space → Regression, Principal Components Analysis
- Measure Space → Probability, Distribution, Testing

Which ingredients to make statistics?

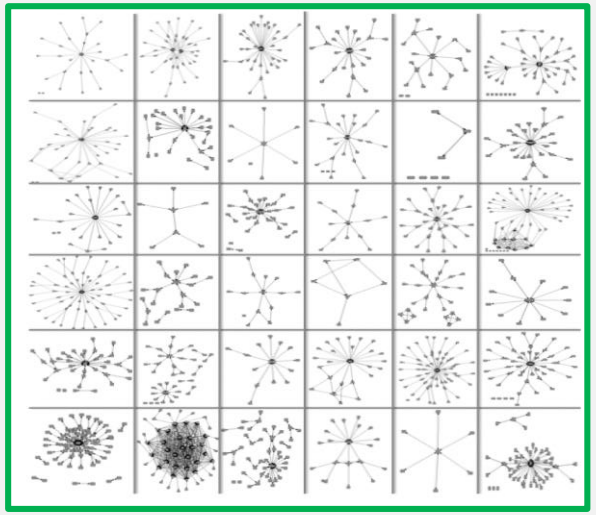
Strongly non Euclidean
(networks, trees, ...)

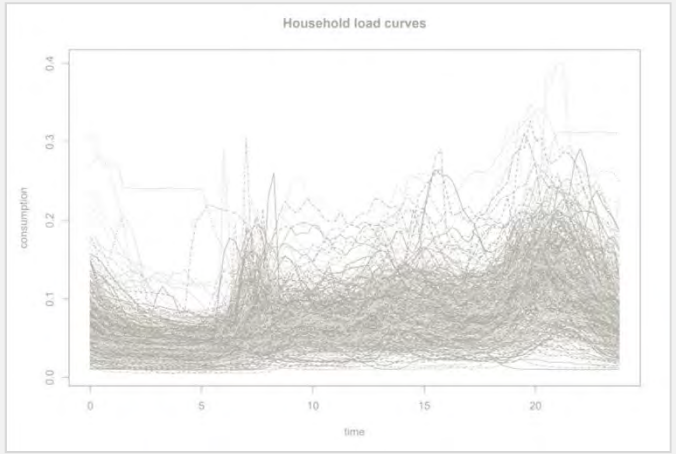
↘ Metric Space → Cluster Analysis

? Geodesics Space → Regression, Principal Components Analysis

? Measure Space → Probability, Distribution, Testing

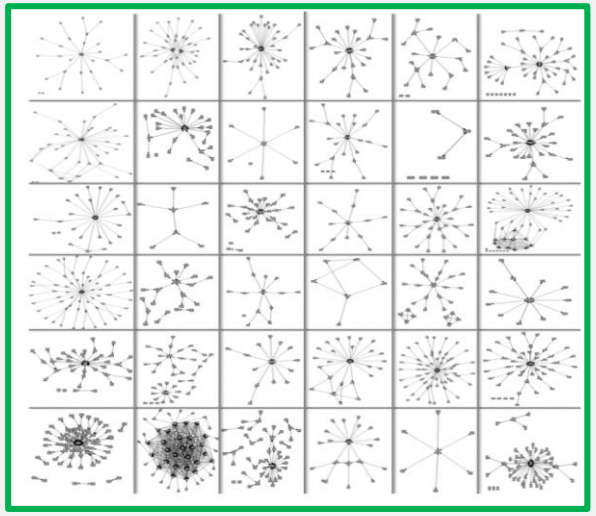
Complex Data

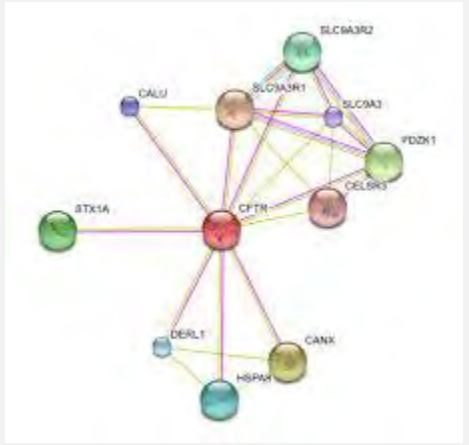


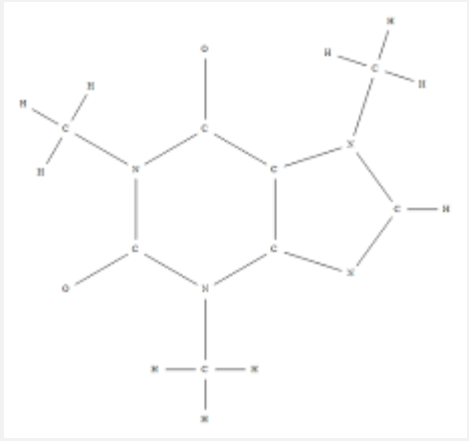


text
RT @simopieranni: I cinesi non sembrano molto #Expottimisti Problem-plagued Milan Expo battles a long list of headaches http://t.co/HlnhtRXâ€¦
RT @eu_echo: Follow us and @EUExpo2015 next week as we focus on EU #foodassistance in lead-up to #Expo2015: http://t.co/IxrxPtBst9 http://tâ€¦
RT @4expo: Five questions for IFAD. An original technique for growing rice, that is spreading #Expo2015 http://t.co/5WRxIAOQDZ http://t.co/â€¦
RT @Expo2015Milano: Shortly press conference to present @OVSofficial and @excelsiormilano as #Expo2015's Official Retailers
RT @MatchforExpo: A Champion on and off the pitch: Hernan #Crespo takes part to the #MatchForExpo as well! #Expo2015 http://t.co/5XM7gT4LIf
RT @PUrquiola: Readings before #Expo2015 #we4expo @Expo2015Milano @WEWomenForExpo http://t.co/HczeEZxNUa
RT @EUExpo2015: Why is the #EU at #Expo2015? Discover our pavilion http://t.co/CgvZjSyJi6 & plan a visit to our pavilion: http://t.co/34y9Jâ€¦
RT @israelexpo2015: Todat is #WorldHealthDay ! @Expo2015Milano @UN #Safefood #Expo2015 #Israel #expottimisti @WEWomenForExpo #foodsafety htâ€¦
#mostraMI #folli50.0 #expo2015 #fuorisalone2015 #eventi #bracco #instart #instapicture @ Bracco Imaging https://t.co/1CjQYGiUVF
Discover the beautiful Milan... #Expo2015 #Milano #promotion http://t.co/9HRCFQtMpG
RT @AmbTelAviv: Chef @craccocarlo on Yedioth Ahronot after #ItalyFoodDesign on #Israel #kosher cuisine & #Expo2015 http://t.co/H78PLaCFO2
"Osteria" with view just launched new Ristorante Terrazza @Triennale #chef @StefanoCerveri #Expo2015 #food #tradition http://t.co/LQ6Kvv7bRG
RT @EUExpo2015: With whom do we share our food every day? Discover #WeEatTogether: http://t.co/itrVistnq4 @WEWomenForExpo #Expo2015 http://â€¦
RT @FraOrlandi_: Expo quando arrivi?! Can't wait! #new #experience #waiting #Expo2015
RT @RaiExpo: Jamie Oliver @jamieoliver: teach your child about food #CountdownToExpo2015/17 #Expo2015
Wal-Mart support of gay rights turns on business: CHICAGO (Reuters) - The chief executive... http://t.co/HB8PBm9Sj4 #expo #expo2015 #biz
EFSAExpo2015 | Shaping the Future of Food Safety, Together #EXPO2015 #Expottimisti #Milan #EFSA #Food http://t.co/18JbXdafKb
#ExpoMilano2015 all in BOTOX MATINEE the friendly after party every sunday morning at Q21 Milan
Follow the adventures of #Students4Expo! Our markoshughes blogs about ESNRomania. ESNITALIA #EUExpo2015 #Expo2015 http://t.co/BUYmjAxC2a
Exxon rejects Beaumont refinery union proposal; offers six-year pact: HOUSTON (Reuters) -... http://t.co/u2O2OWeCFC #expo #expo2015 #biz
Farming First | A global coalition for sustainable agricultural development #EUExpo2015 #Expo2015 http://t.co/WERbYx5Wxi
#HappyEaster #Expo2015 Raisins nutrition facts and health benefits.. #foodie #foodie Dried fruit: http://t.co/hOlqcIJXoJ

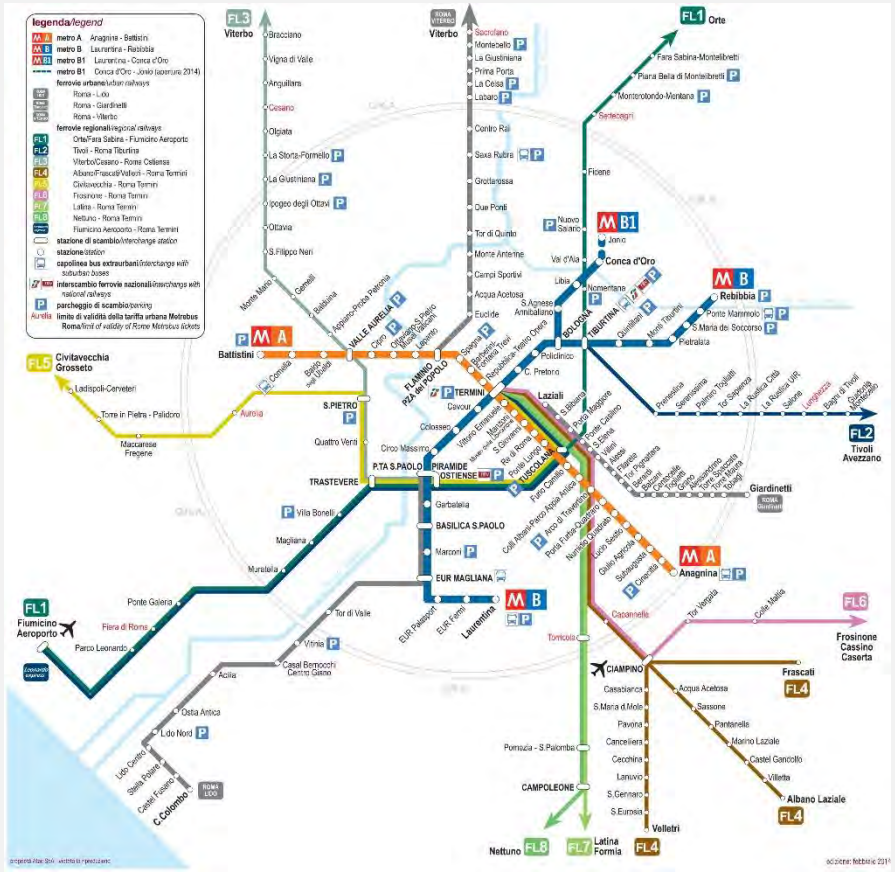
Complex Data







Network Analysis



Rome Public Transport System Network

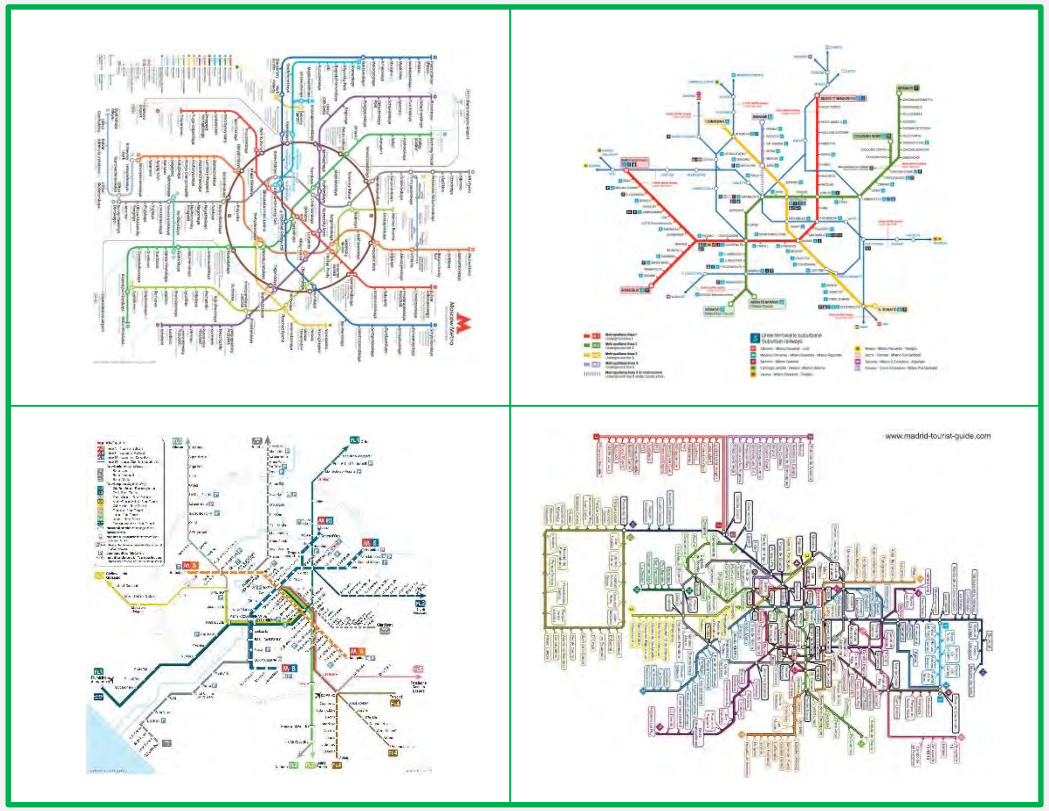
Datum: $x = (P, R, \alpha)$

P- set of nodes

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

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

Network Analysis: from one to many

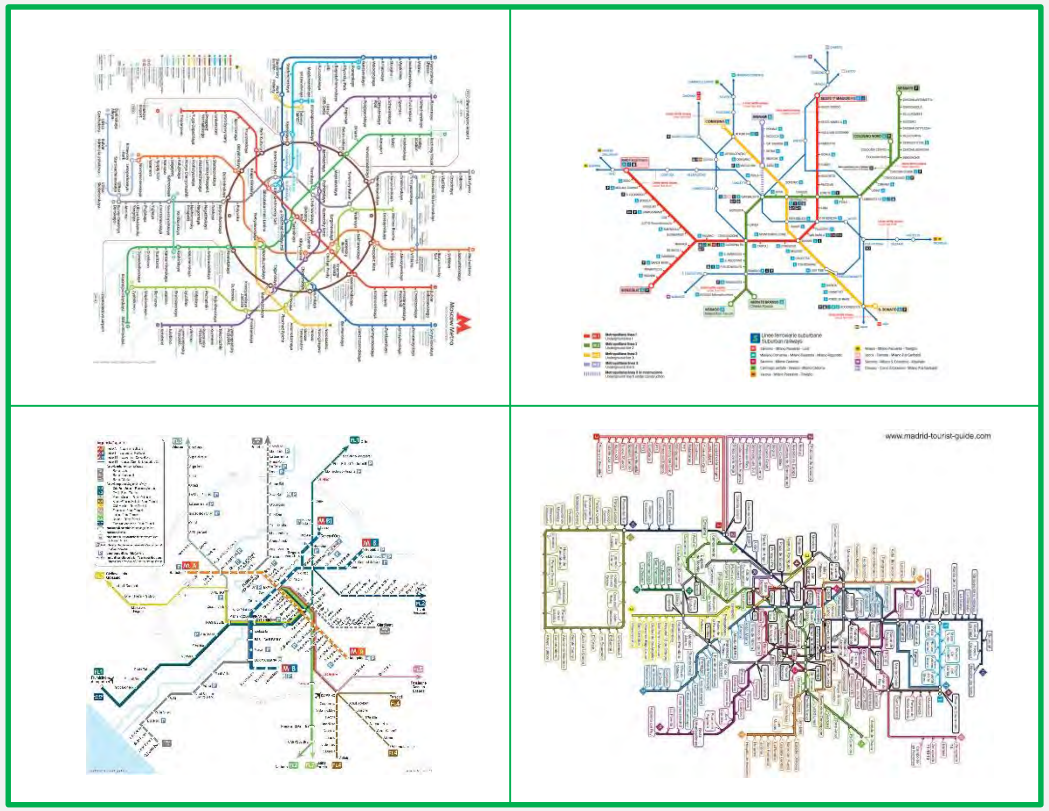


Public Transportation Network in different cities

Data Points are Networks

$$X_1 = (P, R_1), \ldots, X_n = (P, R_n)$$

Network Analysis: from one to many



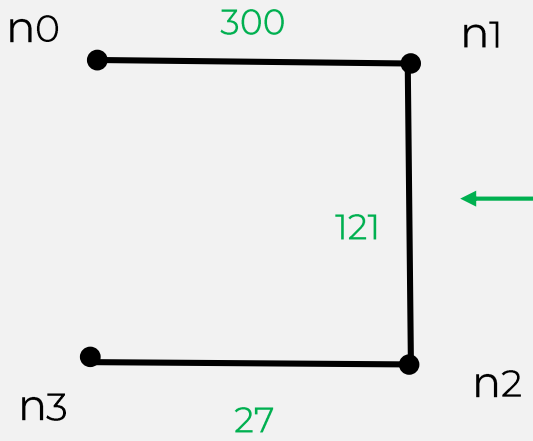
Public Transportation Network in different cities

Many new questions arise:

- How can we describe these networks?
- How can we relate nodes in networks?
- Along which relations/features are different?
- **How can we make statistics?**

Graph Space

Graph Space



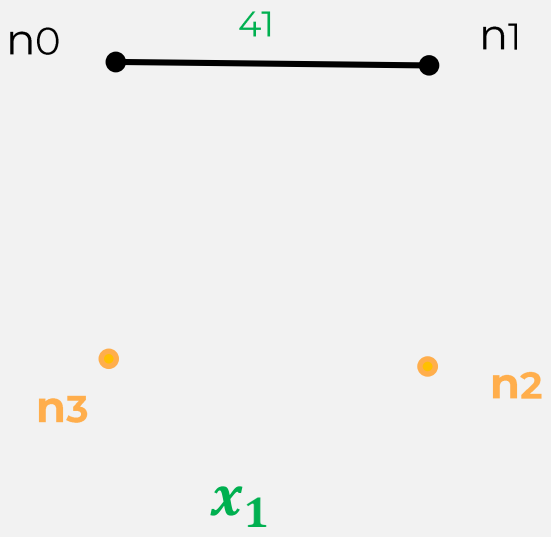
n0	300	0	0
300	n1	121	0
0	121	n2	27
0	0	27	n3

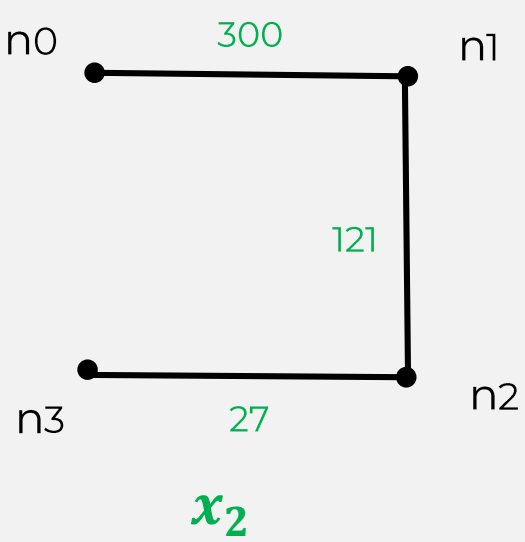
n0	300	0	0	...	n1	121	0	...	27	n3

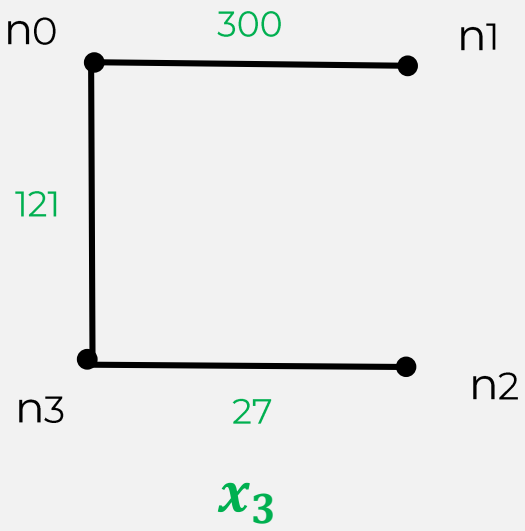
Data Point in

$$\mathbf{X} = \mathbb{R}^{16}$$

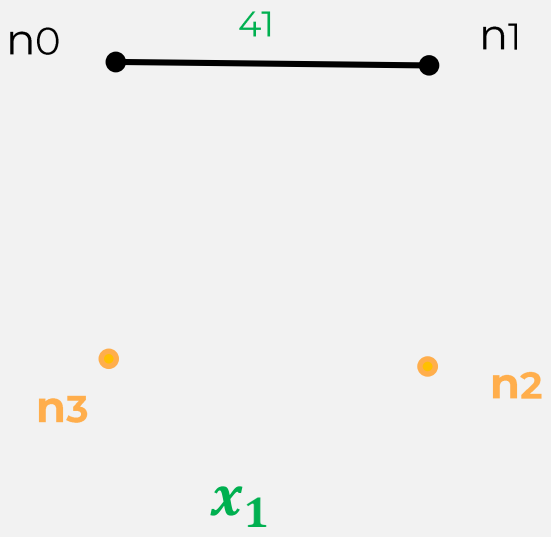
Graph Space



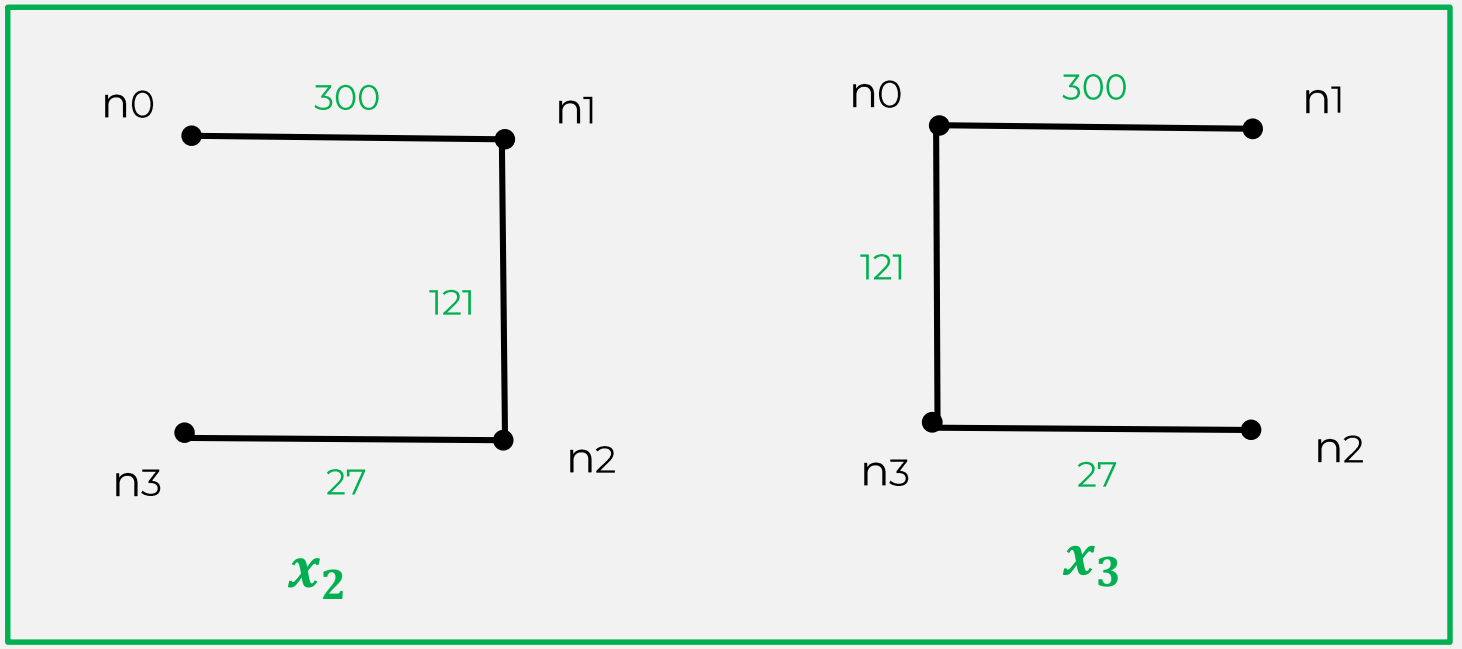




Graph Space



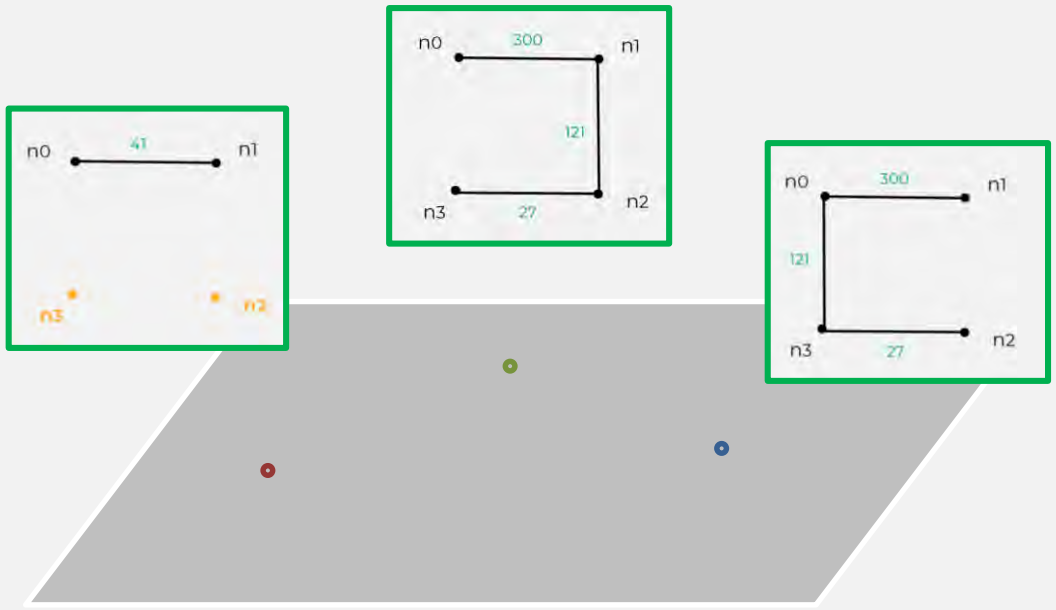
x_1



x_2 x_3

$$x_2 = x_3$$

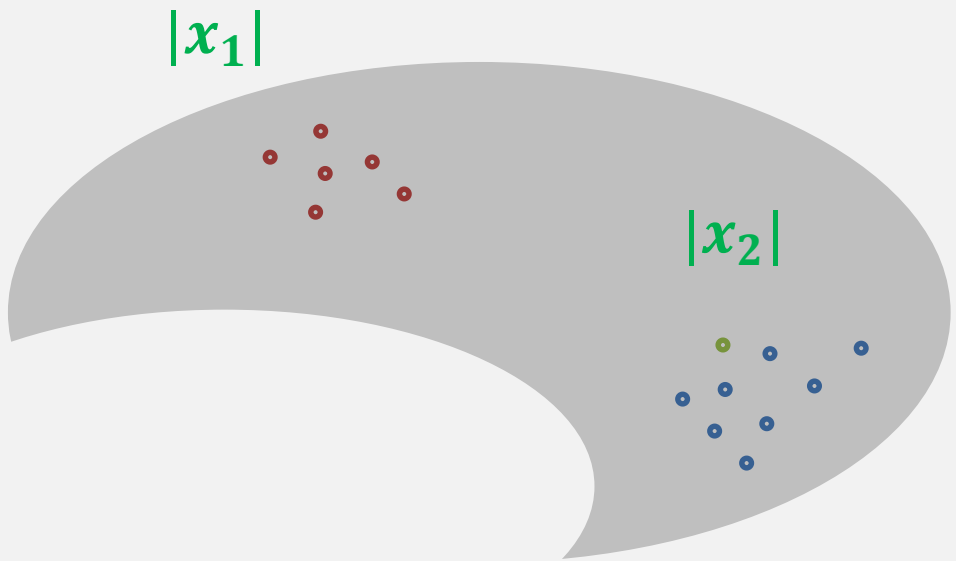
Allowing permutation of nodes



Space: X

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

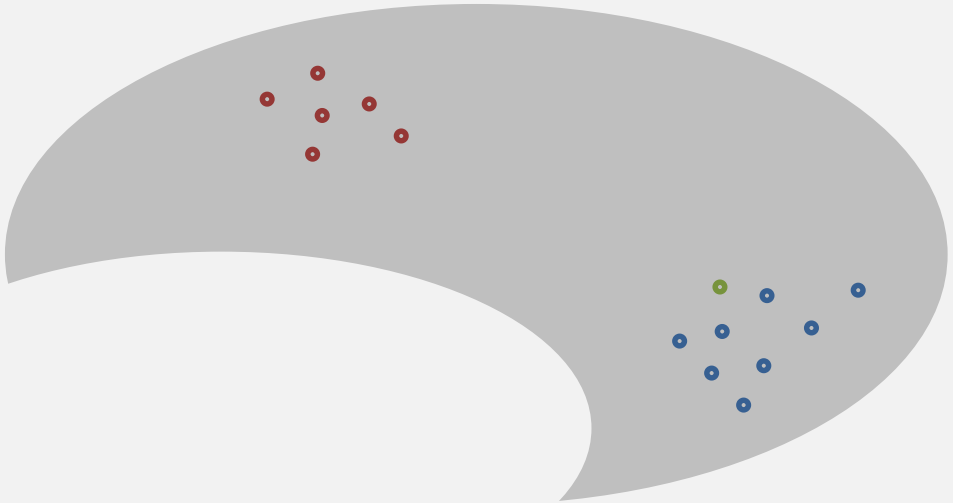
Permuting nodes



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

Quotient Space

Graph Space



Metric Space (X, d_x)
With d_X invariant with respect to permutation

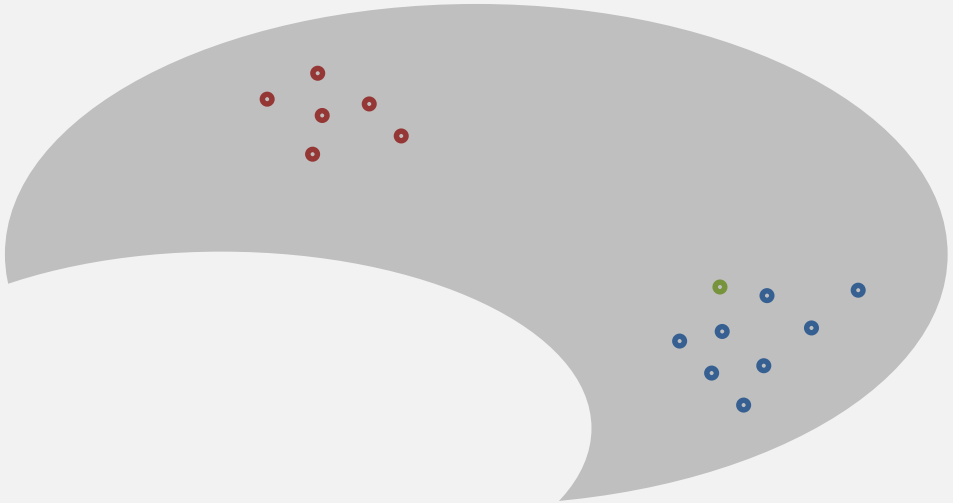
Metric Space $(X/\mathcal{T}, d)$

$$d(|x_1|, |x_2|) = \inf_{t \in \mathcal{T}} d_x(tx_1, x_2)$$

Graph Space

Manifold

$\mathcal{T}$ **action is not Free**



Not a Manifold

Exclude all the literature about
Statistics on Manifold

Graph Space: Novelty

Curvature of the space: unbounded

- → No uniqueness of the geodesics even locally.
- → No uniqueness of the Fréchet Mean

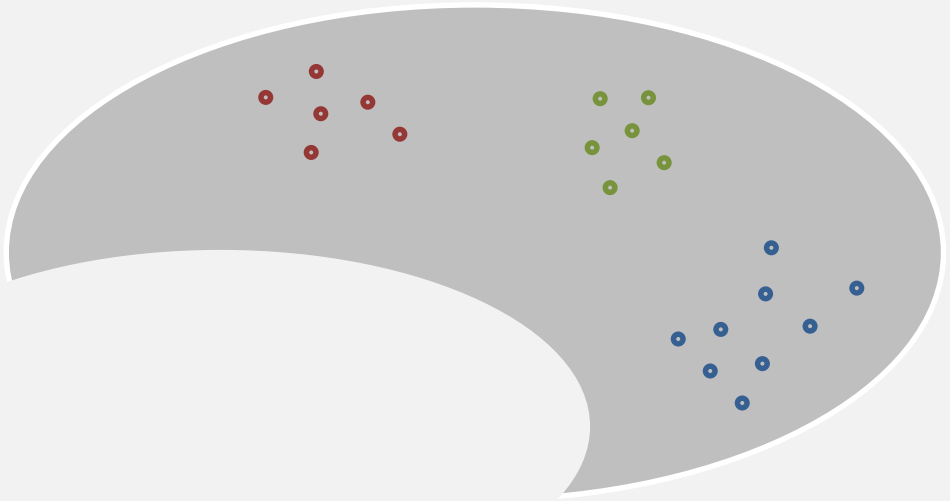
Isometric and Finite Dimension Action:

- → Allows to transfer easily computation from $\mathbb{X}$ to $\mathbb{X}/\mathbb{T}$

Define Geodesic Principal Components and introduce the Align All and Compute Algorithm

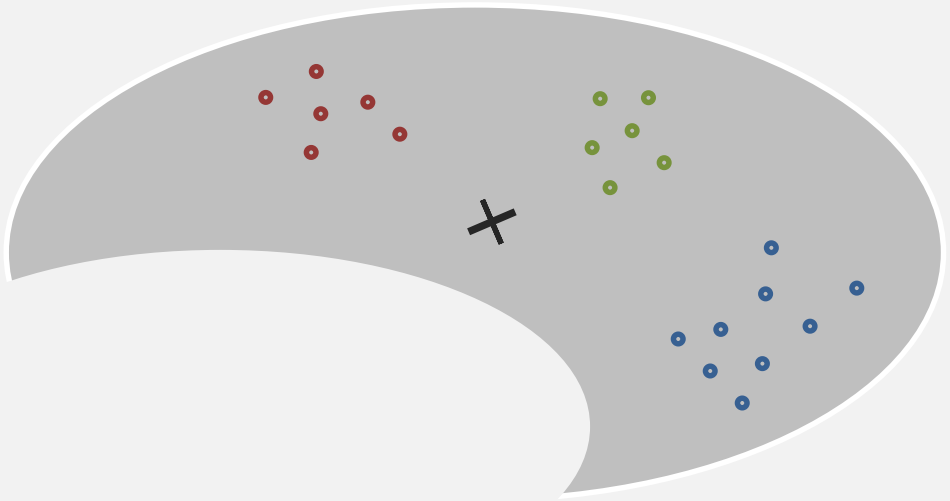
Statistic on Graph Space

Align All and Compute Algorithm



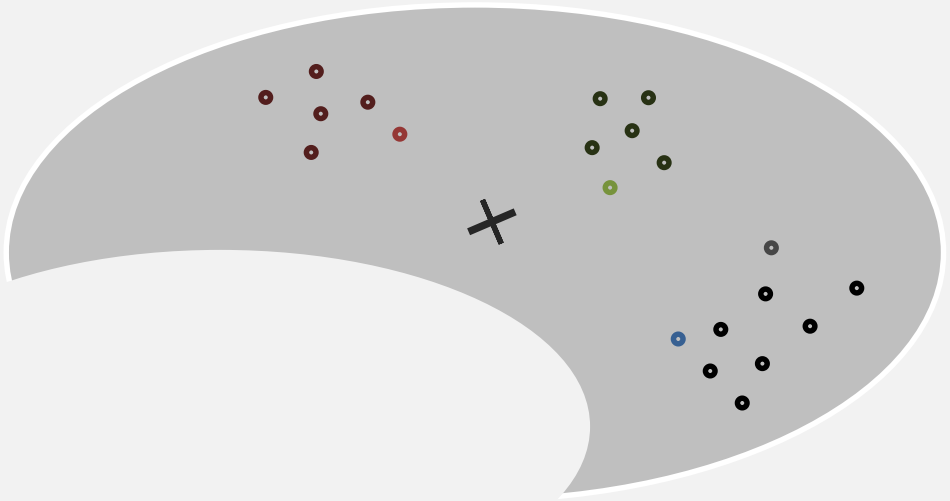
Graph Space: $\mathcal{X}/\mathcal{T}$

Align All and Compute Algorithm



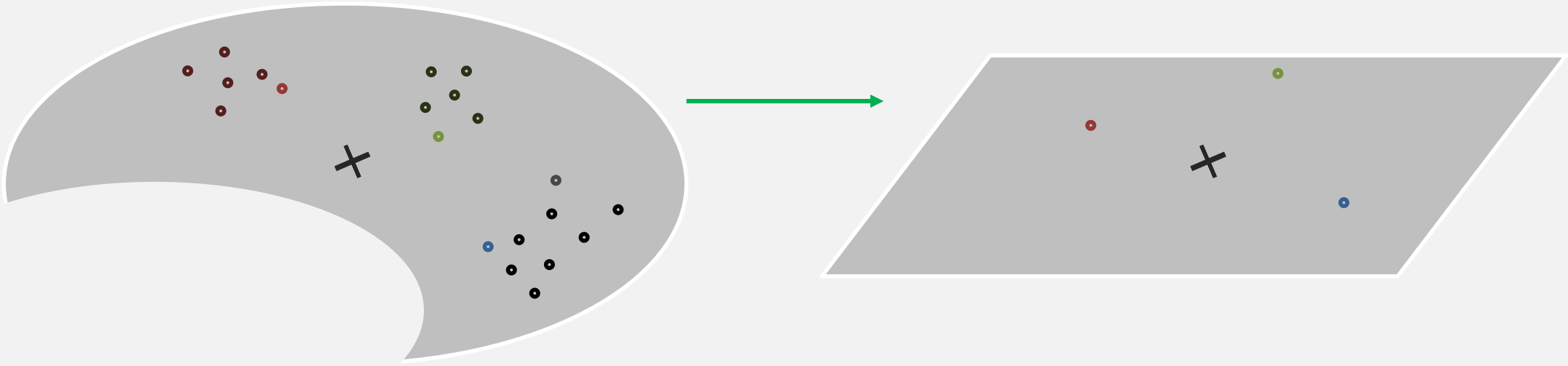
Graph Space: $X/\mathcal{T}$

Align All and Compute Algorithm



Graph Space: X/T

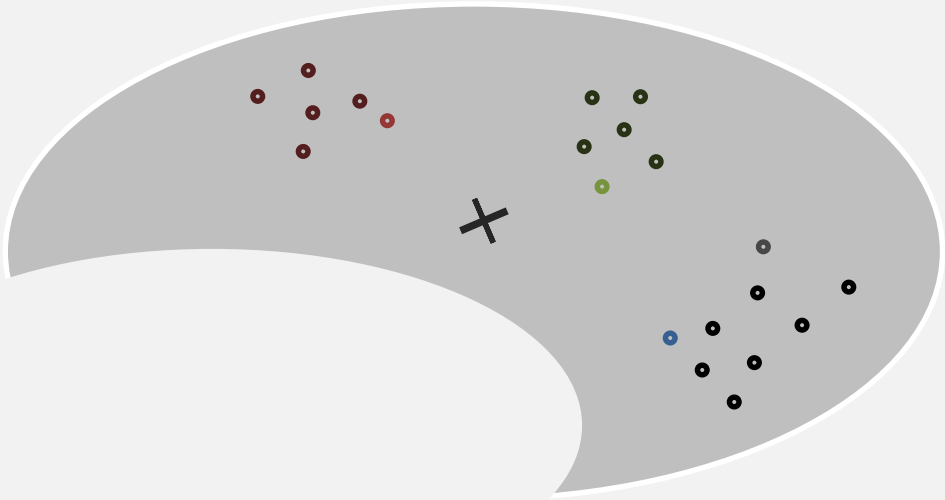
Align All and Compute Algorithm



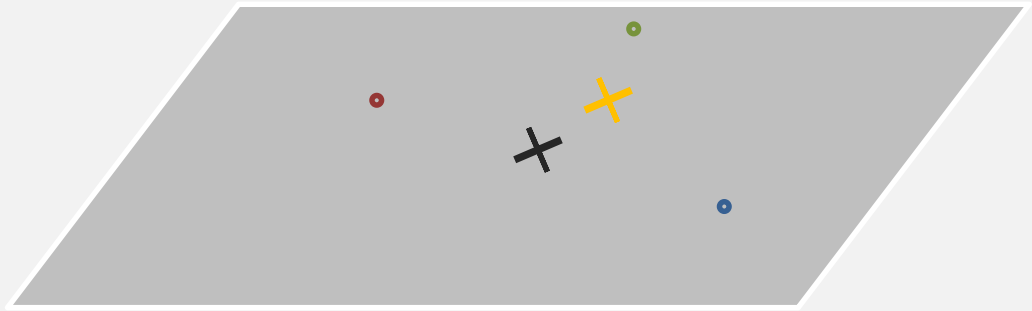
Graph Space: $X/\mathcal{T}$

Space: X

Align All and Compute Algorithm

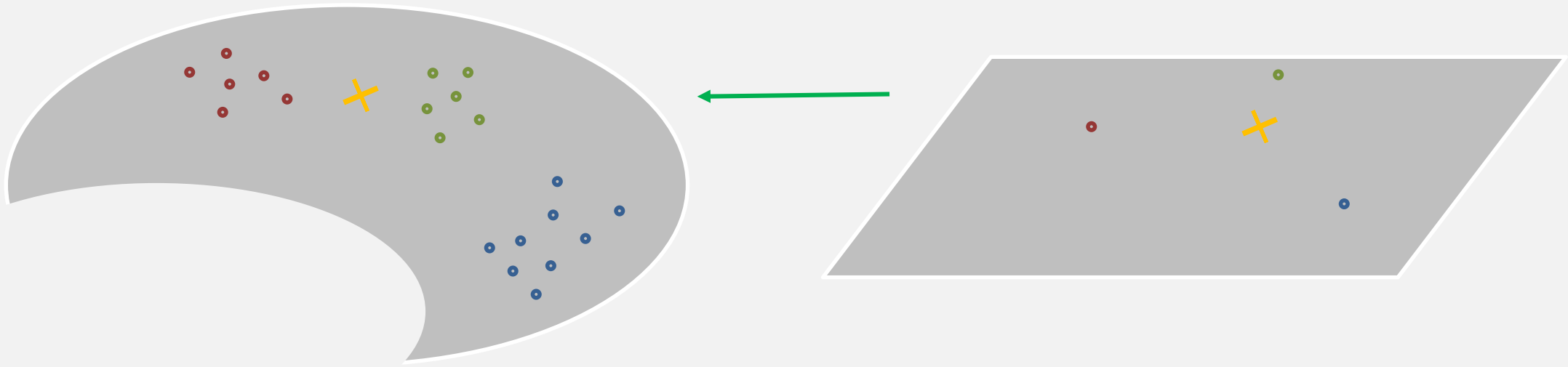


Graph Space: X/T



Space: X

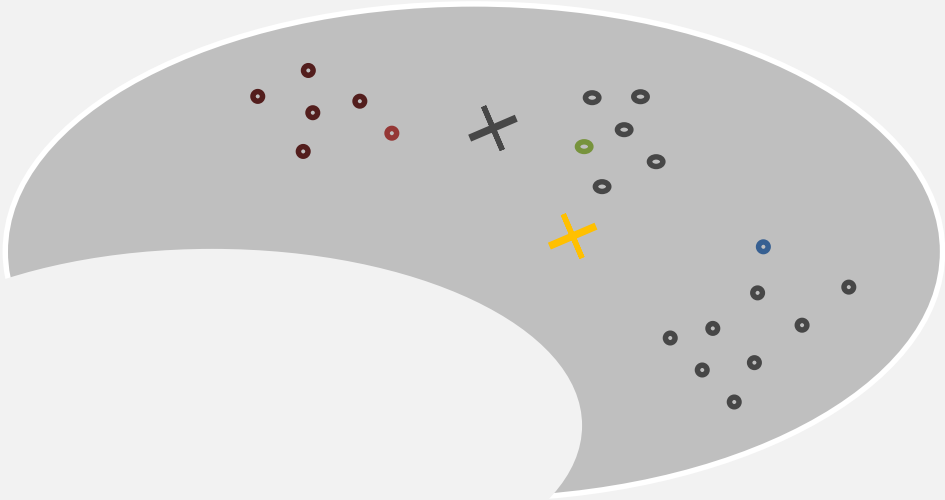
Align All and Compute Algorithm



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

Space: X

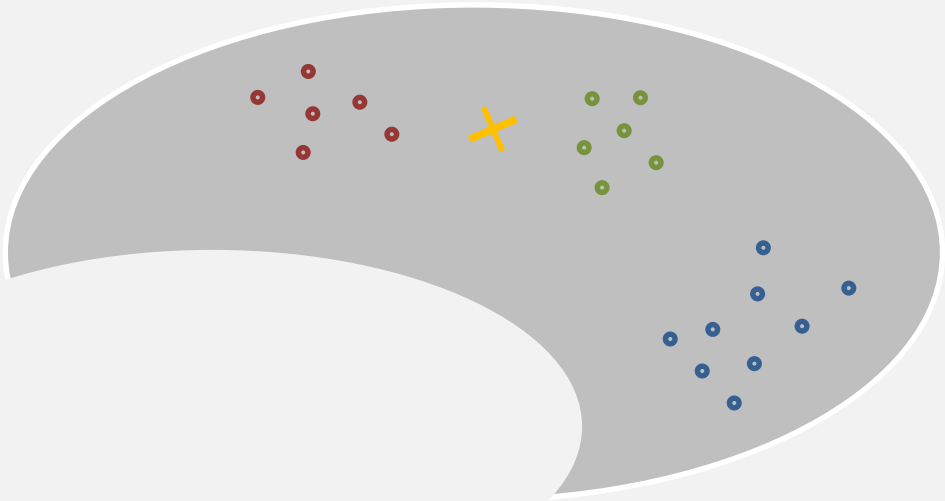
Align All and Compute Algorithm



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

Space: X

Align All and Compute Algorithm



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

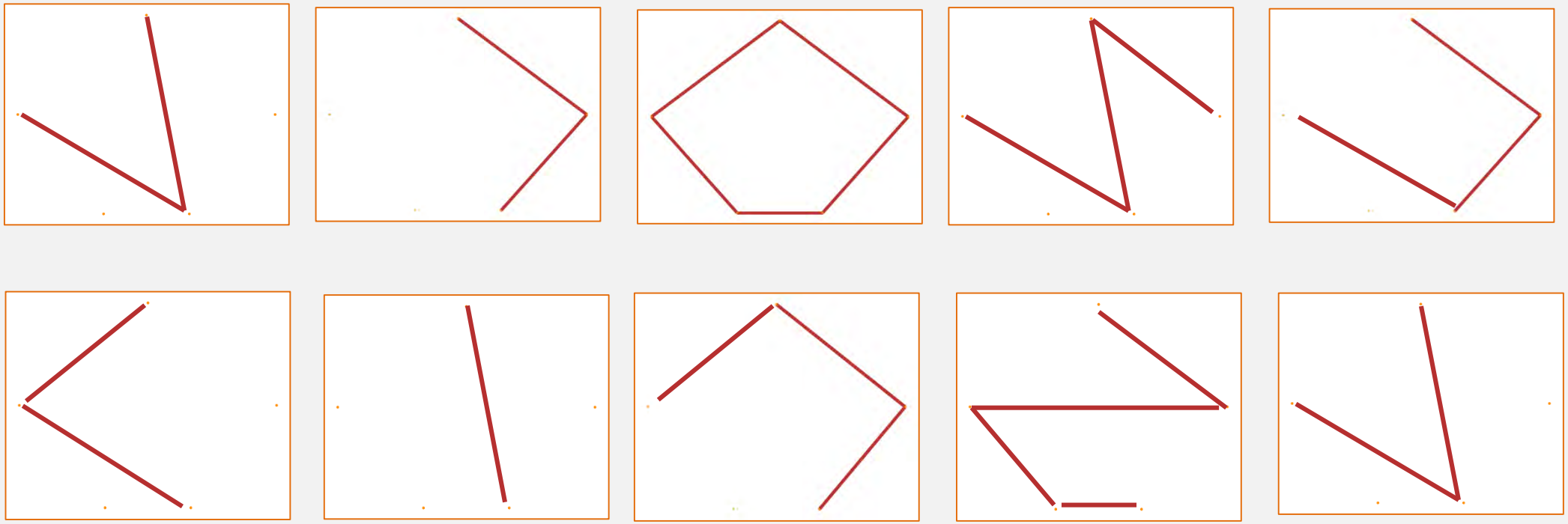
Can be a Fréchet Mean, a Geodesic, etc.

Example 1: Simulation

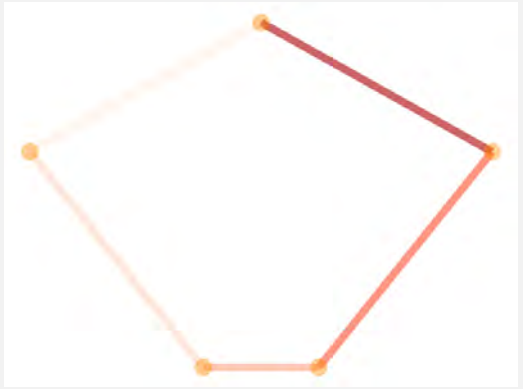
Topological Variation

Topological Variation

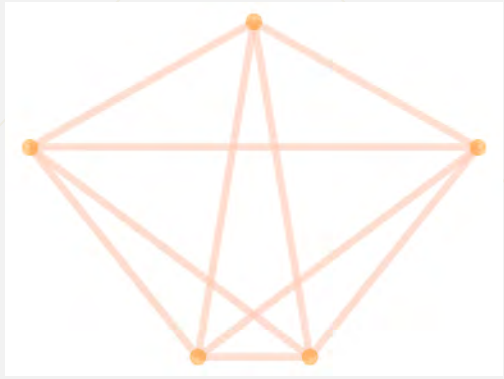
Generated 5000 observation randomly sampled from the 5 orbits. Here a subsample:



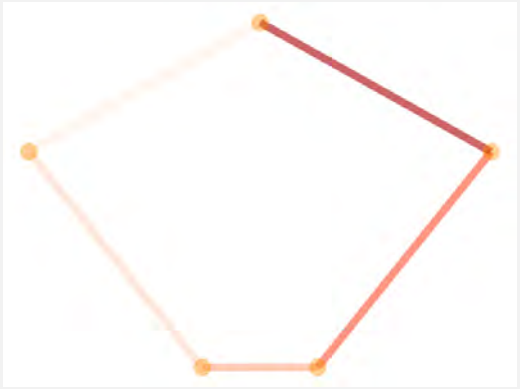
A Frèchet mean



Theoretical Fréchet Mean

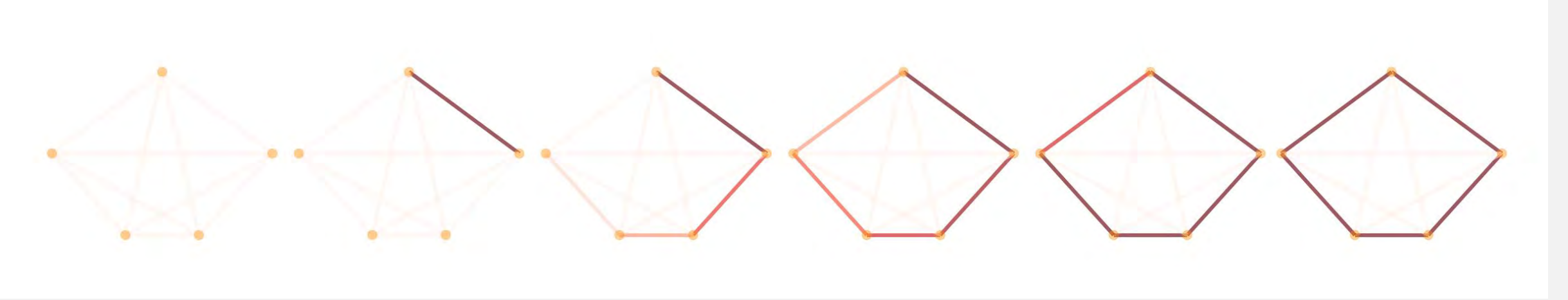


Estimation of Fréchet Mean without Alignment



Estimation of Fréchet Mean with Alignment

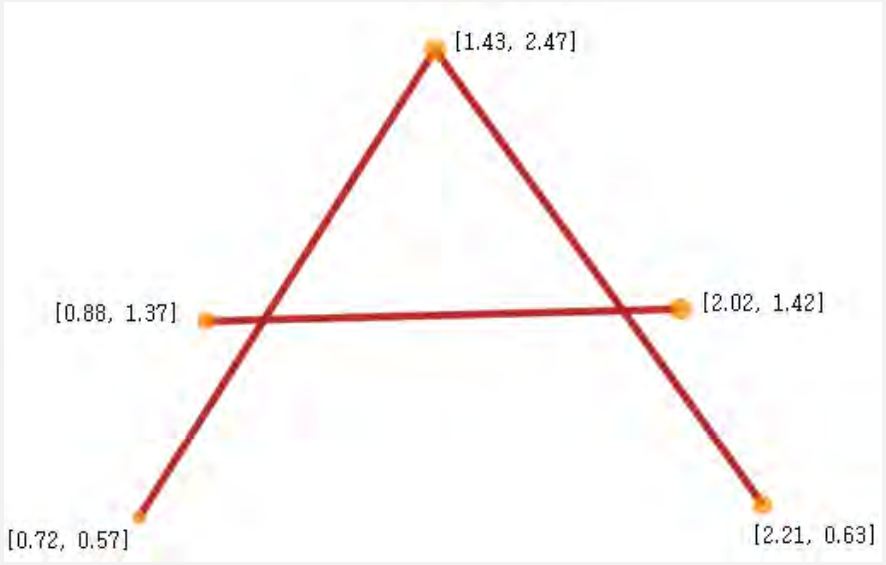
First Geodesic Principal Component



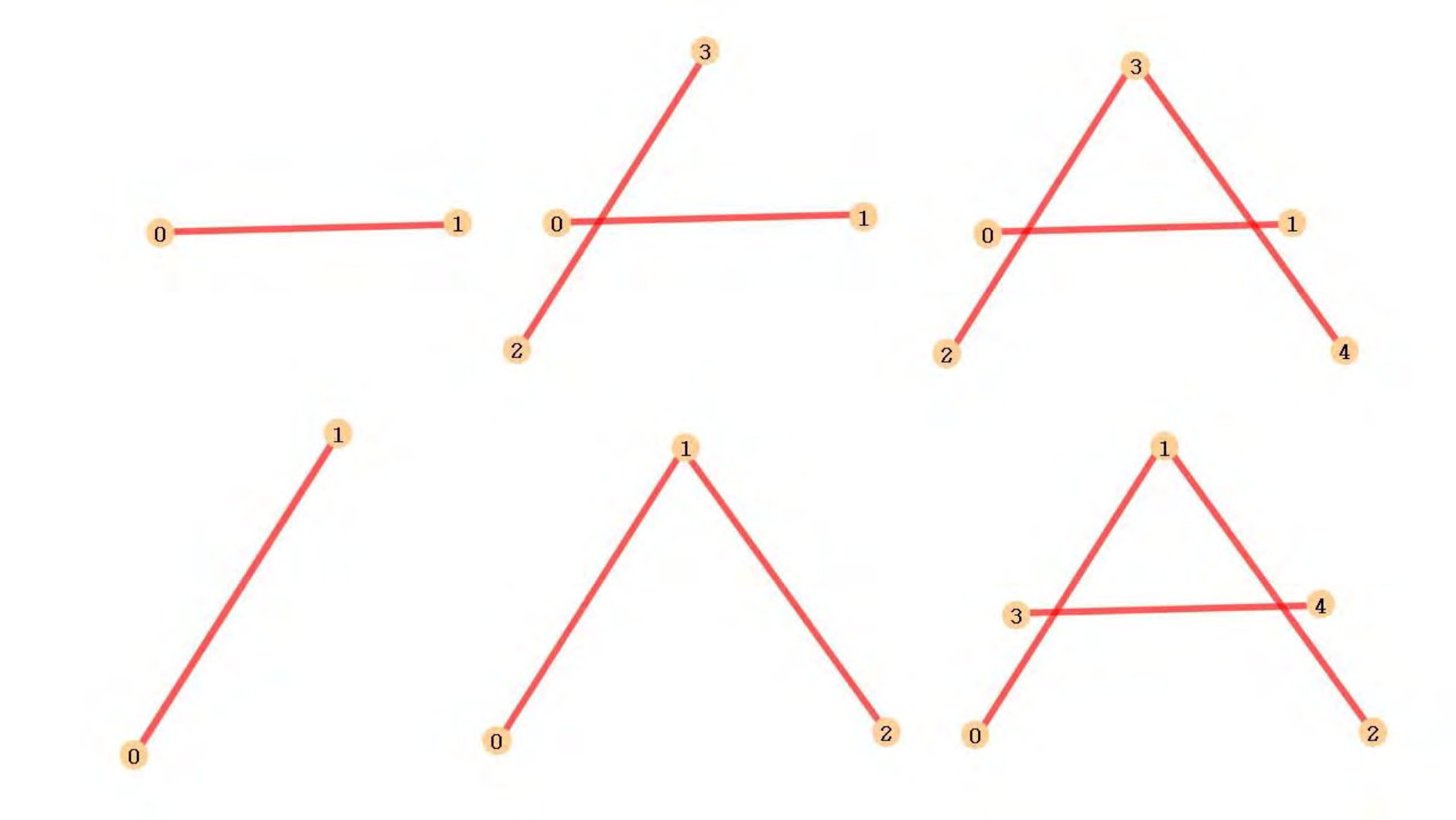
Projection along the 1st GPCA explaining
60% of the total variance

Example 2: Letters

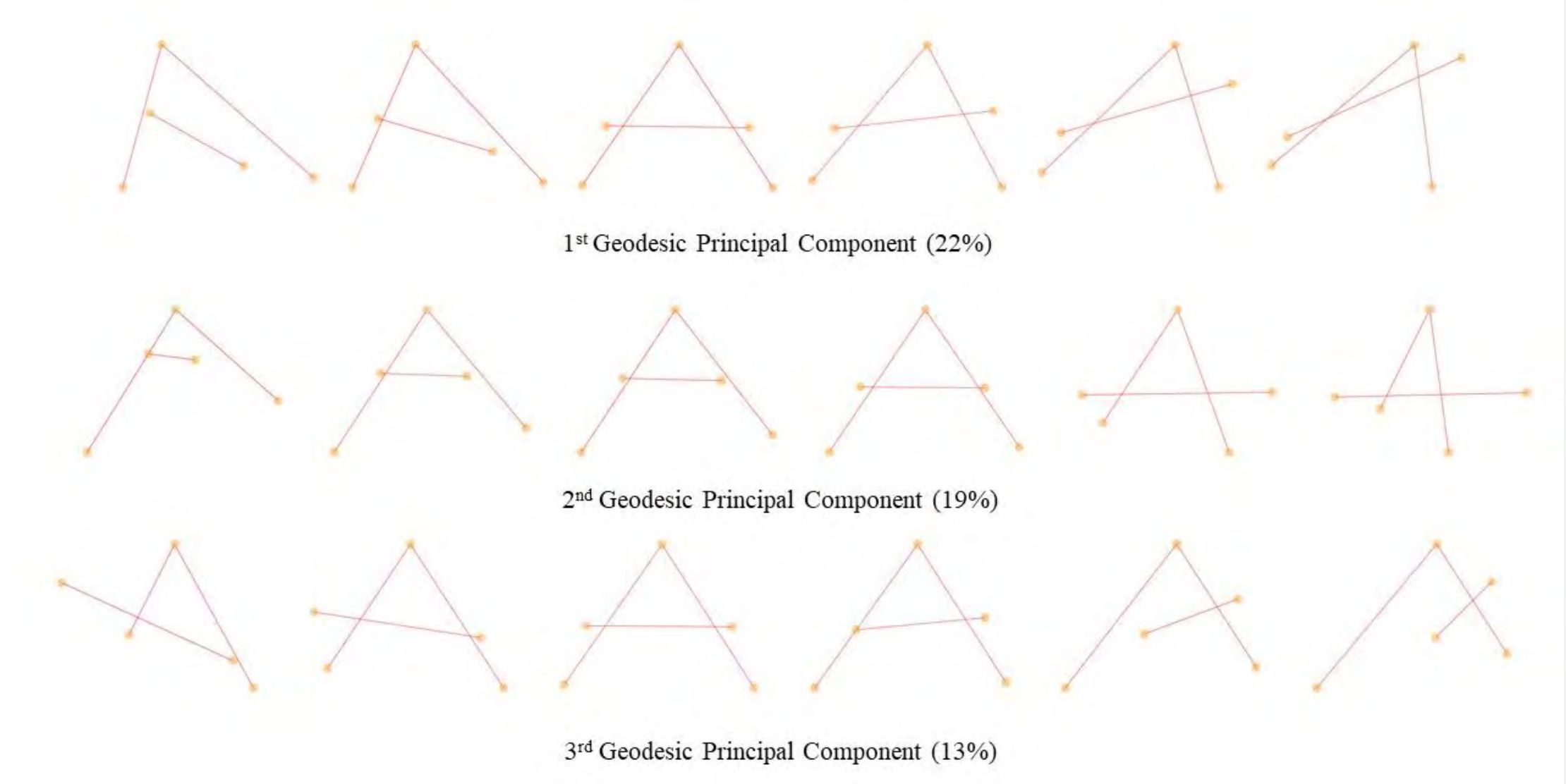
Letters



Letters



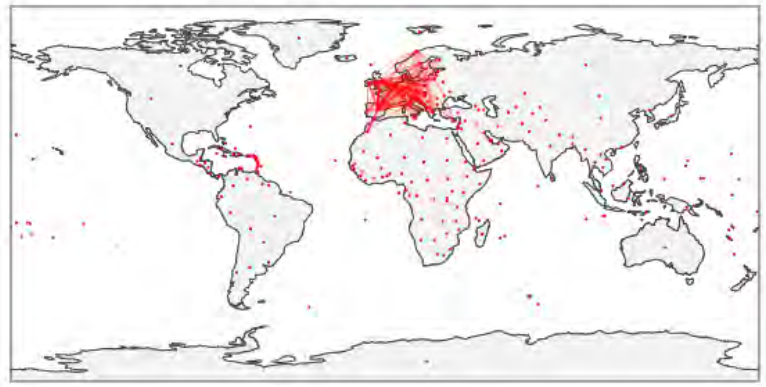
Letters GPCAnalysis



Example 2: Clustering

Airlines Dataset

Alitalia



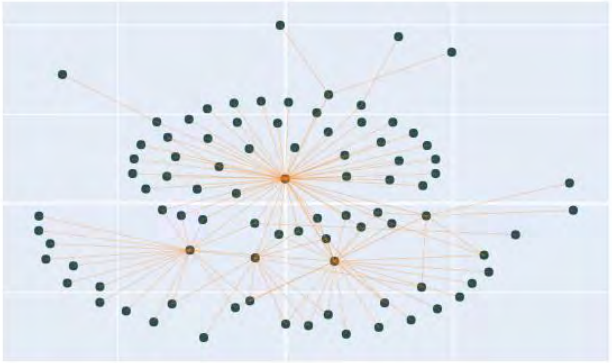
RyanAir

Luftansa

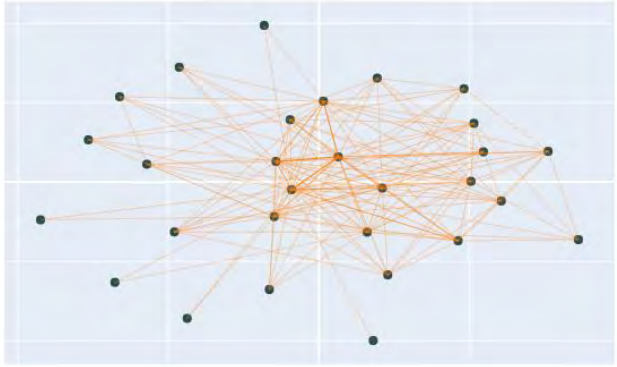
KLM

12 airlines companies OpenFlights Project

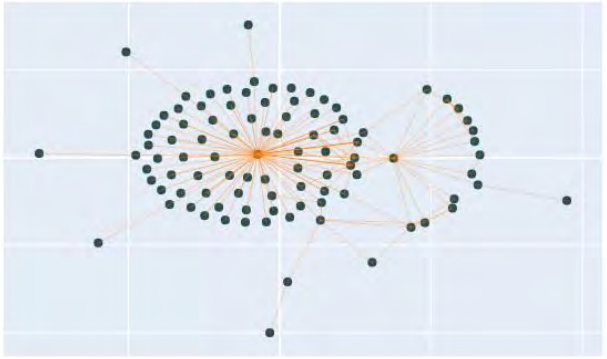
Airlines Dataset



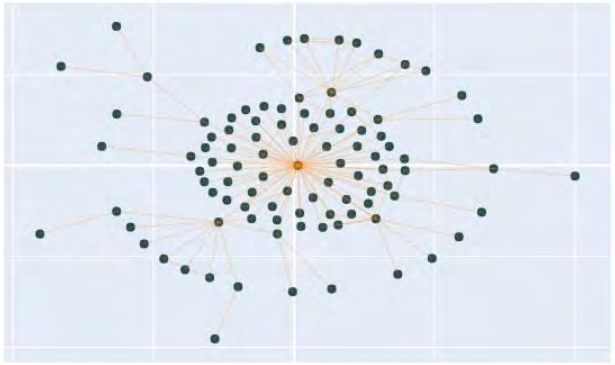
Alitalia



RyanAir

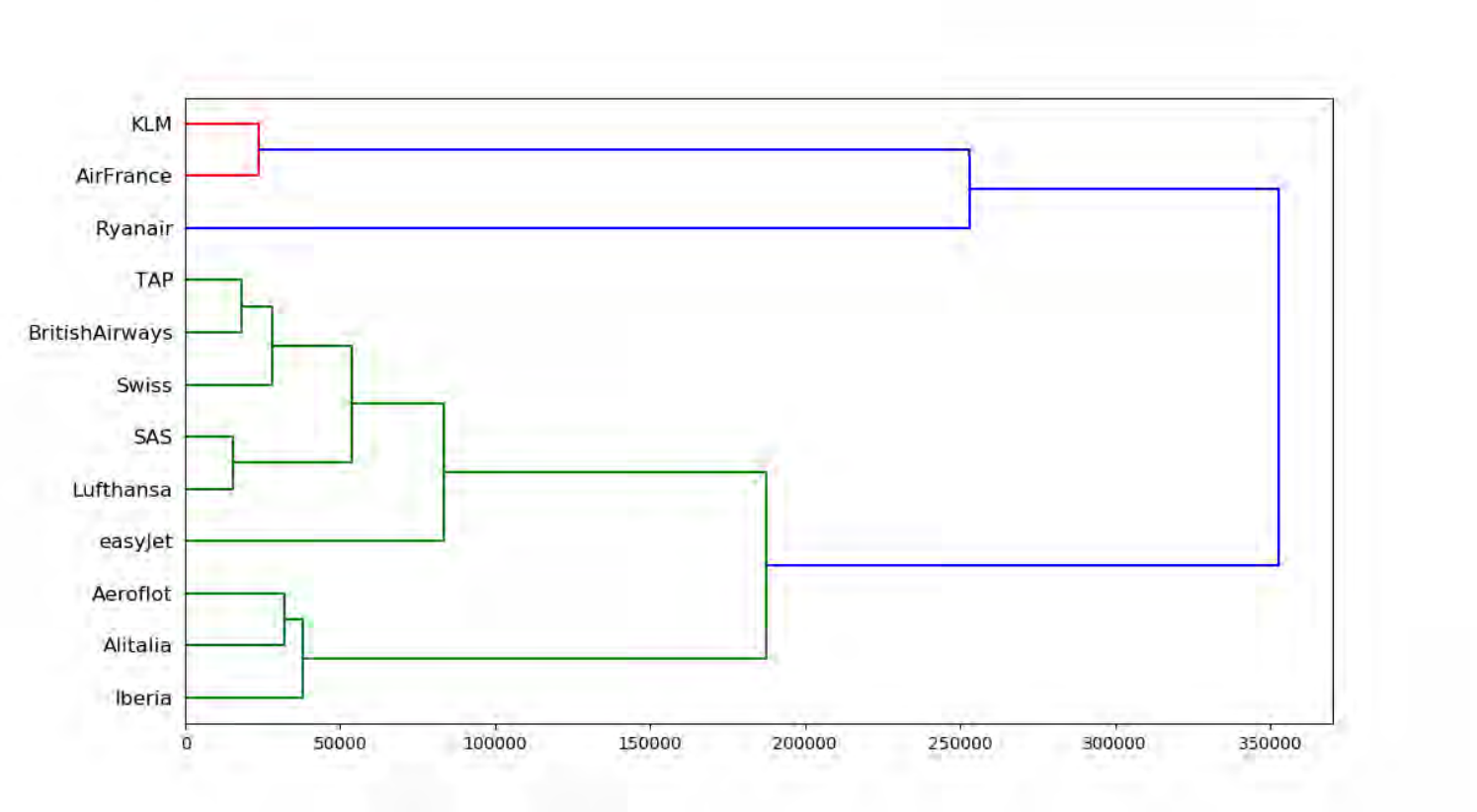


Luftansa



KLM

Cluster



Graduate Assignment Graph Matching, Euclidean Distance, Ward Linkage

Conclusions

Conclusions

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

⟹ Geometrical Properities of the Space (e.g. Curvature)

⟹ Geodesic Principal Component

⟹ Align All and Compute Algorithm

Next Step:

⟹ Introduce a Measure on the Space

⟹ Define a Regression Model Network-on-Network

Object Oriented Data Analysis of Population of Networks

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.

Brenner and Katsikis, Nighttime Lights and Population Density in North Dakota, 2016

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.

Gehl and Svarre. How to study public life. Island press, 2013.

Some References

Guo, Xiaoyang, Anuj Srivastava, and Sudeep Sarkar. "A Quotient Space Formulation for Statistical Analysis of Graphical Data." arXiv preprint arXiv:1909.12907 (2019).

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.

Jain, Brijnesh, and Klaus Obermayer. "On the sample mean of graphs." *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, 2008.

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

Nathan Keyfitz. Individual mobility in a stationary population. Population Studies, 27(2):335–352, 1973

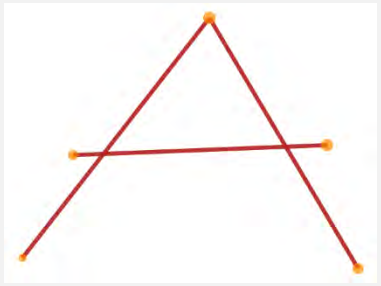
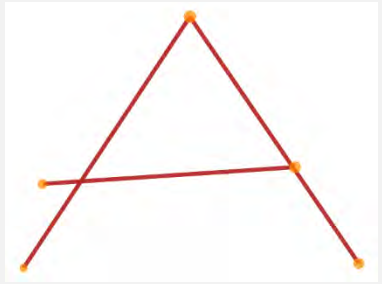
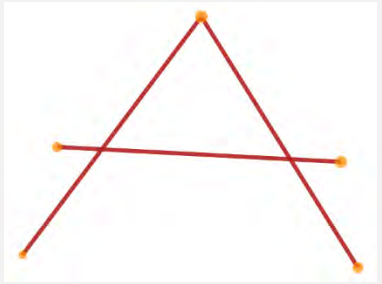
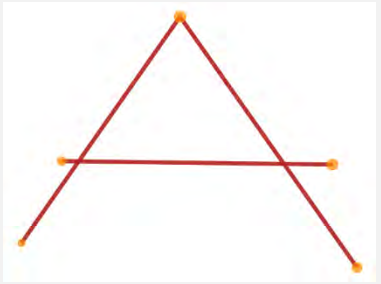
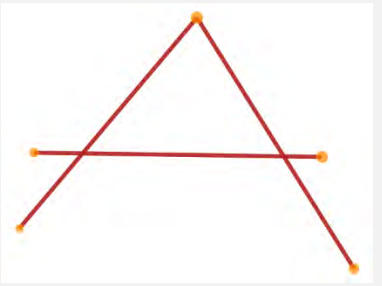
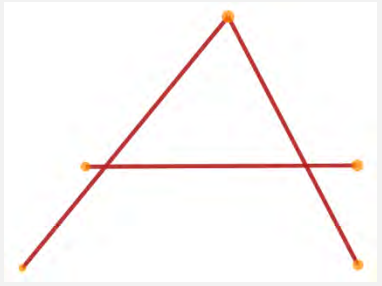
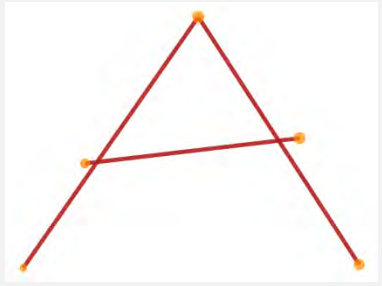
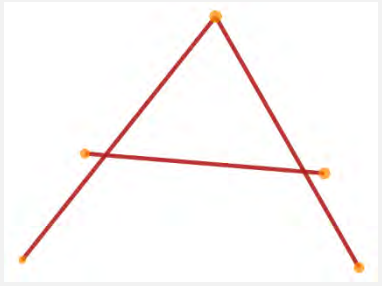
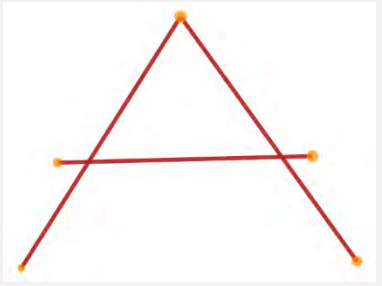
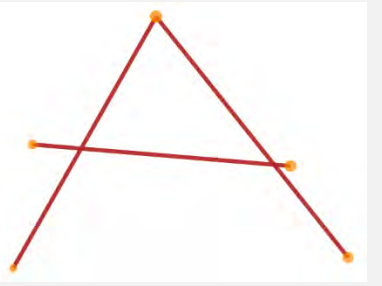
Nye, Tom MW, et al. "Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees." *Biometrika* 104.4 (2017): 901-922

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

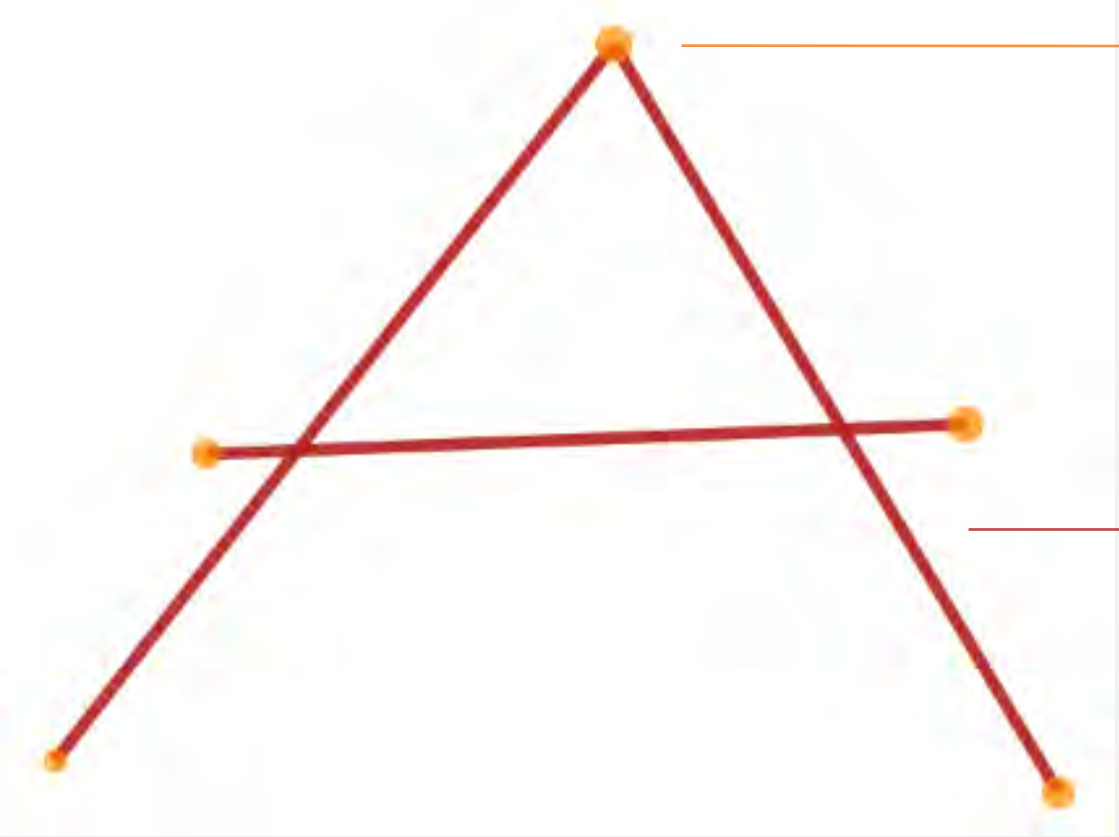
Simpson, Sean L., F. DuBois Bowman, and Paul J. Laurienti. "Analyzing complex functional brain networks: fusing statistics and network science to understand the brain." *Statistics surveys 7* (2013): 1.

Examples

Example: Attributes and Topology

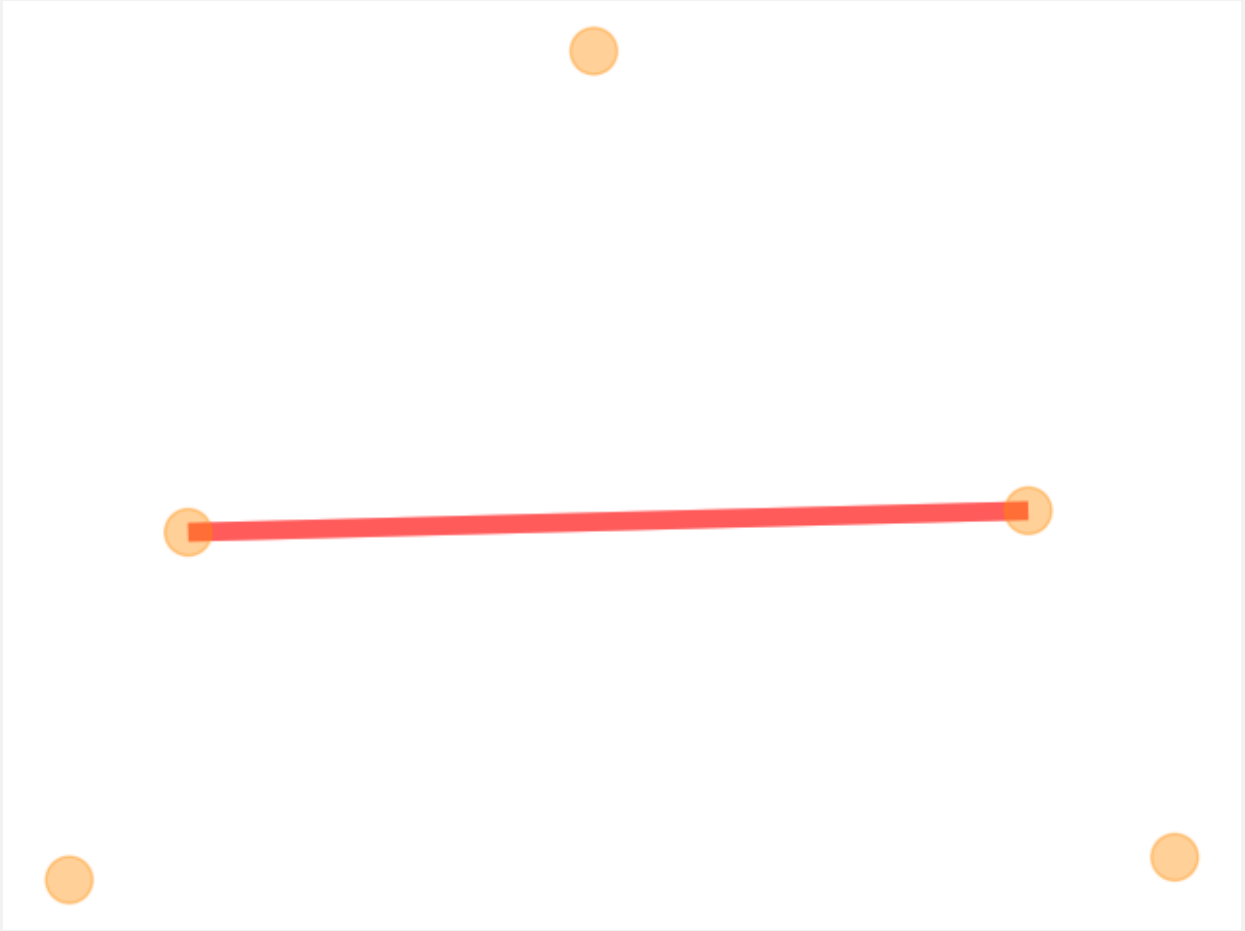
Letters DataSet



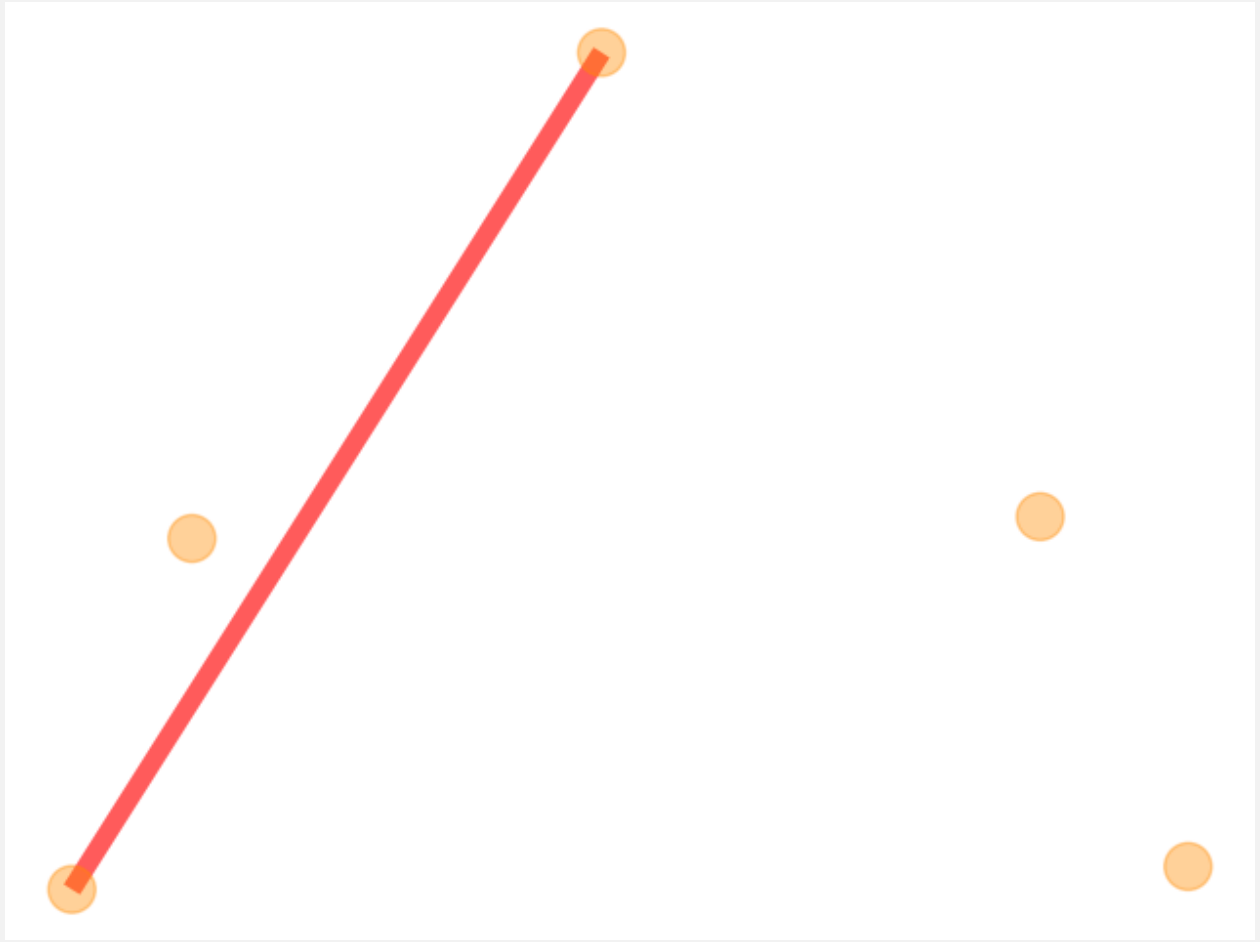
(x,y) coordinates

Presence of a drawn line

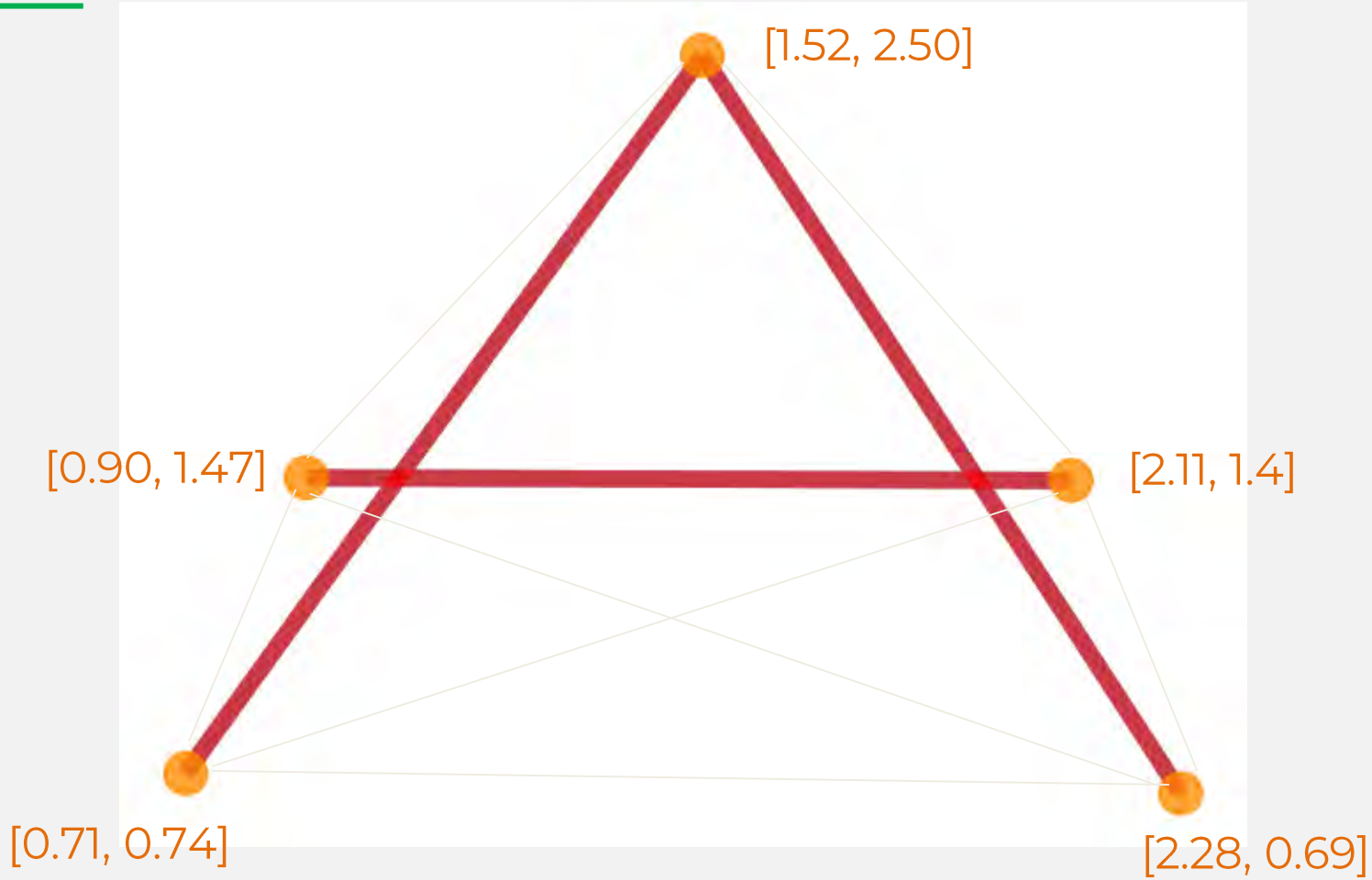
Letters DataSet



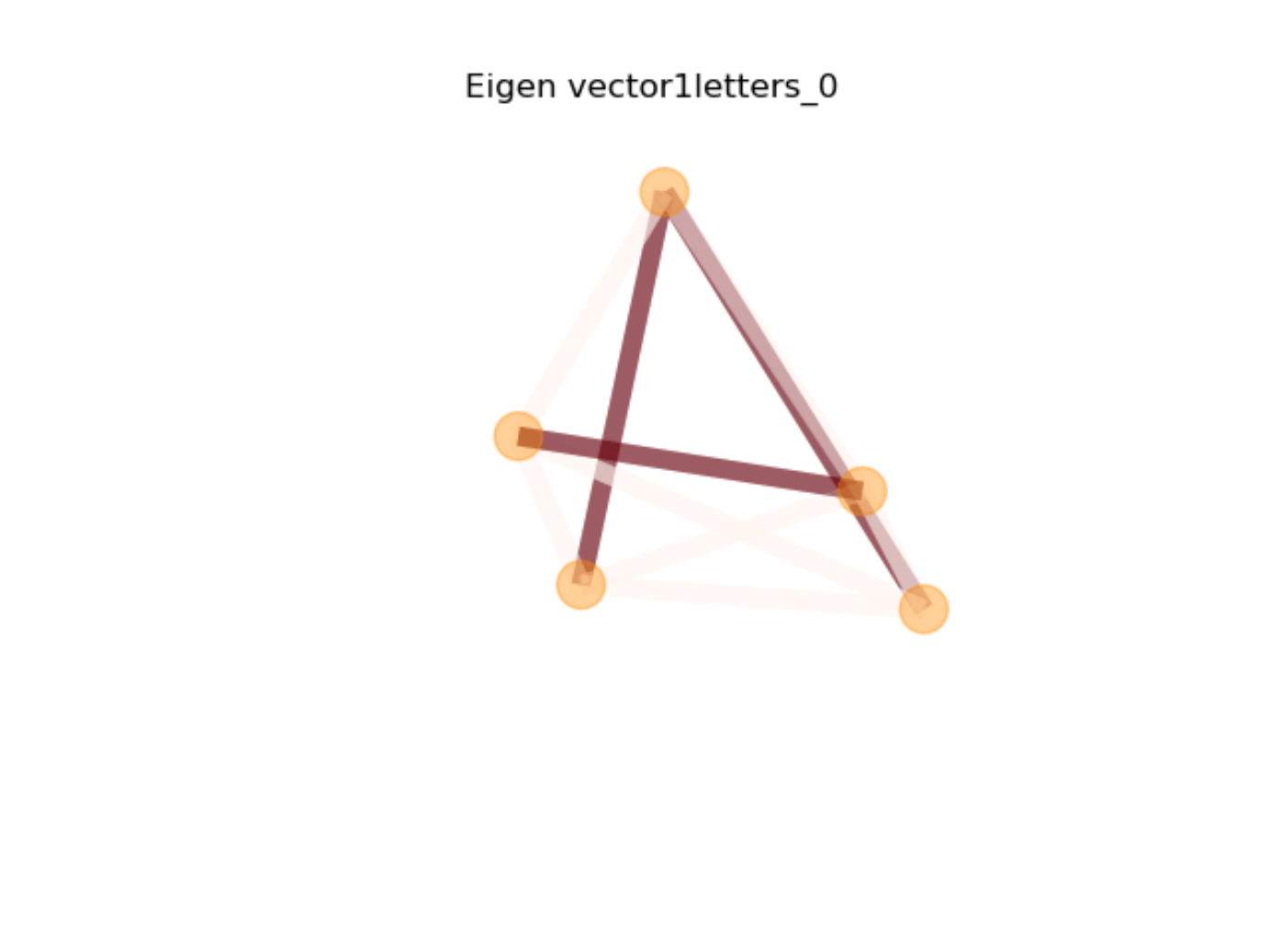
Letters DataSet



A mean



Geodesic Principal Component Analysis

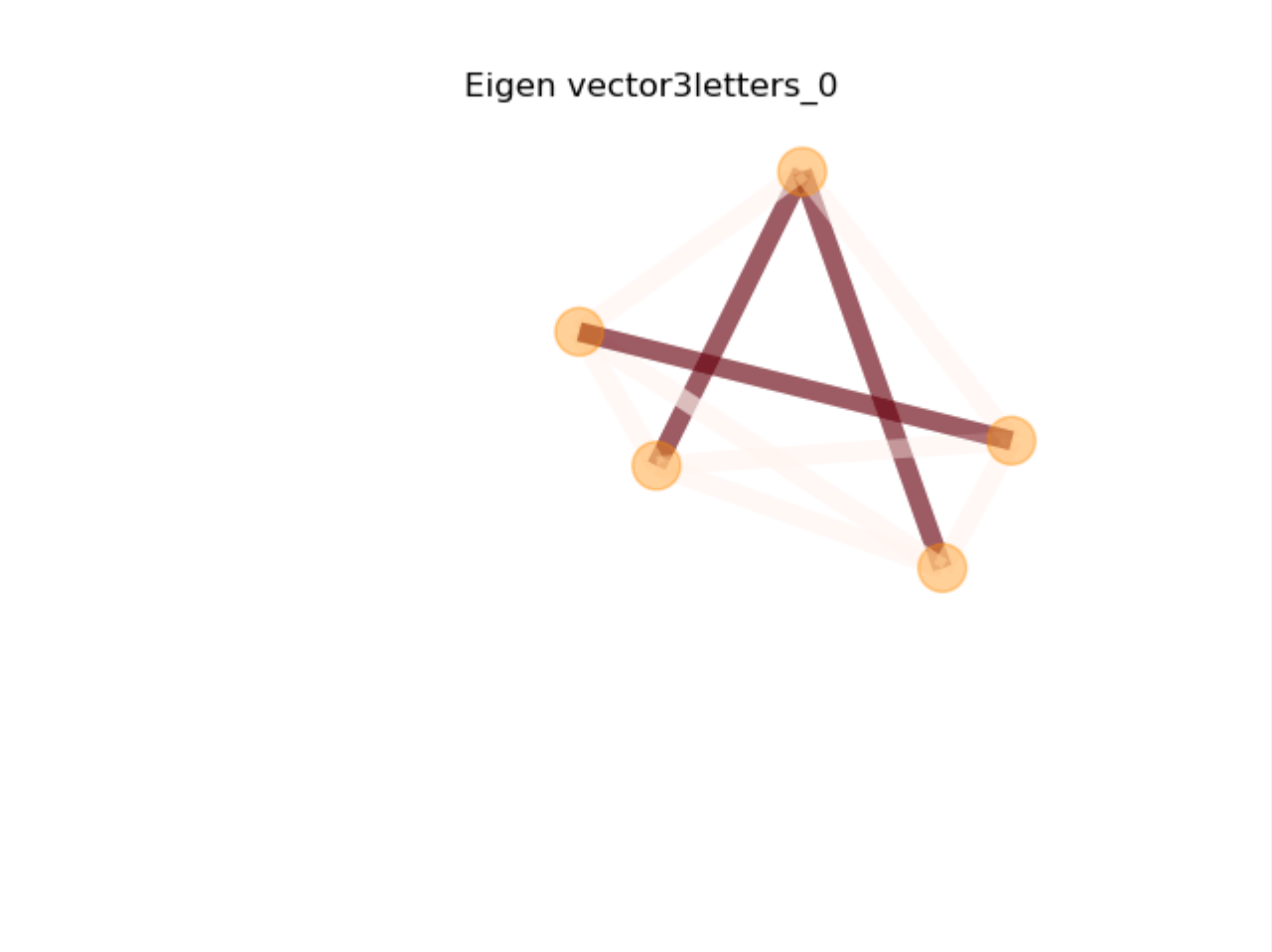


Capturing 18% of the total variance

Geodesic Principal Component Analysis

Capturing 16% of the total variance

Geodesic Principal Component Analysis



Capturing 14% of the total variance

Observation about different cases

	EDGES: All	EDGES: Not All
NODES: All same number	✓	✓ With intensity edge attributes or no attribute
NODES: Different number	✗	✓ With intensity edge and nodes attributes or no attribute