Predicting Graphs

Aasa Feragen

Section for Visual Computing, DTU Compute

afhar@dtu.dk

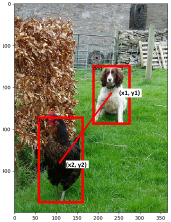
`http://www2.compute.dtu.dk/~afhar`

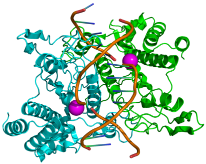
Computational Mathematics for Data Science

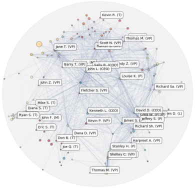
DTU, 15.11.2023

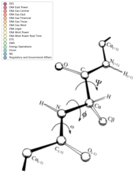
Danmarks
Tekniske
Universitet

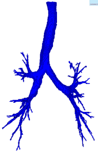
Graphs as data objects – they are everywhere!

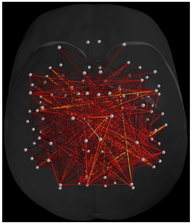








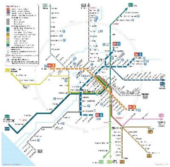




- Scene understanding, social networks, chemo/bioinformatics, brain connectivity

Graphs as data objects – they are everywhere!

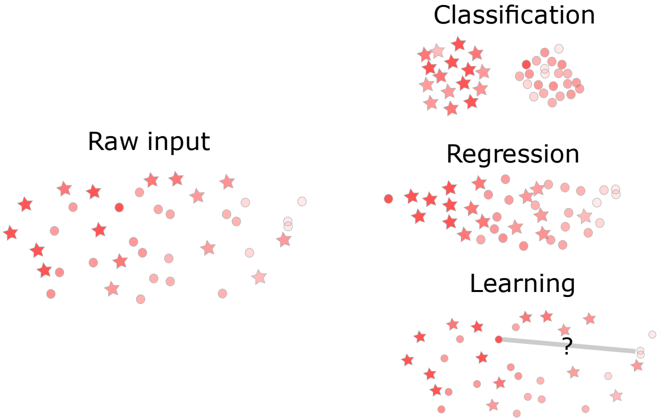




- Variable nodes, variable edges, attributes on nodes and edges

Predicting graphs

Predicting graph-structured output requires more from your learned representation than predicting a class or a single real response



Starting point: Via geometric statistics

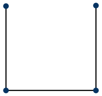
Collaboration with Anna Calissano and Simone Vantini, MOX, Politecnico Milano

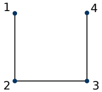
Populations of Unlabelled Networks: Graph Space Geometry and Generalized Geodesic Principal Components, Biometrika, 2023

Graph-valued regression: Prediction of unlabelled networks in a non-Euclidean graph space, Journal of Multivariate Analysis, 2022

A space of graphs

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



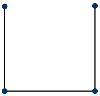


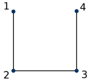
$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Jain et al, JMLR 2009

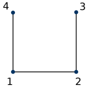
A space of graphs

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





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



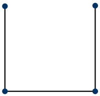
$$\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

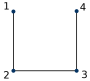
- Each graph has multiple adjacency matrix representations

Jain et al, JMLR 2009

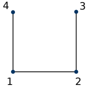
A space of graphs

- 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

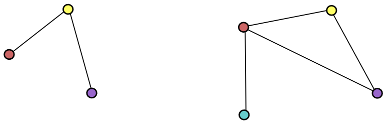
$$\mathcal{G} := \mathcal{A}/S_n$$

with respect to the node permutation group S_n

Jain et al, JMLR 2009

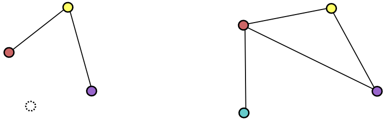
A general space of graphs

- Easy to accommodate graphs with different numbers of nodes:



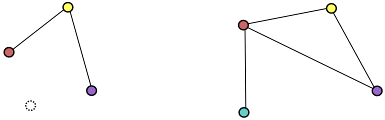
A general space of graphs

- 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

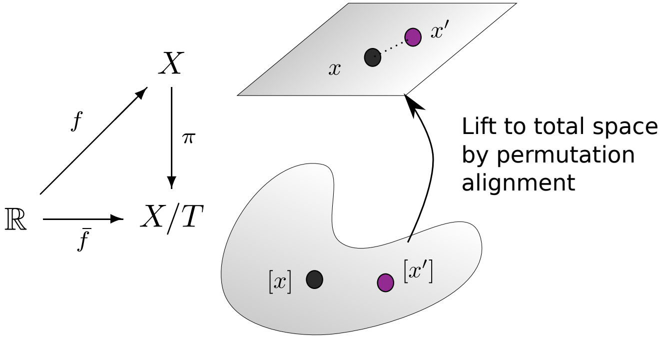
$$A \in \mathbb{R}^{n \times n} =: \mathcal{A} \quad \rightsquigarrow \quad A \in (\mathbb{R}^d)^{n \times n} =: \mathcal{A}$$

Existing work on statistics in Jain's graph space

- 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. JMIV (2021).
- Kolaczyk, Lin, Rosenberg, Xu, Walters: Averages of Unlabeled Networks: Geometric Characterization and Asymptotic Behavior. Annals of Statistics (2020).

Graph Space Geometry

Graph Space is a Quotient Space

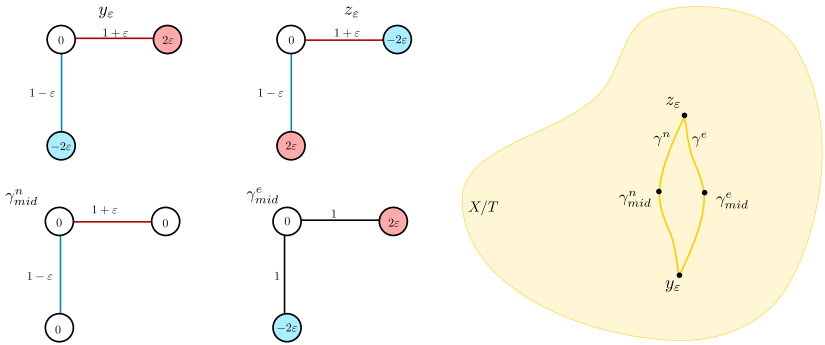


- Graph space X/T inherits a metric from the Euclidean metric on $X = \mathbb{R}^{n \times n}$
- Graph space is a geodesic metric space – any two points joined by a shortest path
- The total space X can be thought of as a "tangent space" where a "log map" at any base point graph $[x]$ is equivalent to alignment to its fixed representative x

The geometry of graph space

Theorem

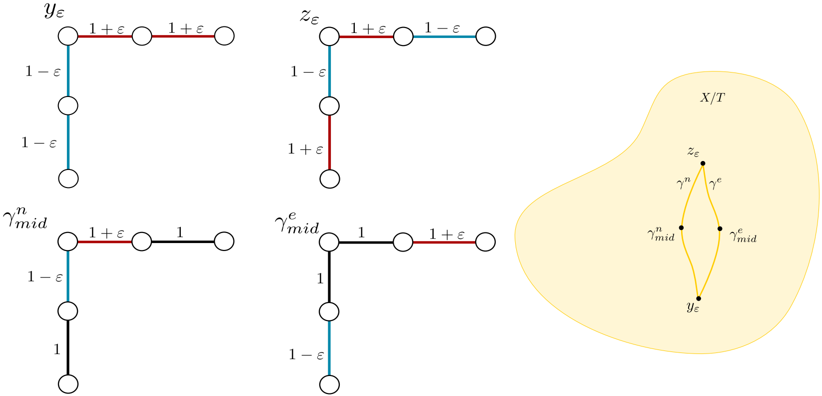
Graph-space geodesics are not necessarily unique.



Theorem

Graph-space curvature is unbounded from above.

The counterexample is not dependent on having node attributes



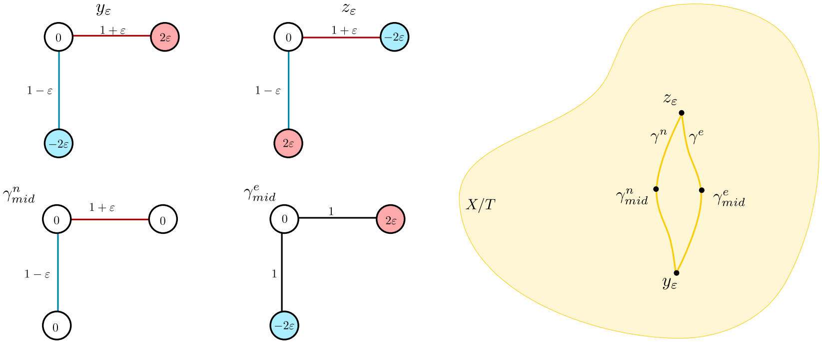
Statistics

First statistic: Fréchet mean

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

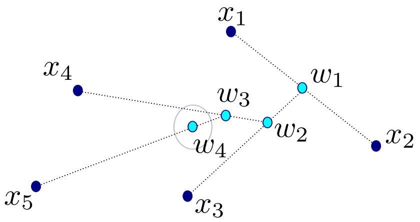
Theorem

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



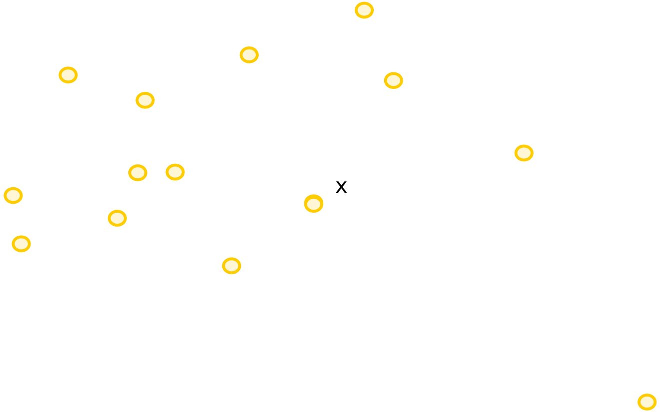
Note: Proofs of convergence usually require being able to work in a neighborhood with unique geodesics.

Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose an analogy with Generalized Procrustes Analysis, and call the general strategy "Align all and compute" (AAC).

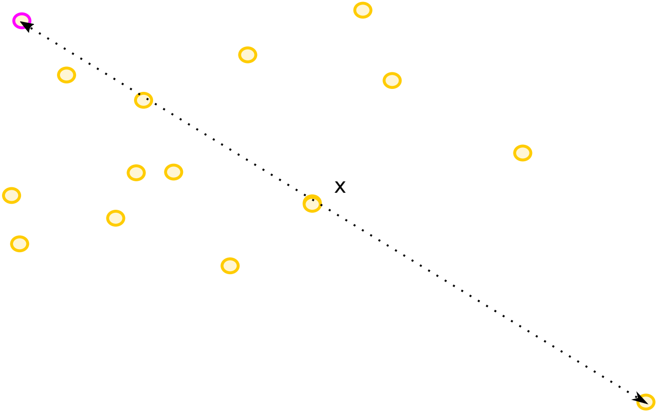
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose an analogy with Generalized Procrustes Analysis, and call the general strategy "Align all and compute" (AAC).



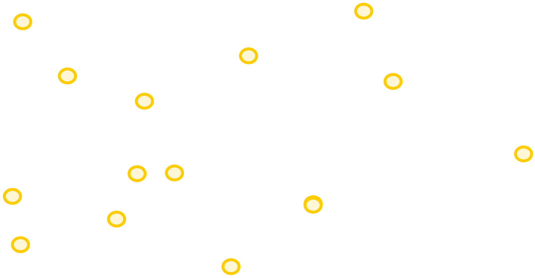
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose an analogy with Generalized Procrustes Analysis, and call the general strategy "Align all and compute" (AAC).



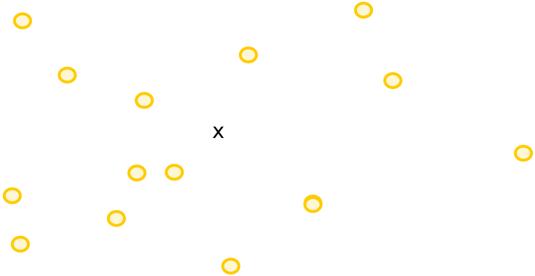
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose an analogy with Generalized Procrustes Analysis, and call the general strategy "Align all and compute" (AAC).



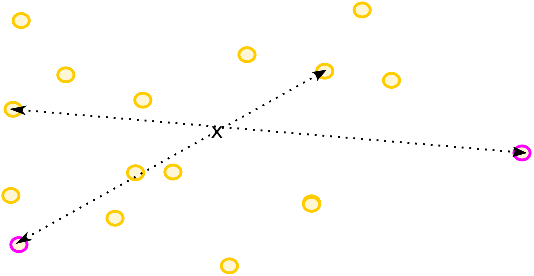
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose an analogy with Generalized Procrustes Analysis, and call the general strategy "Align all and compute" (AAC).



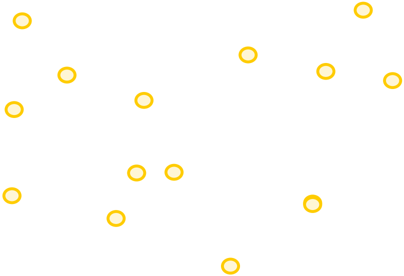
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose an analogy with Generalized Procrustes Analysis, and call the general strategy "Align all and compute" (AAC).



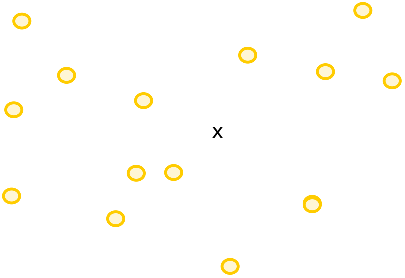
Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

We choose an analogy with Generalized Procrustes Analysis, and call the general strategy “Align all and compute” (AAC).



Existing algorithms and heuristics for computing Fréchet means in nonlinear spaces

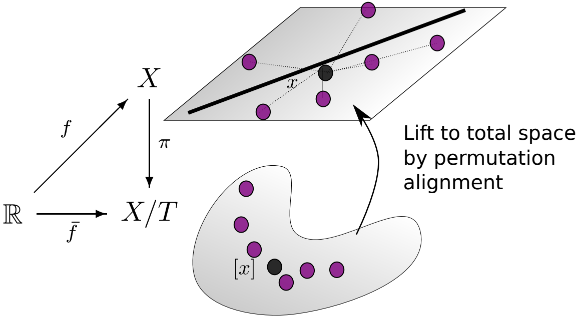
We choose an analogy with Generalized Procrustes Analysis, and call the general strategy “Align all and compute” (AAC).



Guarantee convergence to local minimum in finite time for generic dataset.

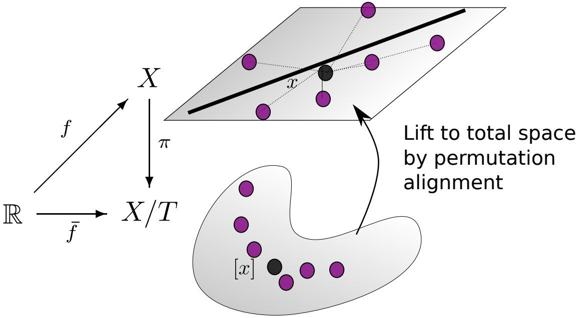
Higher order statistics: “Tangent space” approach

- ”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).
- Guo, Srivastava, Sarkar (2021): Tangent space PCA



Higher order statistics: “Tangent space” approach

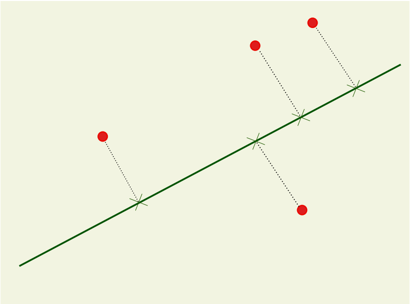
- ”Tangent space“ approach: Align all points with a representative of the mean and perform statistics in the total space ($\Leftrightarrow$ tangent space statistics in manifolds).
- Guo, Srivastava, Sarkar (2021): Tangent space PCA
- Risk distorted residuals away from the base point



Intrinsic dimensionality reduction: Generalized Geodesic PCA via AAC

- **Data:** $[x_1], \ldots, [x_k] \in X/T$
- **Model:** Optimize over generalized geodesics $\delta = \pi \circ \gamma$ for geodesic γ in space of adjacency matrices X
- **Task:** Find generalized geodesic $\delta_1 \in \Gamma(X/T)$ such that

$$\delta_1 = \mathrm{argmin}_{\delta \in \Gamma(X/T)} \sum_{i=1}^{k} (d^2_{X/T}([x_i], \delta))$$



AAC for Generalized Geodesic Principal components

Algorithm 1: PCA 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 $\mathbb{R}^{n \times n}$;

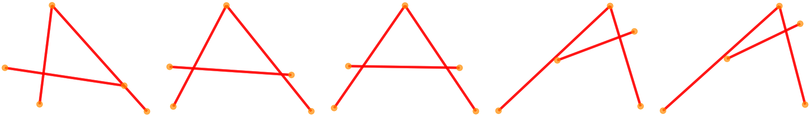
choose representatives of all data graphs in optimal position with the first PC in $\mathbb{R}^{n \times n}$.

end

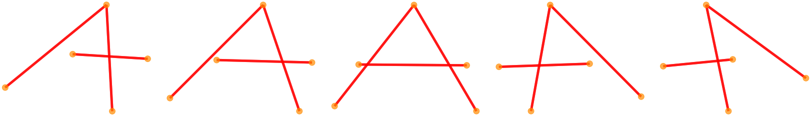
Theorem

The AAC algorithm converges to a local minimum in finite time.

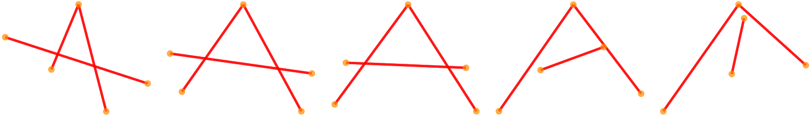
Example: Handwritten letter “A”



1^{st} GGPC (27▷3%)

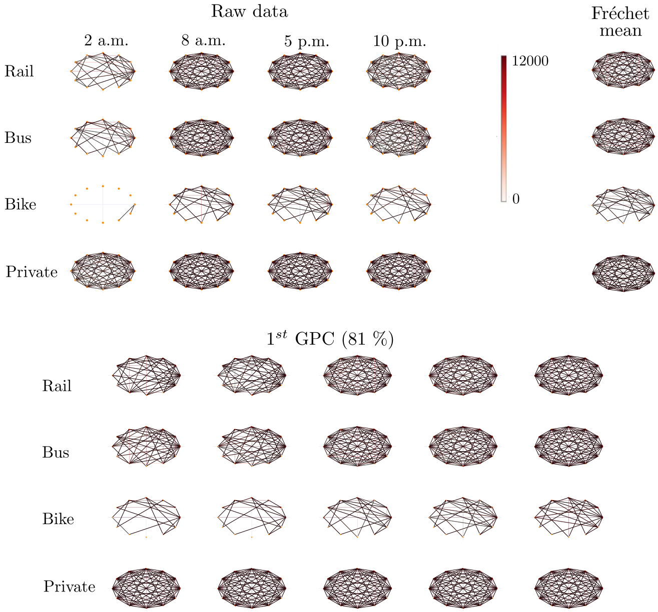


2^{nd} GGPC (18▷3%)



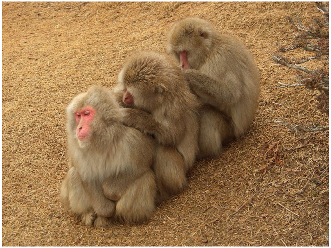
3^{rd} GGPC (15▷6%)

Mobility networks Lombardia region

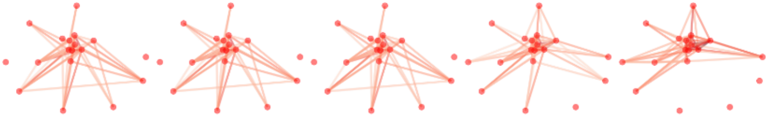


Mammals grooming

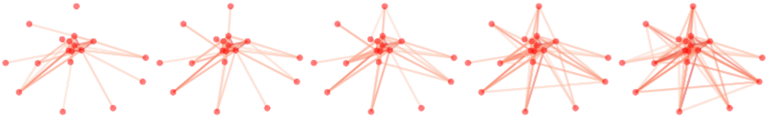
- Networks represent grooming interaction/association between baboons; they can have different roles (alpha/beta) over time $\Rightarrow$ naturally unlabeled networks
- Each network summarizes data collected within 30 days before and 90 days after each knockout.
- Knockout = when a given alpha or beta male leaves the group.



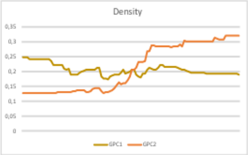
Mammals grooming

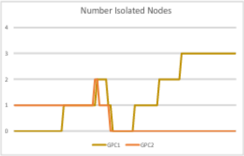


1^{st} GPC (18 %)



2^{nd} GPC (13%)



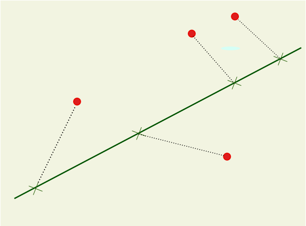


Graph-valued regression models

Regression models taking values in Graph Space

- **Data:** $(s_1, [x_1]), \dots, (s_k, [x_k])$, where $(s_i, [x_i]) \in \mathbb{R}^p \times X/T$
- **Model:** Optimize over $\bar{f} = \pi \circ f$ for linear regression models f on the space X of adjacency matrices
- **Task:** Describe the relationship $f \colon \mathbb{R}^p \to X/T$ minimizing

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



- **Note:** Residuals measured between predicted and true output.

AAC for Linear Regression models taking values in Graph Space

Algorithm 2: Linear Regression Models in Graph Space via AAC

Result: Predictions $f(s) \in \mathbb{R}^{n \times n}$ for any $s \in \mathbb{R}^d$

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

while *While not converged* **do**

- estimate linear regression model in $\mathbb{R}^{n \times n}$;
- choose representatives of all data graphs in optimal position with current estimate of the regression line in $\mathbb{R}^{n \times n}$.

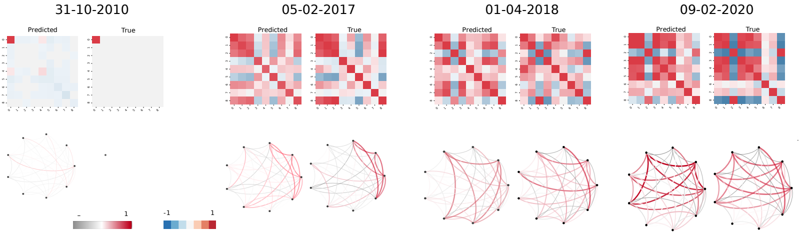
end

Theorem

The AAC algorithm converges to a local minimum in finite time.

Example: Cryptocurrency correlation networks

Data: Correlation networks for price in USD of Bitcoin, Dash, Digibyte, Dogecoin, Litecoin, Vertcoin, Stellar, Monero, Verge from July 18^{th} 2010 (first record of bitcoin) - until April 3^{rd} 2020, for 20-day time windows.
Task: Predict network from time



Example:
Public Transport and Covid-19 in Copenhagen, Denmark

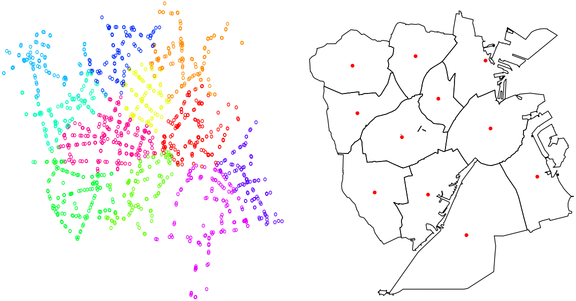


Figure: Bus stops in the different areas of Copenhagen and Frederiksberg.

Example:
Public Transport and Covid-19 in Copenhagen, Denmark

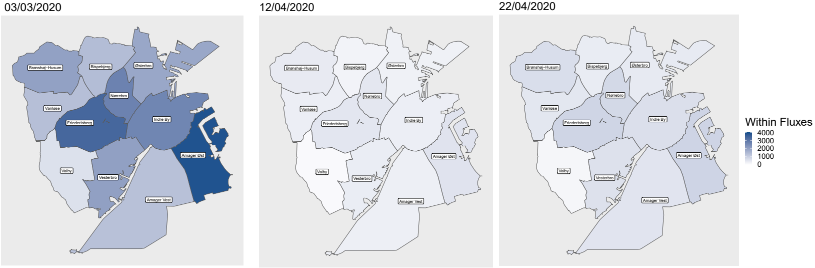


Figure: Prediction of the within area fluxes (i.e. the nodes attributes) of three days randomly sampled from the three periods: 03/03/2020, 12/04/2020, and 22/04/2020

Example:
Public Transport and Covid-19 in Copenhagen, Denmark

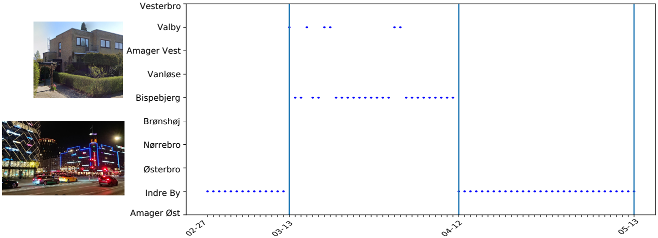


Figure: Most popular optimal matching with the node Indre By as a function time. Vertical Lines describe transitions between different phases of the lockdown.

Discussion

Advantages:

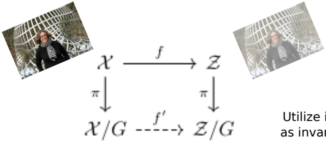
- Total space is $(\mathbb{R}^d)^{n\times n}$, easing generalization of Euclidean methods
- The AAC approach gives intrinsic statistics while retaining the computational advantages of Euclidean statistics

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.

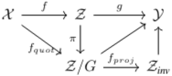
Graph-valued deep learning

Another avenue: Utilizing the empirical strengths of deep learning – equivariant models are actually operating on quotients such as graph space.
NB! Need to study representation in the quotient!



Equivariant model operates implicitly on quotient

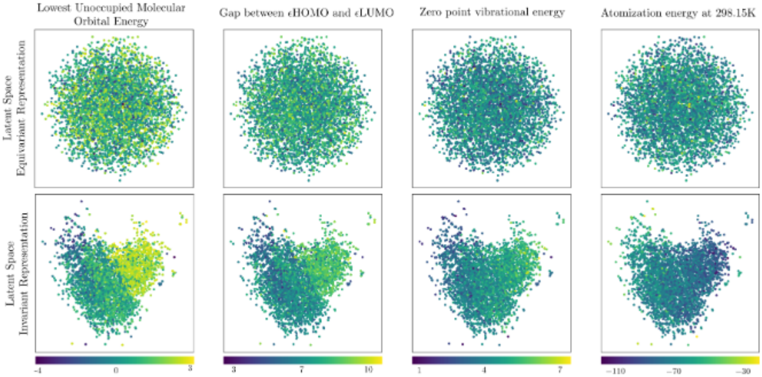
Utilize invariant projections as invariant representations



Hansen, Calissano, Feragen, Interpreting Equivariant Representations, 2023

Graph-valued deep learning

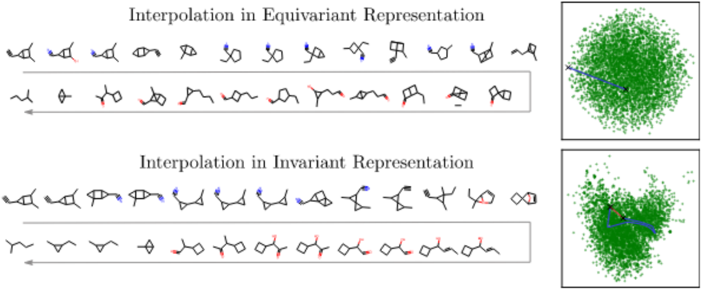
Permutation equivariant graph VAE – analysis and visualization of latent space via invariant representations



Hansen, Calissano, Feragen, Interpreting Equivariant Representations, 2023

Graph-valued deep learning

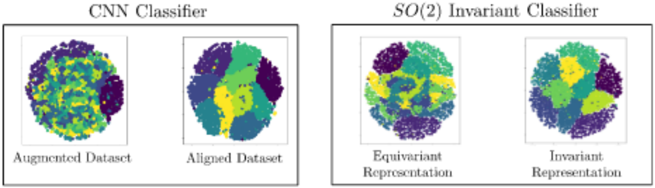
Permutation equivariant graph VAE – analysis and visualization of latent space



Hansen, Calissano, Feragen, Interpreting Equivariant Representations, 2023

Another example: Rotation action on images

Rotation invariant classifier on rotated MNIST with equivariant intermediate representations, or encouraged equivariance via augmentation – analysis and visualization of latent space



Hansen, Calissano, Feragen, Interpreting Equivariant Representations, 2023

Discussion

We have seen two families of methods:

- Classical geometric statistics intrinsic on a quotient space,
- Equivariant neural networks, secretly also working as intrinsic quotient space models
- Glimpse: Using equivariant representations without understanding the quotient space can get you into trouble
- Glimpse: Augmentation leads to similar empirical behavior

Open problems:

- Finding good invariant representations – because studying quotients remains hard
- Handling approximate equivariance encouraged by augmentation – esp when augmentations don't form a group