The geometry of graph space: Towards graph-valued statistics

Anna Calissano[1], **Aasa Feragen**[2], Simone Vantini[1]

[1] MOX - Modelling and Scientific Computing - Politecnico di Milano
[2] Section for Image Analysis & Computer Graphics, DTU Compute
afhar@dtu.dk

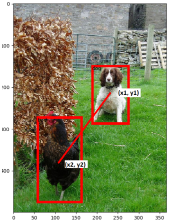
Session on Statistical methods for Non-Euclidean Data @CMStatistics

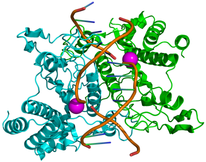
London, 16.12.2019

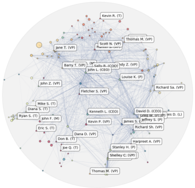
Danmarks Tekniske Universitet DTU

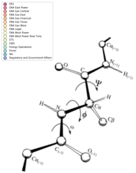
This work belongs to the PhD of Anna Calissano

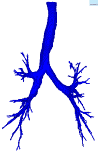
Graphs as data objects – they are everywhere!

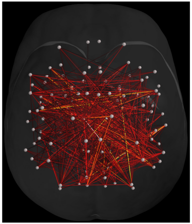






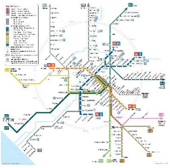






- Scene understanding, social networks, chemo/bioinformatics, brain connectivity

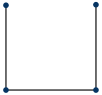
Graphs as data objects – they are everywhere!

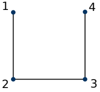
   

- Variable nodes, variable edges, attributes on nodes and edges

A space of graphs[1]

- A (weighted) graph can be represented by its adjacency matrix $A \in \mathbb{R}^{n \times n} =: \mathcal{A}$



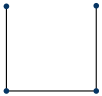


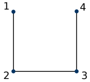
$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

[1] Jain et al, JMLR 2009, Pattern Recognition 2016

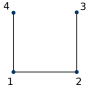
A space of graphs[1]

- A (weighted) graph can be represented by its adjacency matrix $A \in \mathbb{R}^{n \times n} =: \mathcal{A}$





$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$



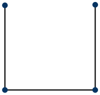
$$\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

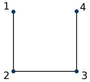
- Each graph has multiple adjacency matrix representations

[1] Jain et al, JMLR 2009, Pattern Recognition 2016

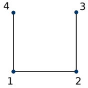
A space of graphs[1]

- A (weighted) graph can be represented by its adjacency matrix $A \in \mathbb{R}^{n \times n} =: \mathcal{A}$





$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$



$$\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

- Each graph has multiple adjacency matrix representations
- A graph space with unique graph representations: The quotient

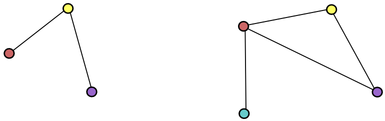
$$\mathcal{G} := \mathcal{A}/S_n$$

with respect to the node permutation group S_n

[1] Jain et al, JMLR 2009, Pattern Recognition 2016

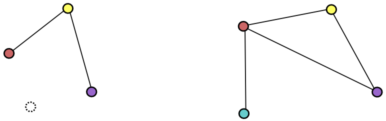
A general space of graphs

- Easy to accommodate graphs with different numbers of nodes:



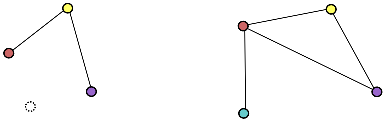
A general space of graphs

- Easy to accommodate graphs with different numbers of nodes:



A general space of graphs

- Easy to accommodate graphs with different numbers of nodes:



- Easy to extend to vector valued node and edge weights

$$A \in \mathbb{R}^{n \times n} =: \mathcal{A} \quad \rightsquigarrow \quad A \in (\mathbb{R}^d)^{n \times n} =: \mathcal{A}$$

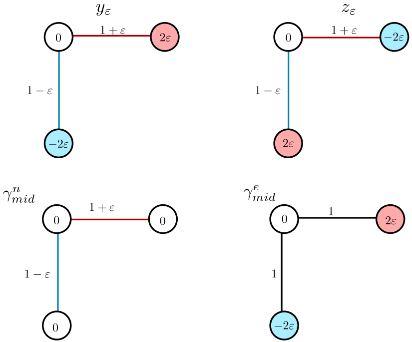
Use of Jain's graph space for statistics

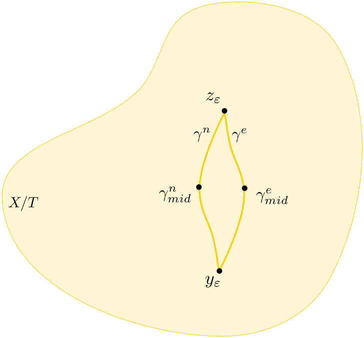
- Jain, Obermayer: Structure Spaces. Journal of Machine Learning Research. (2009)
- Jain, Obermayer: Large Sample Statistics in the Domain of Graphs. SSPR/SPR (2010)
- Jain, Obermayer: Maximum Likelihood for Gaussians on Graphs. GbRPR (2011)
- Jain: Maximum likelihood method for parameter estimation of bell-shaped functions on graphs. Pattern Recognition Letters (2012)
- Calissano, Feragen, Vantini: Analysis of Populations of Networks: Structure Spaces and the Computation of Summary Statistics. ICSA (2019)
- Guo, Srivastava, Sarkar: A Quotient Space Formulation for Statistical Analysis of Graphical Data. arXiv preprint arXiv:1909.12907 (2019).
- Kolaczyk, Lin, Rosenberg, Xu, Walters: Averages of Unlabeled Networks: Geometric Characterization and Asymptotic Behavior. arXiv preprint arXiv:1709.02793 (2019).

The geometry of graph space

Theorem

Graph-space geodesics are not necessarily unique.

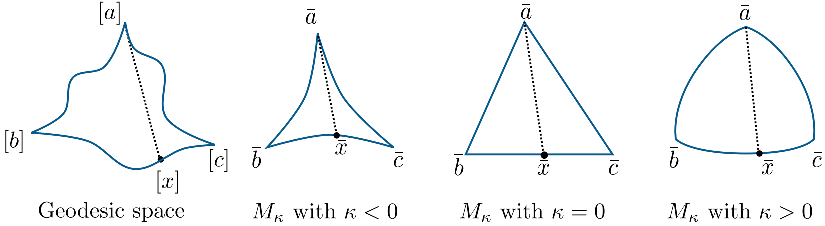




The geometry of graph space

Theorem

Graph-space curvature is unbounded from above.



Curvature bounded from above by κ if for all geodesic triangles in X/G are *thinner* than their corresponding comparison triangles in M_κ.

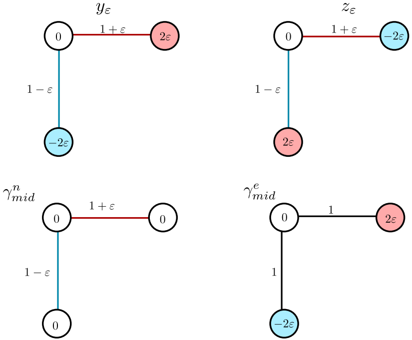
$$d_{X/G}([x],[a]) \leq d_{M_\kappa}(\bar{x},\bar{a}) :$$

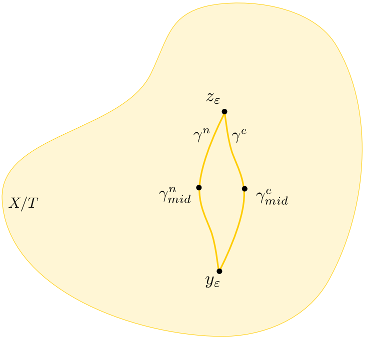
First statistic: Fréchet mean

$$[m] = \mathrm{argmin}_{[x]\in X/G} \sum_{i=1...n} d^2_{X/G}([x],[x_i])$$

Theorem

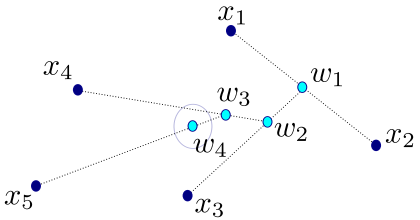
Fréchet means are not generally unique in graph space X/G.





Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

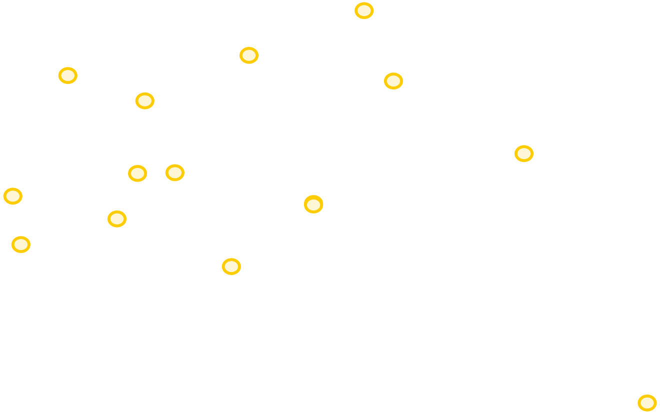
Iterative weighted midpoints / stochastic gradient descent



Note: Proofs of convergence usually require being able to work in a neighborhood with unique geodesics.

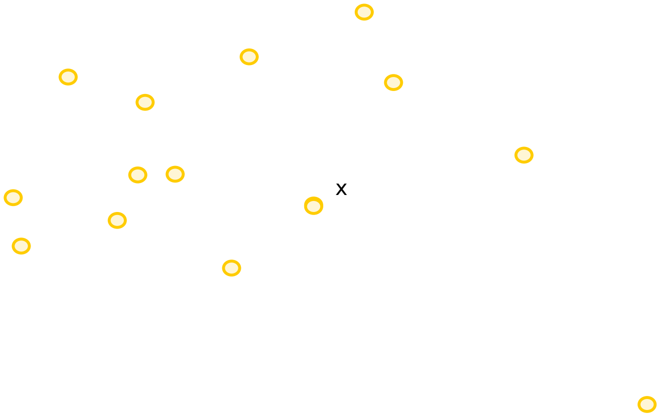
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose: Align all and compute (AAC) / Generalized Procrustes analysis



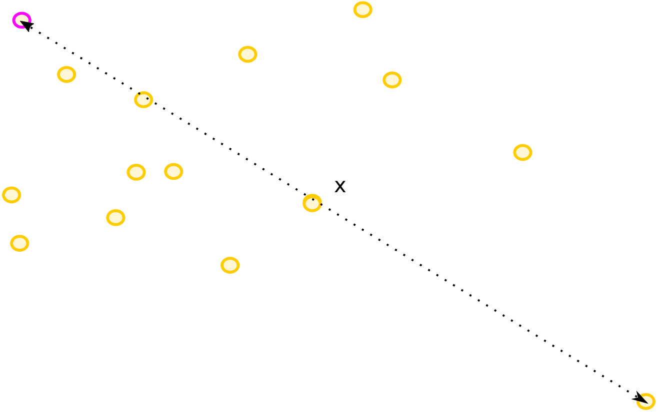
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose: Align all and compute (AAC) / Generalized Procrustes analysis



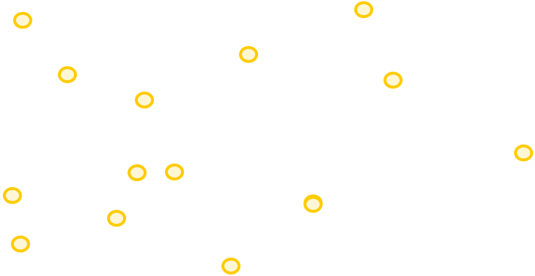
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose: Align all and compute (AAC) / Generalized Procrustes analysis



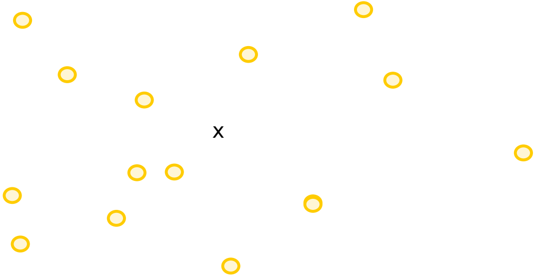
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose: Align all and compute (AAC) / Generalized Procrustes analysis



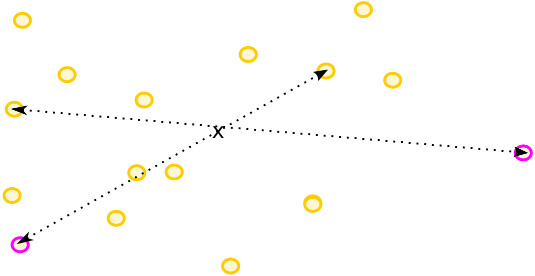
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose: Align all and compute (AAC) / Generalized Procrustes analysis



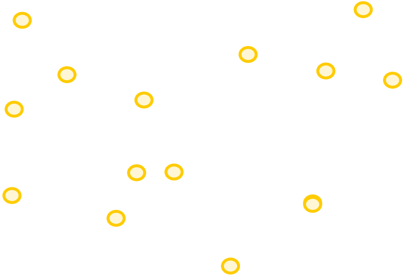
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose: Align all and compute (AAC) / Generalized Procrustes analysis



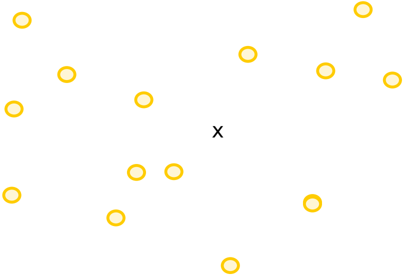
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose: Align all and compute (AAC) / Generalized Procrustes analysis



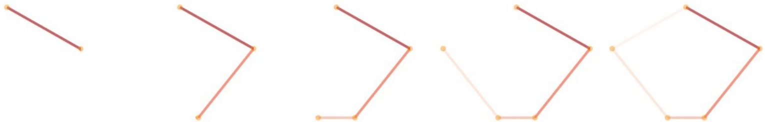
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose: Align all and compute (AAC) / Generalized Procrustes analysis



Guarantee convergence to local minimum for generic dataset.

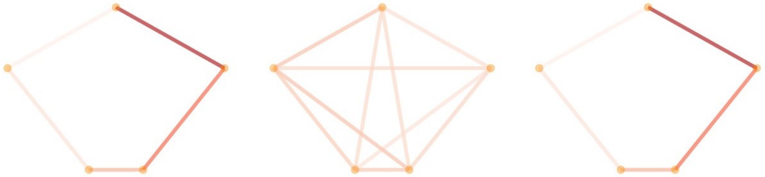
Examples: Synthetic dataset



Sample randomly from five shown equivalence classes

Examples: Synthetic dataset

Fréchet mean:



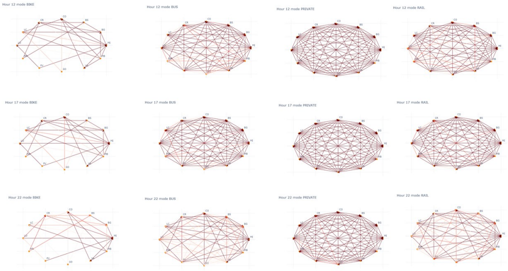
Theoretical Fréchet Mean

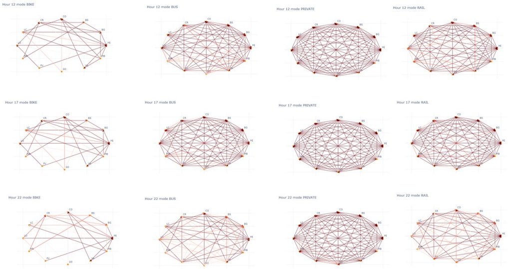
Estimation of Fréchet Mean on X space

Estimation of Fréchet Mean on Graph Space with AAC

Mobility networks Lombardia region

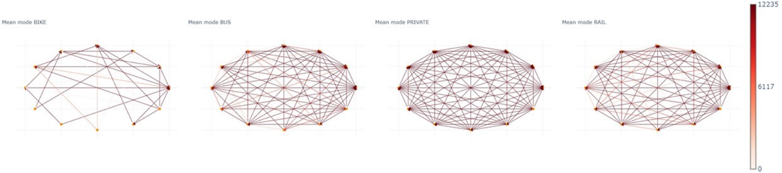
Dataset:





Mobility networks Lombardia region

Fréchet mean:



Next: Principal components via AAC

Result: Principal components 1-d represented as subspaces of $\mathbb{R}^{n\times n}$

initialize by aligning all data graphs to a random adjacency matrix;

while *While not converged* **do**

  perform PCA in $\mathbb{R}^{n\times n}$;

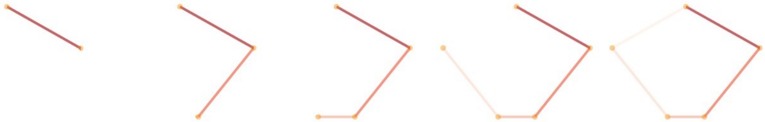
  choose representatives of all data graphs in optimal position with the first PC in $\mathbb{R}^{n\times n}$.

end

Algorithm 1: PCA via AAC

Example: Synthetic dataset

Dataset

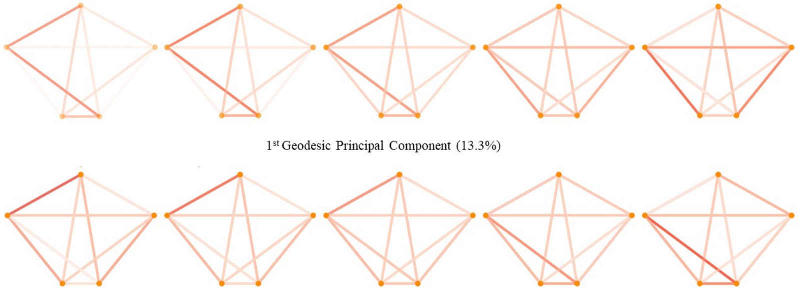


Sampled randomly from each of the 5 shown equivalence classes.

Example: Synthetic dataset

PCA in space of adjacency matrices.

Geodesic Principal Components in X with AAC

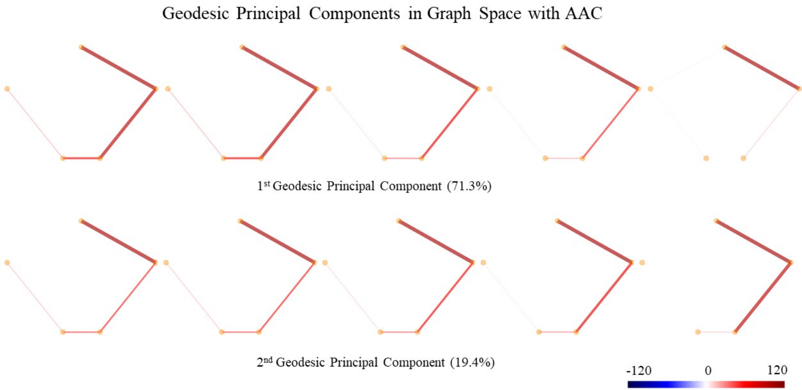


1st Geodesic Principal Component (13.3%)

2nd Geodesic Principal Component (12.3%)

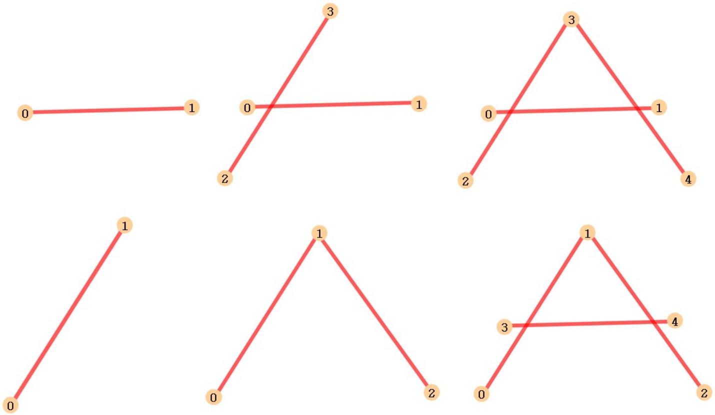
Example: Synthetic dataset

PCA in graph space (with AAC algorithm):



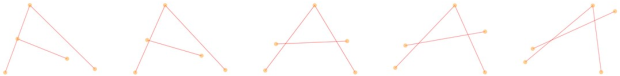
Example: Handwritten letter “A”

Dataset:

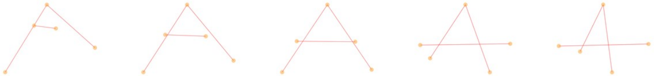


Example: Handwritten letter "A"

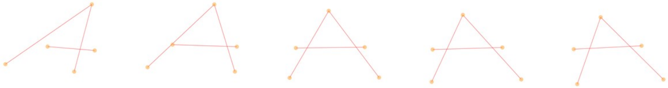
Principal components in graph space:



1st Geodesic Principal Component (22%)



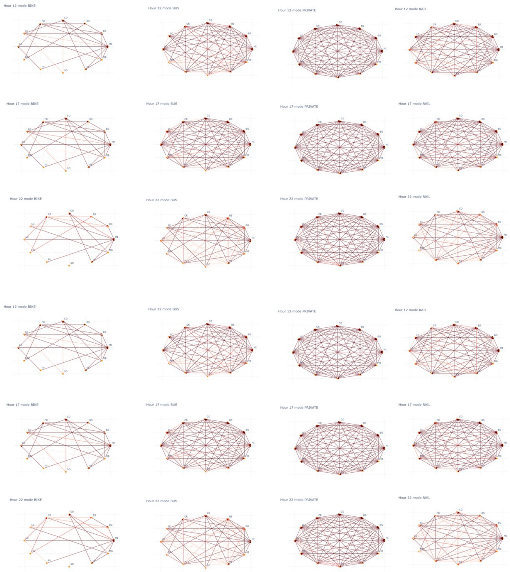
2nd Geodesic Principal Component (19%)



3rd Geodesic Principal Component (13%)

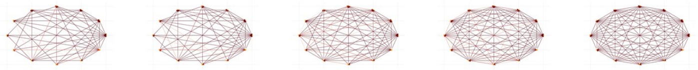
Mobility networks Lombardia region

Dataset:

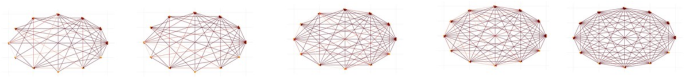


Mobility networks Lombardia region

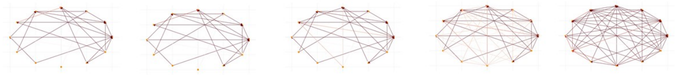
Principal components in graph space:



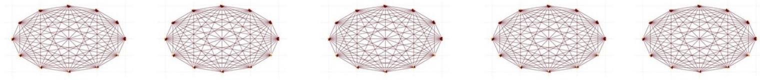
Variation of Rail along the 1st GPCA



Variation of Bus along the 1st GPCA



Variation of Bike along the 1st GPCA



Variation of Private along the 1st GPCA

Explaining 71,9% of the total variability

Discussion

Advantages:

- Total space is $(\mathbb{R}^d)^{n \times n}$, easing generalization of Euclidean methods

Limitations:

Discussion

Advantages:

- Total space is $(\mathbb{R}^d)^{n \times n}$, easing generalization of Euclidean methods
- Jain et al, and by now others, have developed a wide range of theoretical and practical tools for distance-based statistics and learning on graphs in these spaces

Limitations:

Discussion

Advantages:

- Total space is $(\mathbb{R}^d)^{n \times n}$, easing generalization of Euclidean methods
- Jain et al, and by now others, have developed a wide range of theoretical and practical tools for distance-based statistics and learning on graphs in these spaces

Limitations:

- Distances are generally NP-complete due to graph matching problem $\rightsquigarrow$ approximations

Discussion

Advantages:

- Total space is $(\mathbb{R}^d)^{n \times n}$, easing generalization of Euclidean methods
- Jain et al, and by now others, have developed a wide range of theoretical and practical tools for distance-based statistics and learning on graphs in these spaces

Limitations:

- Distances are generally NP-complete due to graph matching problem $\rightsquigarrow$ approximations
- Graph space geometry is highly non-Euclidean (and it is not a manifold); this is likely to affect statistics in ways we do not yet understand.

Discussion: Align all and compute

- ''Tangent space" approach: Align all points with a representative of the mean and perform statistics in the total space ($\Leftrightarrow$ tangent space statistics in manifolds). This is done in Guo et al (2019).

Discussion: Align all and compute

- "Tangent space" approach: Align all points with a representative of the mean and perform statistics in the total space ($\Leftrightarrow$ tangent space statistics in manifolds). This is done in Guo et al (2019).
- As we know it from manifolds, such an approach leads to increased distortion with distance to mean.

Discussion: Align all and compute

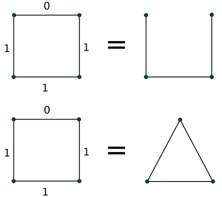
- "Tangent space" approach: Align all points with a representative of the mean and perform statistics in the total space ($\Leftrightarrow$ tangent space statistics in manifolds). This is done in Guo et al (2019).
- As we know it from manifolds, such an approach leads to increased distortion with distance to mean.
- Non-uniqueness of geodesics lead to further issues as you move outside the injectivity radius from the mean.

Discussion: Align all and compute

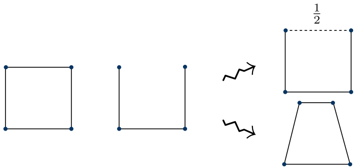
- ''Tangent space" approach: Align all points with a representative of the mean and perform statistics in the total space ($\Leftrightarrow$ tangent space statistics in manifolds). This is done in Guo et al (2019).
- As we know it from manifolds, such an approach leads to increased distortion with distance to mean.
- Non-uniqueness of geodesics lead to further issues as you move outside the injectivity radius from the mean.
- Aligning with higher principal components, as we do, can help alleviate this problem – but NB! Don't go too high... if you include them all you will stay put. So how many?

Open problems

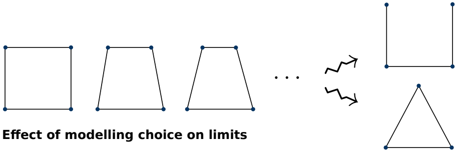
- For geometric graphs, Jain's model does not capture well contracting branches:
- This is likely to significantly complicate geometry and computation.



Modelling choice:
What does the 0 attribute represent?



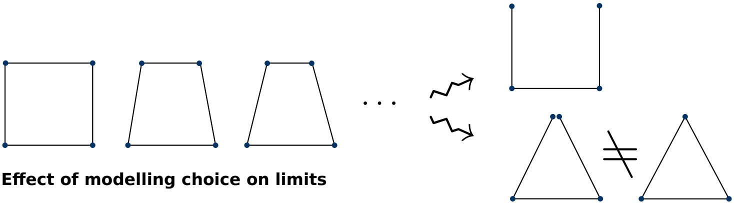
Effect of modelling choice on Fréchet mean



Effect of modelling choice on limits

Open problems

- Ability to merge nodes



Effect of modelling choice on limits