The geometry of synchronisation problems and group actions

Jacek Brodzki
Centre for Geometry, Topology, and Applications
University of Southampton

Análisis Topológico de Datos November 2018

Supported by the EPSRC: Joining the Dots: from data to insight EP/N014189/1

Plan

- Synchronisation in examples
- Spectral properties of graphs
- Spectral clustering of graphs
- The geometry and topology of synchronisation
- Learning group actions
- Persistence

Synchronisation in examples

Synchronisation

Synchronisation problems deal with questions regarding aligning a collection of objects in a consistent manner.

Given a sequence of objects $(o_1, ..., o_n)$ and a group (or collection of groups) G the objective is to learn a collection of group elements ρ_{ij} that would transform object o_i to o_j in a consistent way.

Synchronisation

Synchronisation problems deal with questions regarding aligning a collection of objects in a consistent manner.

Given a sequence of objects $(o_1, ..., o_n)$ and a group (or collection of groups) G the objective is to learn a collection of group elements ρ_{ij} that would transform object o_i to o_j in a consistent way.

We'd like to get as close as possible to a group action, so if $\rho_{ij} : o_i \mapsto o_j$ and $\rho_{jk} : o_j \mapsto o_k$ then

$$\rho_{ik} = \rho_{ij} \circ \rho_{jk} : o_i \mapsto o_k.$$

Synchronisation

Synchronisation problems deal with questions regarding aligning a collection of objects in a consistent manner.

Given a sequence of objects $(o_1, ..., o_n)$ and a group (or collection of groups) G the objective is to learn a collection of group elements ρ_{ij} that would transform object o_i to o_j in a consistent way.

We'd like to get as close as possible to a group action, so if $\rho_{ij} : o_i \mapsto o_j$ and $\rho_{jk} : o_j \mapsto o_k$ then

$$\rho_{ik} = \rho_{ij} \circ \rho_{jk} : o_i \mapsto o_k.$$

If it is possible to find a sequence of group elements that allow for an accurate transformation between objects then this set can be *synchronised*.

Examples of synchronisation

- Orientability of manifolds [Singer, Wu (2011)]: $G = O(1)$;
- Angular synchronisation [Singer (2011)]: $G = U(1)$;
- Vector diffusion maps [Singer, Wu (2012)]: $G = O(d)$;
- Collection Shape Matching [Nguyen et al. (2011)], [Huang, Guibas (2013)], [Chen et al. (2014)], [Maron et al. (2016)]: $G = \mathbb{E}^d$;
- Cryo-EM Structural Reconstruction [Singer et al. (2011)], [Shkolnisky, Singer (2012)], [Zhao, Singer (2014)], [Bandeira et al. (2015)]: $G = SO(3)$

Differentiable manifolds

A differentiable manifold M is a topological space equipped with an atlas $\{U_\alpha\}$ which is an open cover of M such that for every α there is a homeomorphism

$$\phi_\alpha : U_\alpha \to \mathbb{R}^n$$

which are such that for every nonempty overlap $U_\alpha \cap U_\beta$ the transition map

$$g_{\alpha\beta} = \phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cup U_\beta) \to \phi_\alpha(U_\alpha \cup U_\beta)$$

is a diffeomorphism of open subsets of $\mathbb{R}^n$.

Orientable manifolds

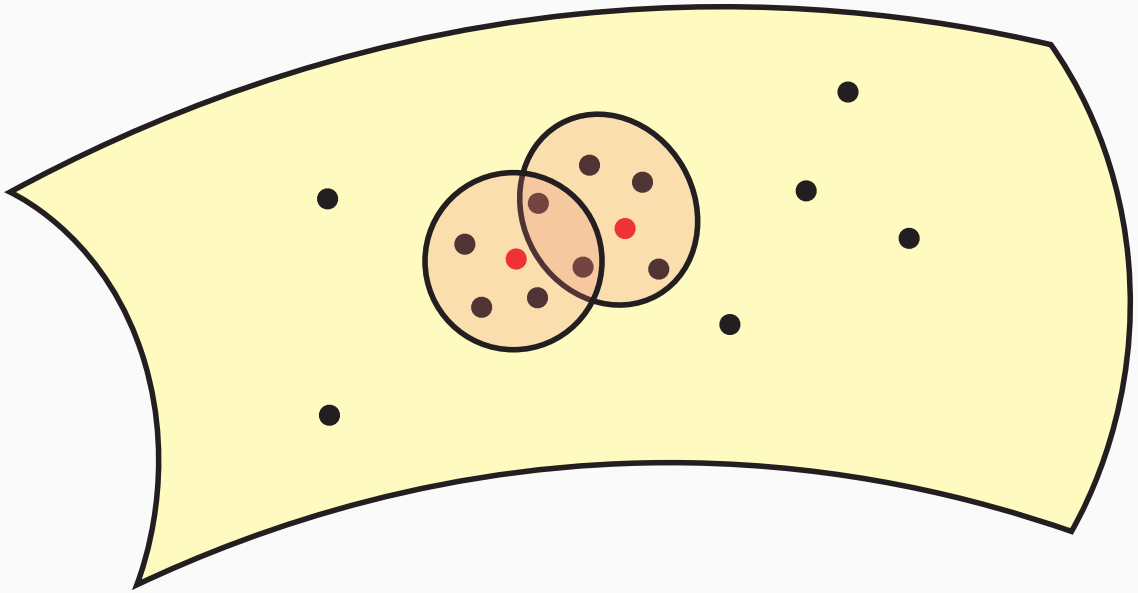
A manifold is orientable iff it admits an oriented atlas, i.e., an atlas $\{U_\alpha\}$ such that all the transition maps $g_{\alpha\beta}$ are orientable.

This in turn means that the Jacobian of each such map has an everywhere positive determinant.

Question:
How can one determine if a given manifold is orientable starting from a sample of points of M?

Orientable diffusion maps, Singer and Wu 2011

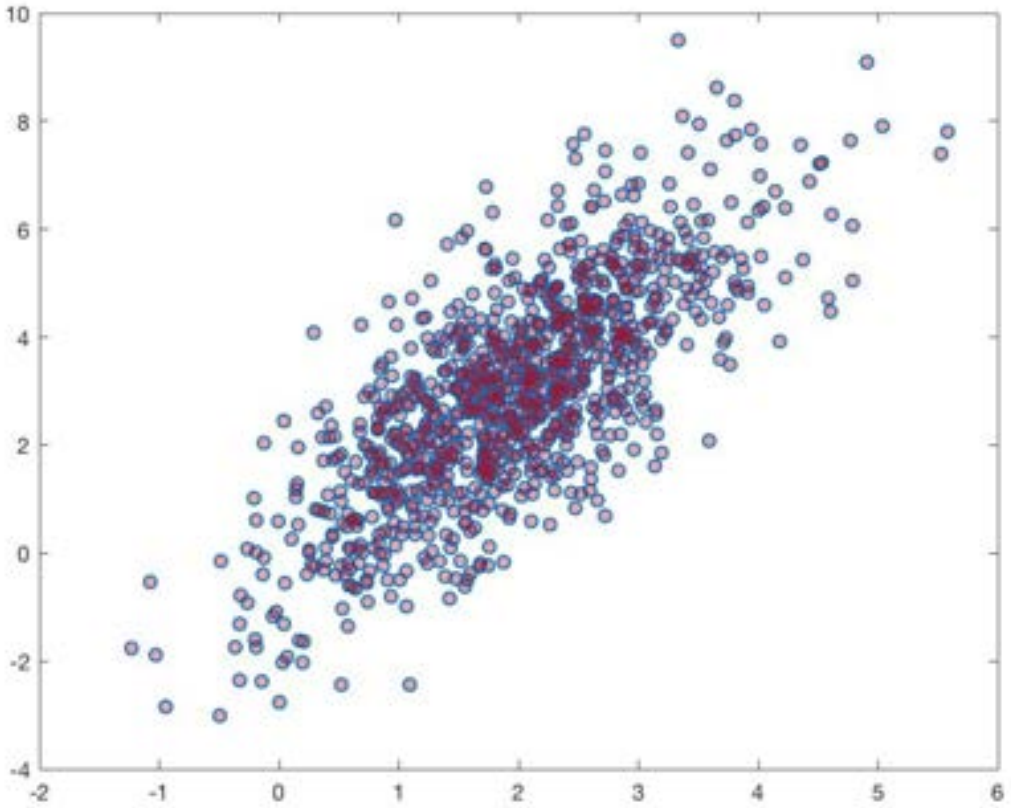
Assume given points $S = \{x_1, x_2, \ldots, x_n\}$ sampled from some $\mathbb{R}^p$; we assume that these points lie on a submanifold M of $\mathbb{R}^p$.



What is a good notion of tangent space in this case?

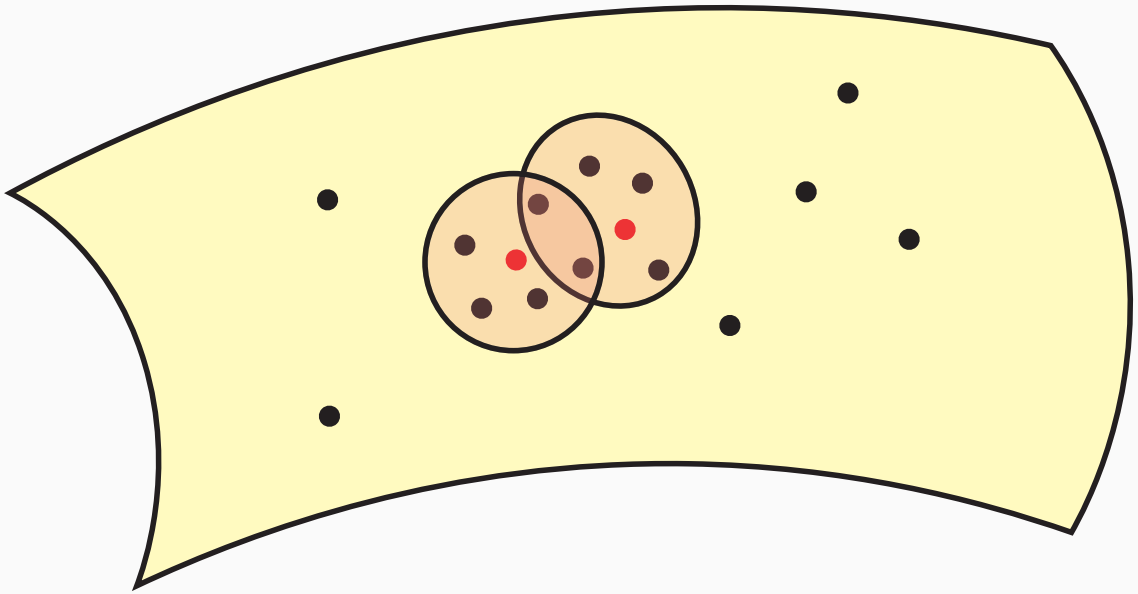
Principal Component Analysis PCA

- PCA was invented by Karl Pearson in 1901
- It provides a way of reducing the dimension of data, which is particularly useful when trying to visualise data in two or three dimensions.



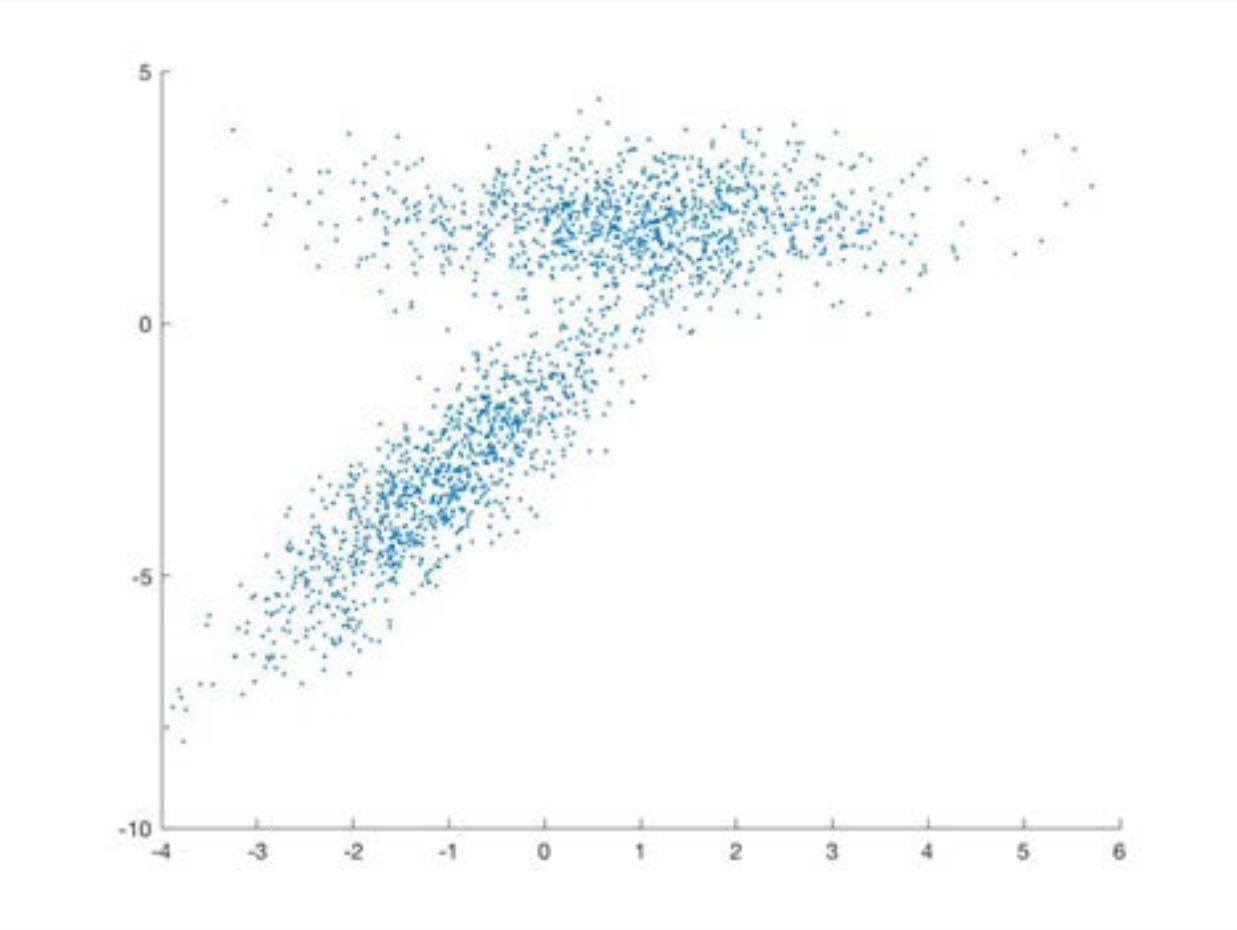
Local PCA

Choose $x_i \in S$ and identify its K nearest neighbours, where $K << n$. These points $x_{i_1}, \ldots, x_{i_K}$ are located near the tangent space $T_{x_i}M$. There may of course be deviations due to noise or curvature of M.



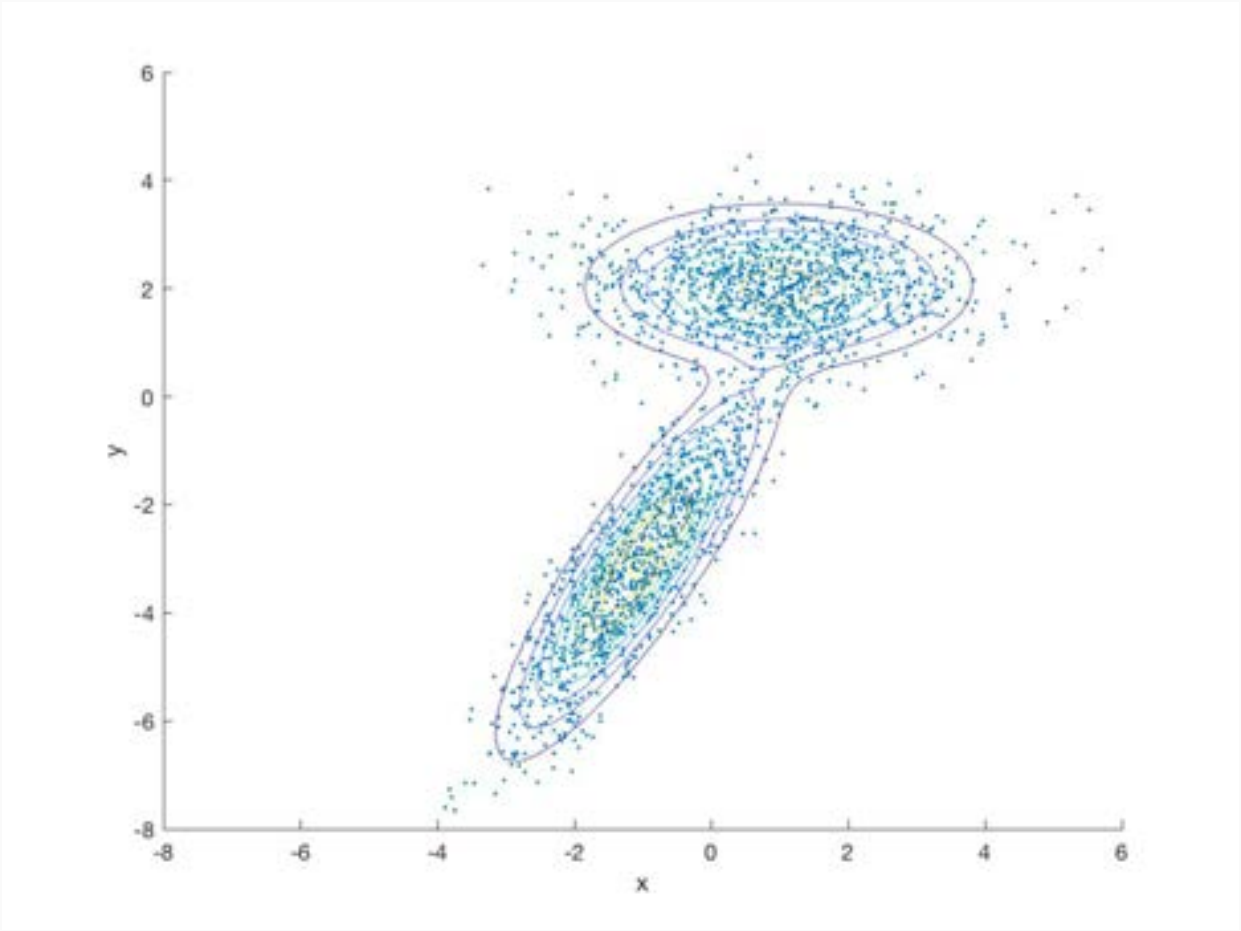
Local PCA

Choose $x_i \in S$ and identify its K nearest neighbours, where $K << n$. These points $x_{i_1}, \ldots, x_{i_K}$ are located near the tangent space $T_{x_i}M$. There may of course be deviations due to noise or curvature of M.



Local PCA

Choose $x_i \in S$ and identify its K nearest neighbours, where $K << n$. These points $x_{i_1}, \ldots, x_{i_K}$ are located near the tangent space $T_{x_i}M$. There may of course be deviations due to noise or curvature of M.



Local PCA

Let μ be the average of the nearest neighbours of x_i:

$$\mu = \frac{1}{K} \sum_{k=1}^{K} x_{i_k}$$

and form the matrix of vectors

$$X_i = [x_{i_1} - \mu \quad x_{i_2} - \mu \quad \dots \quad x_{i_K} - \mu].$$

So the k-th column is the vector $x_{i_k} - \mu$ and the matrix is of size $p \times K$.

Local PCA

$$X_i = [x_{i_1} - \mu \quad x_{i_2} - \mu \quad \dots \quad x_{i_K} - \mu].$$

Assume that the dimension of M from which the points x_i are sampled is d.

If the nearest points x_{i_k} to x_i are elements of $T_{x_i}M$ (more precisely: the vectors $x_{i_k} - x_i$ are elements of $T_{x_i}M$) then the matrix X will have d linearly independent columns, i.e., its rank is d.

This can be detected through the singular values of X_i.

Singular Value Decomposition

Assume A is a rectangular $m \times n$ matrix; $A \in \mathbb{C}^{m \times n}$, with $m \geq n$.

A singular value decomposition of A is

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^*$$

where

$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}, \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0,$$

and $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ are unitary. The numbers σ_k are called *singular values*.

The matrix U is a left singular vector matrix, and V is the right singular vector matrix.

Singular Value Decomposition

- A $m \times n$ matrix has $\min\{m, n\}$ singular values.

Singular Value Decomposition

- A $m \times n$ matrix has $\min\{m, n\}$ singular values.
- The singular values are unique, but the singular vector matrices are not.

Singular Value Decomposition

- A $m \times n$ matrix has $\min\{m, n\}$ singular values.
- The singular values are unique, but the singular vector matrices are not.
- If $A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^*$ is an SVD decomposition of A then

$$A^* = V(\Sigma \quad 0)U^*$$

is a SVD of A^*. So A and A^* have the same singular values.

Singular Value Decomposition

- A $m \times n$ matrix has $\min\{m, n\}$ singular values.
- The singular values are unique, but the singular vector matrices are not.
- If $A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^*$ is an SVD decomposition of A then

$$A^* = V(\Sigma \quad 0)U^*$$

  is a SVD of A^*. So A and A^* have the same singular values.
- A is nonsingular iff all its singular values are nonzero.

Singular Value Decomposition

- A $m \times n$ matrix has $\min\{m, n\}$ singular values.
- The singular values are unique, but the singular vector matrices are not.
- If $A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^*$ is an SVD decomposition of A then

$$A^* = V(\Sigma \quad 0)U^*$$

  is a SVD of A^*. So A and A^* have the same singular values.
- A is nonsingular iff all its singular values are nonzero.
- The rank of a matrix is the number of its nonzero singular values.

Singular Value Decomposition

- A $m \times n$ matrix has $\min\{m, n\}$ singular values.
- The singular values are unique, but the singular vector matrices are not.
- If $A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^*$ is an SVD decomposition of A then

$$A^* = V(\Sigma \quad 0)U^*$$

  is a SVD of A^*. So A and A^* have the same singular values.
- A is nonsingular iff all its singular values are nonzero.
- The rank of a matrix is the number of its nonzero singular values.
- Vectors of U corresponding to the nonzero singular values span the column space of A (the image of A regarded as a linear operator).

SVD example

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} -0.1525 & -0.8226 & -0.3945 & -0.3800 \\ -0.3499 & -0.4214 & 0.2428 & 0.8007 \\ -0.5474 & -0.0201 & 0.6979 & -0.4614 \\ -0.7448 & 0.3812 & -0.5462 & 0.0407 \end{pmatrix}$$

$$\times \begin{pmatrix} 14.2691 & 0 \\ 0 & 0.6268 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\times \begin{pmatrix} -0.6414 & 0.7672 \\ -0.7672 & -0.6414 \end{pmatrix}$$

Estimating the dimension

The number of nonzero singular values $\sigma_{i,k}$ of X_i can be taken to be the dimension of $T_{x_i}(M)$. This however, is not quite accurate because of noise and curvature of M.

Estimating the dimension

The number of nonzero singular values $\sigma_{i,k}$ of X_i can be taken to be the dimension of $T_{x_i}(M)$. This however, is not quite accurate because of noise and curvature of M.

Choose a threshold $\gamma \in [0, 1]$. In practice γ is closer to 1 than 0.

We estimate the dimension d_i of $T_{x_i}M$ as the smallest integer d_i for which

$$\frac{\sum_{k=1}^{d_i} \sigma_{i,k}^2}{\sum_{k=1}^{K} \sigma_{i,k}^2} > \gamma.$$

The intuition here is that we want to consider the singular values that account for a high enough percentage of the variability of the data.

Estimating the dimension

The number of nonzero singular values $\sigma_{i,k}$ of X_i can be taken to be the dimension of $T_{x_i}(M)$. This however, is not quite accurate because of noise and curvature of M.

Choose a threshold $\gamma \in [0, 1]$. In practice γ is closer to 1 than 0.

We estimate the dimension d_i of $T_{x_i}M$ as the smallest integer d_i for which

$$\frac{\sum_{k=1}^{d_i} \sigma_{i,k}^2}{\sum_{k=1}^{K} \sigma_{i,k}^2} > \gamma.$$

The intuition here is that we want to consider the singular values that account for a high enough percentage of the variability of the data.

E.g., if $\gamma = 0.9$ then d_i singular values account for at least 90% of the variability of the data, but $d_i - 1$ singular values will account for less than 90%.

Estimating the dimension

Having estimated the local dimension d_i for all $i = 1, \ldots, n$ we need an estimate of the global dimension. Possible options are:

The mean

$$d = \frac{1}{n} \sum_{i=1}^{n} d_i$$

rounding to the nearest integer. This minimises the sum of squared errors

$$\sum_{i=1}^{n} (d_i - d)^2.$$

Estimating the dimension

Having estimated the local dimension d_i for all $i = 1, \ldots, n$ we need an estimate of the global dimension. Possible options are:

The median

$$d = \text{median}\,(d_1, d_2, \ldots, d_n).$$

This minimises the ℓ^1-error

$$\sum_{i=1}^{n} |d_i - d|.$$

This is more robust with respect to outliers than the mean estimator.

Local coordinate frame

Consider again the matrix describing the shifted local neighbourhood of x_i:

$$X_i = [x_{i_1} - \mu \quad x_{i_2} - \mu \quad \dots \quad x_{i_K} - \mu].$$

Assume that the matrix X_i has an SVD decomposition

$$X_i = U_i \Sigma_i V_i^t,$$

where the columns of the $p \times K$ matrix U_i (the left singular vectors) are orthonormal.

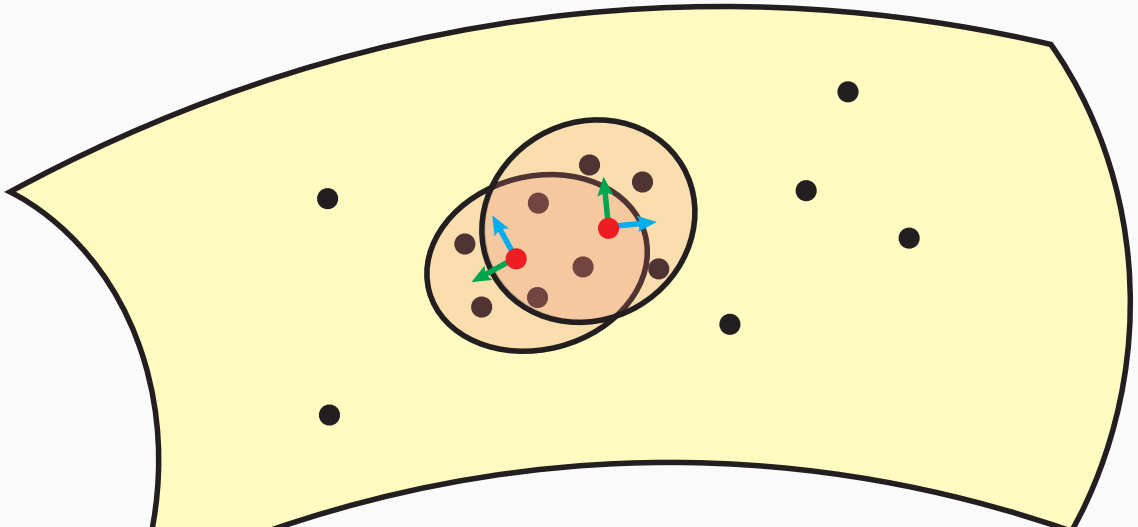
Local coordinate frame

Define the matrix O_i by

$$O_i = [u_{i_1} u_{i_2} \dots u_{i_d}]$$

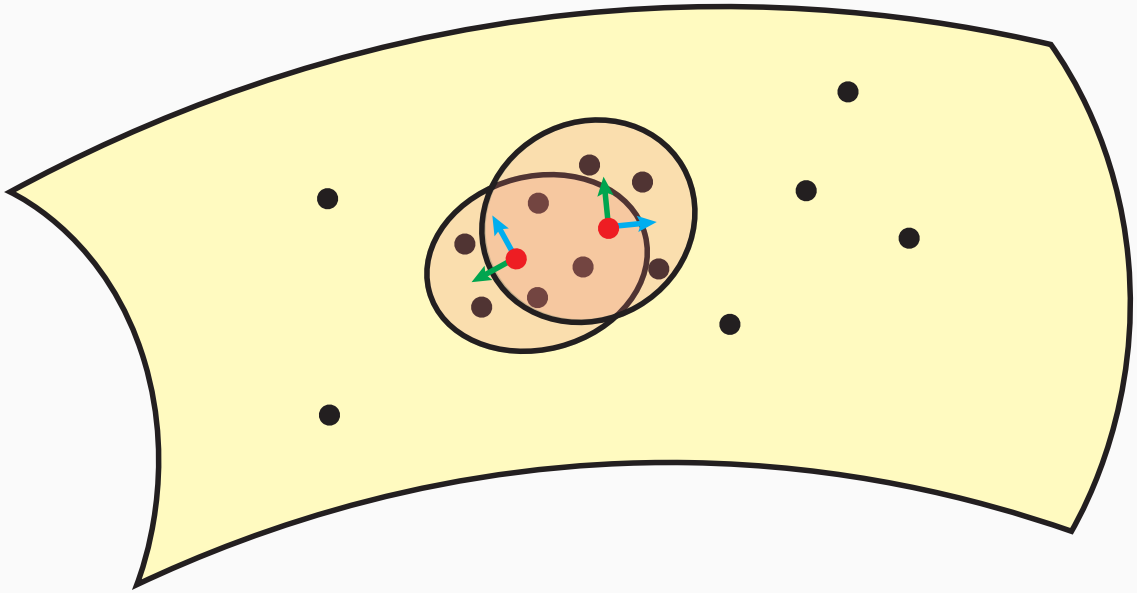
where the column vectors u_{i_j} are the first d left singular vectors of the matrix X_i (arranged in the decreasing order of the singular values)

The columns of O_i provide an approximation of an orthonormal basis of the tangent space $T_{x_i}(M)$.



Alignment

Now consider two nearby points x_i and x_j which which are each other's nearest neighbours. We have identified their local frames, can we map one to the other?



Alignment

We have found matrices O_i and O_j approximating the tangent spaces at each point. Because of possible noise or curvature effects, the matrix $O_i^T O_j$ need not be orthogonal, and we define O_{ij} to be an orthogonal matrix which is closest to $O_i^T O_j$:

$$O_{ij} = \text{argmin} \, \|O - O_i^T O_j\|$$

where the minimum is taken over all orthogonal matrices $O \in O(d)$.

Alignment

We have found matrices O_i and O_j approximating the tangent spaces at each point. Because of possible noise or curvature effects, the matrix $O_i^T O_j$ need not be orthogonal, and we define O_{ij} to be an orthogonal matrix which is closest to $O_i^T O_j$:

$$O_{ij} = \text{argmin} \, \|O - O_i^T O_j\|$$

where the minimum is taken over all orthogonal matrices $O \in O(d)$.

This has the following solution.

If the SVD of $O_i^T O_j$ is

$$O_i^T O_j = U \Sigma V^T$$

then

$$O_{ij} = U V^T.$$

Alignment

The process of finding optimal orthogonal transformations between bases is called *alignment*.

Given that $O_{ij} \in O(d)$, we have that $\det O_{ij} = \pm 1$ where

- $\det O_{ij} = 1$ iff $O_{ij} \in SO(d)$, so it is a 'rotation'
- $\det O_{ij} = -1$ iff O_{ij} is a composition of a rotation and a reflection.

Alignment

The process of finding optimal orthogonal transformations between bases is called *alignment*.

Given that $O_{ij} \in O(d)$, we have that $\det O_{ij} = \pm 1$ where

- $\det O_{ij} = 1$ iff $O_{ij} \in SO(d)$, so it is a 'rotation'
- $\det O_{ij} = -1$ iff O_{ij} is a composition of a rotation and a reflection.

Note that this only works between nearby points, and the outcome can be encoded as a graph $\mathcal{G} = (V, E)$ with vertices corresponding to the n data points, and where an edge (i, j) exists iff there is a corresponding matrix O_{ij}.

Synchronisation

We can encode the results of our investigation so far by defining a matrix

$$Z_{ij} = \begin{cases} \det O_{ij}, & \text{if } (i,j) \in E \\ 0, & \text{otherwise} \end{cases}$$

For each point x_i we have a basis defined by local *PCA* as well as a basis given by the ambient manifold.

Define:

$$z_i = \begin{cases} 1, & \text{if the orientations agree} \\ -1, & \text{otherwise.} \end{cases}$$

Clearly,

$$z_i z_j^{-1} = Z_{ij}.$$

Synchronisation

In practice, because of the curvature and noise, the relation

$$z_i z_j^{-1} = Z_{ij}.$$

will not necessarily hold for all edges.

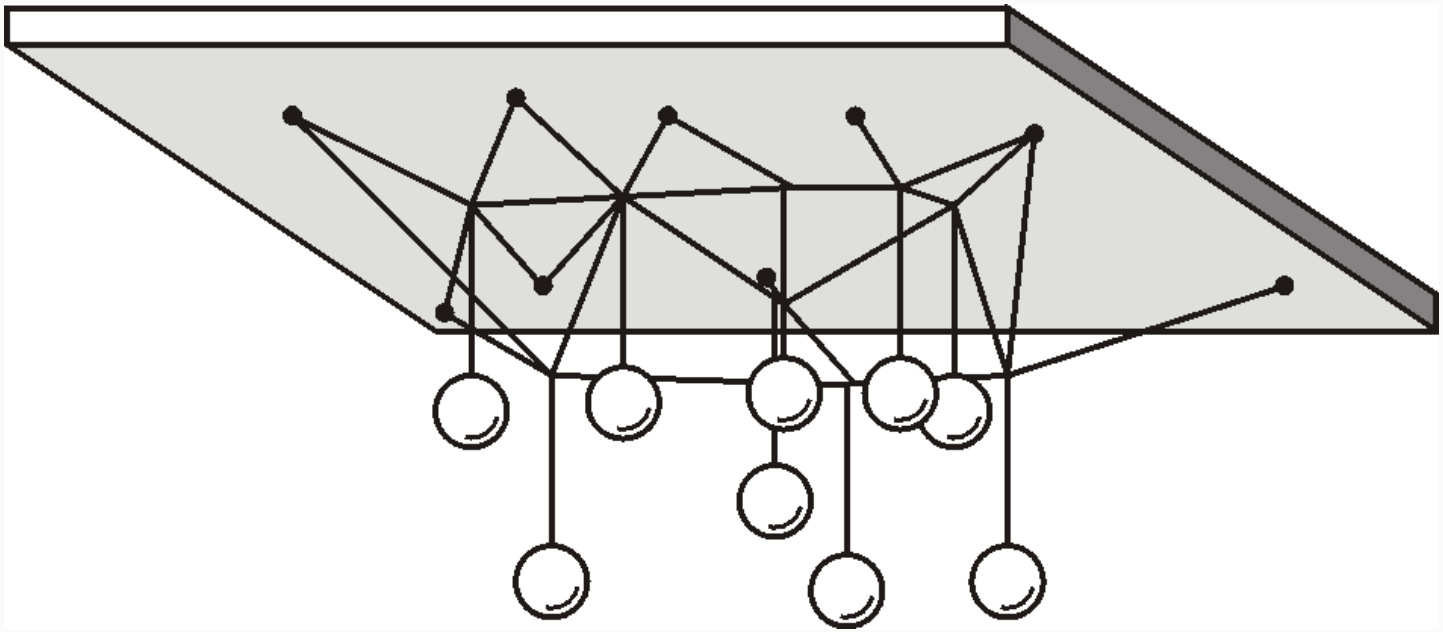
Synchronisation in this case refers to finding the function z on the vertices (z takes values in $\mathbb{Z}_2$) for which this equation holds. When the manifold is orientable, this solution exists, while for nonorientable manifolds we will find two classes of frames connected by matrices of determinant 1.

Angular synchronisation

Problem

Find an accurate estimation of a set of unknown angles $\theta_1, \ldots, \theta_n$ from m noisy measurements of the differences $\theta_i - \theta_j$.

A main motivation for this problem comes from systems of coupled oscillators. This is relevant, e.g., in power grids where generators on the network need to be synchronised very precisely.



Angular synchronisation

Other applications include

- Time synchronisation of distributed networks given the differences $t_i - t_j$ from which we need to determine the times $t_1, \ldots, t_n \in \mathbb{R}$.
- Surface reconstruction in computer vision and optics, where the surface is reconstructed from noisy measurements of the gradient of the surface.

Angular synchronisation

Main problem: given the set of differences $\theta_i - \theta_j$ (call them offsets), we do not know which of them are reliable, and which are noise.

So there are two kinds of edges in the synchronisation graph $\mathcal{G}$:

$$\begin{aligned} \delta_{ij} &= \theta_i - \theta_j && (i,j) \in E_{\text{good}} \\ \delta_{ij} &\sim \text{Uniform}([0, 2\pi)) && (i,j) \in E_{\text{bad}}. \end{aligned}$$

Angular synchronisation

Main problem: given the set of differences $\theta_i - \theta_j$ (call them offsets), we do not know which of them are reliable, and which are noise.

So there are two kinds of edges in the synchronisation graph $\mathcal{G}$:

$$\begin{aligned} \delta_{ij} &= \theta_i - \theta_j && (i,j) \in E_{\text{good}} \\ \delta_{ij} &\sim \text{Uniform}([0, 2\pi)) && (i,j) \in E_{\text{bad}}. \end{aligned}$$

We note that cycles in the graph $\mathcal{G}$ satisfy the following consistency relation. For a triangle of 'good' edges (i,j), (j,k), and (k,i) we have

$$\delta_{ij} + \delta_{jk} + \delta_{ki} = 0 \qquad \text{mod } 2\pi.$$

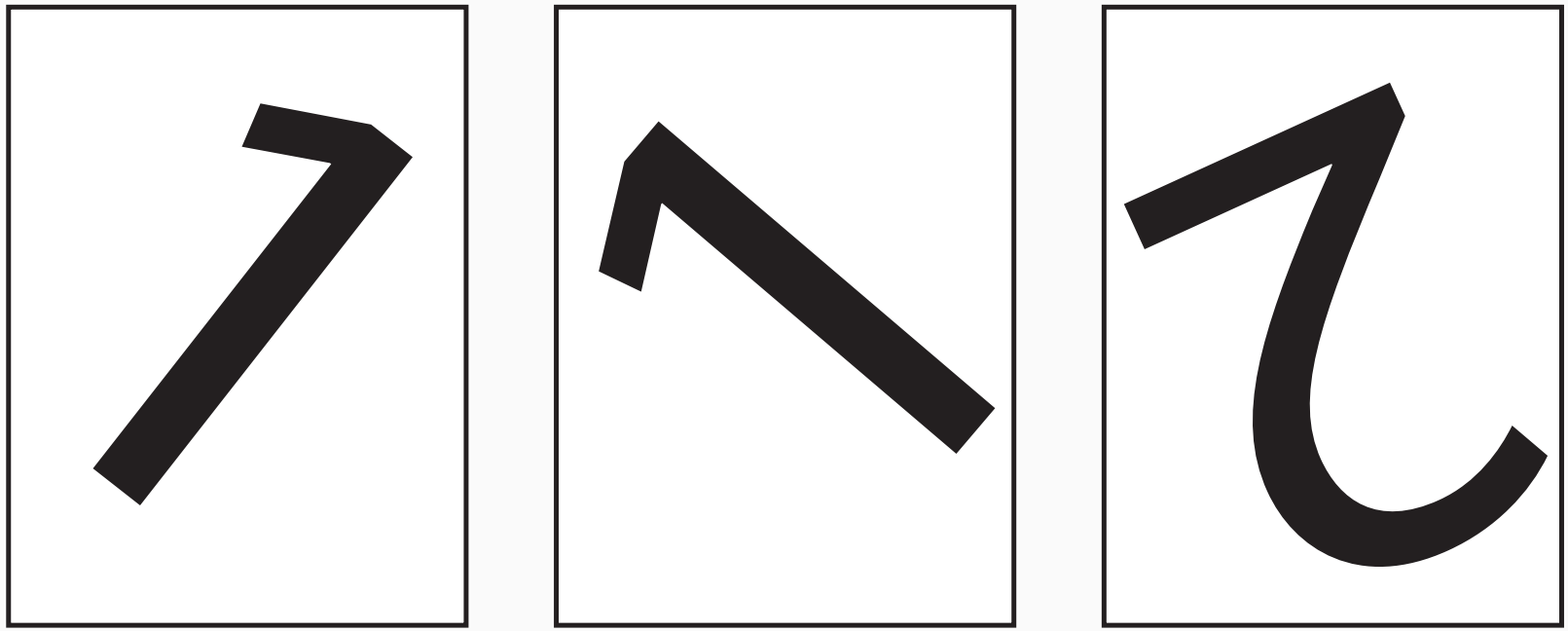
This will be quite important in later lectures.

Measuring similarity through orthogonal transformations

The following new framework emerges from our examples:

- Similarity between data points is described by a graph of relationships $\mathcal{G}$;
- If two points are similar enough, they are connected by and edge with a weight that quantifies their affinity.
- Many examples show that apart from a weight, it may be useful to have an orthogonal transformation attached to an edge.

Image classification



To each pair we would like to assign an orthogonal transformation O_{ij} that best aligns the images and a weight w_{ij} describing their measure of affinity.

Vector diffusion maps (Singer, Wu 2012)

Given a graph $\mathcal{G} = (V, E)$ of relationships define

$$S_{i,j} = \begin{cases} w_{ij} O_{ij}, & (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

where $O_{ij} \in O(d)$, $0 \in O(d)$. We assume $w_{ij} = w_{ji}$ and $O_{ji} = O_{ij}^T$.

Vector diffusion maps (Singer, Wu 2012)

Given a graph $\mathcal{G} = (V, E)$ of relationships define

$$S_{i,j} = \begin{cases} w_{ij} O_{ij}, & (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

where $O_{ij} \in O(d)$, $0 \in O(d)$. We assume $w_{ij} = w_{ji}$ and $O_{ji} = O_{ij}^T$.

As previously, we assume that we have identified tangent spaces at data points in V and want to create a suitable 'parallel transport' operator to compare the tangent spaces at nearby points. These are called *diffusion maps*.

Diffusion maps

Vertex degree:

$$\deg(i) = \sum_{j \sim i} w_{ij}.$$

$$D(i,i) = \deg(i) I_{d \times d}.$$

Now for $x_i \in V$ consider a vector $v(i) \in \mathbb{R}^d$ regarded as an element of $\mathbb{R}^d$.

'Divergence'

$$(D^{-1} S v)(i) = \frac{1}{\deg(i)} \sum_{j \sim i} w_{ij} O_{ij} v(j).$$

Affinity

To encode affinity between vertices i and j (we will frequently simplify the notation x_i and x_j for vertices) consider a path γ of length t connecting i and j, e.g., a sequence of consecutive edges of length t, and take

$$O(\gamma) = O_{i=i_0,i_1} O_{i_1 i_2} \dots O_{i_{t-1},i_t=j}$$

This orthogonal transformation will a priori depend on the path γ, and we will return to this point later.

A main point of the VDM approach is the characterisation of the size of this orthogonal transformation.

Affinity estimation

Recall that we have defined a $d \times d$ matrix

$$S_{i,j} = \begin{cases} w_{ij} O_{ij}, & (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

where $O_{ij} \in O(d)$, $0 \in O(d)$.

Affinity is a measure of compatibility between transformations obtained by following different paths from i to j.

This is measured using the matrix

$$\widetilde{S} = D^{-1/2} S D^{1/2}$$

which is similar to the matrix $D^{-1}S$ introduced earlier.

Affinity

The affinity between vertices i and j is defined as follows. For a natural number t consider all paths from i to j of length $2t$, and put:

$$A_t(i,j) = \|\widetilde{S}^{2t}(i,j)\|^2_{HS}$$

where $\|-\|$ is the Hilbert-Schmidt norm of this matrix:

$$\|T\|_{HS} = \left(\sum_{i,j} |T_{ij}|^2\right)^{1/2}$$

Affinity

The matrix $\widetilde{S}$ is symmetric, so it has a complete set of eigenvectors $v_1, \ldots, v_{nd}$ and eigenvalues $\lambda_1, \ldots, \lambda_{nd}$.

Then

$$\widetilde{S}(i,j) = \sum_{l=1}^{nd} \lambda_l v_l(i) v_l(j)^T; \quad \widetilde{S}^{2t}(i,j) = \sum_{l=1}^{nd} \lambda_l^{2t} v_l(i) v_l(j)^T$$

where $v(i) \in \mathbb{R}^d$.

Affinity

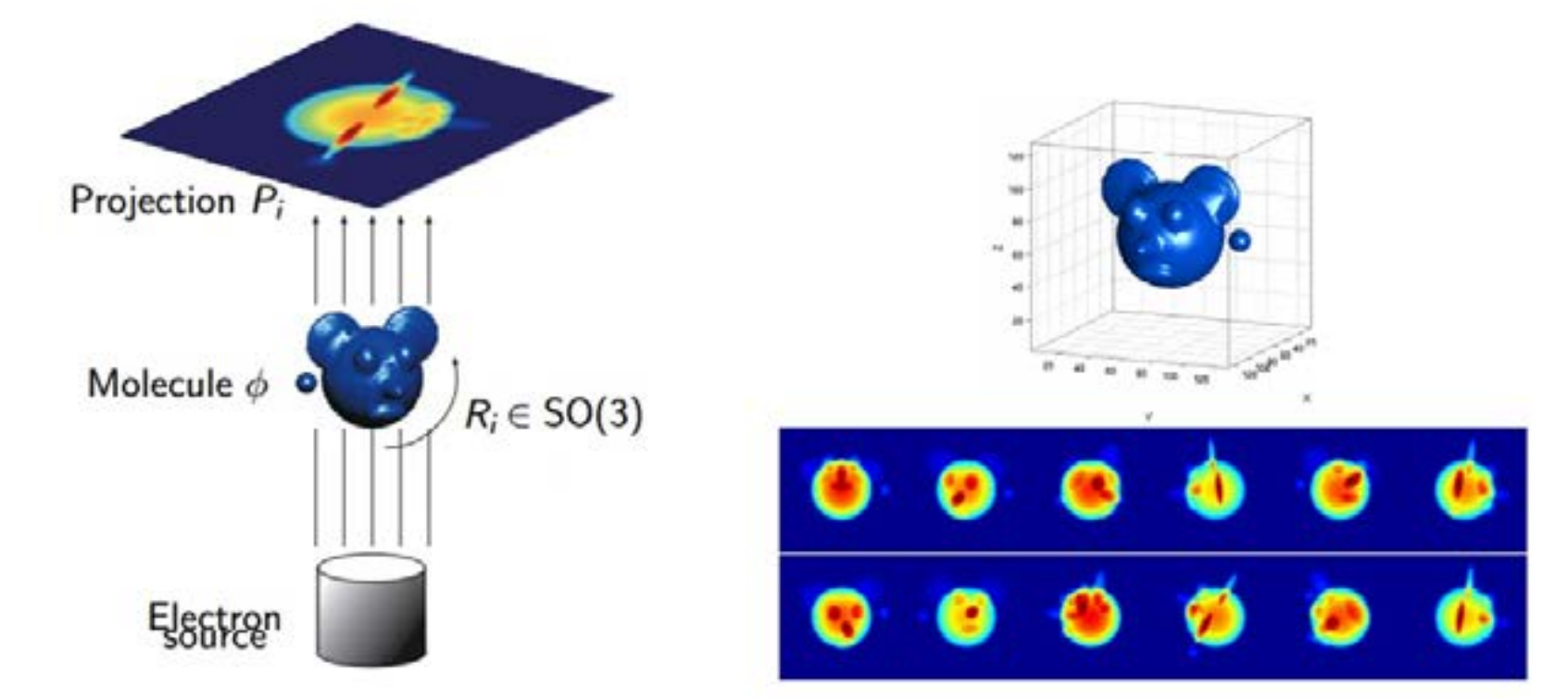
$$\begin{aligned}\|\widetilde{S}^{2t}(i,j)\|_{HS}^2 &= \mathrm{Tr}(\widetilde{S}^{2t}(i,j)\widetilde{S}^{2t}(i,j)^T) \\ &= \sum_{l,r=1}^{nd} (\lambda_l\lambda_r)^{2t}\langle v_l(i), v_r(i)\rangle\langle v_l(j), v_r(j)\rangle \\ &= \langle V_t(i), V_t(j)\rangle\end{aligned}$$

where

$$V_t : i \mapsto (\lambda_l\lambda_r)^{2t}\langle v_l(i), v_r(i)\rangle$$

and $l, r = 1, \ldots, nd$.

Cryoelectron microscopy



Picture from Picture from Singer, Shkolnitsky via Baideira *Ten lectures and forty two problems . . .*

Cryoelectron microscopy

The pixel intensity of the projection is described as

$$P_i(x, y) = \int_{-\infty}^{\infty} \phi_i(x, y, z) dz$$

where z is the direction of the imaging electrons. Here ϕ_i is defined as follows.

Let $\phi(x, y, z)$ be the electric potential of the 'laboratory molecule', then

$$\phi_i(x, y, z) = \phi(R_i^{-1}(x, y, z)),$$

where R_i is the $SO(3)$ matrix moving the laboratory molecule to its position when it is X-rayed.

Problem: Find the rotation matrices R_i.

Cryoelectron microscopy

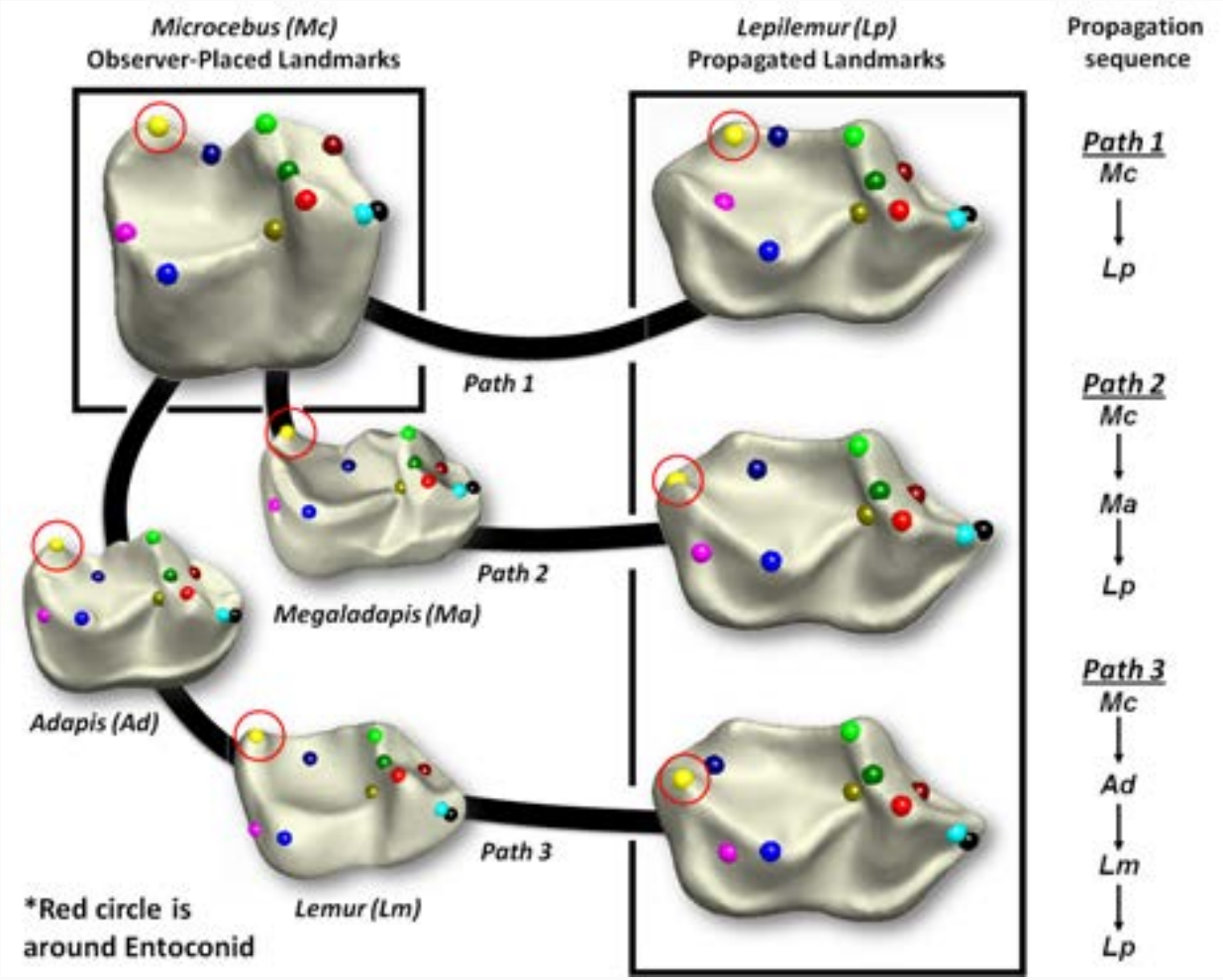
Note that for two molecules i and j the element of $SO(3)$ given by

$$\rho(i,j) = R_j R_i^{-1}$$

represents a measure of similarity between two projections.

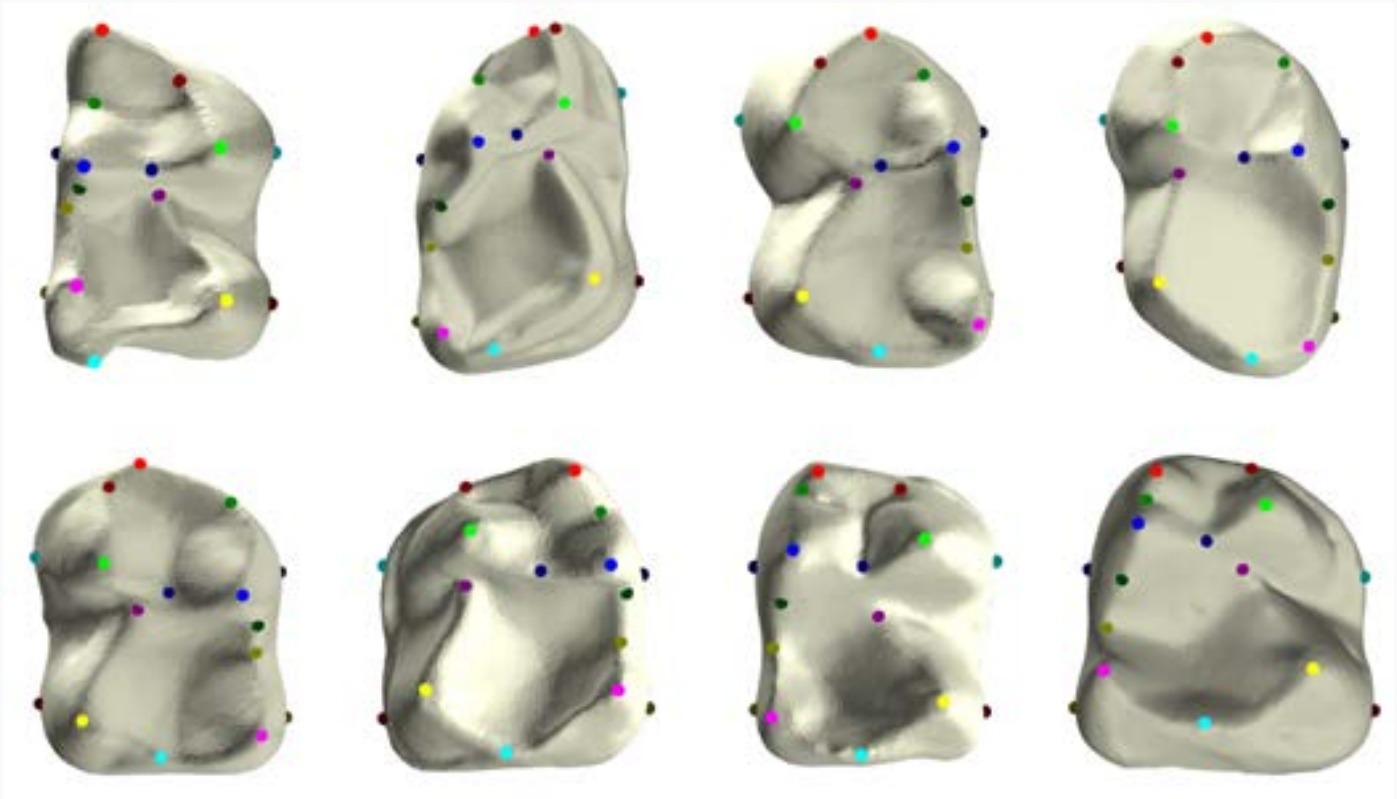
A different version of the same question: can we work out the necessary group actions $\rho(i,j)$?

Shape transport



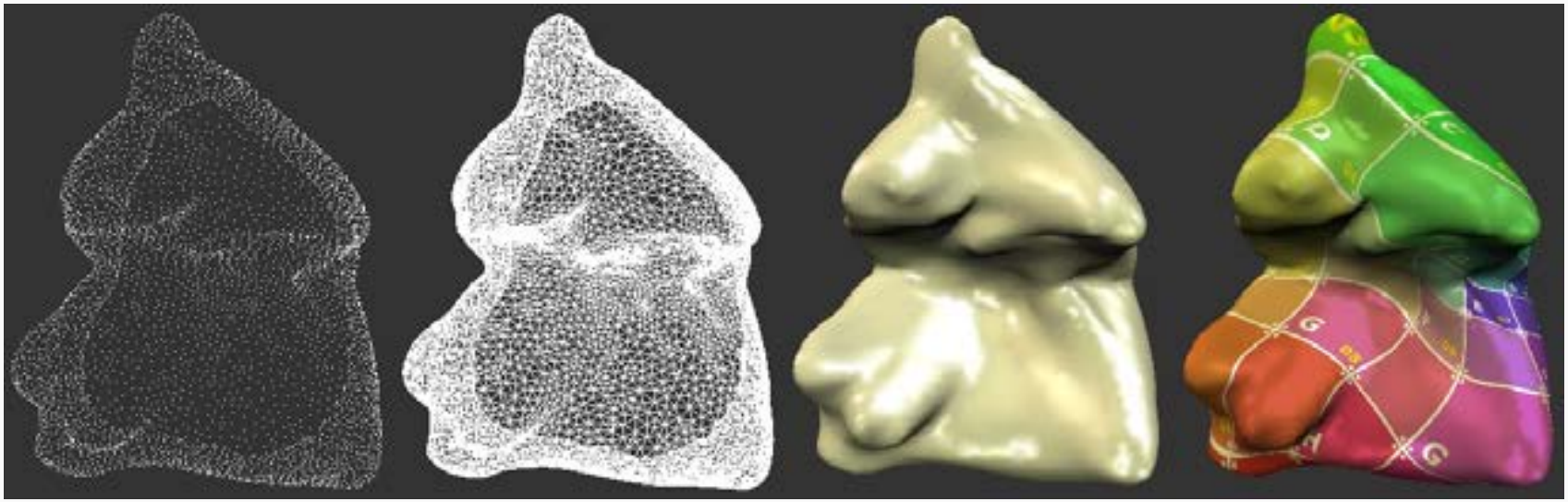
Picture from Sayan Muhkerjee

Object alignment



Picture from Sayan Muhkerjee

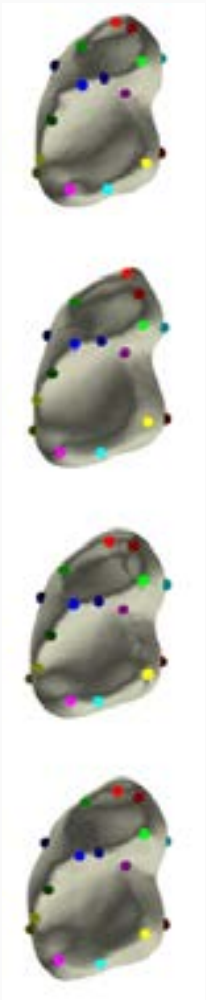
Data Acquisition: microCT (High Resolution X-ray CT)

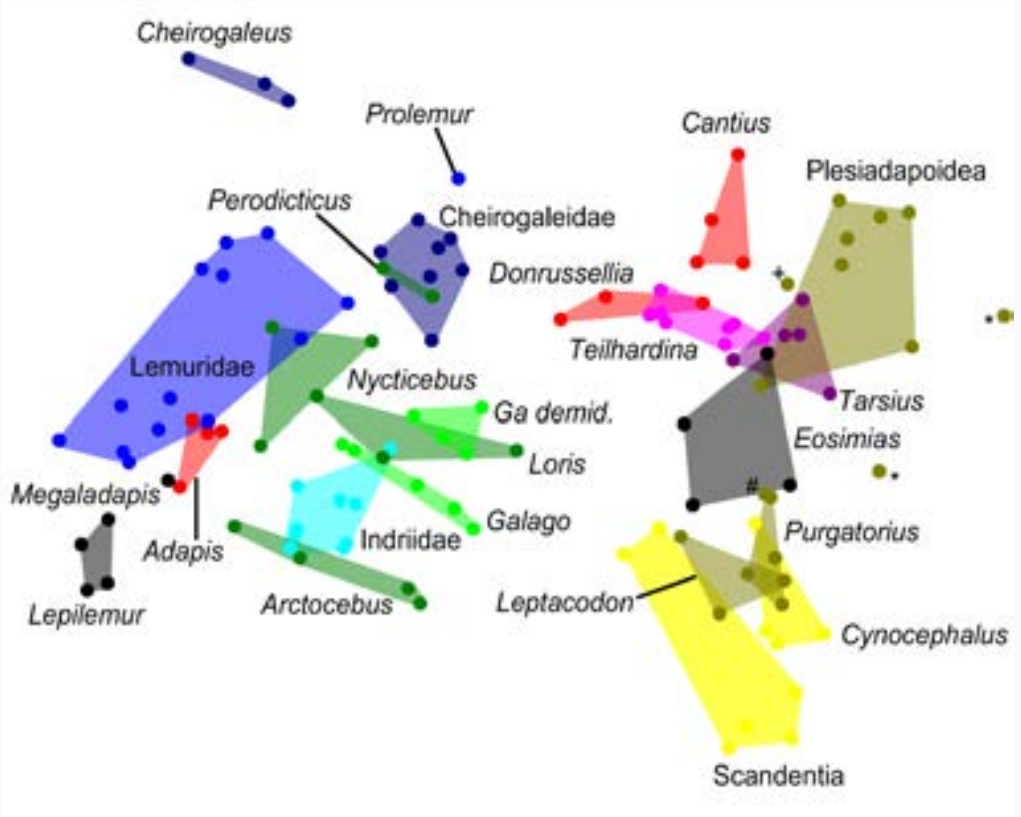


Surface reconstructed from μCT-scanned voxel data

Landmarked Teeth $\longrightarrow$

$$d^2_{Procrustes}(S_1, S_2) = \min_{R \text{ rigid motion}} \frac{1}{k} \sum_{j=1}^{k} \|R(x_j) - y_j\|^2$$





Boyer et al. "Algorithms to Automatically Quantify the Geometric Similarity of Anatomical Surfaces." *Proceedings of the National Academy of Sciences* 108.45 (2011): 18221-18226.

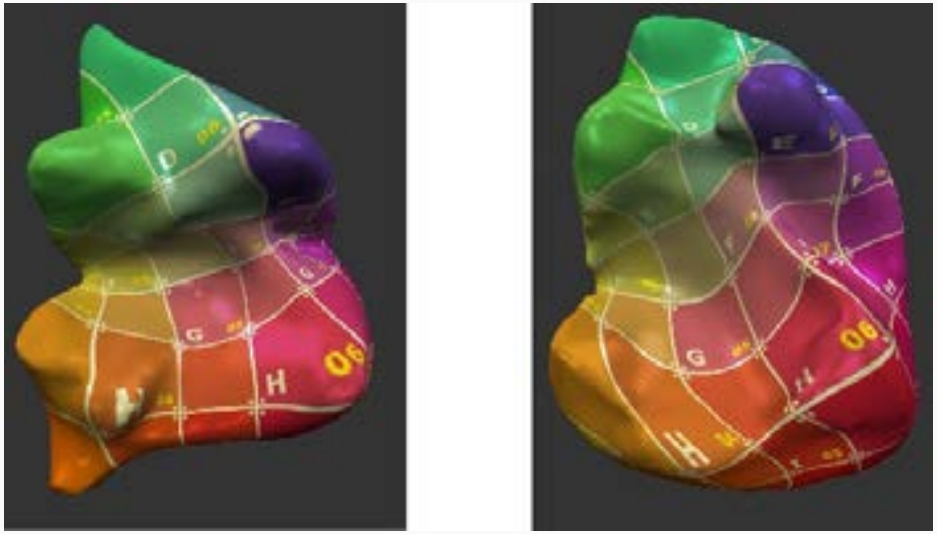
Myriads of Shape Distances...

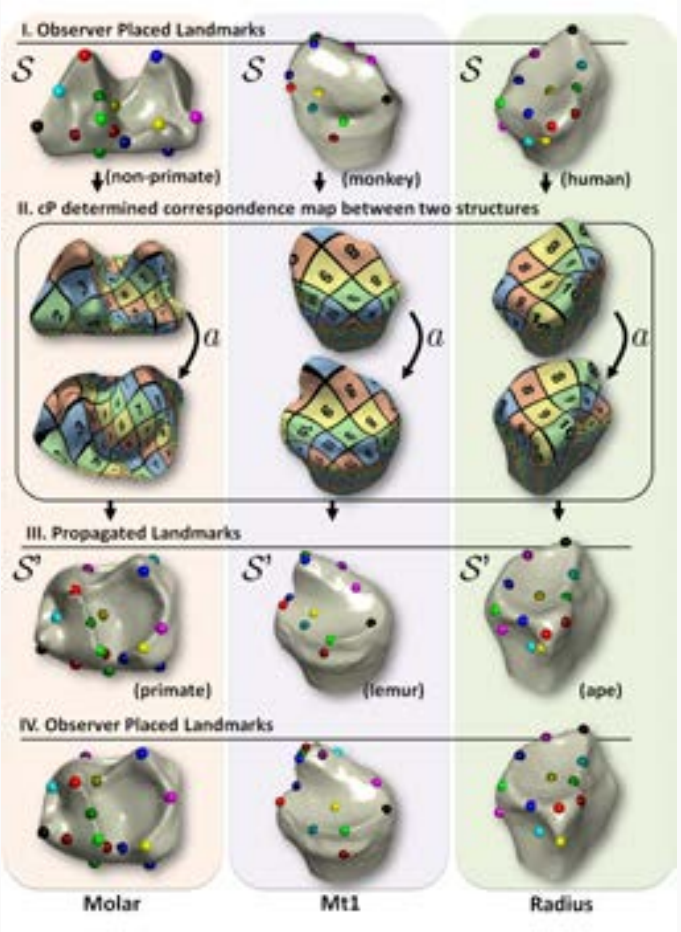
$d_{\mathrm{cWn}}(S_1, S_2)$: Conformal Wasserstein Distance (CWD)

$d_{\mathrm{cP}}(S_1, S_2)$: Continuous Procrustes Distance (CPD)

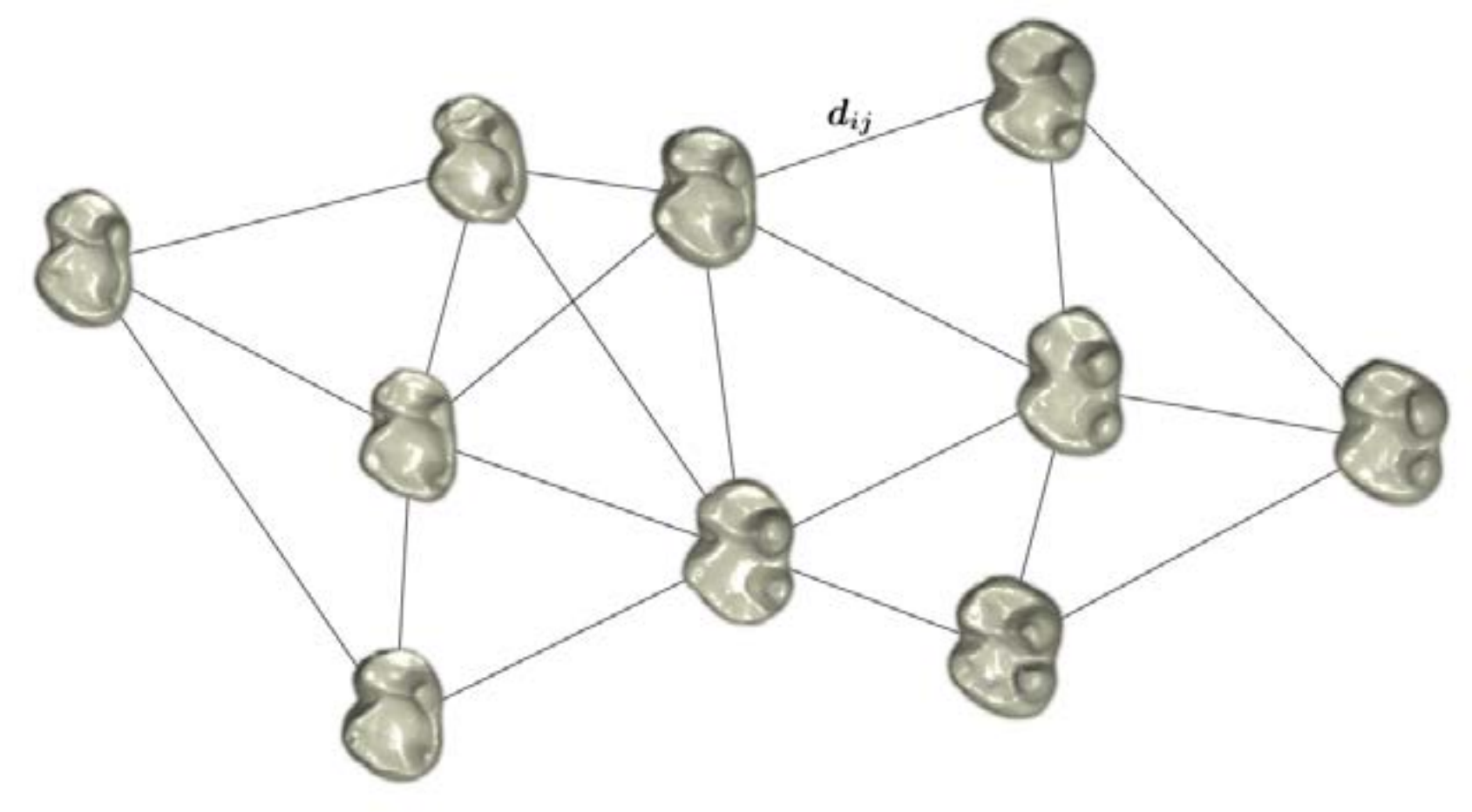
$d_{\mathrm{cKP}}(S_1, S_2)$: Continuous Kantorovich-Procrustes Distance (CKPD)

$$d_{\mathrm{cP}}(S_1, S_2) = \inf_{C \in \mathcal{A}(S_1, S_2)} \inf_{R \in \mathbb{E}(3)} \left(\int_{S_1} \| R(x) - C(x) \|^2 \, d\mathrm{vol}_{S_1}(x) \right)^{\frac{1}{2}}$$



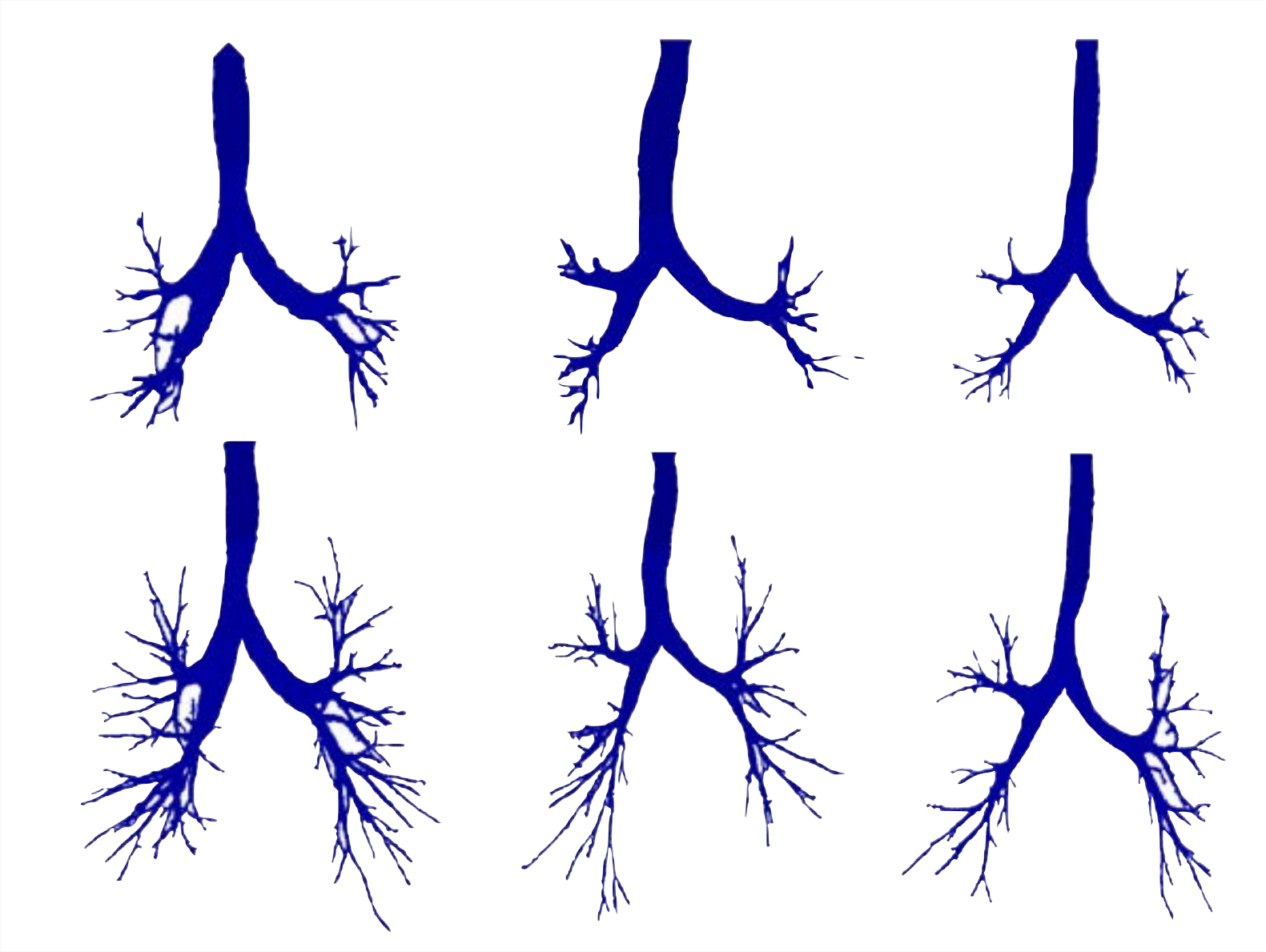


Shape Space



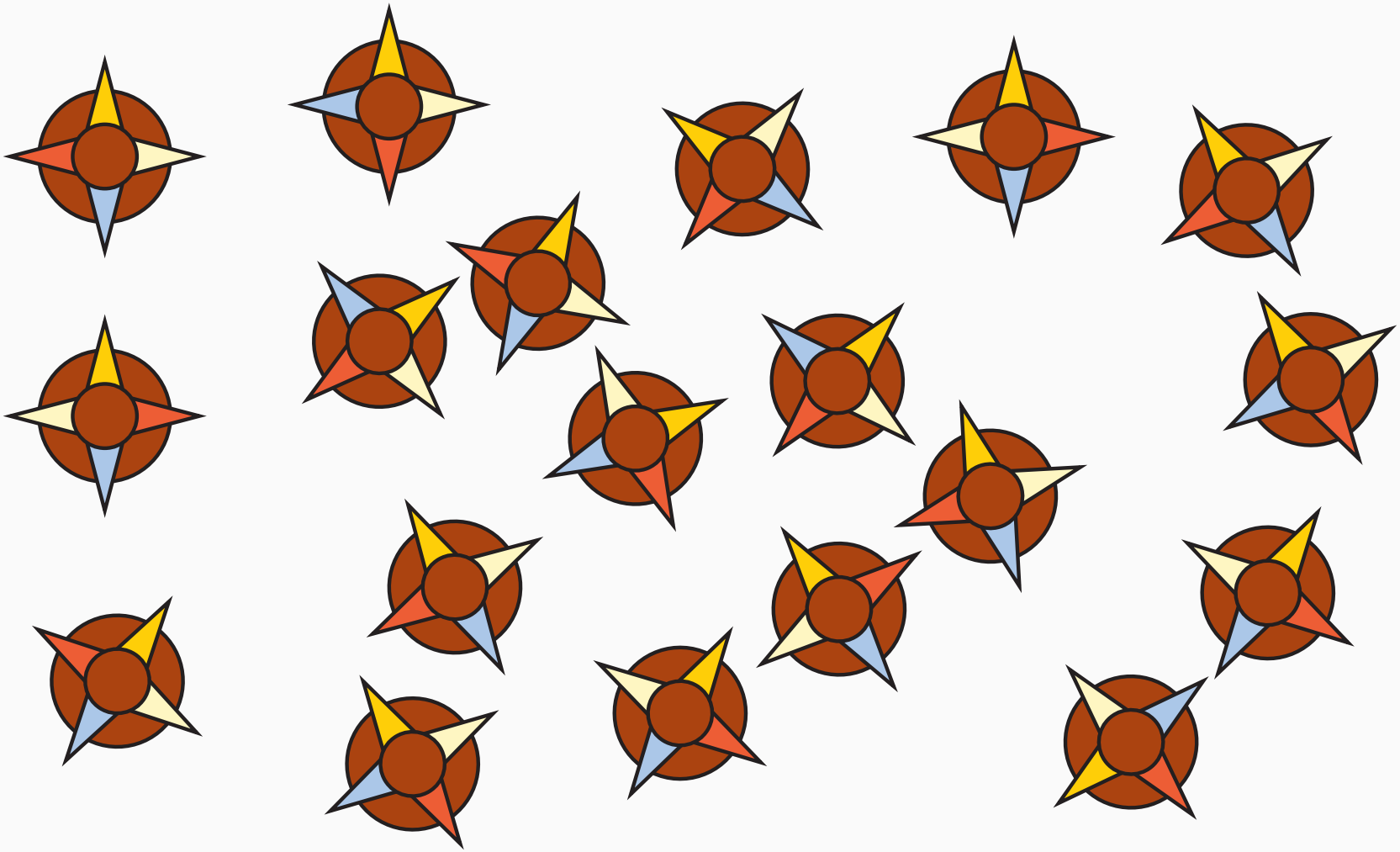
The comparisons are mostly local and work best over small distances.

CT scans

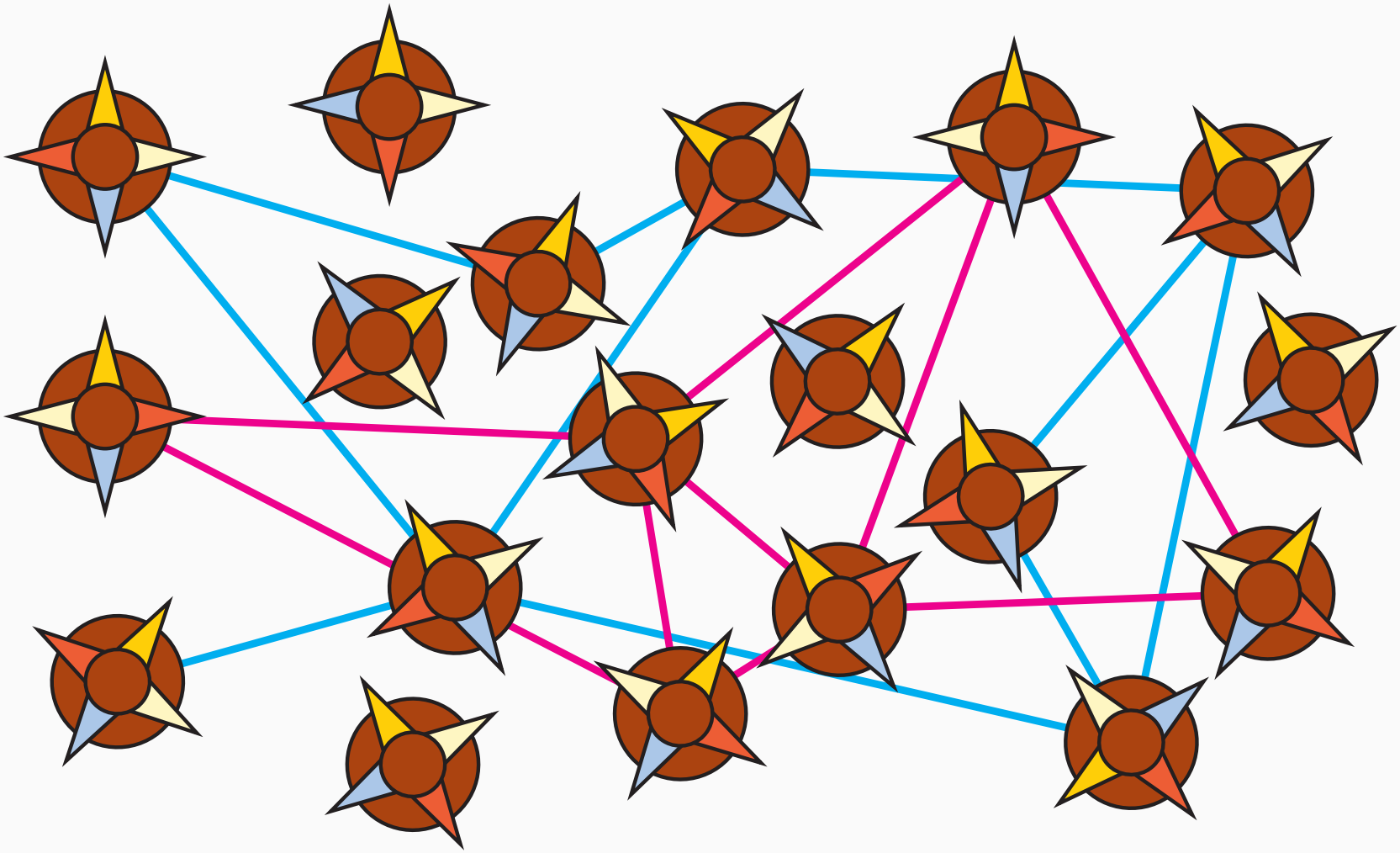


Geometry of synchronisation

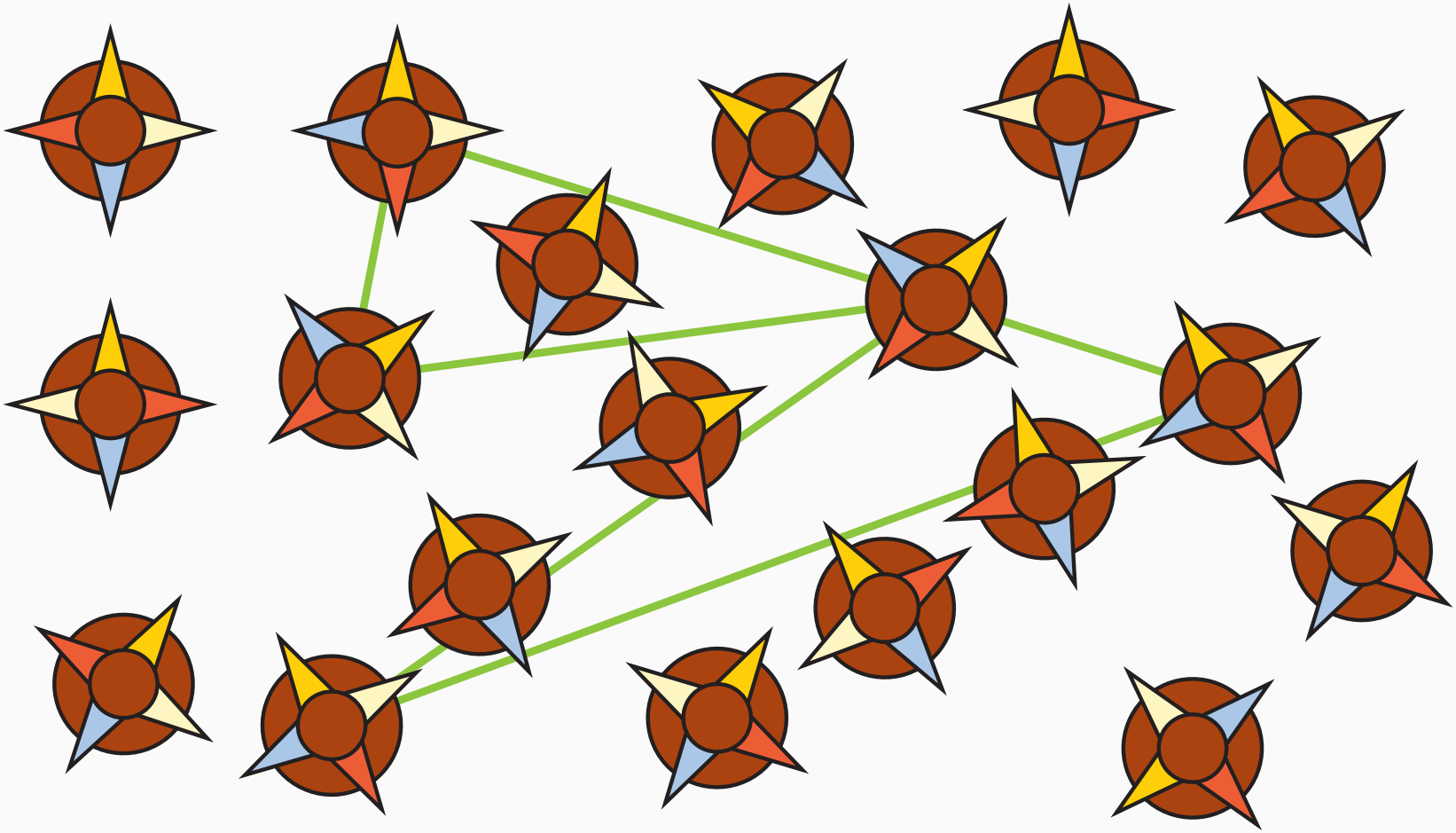
Classifying objects



Classifying objects



Classifying objects



Synchronisation

Synchronisation problems deal with questions regarding aligning a collection of objects in a consistent manner.

Given a sequence of objects $(o_1, ..., o_n)$ and a group (or collection of groups) G the objective is to learn a collection of group elements ρ_{ij} that would transform object o_i to o_j.

Synchronisation

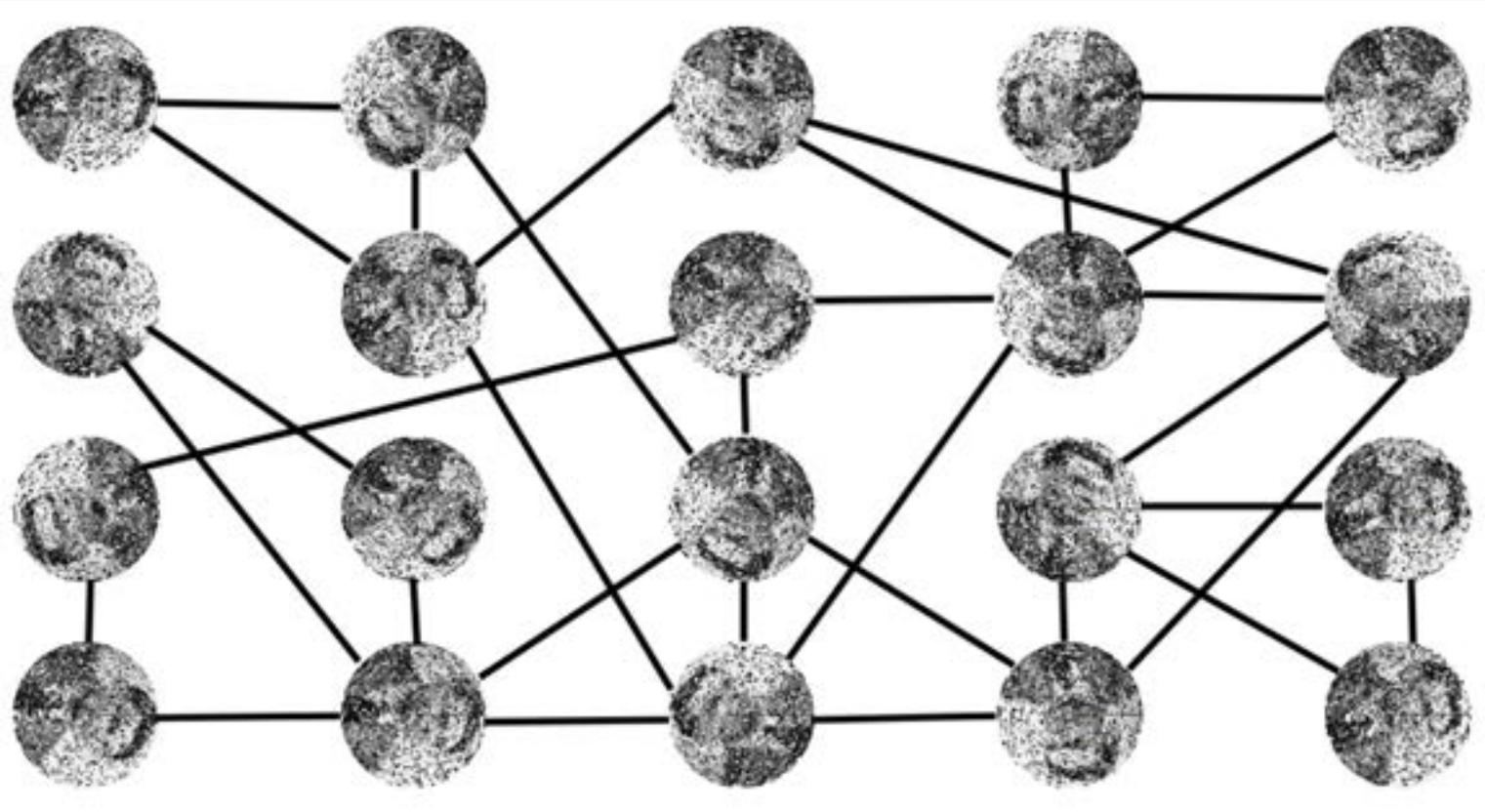
Synchronisation problems deal with questions regarding aligning a collection of objects in a consistent manner.

Given a sequence of objects $(o_1, ..., o_n)$ and a group (or collection of groups) G the objective is to learn a collection of group elements ρ_{ij} that would transform object o_i to o_j.

If it is possible to find a sequence of group elements that allow for an accurate transformation between objects then this set can be *synchronised*.

Synchronization problem with noise

$$y_i = R_i x + \xi_i$$
$$R_i \in O(d), \quad \xi_i \sim \text{i.i.d. noise}$$



Afonso S. Bandeira. "Ten Lectures and Forty-Two Open Problems in the Mathematics of Data Science." (2015).

Spectral properties of graphs

Basics of graph theory

- Assume given a graph G, with a set of vertices V and a set of edges E. For simplicity we assume that G has no loops or multiple edges. (This assumption can be removed.) For a vertex v we denote its degree (or valency) by d_v: this is the number of edges that meet at v.
- Is it possible to measure the *level of complexity of the graph*?
- For example, how *well connected* is the graph? Does it have bottlenecks? How easy is it to disconnect it into components?

Basics of graph theory

The topological structure of the graph does not capture the functional information about the system it describes. To include this information, we use edge weights. A typical example of an *edge weight* is a function $w\colon V \times V \to \mathbb{R}^{\geq 0}$ such that

1. $w(i,j) = w(j,i)$ for all i,j;
2. $w(i,j) = 0$ if $(i,j) \notin E$;
3. $w(i,i) = \sum_{j=1}^{N} w(i,j)$.

In what follows, we will use more general weights that will be allowed to take values in groups, modules over groups, etc.

The discrete Laplacian

The discrete Laplace operator is defined on the space of functions $g : V \to \mathbb{C}$ on the set of vertices V.

$$\Delta g(u) = \sum_{v \sim u} (g(v) - g(u))$$

where $v \sim u$ means that the two vertices are adjacent.

The Laplacian is the composition of two operations:

- 'gradient' (or derivative): $d : C(V) \to C(E)$, $df(u, v) = f(v) - f(u)$, where u, v are ends of one edge;
- 'divergence': $\partial : C(E) \to C(V)$;

$$\partial\phi(u) = \sum_{u \sim v} \phi(u, v).$$

The discrete Laplacian

Since the graph is finite, $C(V) = \mathbb{R}^{\text{number of vertices}}$, $C(E) = \mathbb{R}^{\text{number of edges}}$, and the operators d, ∂, Δ are given by matrices of a suitable size.

The discrete Laplacian

Since the graph is finite, $C(V) = \mathbb{R}^{\text{number of vertices}}$, $C(E) = \mathbb{R}^{\text{number of edges}}$, and the operators d, ∂, Δ are given by matrices of a suitable size.

The *Laplacian matrix* of $\mathcal{G}$ is the $N \times N$ matrix L defined as

$$[L]_{ij} = \begin{cases} d_i, & \text{if } i = j; \\ -w_{ij}, & \text{if } i \neq j \text{ and } (i,j) \in E; \\ 0, & \text{otherwise.} \end{cases}$$

It is a real symmetric matrix with non-positive entries outside the diagonal, and the sum of each column (or row) is zero.

Normalized Laplacian

The *normalized Laplacian* is the matrix $L_{\mathbf{n}} = D^{-1/2} L D^{-1/2}$, where D is the diagonal matrix with nonzero entries

$$d_i = \sum_{i \sim j} w(i, j).$$

Normalized Laplacian

The *normalized Laplacian* is the matrix $L_{\mathbf{n}} = D^{-1/2} L D^{-1/2}$, where D is the diagonal matrix with nonzero entries

$$d_i = \sum_{i \sim j} w(i,j).$$

That is,

$$[L_{\mathbf{n}}]_{ij} = \begin{cases} 1, & \text{if } i = j; \\ \frac{-w_{ij}}{\sqrt{d_i}\sqrt{d_j}}, & \text{if } i \neq j \text{ and } (i,j) \in E; \\ 0, & \text{otherwise.} \end{cases}$$

The normalized Laplacian is scale-independent, and it is more advantageous for clustering purposes.

Some properties of the Laplacian

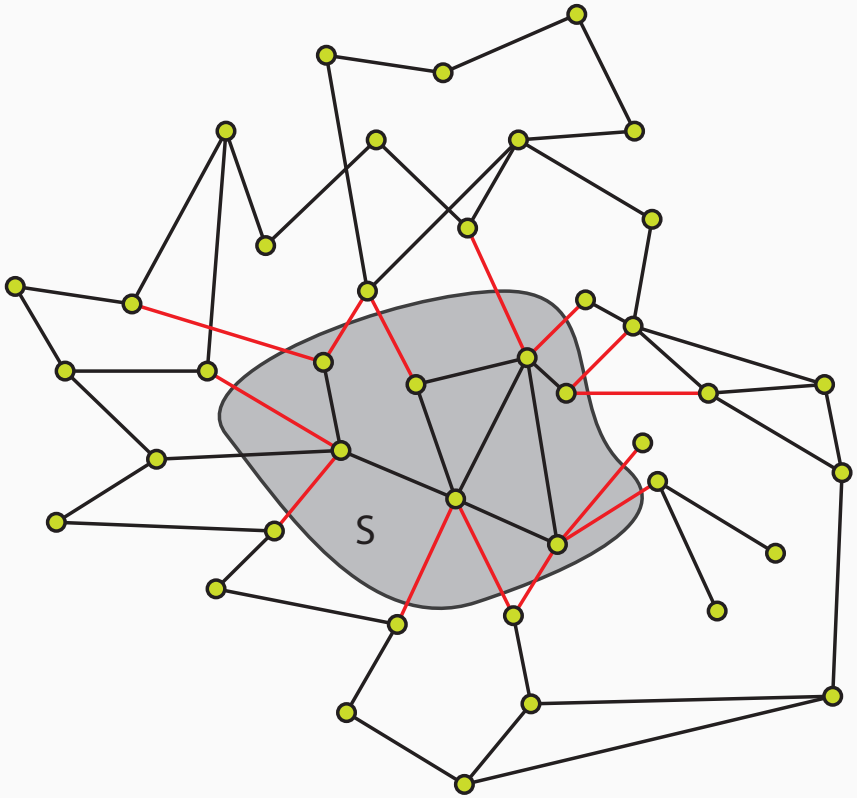
1. When the graph G is connected, the kernel of the Laplacian consists of constant functions. So while Δ is not invertible, it fails to be invertible on a one-dimensional space.
2. The operator Δ is positive, so it has real, nonnegative eigenvalues $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{n-1}$.
3. The first non-zero eigenvalue λ_1 is particularly interesting. It can measure 'the growth' of the graph; an example of this is the edge expansion.

Isoperimetric problems

- A classical problem in graph theory: Remove as few edges as possible to separate the graph into two parts of almost equal size.
- There are many problems of this general type: Remove as little of the graph as possible to separate out a subset of vertices of some desired 'size'. (These are called 'separator problems'.)
- An *edge-cut* is a subset of a set of edges which disconnects the graph.
- Cheeger constants appear in isoperimetric problems involving edge-cuts. They are also important in flow or routing problems on graphs.

The edge expansion of a graph

Let S be a subset of the vertex set of G. The edge boundary ∂S of S consists of all edges with exactly one end in S. This is the same as $\partial\overline{S}$, where $\overline{S} = G \setminus S$.
What is the typical size of the boundary ∂S? This information indicates the level of connectedness of the graph.



Edge expansion

The *volume* of a subgraph spanned by a subset $S \subset V$ is:

$$\text{vol}(S) = \sum_{i \in S} d_i = \sum_{i \in S} w_{ii}.$$

The *boundary* of S is the sum of the weights of the edges between vertices in S and vertices not in S:

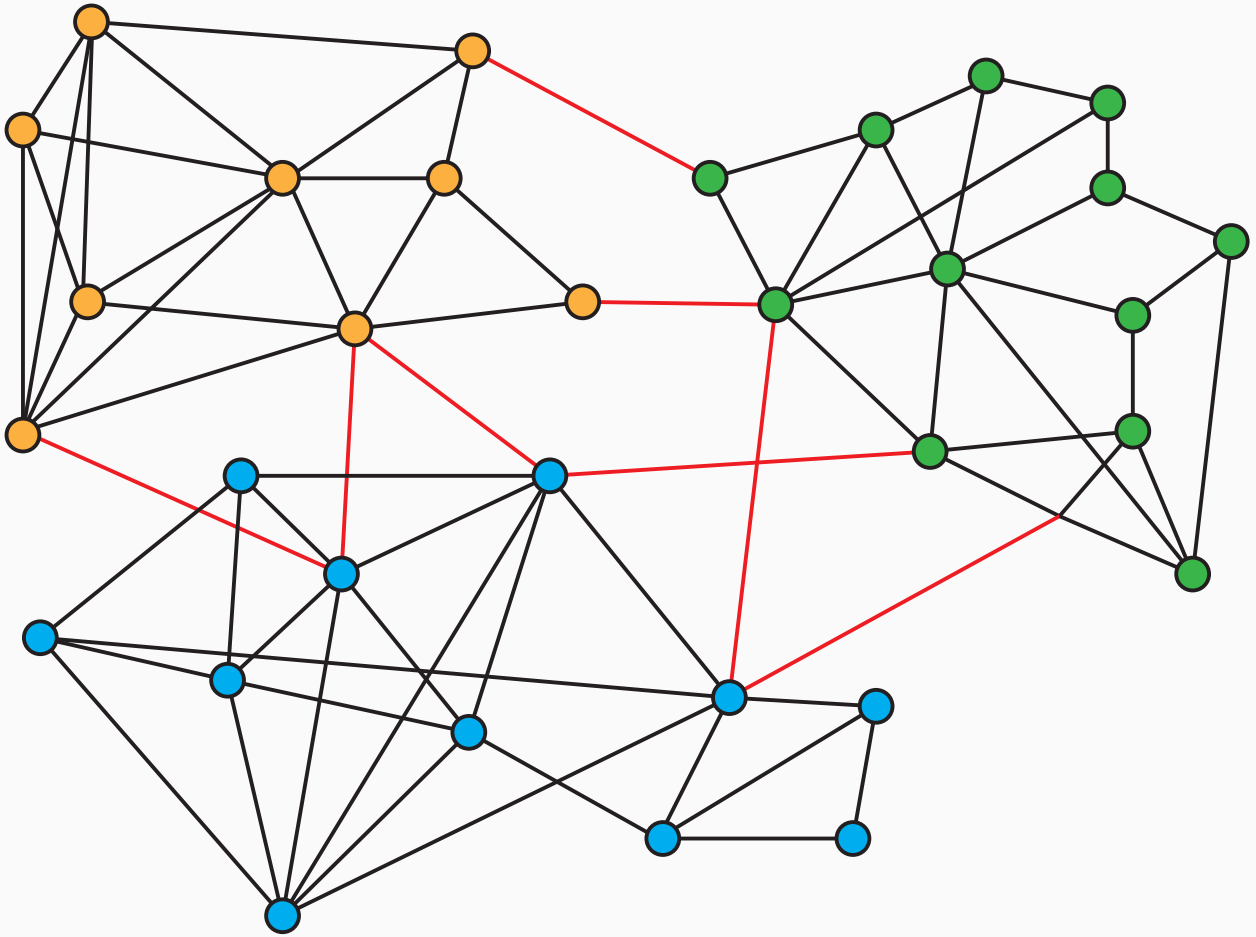
$$\partial(S) = \sum_{i \in S, j \notin S} w_{ij}.$$

The *edge expansion* of S as the quotient

$$\phi(S) = \frac{\partial(S)}{\text{vol}(S)}.$$

Spectral clustering

We will be interested in the question of clustering on graphs, that is, a partition of the set of vertices V



Spectral clustering

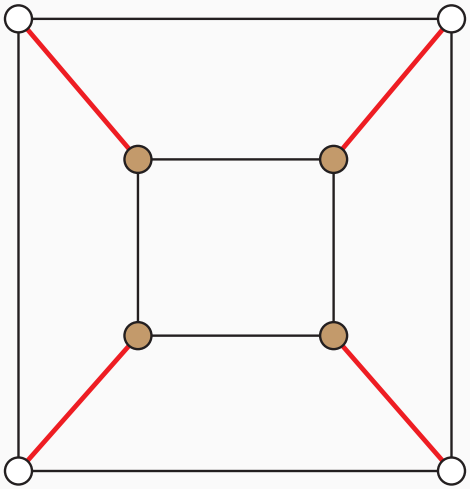
We will be interested in the question of clustering on graphs, that is, a partition of the set of vertices V

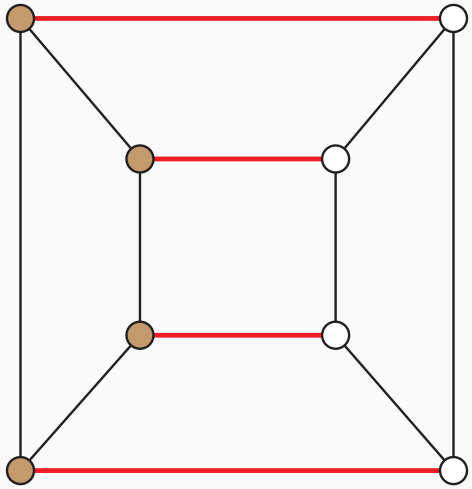
For this, the edge expansion is a useful indicator of quality:

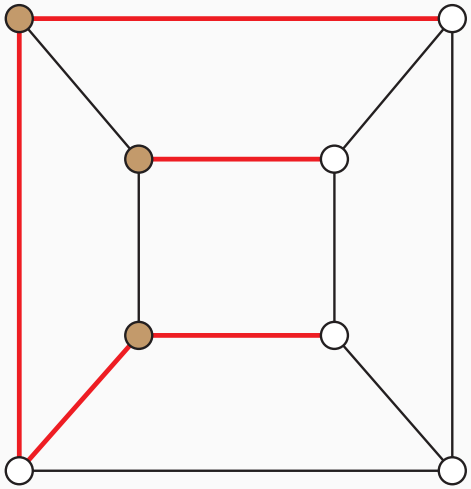
$$\phi(S) = \frac{\partial(S)}{\text{vol}(S)}.$$

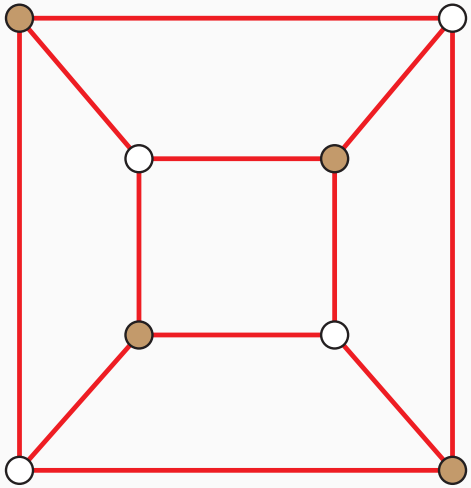
$\phi(S)$ is small if the vertices of S are more connected among themselves than they are to the complement of S.

Edge expansion





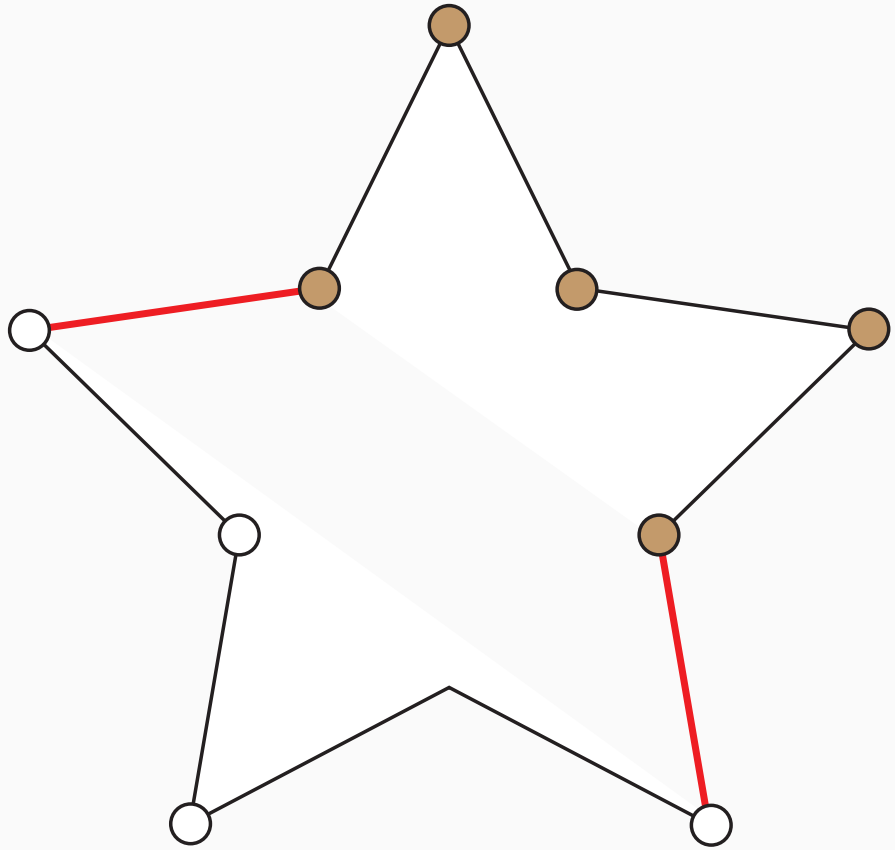




The Cheeger constant

$$h_G = \min \left\{ \frac{|\partial S|}{\operatorname{vol} S} \right\}$$

For graphs with bottlenecks, the Cheeger constant will be small, close to zero.
Graphs with well connected components will have a larger Cheeger constant.
Here $|\partial S| = 2$ so $|\partial S| / \operatorname{vol} S \to 0$ as the size of S grows.

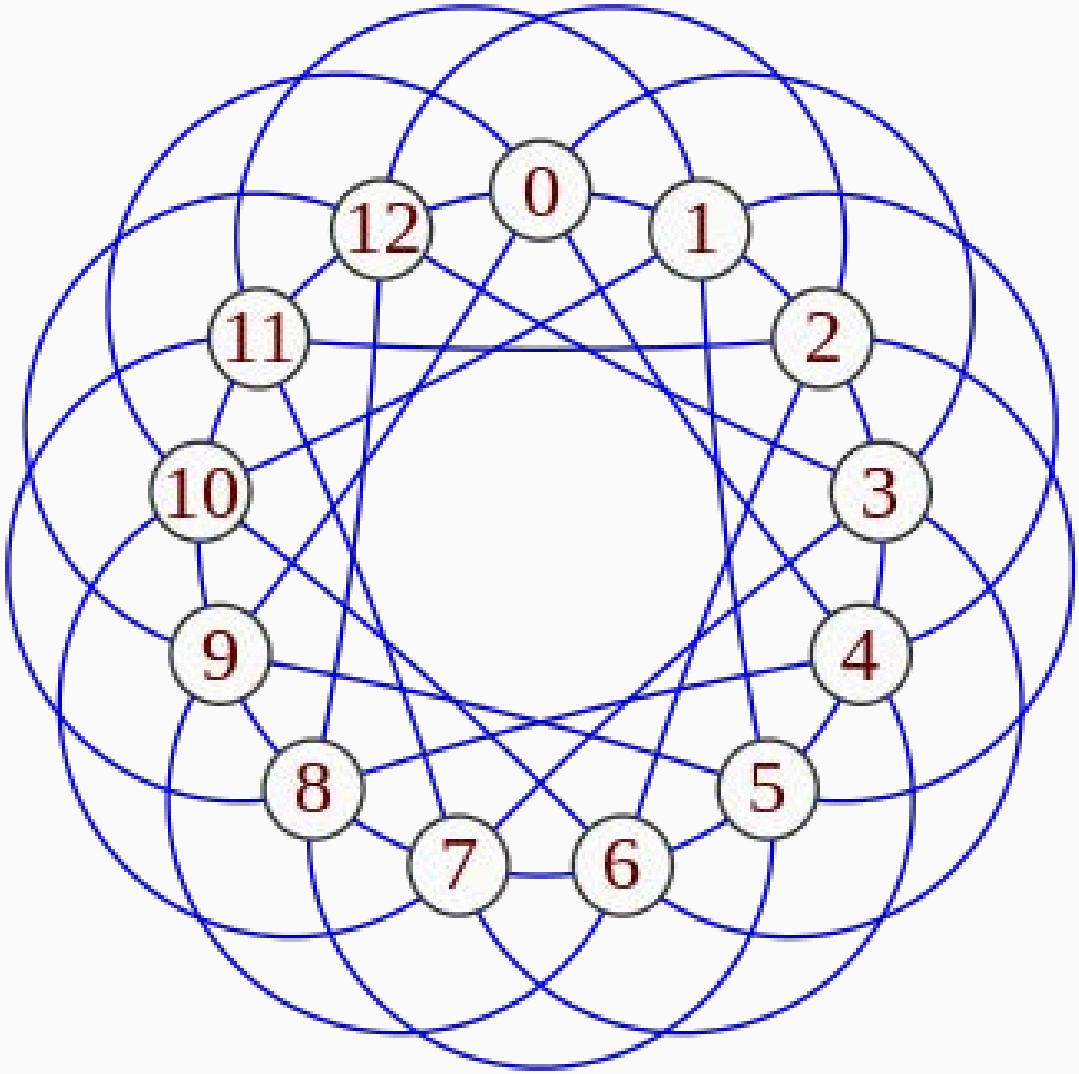


Cheeger inequality

We have the *isoperimetric inequality*

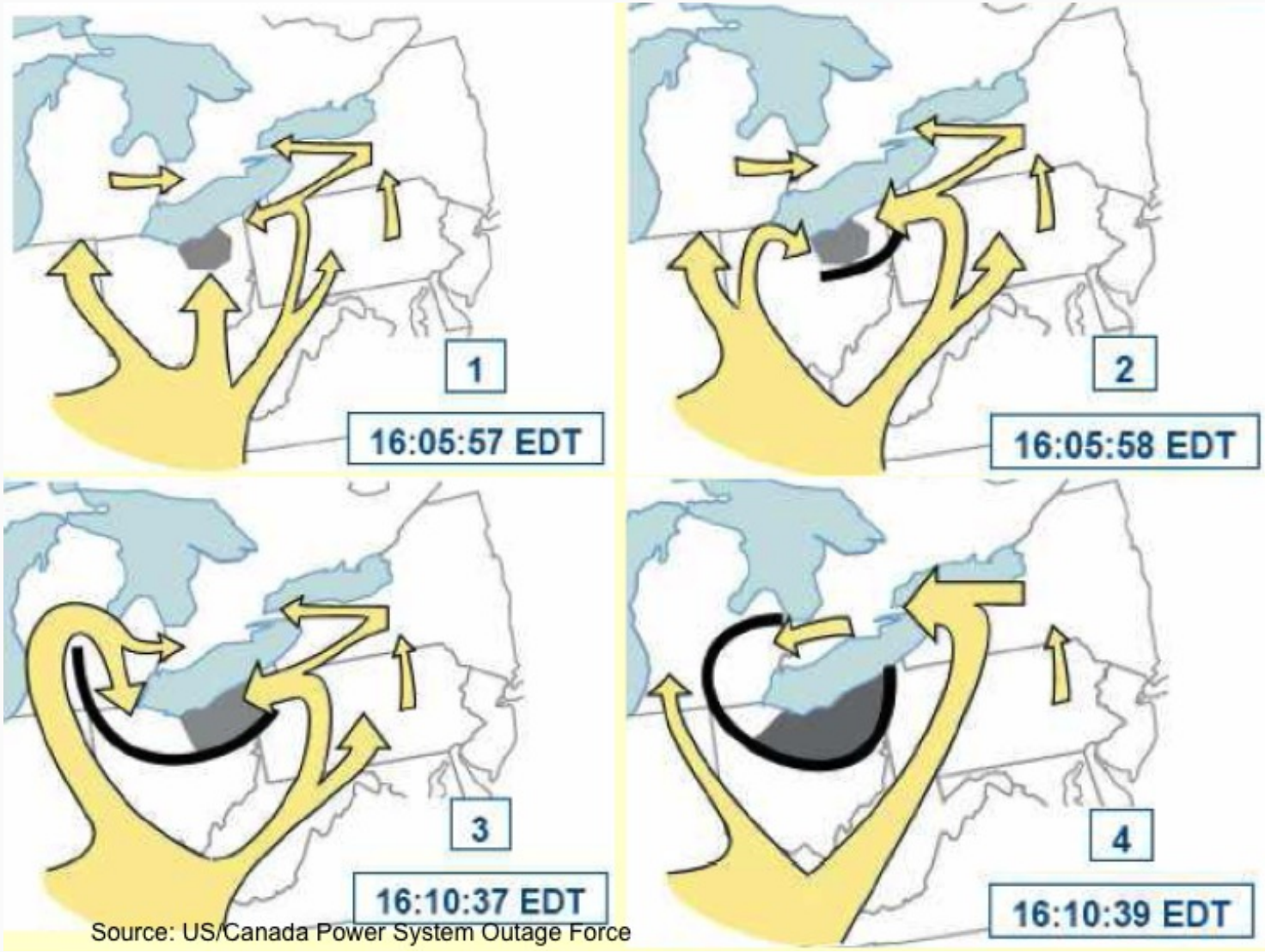
$$\lambda_1/2 \leq h_G < \sqrt{2\lambda_1}.$$

Expander graphs

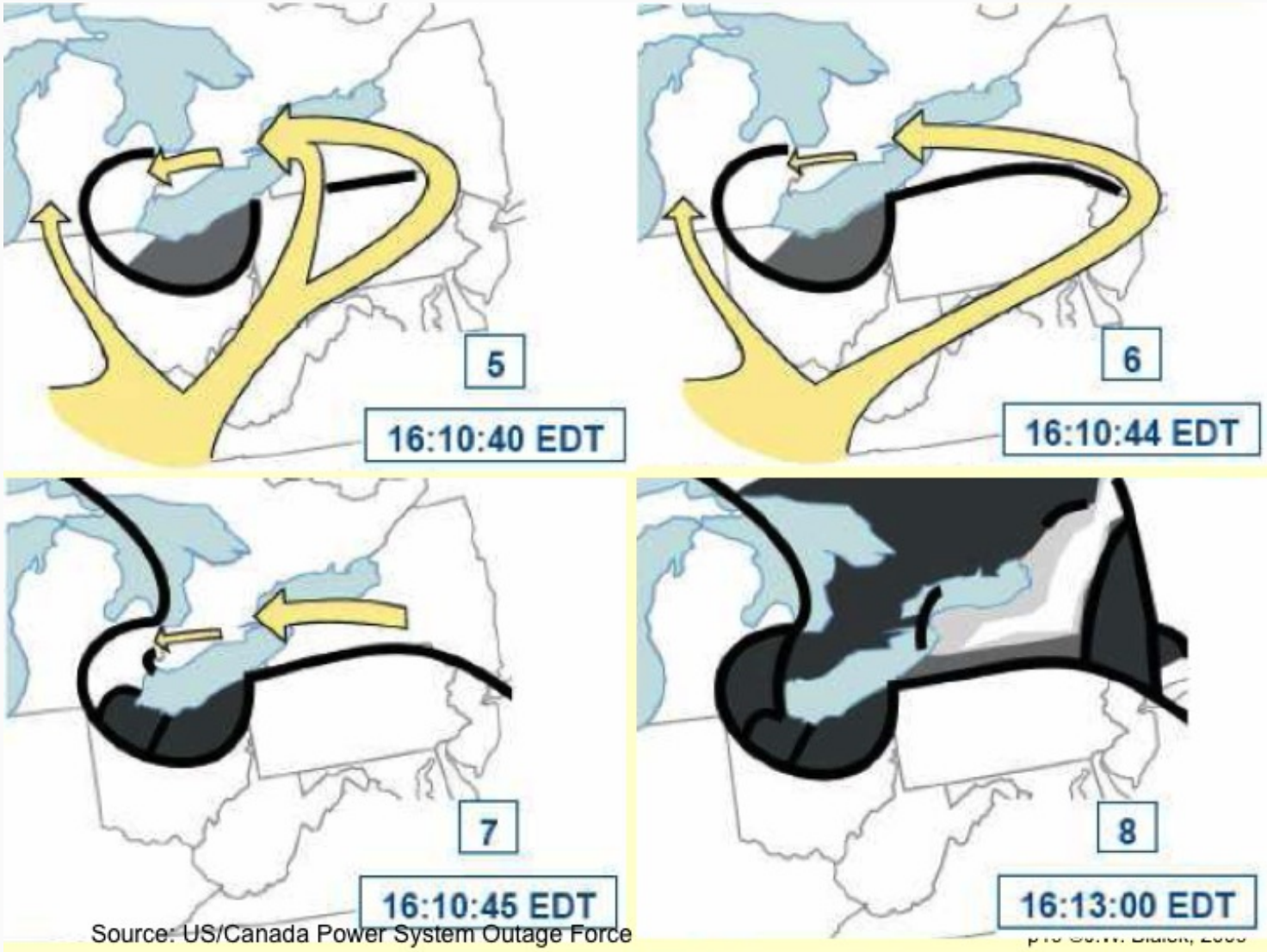


Applications of spectral clustering

Wide-area blackouts in power grids



Wide-area blackouts in power grids



Wide-area blackouts in power grids

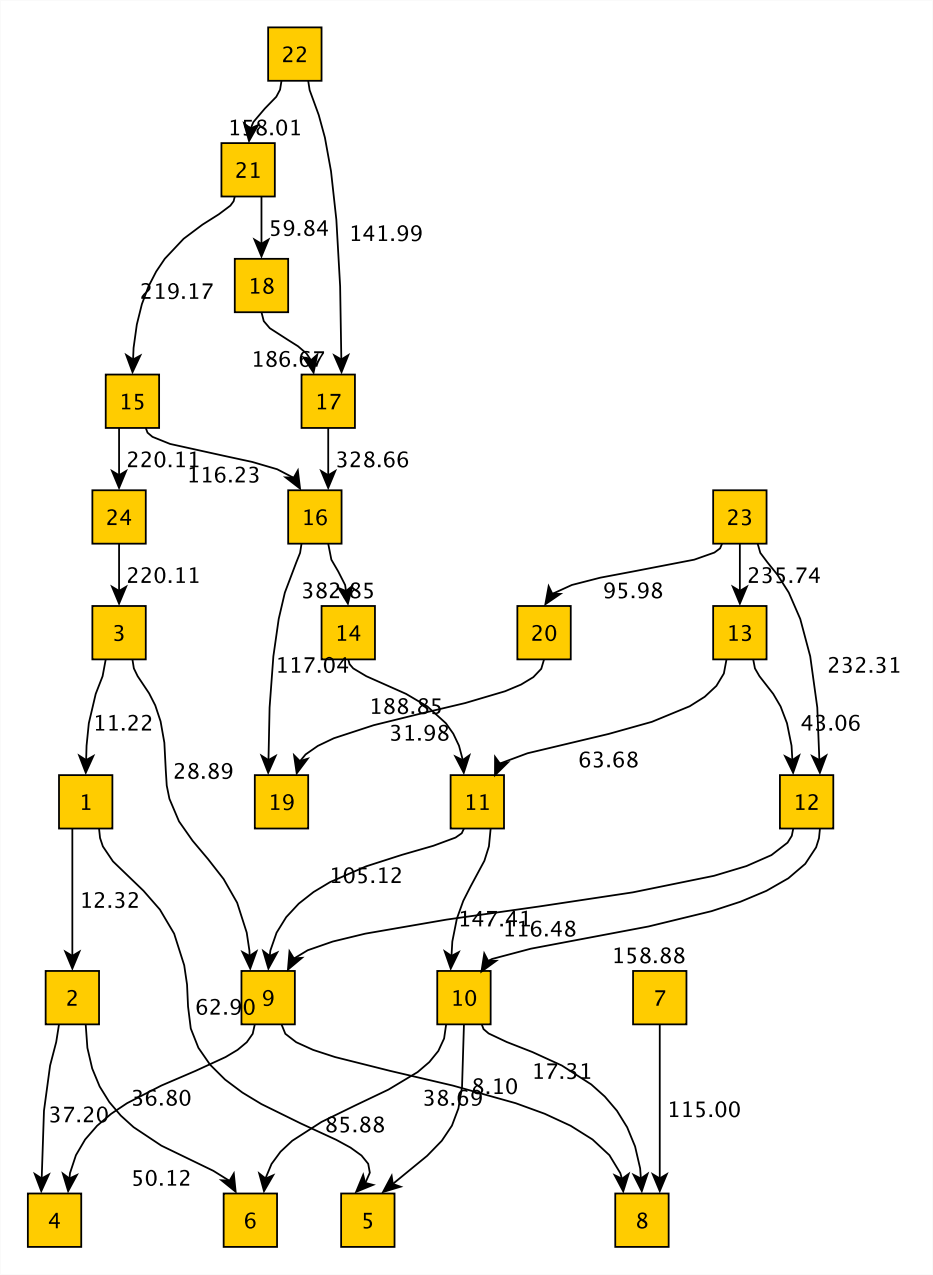
Wide-area blackouts in power grids

Power grids as graphs

The main high voltage part (above 132 kV) of the UK power grid can be represented as a graph with 810 vertices and 1194 edges.

If lower voltage nodes were included the graph would have many thousands of vertices.

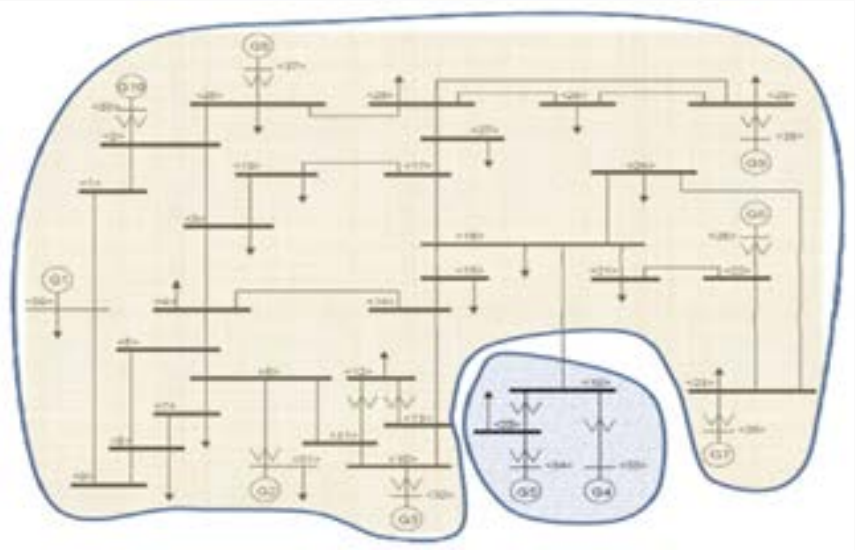
Functional structure



It is crucial that the graph model captures the functional and physical structure of the grid: for example, power flow through the edges.

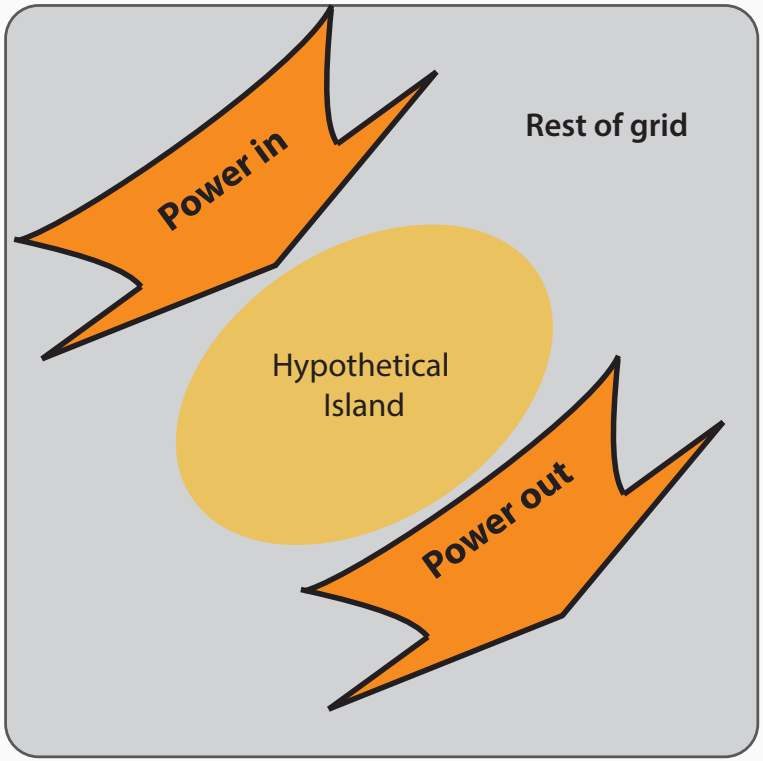
The islanding problem

- Find a way to identify energetically balanced islands within a power network (preferably in real time);
- The islands should have a small overall powerflow through the boundary
- Understand the effect of removing a part of the grid on the rest of the network. This requires the use of realistic constraints in all models.



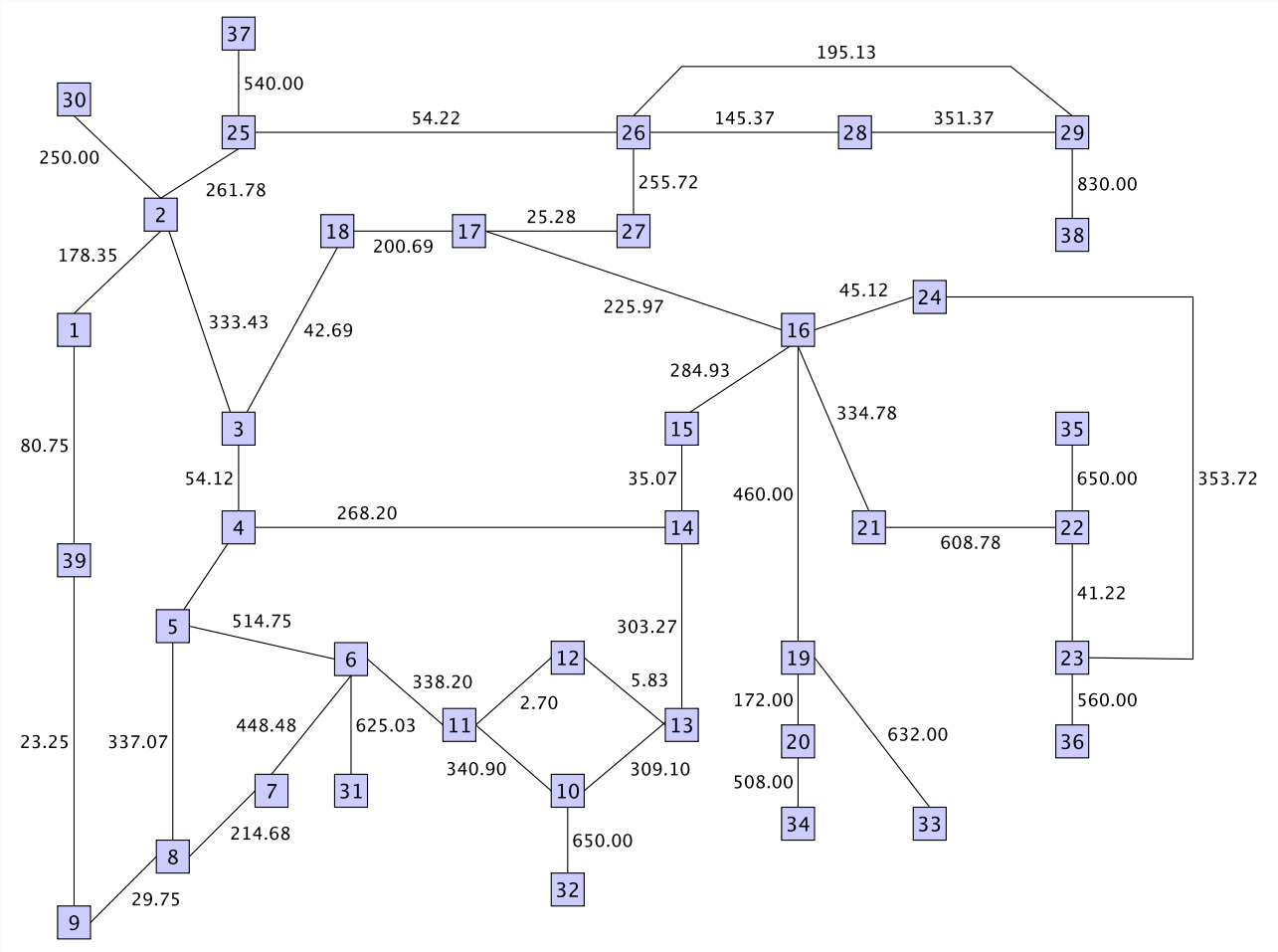
The islanding problem

- Find a way to identify energetically balanced islands within a power network (preferably in real time);
- The islands should have a small overall powerflow through the boundary
- Understand the effect of removing a part of the grid on the rest of the network. This requires the use of realistic constraints in all models.



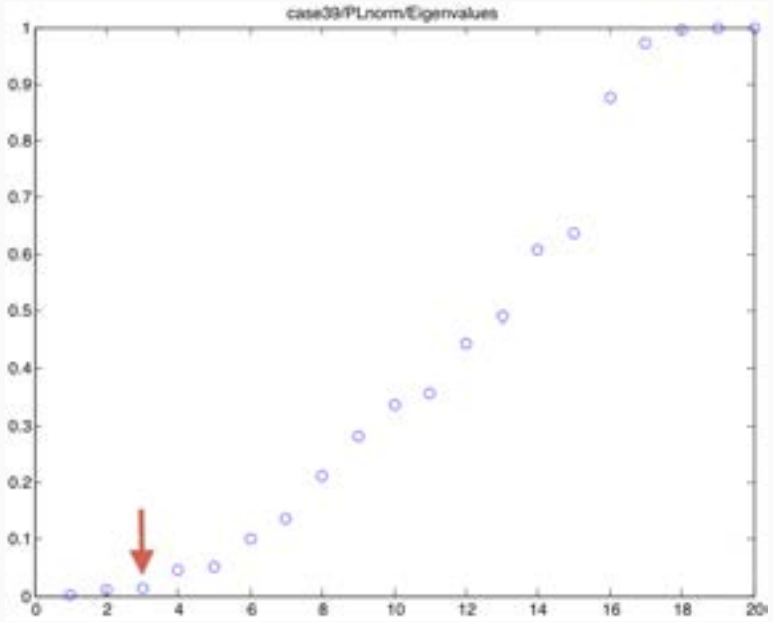
Spectral clustering of power grids

Step 0 (Spectral dimension) Identify gaps in the spectrum of the Laplacian weighted by the powerflow



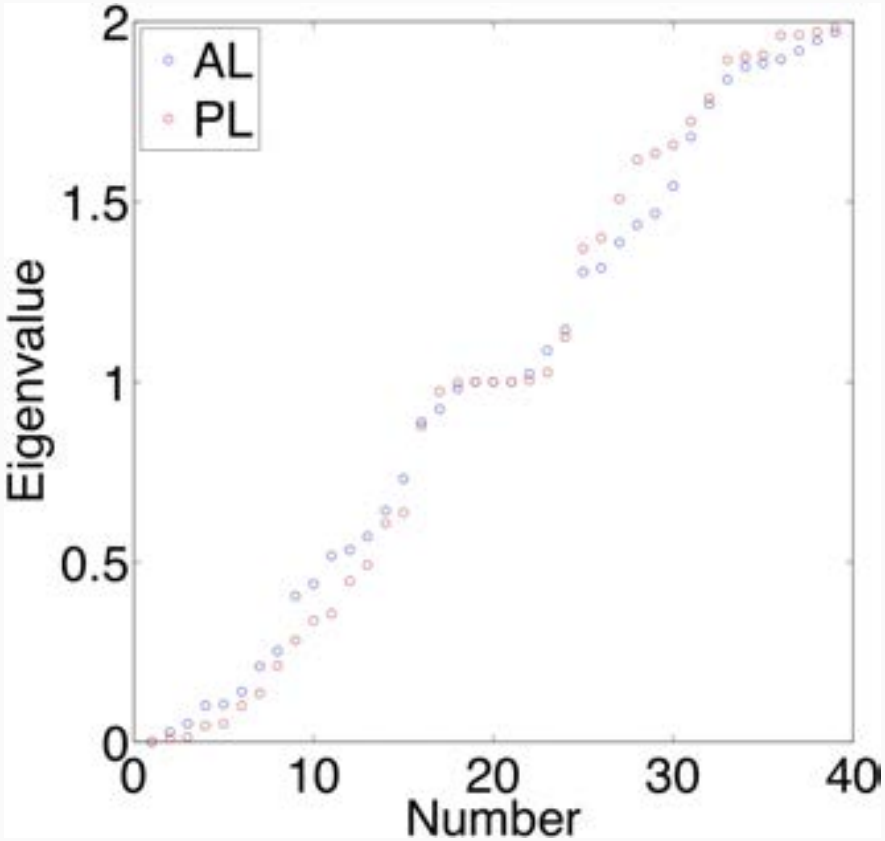
Spectral clustering of power grids

Step 0 (Spectral dimension) Identify gaps in the spectrum of the Laplacian weighted by the powerflow



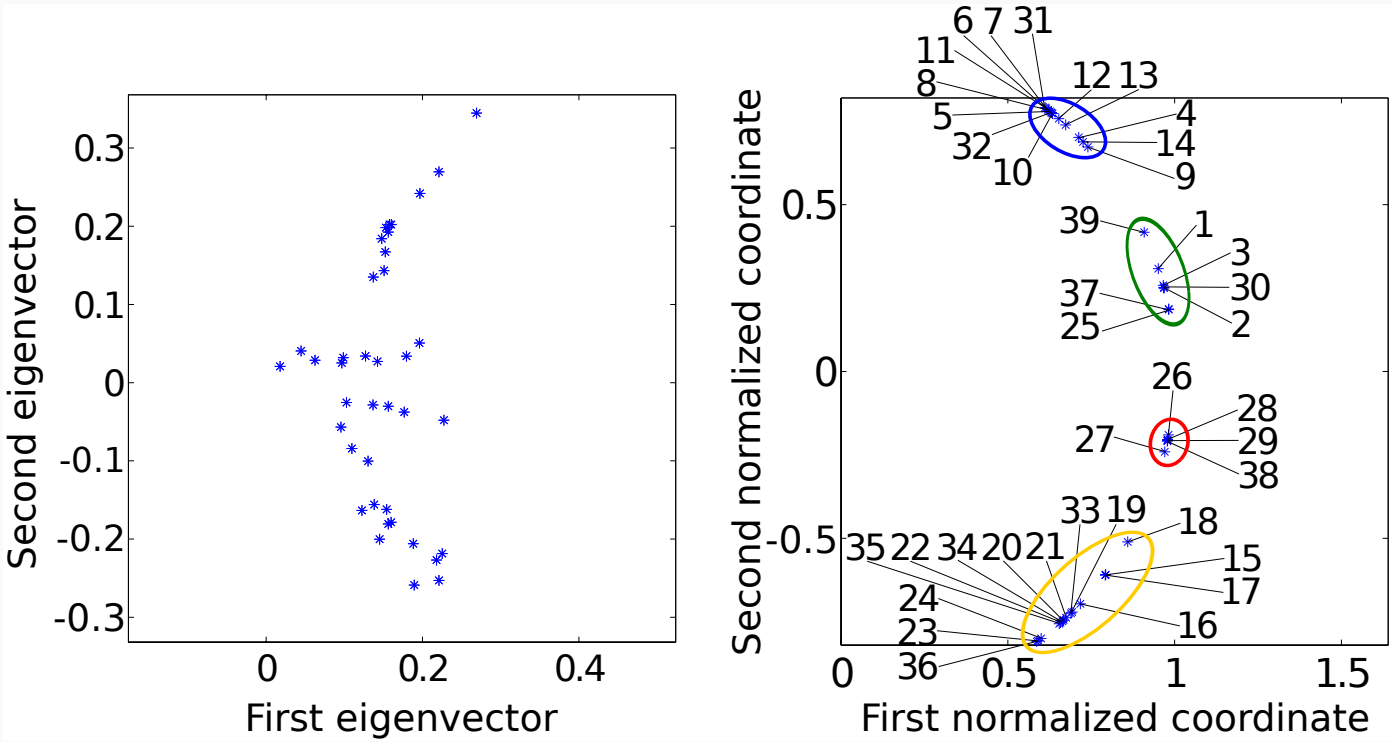
Spectral clustering of power grids

Step 0 (Spectral dimension) Identify gaps in the spectrum of the Laplacian weighted by the powerflow



Steps 1, 2 (Normalised Laplacian, Spectral embedding)

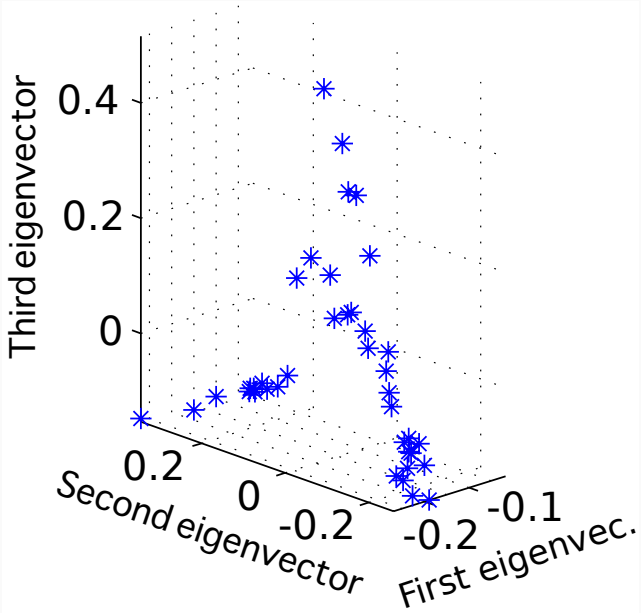
Use the first k eigenvectors to provide coordinates to the vertices in $\mathbb{R}^k$, for $k = 2$ and 3 to embed in $\mathbb{R}^2$ or $\mathbb{R}^3$.



Spectral clustering of power grids

Steps 1, 2 (Normalised Laplacian, Spectral embedding)

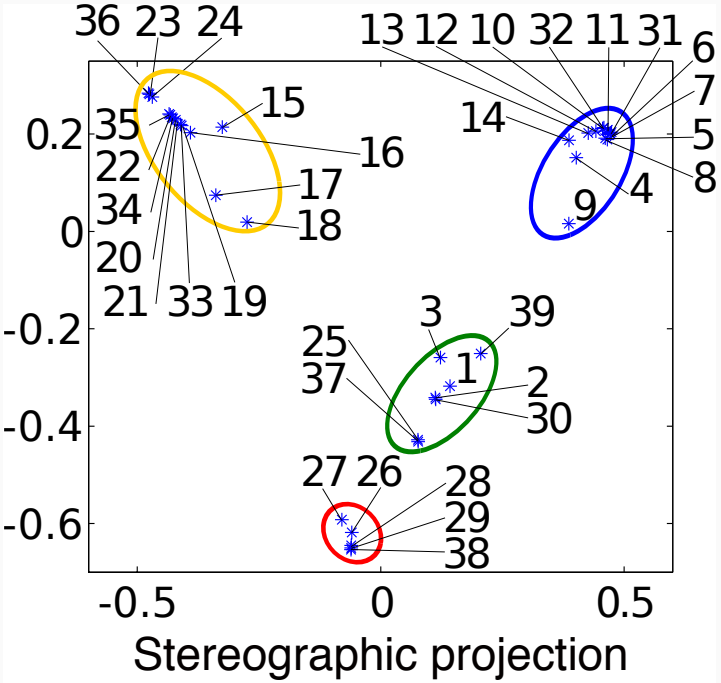
Use the first k eigenvectors to provide coordinates to the vertices in $\mathbb{R}^k$, for $k = 2$ and 3 to embed in $\mathbb{R}^2$ or $\mathbb{R}^3$.



Spectral clustering of power grids

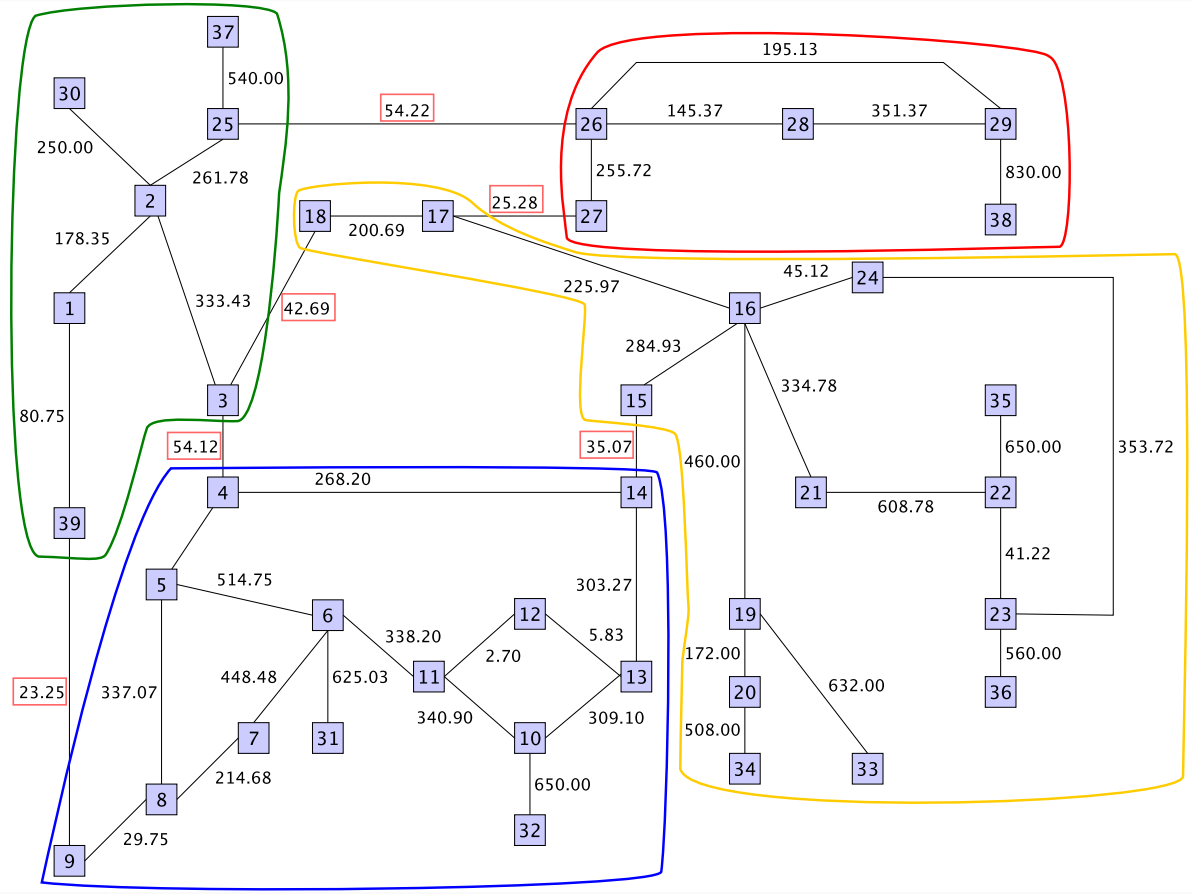
Steps 1, 2 (Normalised Laplacian, Spectral embedding)

Use the first k eigenvectors to provide coordinates to the vertices in $\mathbb{R}^k$, for $k = 2$ and 3 to embed in $\mathbb{R}^2$ or $\mathbb{R}^3$.



Spectral clustering of power grids

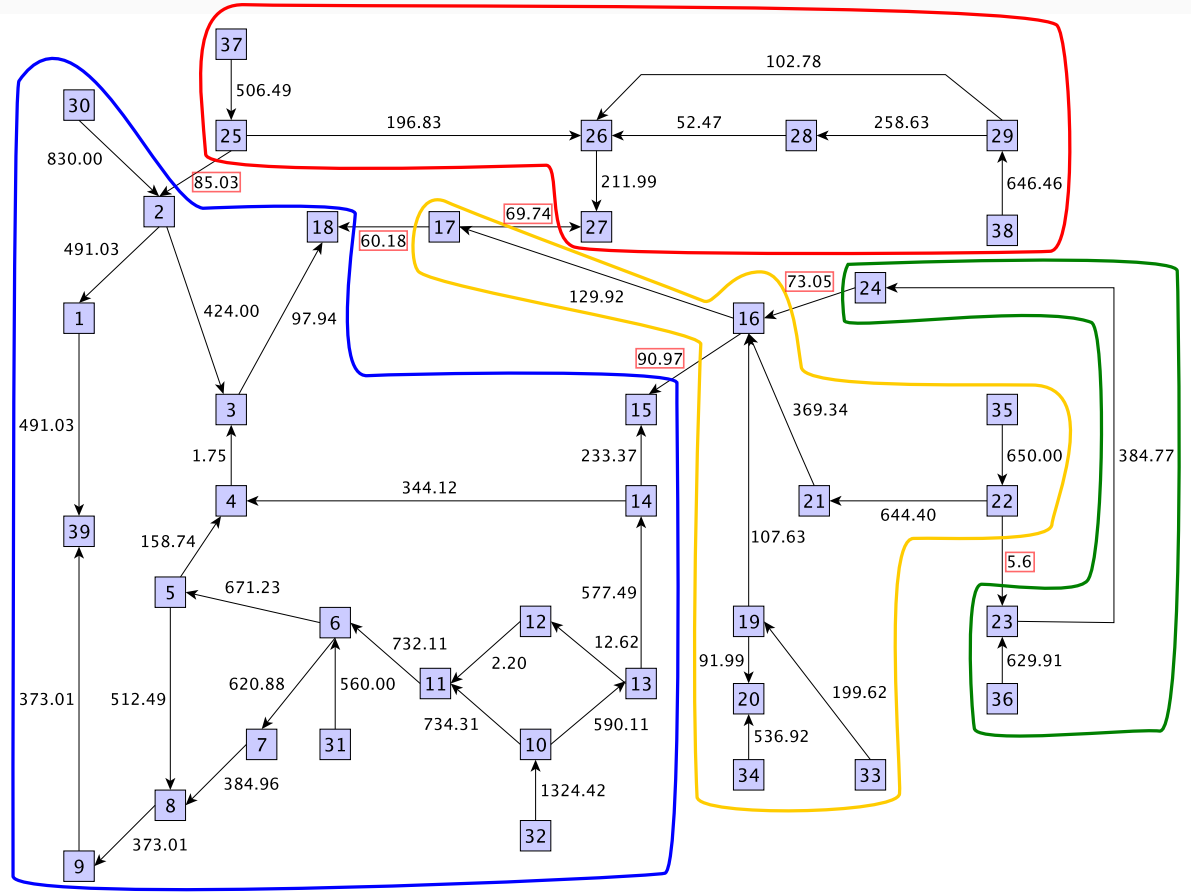
Step 3 (Power flows)



Observe that the island with just bus 12 has the smallest boundary $\partial(S) = 8.53$ but maximal expansion $\phi(S) = 100\%$, as its boundary amounts to all its (electrical) size.

Spectral clustering of power grids

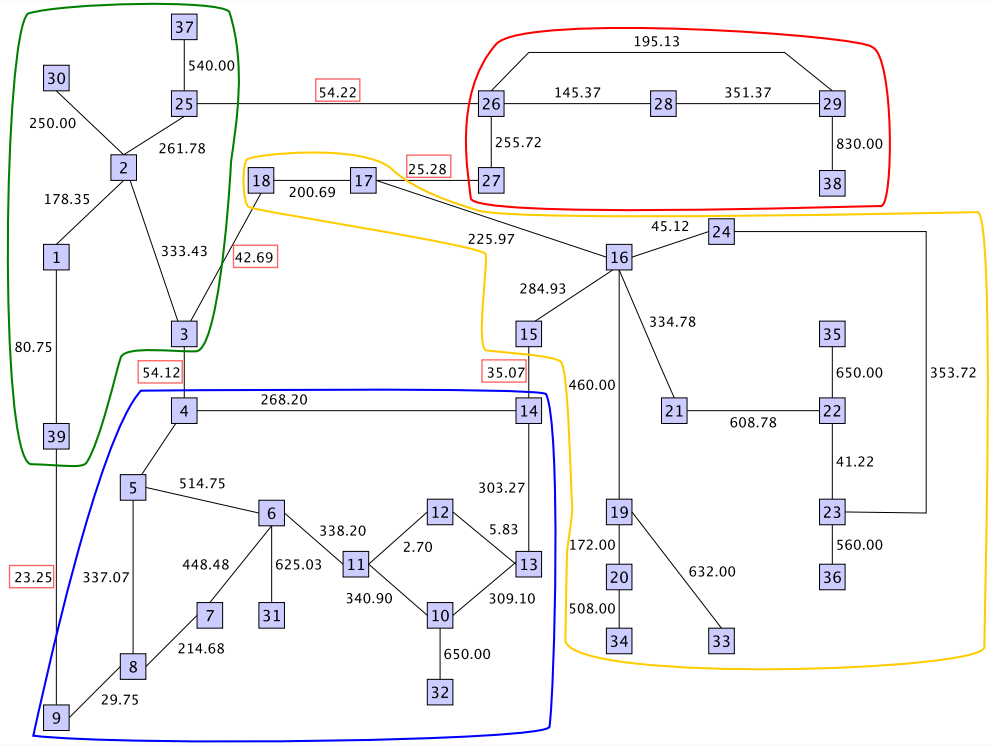
Step 3 (Power flows)

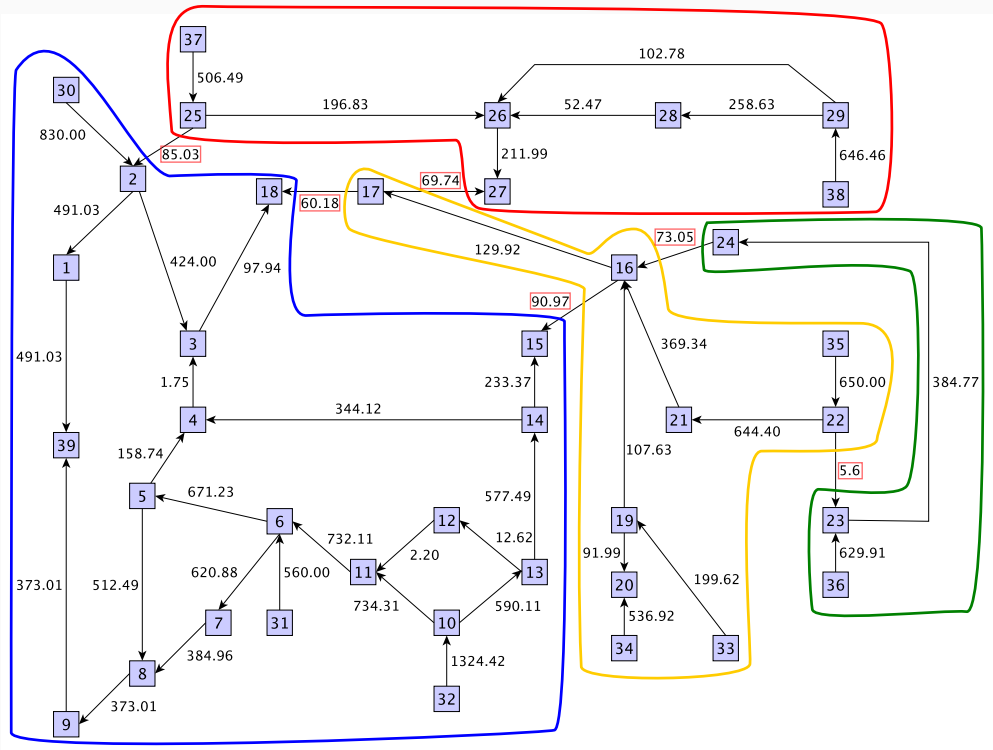


Alternative power flow in the IEEE 39-bus test system and clustering into four islands.

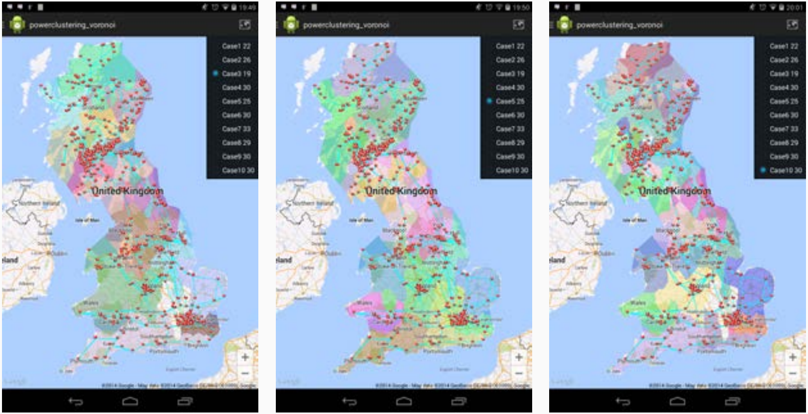
Spectral clustering of power grids

Step 3 (Power flows)





UK power grid



Synchronisation

Synchronisation

Synchronisation problems deal with questions regarding aligning a collection of objects in a consistent manner.

Given a sequence of objects $(o_1, ..., o_n)$ and a group (or collection of groups) G the objective is to learn a collection of group elements ρ_{ij} that would transform object o_i to o_j.

Synchronisation

Synchronisation problems deal with questions regarding aligning a collection of objects in a consistent manner.

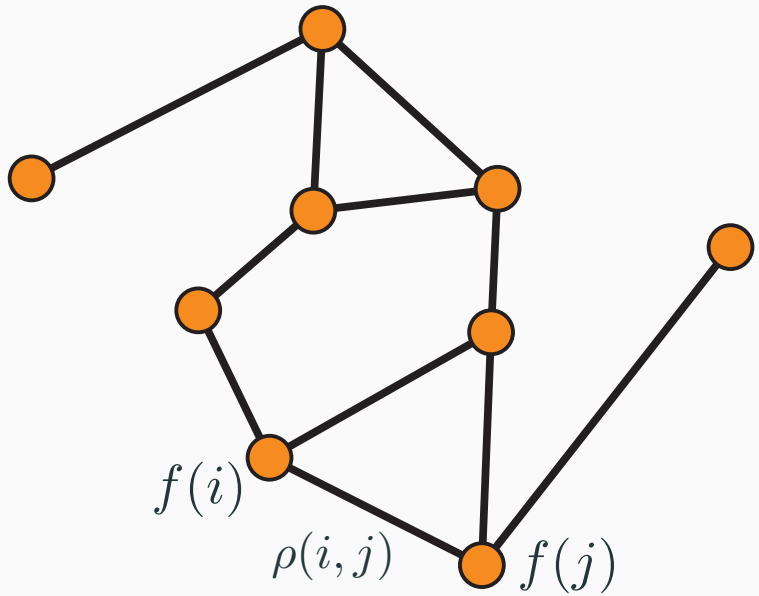
Given a sequence of objects $(o_1, ..., o_n)$ and a group (or collection of groups) G the objective is to learn a collection of group elements ρ_{ij} that would transform object o_i to o_j.

If it is possible to find a sequence of group elements that allow for an accurate transformation between objects then this set can be *synchronised*.

Vertex and edge potentials

Let $\Gamma = (V, E)$ be a graph with vertex set V and edge set E. Let G be a group.

- A *vertex potential* is a map $f : V \to G$.
- An *edge potential* is a function $\rho : E \to G$ which is symmetric: for every edge $(i, j) \in E$ we have $\rho(j, i) = \rho(i, j)^{-1}$

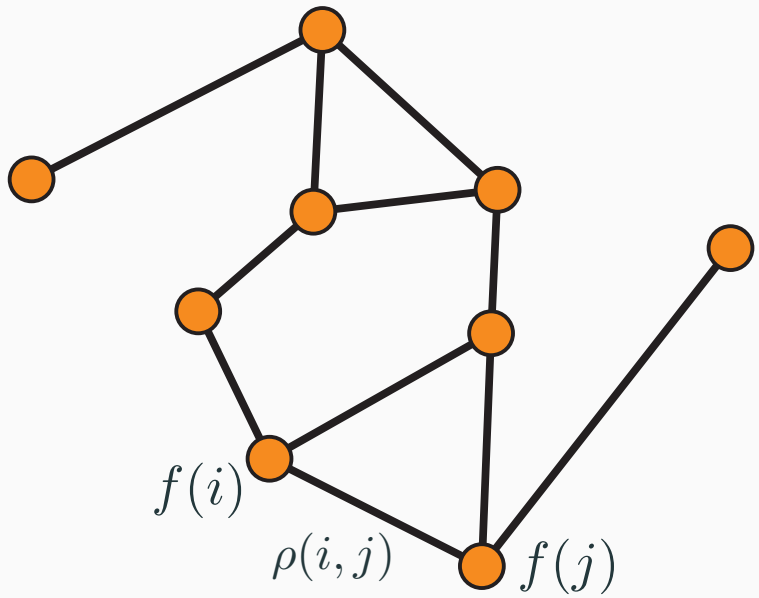


Compatible potentials

- We say that a group potential f and an edge potential ρ are *compatible across an edge* (i, j) iff

$$f(i) = \rho(i, j) f(j).$$

- We say that the two potentials ν and ρ are *compatible* iff they are compatible across every edge $(i, j) \in E$.
- It is natural to extend this definition to the case $f : V \to M$, where M is a G-module.



Synchronisation problem

Bandeira, Singer and Spielman (2013) posed the following question:

Given an edge potential $\rho : E \to G$, does there exits a vertex potential $f : V \to G$ which is compatible with ρ?

Synchronisation problem

Bandeira, Singer and Spielman (2013) posed the following question:

Given an edge potential $\rho : E \to G$, does there exits a vertex potential $f : V \to G$ which is compatible with ρ?

Note that in this formulation the converse problem is easy. Given a vertex potential $f : V \to G$, one can define a compatible edge potential ρ by

$$\rho(i, j) = f(i)f(j)^{-1}.$$

Synchronisation problem

Bandeira, Singer and Spielman (2013) posed the following question:

Given an edge potential $\rho : E \to G$, does there exits a vertex potential $f : V \to G$ which is compatible with ρ?

Note that in this formulation the converse problem is easy. Given a vertex potential $f : V \to G$, one can define a compatible edge potential ρ by

$$\rho(i, j) = f(i)f(j)^{-1}.$$

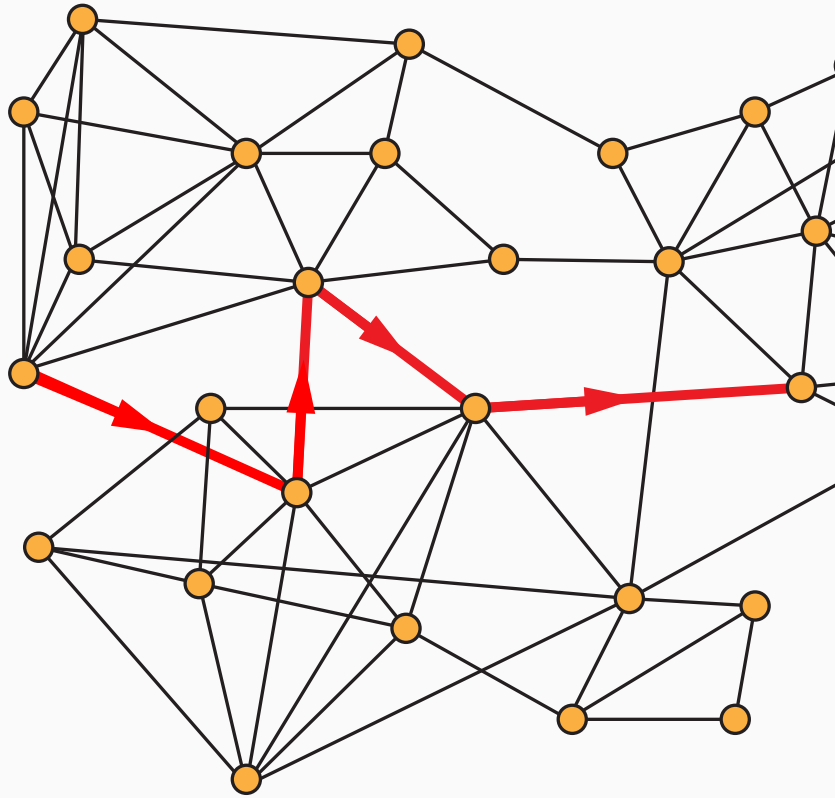
The problem becomes more complex when we consider functions $f : V \to M$, for a G-module M.

Synchronisation along paths

A *path* γ connecting vertices v and w is a sequence of edges

$$\gamma = (e_1, e_2, \ldots, e_n)$$

where for $k = 1, \ldots, n$, $e_k = (i_k, j_k)$, $v = i_1$, $j_k = i_{k+1}$ and $w = j_n$.



An edge potential $\rho : E \to G$ gives rise to a map

$$\{\text{Paths in } \Gamma\} \longrightarrow G, \qquad R : \gamma \mapsto \rho(e_1) \dots \rho(e_n).$$

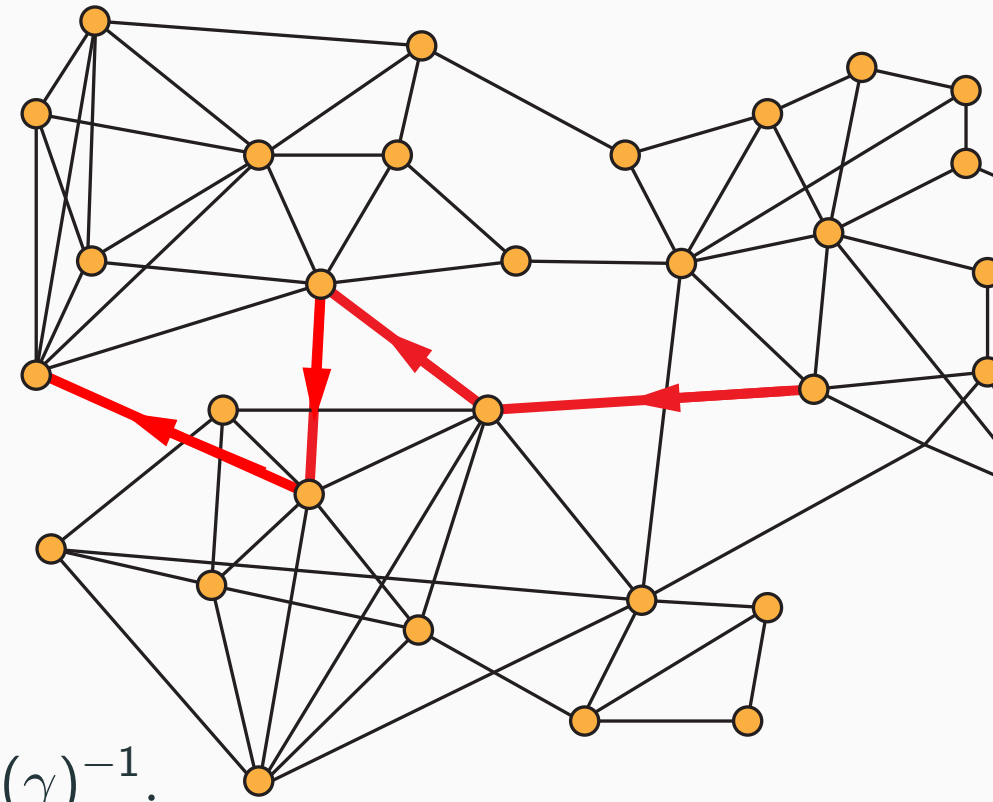
Reversal property

For $e = (i, j)$, let $e^{-1} = (j, i)$, and

$$\gamma^{-1} = (e_n^{-1}, \dots, e_1^{-1}).$$

Because the edge potential ρ is symmetric:

$$\begin{aligned} R(\gamma^{-1}) &= \rho(e_n^{-1}) \dots \rho(e_1^{-1}) \\ &= \rho(e_n)^{-1} \dots \rho(e_1)^{-1} = R(\gamma)^{-1}. \end{aligned}$$

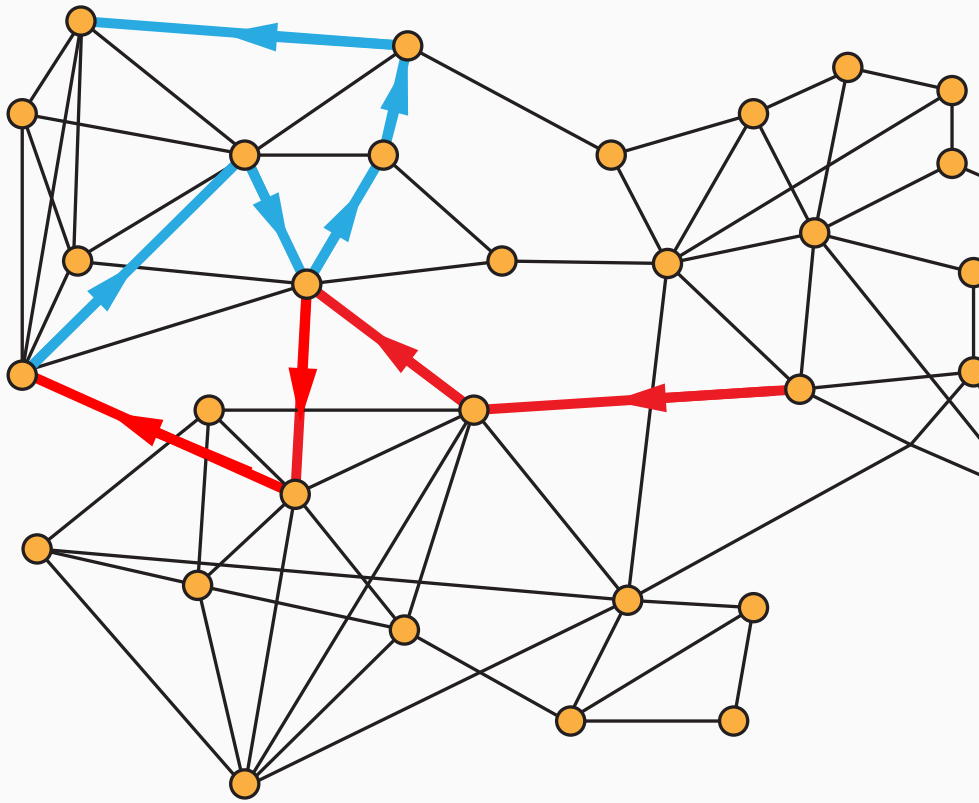


An edge potential $\rho : E \to G$ gives rise to a map

$$\{\text{Paths in } \Gamma\} \longrightarrow G, \qquad R : \gamma \mapsto \rho(e_1) \dots \rho(e_n).$$

Concatenation

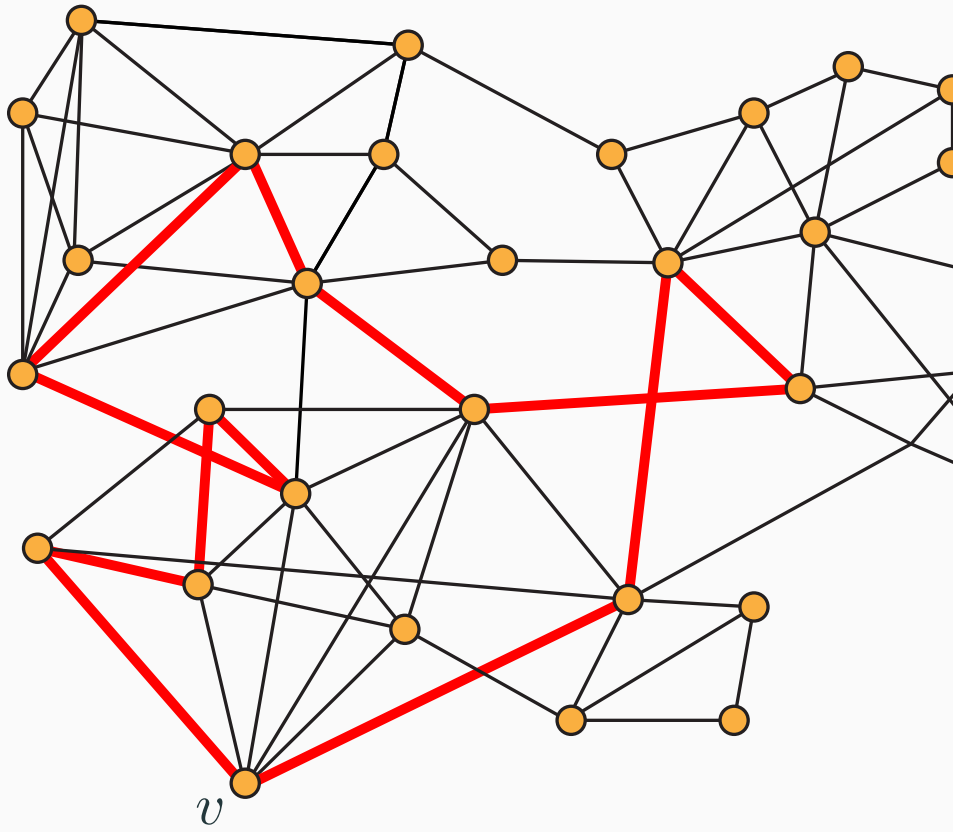
$$R(\gamma_1 \circ \gamma_2) = R(\gamma_1)R(\gamma_2).$$



Holonomy groups

Let ω be a based loop at the vertex v.
Let Ω_v be the space of loops based at v together with the trivial loop $0 = (v, v)$; put $R(0) = 1$.
Then we have a natural map from the space Ω_v of loops based at v to G defined by

$$\omega \mapsto R(\omega)$$



Holonomy

Definition

The group element $R(\omega)$ will be called the *holonomy* of the loop ω. We say that the holonomy of ω is trivial if and only if $R(\omega) = 1$, the identity element of G.

Holonomy group

Lemma

For every vertex v, $H(v) = R(\Omega_v)$ is a subgroup of G, called the holonomy group at v. If the graph $\mathcal{G}$ is connected, for every two vertices v and w, the holonomy groups Ω_v and Ω_w are isomorphic.

Restriction to paths

For a path γ connecting a vertex v to a vertex w, let $V(\gamma)$ be the set of vertices along γ, and let $E(\gamma)$ denote the set of edges comprising γ.

If $f : V \to G$, then f_γ is the restriction of f to $V(\gamma)$. Similarly, ρ_γ is the restriction of ρ to $E(\gamma)$.

We say that f_γ is compatible with ρ_γ along γ if and only if for every edge $e \in E(\gamma)$, $e = (i, j)$, $f_\gamma(i) = \rho(i, j) f_\gamma(j)$.

It follows that

$$f_\gamma(v) = R(\gamma) f_\gamma(w).$$

Synchronisation along loops

Lemma

Let $R(\omega) = 1$ for a based loop ω. Then there exists an edge potential $f_\omega : V_\omega \to G$ along ω which is compatible with the edge potential ρ_ω along ω.

Synchronisation along loops

Lemma

Let $R(\omega) = 1$ for a based loop ω. Then there exists an edge potential $f_\omega : V_\omega \to G$ along ω which is compatible with the edge potential ρ_ω along ω.

Proof.

Let $\omega = (e_1, \ldots, e_n)$, where $i_1 = j_n = v$. We put $f_\omega(v) = 1$, and $f_\omega(j_k) = f(i_{k+1}) = \rho(e_1) \ldots \rho(e_k)$.

Then for every edge $e_k = (i_k, j_k)$, $f_\omega(i_k) = \rho(i_k, j_k) f(j_k)$.

Synchronisation along loops

Lemma

Let $R(\omega) = 1$ for a based loop ω. Then there exists an edge potential $f_\omega : V_\omega \to G$ along ω which is compatible with the edge potential ρ_ω along ω.

Proof.

Let $\omega = (e_1, \ldots, e_n)$, where $i_1 = j_n = v$. We put $f_\omega(v) = 1$, and $f_\omega(j_k) = f(i_{k+1}) = \rho(e_1) \ldots \rho(e_k)$.

Then for every edge $e_k = (i_k, j_k)$, $f_\omega(i_k) = \rho(i_k, j_k) f(j_k)$.

When $k = n$,

$$f_\omega(j_n) = \rho(e_1) \ldots \rho(e_n) = R(\omega) = 1$$

so

$$f_\omega(j_n) = f_\omega(i_1) = 1$$

Synchronisation and holonomy

Theorem (Gao-B-Mukherjee)

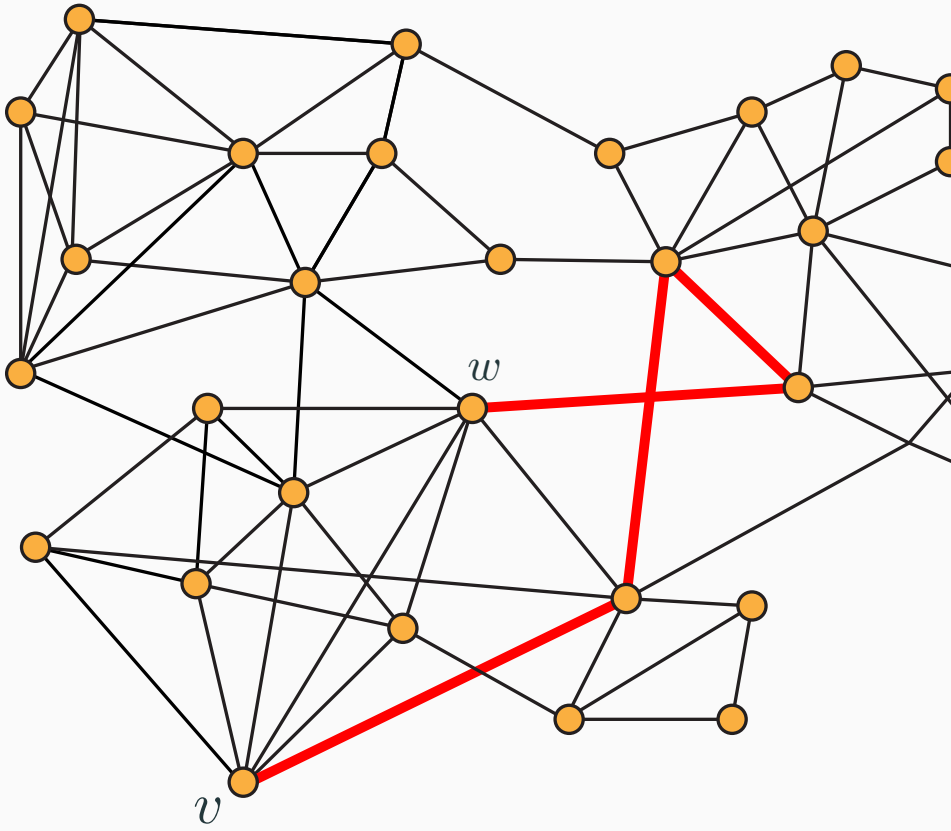
Let $\mathcal{G}(V, E)$ be a connected graph and let ρ be an edge potential on $\mathcal{G}$. Then there exists a vertex potential $f : V \to G$ compatible with ρ if and only if the holonomy group $H(v)$ of some vertex v is trivial. It follows that the holonomy group of every other vertex is also trivial, as is the holonomy of every loop in $\mathcal{G}$.

Basic idea

Assume that the holonomy of every loop is trivial. Choose a base vertex v and put $f(v) = 1$. Then for every vertex w define

$$f(w) = R(\gamma)$$

where γ is a path connecting v to w.

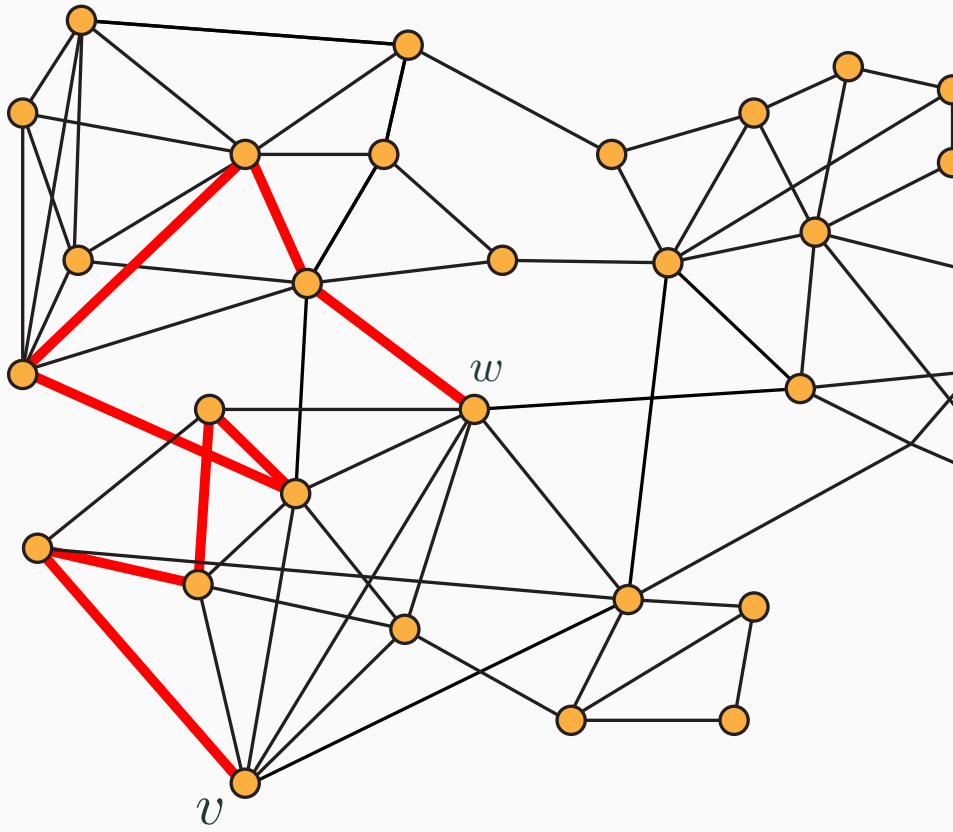


Basic idea

If γ_1 is a different path connecting v to w, then $\omega = \gamma \circ \gamma_1^{-1}$ is a loop based at v, so that

$$1 = R(\omega) = R(\gamma)R(\gamma_1)^{-1}$$

so the construction is independent of the path.

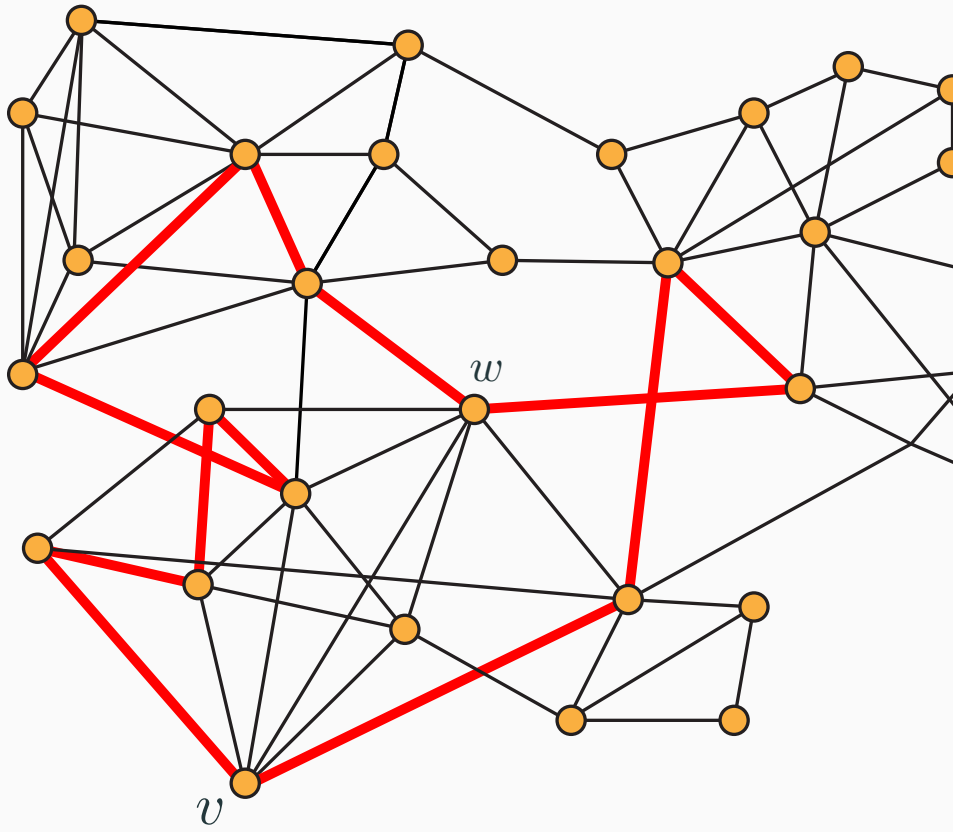


Basic idea

If γ_1 is a different path connecting v to w, then $\omega = \gamma \circ \gamma_1^{-1}$ is a loop based at v, so that

$$1 = R(\omega) = R(\gamma)R(\gamma_1)^{-1}$$

so the construction is independent of the path. The vertex potential f is compatible with ρ by construction.

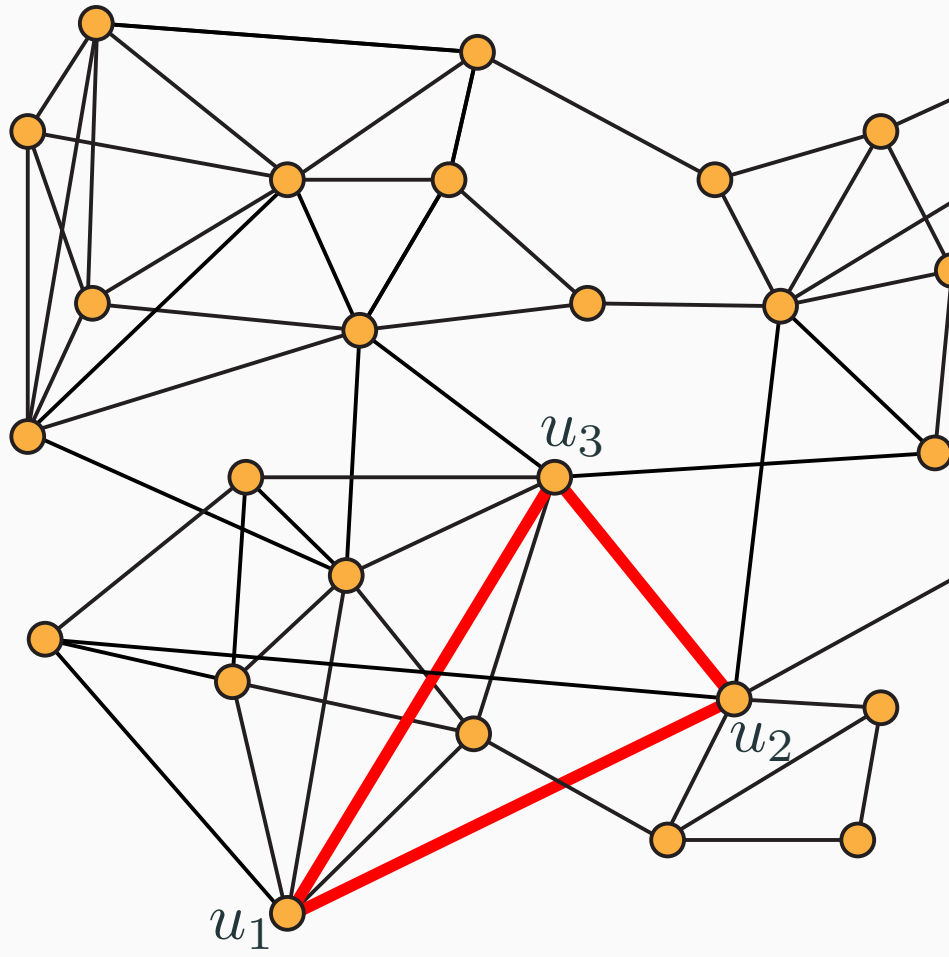


Corollary

Let $\rho \in P(E, G)$ be an edge potential for which the synchronization problem has a solution, that is, there exists a vertex potential $f \in P(V, G)$ compatible with ρ.

Then for every triangle (u_0, u_1, u_2) in $\mathcal{G}$

$$\rho(u_0, u_1)\rho(u_1, u_2)\rho(u_2, u_0) = 1.$$

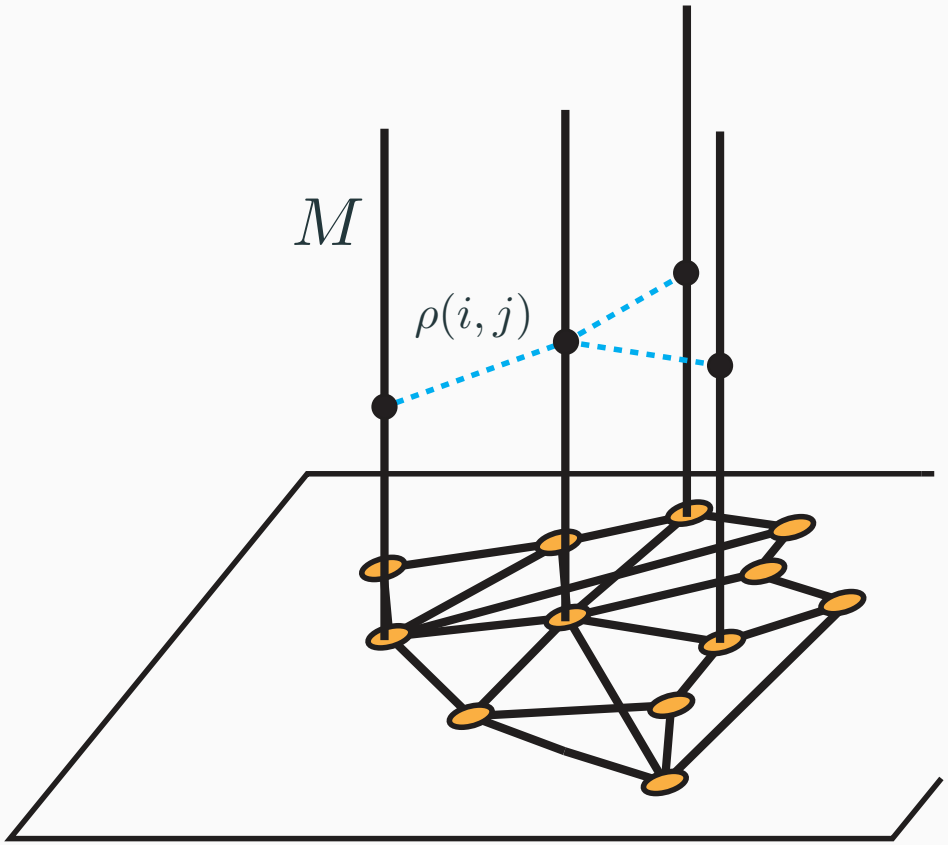


Cycle consistency

In summary, the triviality of holonomy groups for a symmetric edge potential implies the following well-known cycle consistency conditions:

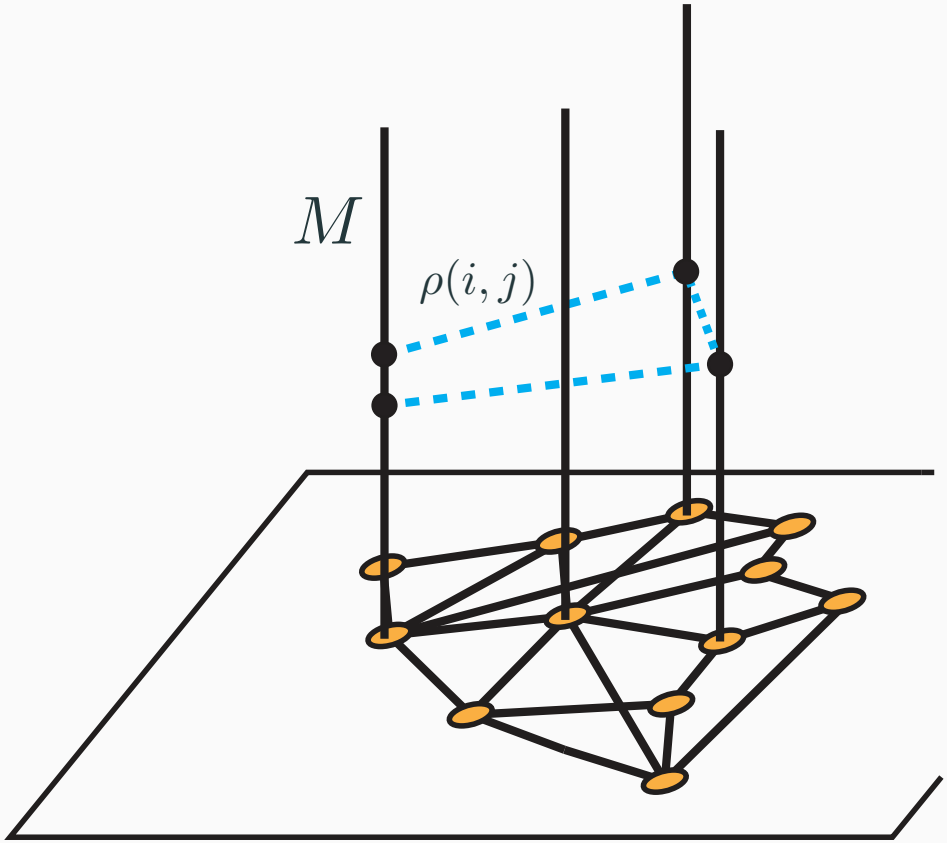
$$\begin{aligned} \rho(i,i) &= 1 \quad \text{for all } i \in V, \\ \rho(i,j)\rho(j,i) &= 1 \quad \text{for all } (i,j) \in E, \\ \rho(i,j)\rho(j,k)\rho(k,i) &= 1 \quad \text{if } (i,j),(j,k),(k,i) \in E. \end{aligned}$$

Fibre bundles and consistency conditions



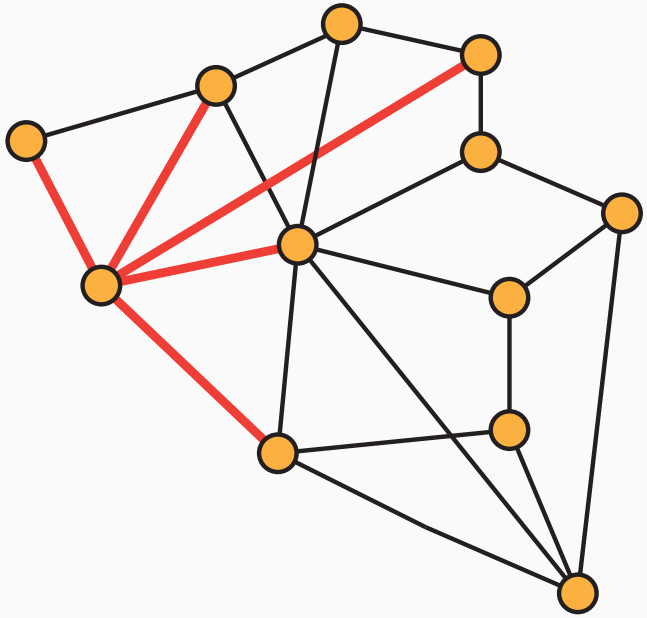
Theorem (Steenrod 1951, §2). If topological group G acts on Y and $\{U_i\}$, $\{\rho_{ij}\}$ is a system of coordinate transformations in the space X satisfying the cycle-consistency conditions then there exists a fibre bundle $\mathscr{B}$ with base space X, fibre Y, group G, and coordinate transforms $\{\rho_{ij}\}$.

Fibre bundles and consistency conditions



Non-trivial holonomy over a based loop in the base graph is represented as a non-trivial transformation of the fibre above the base vertex.

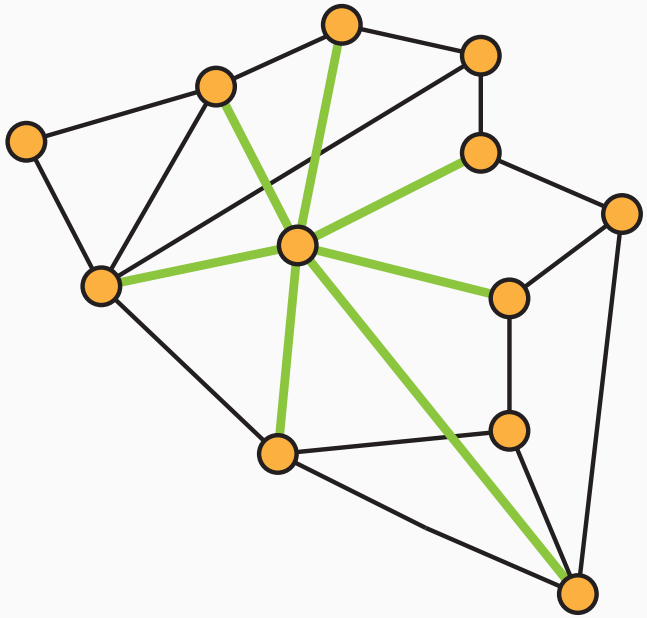
Bundle representation



Theorem (Gao-B-Mukherjee)

Let G be a topological group, $\mathcal{G} = (V, E)$ a connected graph, and $\rho : E \to G$ a map satisfying $\rho_{ij} = \rho_{ji}^{-1}$ for all $(i, j) \in E$. Write $\mathfrak{U} = \{U_i \mid 1 \leq i \leq |V|\}$ for an open cover of $\mathcal{G}$ in which U_i the open star of the vertex i

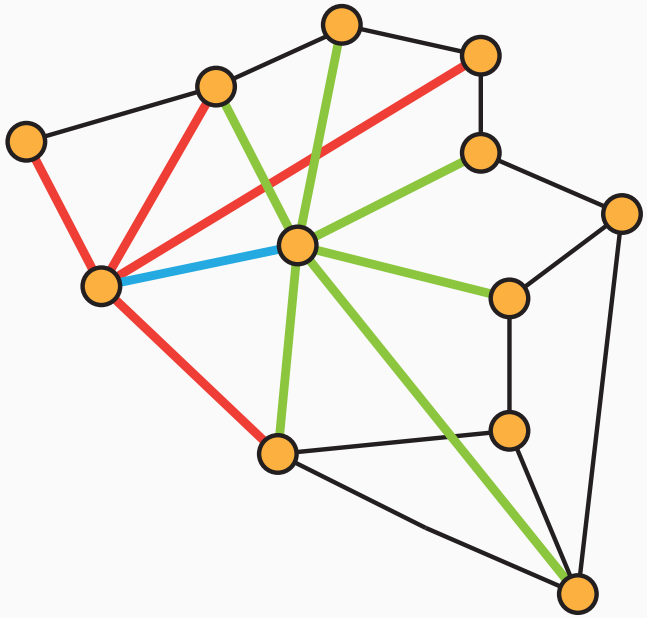
Bundle representation



Theorem (Gao-B-Mukherjee)

Let G be a topological group, $\mathcal{G} = (V, E)$ a connected graph, and $\rho : E \to G$ a map satisfying $\rho_{ij} = \rho_{ji}^{-1}$ for all $(i, j) \in E$. Write $\mathfrak{U} = \{U_i \mid 1 \leq i \leq |V|\}$ for an open cover of $\mathcal{G}$ in which U_i the open star of the vertex i

Bundle representation



Theorem (Gao-B-Mukherjee)
Then ρ defines a flat principal G-bundle $\mathcal{B}_\rho$ over $\mathcal{G}$ with a system of local trivializations defined on the open sets in $\mathfrak{U}$ with constant bundle transition functions ρ_{ij} on non-empty $U_i \cap U_j$. Furthermore, ρ is synchronizable if and only if $\mathcal{B}_\rho$ is a flat principal G-bundle.

Synchronisation bundle

The fibre bundle $\mathcal{B}_\rho$ associated with the connected graph $\mathcal{G}$ and edge potential ρ is called a *synchronization bundle associated with the edge potential ρ over $\mathcal{G}$*.

Synchronisation bundle

The fibre bundle $\mathcal{B}_\rho$ associated with the connected graph $\mathcal{G}$ and edge potential ρ is called a *synchronization bundle associated with the edge potential ρ over $\mathcal{G}$*.

Our theorem establishes a one-to-one correspondence between equivalence classes of synchronisable potentials and flat bundles on $\mathcal{G}$ (up to isomorphism).

Algorithm: SynCut

Input: $\mathcal{G} = (V, E, w)$, $\rho \in C^1(\mathcal{G}; G)$, number of partitions K

Output: Partitions $\{S_1, \cdots, S_K\}$

1. Solve synchronization problem over $\mathcal{G}$ for ρ, obtain $f \in C^0(\mathcal{G}; G)$
2. Compute $d_{ij} = \exp(-w_{ij} \|f_i - \rho_{ij} f_j\|)$ on all edges $(i, j) \in E$
3. Spectral clustering on weighted graph (V, E, d) to get $\{S_1, \cdots, S_k\}$
4. Solve synchronization problem within each partition S_j, "glue up" the local solutions to obtain $f_* \in C^0(\mathcal{G}; G)$
5. $f \leftarrow f_*$, repeat from Step 2

Tingran Gao, Jacek Brodzki, Sayan Mukherjee. "The Geometry of Synchronization Problems and Learning Group Actions," *submitted*, arXiv:1610.09051, 2016

Obstruction to synchronisation

Obstruction to synchronisation

The compatibility condition on edges is

$$f(i) = \rho(i,j)f(j)$$

To construct a useful way to measure the failure to satisfy this condition, we now consider a G-module F (equipped with an inner product), and F-valued functions $f : V \to F$ on the vertex set V.

We want to interpret the difference

$$f(i) - \rho(i,j)f(j).$$

Quantification of obstructions

Let F be a vector space such that $G \subset GL(F)$, and $f \in C^0(\mathcal{G}; F)$.

Quantification of obstructions

Let F be a vector space such that $G \subset GL(F)$, and $f \in C^0(\mathcal{G}; F)$.

Failure of synchronisation can be measured by frustration

$$\eta(f) = \frac{1}{2} \frac{\sum_{i,j \in V} \|f_i - \rho_{ij} f_j\|^2}{\sum_{i \in V} \|f_i\|^2},$$

which can be written as a Raleigh quotient

$$\eta(f) = \frac{\langle f, L_1 f \rangle}{\langle f, f \rangle}$$

where L_1 is the *Graph Connection Laplacian*

$$(L_1 f)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - \rho_{ij} f_j)$$

The standard graph Laplacian

Given a graph $\mathcal{G}(V, E)$ we define the gradient $d : \mathbb{C}[V] \to \mathbb{C}[E]$ by the formula

$$df(i,j) = f(i) - f(j).$$

The divergence operator $\delta : \mathbb{C}[E] \to \mathbb{C}[V]$ is given by

$$\delta\phi(i) = \sum_{j \sim i} \phi(i,j).$$

The graph Laplacian is the operator $\Delta : \mathbb{C}[V] \to \mathbb{C}[V]$ defined by $\Delta_\rho = \delta d$:

$$\Delta f(i) = \sum_{j : j \sim i} (f(i) - f(j))$$

Twisted differential

M is a G-module equipped with an inner product, and ρ a symmetric potential as before.

$$d_\rho : \mathbb{C}(V, M) \longrightarrow \mathbb{C}(E, M)$$

$$d_\rho f(i,j) = f(i) - \rho(i,j) f(j)$$

for every $f \in \mathbb{C}(V, M)$.

Twisted differential

M is a G-module equipped with an inner product, and ρ a symmetric potential as before.

$$d_\rho : \mathbb{C}(V, M) \longrightarrow \mathbb{C}(E, M)$$

$$d_\rho f(i,j) = f(i) - \rho(i,j) f(j)$$

for every $f \in \mathbb{C}(V, M)$.

A solution to the synchronisation problem is an element of the kernel of d_ρ.

Twisted Laplacian

The usual divergence:

$$\delta : \mathbb{C}(E, M) \longrightarrow \mathbb{C}(V, M)$$

$$\delta\phi(i) = \sum_{j:j\sim i} \phi(i, j)$$

together with d_ρ give a twisted Laplacian

$$\Delta_\rho = \delta d_\rho = \sum_{j:j\sim i} (f(i) - \rho(i, j)f(j))$$

The Raleigh quotient

$$\eta_\rho(f) = \frac{\langle f, \Delta_\rho f \rangle}{\langle f, f \rangle} = \frac{\sum_{(i,j)\in E} w_{ij} \left\| f(i) - \rho(i,j) f(j) \right\|^2}{\sum_{i \in V} d_i \left\| f(i) \right\|^2}$$

$\eta_\rho(f)$ is defined in Bandeira et al. (2013) as the *frustration*

Synchronisation and cohomology

We have defined a simple de Rham-type complex

$$\mathcal{C} : 0 \longrightarrow \mathbb{C}(V, M) \xrightarrow{d_\rho} \mathbb{C}_\rho(E, M) \longrightarrow 0.$$

The cohomology of this complex is

$$H^0(\mathcal{C}, d) = \ker d_\rho$$

in degree zero. In degree one, we have that

$$H^1(\mathcal{C}, d) = \mathbb{C}(E, M)/d_\rho(\mathbb{C}(V, M)).$$

Thus solutions to the synchronisation problem form the zeroth cohomology group, which is the same as the kernel of d_ρ.

Harmonic forms

Theorem

1. *The space of solutions of the synchronisation problem given by a unitary edge potential ρ is isomorphic to the space of harmonic functions f, i.e., functions f with the property $\Delta_\rho f = 0$. Moreover, we have the following orthogonal decomposition:*

$$\mathbb{C}(V, M) = \ker d_\rho \oplus \operatorname{Im} \delta.$$

2. *The Laplace operator Δ_ρ is self-adjoint and positive.*

Inner products

The two linear spaces $\Omega^0(\mathcal{G}, M)$ and $\Omega^1_\rho(\mathcal{G}, M)$ are equipped with inner products naturally induced from the G-invariant inner product on M:

$$\langle f, g \rangle = \sum_{i \in V} d_i \langle f_i, g_i \rangle, \quad \forall f, g \in \Omega^0(\mathcal{G}, M),$$

$$\langle \phi, \psi \rangle = \sum_{(i,j) \in E} w_{ij} \langle \phi_{ij}, \psi_{ij} \rangle, \quad \forall \phi, \psi \in \Omega^1(\mathcal{G}, M),$$

Formal adjoints

With respect to these inner products, the twisted differential d_ρ and the divergence δ are adjoints of each other.

Learning group actions

Given a group G acting on a set X, simultaneously learn a new action of G on X and a partition of X into disjoint subsets $X_1, \cdots, X_K$, such that the new action is as close as possible to the given action and cycle-consistent on each X_i $(1 \leq i \leq K)$.

Discrete Hodge theory

$\mathcal{G} = (V, E)$:

Spaces of cochains:

$$\Omega^0(\mathcal{G}) := \{f : V \to \mathbb{K}\},$$
$$\Omega^1(\mathcal{G}) := \{\omega : E \to \mathbb{K} \mid \omega_{ij} = -\omega_{ji} \ \forall i \sim j\},$$

Discrete Hodge theory

$\mathcal{G} = (V, E)$:

Spaces of cochains:

$$\Omega^0(\mathcal{G}) := \{f : V \to \mathbb{K}\},$$
$$\Omega^1(\mathcal{G}) := \{\omega : E \to \mathbb{K} \mid \omega_{ij} = -\omega_{ji} \ \forall i \sim j\},$$

A cochain complex

$$0 \rightleftarrows \Omega^0(\mathcal{G}) \underset{\delta}{\overset{d}{\rightleftarrows}} \Omega^1(\mathcal{G}) \rightleftarrows 0,$$

where

$$(df)_{ij} = f_i - f_j, \quad \forall f \in \Omega^0(\mathcal{G}),$$
$$(\delta\omega)_i = \frac{1}{\deg(i)} \sum_{j \sim i} \omega_{ij}, \quad \forall \omega \in \Omega^1(\mathcal{G}).$$

Discrete Hodge theory

The standard cochain complex

$$0 \rightleftarrows \Omega^0(\mathcal{G}) \underset{\delta}{\overset{d}{\rightleftarrows}} \Omega^1(\mathcal{G}) \rightleftarrows 0,$$

where

$$(df)_{ij} = f_i - f_j, \quad \forall f \in \Omega^0(\mathcal{G}),$$
$$(\delta\omega)_i = \frac{1}{\deg(i)} \sum_{j \sim i} \omega_{ij}, \quad \forall \omega \in \Omega^1(\mathcal{G}).$$

The graph Laplacian:

$$\begin{aligned}(L_0 f)_i &:= (\delta d f)_i \\ &= \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - f_j) \quad \forall i \in V, \ \forall f \in \Omega^0(\mathcal{G}).\end{aligned}$$

Twisted De Rham-Hodge Theory

$$(L_0 f)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - f_j), \quad \forall f : V \to \mathbb{K}$$

$$(L_1 f)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - \rho_{ij} f_j) \quad \forall f : V \to F$$

Twisted De Rham-Hodge Theory

$$(L_0 f)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - f_j), \quad \forall f : V \to \mathbb{K}$$

$$(L_1 f)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - \rho_{ij} f_j) \quad \forall f : V \to F$$

Naïvely:

$$(d_\rho f)_{ij} = f_i - \rho_{ij} f_j, \quad \forall f \in C^0(\mathcal{G}; F)$$

$$(\delta_\rho \omega)_i = \frac{1}{\deg(i)} \sum_{j \sim i} \omega_{ij}, \quad \forall \omega \in C^1(\mathcal{G}; F)$$

then $L_1 = \delta_\rho d_\rho$.

We have defined:

$$(d_\rho f)_{ij} \sim f_i - \rho_{ij} f_j, \quad \forall f \in C^0(\mathcal{G}; F)$$
$$(\delta_\rho \omega)_i \sim \frac{1}{\deg(i)} \sum_{j \sim i} \omega_{ij}, \quad \forall \omega \in C^1(\mathcal{G}; F)$$

There is a problem: d_ρ does not map into $C^1(\mathcal{G}; F)$ (no skew-symmetry).

$$f_j - \rho_{ji} f_j = -\rho_{ji}(f_i - \rho_{ij} f_j) \neq -(f_i - \rho_{ij} f_j).$$

We have defined:

$$(d_\rho f)_{ij} \sim f_i - \rho_{ij} f_j, \quad \forall f \in C^0(\mathcal{G}; F)$$
$$(\delta_\rho \omega)_i \sim \frac{1}{\deg(i)} \sum_{j \sim i} \omega_{ij}, \quad \forall \omega \in C^1(\mathcal{G}; F)$$

There is a problem: d_ρ does not map into $C^1(\mathcal{G}; F)$ (no skew-symmetry).

$$f_j - \rho_{ji} f_j = -\rho_{ji}(f_i - \rho_{ij} f_j) \neq -(f_i - \rho_{ij} f_j).$$

Fix: Interpret $f_i - \rho_{ij} f_j$ as the "local expression" of $(d_\rho f)_{ij}$ in a local trivialization over $\mathfrak{U} = \{U_i \mid 1 \leq i \leq |V|\}$ of the associated F-bundle of $\mathscr{B}_\rho$, denoted as $\mathscr{B}_\rho[F]$, such that the extra ρ_{ji} factor encodes a bundle transformation from U_i to U_j.

Twisted de Rham-Hodge Theory

- Combinatorial Hodge Theory:

$$0 \rightleftarrows \Omega^0(\mathcal{G}) \underset{\delta}{\overset{d}{\rightleftarrows}} \Omega^1(\mathcal{G}) \rightleftarrows 0,$$

- *Twisted* Combinatorial Hodge Theory:

$$0 \rightleftarrows C^0(\mathcal{G}; F) \underset{\delta_\rho}{\overset{d_\rho}{\rightleftarrows}} \Omega^1(\mathcal{G}; \mathscr{B}_\rho[F]) \rightleftarrows 0.$$

Twisted de Rham-Hodge Theory

Theorem (Gao, B, Mukherjee (2016)))

Define

$$\Delta_\rho^{(0)} := \delta_\rho d_\rho, \quad \Delta_\rho^{(1)} := d_\rho \delta_\rho$$

then the following Hodge-type decomposition holds:

$$C^0(\mathcal{G}; F) = \ker \Delta_\rho^{(0)} \oplus \operatorname{Im} \delta_\rho = \ker d_\rho \oplus \operatorname{Im} \delta_\rho,$$
$$\Omega^1(\mathcal{G}; \mathscr{B}_\rho[F]) = \operatorname{Im} d_\rho \oplus \ker \Delta_\rho^{(1)} = \operatorname{Im} d_\rho \oplus \ker \delta_\rho.$$

Solutions to the synchronisation problem are elements of $\ker d_\rho$.

Learning group actions

Learning group actions

Given a group G acting on a set X, simultaneously learn a new action of G on X and a partition of X into disjoint subsets $X_1, \cdots, X_K$, such that

- the new action is as close as possible to the given action;
- and is cycle-consistent on each X_i $(1 \leq i \leq K)$.

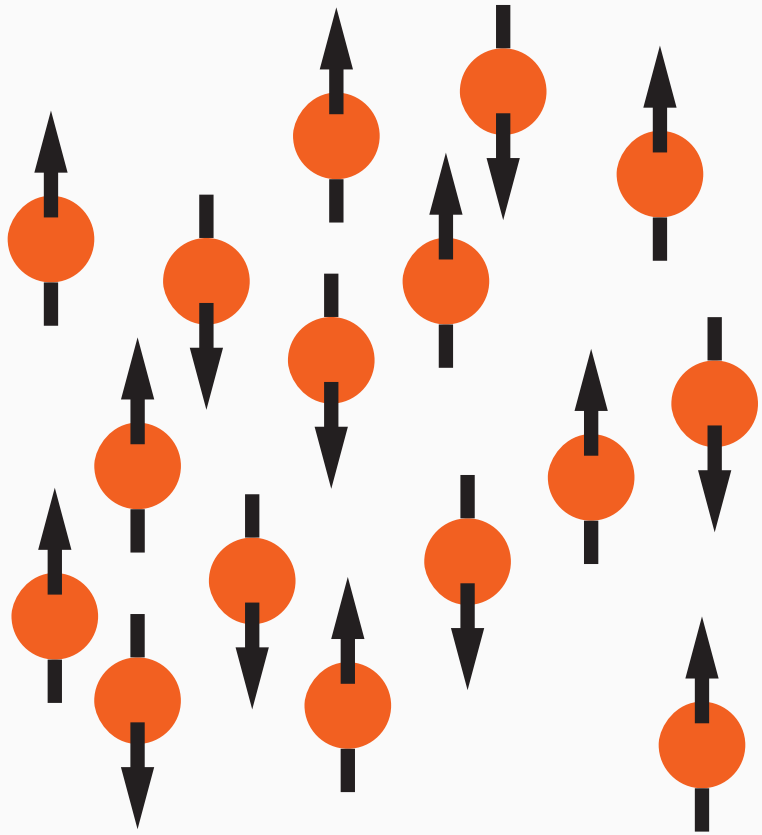
Example: Representations

If the set X is a vector space and we seek a direct sum decomposition $X = \bigoplus_{i=1}^{K} X_i$, the LGA problem reduces to the search for all irreducible G-subrepresentations of X.

Example: Spins

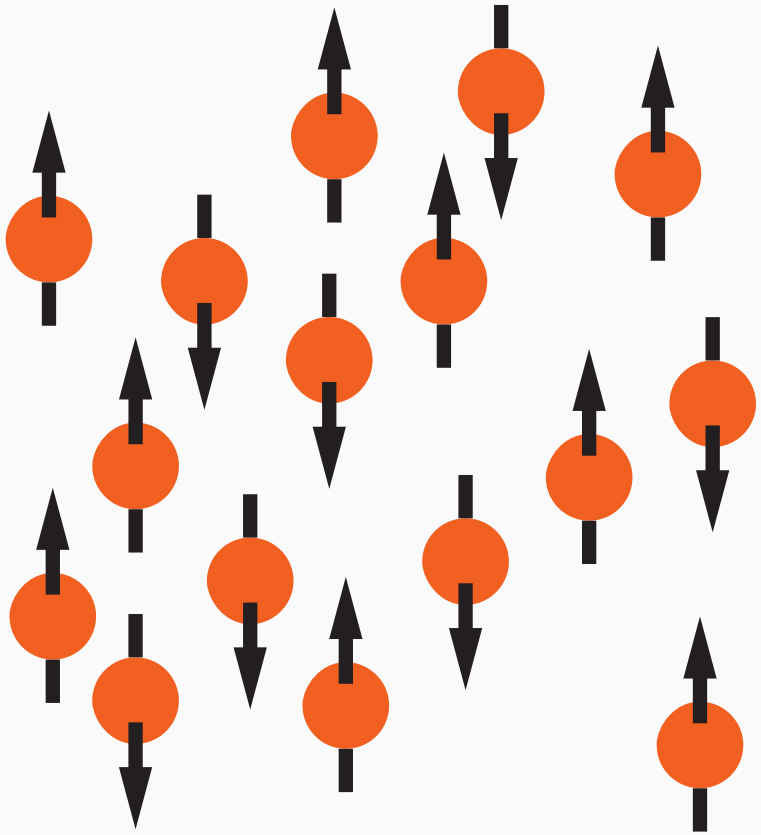
$X = \{x_1, \cdots, x_n\}$ equipped with $S : X \to \{\pm 1\}$.

Let $G = \{\pm 1\}$ act on X transitively as $(g_{ji}, x_i) \mapsto x_j$, $g_{ji} = S(x_j)S(x_i)$.



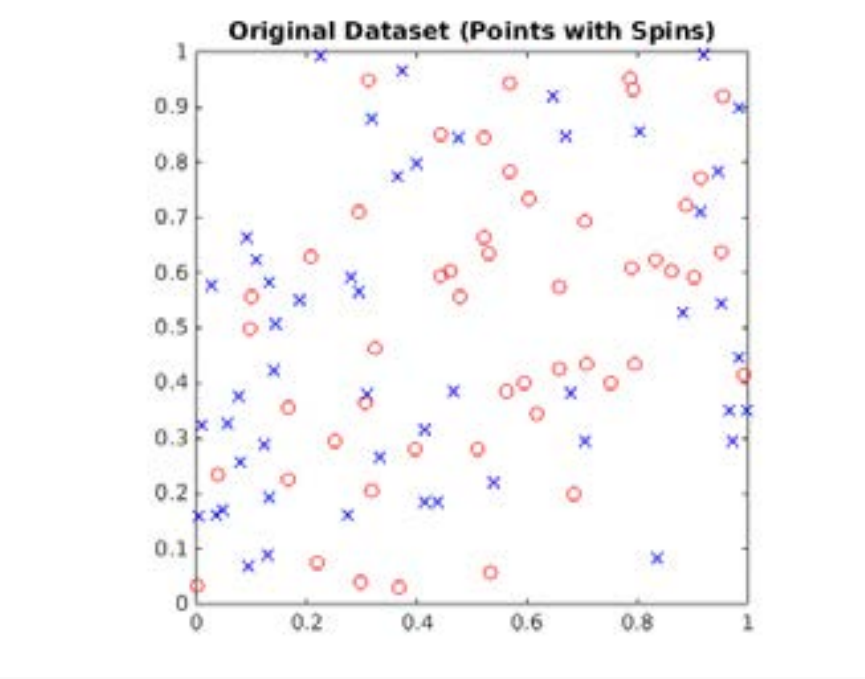
Example: Spins

Suppose the spin of each point in X (i.e. the label map S) is unknown, but we know the group actions $\{g_{ij}\}$
Then we can reconstruct S — by spectral clustering the dataset X, viewed as vertices of a complete graph $\mathcal{G}$ with weight $w_{ij} = g_{ij}$ on the edge connecting x_i and x_j.

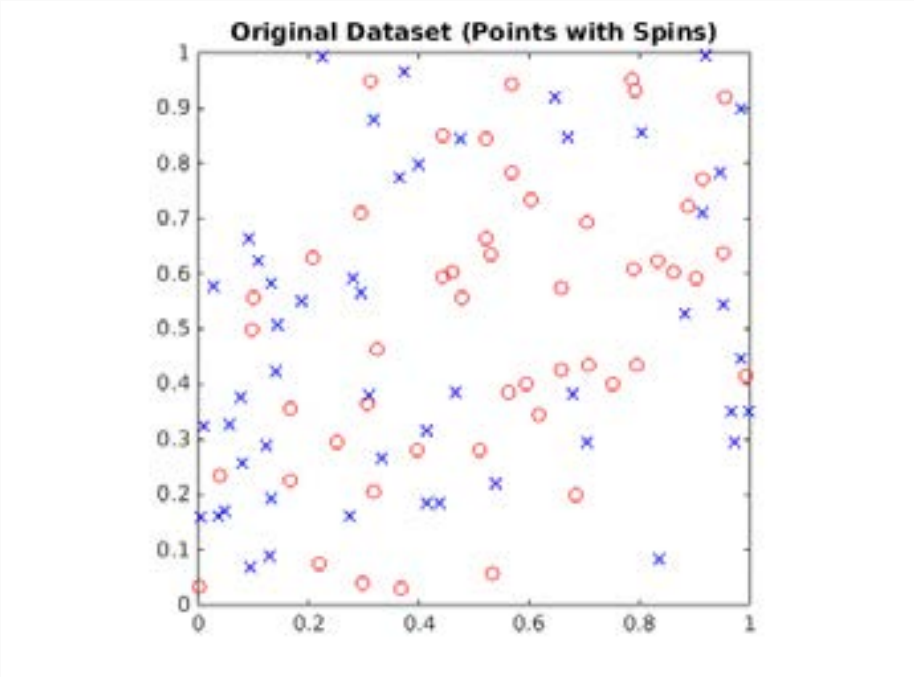


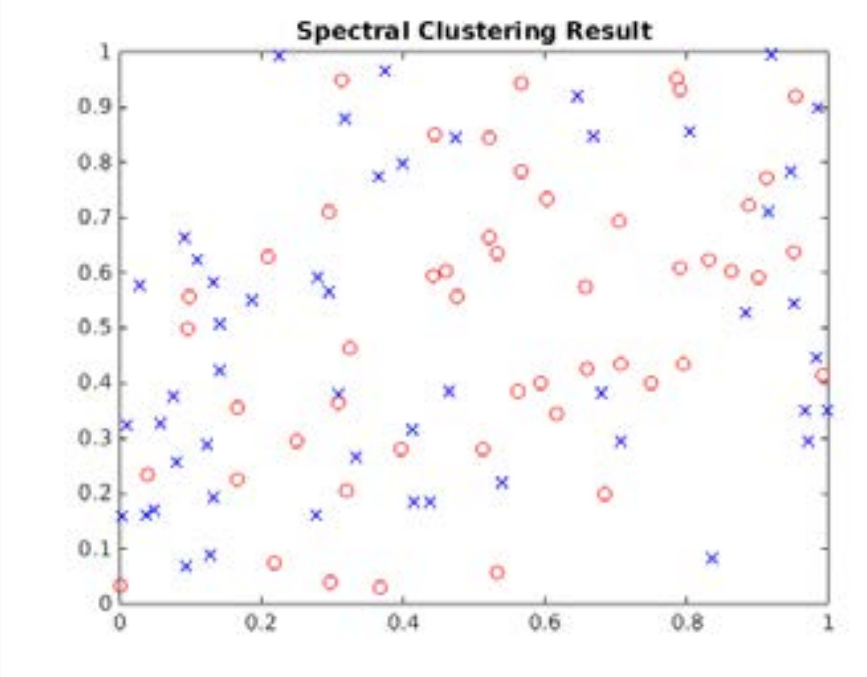
Example: Spins

Suppose the spin of each point in X (i.e. the label map S) is unknown, but we know the group actions $\{g_{ij}\}$
Then we can reconstruct S — by spectral clustering the dataset X, viewed as vertices of a complete graph $\mathcal{G}$ with weight $w_{ij} = g_{ij}$ on the edge connecting x_i and x_j.

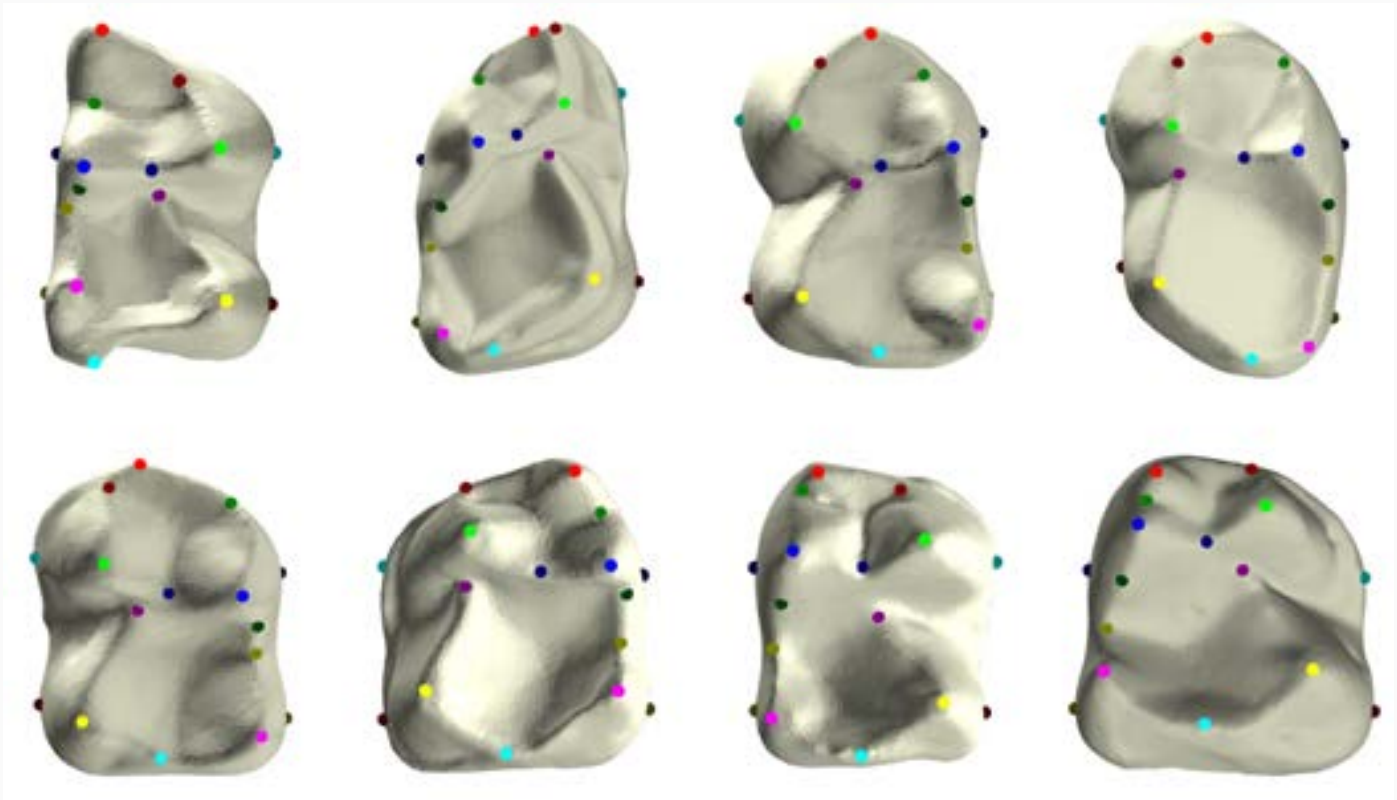


Example: Spins

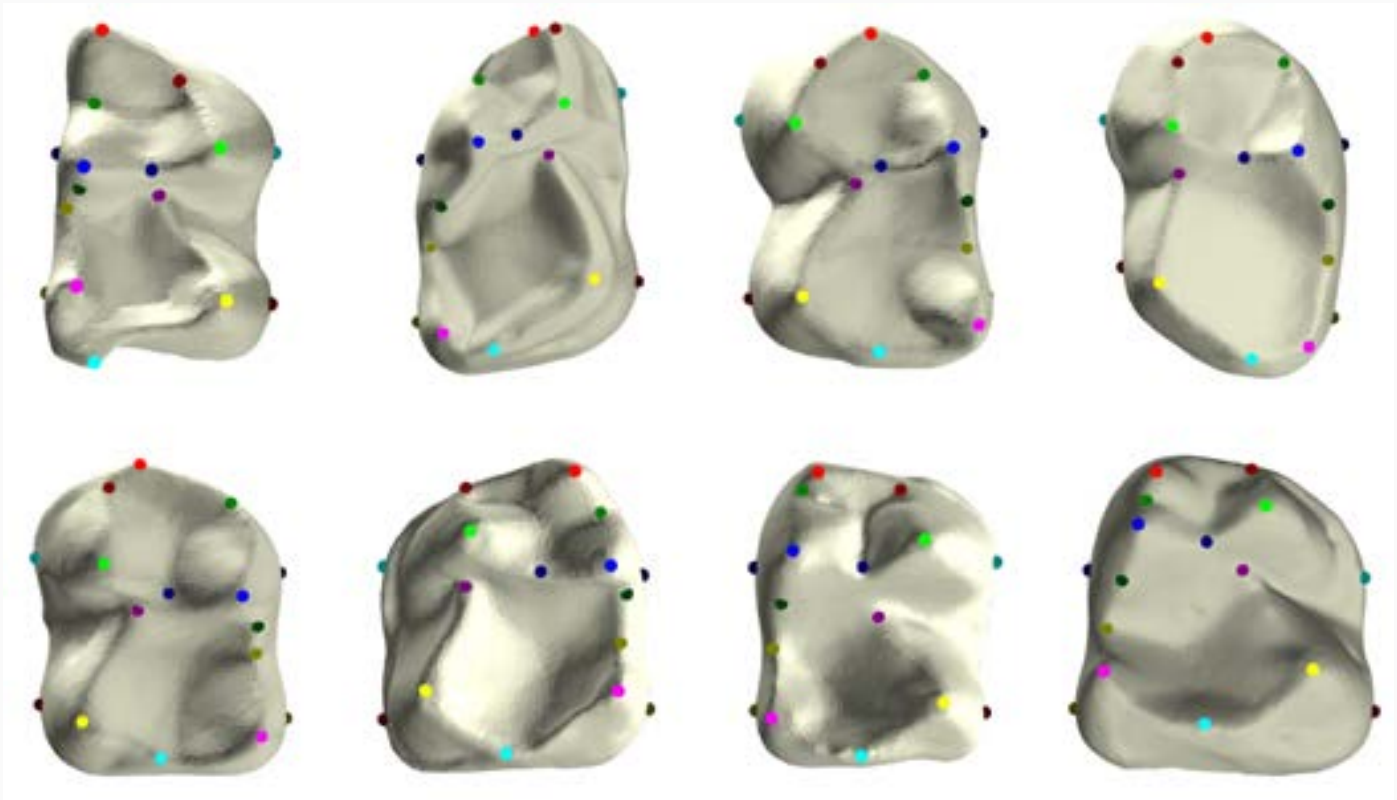




Lemurs

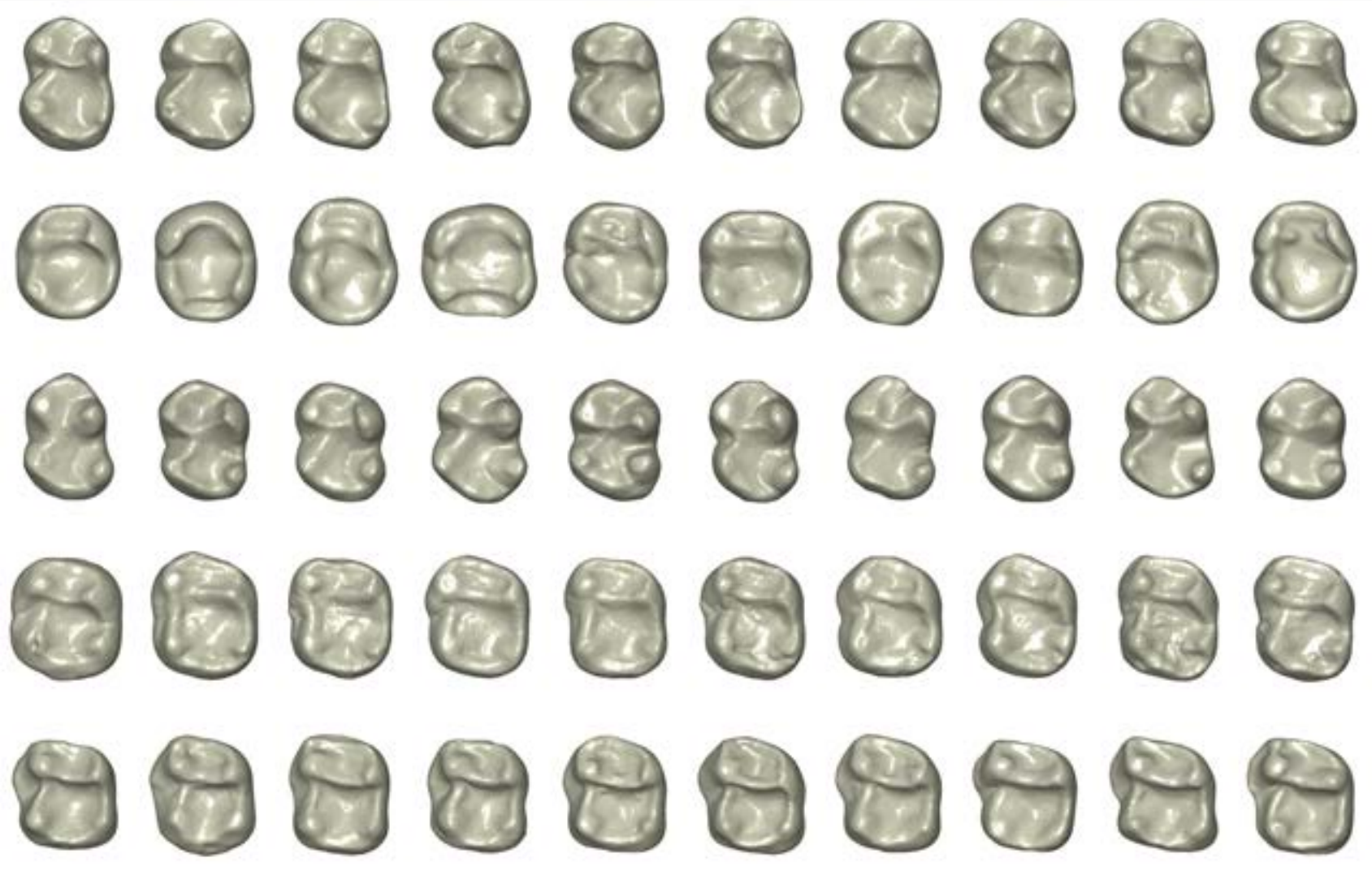


Lemurs

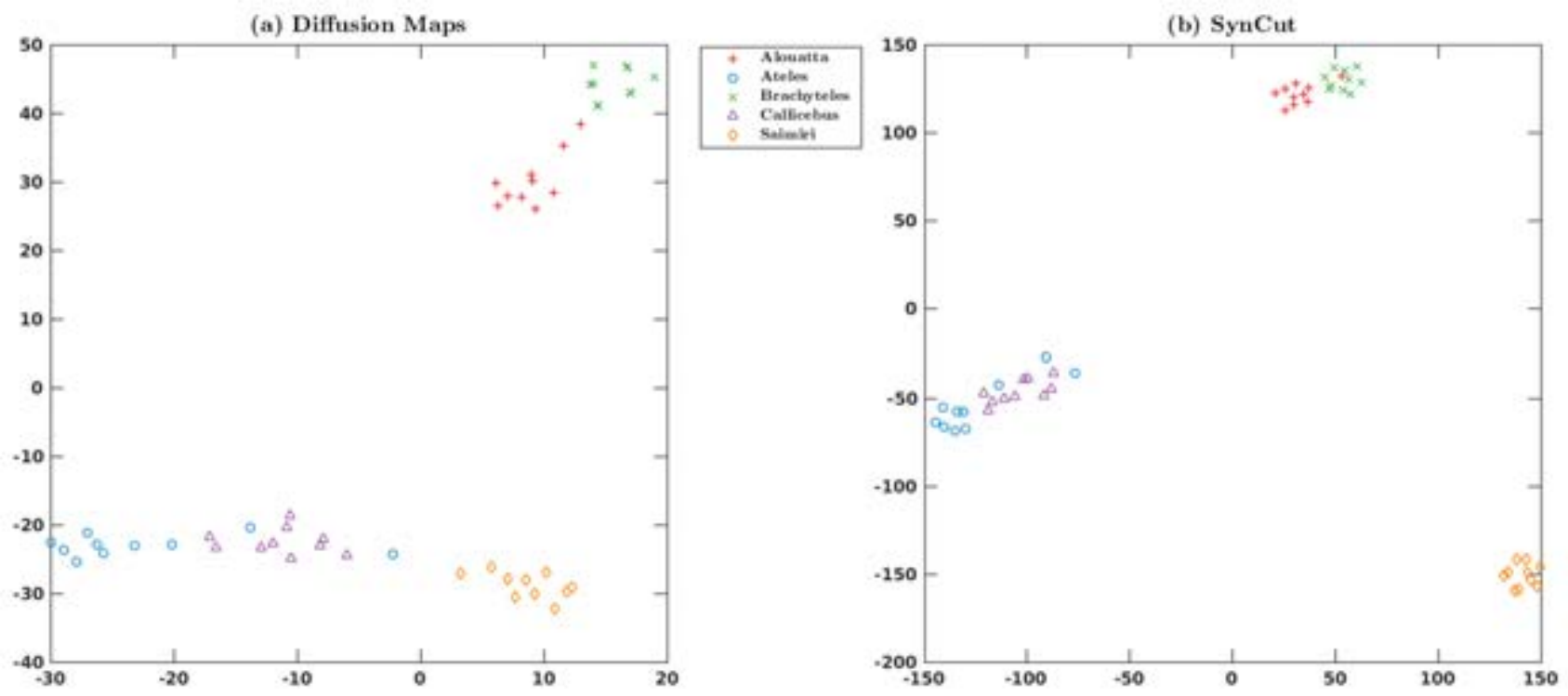


If we restrict attention to the determinant of the transformation providing the best fit then we discover that teeth 1, 2, 7, 8 belong to one cluster (right), while 3, 4, 5, 6 belong to the other cluster (left).

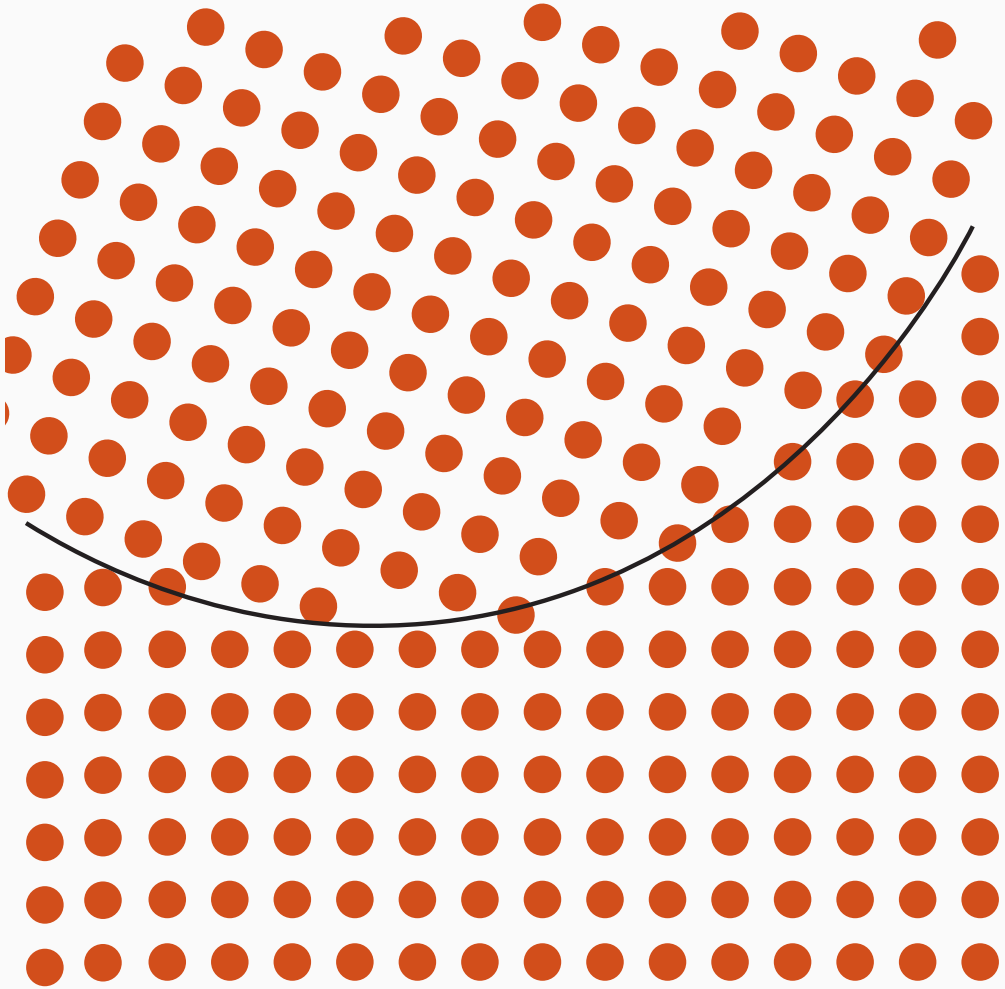
Identifying primate genera



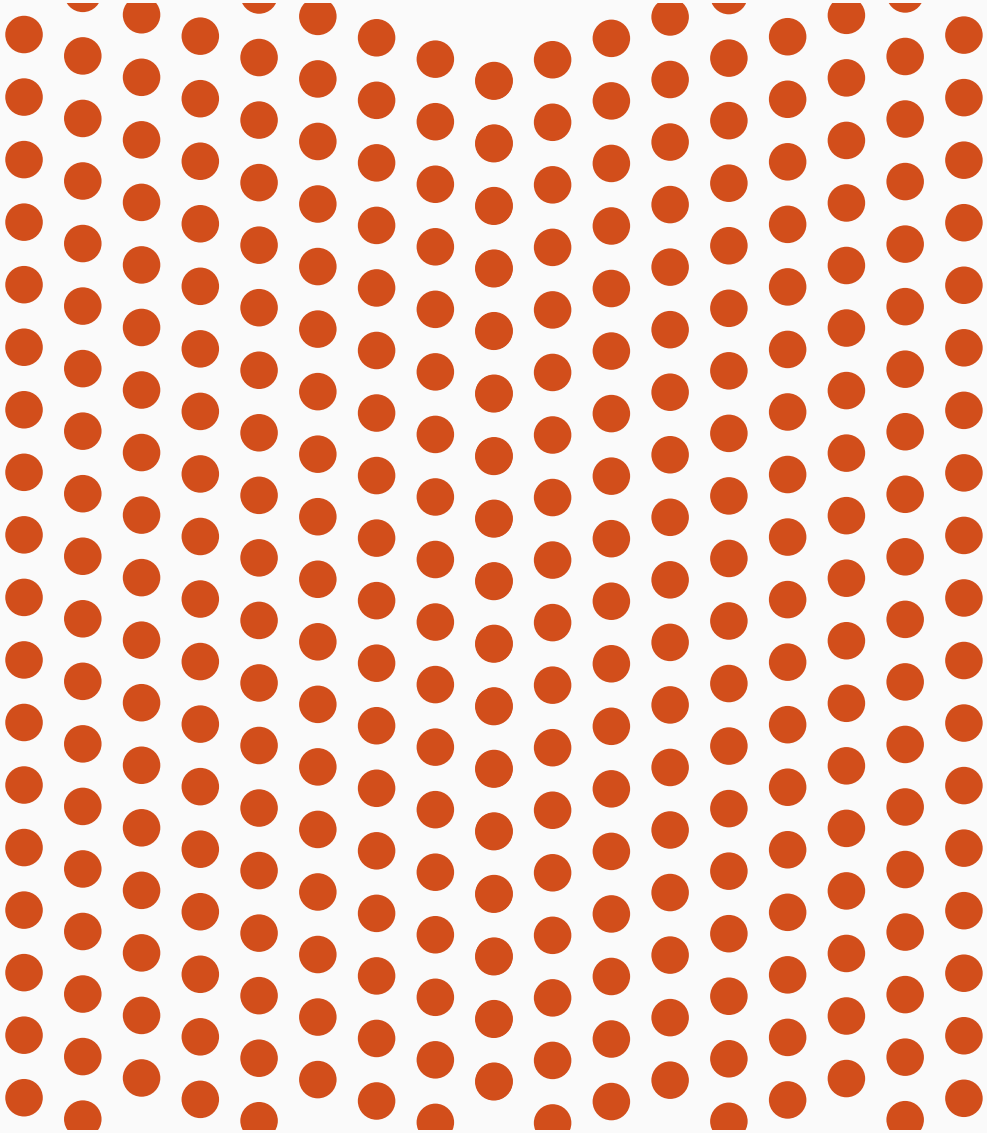
Classification into folivorous, frugivorous, and insectivorous



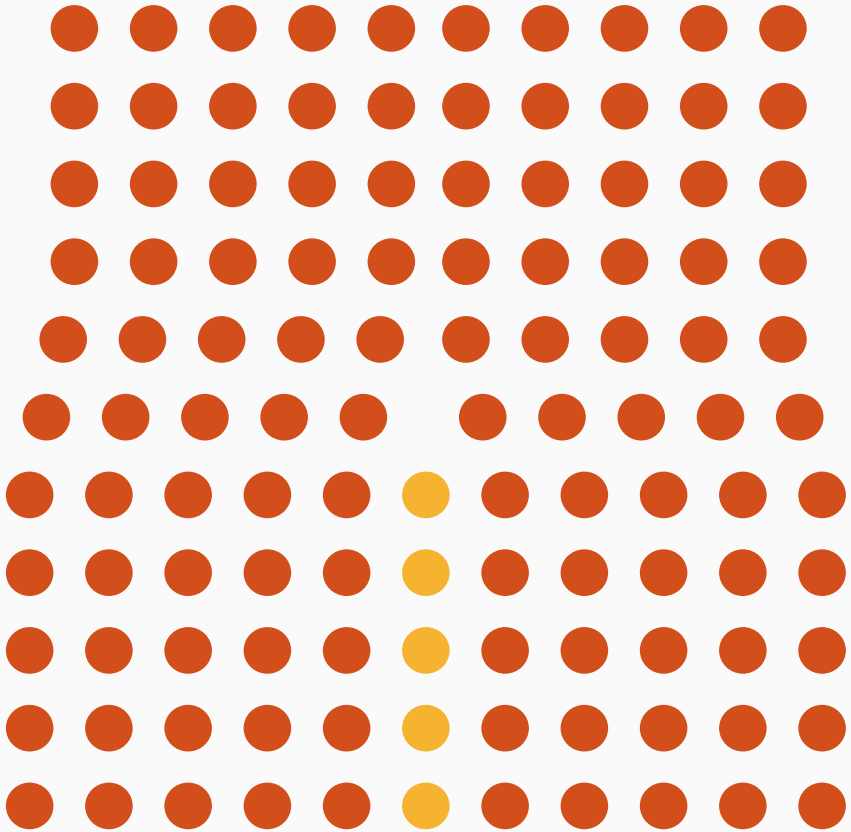
Extensions: defects in crystals



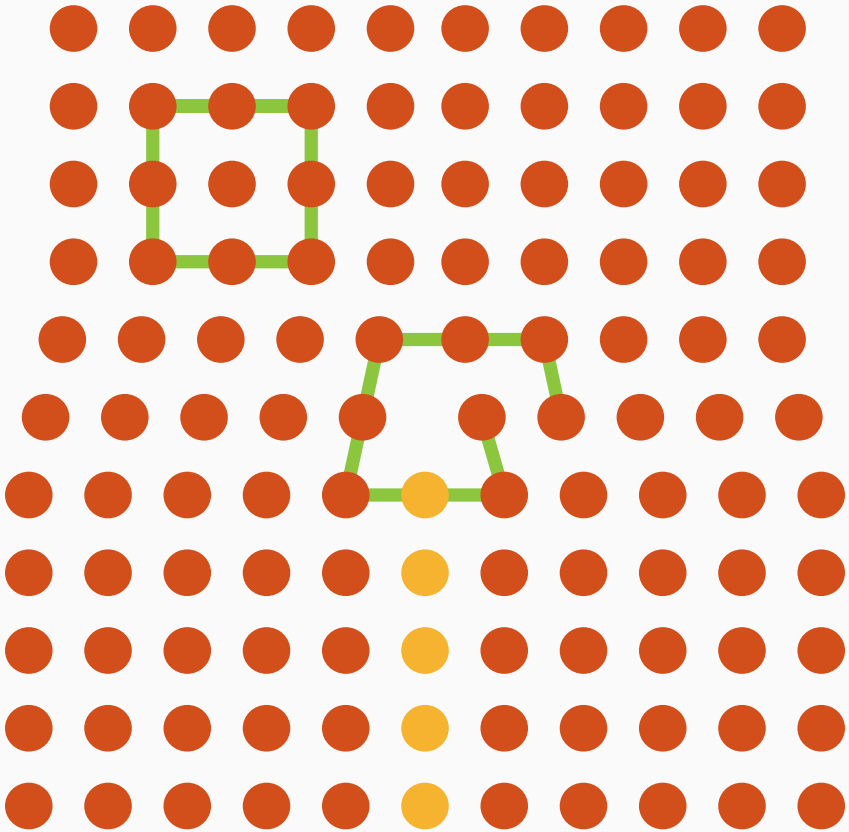
Extensions: defects in crystals



Extensions: defects in crystals

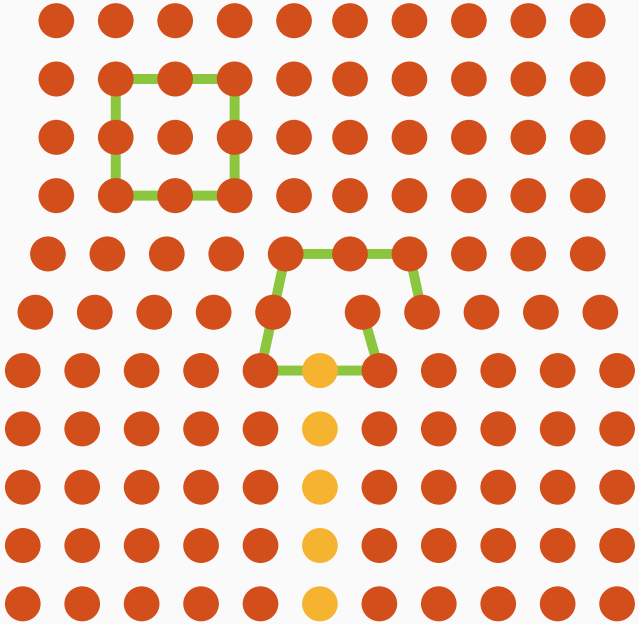


Extensions: defects in crystals



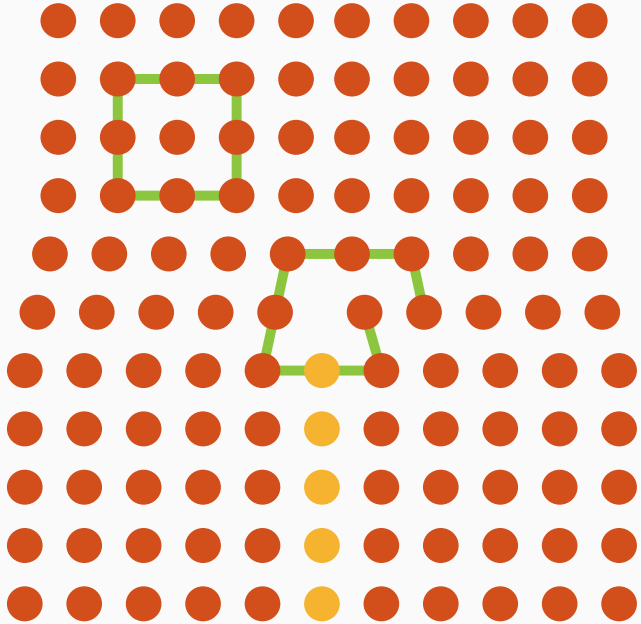
Extensions: defects in crystals

The local structure of the defects is encoded in a comparison map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ which compares an ideal crystal to one with defects.



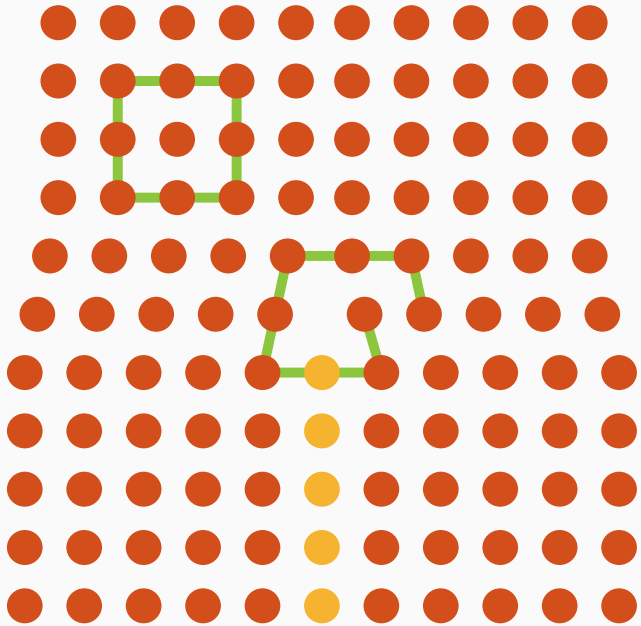
Extensions: defects in crystals

The local structure of the defects is encoded in a comparison map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ which compares an ideal crystal to one with defects.
Note that ϕ is only determined up to action of point groups which determine symmetries of the crystal.



Extensions: defects in crystals

The local structure of the defects is encoded in a comparison map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ which compares an ideal crystal to one with defects.
Note that ϕ is only determined up to action of point groups which determine symmetries of the crystal.
The geometry of a planar crystal is determined by a a two-dimensional de Rham complex.



The shape of lungs in COPD

Rerefence

The shape of lungs in chronic obstructive pulmonary disease, 2017

F. Belchi Guillamon, M. Pirashvili, M. Bennett, J. Conway,
R. Djukanović, J. Brodzki

Scientific Reports, 28 March 2018, supported by the EPSRC
Joining the Dots programme

COPD

- Chronic obstructive pulmonary disease (COPD) is a common progressive disease, affecting more than 200 million individuals worldwide and is the third leading cause of death.

COPD

- Chronic obstructive pulmonary disease (COPD) is a common progressive disease, affecting more than 200 million individuals worldwide and is the third leading cause of death.
- It is characterized by chronic inflammation of the bronchi and the lung parenchyma, resulting in varying degrees of obstructive bronchitis and emphysema

COPD

- Chronic obstructive pulmonary disease (COPD) is a common progressive disease, affecting more than 200 million individuals worldwide and is the third leading cause of death.
- It is characterized by chronic inflammation of the bronchi and the lung parenchyma, resulting in varying degrees of obstructive bronchitis and emphysema
- Although its pathology is heterogeneous, in functional terms, all forms of COPD result in loss of lung function

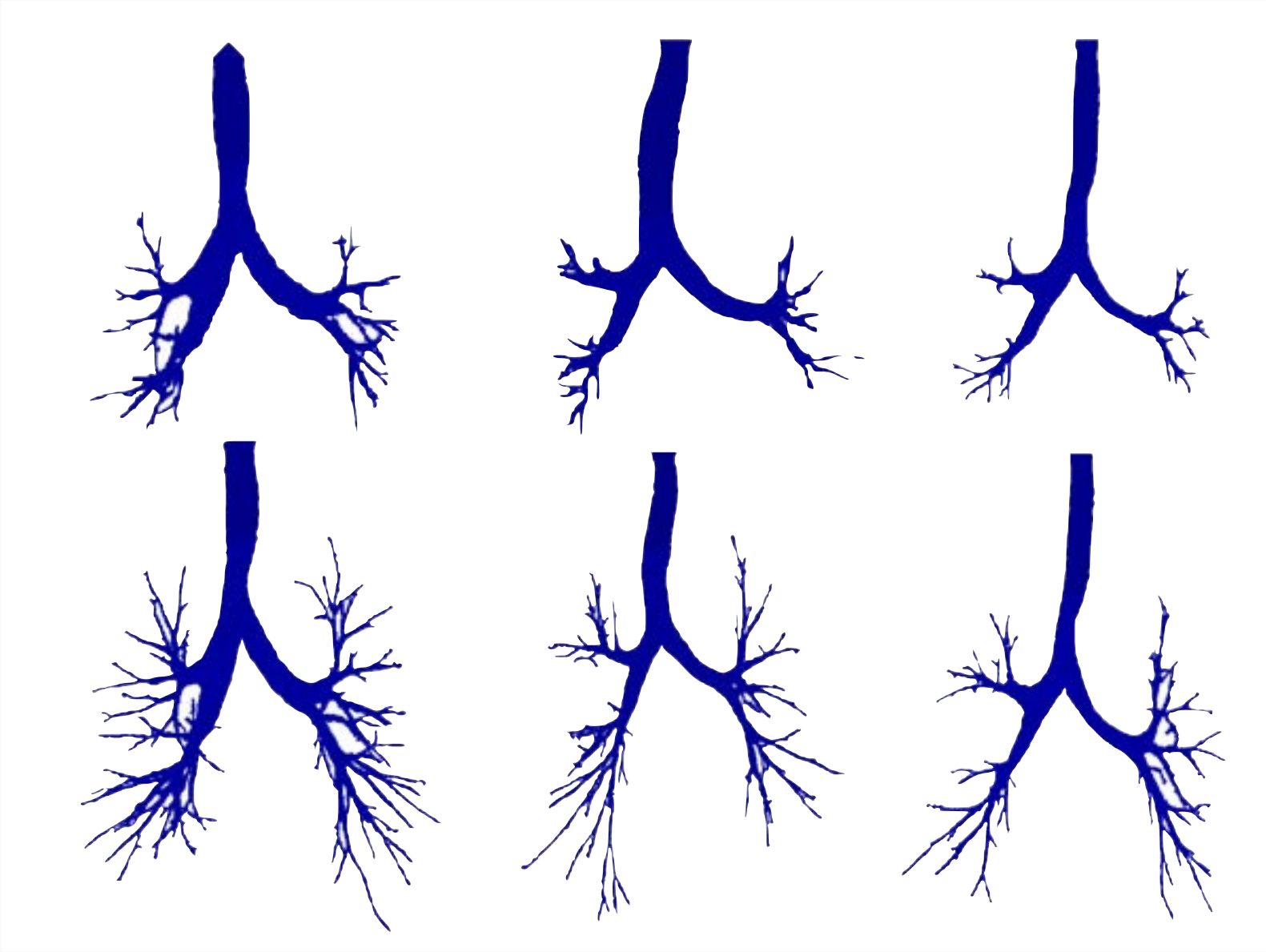
COPD

- Chronic obstructive pulmonary disease (COPD) is a common progressive disease, affecting more than 200 million individuals worldwide and is the third leading cause of death.
- It is characterized by chronic inflammation of the bronchi and the lung parenchyma, resulting in varying degrees of obstructive bronchitis and emphysema
- Although its pathology is heterogeneous, in functional terms, all forms of COPD result in loss of lung function
- Cigarette smoking is the main COPD risk factor, but it is quite frequent among people who never smoked, and many smokers do not develop COPD.

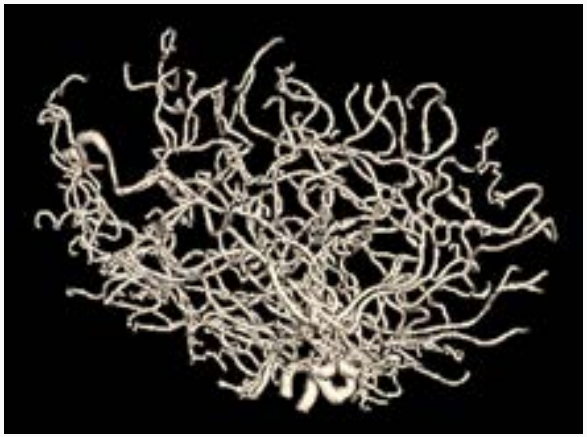
COPD

- Chronic obstructive pulmonary disease (COPD) is a common progressive disease, affecting more than 200 million individuals worldwide and is the third leading cause of death.
- It is characterized by chronic inflammation of the bronchi and the lung parenchyma, resulting in varying degrees of obstructive bronchitis and emphysema
- Although its pathology is heterogeneous, in functional terms, all forms of COPD result in loss of lung function
- Cigarette smoking is the main COPD risk factor, but it is quite frequent among people who never smoked, and many smokers do not develop COPD.
- Many adults who develop COPD have low lung function in the early life.

CT scans

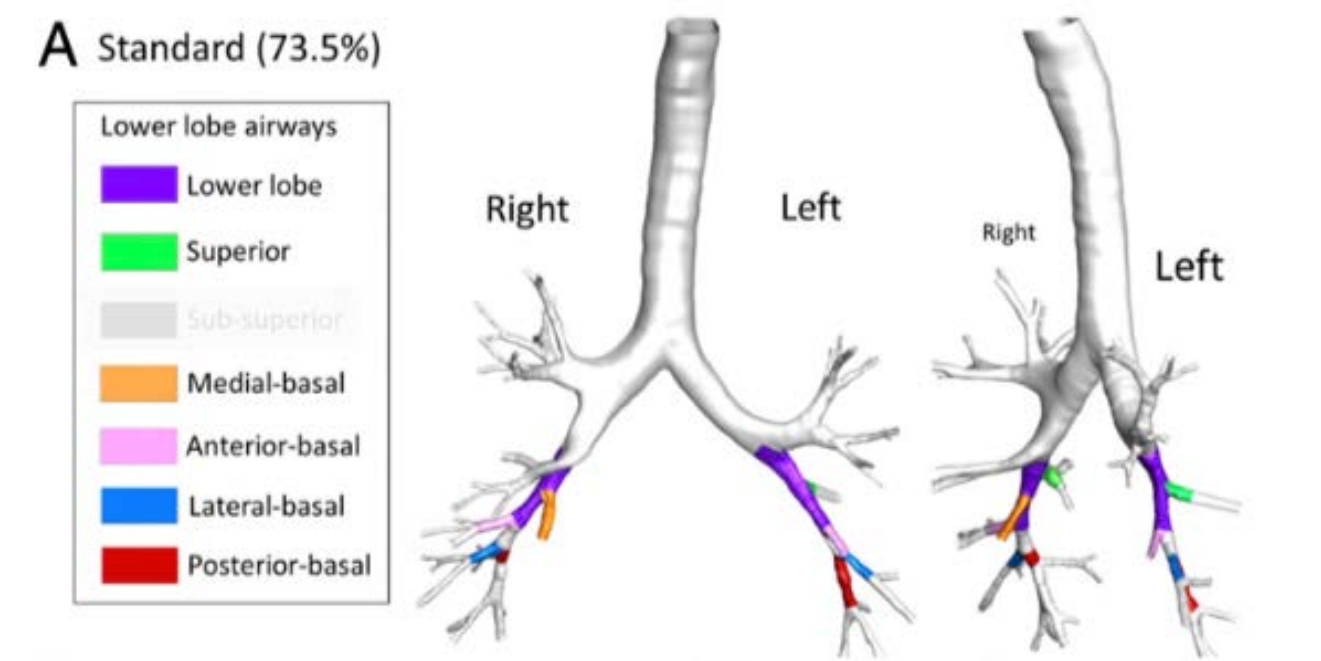


Related work: Arterial trees in the brain



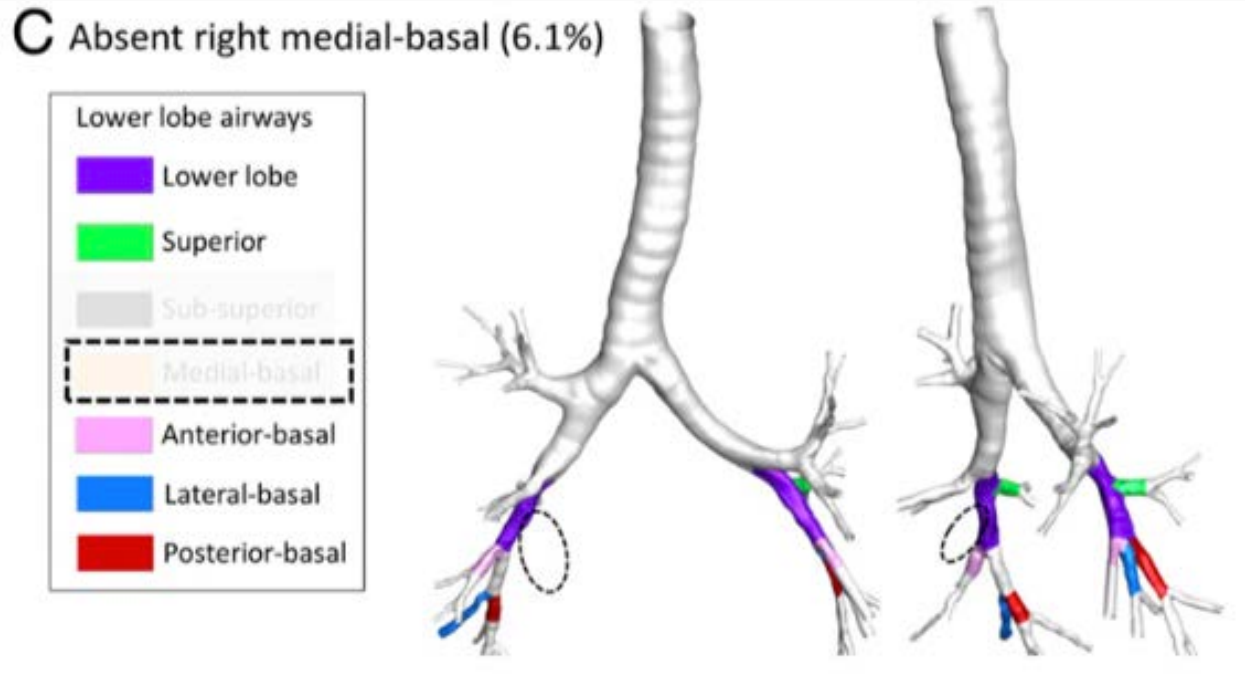
Bendich et al., Ann. Appl. Statistics, 2016

The geometry of the lung



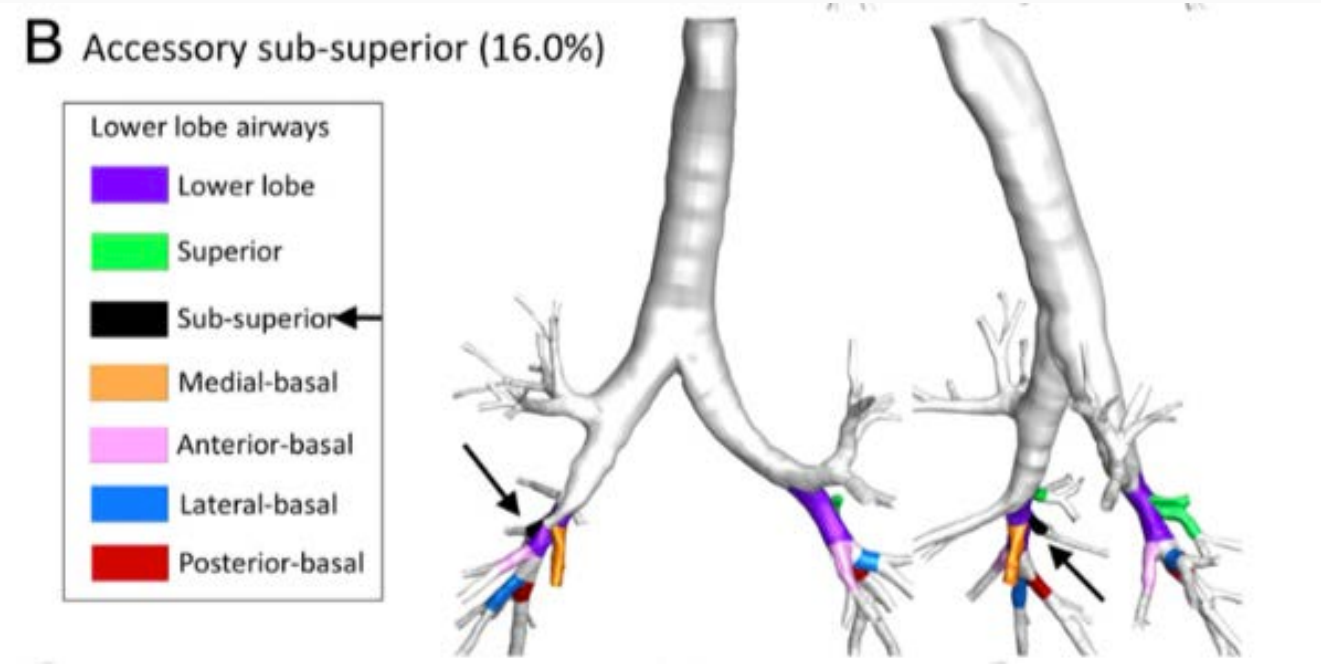
Pictures from Smith et al., Human airway branch variation and chronic obstructive pulmonary disease, PNAS 2018

The geometry of the lung



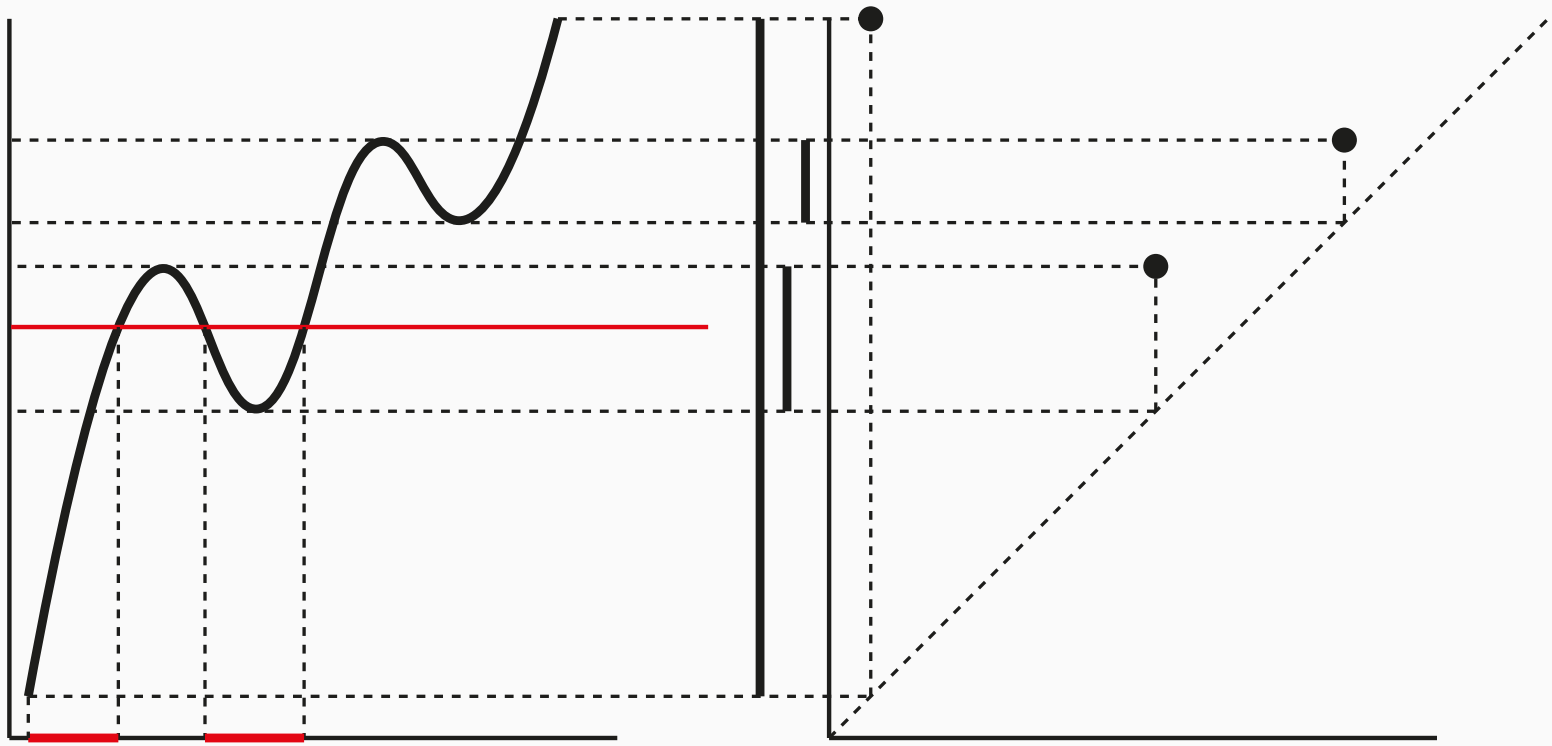
Pictures from Smith et al., Human airway branch variation and chronic obstructive pulmonary disease, PNAS 2018

The geometry of the lung



Pictures from Smith et al., Human airway branch variation and chronic obstructive pulmonary disease, PNAS 2018

Example



Comparing persistence diagrams: the bottleneck distance

Definition

Let X and Y be finite multisets, whose underlying sets are subsets of a metric space Z. Assume that X and Y are of equal size as multisets.

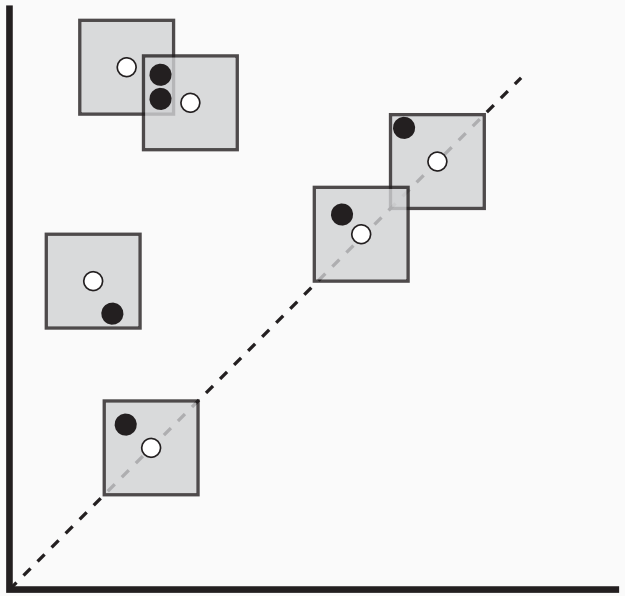
The bottleneck distance between X and Y is defined by

$$d_B(X, Y) = \inf_{\gamma} \sup_{x \in X} d(x, \gamma(x))$$

where γ ranges over all bijections of multisets $\gamma : X \to Y$.

We consider each point of multiplicity k as k points, and γ is a bijection between the resulting sets.

The bottleneck distance



White and black dots belong to persistence diagrams of two different functions. The squares represent a bijection realising the bottleneck distance, which equals half the side length.

Comparing persistence diagrams: the Wasserstein metric

Let X, Y be subsets of $\mathbb{R}^2$ (multisets in general).

Let $\gamma : X \to Y$ be a bijection. For $p \geq 1$ the cost of γ is defined by

$$C_p(\gamma) = \left(\sum_{x \in D(f)} \| x - \gamma(x) \|^p \right)^{1/p}$$

Definition

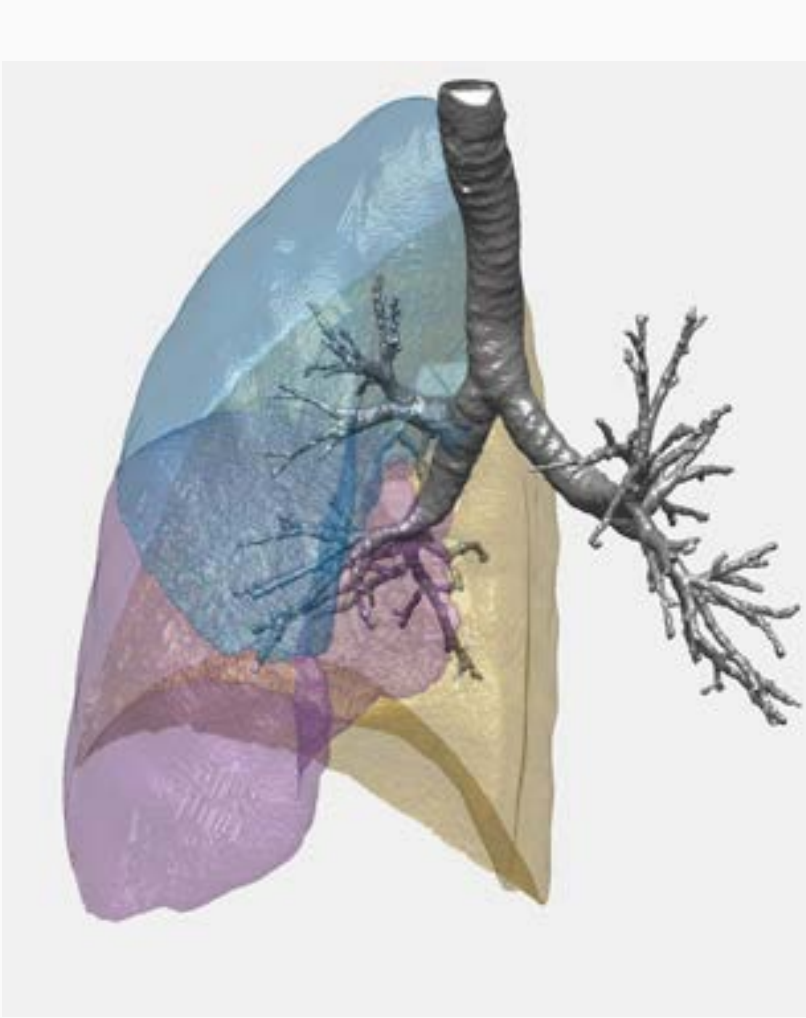
$$W_p(X, Y) = \inf_{\gamma} C_p(\gamma)$$

where the infimum is taken over all bijections $\gamma : X \to Y$.

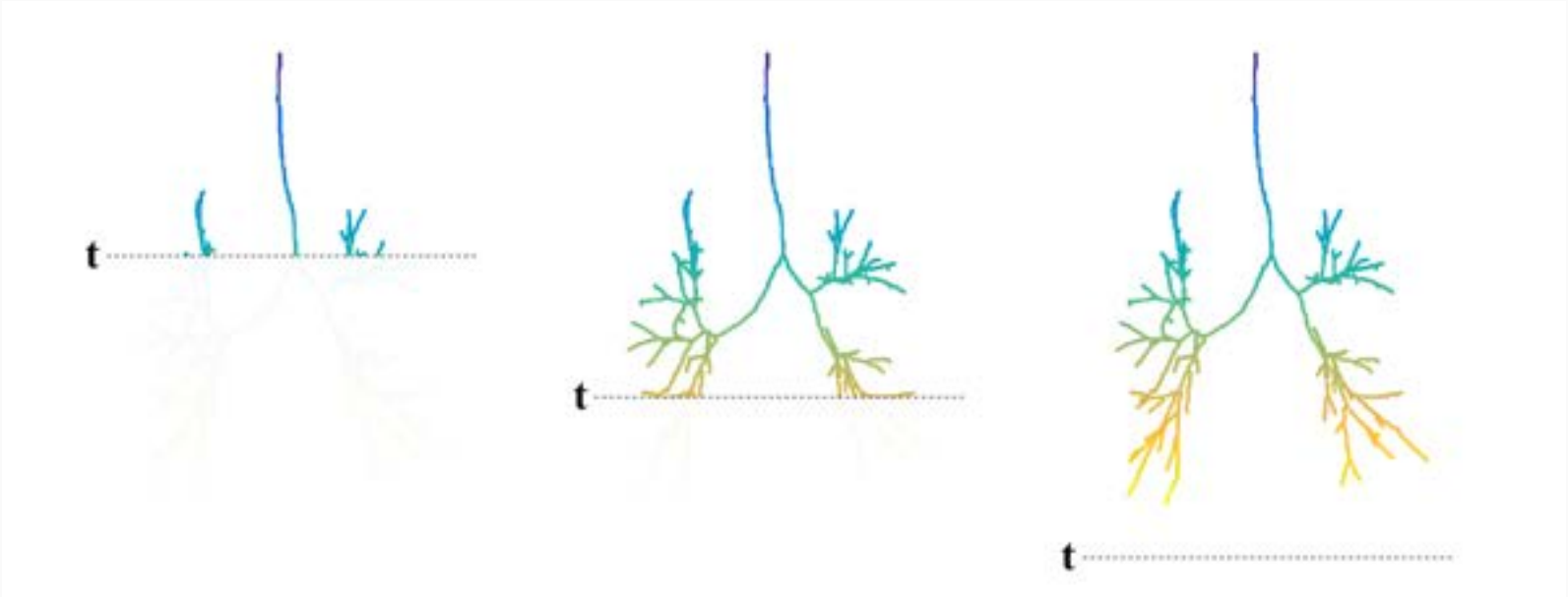
Topological characteristics of lungs

- The overall aim of this study was to develop a set of new radiomic features that can distinguish between healthy non-smokers as well as healthy smokers and patients with COPD.
- The following four study participant groups defined by smoking status and spirometry given by the GOLD guidelines were studied:
  - healthy non-smokers and healthy smokers (both judged as healthy by spirometry showing FEV1 $>$ 80% of predicted and FEV1/FVC $>$ 0.75),
  - mild COPD patients, consisting of GOLD stage 1 (with FEV1 $\geq$ 80% of predicted and FEV1/ FVC $<$ 0.70)
  - moderate COPD patients, consisting of GOLD stage 2 (50% $\leq$ FEV1 $<$ 80% of predicted and FEV1/FVC $<$ 0.70).

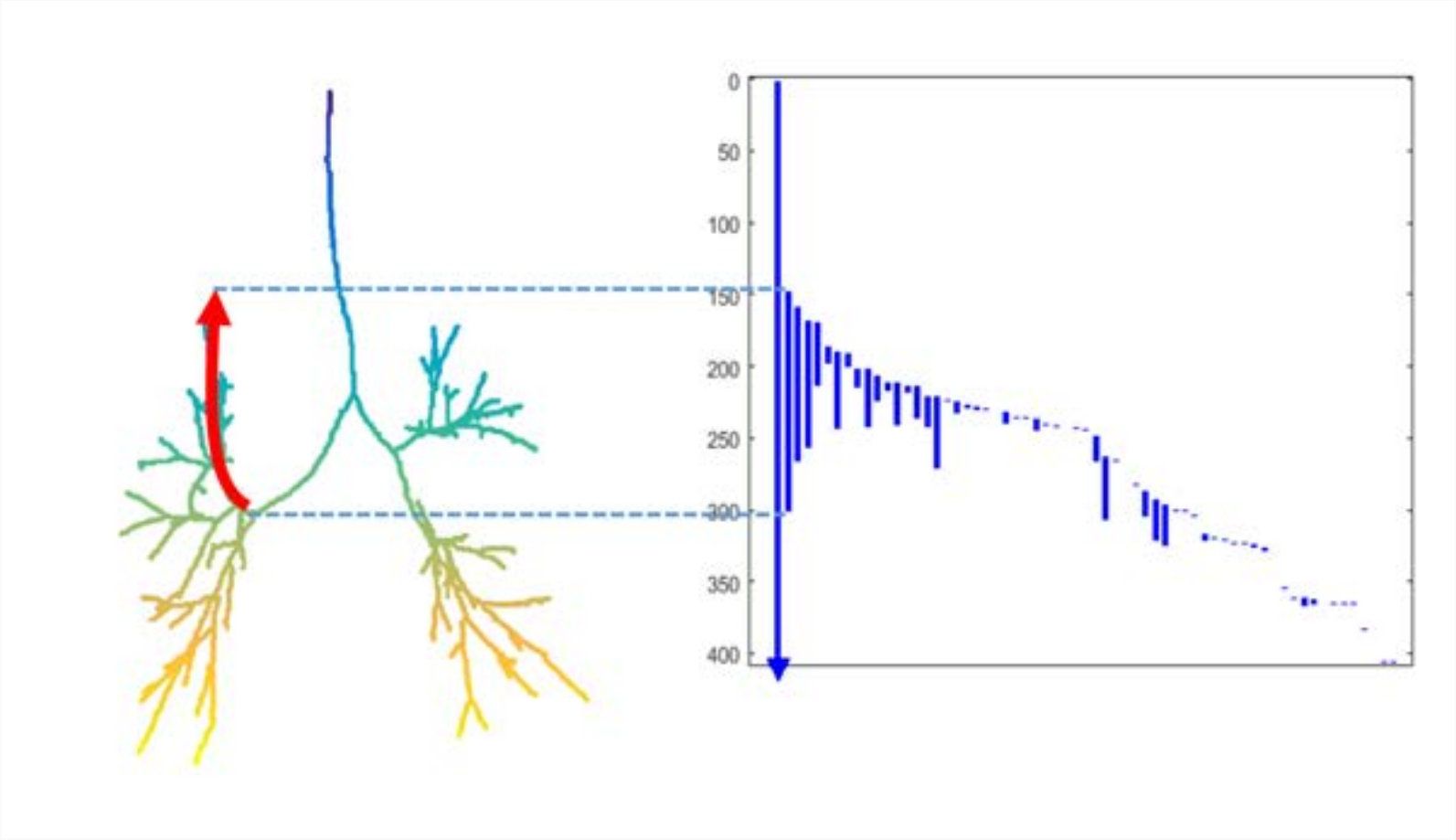
Topological characteristics of lungs



Topological characteristics of lungs

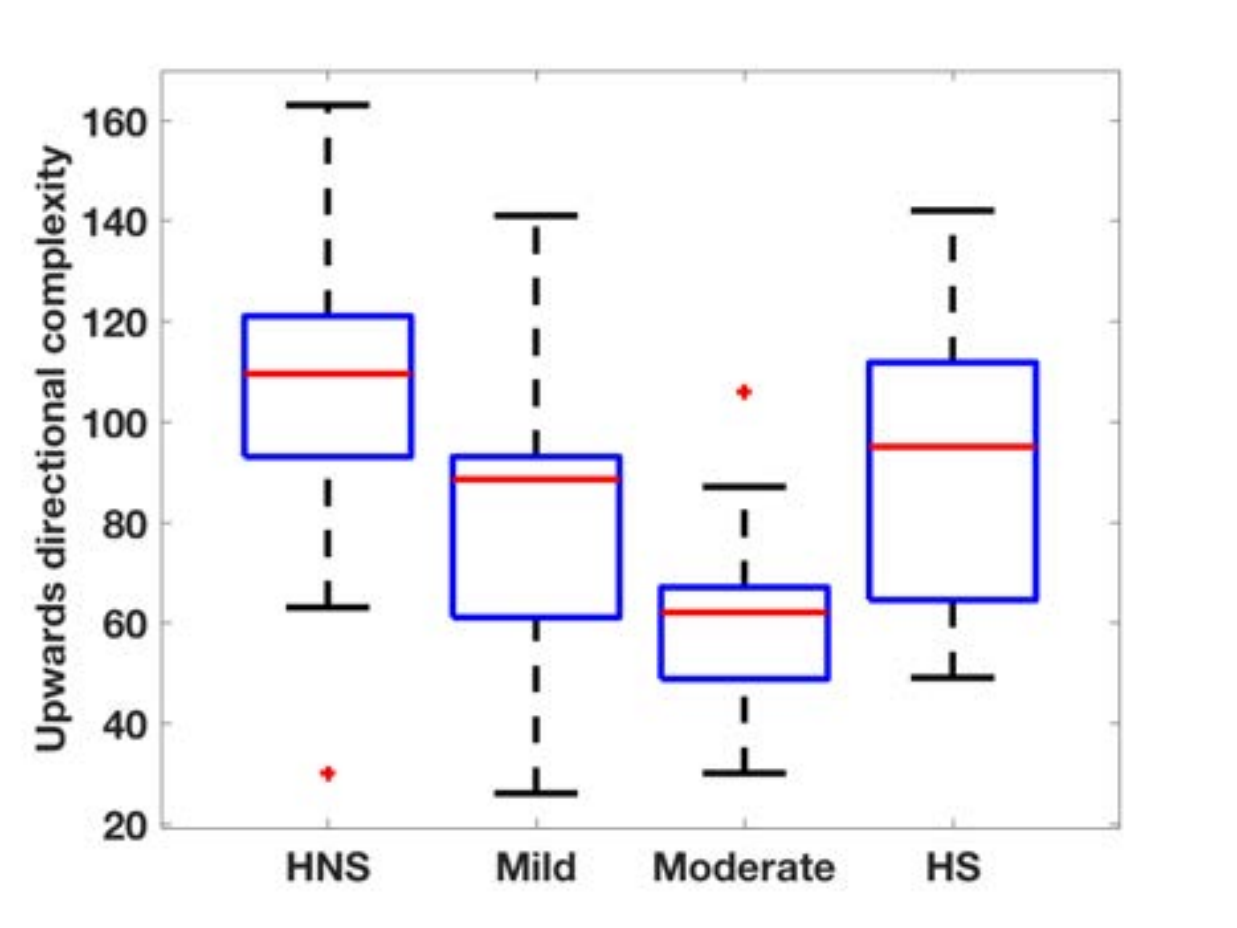


Topological characteristics of lungs



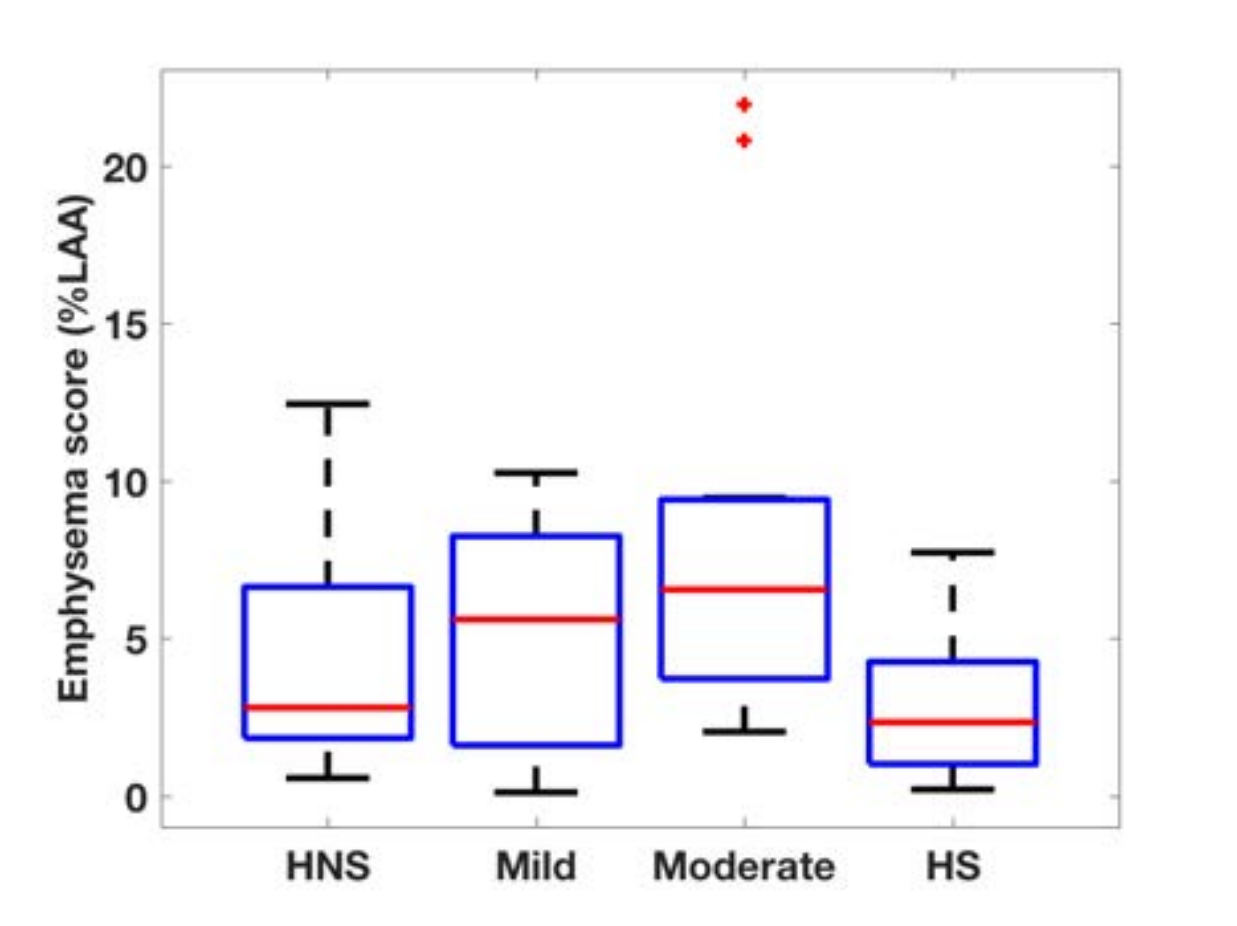
Directional complexity: the sum of the lengths of the bars in the graph (total life span of the persistence diagram).

Topological characteristics of lungs

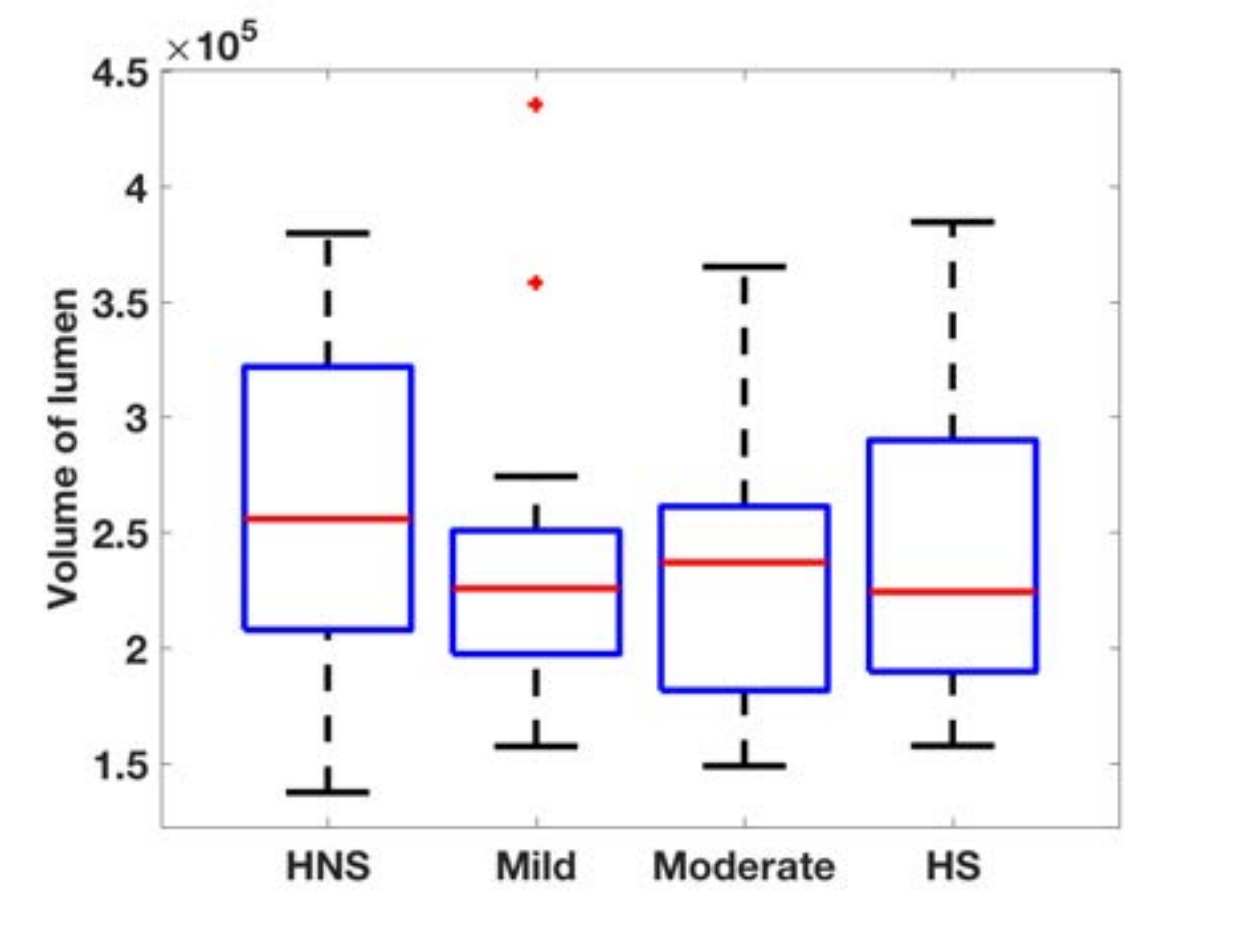


Directional complexity distinguishes healthy subjects from Mild and Moderate COPD

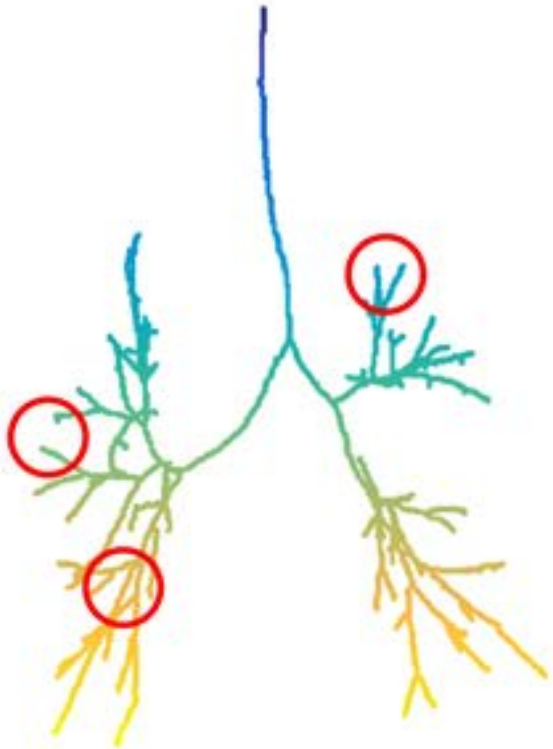
Comparison to other standard measurements



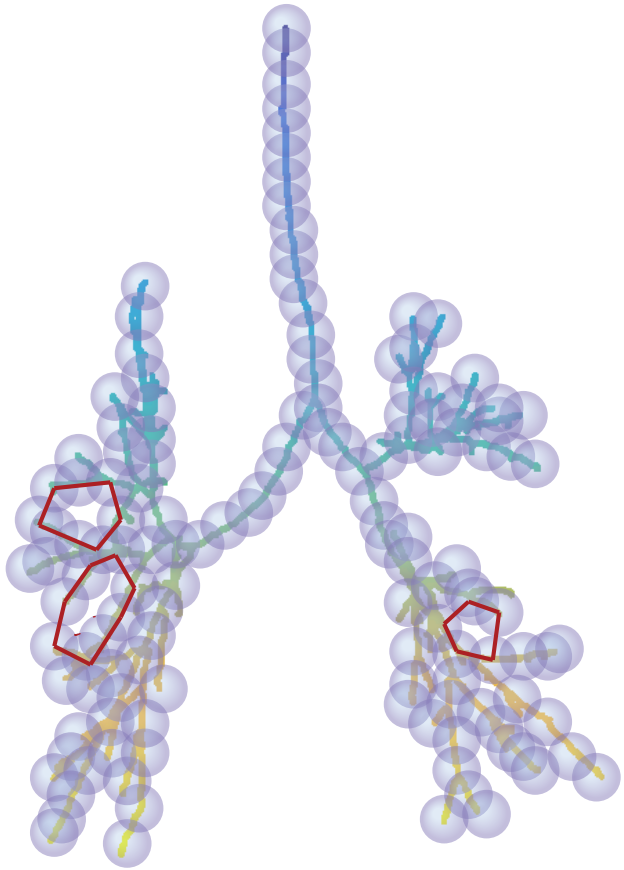
Comparison to other standard measurements



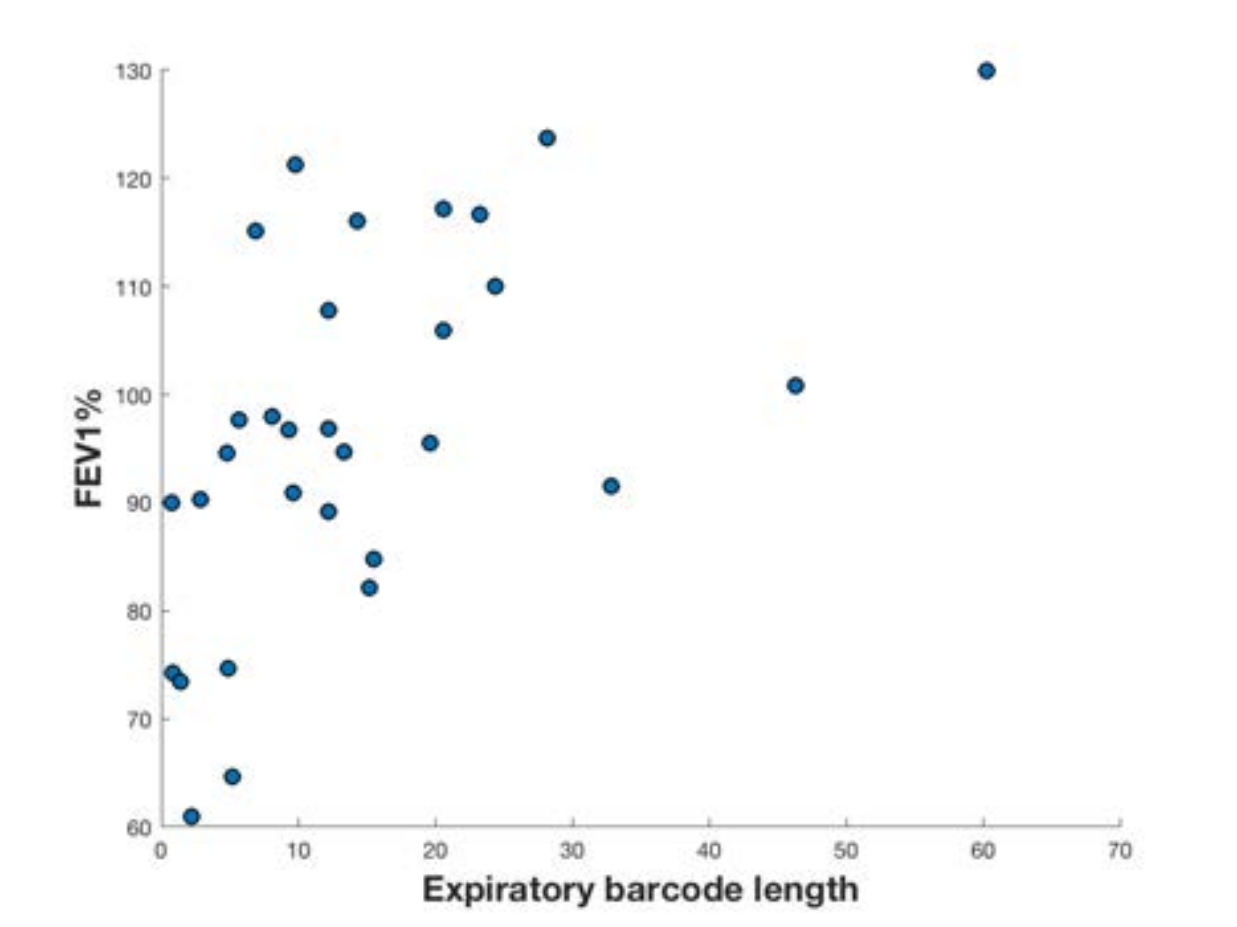
H_1 and the shape of the lung



H_1 and the shape of the lung

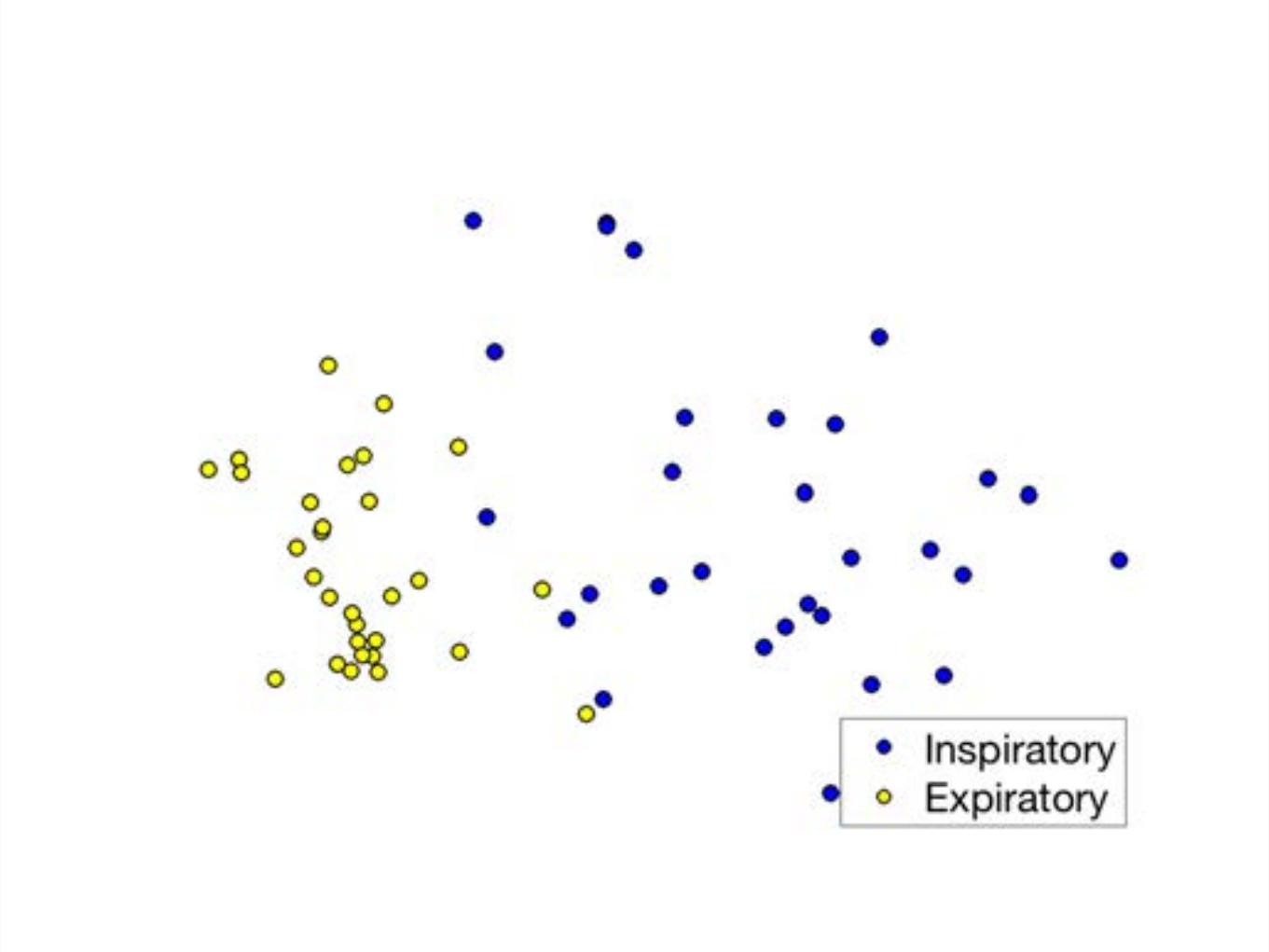


H_1 and the shape of the lung



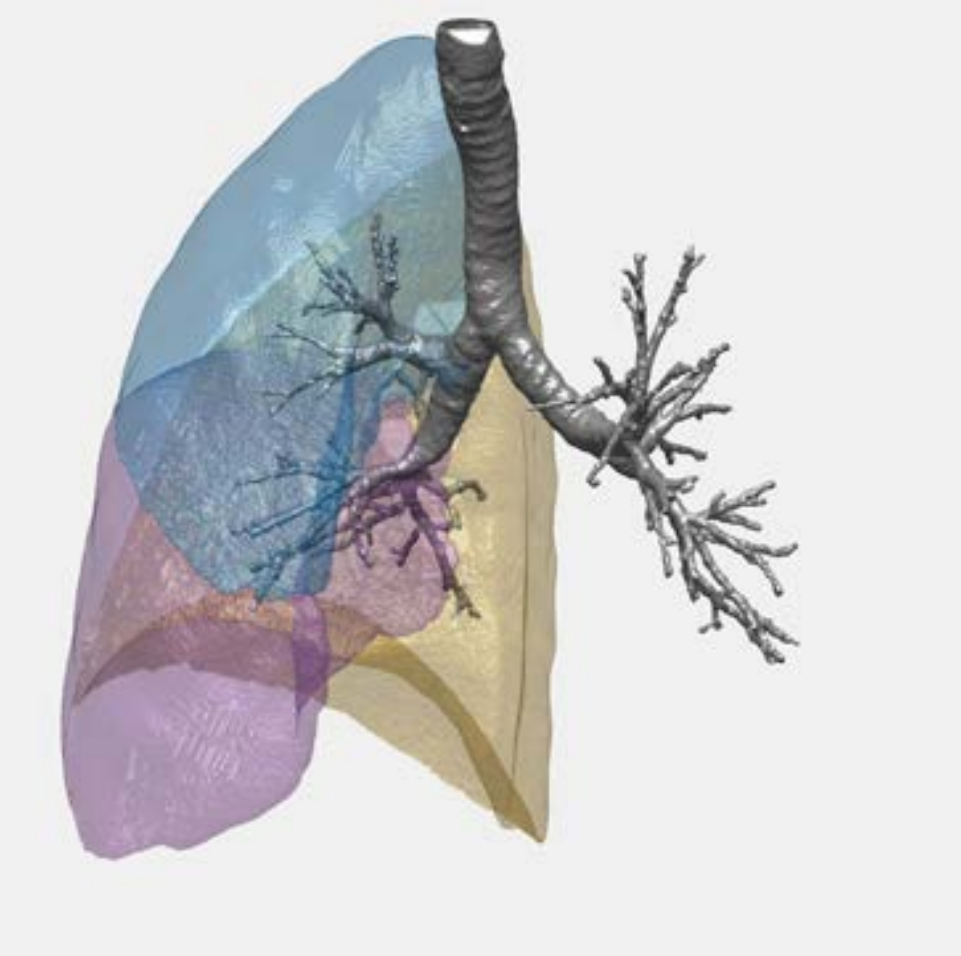
Numerical invariant: total barcode length in the H_1 barcode.

H_1 and the shape of the lung

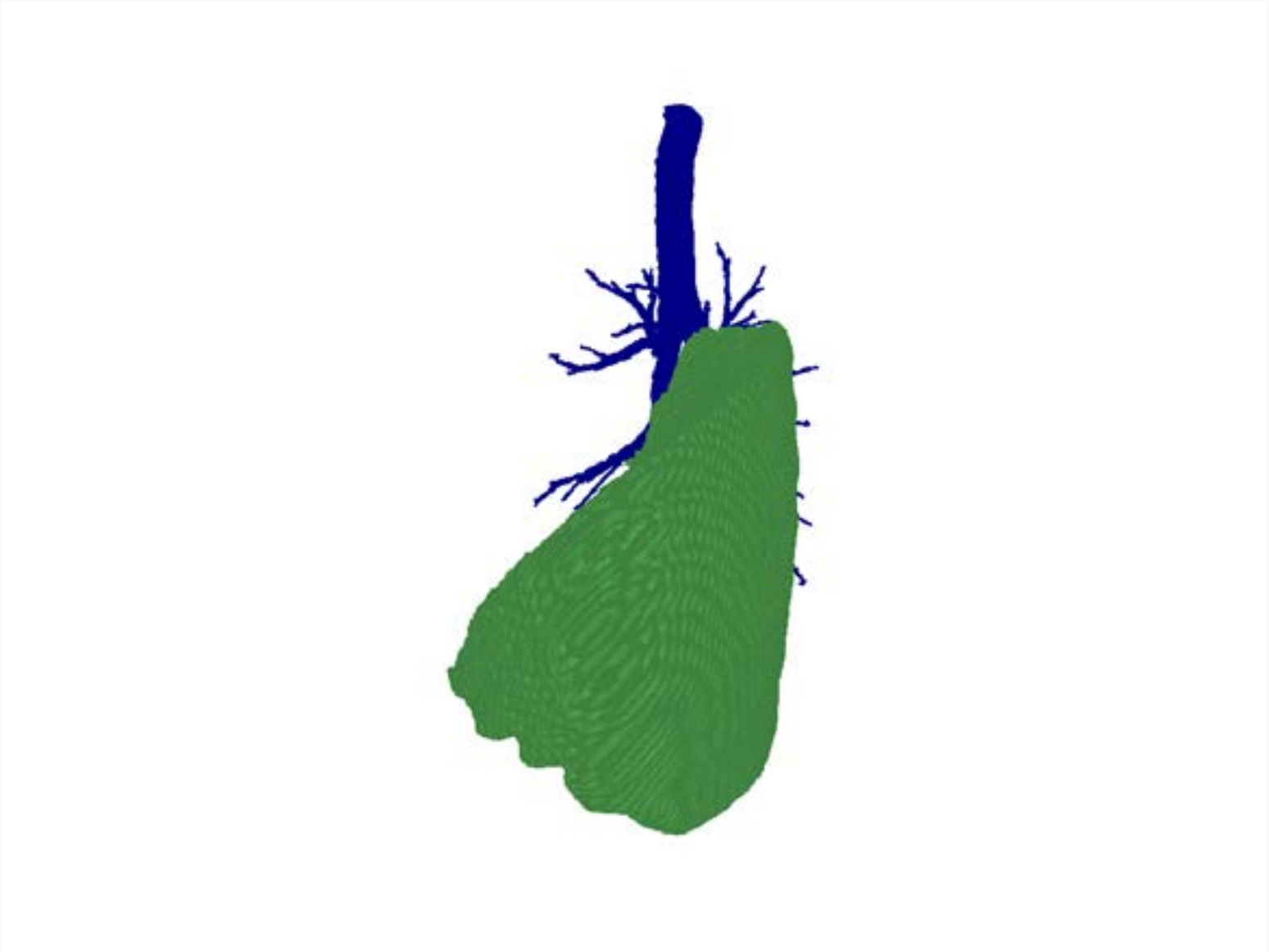


Metric space structure created using the Wasserstein metric and represented through 2-dim MDS projection

H_2 and the spatial structure

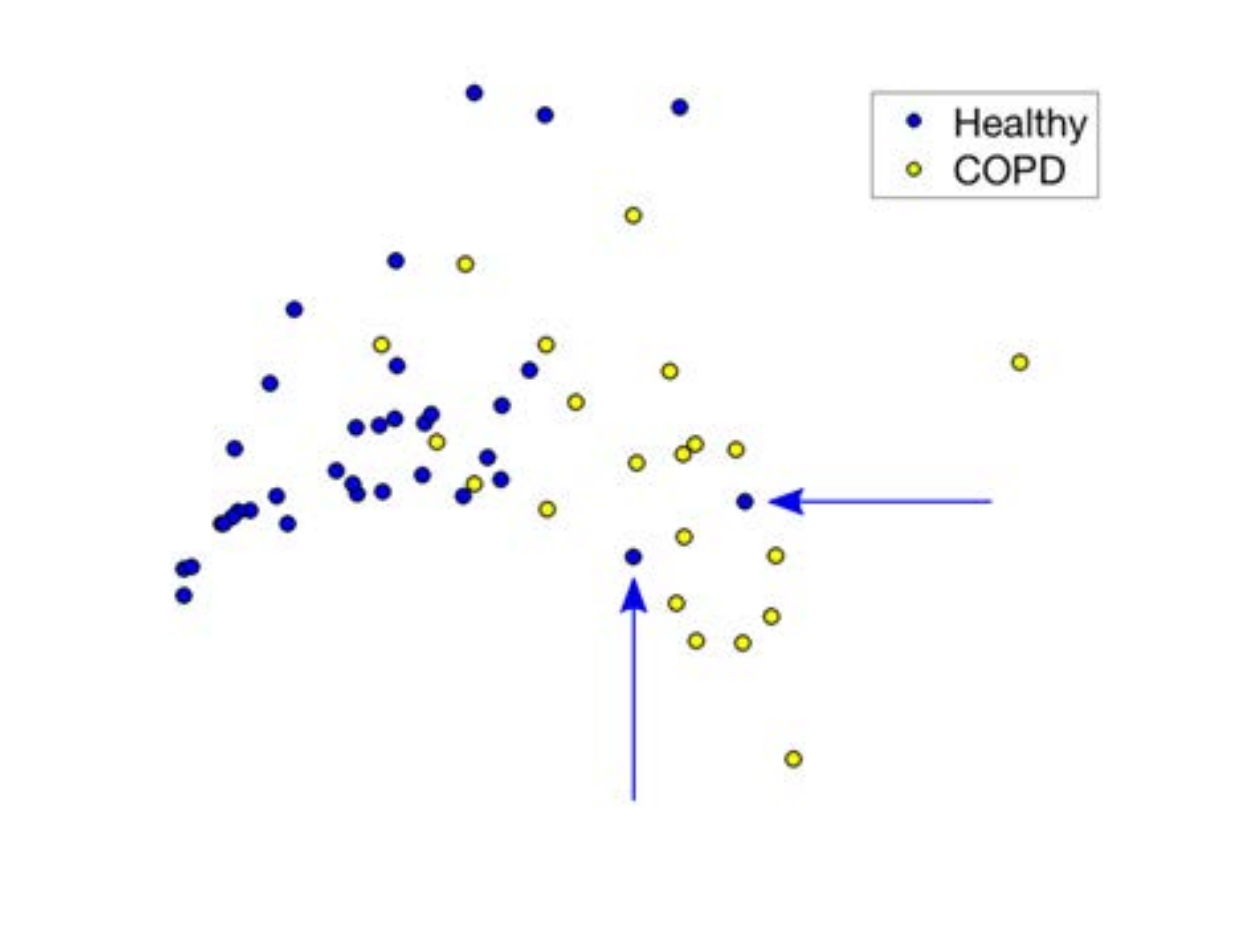


H_2 and the spatial structure



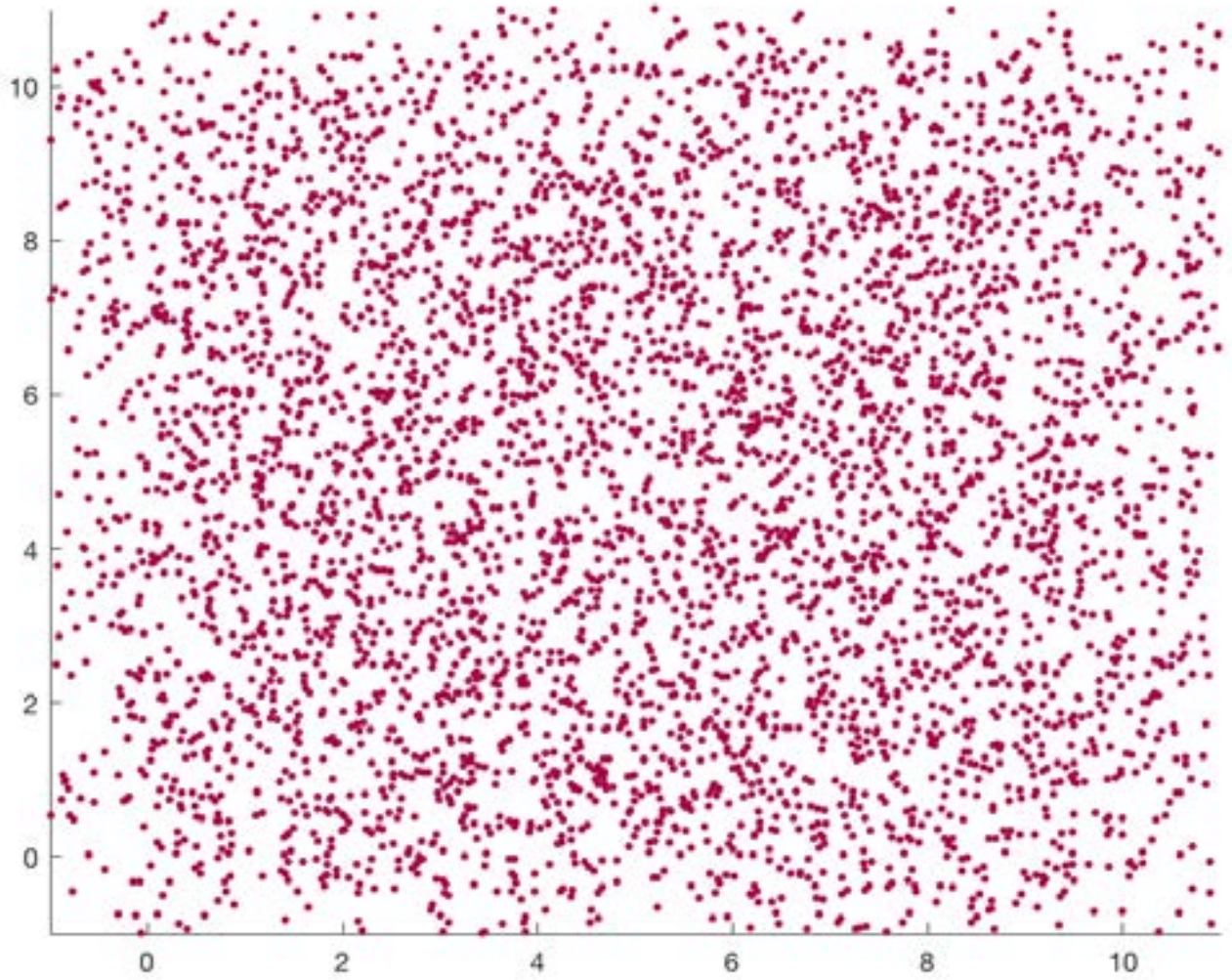
H_2 and the spatial structure

H_2 and the spatial structure

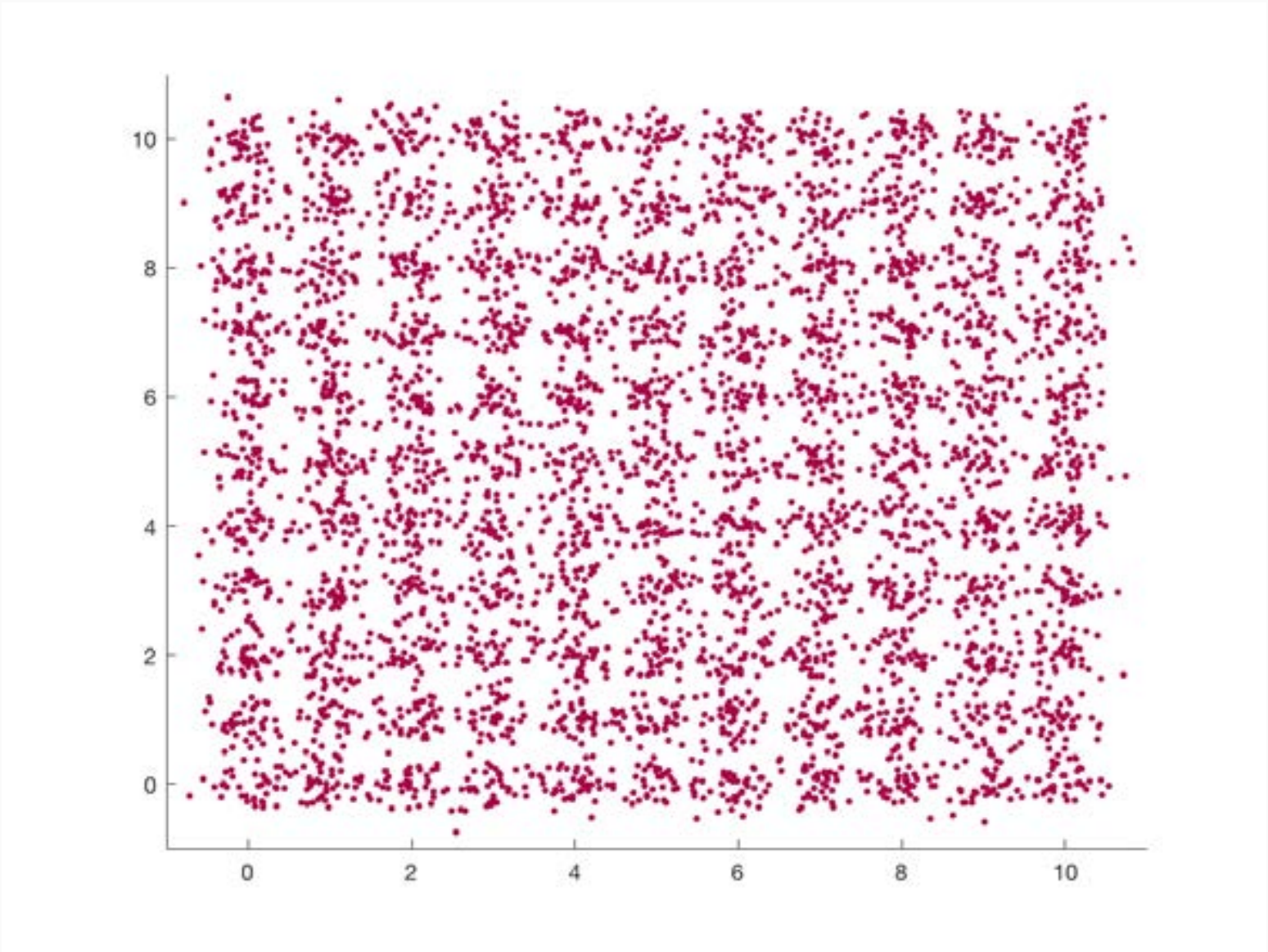


Metric space structure created using the Wasserstein metric; represented here in a 2-dim MDS projection

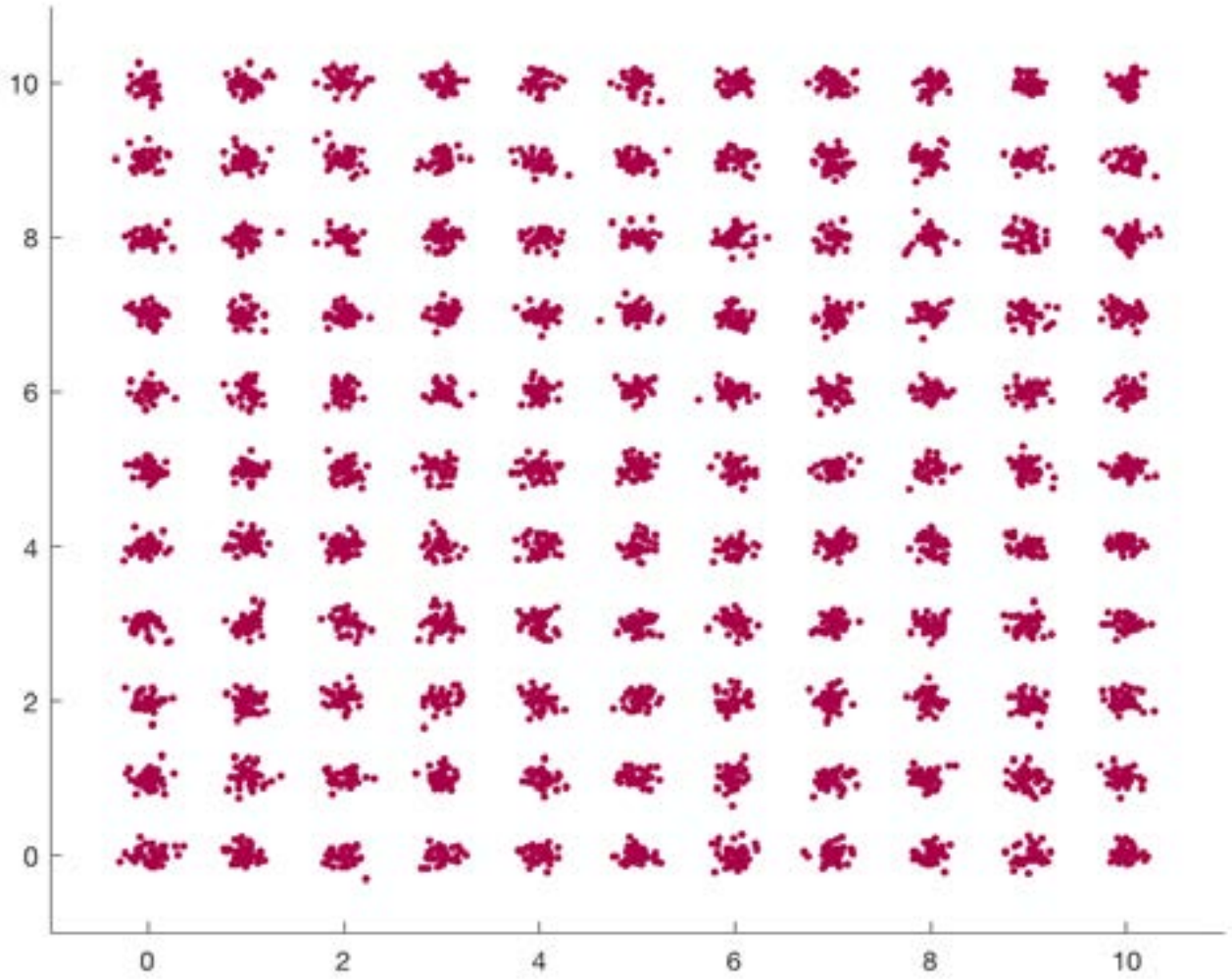
Geometry and persitence



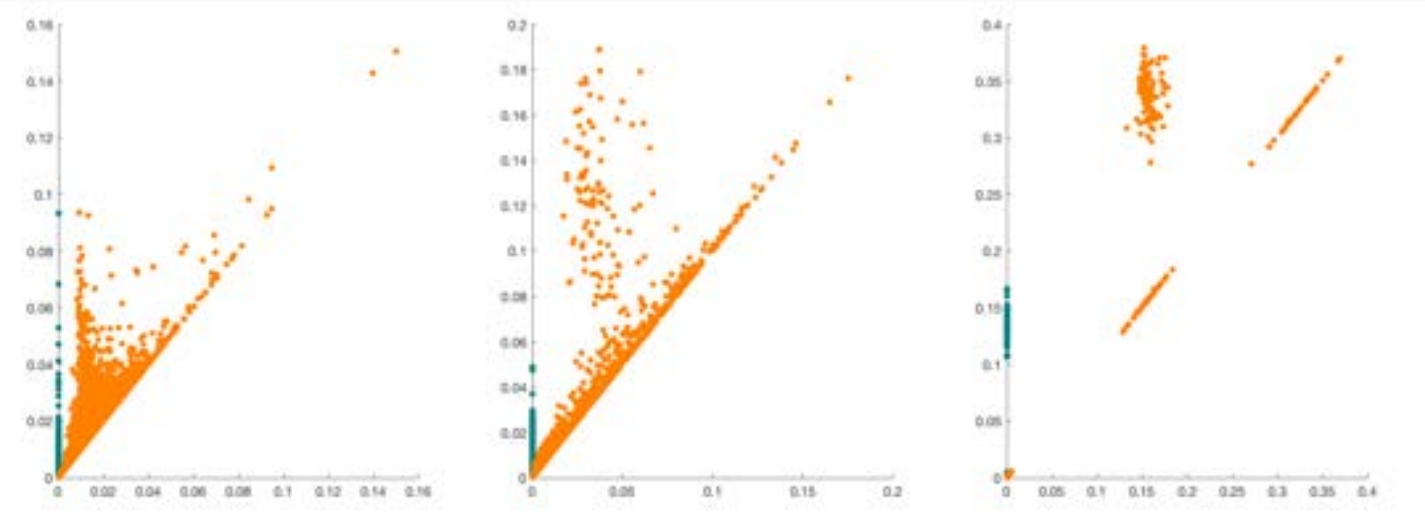
Geometry and persitence



Geometry and persitence



Geometry and persitence



Questions

- Persistence for fibre bundles: a combination of geometric and topological information
- Synchronisation through persistence
- Symmetry detection
- Time evolution: geometric intuition to guide the evolution of persistent homology
- Phase transitions in persistence diagrams

The same equations have the same solutions.

Richard Feynman