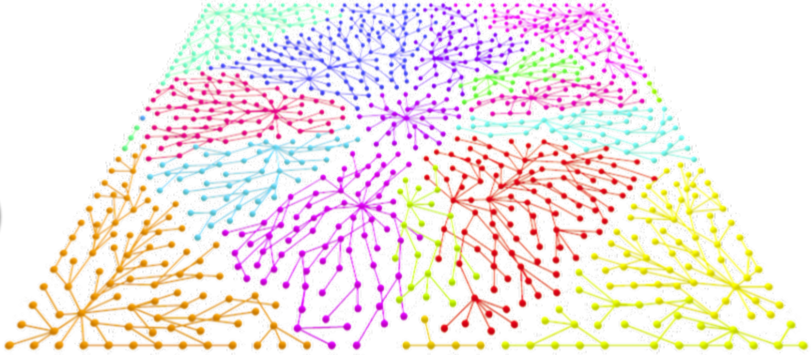
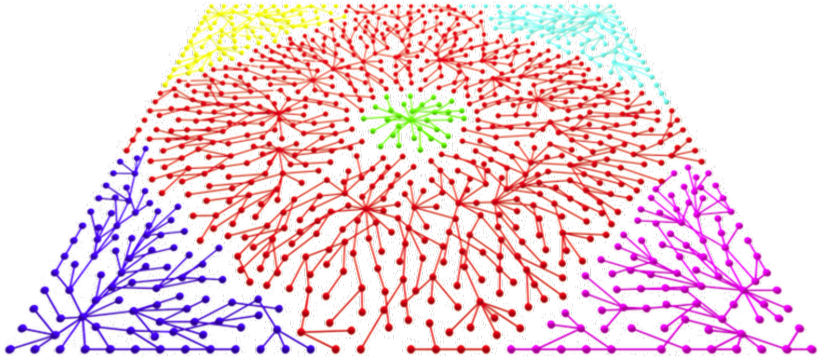
CS233, CME251: Geometric and Topological Data Analysis

Leonidas Guibas
Computer Science Department
Stanford University



Guest lecture: Samir Chowdhury (Stanford)

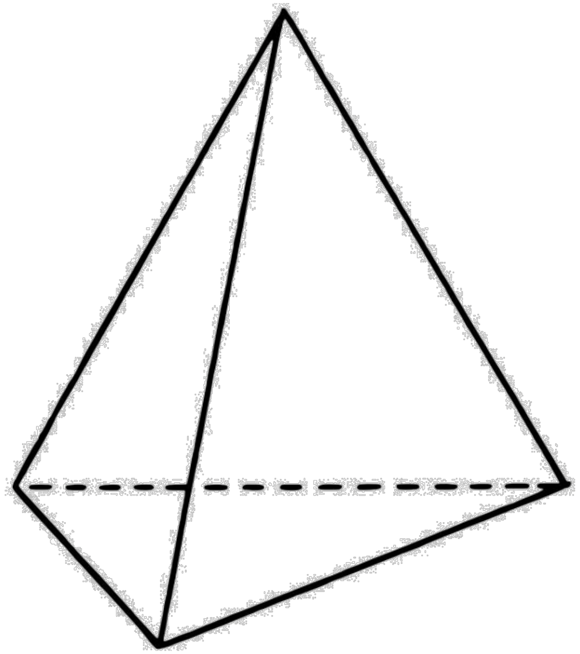
Lecture 9
26 April 2021



Last Time: Persistent Homology

Betti Numbers β_i

- Ranks of the free part of homology groups H_i
- β_0 counts the number of connected components
- β_1 counts the number of independent loops
- β_2 counts the number of independent voids
- …



Topology is fundamentally a tool for classification

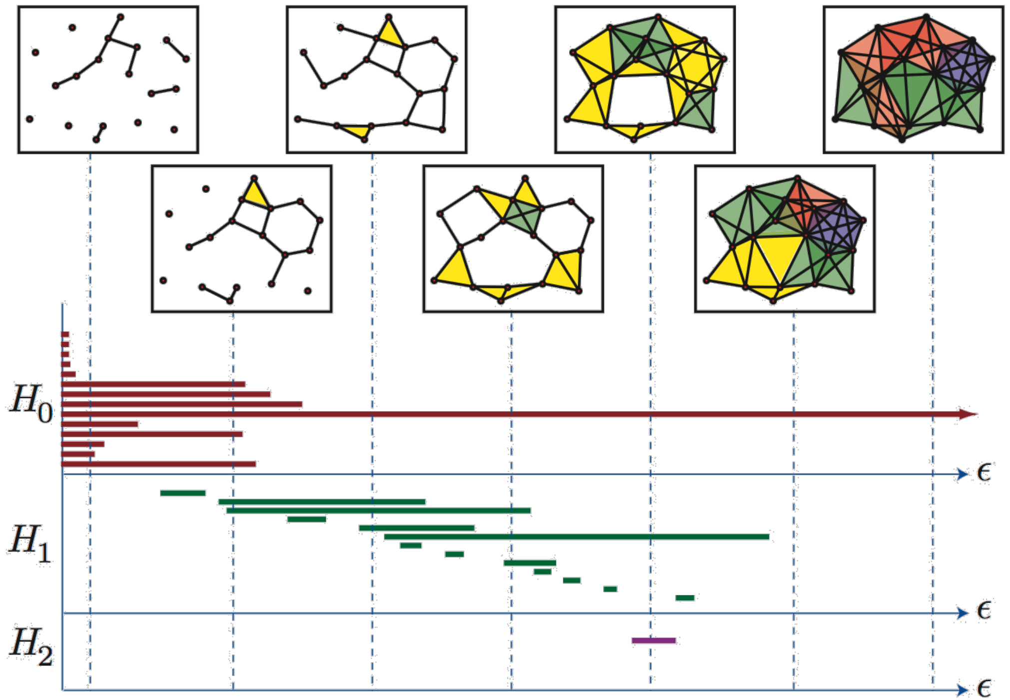
Pipeline

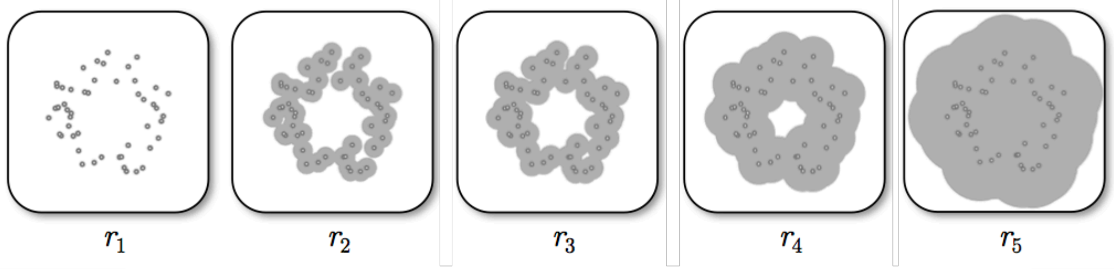
Dataset:
- **Euclidean point cloud data**
- **Metric spaces**
- **Triangulated shapes**
- Graphs, matrices…

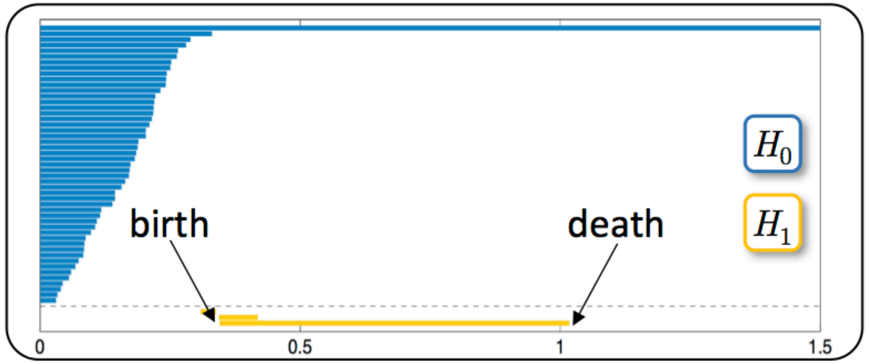
Filtered simplicial complex
- Possible to have both inclusions and deletions (dynamic data)

Vector spaces with linear maps: a **persistent vector space**

Visual summary: a **persistence diagram** or **persistence barcode**
- Postprocessing available for ML-friendly output







Reduction Algorithm for Homology*

*Cohen-Steiner, Edelsbrunner, Morozov - **Vines and Vineyards by Updating Persistence in Linear Time**

To get H_p, we need: (1) a basis for $\ker \partial_p$, and (2) a sub-basis for $im\ \partial_{p+1}$.

	ab	ca	bc	cd	de	eb	abc
a	1	1	0	0	0	0	0
b	1	0	1	0	0	1	0
c	0	1	1	1	0	0	0
d	0	0	0	1	1	0	0
e	0	0	0	0	1	1	0
ab	0	0	0	0	0	0	1
ca	0	0	0	0	0	0	1
bc	0	0	0	0	0	0	1
cd	0	0	0	0	0	0	0
de	0	0	0	0	0	0	0
eb	0	0	0	0	0	0	0

R

Notation: $low_R(j)$= row index of bottom 1 in column j of matrix R

Algorithm:

Initialize boundary information in incidence matrix D

Initialize $R \ := \ D$

For column $j \ = \ 1, 2, \ldots, n$

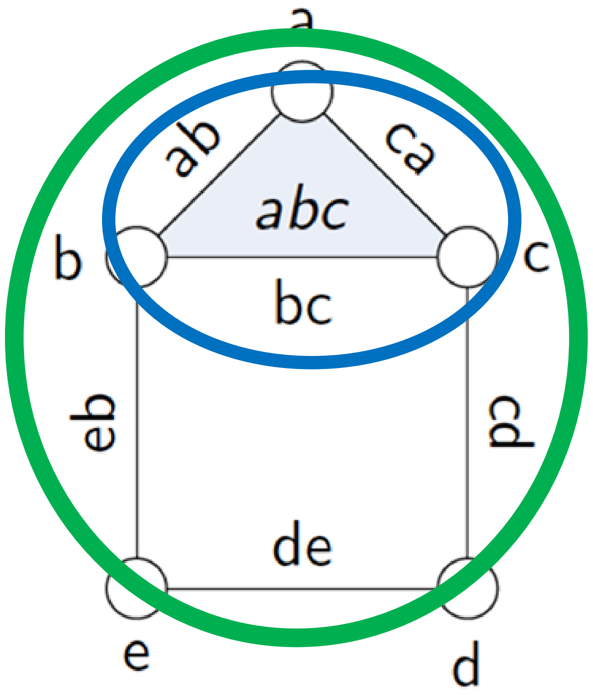
  while ∃j′ < j, $low_R(j') = low_R(j)$ do

    add column j’ to column j

Reduction Algorithm for Homology

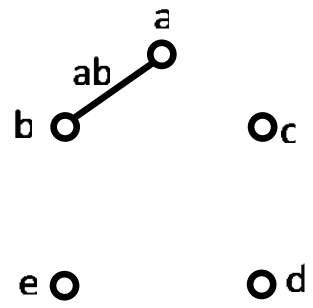
$H_0 \cong \mathbb{F}[a]$
$H_1 \cong \mathbb{F}[eb + de + cd + ca + ab]$

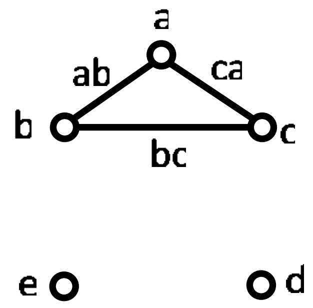
		1	2	3	4	5			8			11	
		a	b	c	d	e	ab	ca	$bc+ca+ab$	cd	de	$eb+de+cd+ca+ab$	abc
1	a	0	0	0	0	0	1	1	0	0	0	0	0
2	b	0	0	0	0	0	1	0	0	0	0	0	0
3	c	0	0	0	0	0	0	1	0	1	0	0	0
4	d	0	0	0	0	0	0	0	0	1	1	0	0
5	e	0	0	0	0	0	0	0	0	0	1	0	0
	ab	0	0	0	0	0	0	0	0	0	0	0	0
	ca	0	0	0	0	0	0	0	0	0	0	0	0
8	$bc+ca+ab$	0	0	0	0	0	0	0	0	0	0	0	1
	cd	0	0	0	0	0	0	0	0	0	0	0	0
	de	0	0	0	0	0	0	0	0	0	0	0	0
11	$eb+de+cd+ca+ab$	0	0	0	0	0	0	0	0	0	0	0	0
	abc	0	0	0	0	0	0	0	0	0	0	0	0

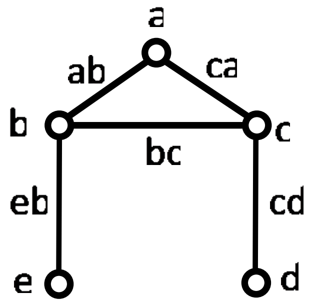


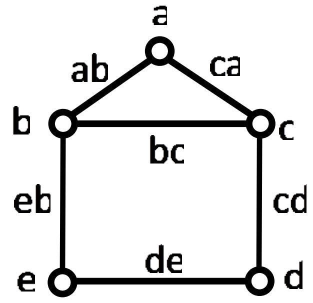
Back to the House Example

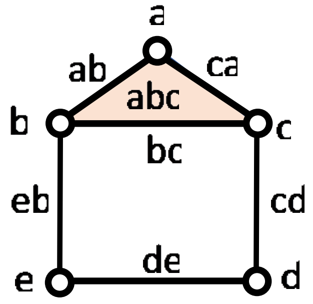
		1	2	3	4	5			8			11	
		a	b	c	d	e	ab	ca	$bc+ca+ab$	cd	eb	$de+eb+cd+ca+ab$	abc
1	a	0	0	0	0	0	1	1	0	0	0	0	0
2	b	0	0	0	0	0	1	0	0	0	1	0	0
3	c	0	0	0	0	0	0	1	0	1	0	0	0
4	d	0	0	0	0	0	0	0	0	1	0	0	0
5	e	0	0	0	0	0	0	0	0	0	1	0	0
	ab	0	0	0	0	0	0	0	0	0	0	0	0
	ca	0	0	0	0	0	0	0	0	0	0	0	0
8	$bc+ca+ab$	0	0	0	0	0	0	0	0	0	0	0	1
	cd	0	0	0	0	0	0	0	0	0	0	0	0
	eb	0	0	0	0	0	0	0	0	0	0	0	0
11	$de+eb+cd+ca+ab$	0	0	0	0	0	0	0	0	0	0	0	0
	abc	0	0	0	0	0	0	0	0	0	0	0	0



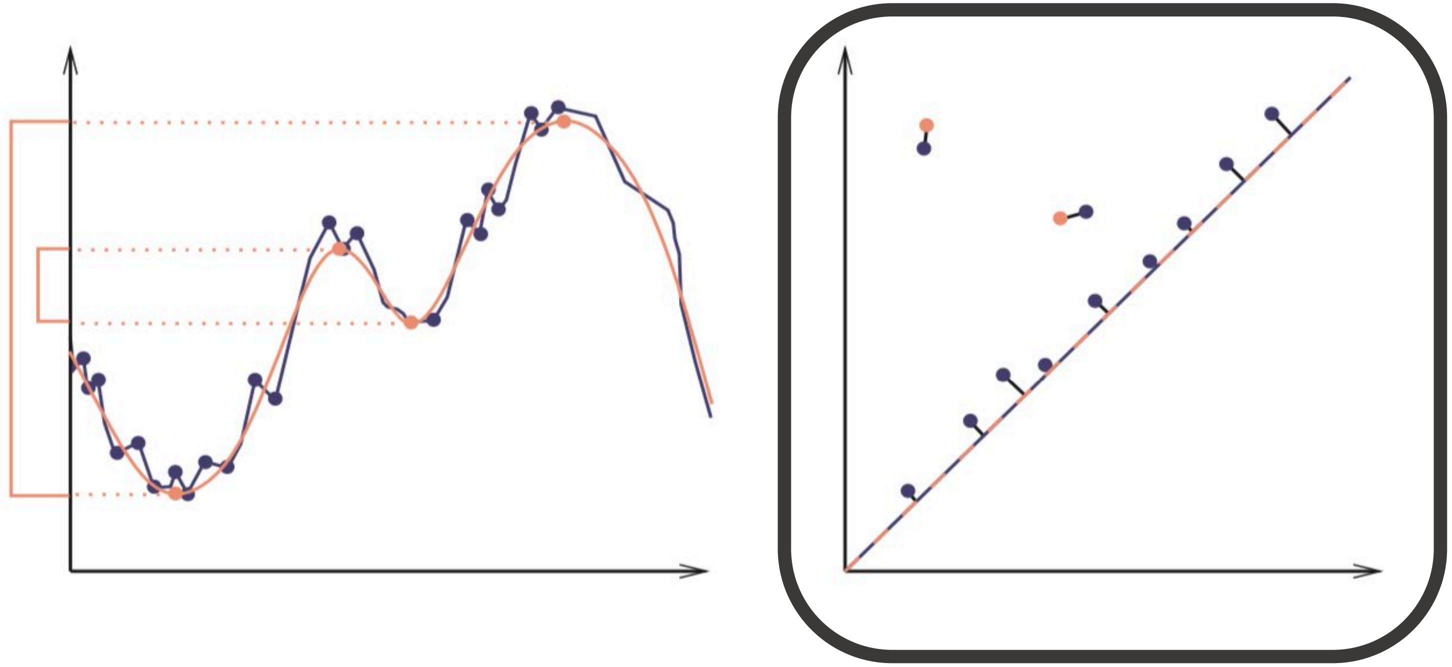








The Bottleneck Distance



A, B two persistence diagrams.

$$d_B(A,B) := \inf \left\{ \sup_{x \in A \cup \Delta} \|x - \phi(x)\|_\infty : \phi : A \cup \Delta \to B \cup \Delta \text{ a bijection} \right\}$$

Stability of Persistent Homology

- Persistent homology outputs are robust to small perturbations of data

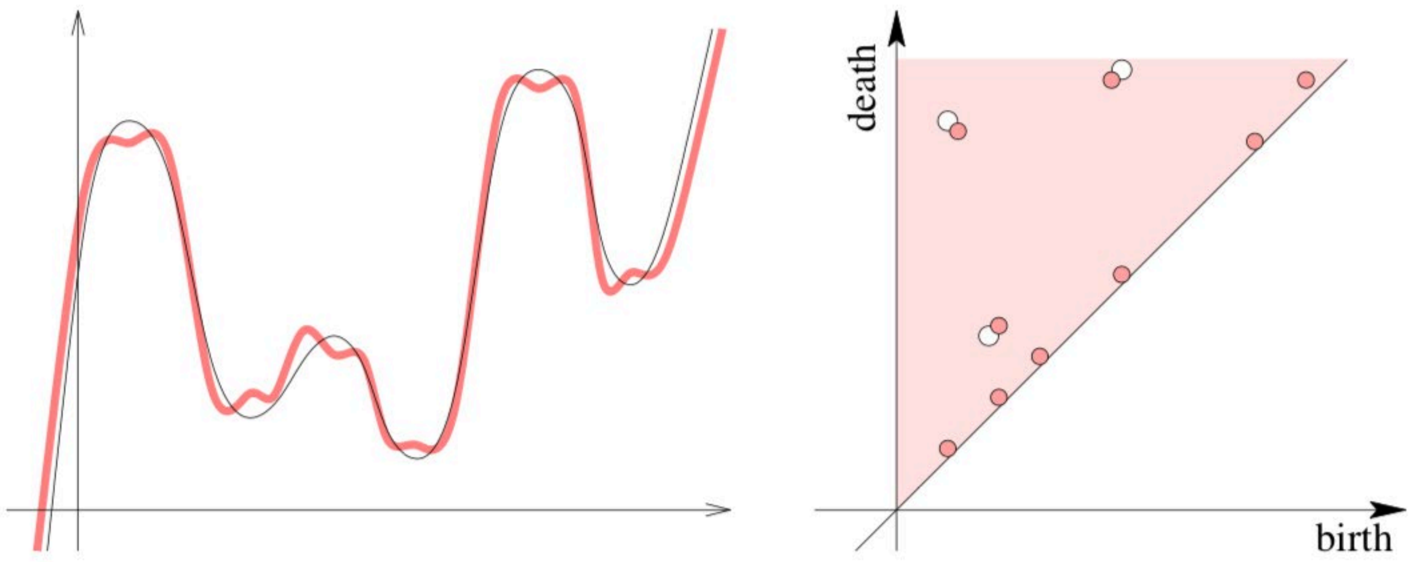


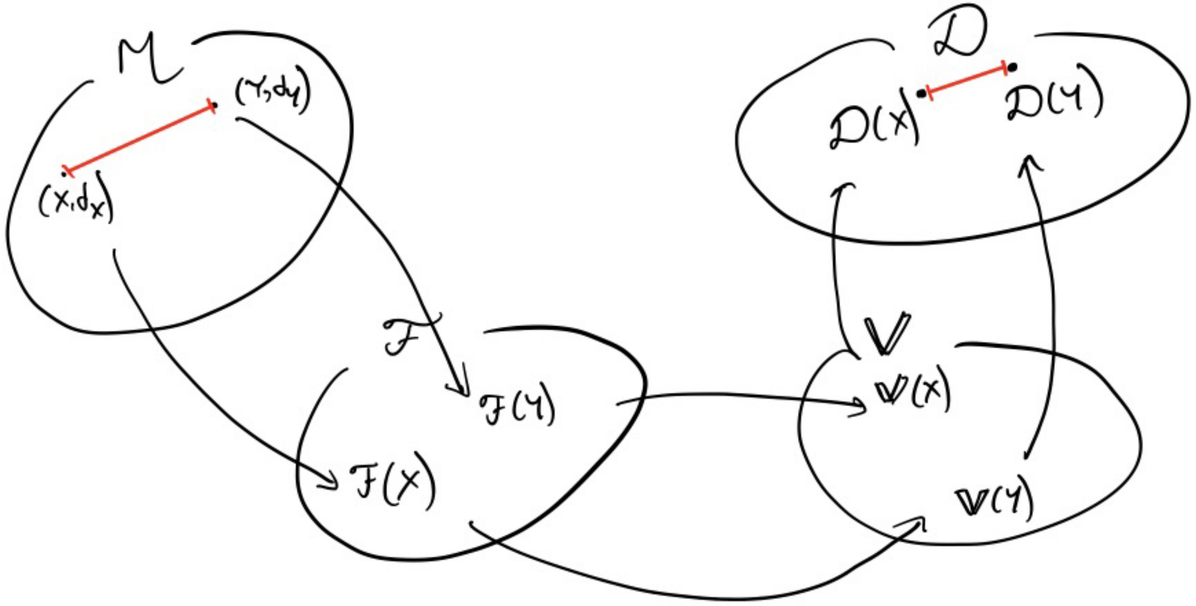
Figure from Edelsbrunner, Harer, *Computational Topology*

- Theorem (Cohen-Steiner, Edelsbrunner, Harer): $\mathcal{M}$ a triangulable space, $f, g: \mathcal{M} \to \mathbb{R}$ "nice". Then,

$$d_B\big(D_k(f), D_k(g)\big) \leq \|f - g\|_\infty.$$

Here D_k is the persistence diagram in dimension k, $\|\cdot\|_\infty$ is the max-norm, and d_B is the bottleneck distance

Gromov-Hausdorff Stability



$\mathcal{M}$ = finite metric spaces
$\mathcal{F}$ = filtered simplicial complexes
$\mathbb{V}$ = persistent vector spaces
$\mathbb{D}$ = persistence diagrams

- Theorem (Chazal, Cohen-Steiner, Guibas, Mémoli, Oudot): $(X, d_X), (Y, d_Y)$ finite metric spaces. Then,

$$\underbrace{d_B\big(D_k(VR(X)), D_k(VR(Y))\big)} \leq \underbrace{2d_{GH}(X, Y)}.$$

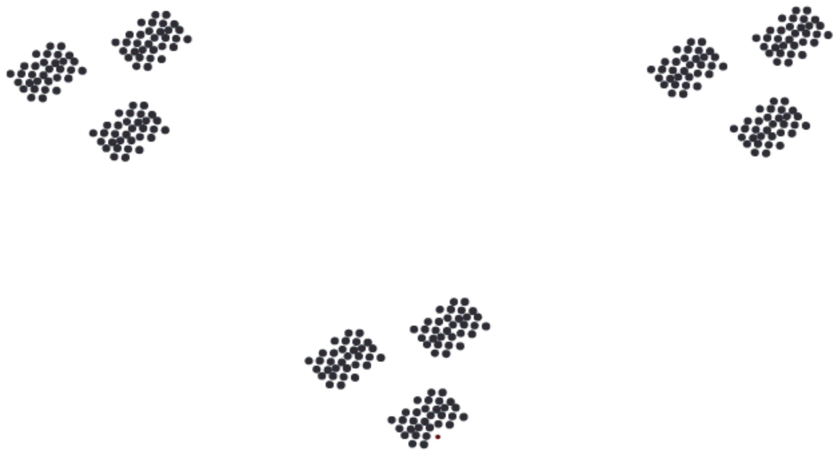
$O(n^3)$ with Hungarian algorithm,
$O(n^{3/2})$ state-of-the-art (Kerber, Morozov, Nigmetov 2017)

NP-hard

0-Dimensional Persistence and Single Linkage Clustering

Hierarchical Clustering

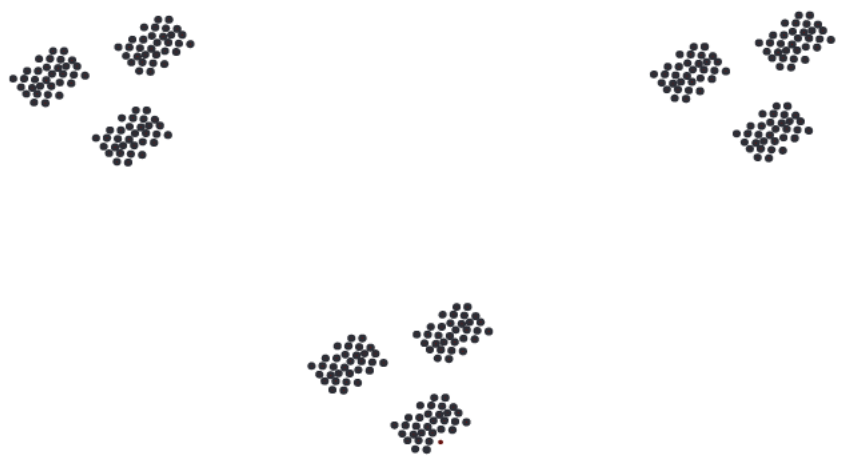
Carlsson, Mémoli, *Characterization, Stability, and Convergence of Hierarchical Clustering Methods*



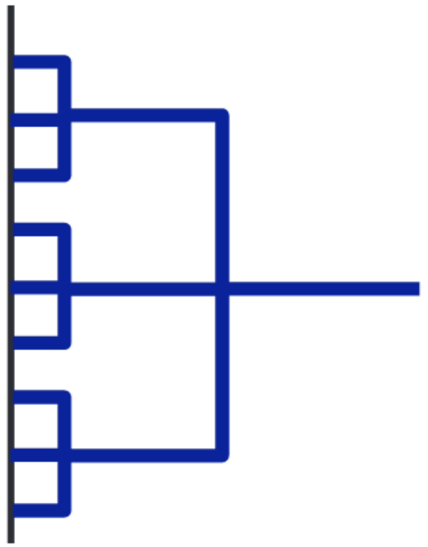
Are there three clusters? Or six?

Hierarchical Clustering

Carlsson, Mémoli, *Characterization, Stability, and Convergence of Hierarchical Clustering Methods*



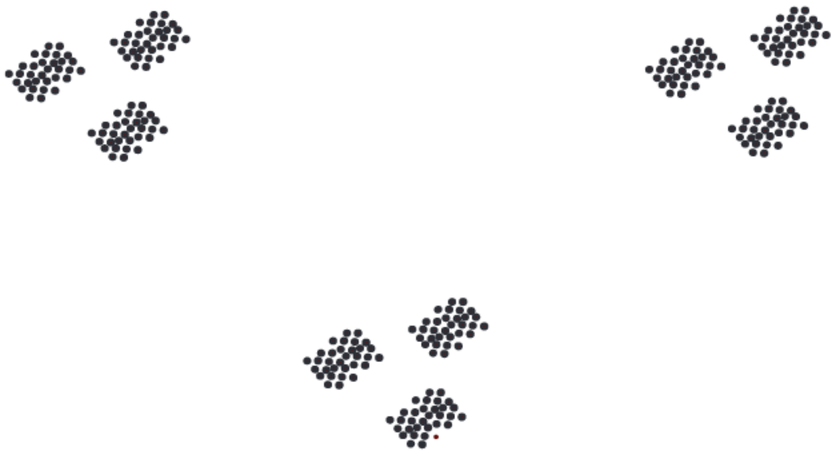
Are there three clusters? Or six?

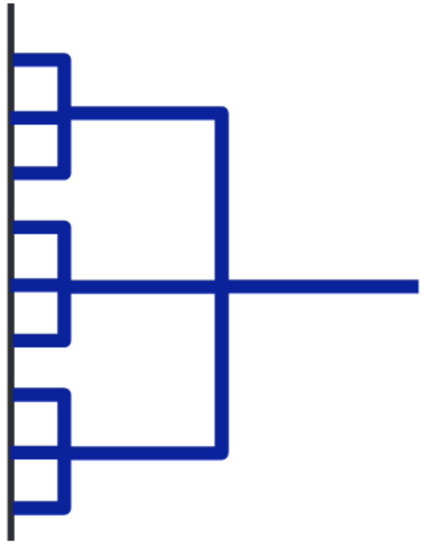


Represent clusters at **all** scales via dendrogram (= "tree" + "drawing")

Hierarchical Clustering

Carlsson, Mémoli, *Characterization, Stability, and Convergence of Hierarchical Clustering Methods*





Recall **metric space** (X, d_X):

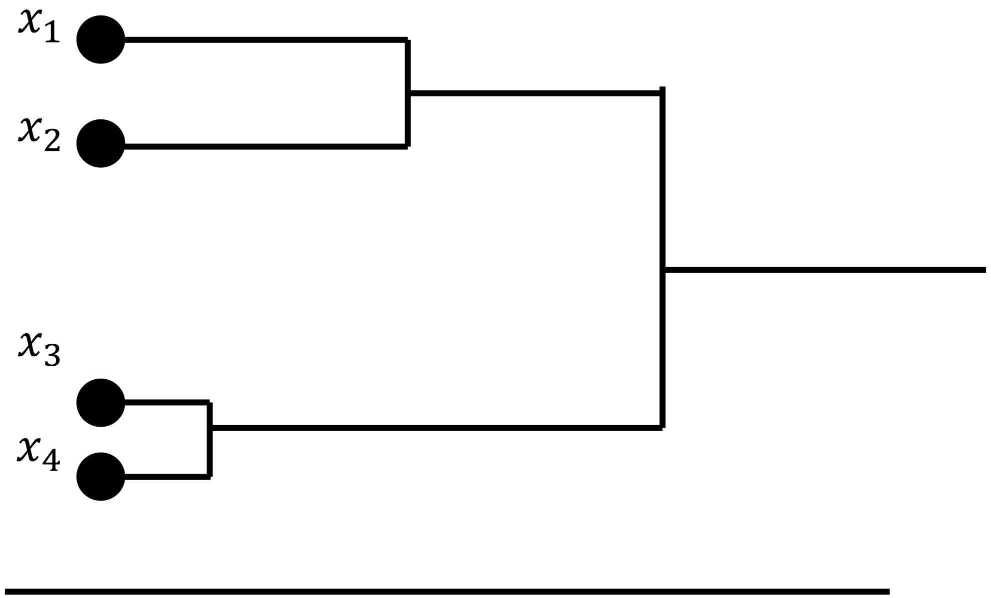
- X a set
- $d_X: X \times X \to [0, \infty)$
    - $d_X(x, x') = 0$ iff $x = x'$
    - (Symmetry) $d_X(x, x') = d_X(x', x)$
    - (Triangle inequality)
      $d_X(x, x') \leq d_X(x, x') + d_X(x', x'')$
      for all $x, x', x'' \in X$

Ultrametric space (X, u_X):

- X a set
- $u_X: X \times X \to [0, \infty)$
    - $u_X(x, x') = 0$ iff $x = x'$
    - (Symmetry) $u_X(x, x') = u_X(x', x)$
    - (Strong Triangle inequality)
      $u_X(x, x') \leq \max(u_X(x, x'), u_X(x', x''))$
      for all $x, x', x'' \in X$

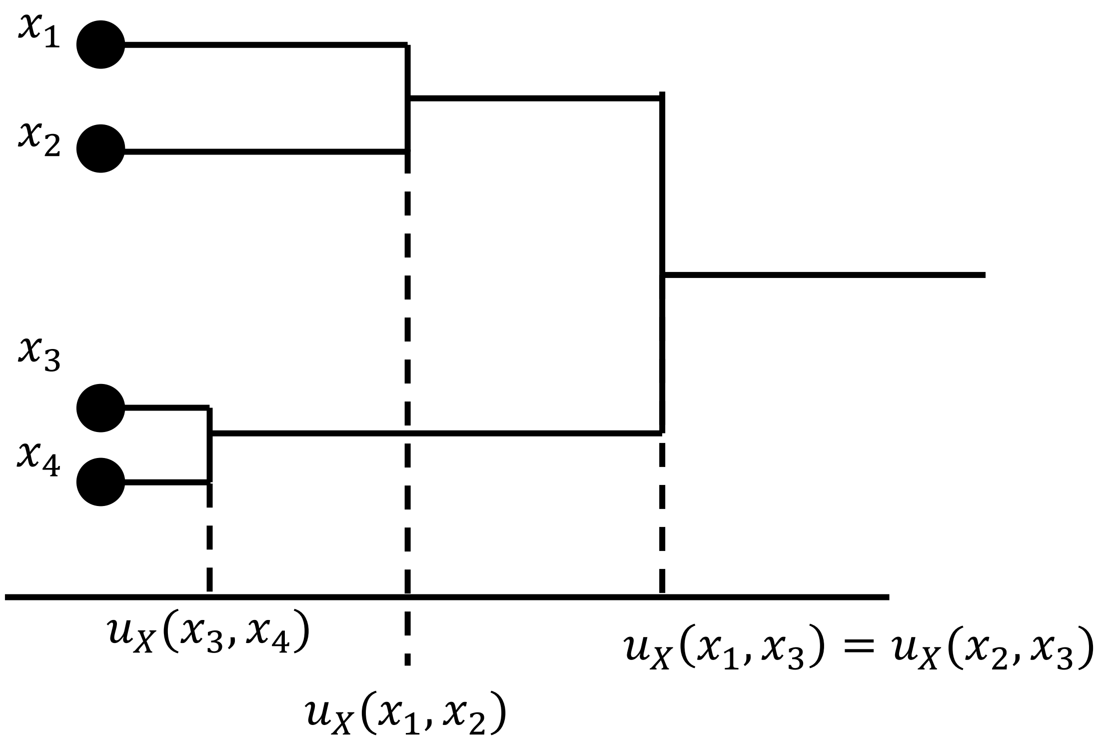
Single Linkage Clustering

Carlsson, Mémoli, *Classifying clustering schemes*



Single Linkage Clustering

Carlsson, Mémoli, *Classifying clustering schemes*

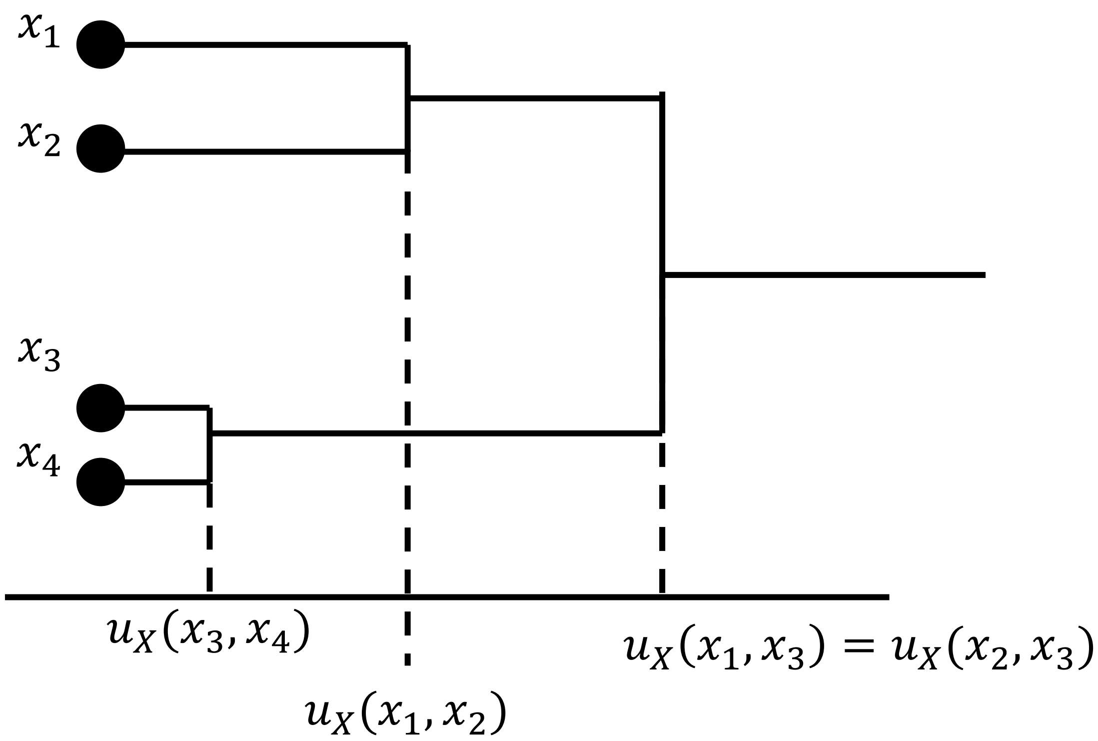


Fact: { dendrograms } ↔ {ultrametric spaces}

Exercise: Given an ultrametric space, draw the corresponding dendrogram

Single Linkage Clustering

Carlsson, Mémoli, *Classifying clustering schemes*

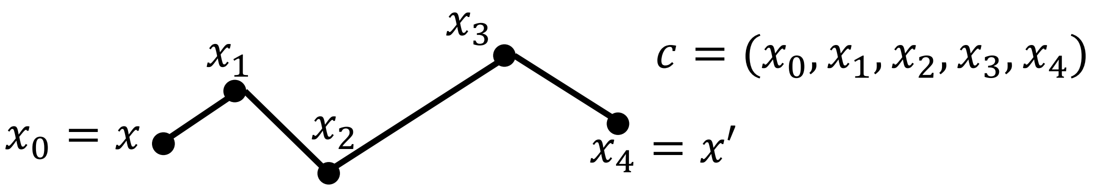


Fact: { dendrograms } ↔ {ultrametric spaces}

Exercise: Given an ultrametric space, draw the corresponding dendrogram

Example: **Single linkage clustering**

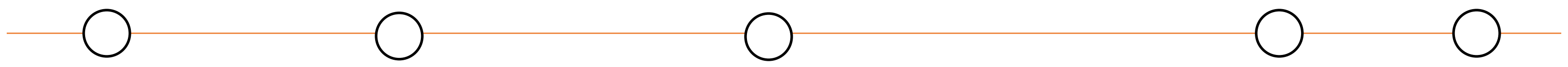
- Input: (X, d_X) finite metric space
- $C(x, x')$ = paths starting at x and ending at x'



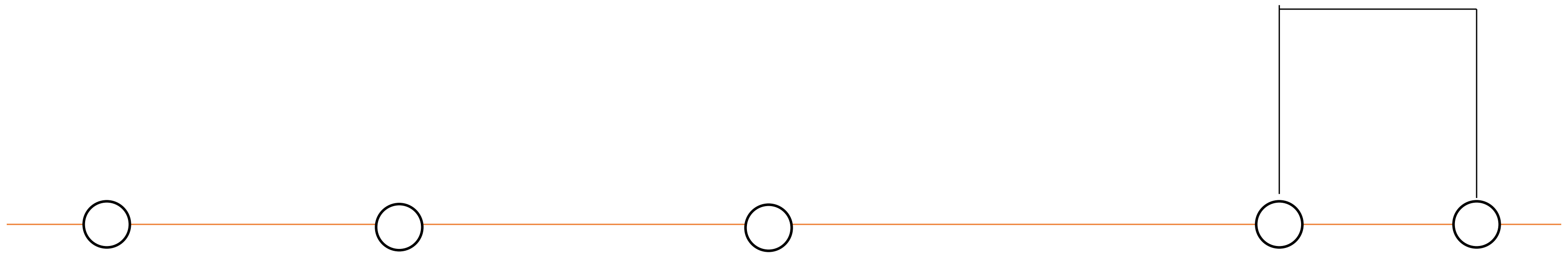
- For $c \in C(x, x')$, $cost(c) = \max_{x_i, x_{i+1} \in c} d_X(x_i, x_{i+1})$
- $u_X(x, x') := \min_{c \in C(x,x')} \text{cost}(c)$
- Output: (X, u_X) **ultrametric** space ↔ dendrogram

- Hierarchical clustering methods can be viewed as functors $\mathcal{H}: \mathcal{M} \to \mathcal{M}^{ult}$ (metric → ultrametric)
- Vietoris-Rips filtration + 0-dimensional persistence ≈ single linkage clustering!

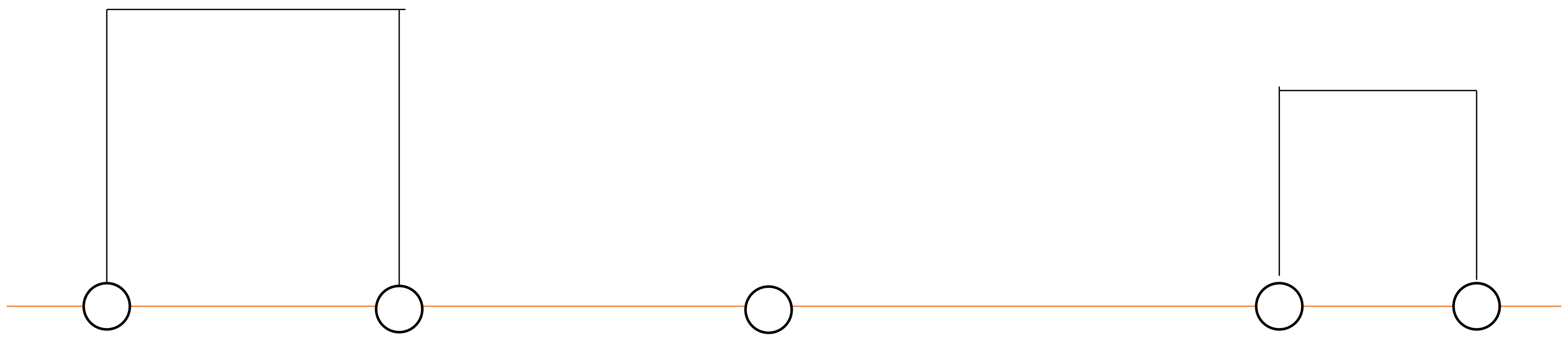
0-Dimensional VR Persistence and Single Linkage Clustering



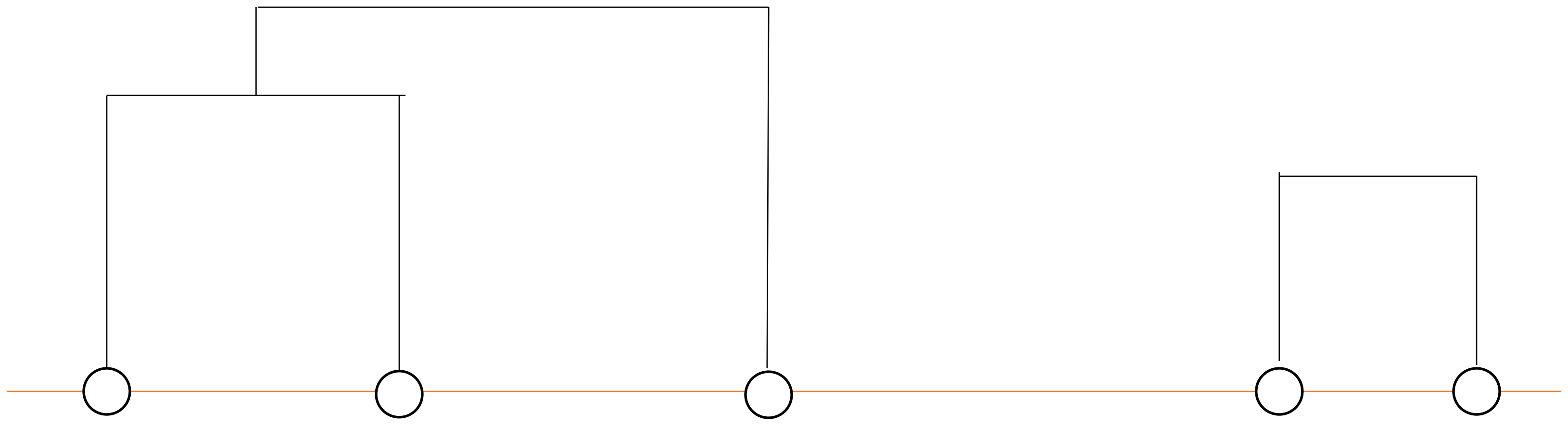
0-Dimensional VR Persistence and Single Linkage Clustering



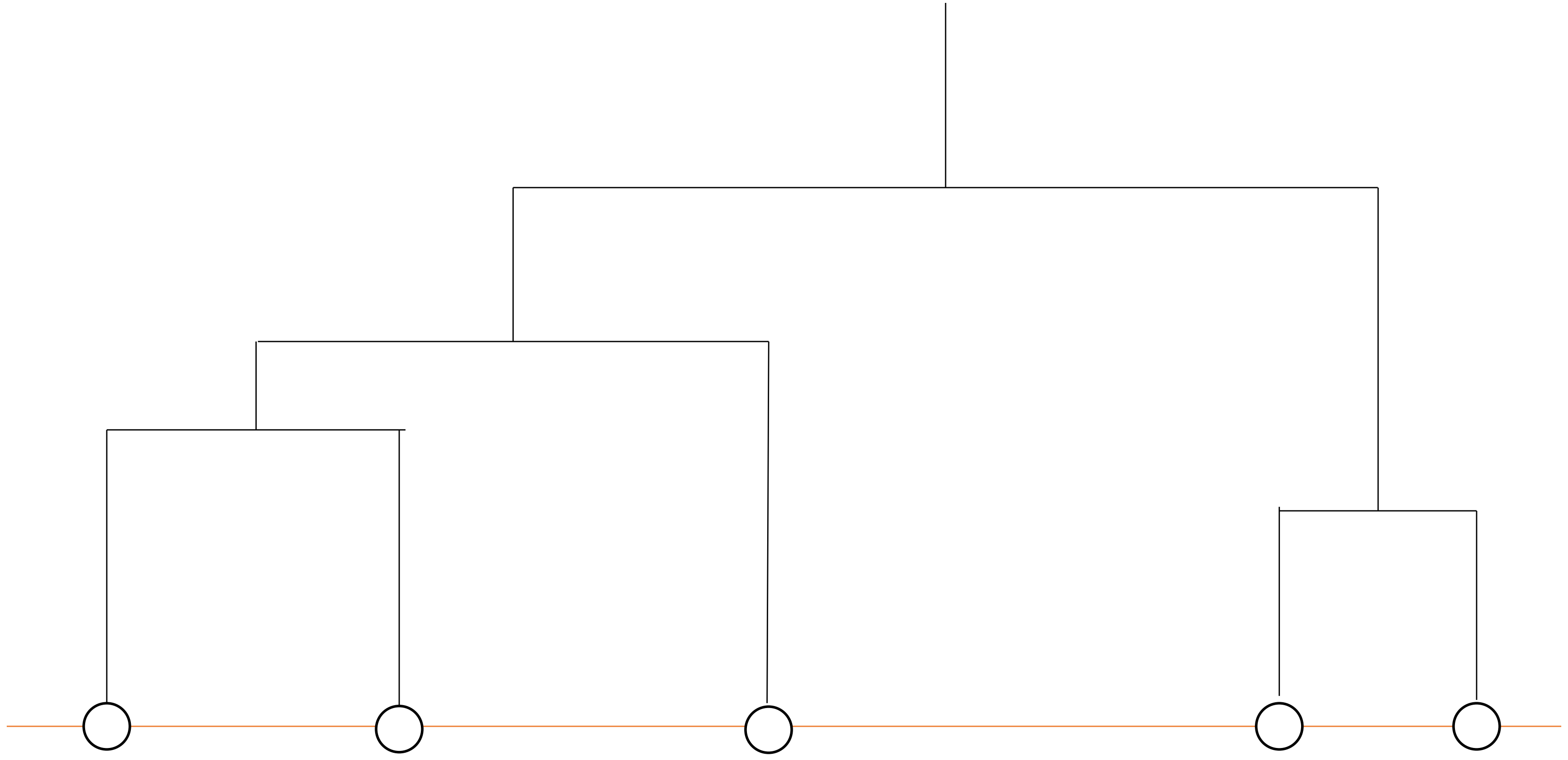
0-Dimensional VR Persistence and Single Linkage Clustering



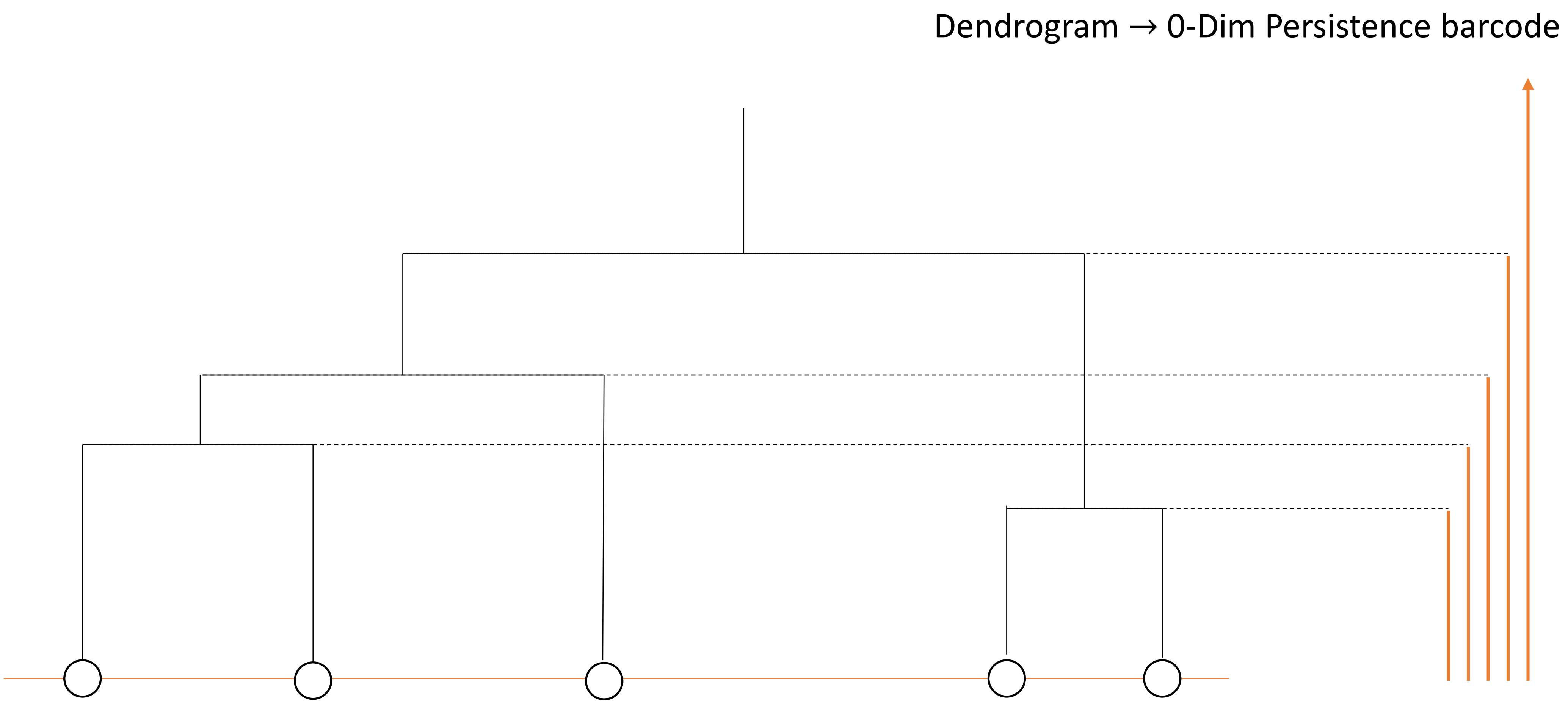
0-Dimensional VR Persistence and Single Linkage Clustering



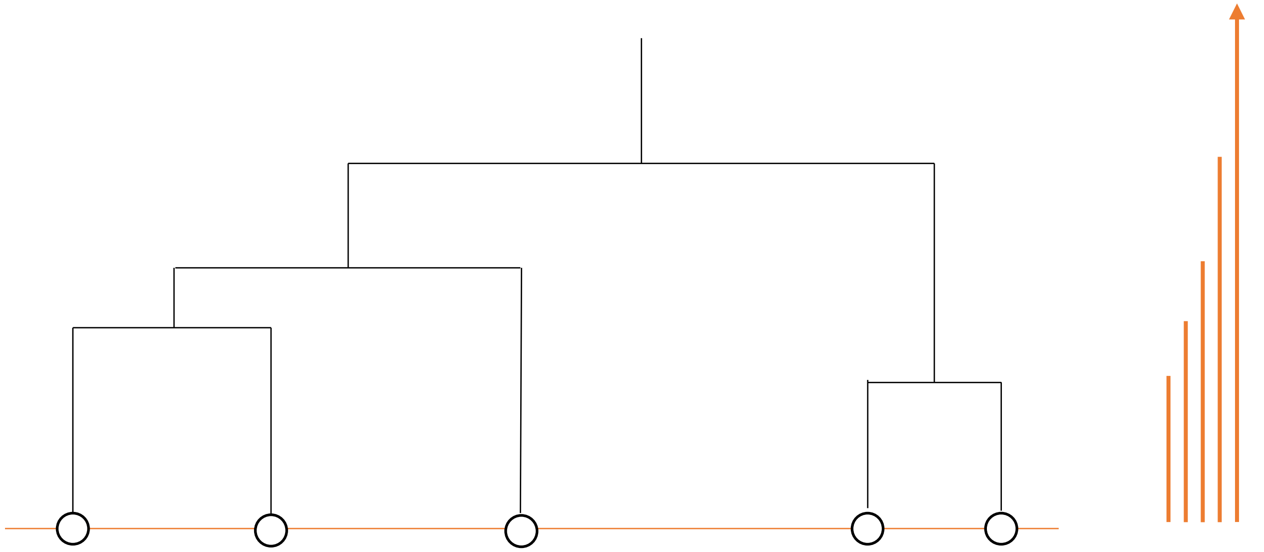
0-Dimensional VR Persistence and Single Linkage Clustering



0-Dimensional VR Persistence and Single Linkage Clustering



0-Dimensional VR Persistence and Single Linkage Clustering



- Stability theorem (Carlsson, Mémoli): Let (X, u_X) denote the output of SLHC on (X, d_X). Then,

$$d_{GH}\big((X, u_X), (Y, u_Y)\big) \leq d_{GH}\big((X, d_X), (Y, d_Y)\big).$$

Why compute 0-dimensional persistence?

- (Interpretation-computation tradeoff) 0-dimensional persistence loses label information but permits comparison via d_B (easier than d_{GH})

Scalar Field Analysis and Persistence-Based Clustering

Underlying Problem: Characterizing a Landscape from Limited Data

Chazal, Guibas, Oudot, Skraba, *Scalar Field Analysis over Point Cloud Data*. DCG 2011

Problem setup

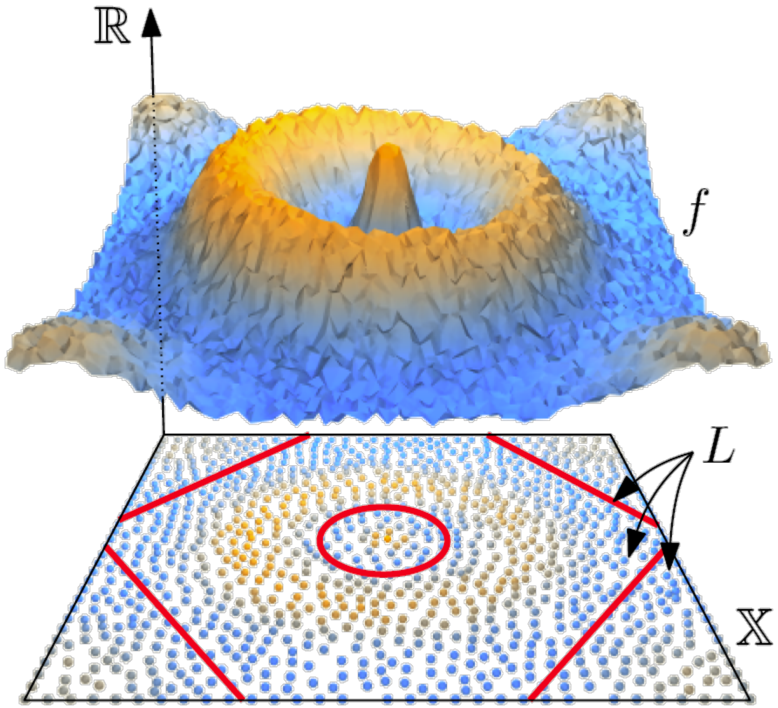
(X, d) a metric space, $f: X \to \mathbb{R}$ some function (say Lipschitz)

Given: $L \subseteq X$ a finite subsample, $f|_L$ restriction of f to L, $d|_{L \times L}$ restriction of metric to L

(not given: relative locations/coordinates of L to X, values of f outside L)

Goal: Recover salient properties of the graph of f, namely:

- peaks, valleys
- basins of attraction

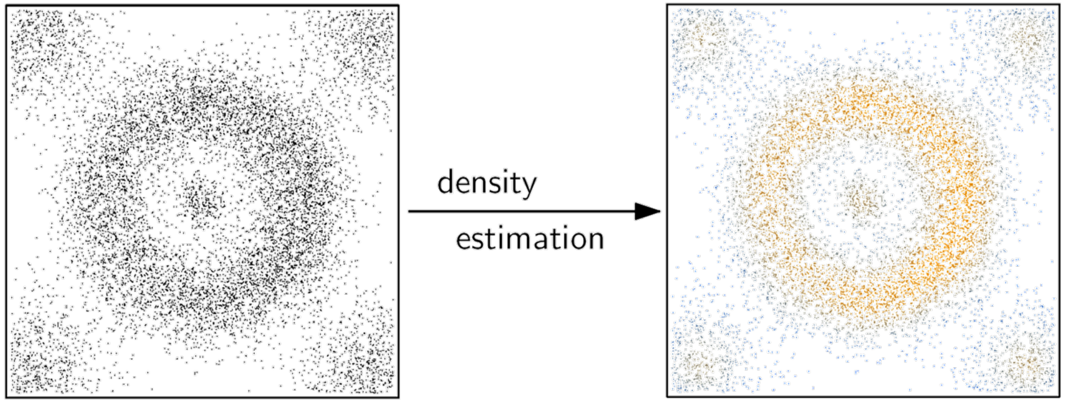


Motivating Applications

Chazal, Guibas, Oudot, Skraba, *Scalar Field Analysis over Point Cloud Data*. DCG 2011

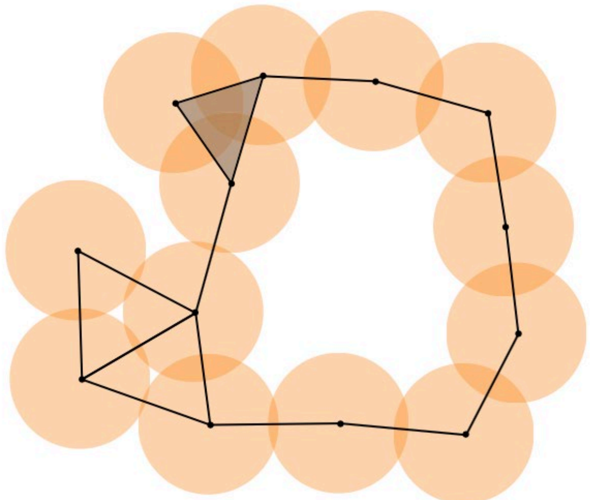
Unsupervised learning

- Data (and function values) sampled from some unknown underlying distribution f
- Obtain approximation through density estimator $\hat{f}$
- Characterize peaks, valleys, basins of attraction



Sensor networks

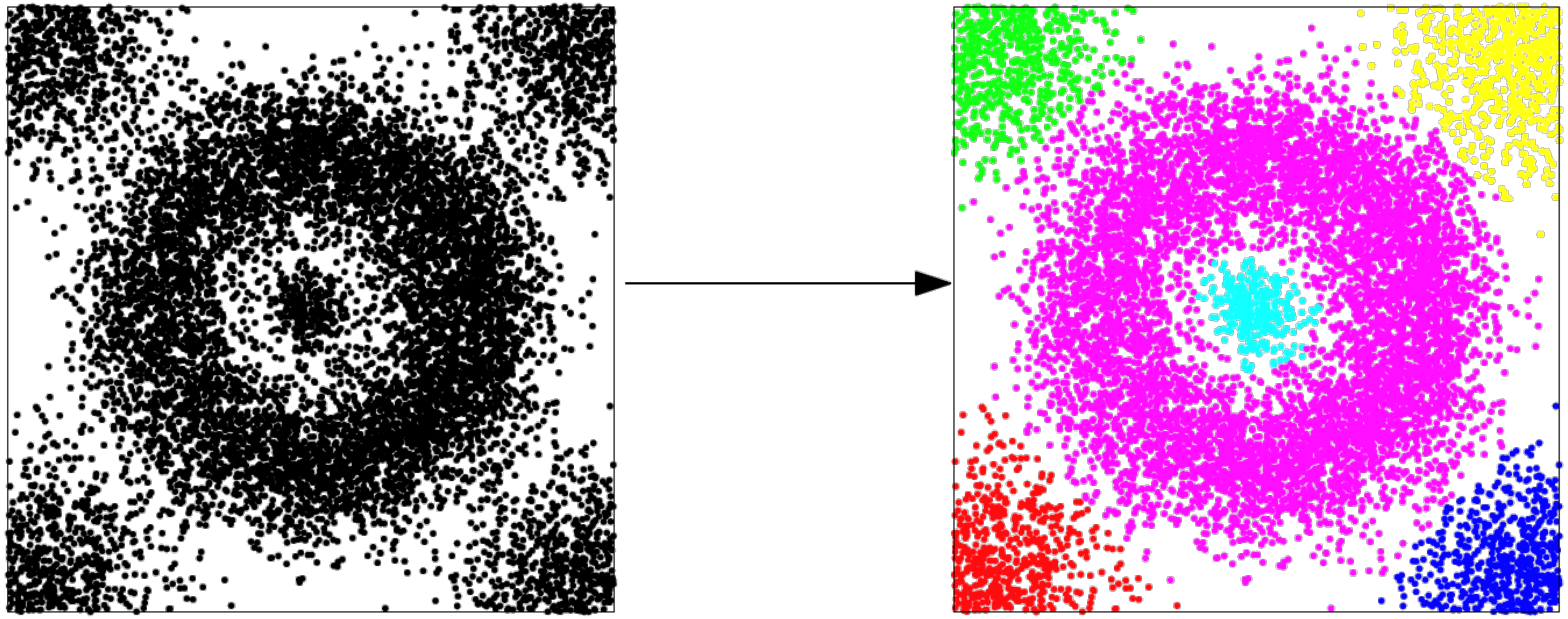
- Sensors distributed in some unknown area
- Measure some quantity f
- Communicate with nearby sensors
- **Want**: to understand landscape of f



Cluster Analysis

Input: a finite set of observations: - point cloud with coordinates

- distance / (dis-)similarity matrix



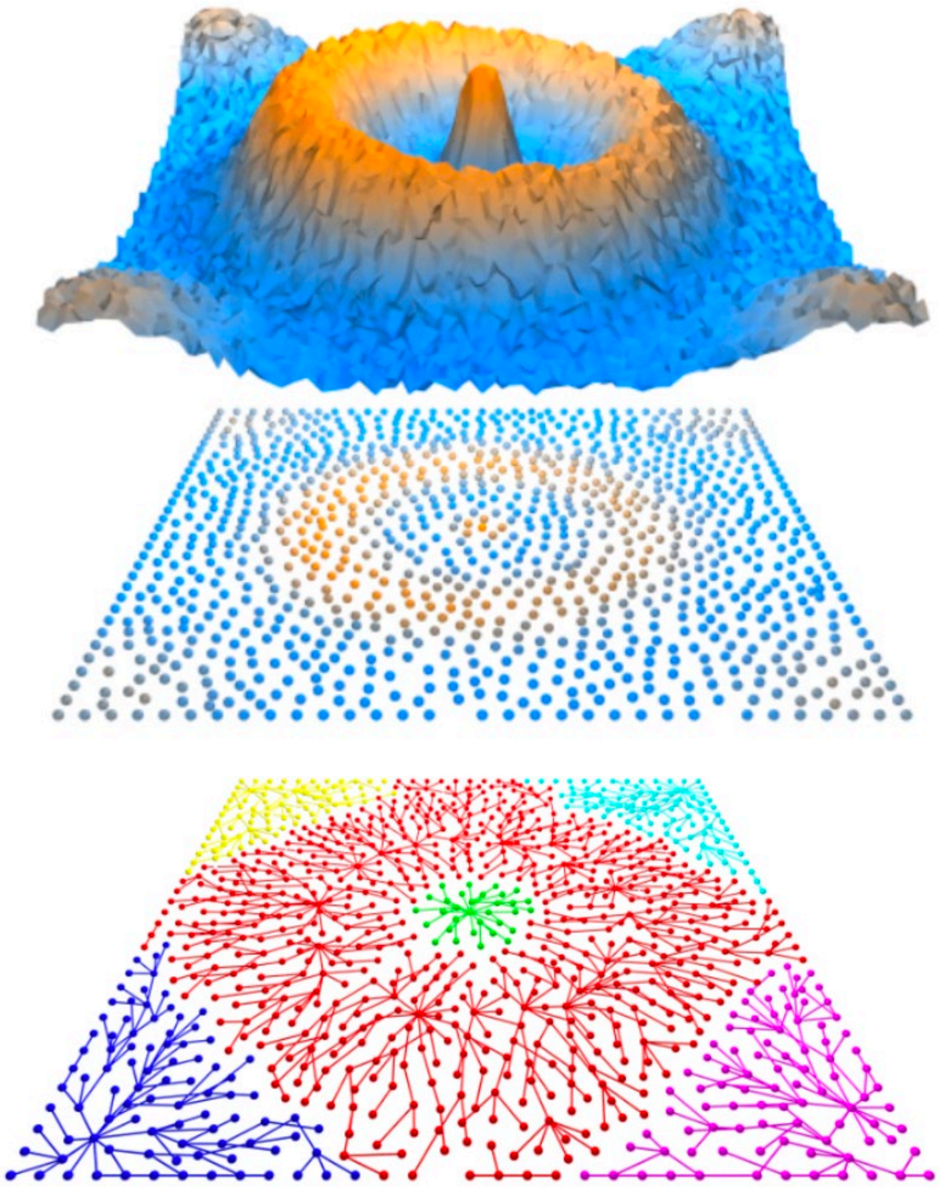
Task:

partition the data points into a collection of *relevant* subsets called clusters

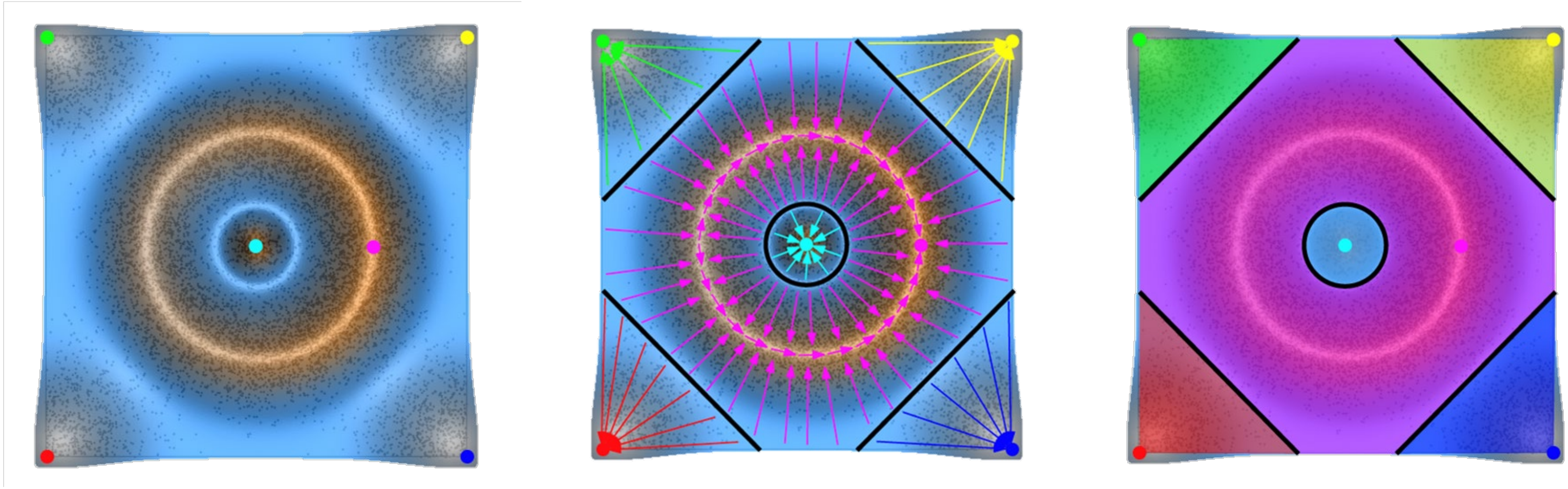
Overview

Mode-seeking paradigm (Mean-Shift and variants)

- Identify clusters with *basins of attraction*
- Problem: can be highly unstable/unpredictable
- Proactive solution: apply smoothing before hill-climbing (but unclear how this affects output, e.g. number of clusters)
- *Reactive* solution: Use topological persistence to detect and merge clusters *after* computation

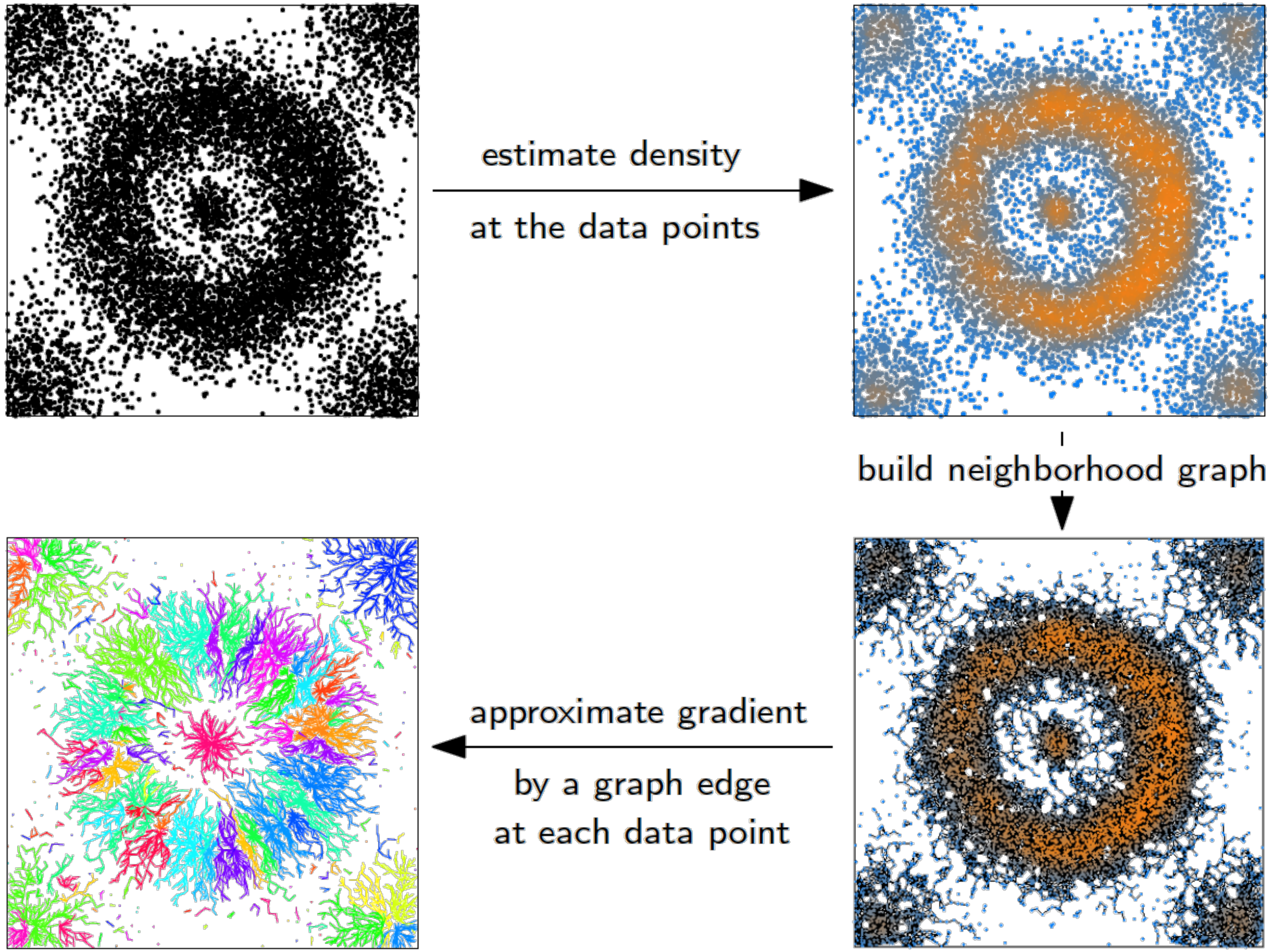


Mode-Seeking Paradigm

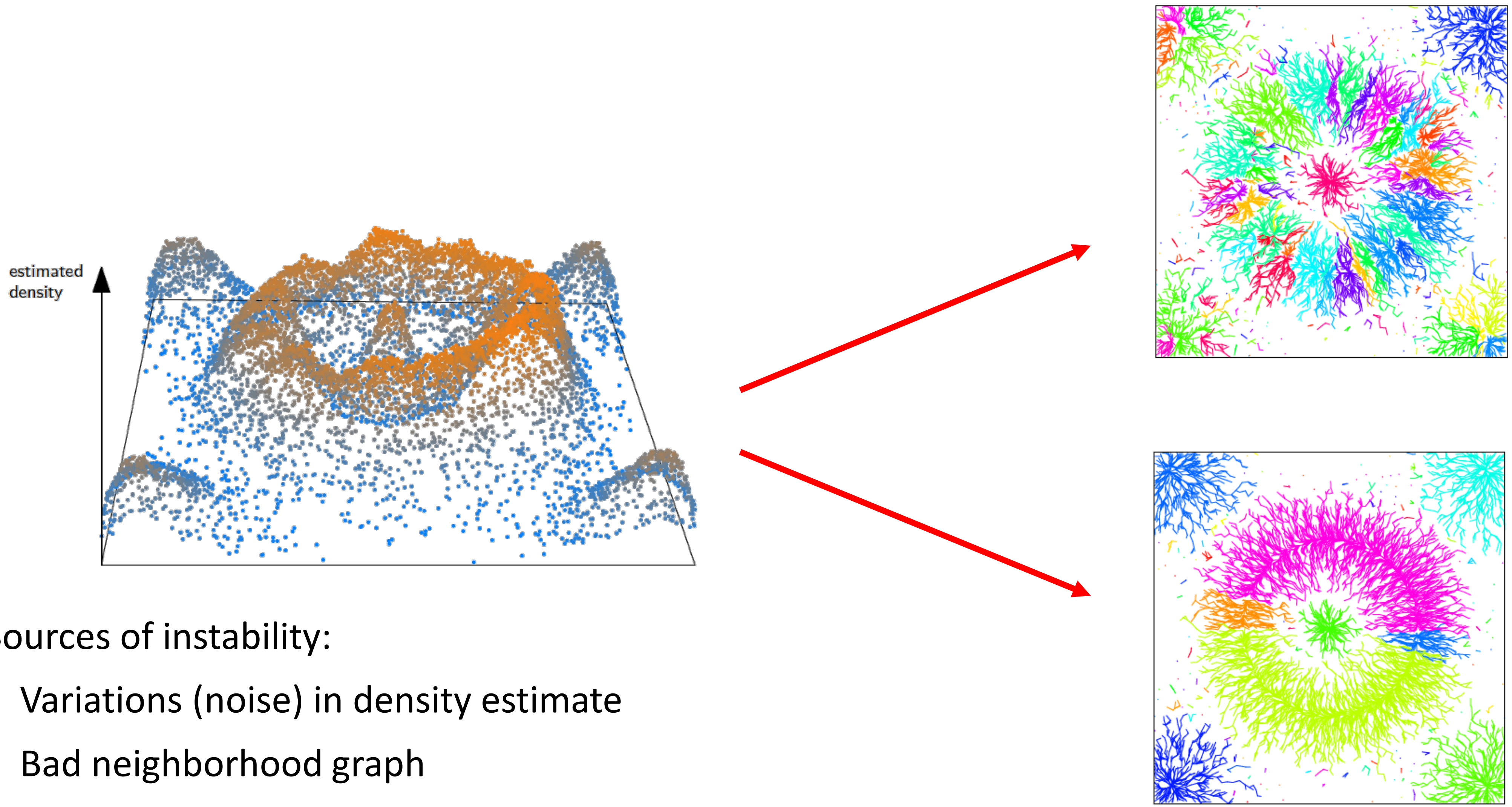


Mean-Shift and Variants

- Estimate density f
- Build neighborhood graph (e.g. via Vietoris-Rips at scale δ)
- Build spanning forest (= union of spanning trees) as follows: at each vertex v, either connect v to neighbor with highest f value or declare v to be a peak
- Trees in the forest become clusters with peaks as cluster centers



Instability in Density Estimate can Cause Unstable Output



Sources of instability:

- Variations (noise) in density estimate
- Bad neighborhood graph

Topological Mode Analysis Tool (ToMATo)

Chazal, Oudot, Skraba, Guibas *Persistence-Based Clustering in Riemannian Manifolds*

- Density estimator $\hat{f}$ defines an order on the point cloud

  (sort data points by **decreasing** estimated density values)

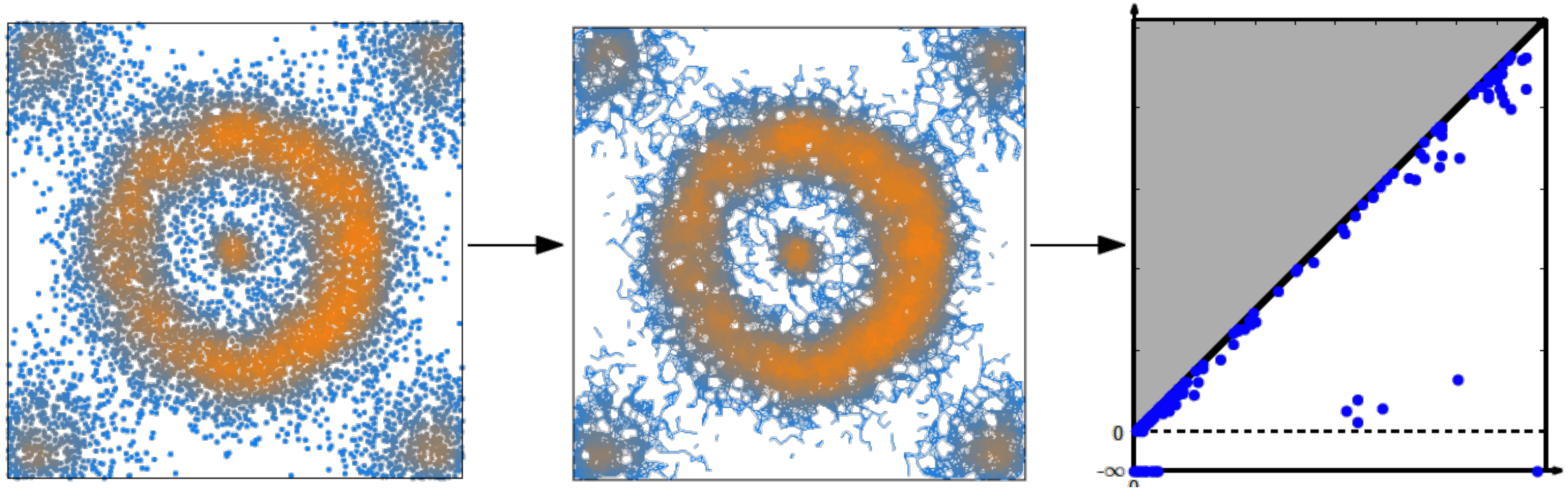
- Extend order to the graph edges $\rightarrow$ *upper-star filtration*

  $(\hat{f}([u,v]) = \min\{\hat{f}(u),\ \hat{f}(v)\})$

- Compute the 0-dimensional persistence diagram of this filtration

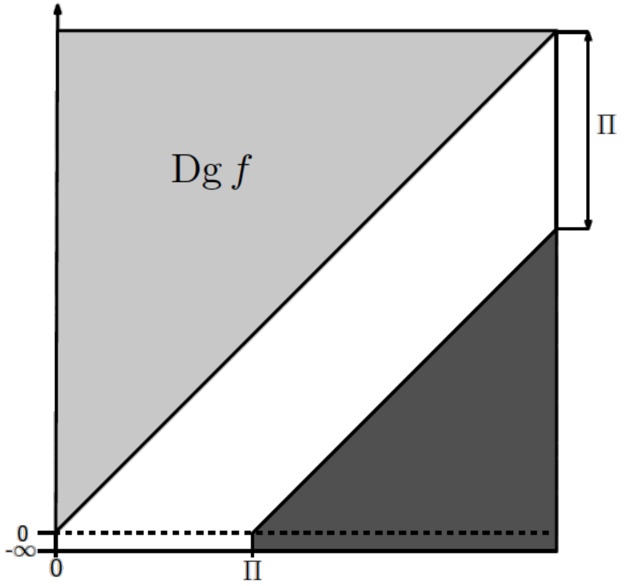
  (apply 0-dimensional persistence algorithm $\rightarrow$ union-find data structure)

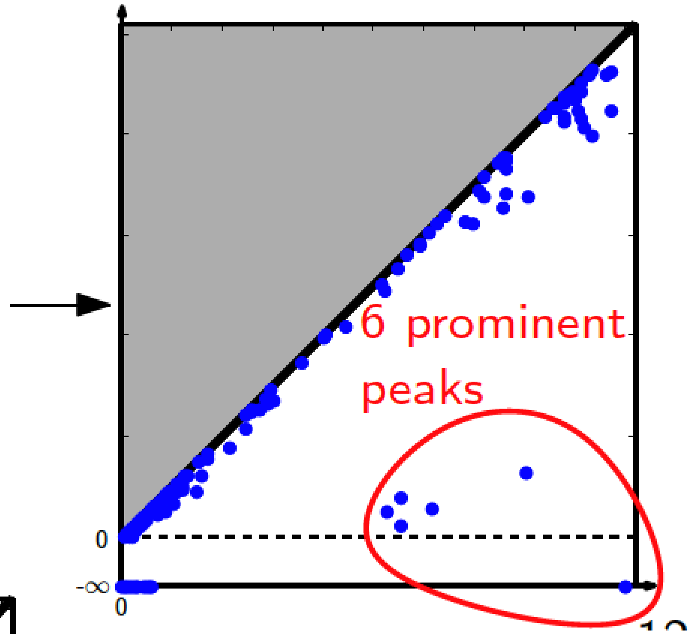
0-d persistence can be computed via Kruskall's minimum spanning tree algorithm + union-find data structure for $O(\alpha(n) \cdot n^2)$ complexity, where $\alpha(n)$ is an inverse Ackermann function

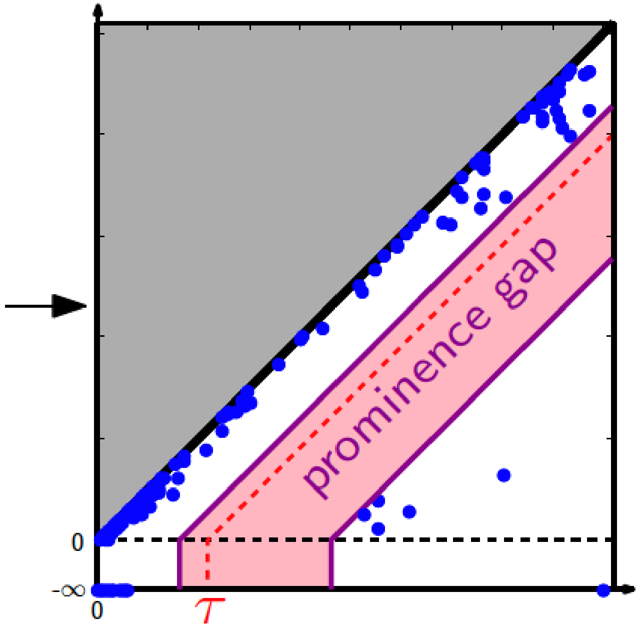


Q: Why the flipped diagram?

Estimating the Merge Parameter





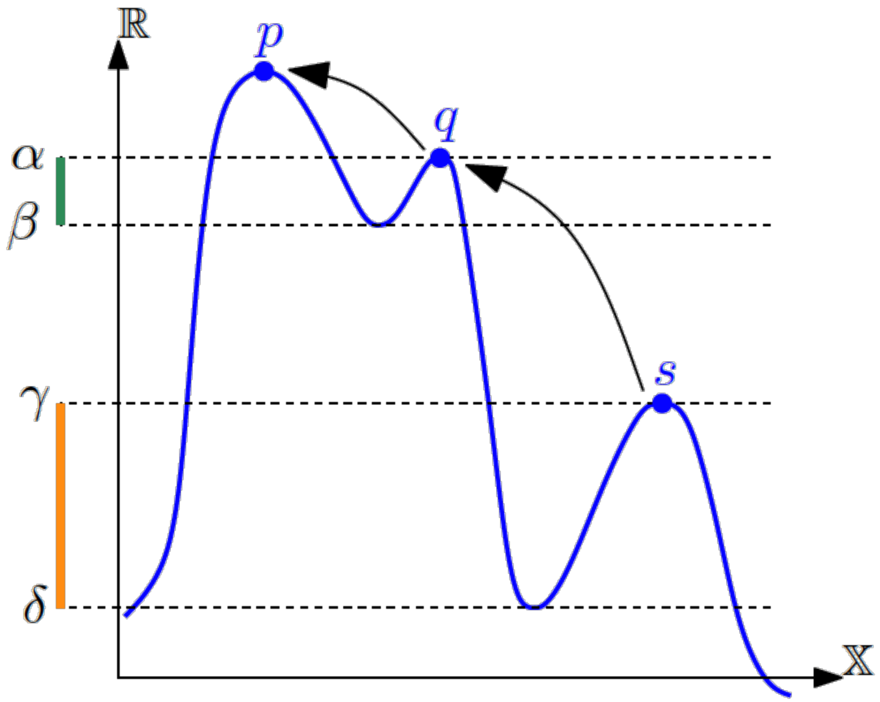


The gap parameter τ

Merging Clusters

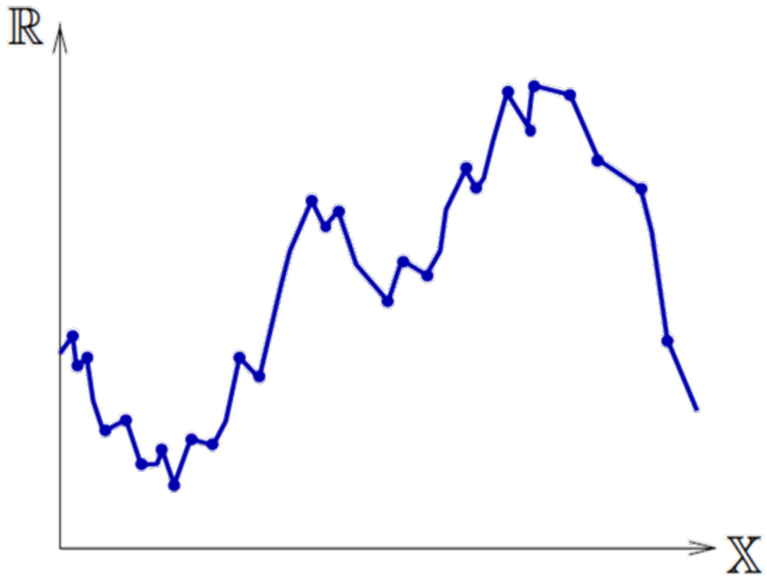
- 0-dimensional persistence builds a hierarchy of the peaks of $\hat{f}$ (merge tree)
- merge clusters according to the hierarchy (merge each cluster into its parent)
- given a fixed threshold $\tau \geq 0$, only merge those clusters of prominence $< \tau$

$$\gamma - \delta < \tau \leq +\infty$$



Prominence of a peak: Difference in height of peak and the level at which its basin of attraction meets that of a higher peak

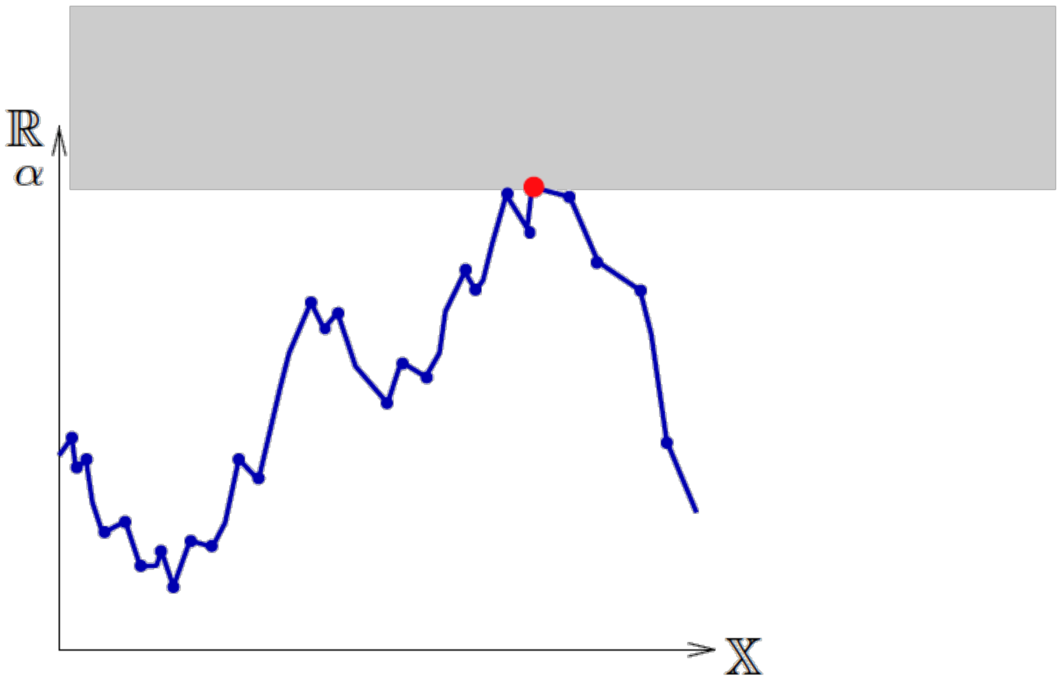
Computing Segments – Another Illustration



Merge clusters if prominence < τ

Keep separate otherwise

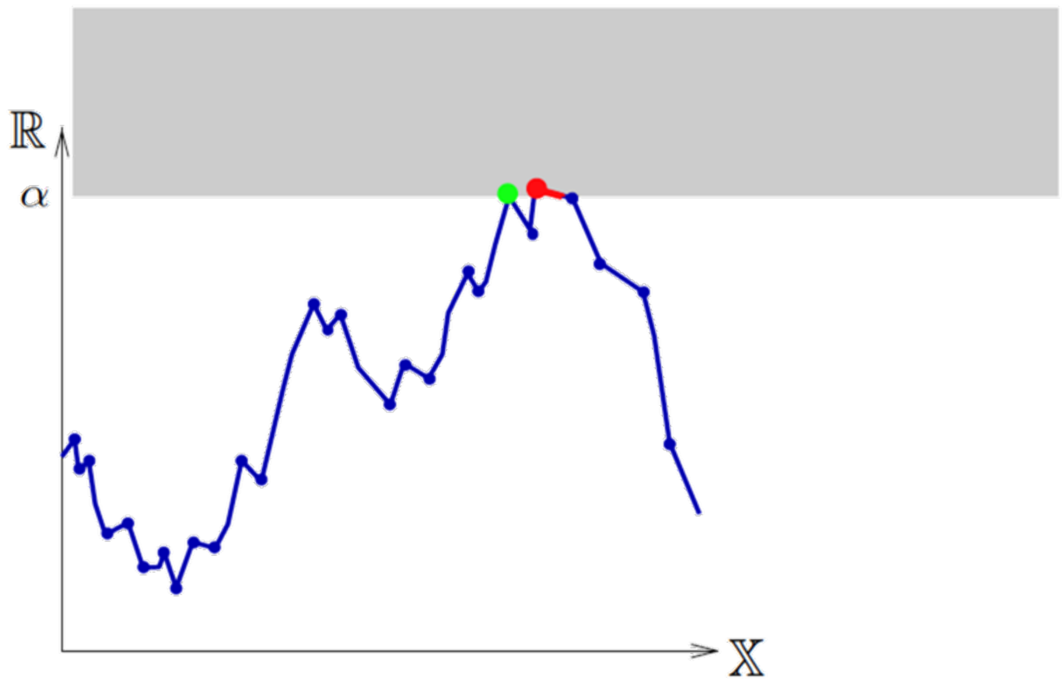
Computing Segments – Another Illustration



Merge clusters if prominence < τ

Keep separate otherwise

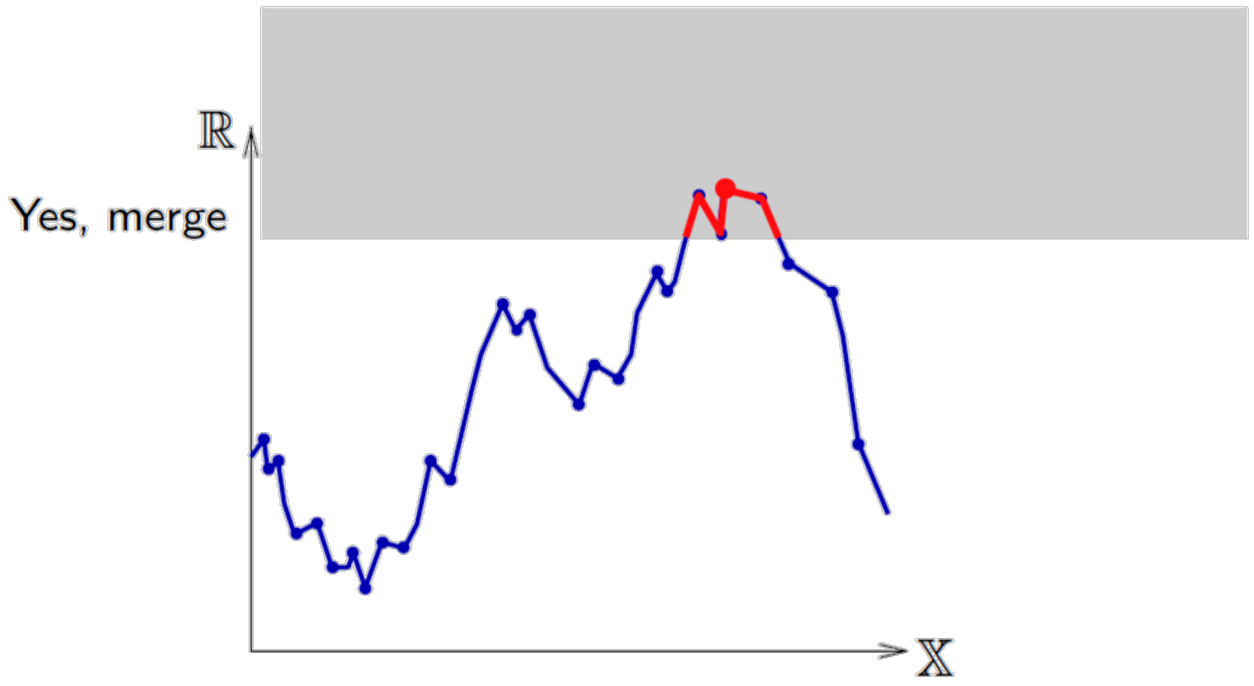
Computing Segments – Another Illustration



Merge clusters if prominence < τ

Keep separate otherwise

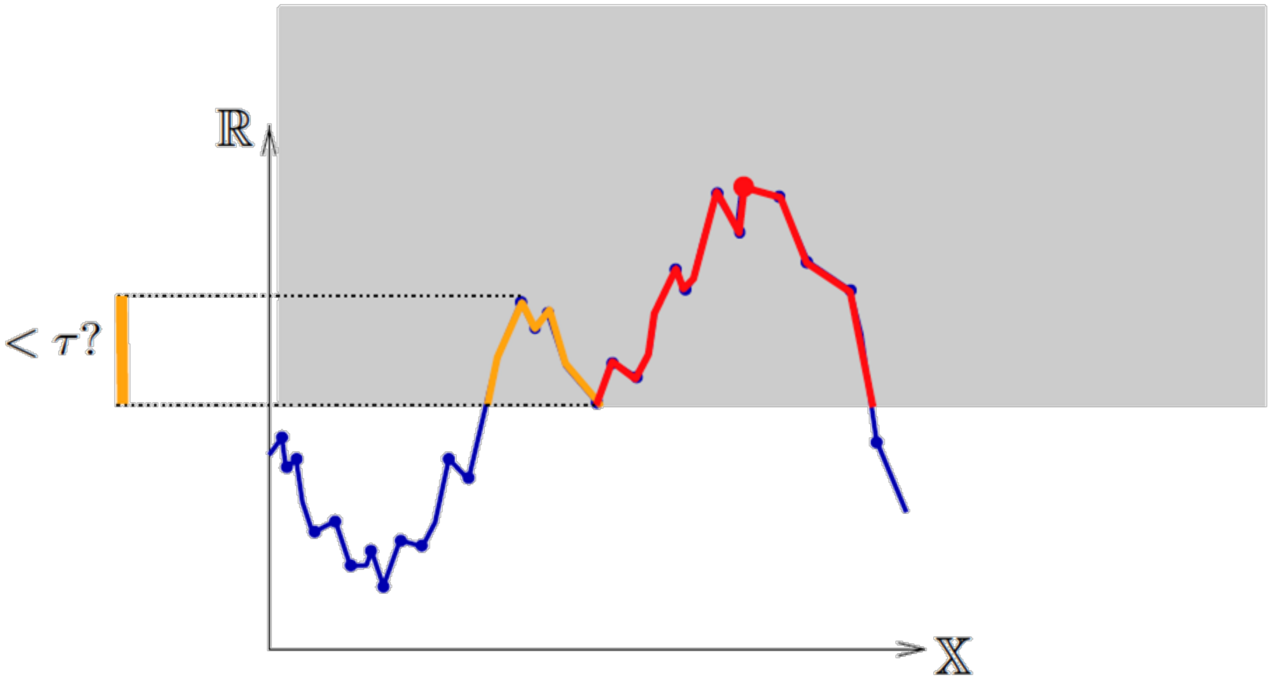
Computing Segments – Another Illustration



Merge clusters if prominence < τ

Keep separate otherwise

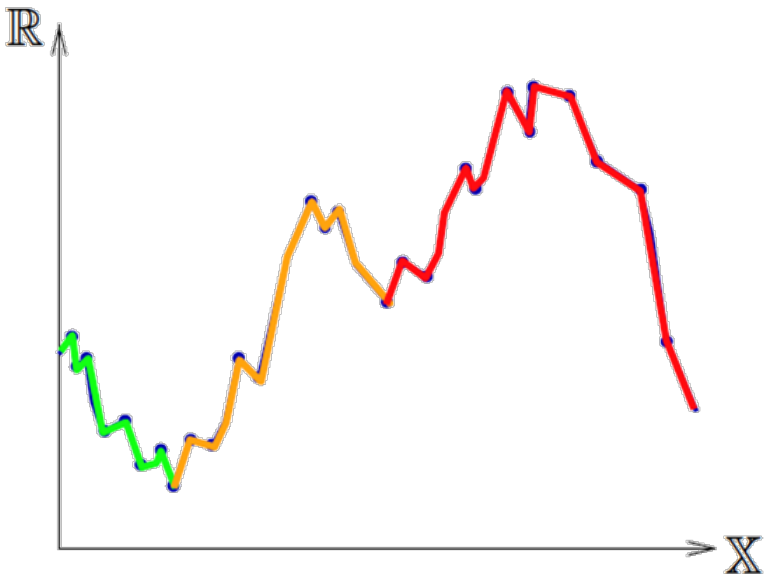
Computing Segments – Another Illustration

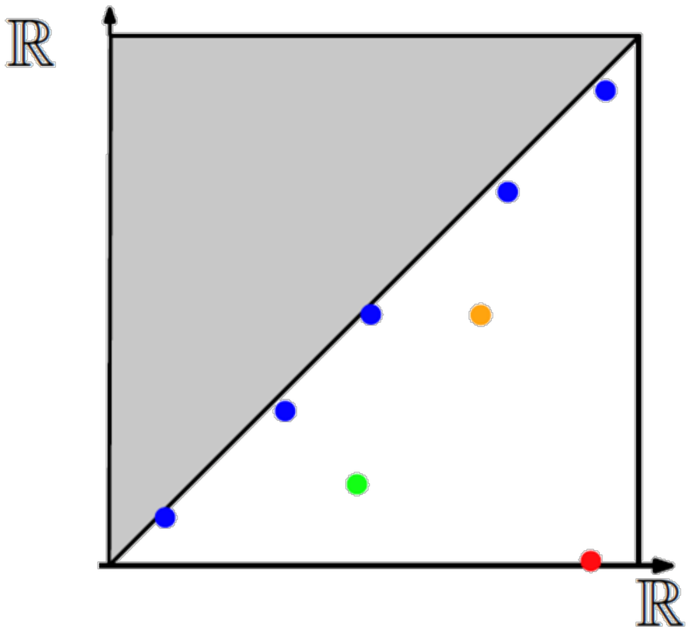


Merge clusters if prominence < τ

Keep separate otherwise

Computing Segments – Another Illustration

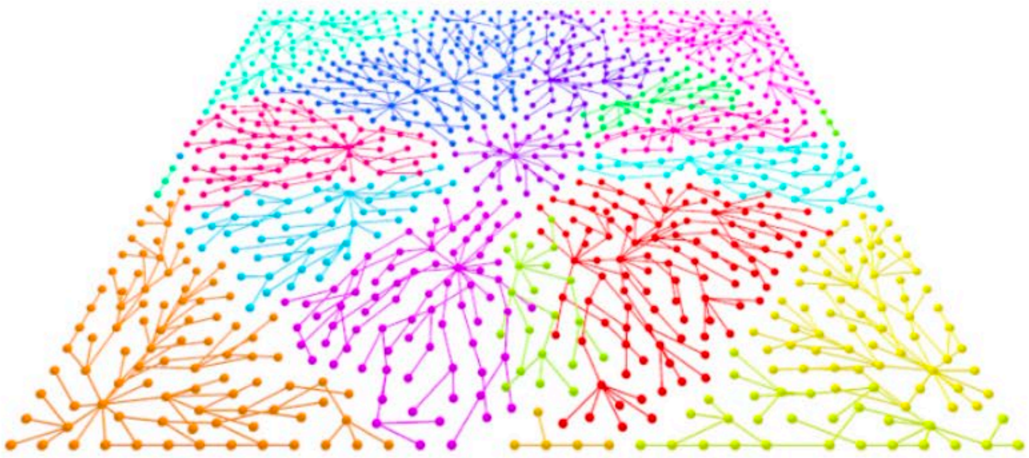




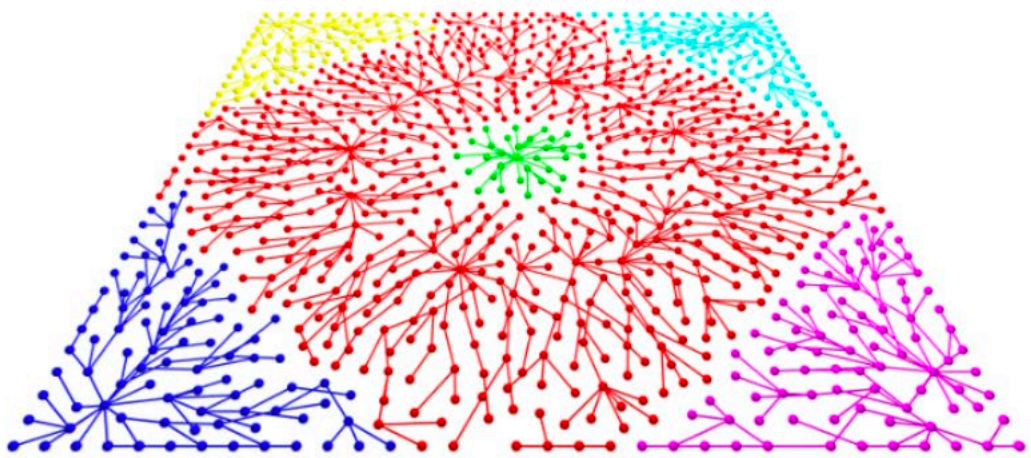
Merge clusters if prominence < τ

Keep separate otherwise

Effect of Persistence-Guided Cluster Merging

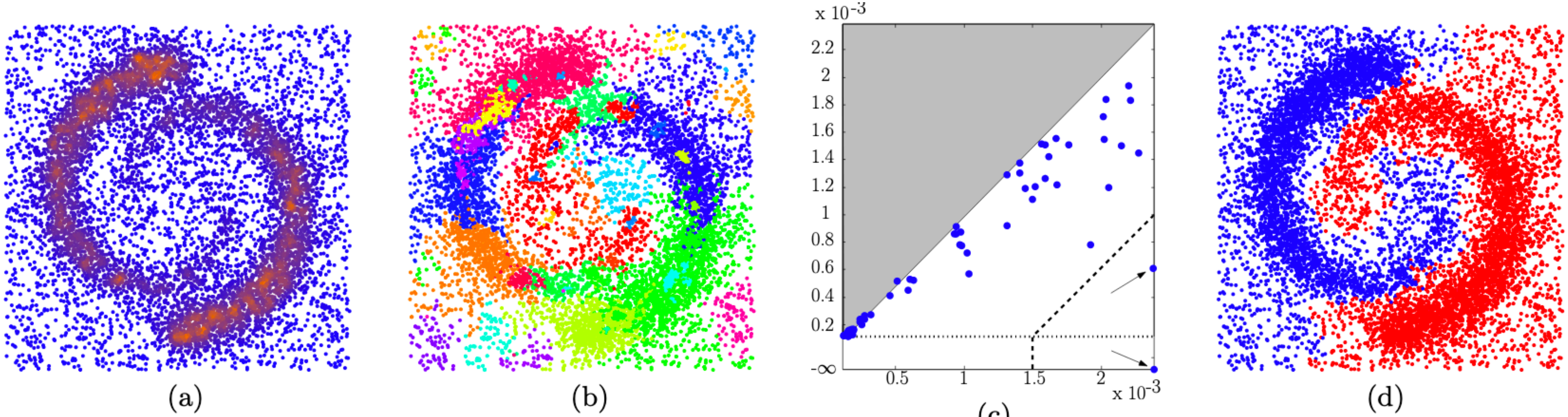


Output of hill climbing, before merge



After merge

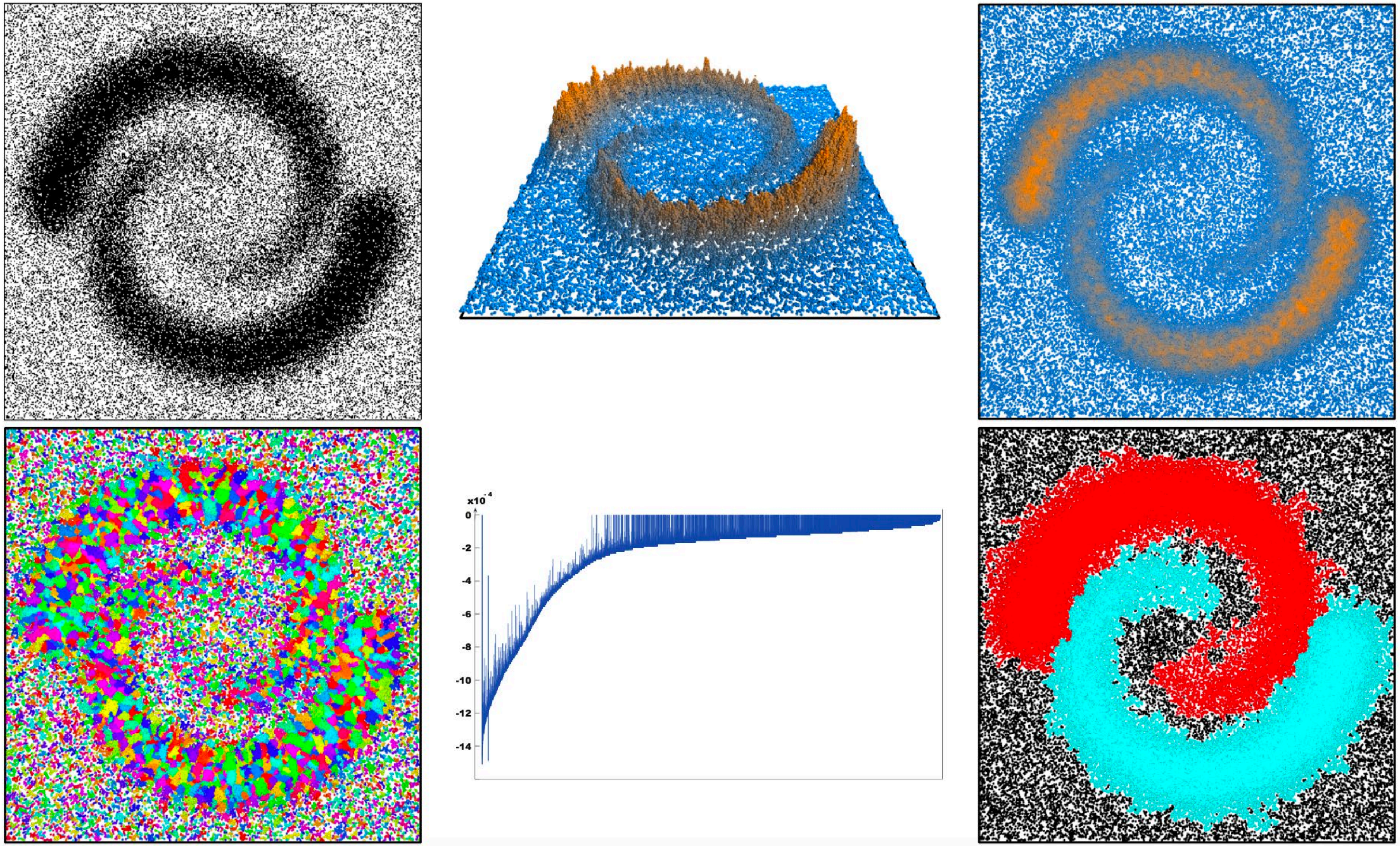
Effect of Persistence-Guided Cluster Merging



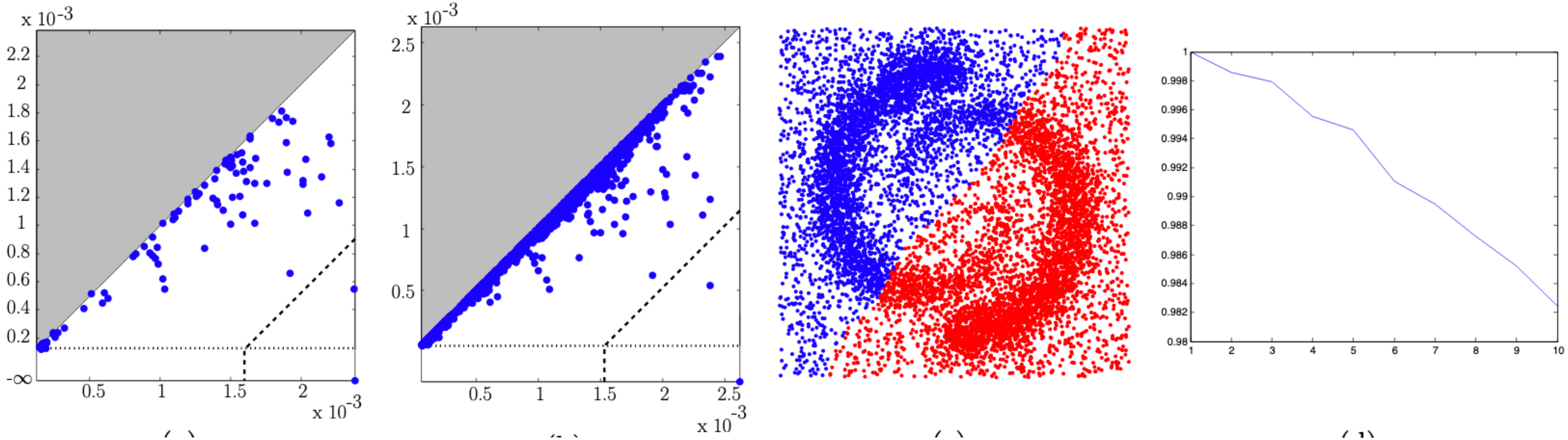
(a) (b) (c) (d)

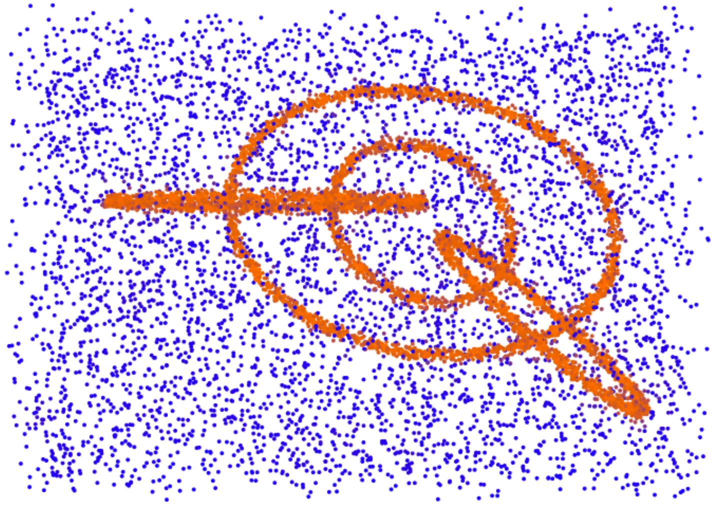
Theoretical guarantees available!

Effect of Persistence-Guided Cluster Merging

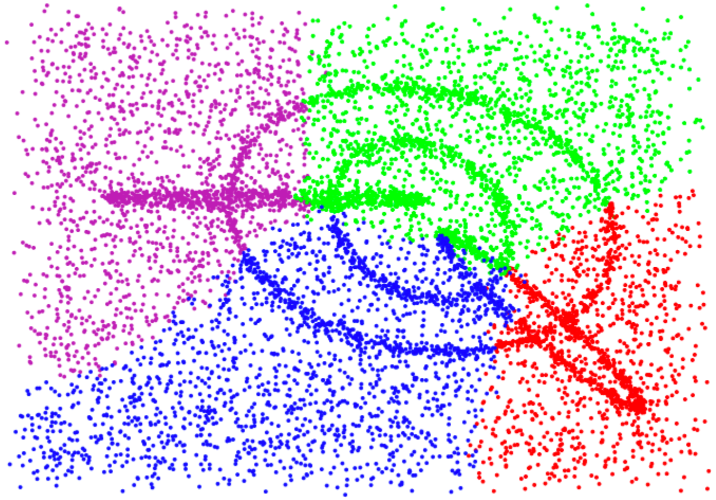


Comparison to Spectral Clustering

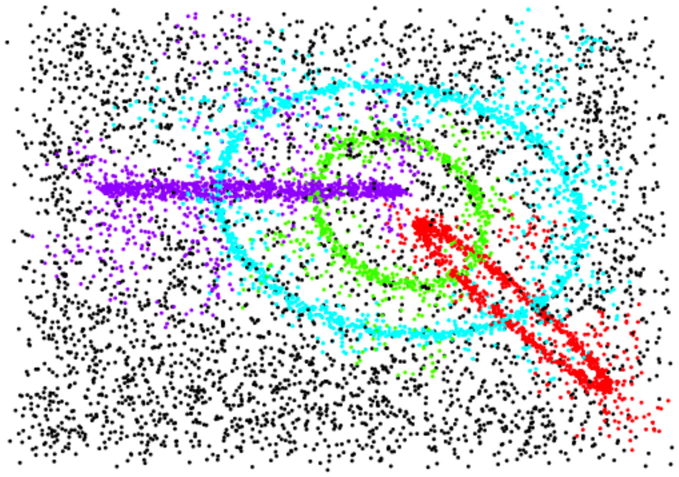




Data



Spectral

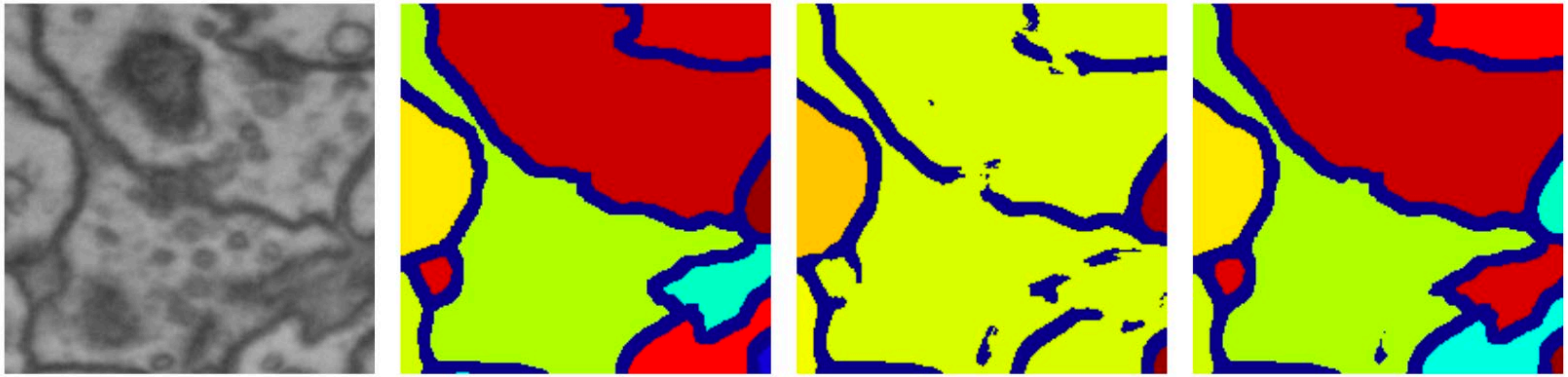


ToMATo

Topological Regularization for Deep Learning Applications

Topology for Image Segmentation

Hu, Fuxin, Samaras, Chen. *Topology-preserving deep image segmentation. NeuRIPS 2019*

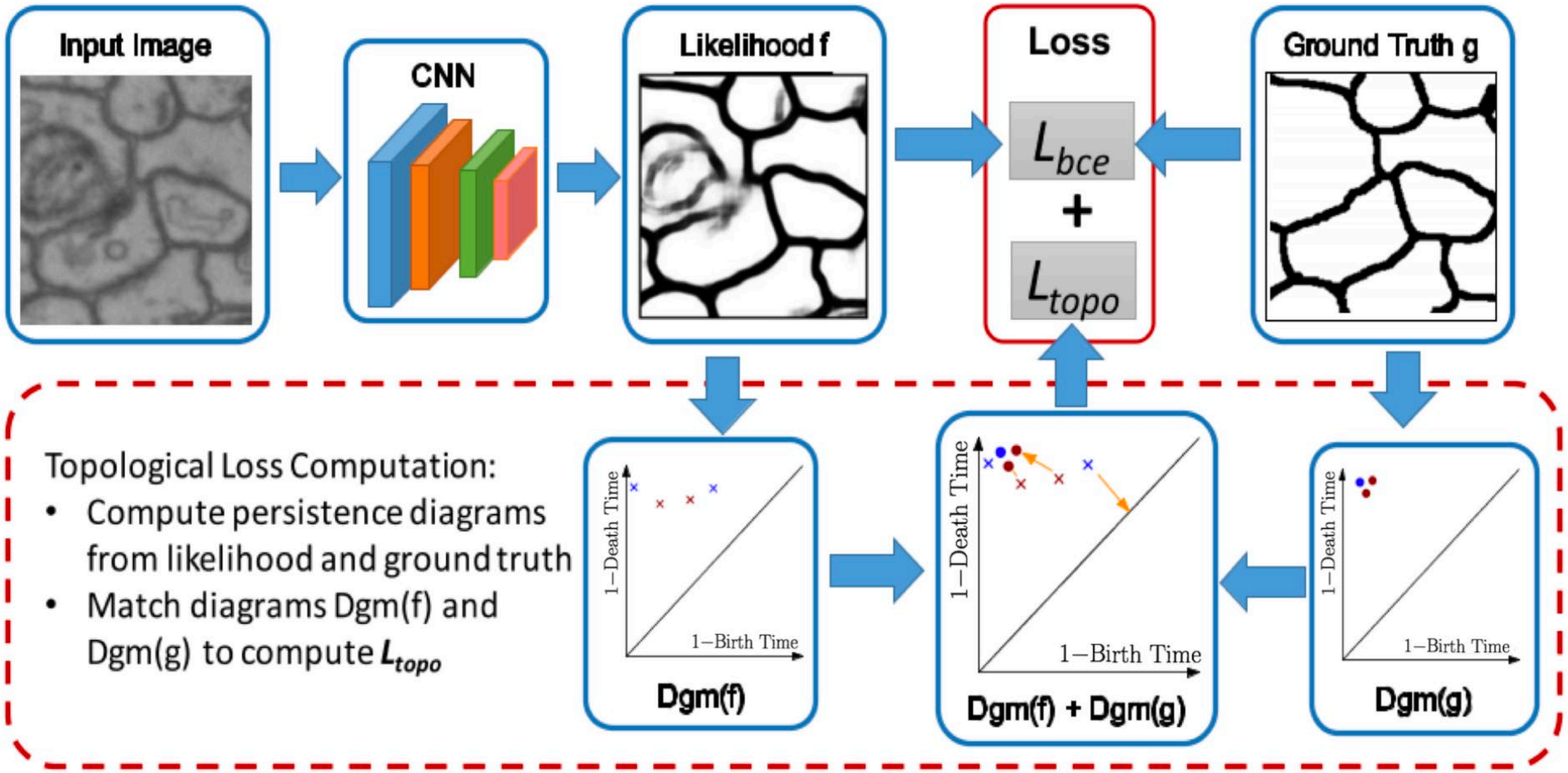


(a) (b) (c) (d)

- Segmentation errors in fine structures (e.g. membranes and vessels in biomedical imaging) introduce marginal per-pixel error, but major functional error
- Figure above: (a) input neuron image, (b) ground-truth segmentation, (c) baseline method without topological guarantees, (d) TopoLoss reconstruction
- Why *persistent* homology was needed: Betti numbers are discrete, but need a continuous, differentiable function for backprop---provided by persistent homology

Pipeline

Hu, Fuxin, Samaras, Chen. *Topology-preserving deep image segmentation. NeuRIPS 2019*

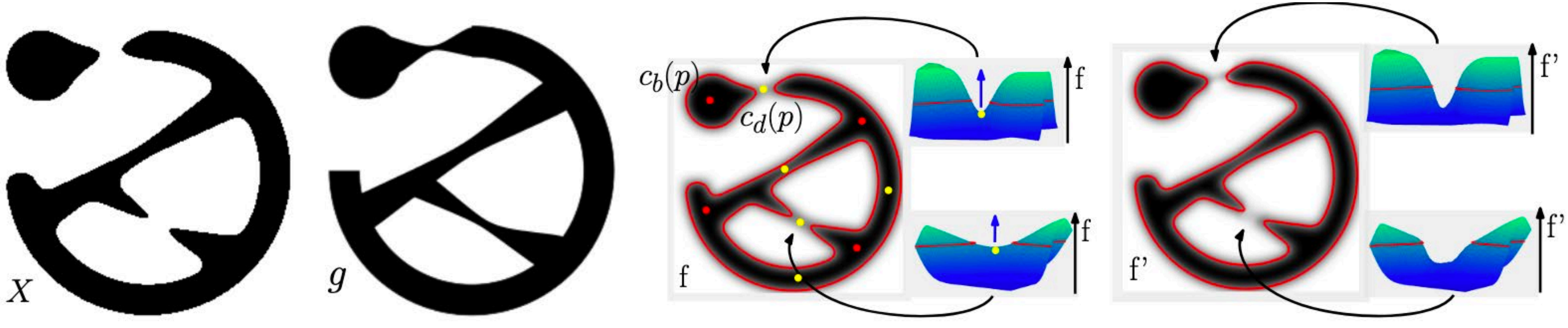


f the likelihood map predicted by the network, g the ground truth, L_{ce} the cross-entropy loss, L_{top} the topological loss, λ a regularization parameter. Loss function:

$$L(f,g) = L_{ce}(f,g) + \lambda L_{top}(f,g)$$

Superlevel-Set Filtration on Likelihood

Hu, Fuxin, Samaras, Chen. *Topology-preserving deep image segmentation. NeuRIPS 2019*



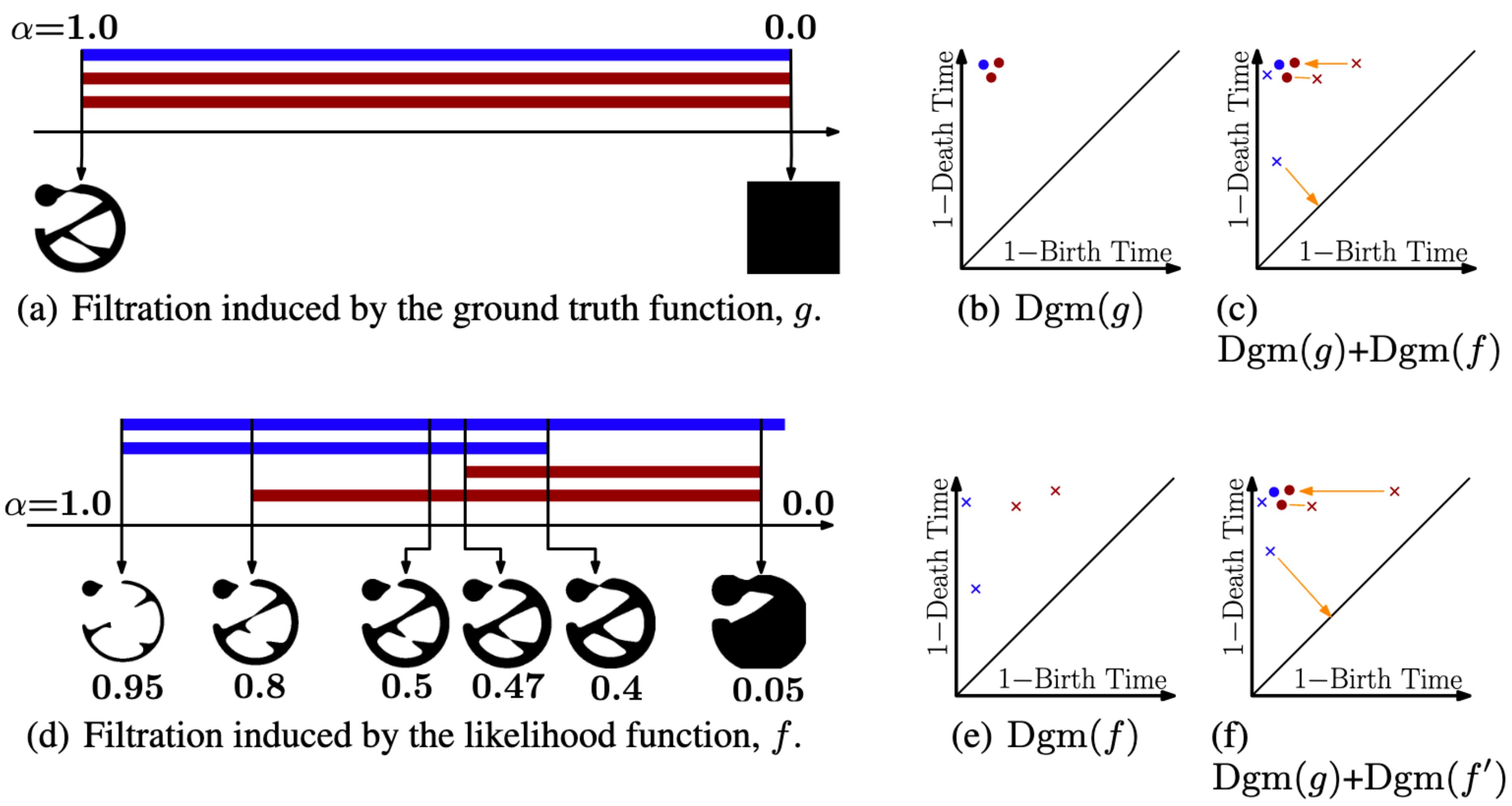
(a) (b) (c) (d)

Note: likelihood maps are valued in [0,1]

- (a) example segmentation X with one loop and two connected components
- (b) ground truth segmentation with two loops and one connected component---view g as binary $\{0,1\}$-valued map
- (c,d) likelihood maps f, f', both produce segmentation X if we threshold at $\alpha = 0.5$. But f is closer to being correct, due to shallower gaps
- Conclusion: single threshold $\alpha = 0.5$ not enough, but persistent homology $\rightarrow$ all possible threshold values

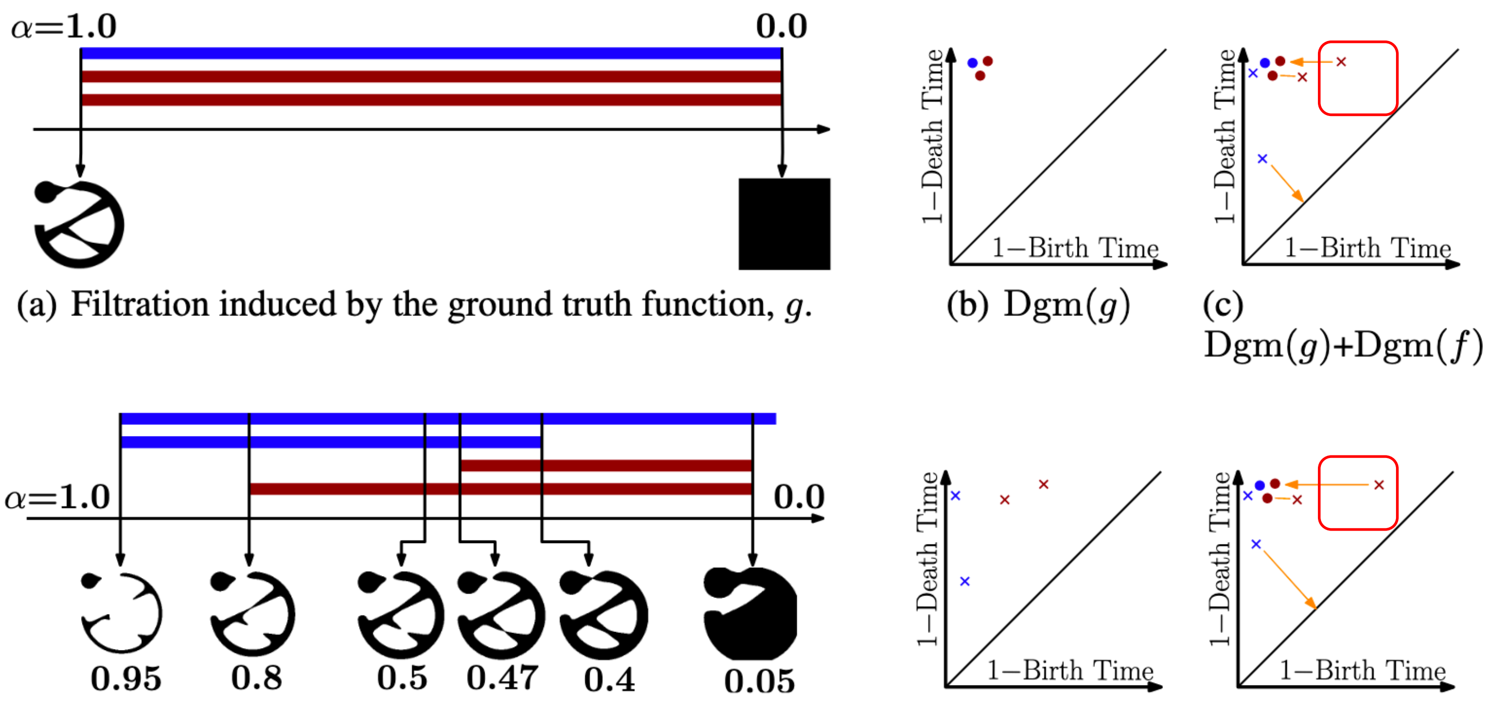
Topological Loss

Hu, Fuxin, Samaras, Chen. *Topology-preserving deep image segmentation. NeuRIPS 2019*



Topological Loss

Hu, Fuxin, Samaras, Chen. *Topology-preserving deep image segmentation. NeuRIPS 2019*



(a) Filtration induced by the ground truth function, g.
(b) Dgm(g)
(c) Dgm(g)+Dgm(f)
(d) Filtration induced by the likelihood function, f.
(e) Dgm(f)
(f) Dgm(g)+Dgm(f')

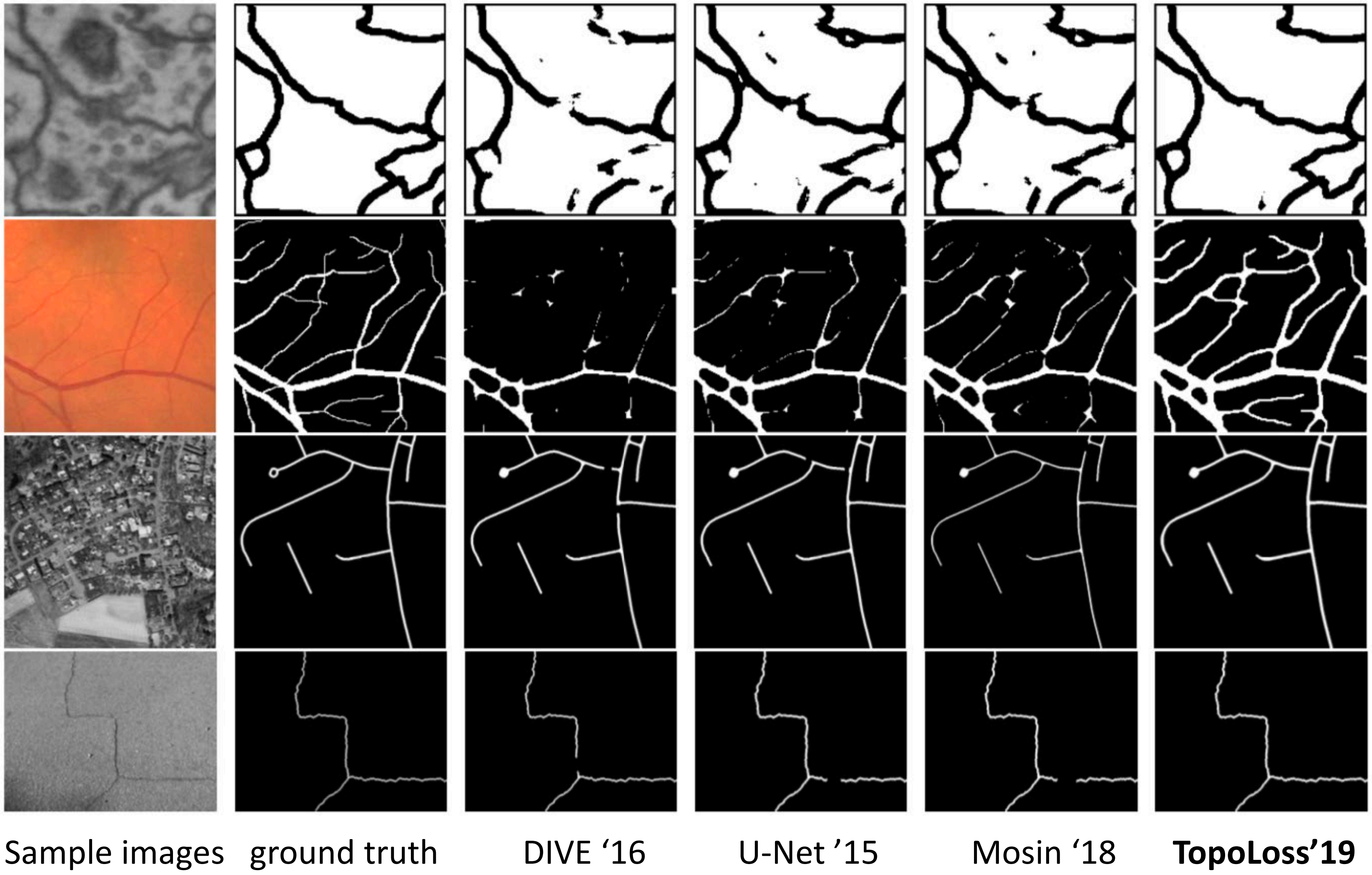
A slightly modified bottleneck distance: γ^* an optimal matching $Dgm_k(f) \to Dgm_k(g)$, $k \geq 0$,

$$L_{top}(f,g) = \sum_k \sum_{p \in Dgm_k(f)} \|p - \gamma^*(p)\|_2^{\,2} = \sum_k \sum_{p \in Dgm_k(f)} [birth(p) - birth(\gamma^*(p)]^2 + \left[death(p) - death\left(\gamma^*(p)\right)\right]^2$$

Intuition: $L_{top}(f,g)$ measures the cost of changing f to look like g.

Results

Hu, Fuxin, Samaras, Chen. *Topology-preserving deep image segmentation. NeurIPS 2019*



A closer Look at the Topological Loss

Poulenard, Skraba, Ovsjanikov. *Topological Function Optimization for Continuous Shape Matching. SGP '18*

- Neural network parameters α produced a likelihood function f
- Superlevel set filtration of f created a persistence diagram $\mathrm{Dgm}(f)$
- Consider functional $\mathcal{L} \coloneqq L_{top}(-, g)$, where g is the ground truth function and L_{top} measures the cost of matching $\mathrm{Dgm}(-)$ to $\mathrm{Dgm}(g)$
- Want to optimize f to minimize $L_{top}(f, g)$, i.e. need gradient $\nabla_\alpha \mathcal{L}$
- First step: for a point (b, d) in a persistence diagram, compute:

$$\frac{\partial b}{\partial \alpha} \quad \text{and} \quad \frac{\partial d}{\partial \alpha}$$

- Strategy: (b, d) came from a pairing (σ, τ)---use this!

Recall: Simplices are Creators or Destroyers

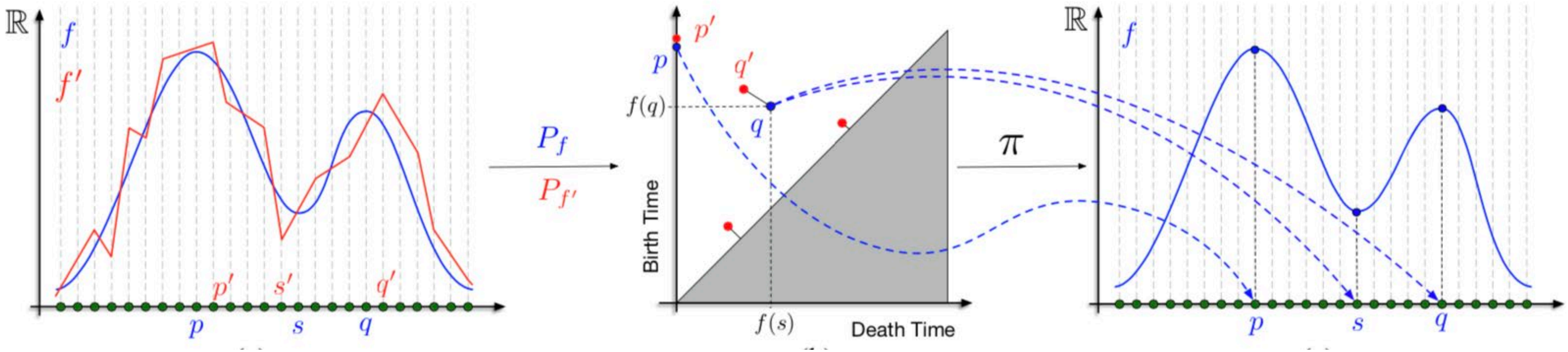
- A k-simplex either creates a k-dimensional feature or destroys a $(k-1)$-dimensional feature
- Persistence pairs a creator simplex σ with a destroyer simplex τ, and records the birth-death times (b, d) as the appearance times of (σ, τ).

Proof: Suppose we add a p-simplex σ. Assume simplices are added incrementally, so we have not yet added a $(p+1)$-simplex with σ as a face.

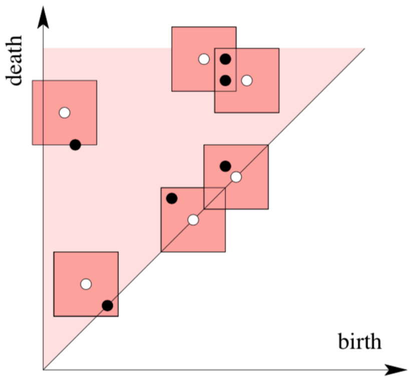
- Case 1: $\partial_p(\sigma) \in im(\partial_p)$, i.e. $\partial_p(\sigma) = \partial_p(\tau)$ for some τ. Then $\partial_p(\sigma - \tau) = 0$, so $\ker(\partial_p)$ grows by one dimension. Because we don't yet have a $(p+1)$-simplex with σ as a face, $im(\partial_{p+1})$ is unchanged. Therefore H_p grows by one dimension, i.e. we created a p-dimensional feature.
- Case 2: $\partial_p(\sigma) \notin im(\partial_p)$, i.e. $im(\partial_p)$ grows by one dimension. But $\partial_{p-1}\partial_p(\sigma) = 0$, and the elements in $\partial_p(\sigma)$ had already been added by the incremental addition assumption, so $\ker(\partial_{p-1})$ remains unchanged. Thus H_{p-1} loses one dimension, i.e. we destroyed a $(p-1)$-dimensional feature.

Gradients for Persistence Diagrams

Poulenard, Skraba, Ovsjanikov. *Topological Function Optimization for Continuous Shape Matching. SGP '18*



$\zeta: (b, d) \to (\sigma, \tau), \quad \eta: \sigma \to v_\sigma$ where $v_\sigma \in \arg\min_{v \in \sigma} f(v)$



$\pi := (\eta, \eta) \circ \zeta$ "reverse" map from persistence diagram to pairs of vertices

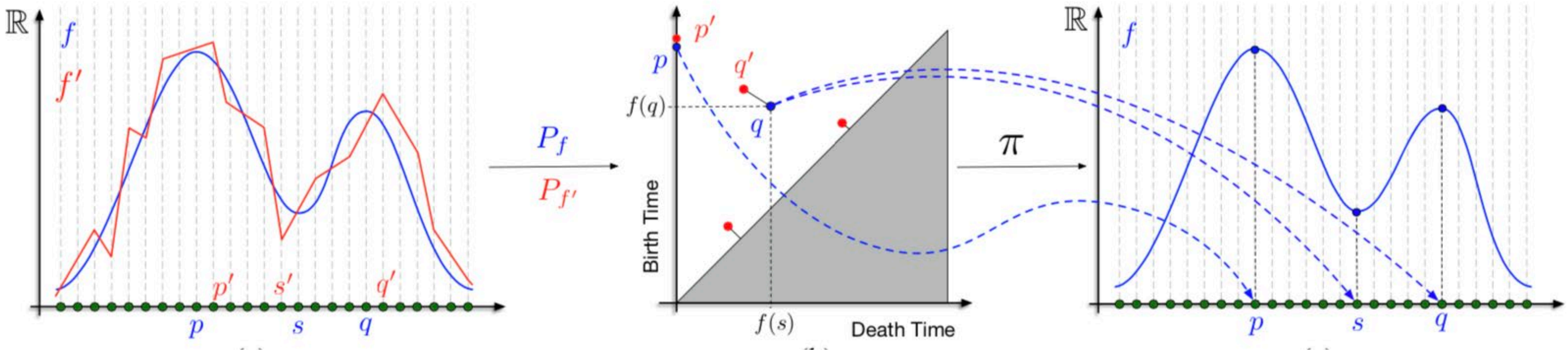
By stability, π remains unchanged for small perturbations of the diagram, and so:

$\frac{\partial b}{\partial \alpha} = \frac{\partial f(\pi(b))}{\partial \alpha} = \frac{\partial f(v_\sigma)}{\partial \alpha}$, and similarly $\frac{\partial d}{\partial \alpha} = \frac{\partial f}{\partial \alpha}(v_\tau)$

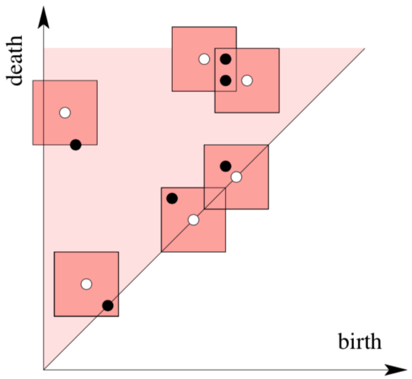
Conclusion: "the derivative (of the diagram) is equivalent to the derivative of the function evaluated at the image of the map π"

Gradients for Persistence Diagrams

Poulenard, Skraba, Ovsjanikov. *Topological Function Optimization for Continuous Shape Matching. SGP '18*



By chain rule, we have $\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{(b_i, d_i) \in Dgm(f)} \frac{\partial \mathcal{L}}{\partial b_i}\frac{\partial b_i}{\partial \alpha} + \frac{\partial \mathcal{L}}{\partial d_i}\frac{\partial d_i}{\partial \alpha}$



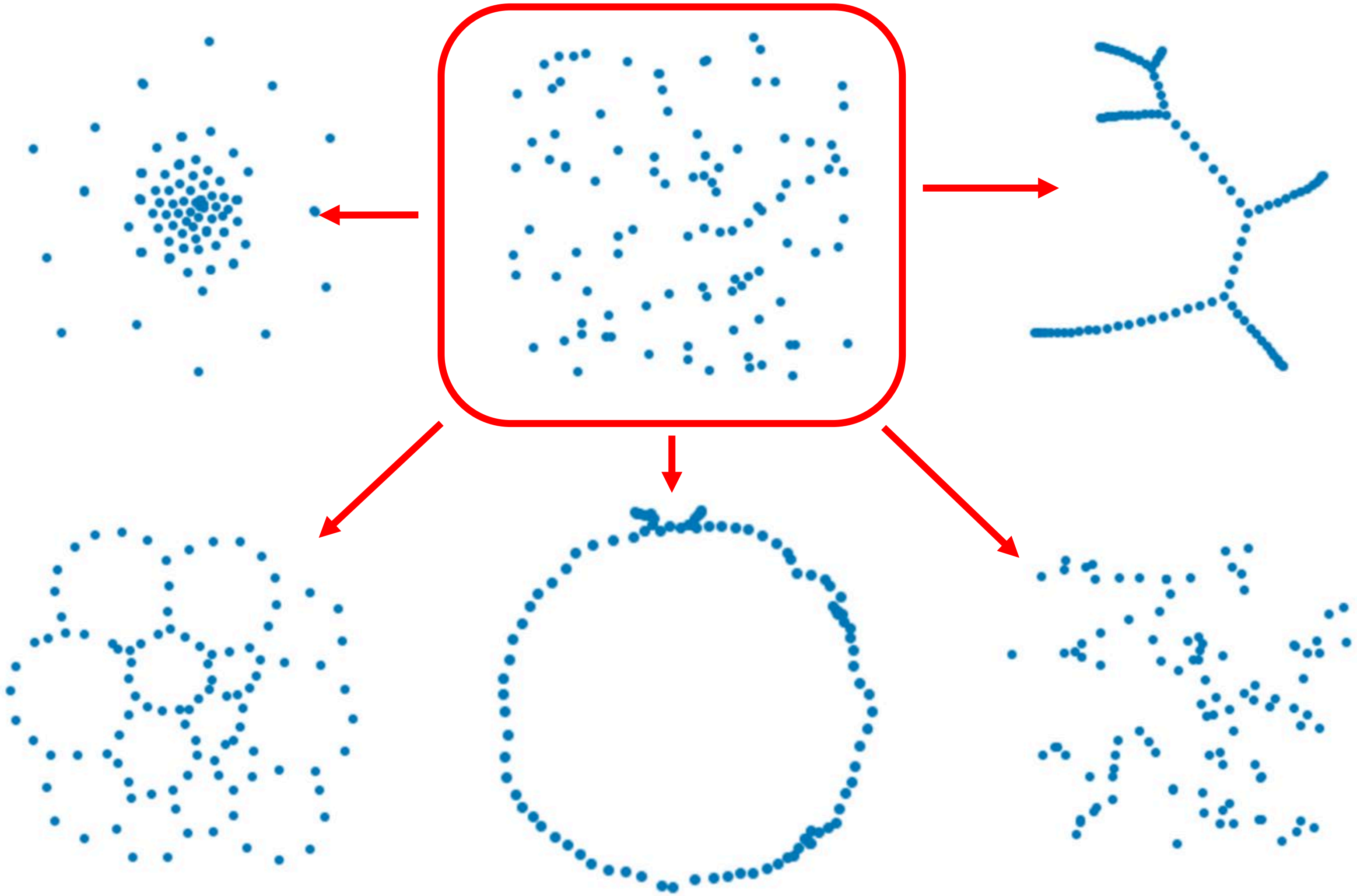
Let $\gamma^*(b_i), \gamma^*(d_i)$ denote the birth-death pair of the ground truth g that (b_i, d_i) is matched to. From before, we had:

$\frac{\partial \mathcal{L}}{\partial b_i} = \frac{\partial}{\partial b_i}\sum_i [b_i - \gamma^*(b_i)]^2 = 2[b_i - \gamma^*(b_i)]$, and likewise for $\frac{\partial \mathcal{L}}{\partial d_i}$. Summarizing,

$$\frac{\partial \mathcal{L}}{\partial \alpha} = 2\sum_i \left(b_i - \gamma^*(b_i)\right)\frac{\partial f}{\partial \alpha}(v_\sigma) + 2\sum_i \left(d_i - \gamma^*(d_i)\right)\frac{\partial f}{\partial \alpha}(v_\tau)$$

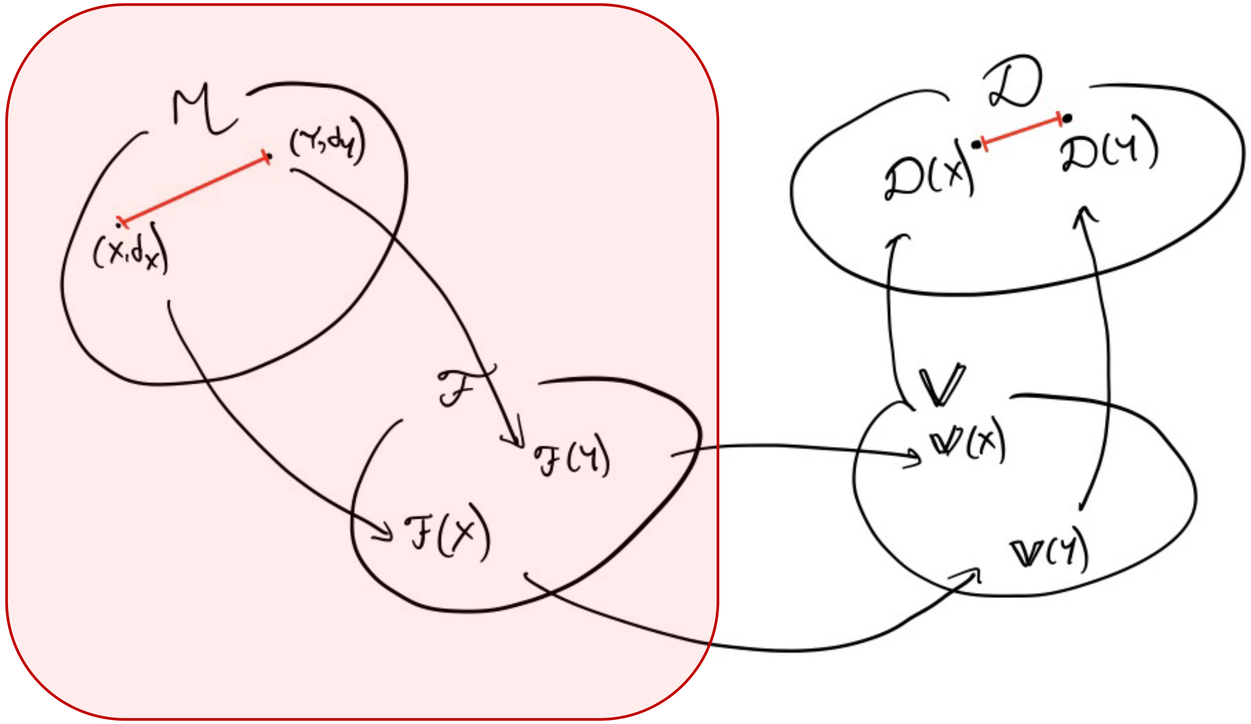
Topological Noise Reduction

Bruel-Gabrielsson, Nelson, Dwaraknath, Skraba, Guibas, Carlsson, *A topology layer for machine learning*. AISTATS 2020



Adding Geometry to Topology

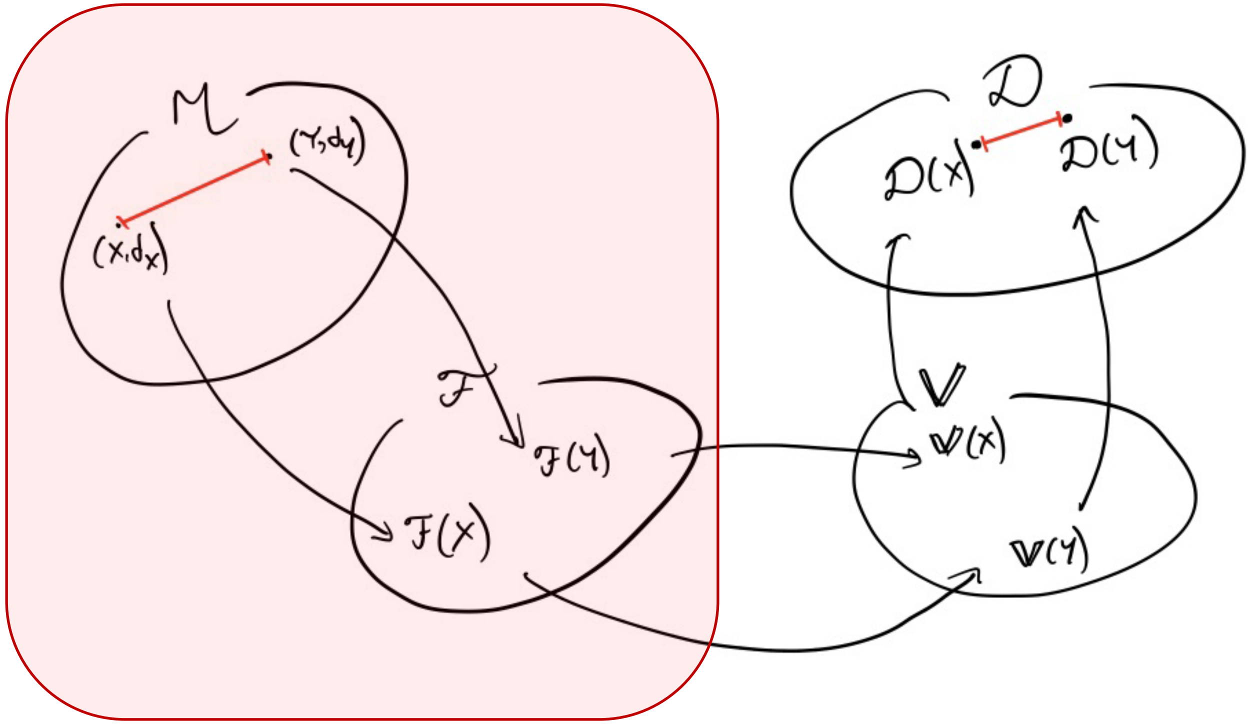
Synthetic Filtrations are Possible, and Informative



Why only build complexes capturing proximity?

- Possible to create “derived complexes” that capture different information
- Natural to consider geometry-dependent filtrations

Synthetic Filtrations are Possible, and Informative



Why only build complexes capturing proximity?

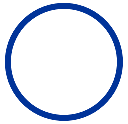
- Possible to create "derived complexes" that capture different information
- Natural to consider geometry-dependent filtrations

Sharp features

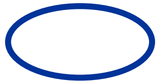
U $\beta_1 = 0$

V $\beta_1 = 0$

Soft features



$\beta_1 = 1$

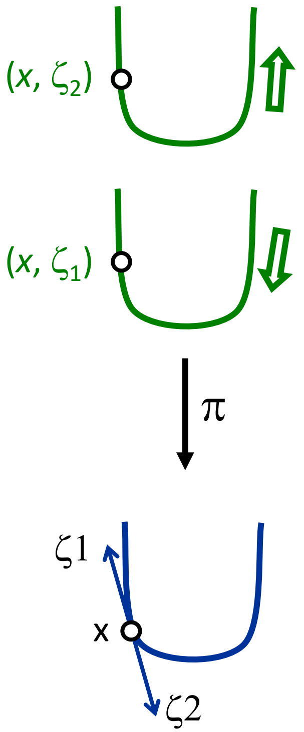


$\beta_1 = 1$

Geometry discriminating

Topology classifying

2-D Curve Tangent Complex



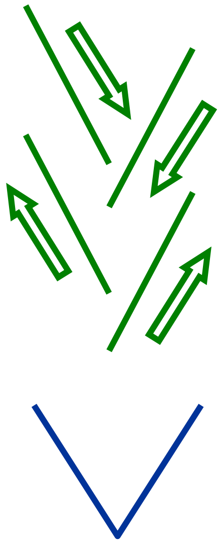
Covering space

T(X) has two components:
$\beta_0(T(X)) = 2$

There are two points in its fiber $\pi^{-1}(x)$

Every point x on a smooth curve X has two tangent directions.

A corner point has four tangent directions:
$\beta_0(T(X)) = 4$

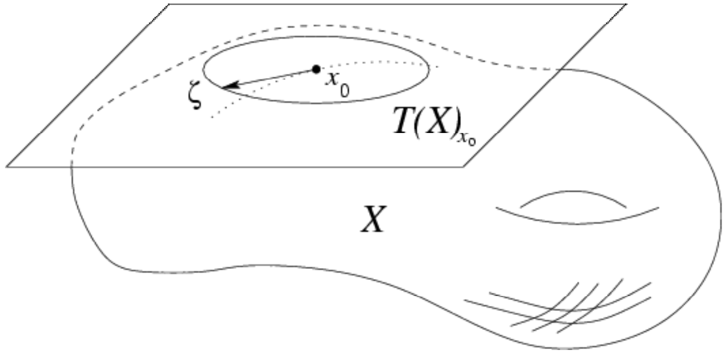


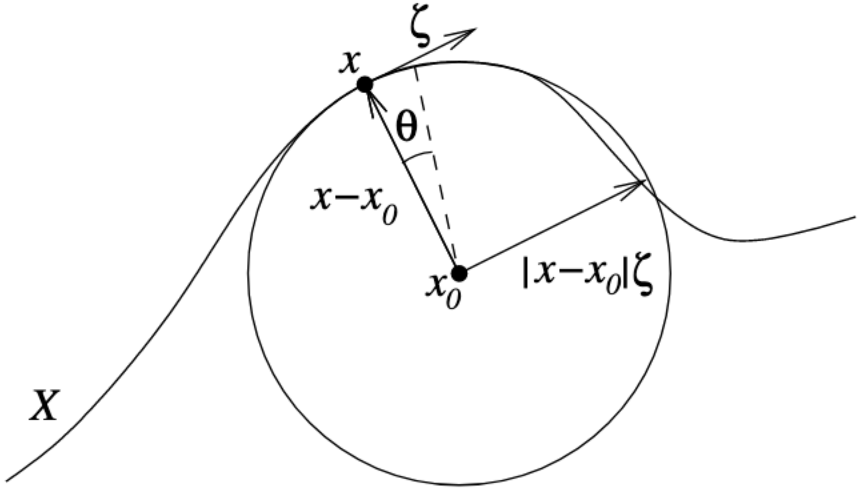
The (Filtered) Tangent Complex

Collins, Zomorodian, Carlsson, Guibas, *A barcode shape descriptor for curve point cloud data*, CaG'04

Setup: $X \subseteq \mathbb{R}^n$

- Tangent complex $T(X) \subseteq X \times \mathbb{S}^{n-1}$:
  - unit tangent vectors at all points of X, i.e. pairs (x, ζ) where $x \in X$ and ζ is a tangent vector
- Filtered tangent complex: filter the points in the tangent complex by increasing curvature
  - Curvature at point (x, ζ) is $1/\rho$, where ρ is radius of osculating circle
  - $T_\delta(X) =$ points with curvature $< \delta$
  - Filtered tangent complex $T^{filt}(X) \coloneqq \{T_\delta(X)\}_{\delta>0}$





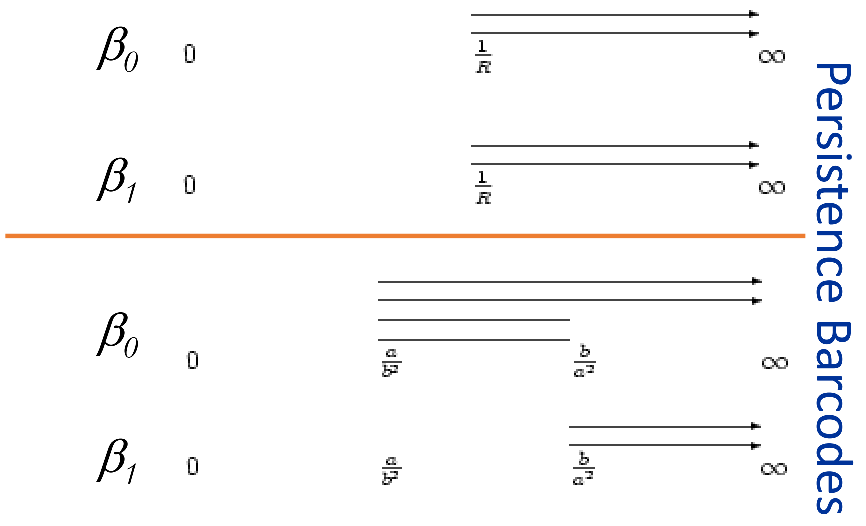
Examples: Circle vs Ellipse

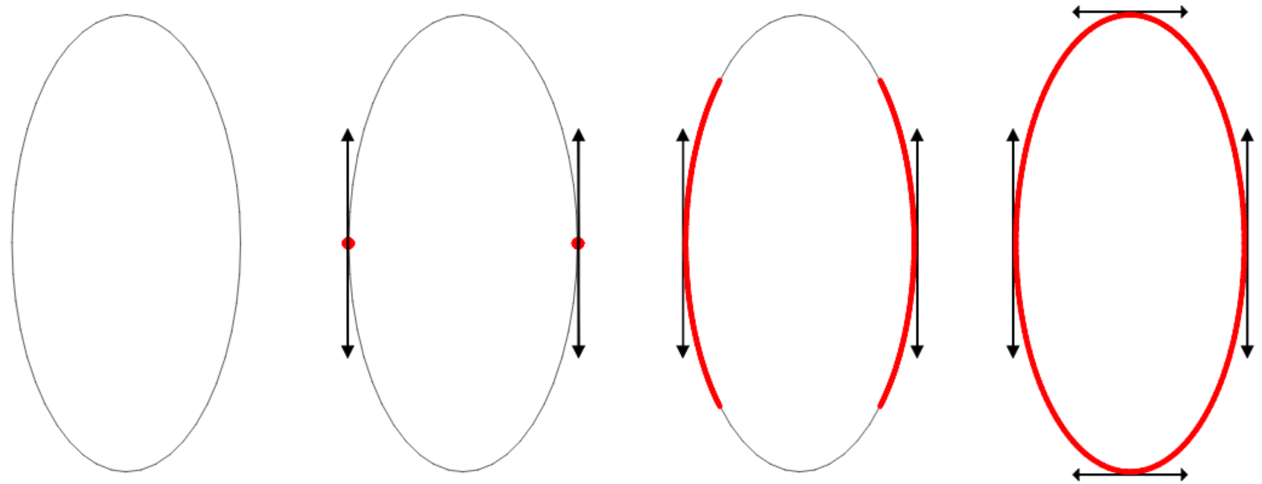
Circle of radius R: Tangent complex is $S^1 \times S^0$**,** as there are two tangent directions at each point.

- T^{filt} is empty until $1/R$, then full complex enters at once

Ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$:

- evolves through four stages: points at *lower* curvature appear earlier.





$0 \leq \delta < \frac{a}{b^2}$ $\quad$ $\frac{a}{b^2} = \delta$ $\quad$ $\frac{a}{b^2} < \delta < \frac{b}{a^2}$ $\quad$ $\frac{b}{a^2} \leq \delta$

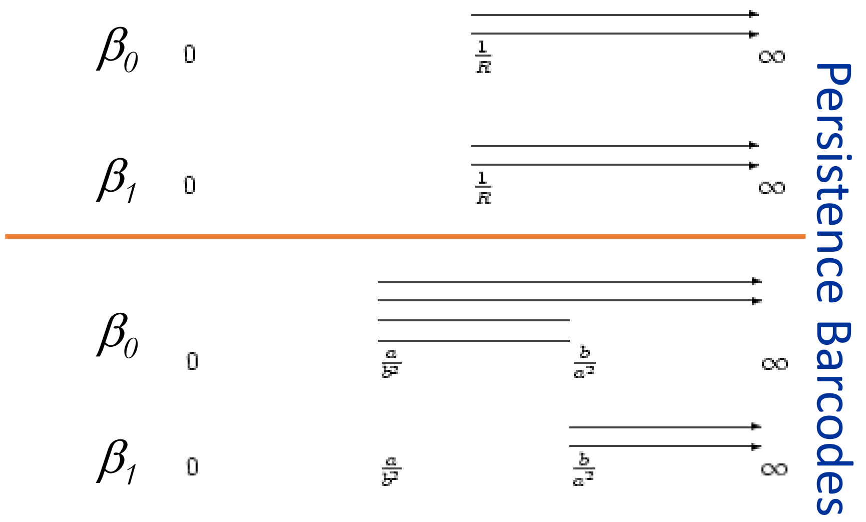
Examples: Circle vs Ellipse

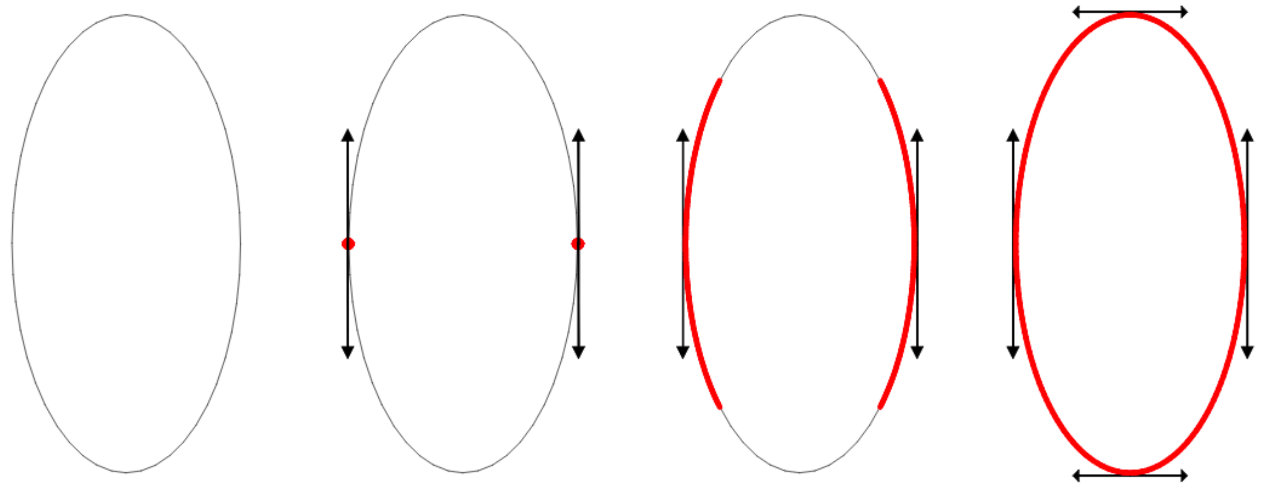
Circle of radius R: Tangent complex is $S^1 \times S^0$**,** as there are two tangent directions at each point.

- T^{filt} is empty until $1/R$, then full complex enters at once

Ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$:

- evolves through four stages: points at *lower* curvature appear earlier.

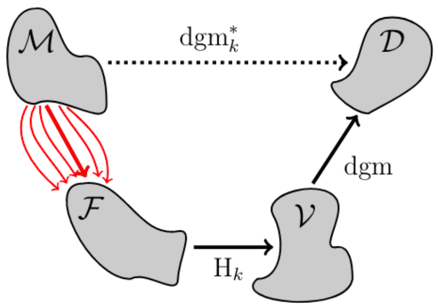




$0 \leq \delta < \frac{a}{b^2}$ $\quad \frac{a}{b^2} = \delta$ $\quad \frac{a}{b^2} < \delta < \frac{b}{a^2}$ $\quad \frac{b}{a^2} \leq \delta$

Proposed framework:

- Information is distributed across a variety of properties
- Use multiple barcodes to interrogate the data
- We will see more of this perspective later…



The Mapper Algorithm

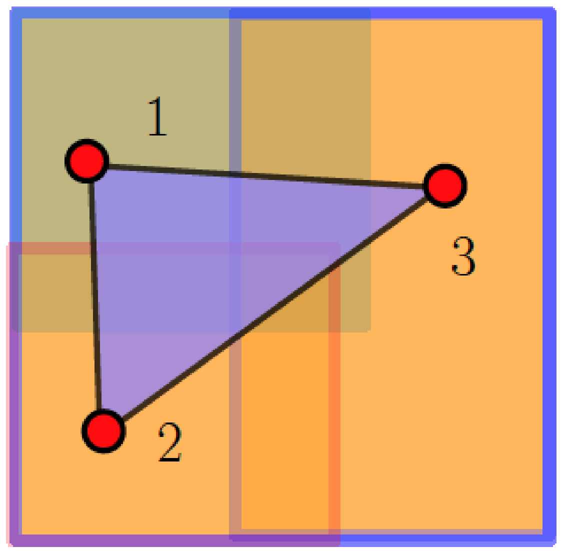
Recall: Covers and Nerves

Finite cover of a topological space X

- $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ for a finite index set A.
- each $U_\alpha \subseteq X$ is open and $X = \bigcup_{\alpha \in A} U_\alpha$

Nerve of a cover

- Simplicial complex: $N(\mathcal{U})$ with vertex set A.
- simplices: $A \supseteq \sigma \in N(\mathcal{U}) \Leftrightarrow \bigcap_{\alpha \in \sigma} U_\alpha \neq \emptyset$.

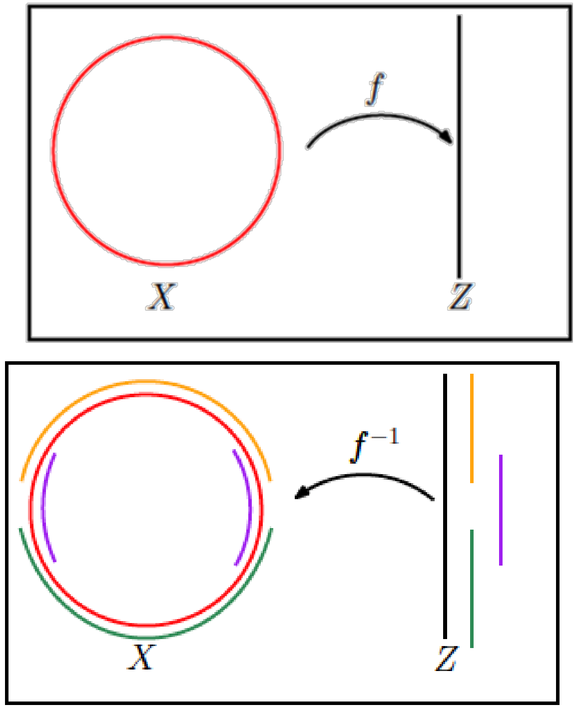


Pullback Covers and their Nerves

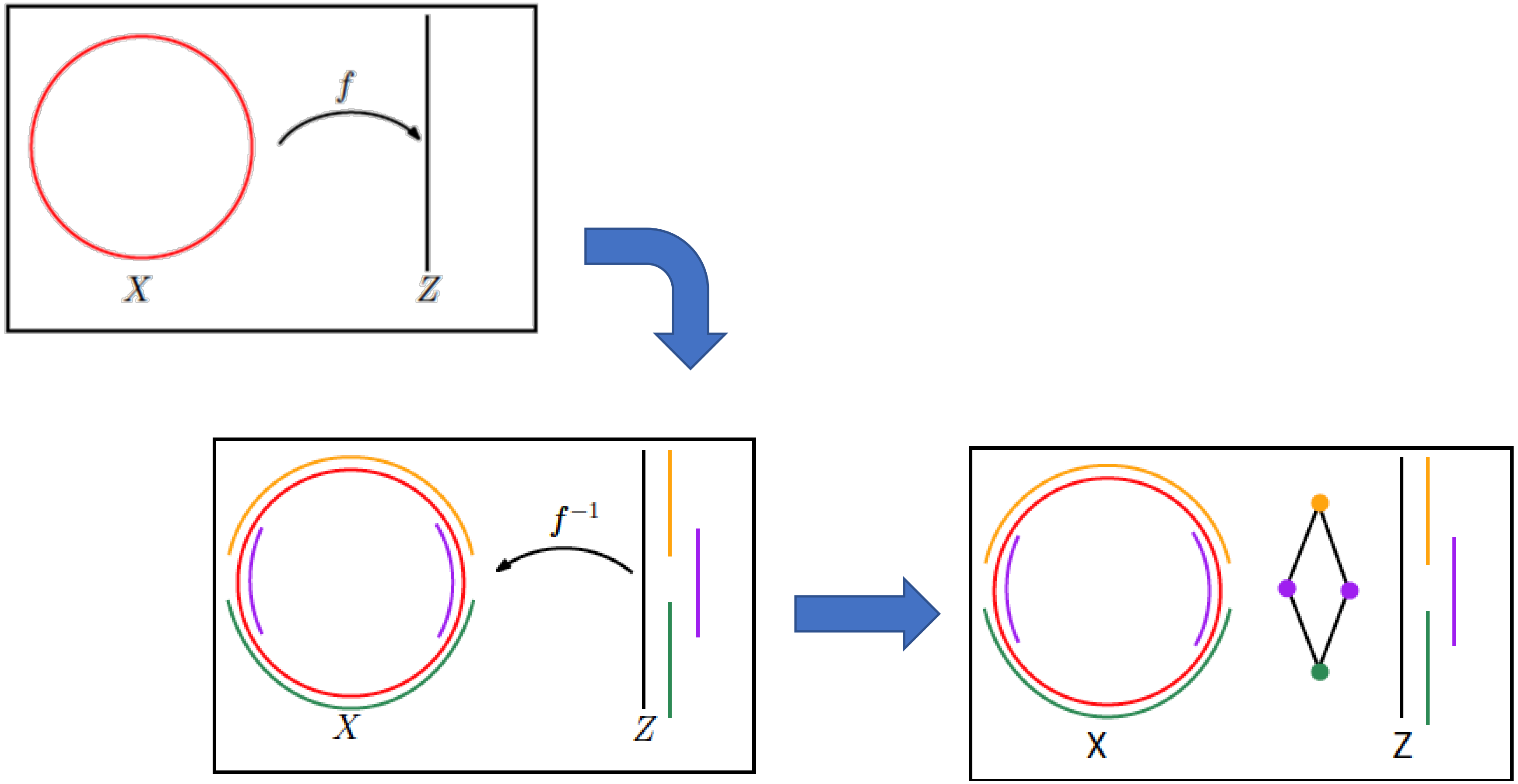
- Assume you have $f : X \to Z$ *well behaved* continuous function and $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ finite open cover of Z.
- For each $\alpha \in A$ consider the connected components of $f^{-1}(U_\alpha) = \{V_{i,\alpha},\ 1 \leq i \leq j_\alpha\}$.
- Let $f^*(\mathcal{U})$ be the (finite) open cover of X thus induced:

$$f^*(\mathcal{U}) := \{V_{i,\alpha},\ 1 \leq i \leq j_\alpha,\ \alpha \in A\}.$$

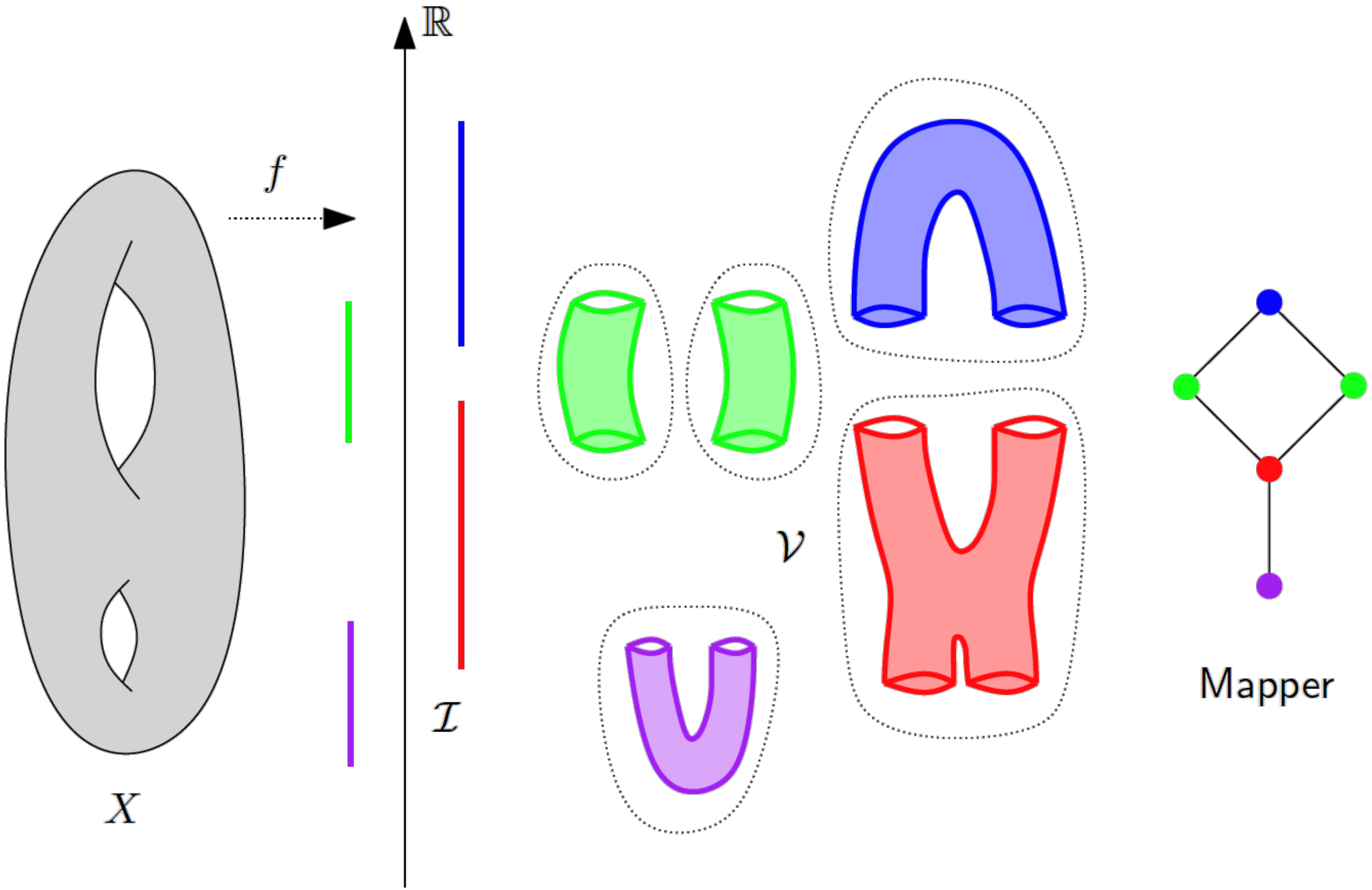
  This is the pullback of $\mathcal{U}$ via f.
- Now consider the nerve of the pullback: $N(f^*(\mathcal{U}))$. This complex often retains structural information about underlying space X.



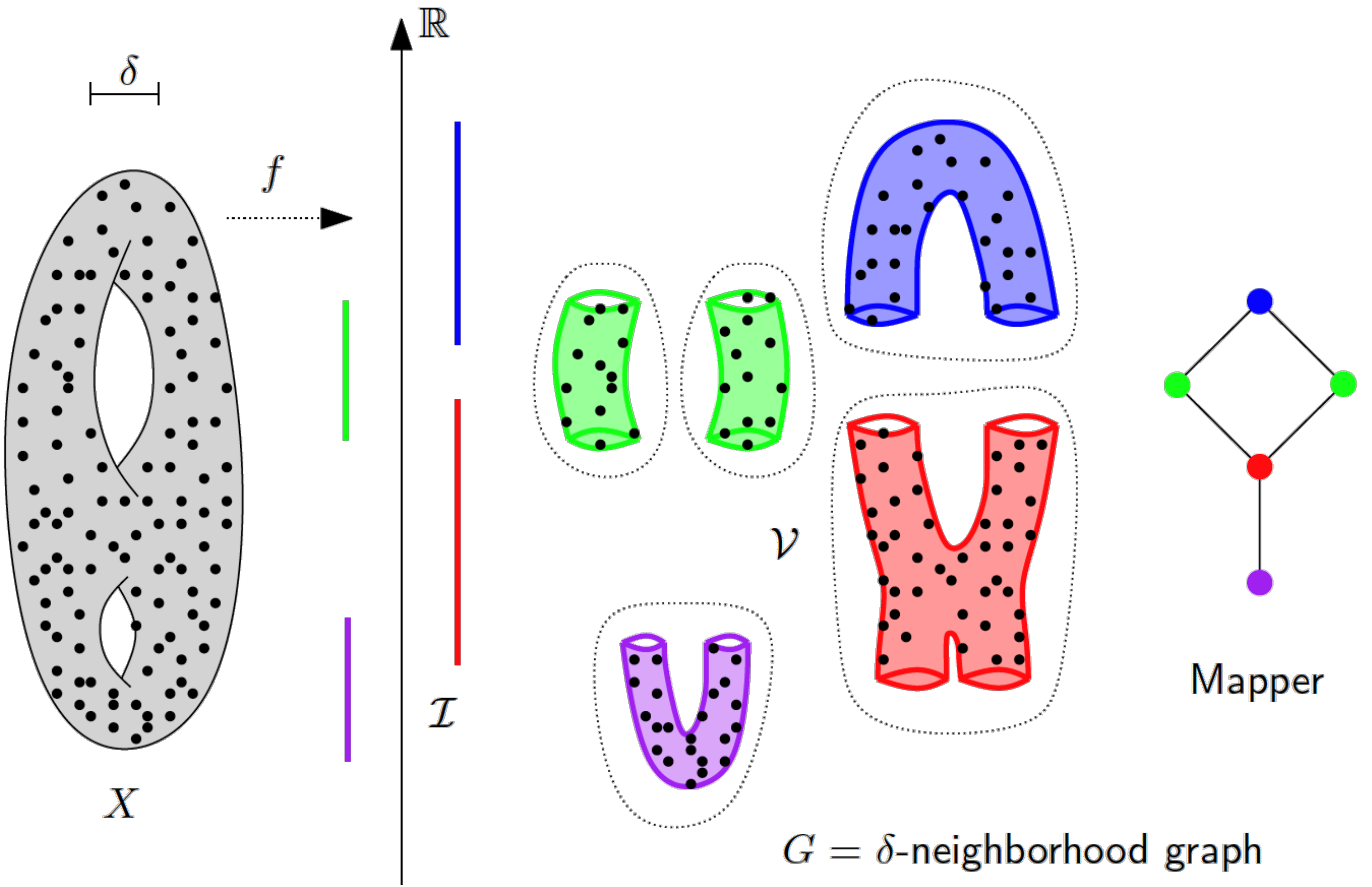
Pullback Covers and their Nerves



Choice of Lens



Choice of Lens

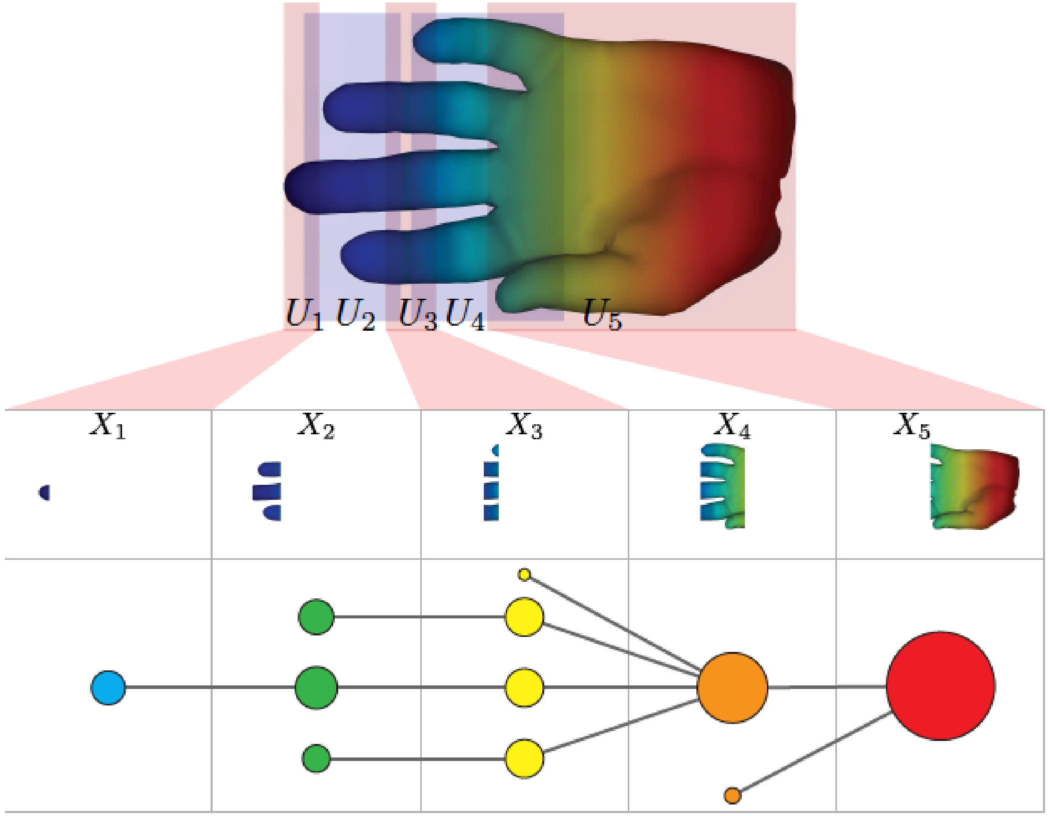


The Mapper Algorithm

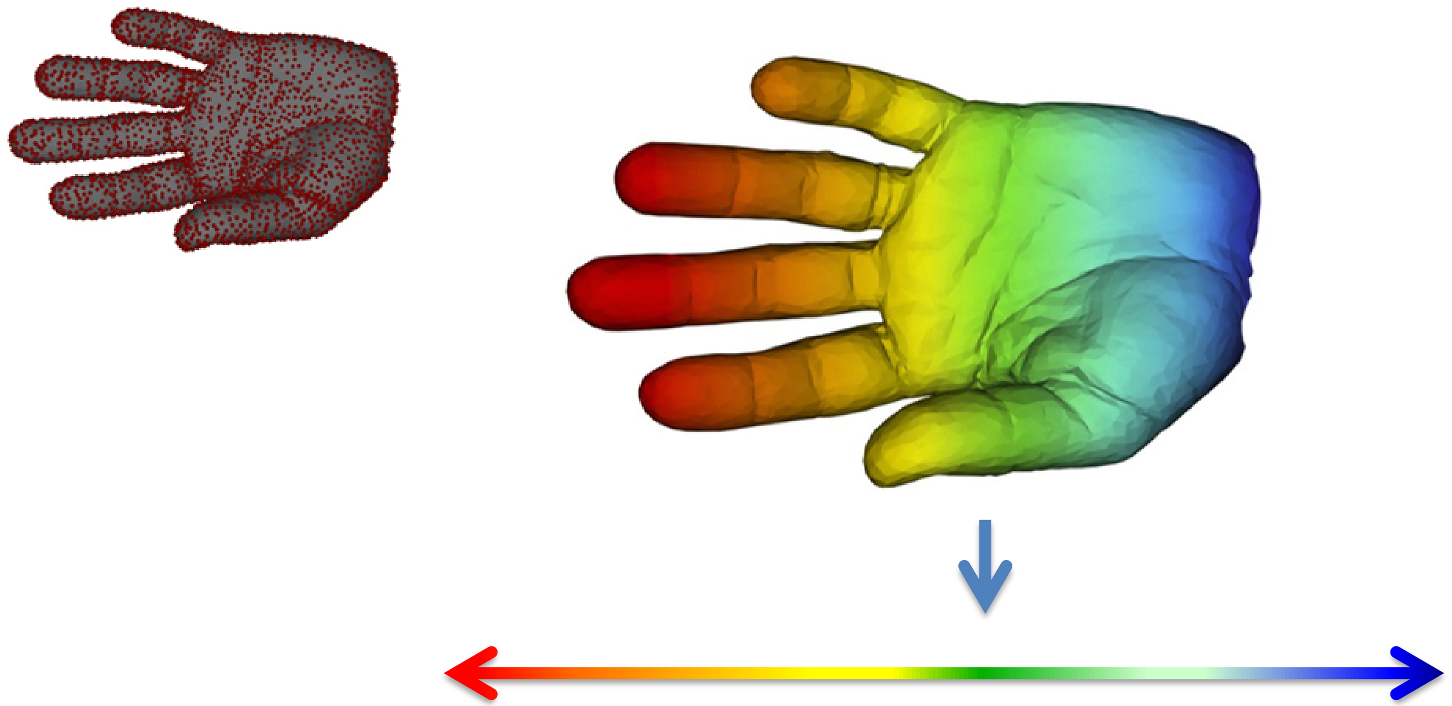
Singh, Mémoli, Carlsson - *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*

Let $f : X \to Z$ be well behaved and continuous and $\mathcal{U}$ be finite open cover of Z, then the Mapper output corresponding to $\mathcal{U}$ and f is

$$\mathrm{M}(\mathcal{U}, f) := N(f^*(\mathcal{U})).$$



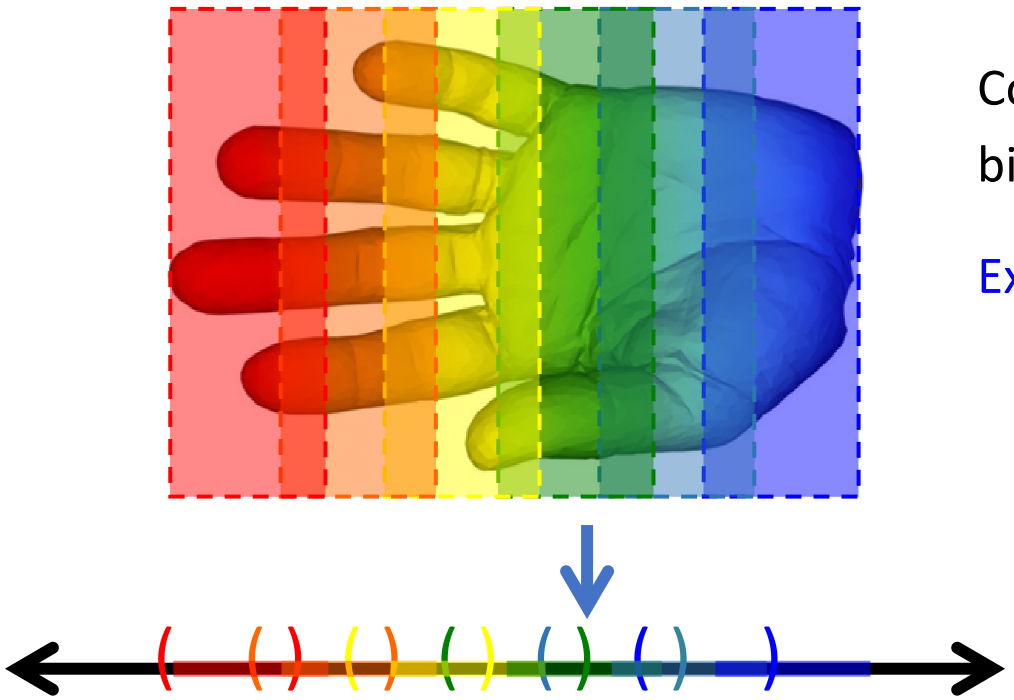
Step 1: Choose a Lens / Filter Function



Function f : Data Set → **R**

Ex 1: x-coordinate f : (x, y, z) → x

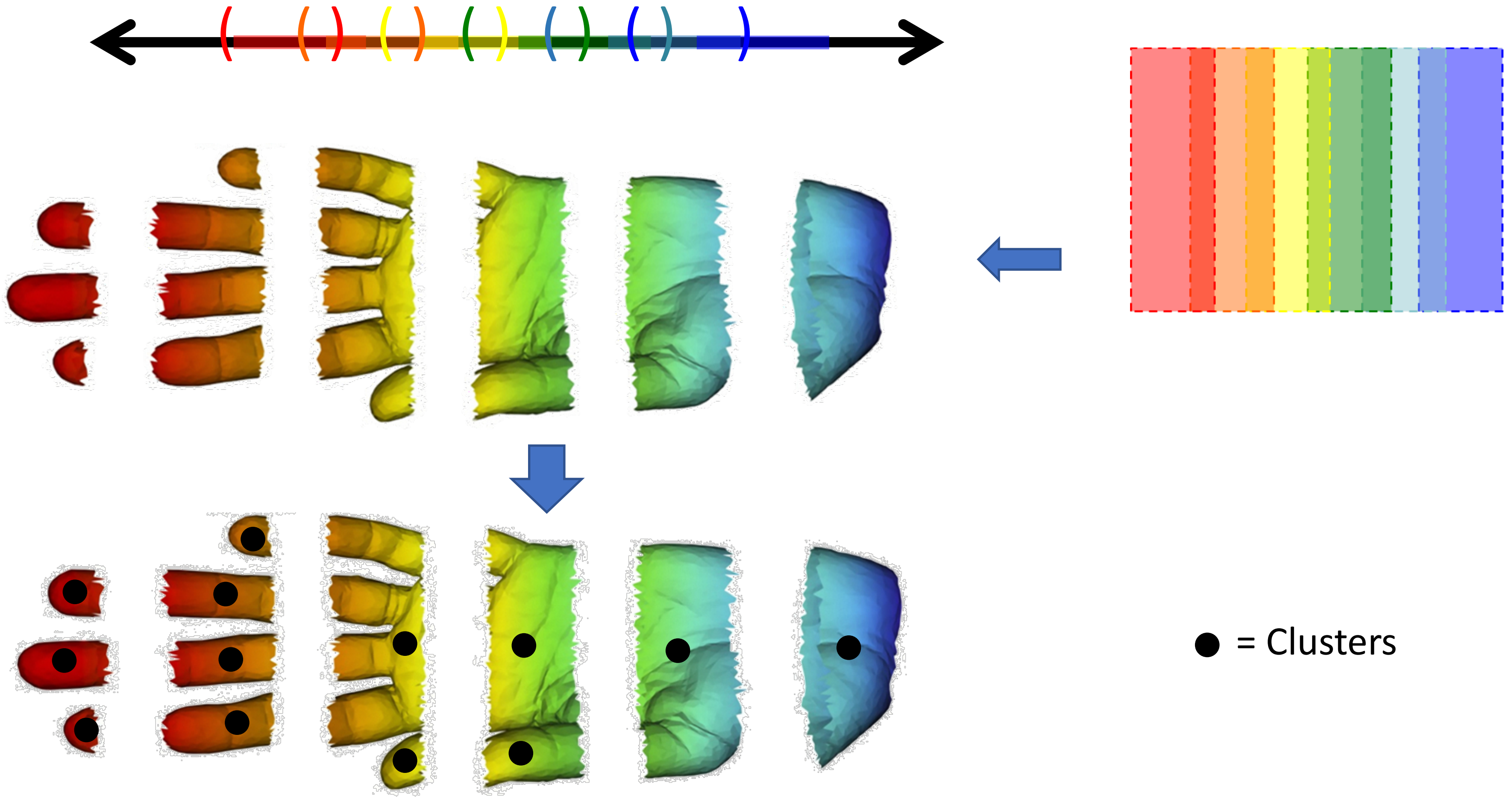
Step 2: Partition into Overlapping Bins



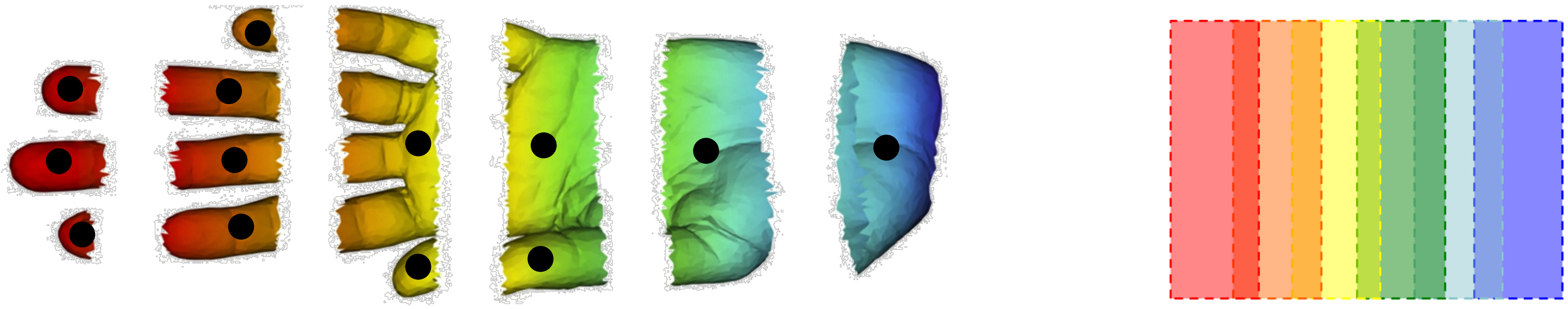
Cover data via overlapping bins.

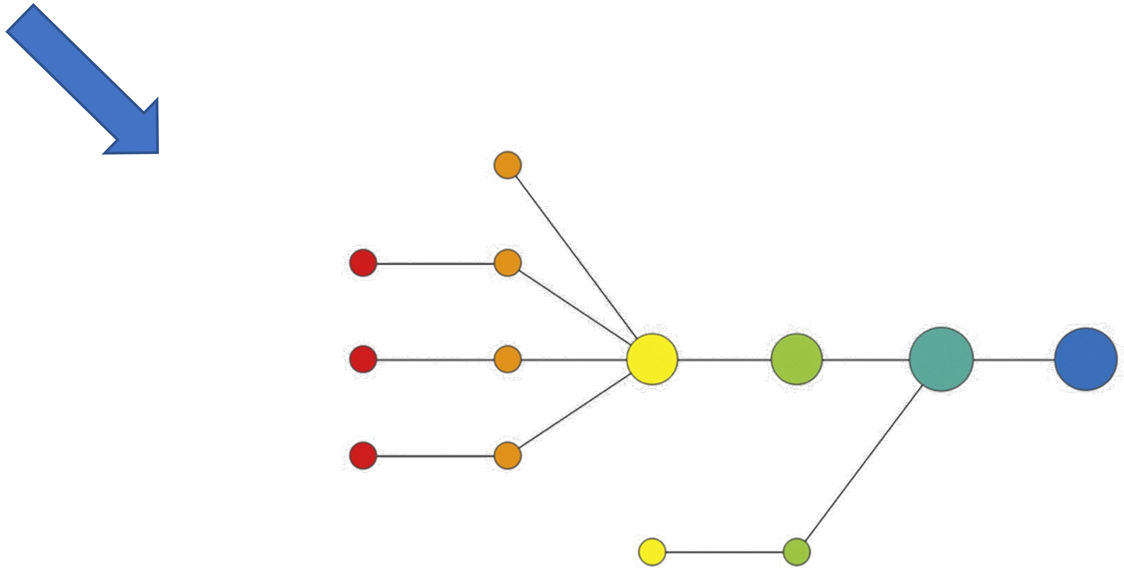
Example: $f^{-1}(a_i, b_i)$

Step 3: Form Connected Components in the Bins



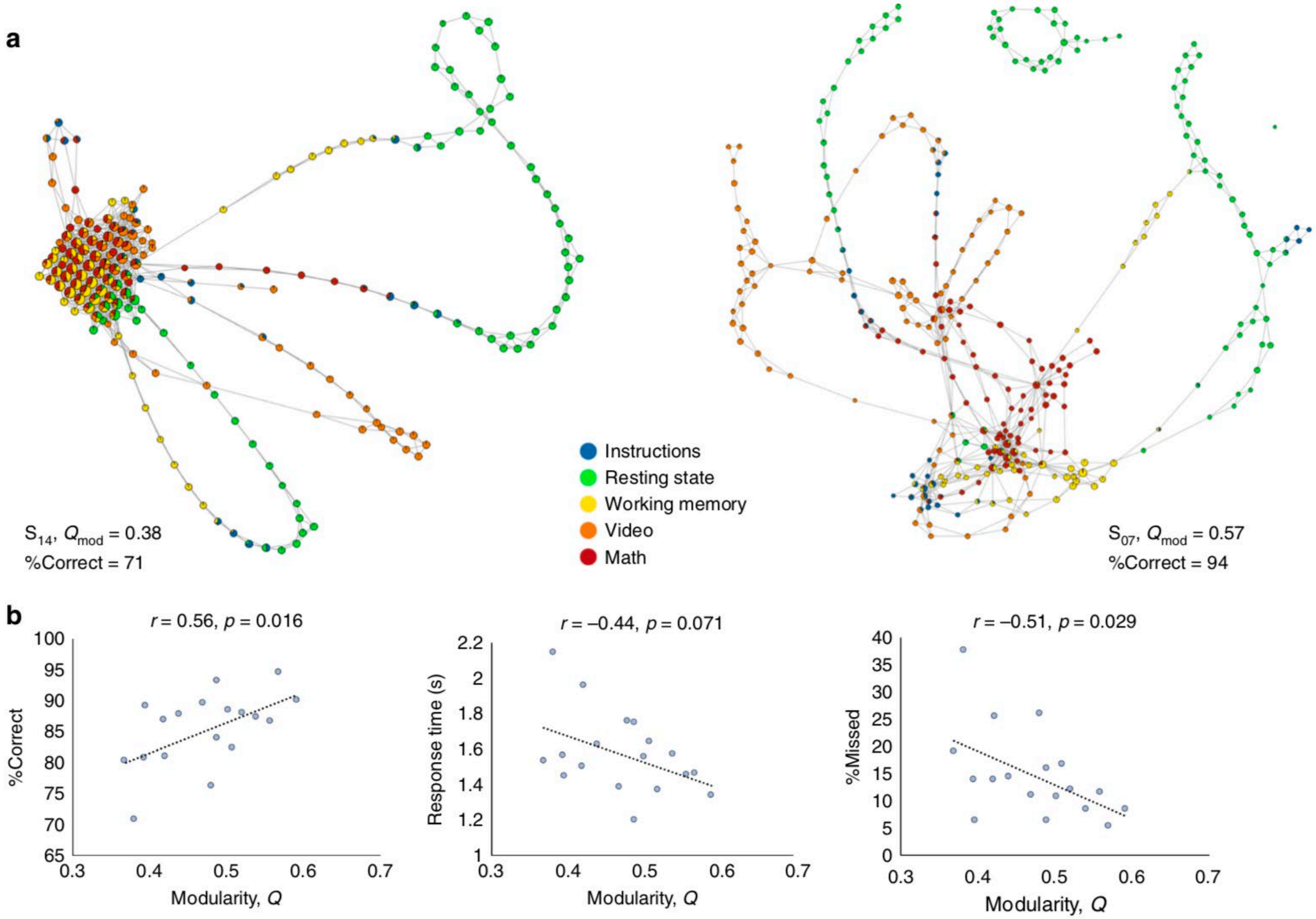
Step 4: Form a Network of Intersecting Clusters





Mapper in Computational Psychiatry

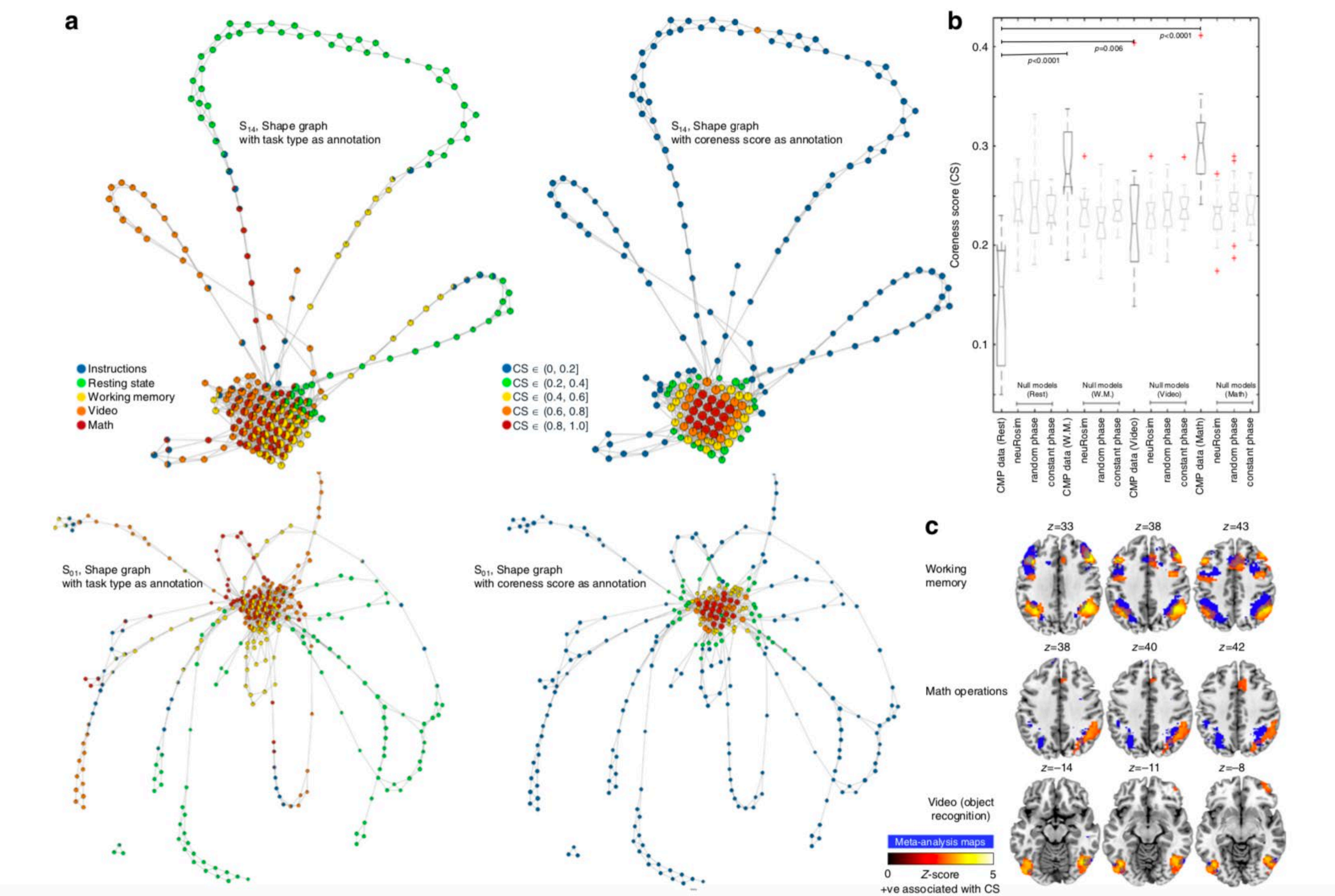
Saggar et al., *Towards a new approach to reveal dynamical organization of the brain*



- Community structure of the mapper graph shown to be positively correlated with performance on a task

Mapper in Computational Psychiatry

Saggar et al., *Towards a new approach to reveal dynamical organization of the brain*



- Community structure of the mapper graph shown to be positively correlated with performance on a task

- Nodes related to high cognitive load buried deep in the core of the graph

- Nodes related to rest stay in the periphery

Topological Time Series Analysis for Video Data

Sliding Window Embeddings

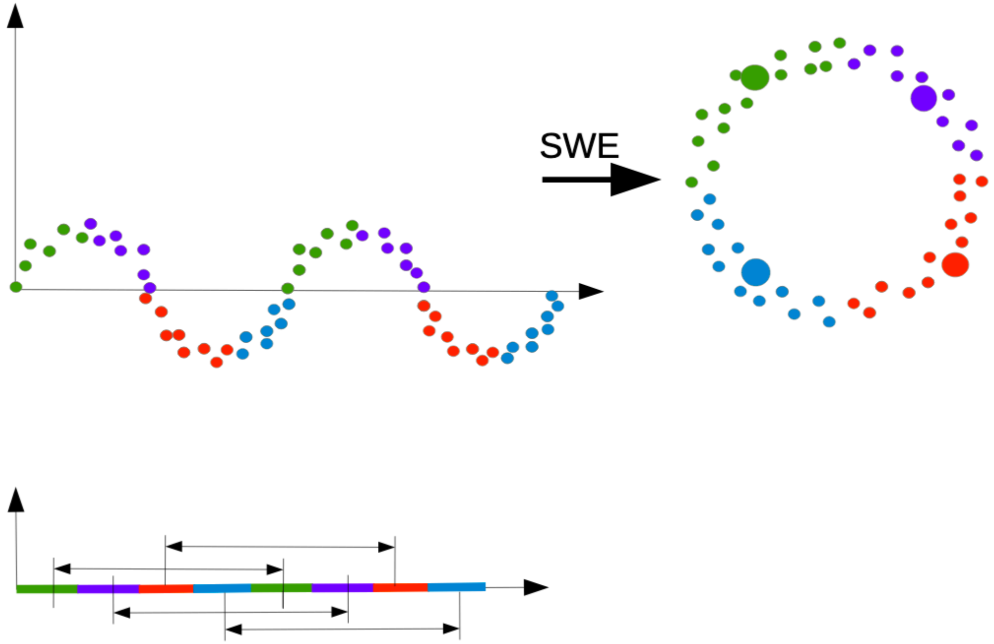
Chris Tralie and Jose Perea, *(Quasi)Periodicity Quantification in Video Data, Using Topology*

- Consider a video dataset as time-series data. I.e. fix integers W (width) and H (height), and consider the function (for grayscale videos):

$$X: \mathbb{N} \to \mathbb{R}^{W\times H}\text{, or, via interpolation, } X: \mathbb{R}_+ \to \mathbb{R}^{W\times H}$$

- Fix $\tau > 0$ (delay) and $d \in \mathbb{Z}_{\geq 0}$ (dimension). The sliding window embedding of X at time t and with parameters d and τ is:

$$SW_{d,\tau}X(t) = \begin{bmatrix} X(t) \\ X(t+\tau) \\ X(t+2\tau) \\ . \\ . \\ . \\ X(t+d\tau) \end{bmatrix} \in \mathbb{R}^{W\times H \times (d+1)}$$



- Sliding window embedding converts time series data into point cloud data

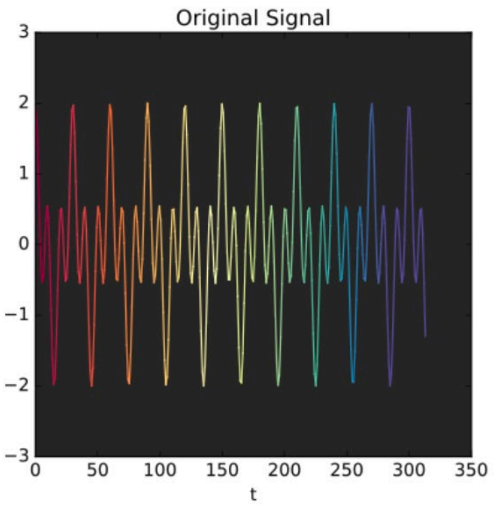
Figure from Dlotko, Qiu, Rudin, *Cyclicity, Periodicity, and the Topology of Time Seres*

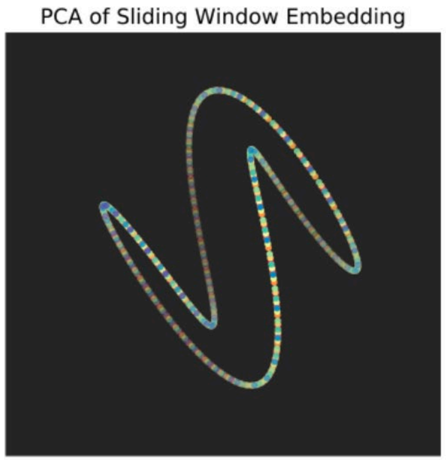
Geometry Missed by the Power Spectrum

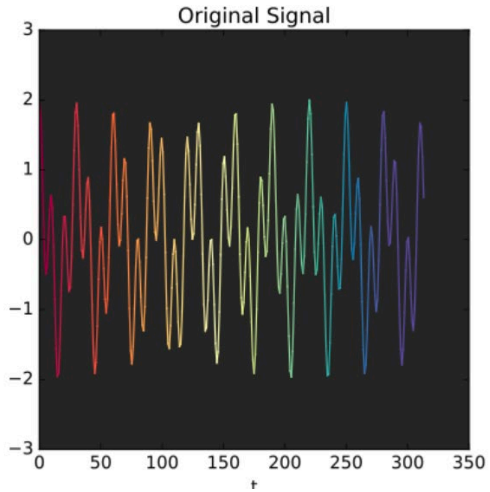
Chris Tralie and Jose Perea, *(Quasi)Periodicity Quantification in Video Data, Using Topology*

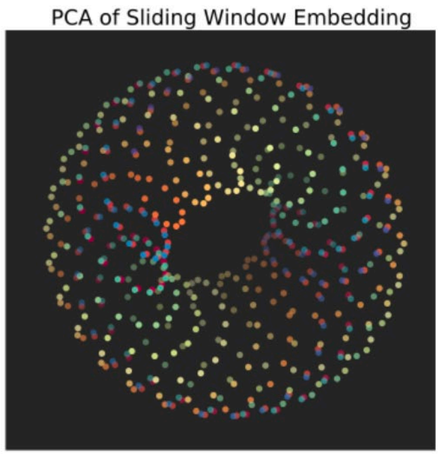
- SWEs reveal topology and geometry that are ignored by conventional methods. Consider two functions

$$f_h(t) = \cos\left(\frac{\pi}{5}t\right) + \cos\left(\frac{\pi}{15}t\right), \qquad f_q(t) = \cos\left(\frac{\pi}{5}t\right) + \cos\left(\frac{1}{5}t\right)$$

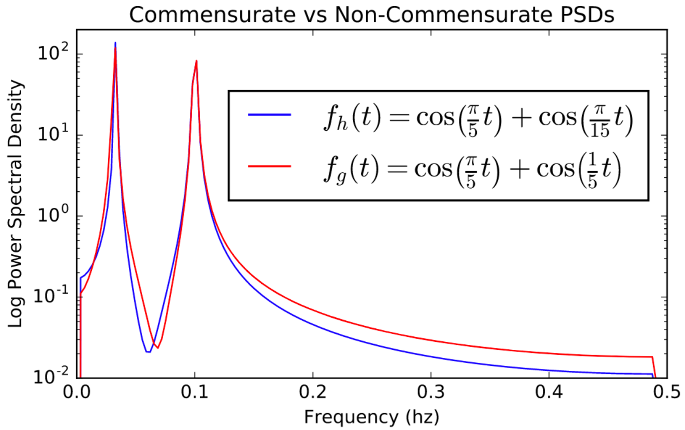








- Clear differences in topology and geometry, but the power spectrum may not see this (depending on who you ask)

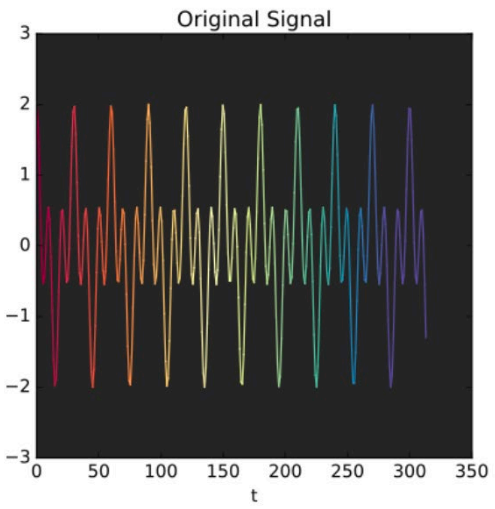


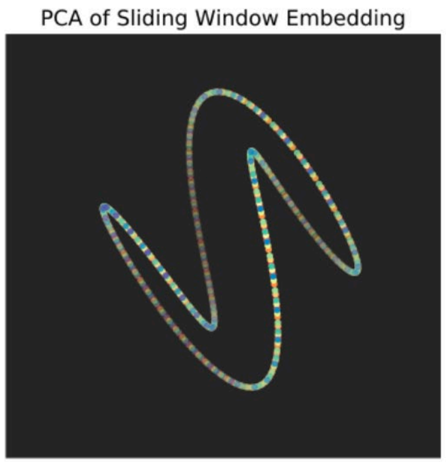
Periodic and Quasiperiodic Functions Have Distinct Topology

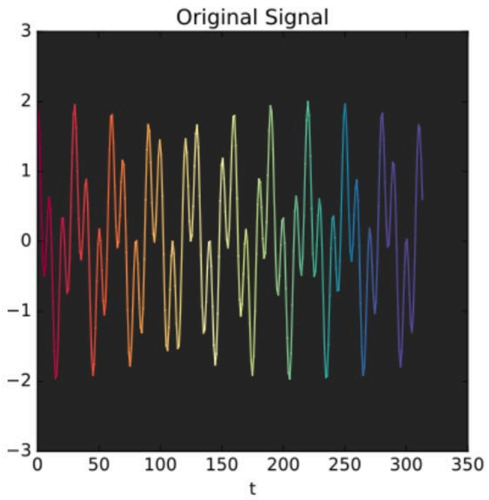
Chris Tralie and Jose Perea, *(Quasi)Periodicity Quantification in Video Data, Using Topology*

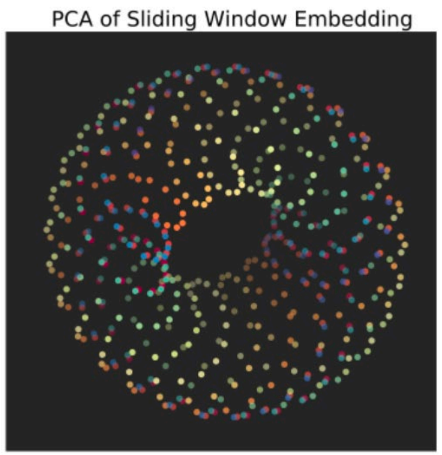
$$f_h(t) = \cos\left(\frac{\pi}{5}t\right) + \cos\left(\frac{\pi}{15}t\right), \qquad f_q(t) = \cos\left(\frac{\pi}{5}t\right) + \cos\left(\frac{1}{5}t\right)$$

- To say f_h is periodic means the constituent frequencies $\frac{1}{10}$ and $\frac{1}{30}$ are linearly dependent (commensurate) over $\mathbb{Q}$, and to say f_q is quasiperiodic means the constituent frequencies $\frac{1}{10}, \frac{1}{10\pi}$ are linearly independent (incommensurate) over $\mathbb{Q}$





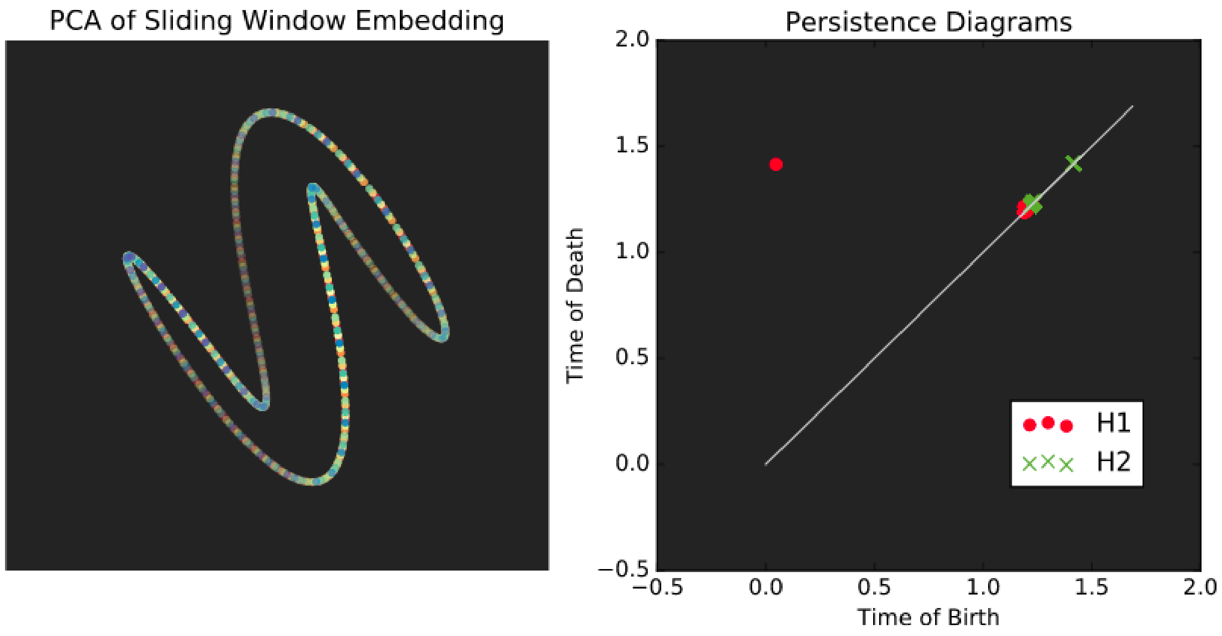


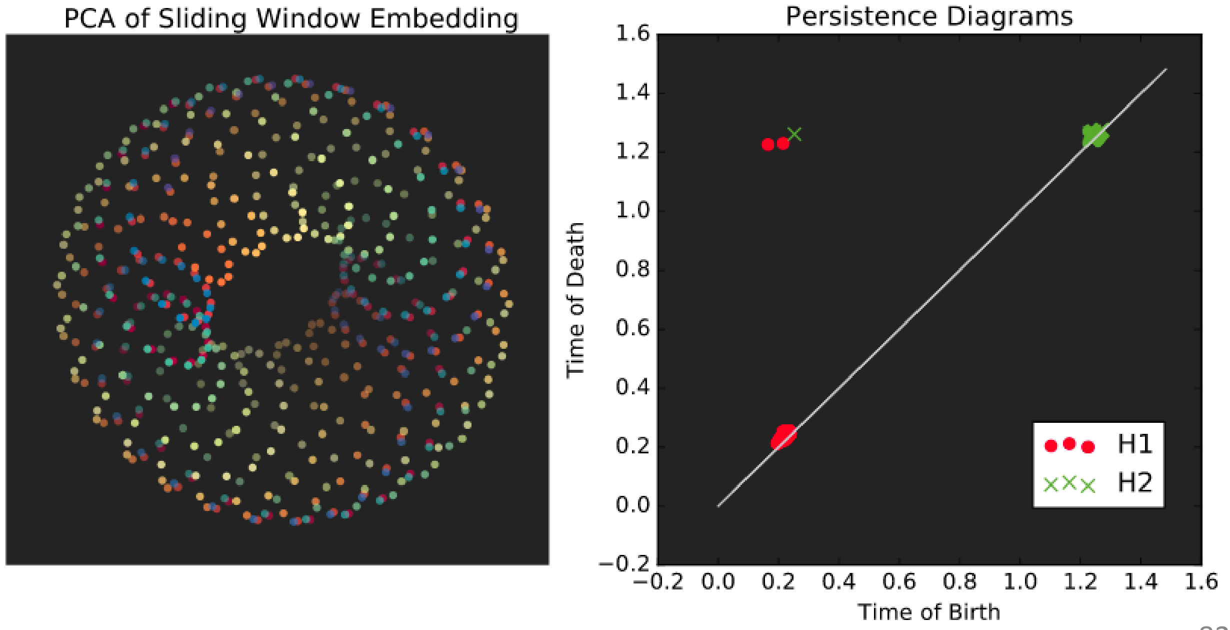


- **Theorem (Perea, Harer '15)**: Given a periodic function $f: [0,2\pi] \to \mathbb{R}$ with N harmonics, the sliding window embedding $SW_{d,\tau}f$ (for d, τ in prescribed ranges) is a topological circle that wraps around an N-dimensional torus $\mathbb{T}^N$. On the other hand, given a quasiperiodic function with N incommensurate harmonics, the sliding window embedding is dense in $\mathbb{T}^N$, i.e. it fills out the N-torus.

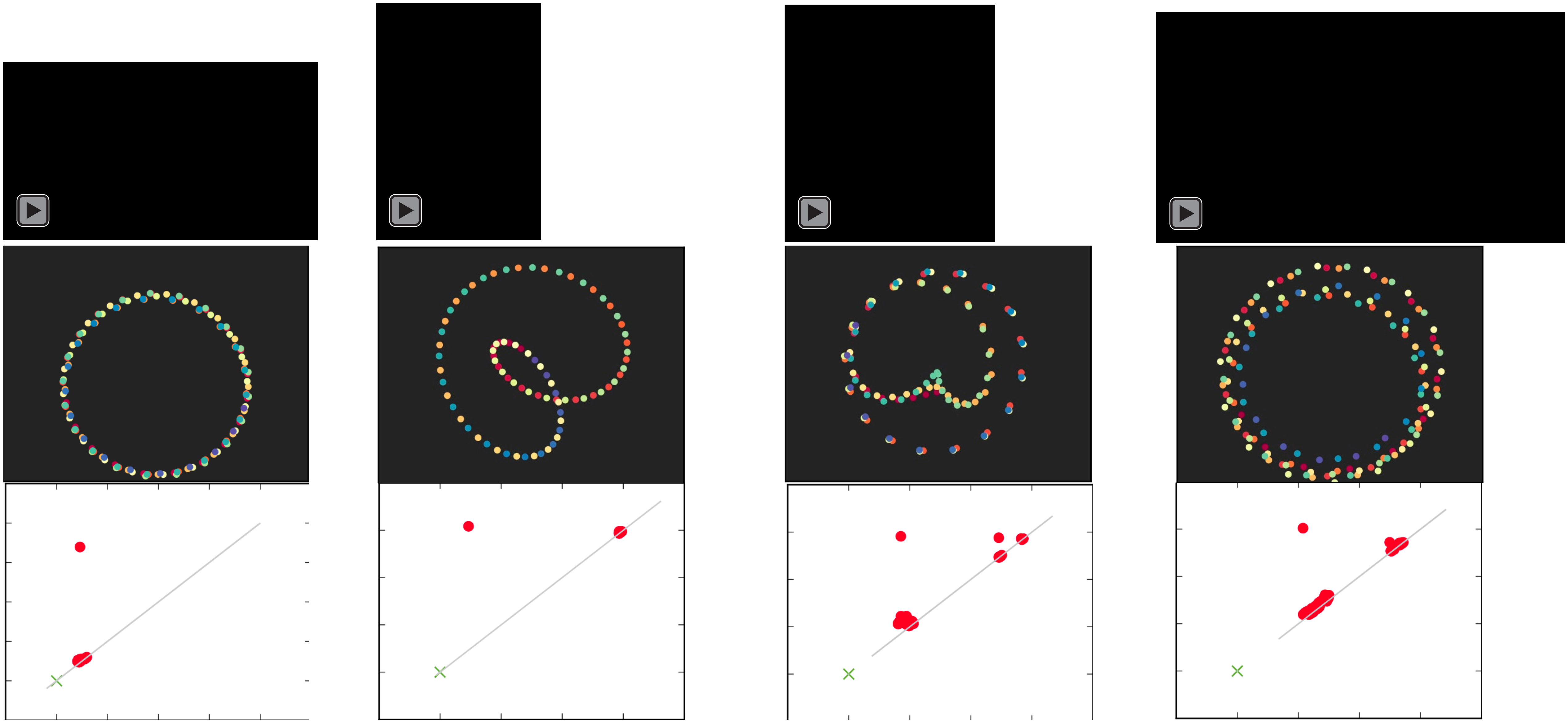
An Application of Persistent H_2

Chris Tralie and Jose Perea, *(Quasi)Periodicity Quantification in Video Data, Using Topology*

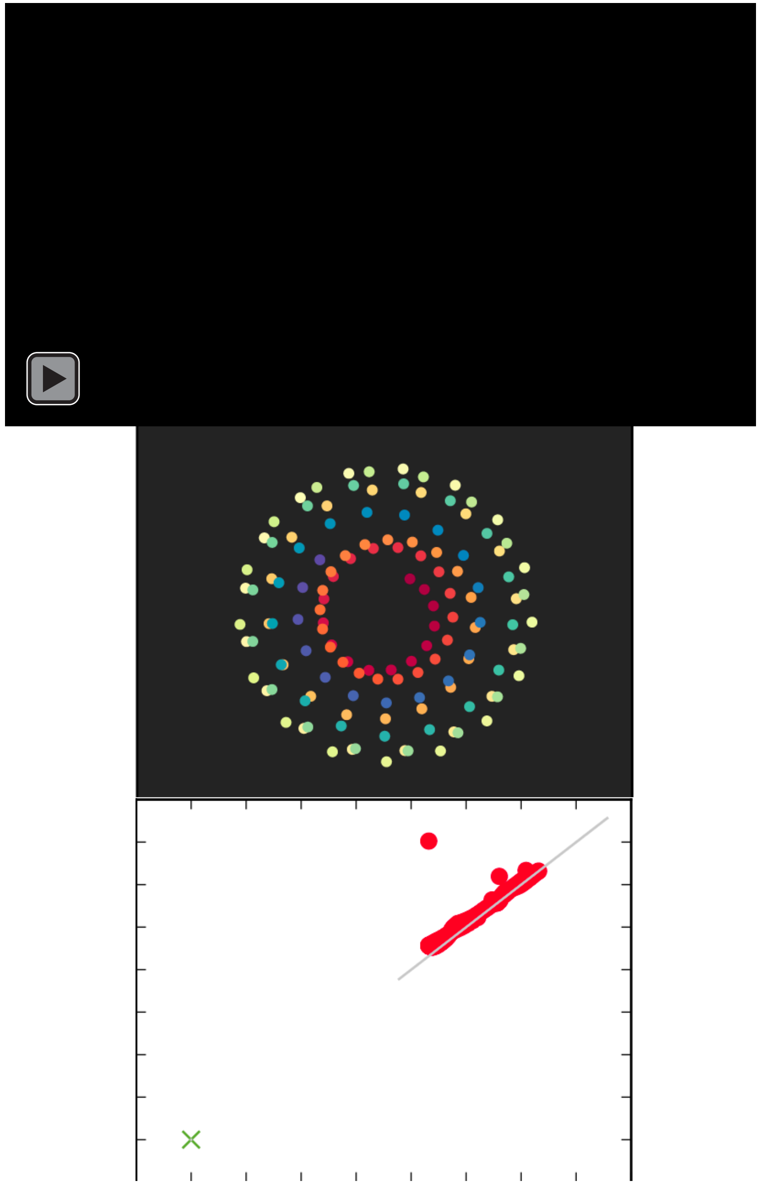


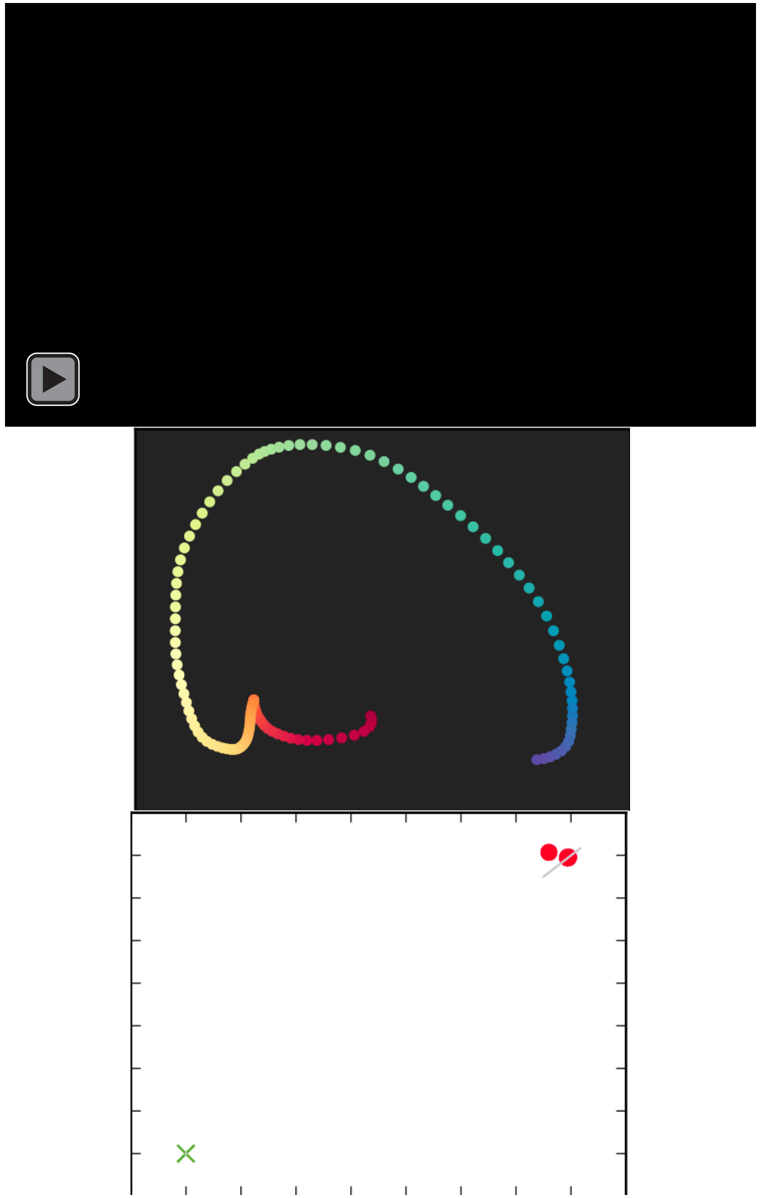


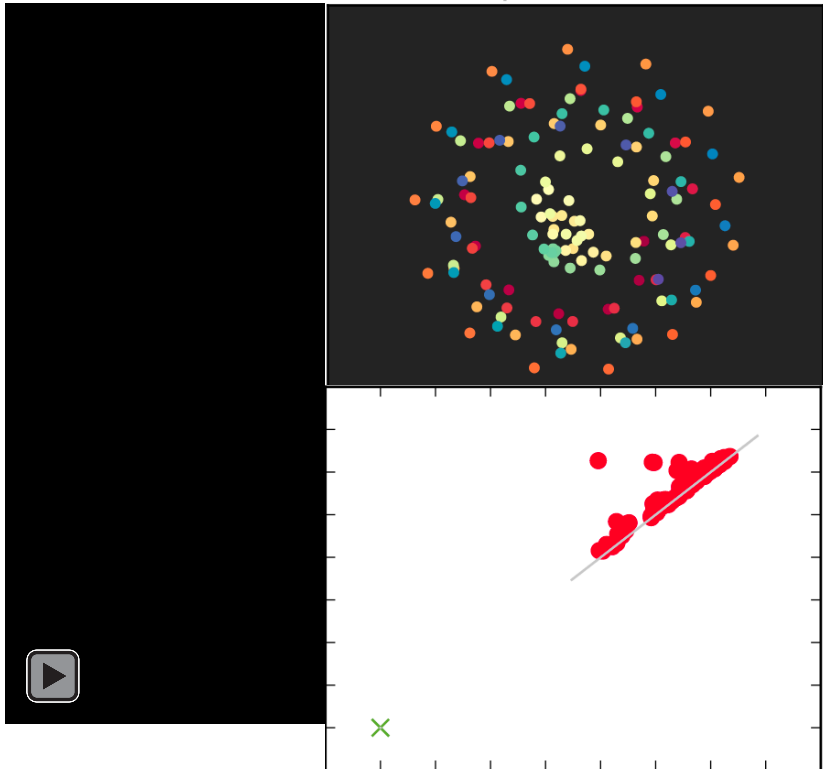
Videos Showing Periodicity



Non-Periodicity and Motion Blur







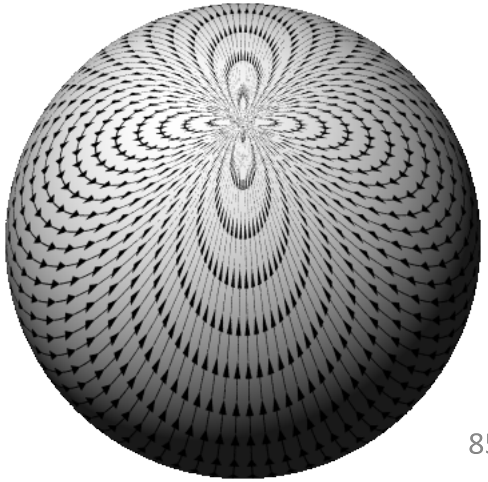
Underlying Theory: Takens' Embedding Theorem

- A remarkable result of Takens uses sliding windows to reconstruct a manifold M simply from a dynamical system defined on that manifold
- Let $\Phi: \mathbb{R} \times M \to M$ be a continuous time dynamical system. Also fix an observation function $\mathrm{F}: \mathrm{M} \to \mathbb{R}$ (e.g. temperature on the earth's surface) and, for any $p \in M$, $\phi_p: \mathbb{R} \to \mathbb{R}$ defined by $t \mapsto F \circ \Phi(t, p)$.
- For $d \geq 2 \dim M, \tau > 0$, consider the map $\phi: M \to \mathbb{R}^{d+1}$ given by $p \mapsto \left(\phi_p(0), \phi_p(\tau), \ldots, \phi_p(d\tau)\right)$.
- Takens' embedding theorem states that (under some extra hypotheses) the map ϕ reconstructs a copy of the manifold M in $\mathbb{R}^{d+1}$.

In practice:

- Dynamical system (M, Φ) unknown, only have time series $f(t) = \phi_p(t)$
- Takens' theorem implies that $SW_{d,\tau} f$ provides a topological copy of

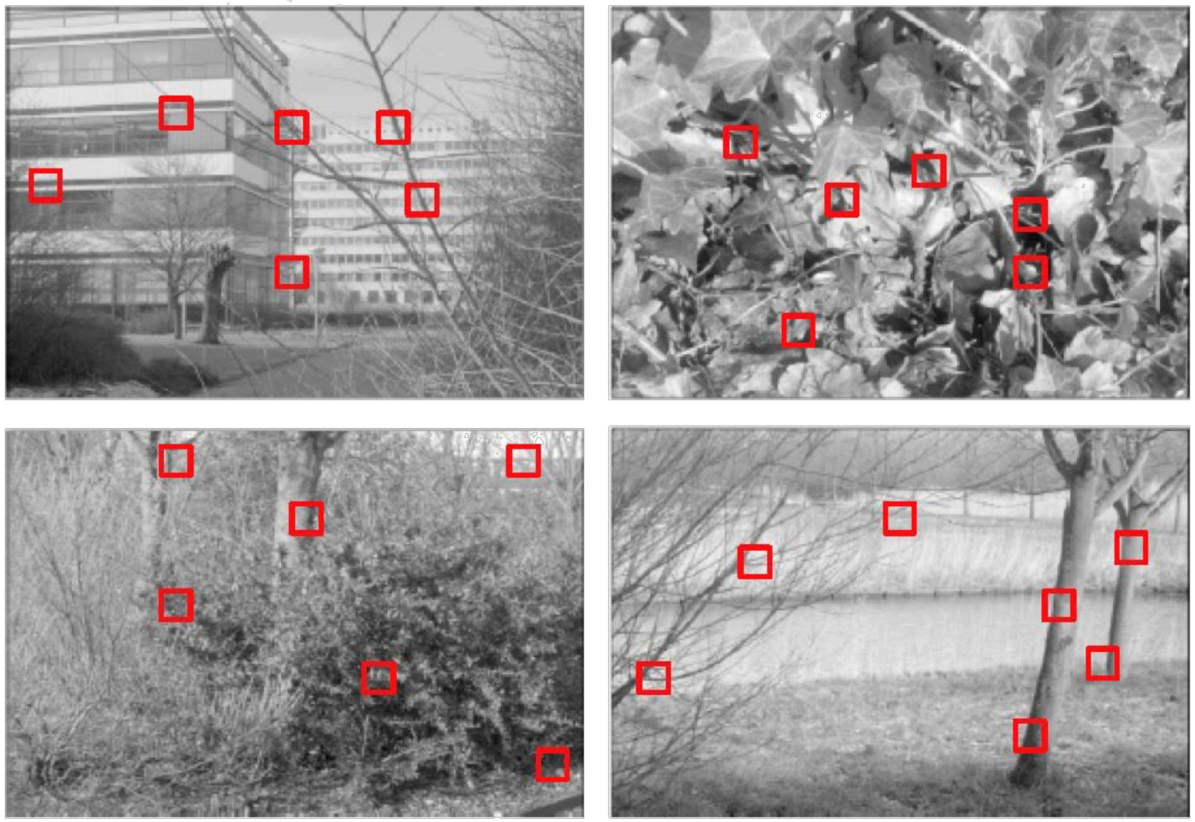
$$\{\Phi(t, p): t \in T\}$$



Topology of Image Patches

Natural Image Statistics

Input: 4 million data points on $\mathbb{S}^7$, coming from high-contrast 3×3 image patches



(source: [Lee, Pederson, Mumford 03])

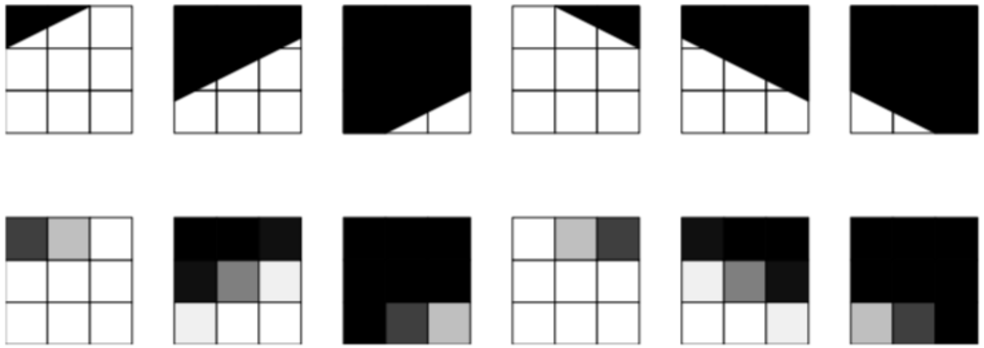
Preprocessing notes:

- Start with vector in $\mathbb{R}^9$
- Subtract mean intensity, rescale to unit length
- Output: points in $\mathbb{S}^7 \subseteq \mathbb{R}^8$

Natural Image Statistics

Expect to see high concentration of edge features

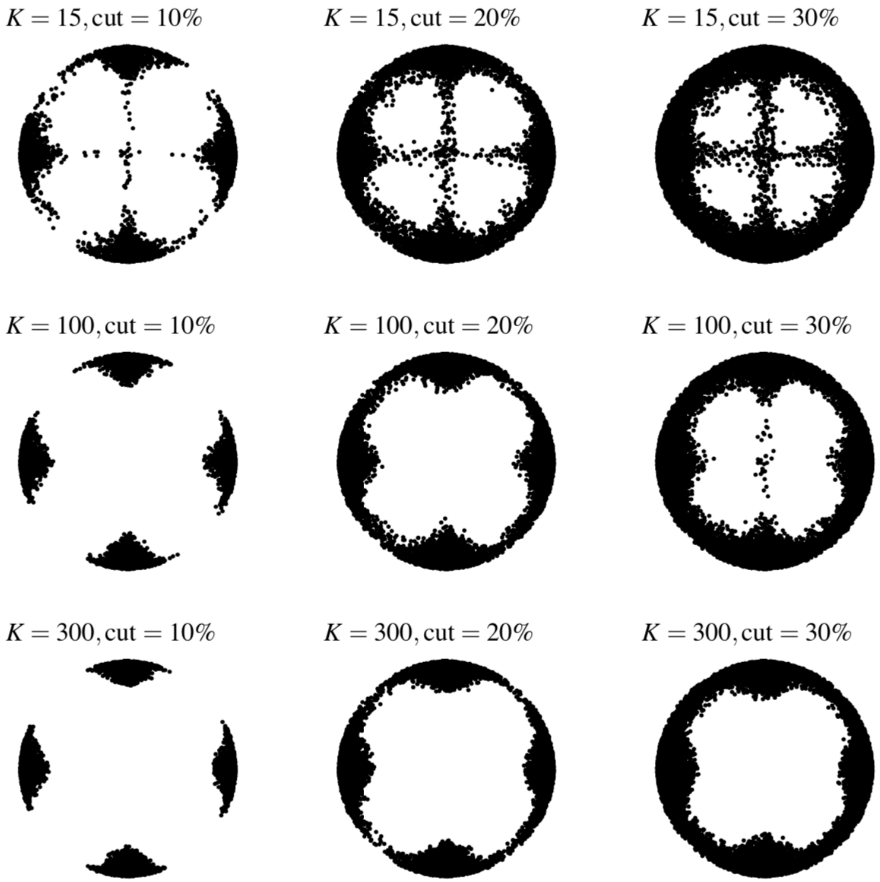
- Parametrized by angle and distance from center, expected to live on an annulus

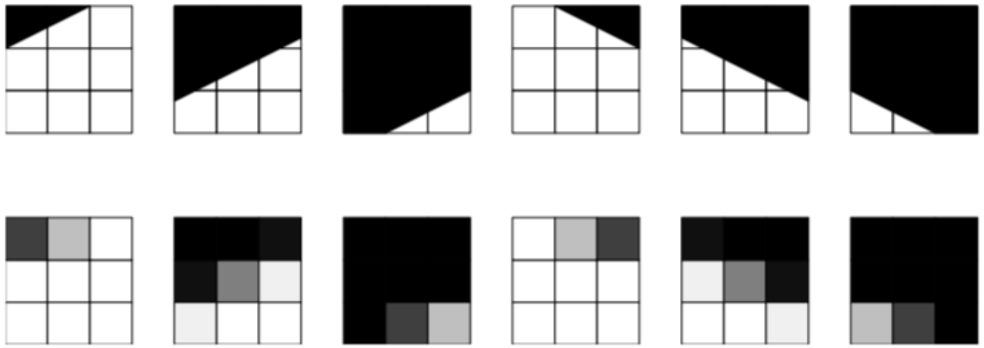


Natural Image Statistics

Expect to see high concentration of edge features

- Parametrized by angle and distance from center, expected to live on an annulus
- 2D view below---why does the cross appear?



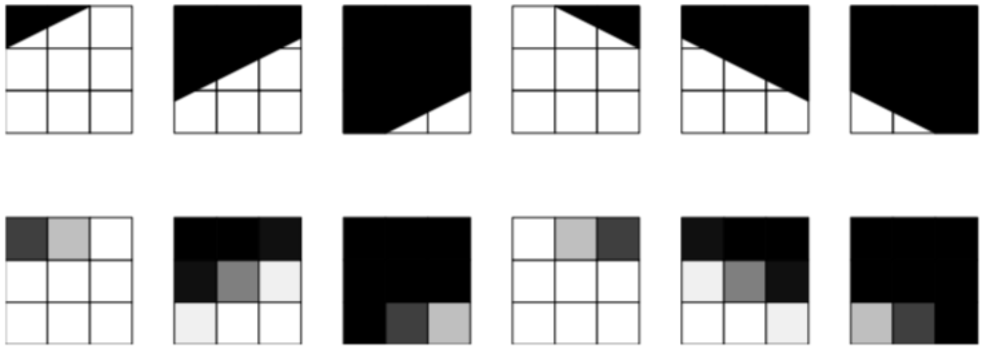


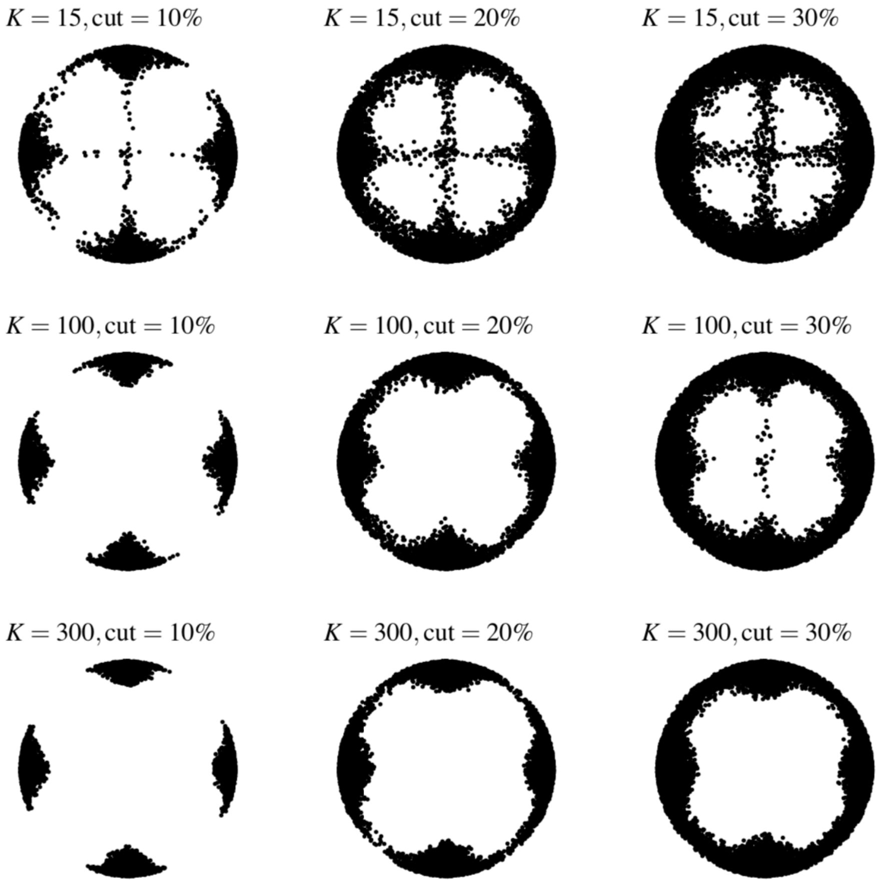
Natural Image Statistics

Carlsson, de Silva, *Topological estimation using witness complexes*

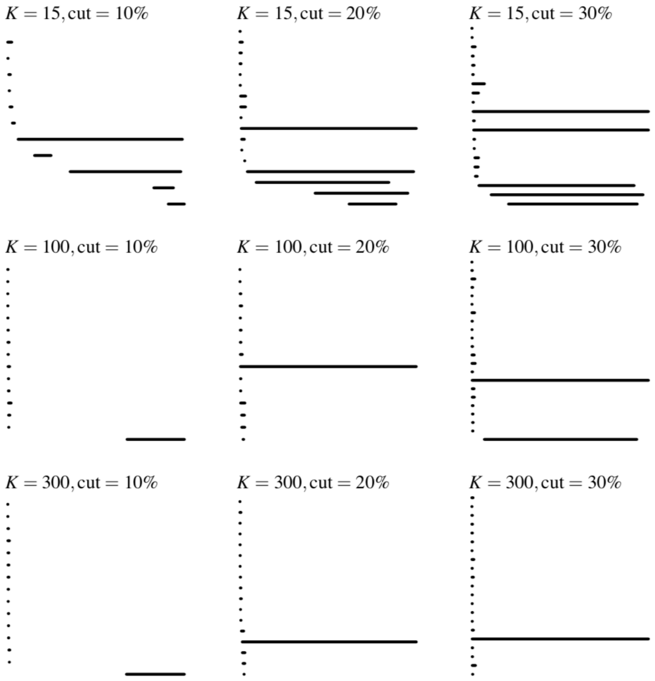
Expect to see high concentration of edge features

- Parametrized by angle and distance from center, expected to live on an annulus
- 2D view below---why does the cross appear?



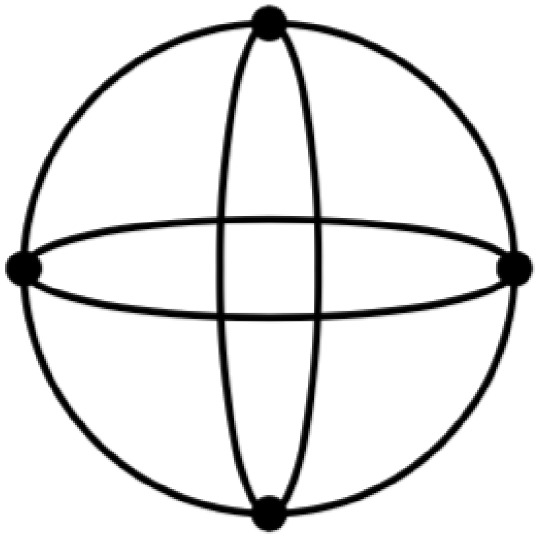


Betti numbers suggest $\beta_1 = 5$. Explanation?

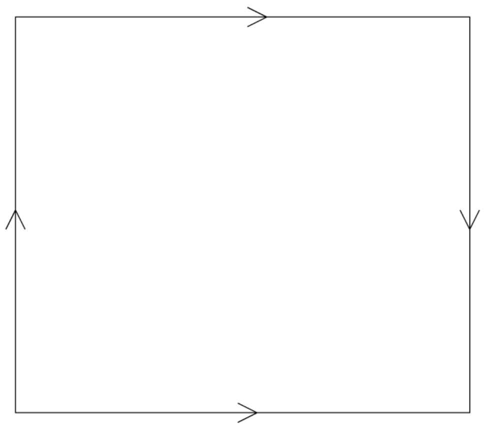


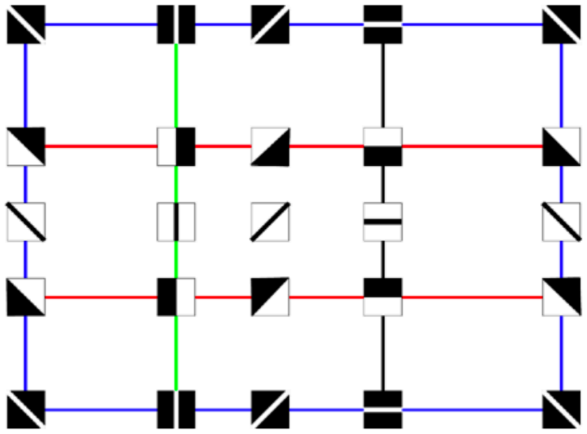
Three Circle Model

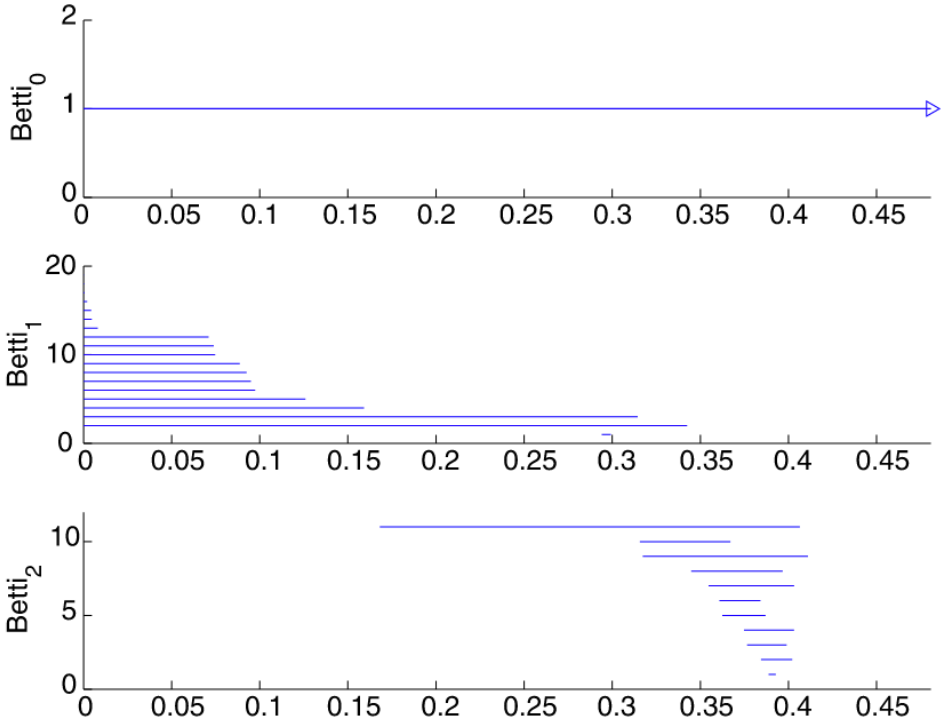
Carlsson, Ishkanov, de Silva, Zomorodian, *On the local behavior of spaces of natural images*



- Vertical and horizontal circles do not intersect
- Euler characteristic computation:
- β_1 = # arcs - # vertices + # connected components = 8-4+1 = 5
- PLEX results suggested 2-manifold. Torus or Klein bottle?

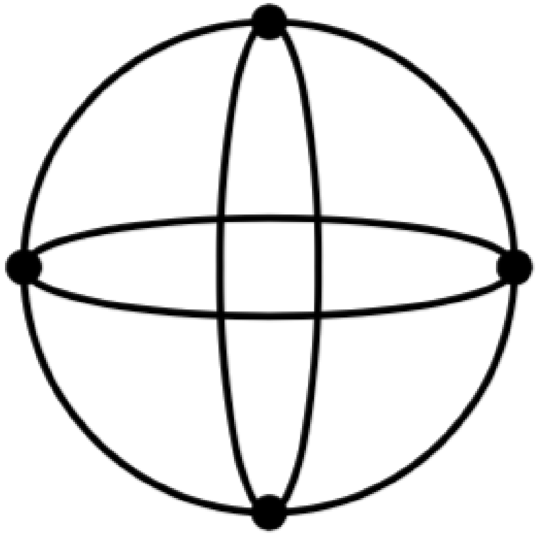




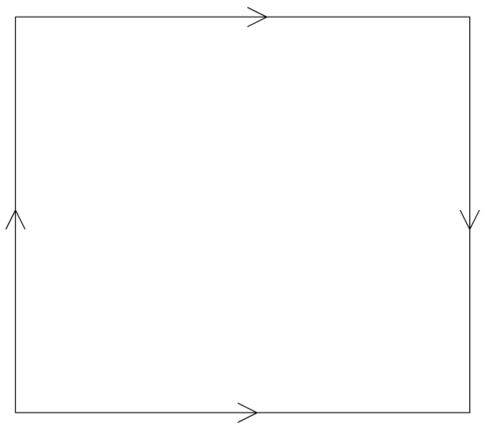


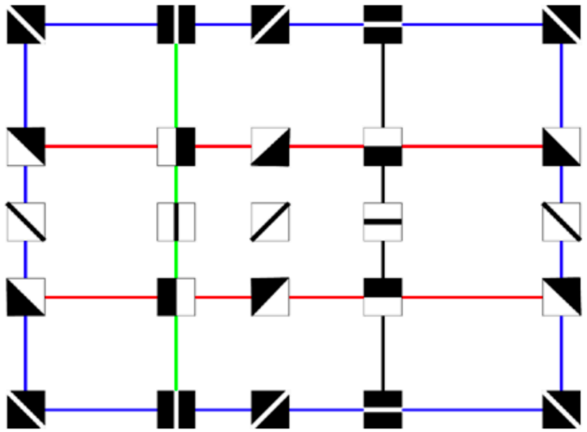
Three Circle Model

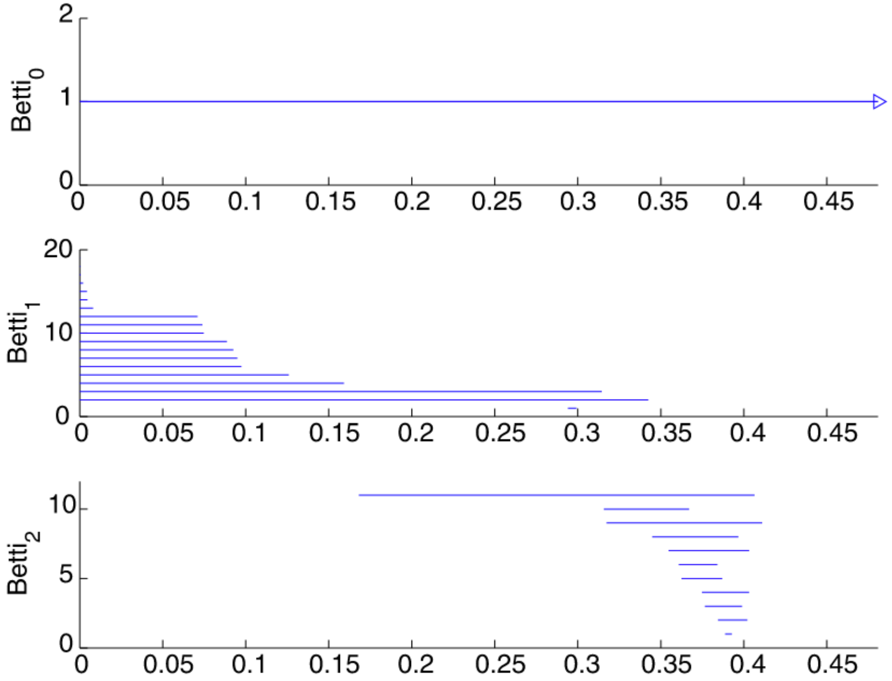
Carlsson, Ishkanov, de Silva, Zomorodian, *On the local behavior of spaces of natural images*



- Vertical and horizontal circles do not intersect
- Euler characteristic computation:
- β_1 = # arcs - # vertices + # connected components = 8-4+1 = 5
- PLEX results suggested 2-manifold.
- Computational trick: augment data with ideal Klein bottle, observe that barcode remains unchanged. So all homology comes from the Klein bottle.







More TDA Info

- https://appliedtopology.org/
- https://scikit-tda.org/
- https://github.com/giotto-ai/giotto-tda
- https://tda-in-ml.slack.com/

That’s All