Advanced Data Visualization

CS 6965

Fall 2019

Prof. Bei Wang Phillips

University of Utah

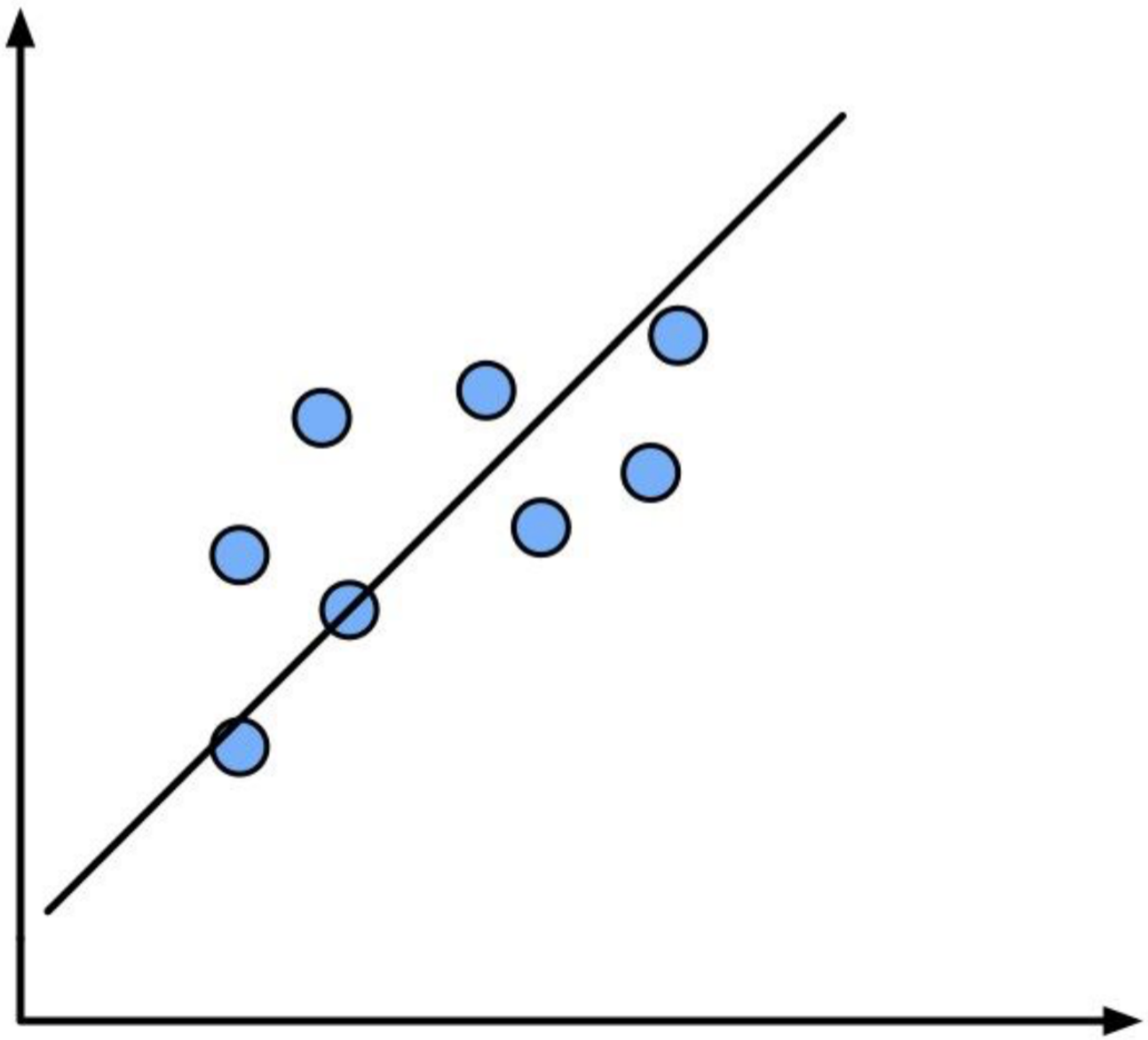
Lecture 26

Structural Inference of High-dim Data

HD+TOPO

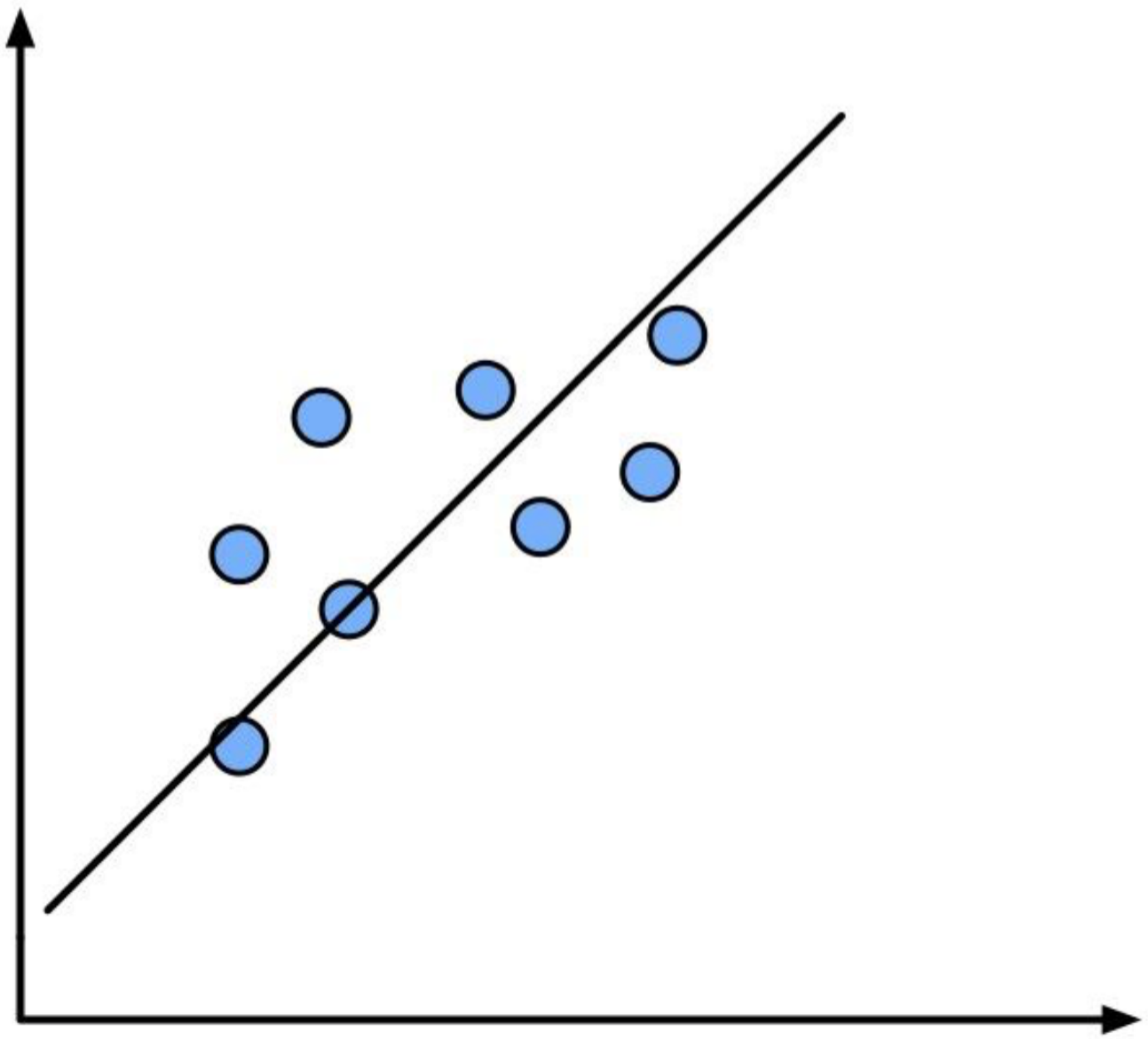
"Data has shape, and shape matters."
– Gunnar Carlsson

Linear regression



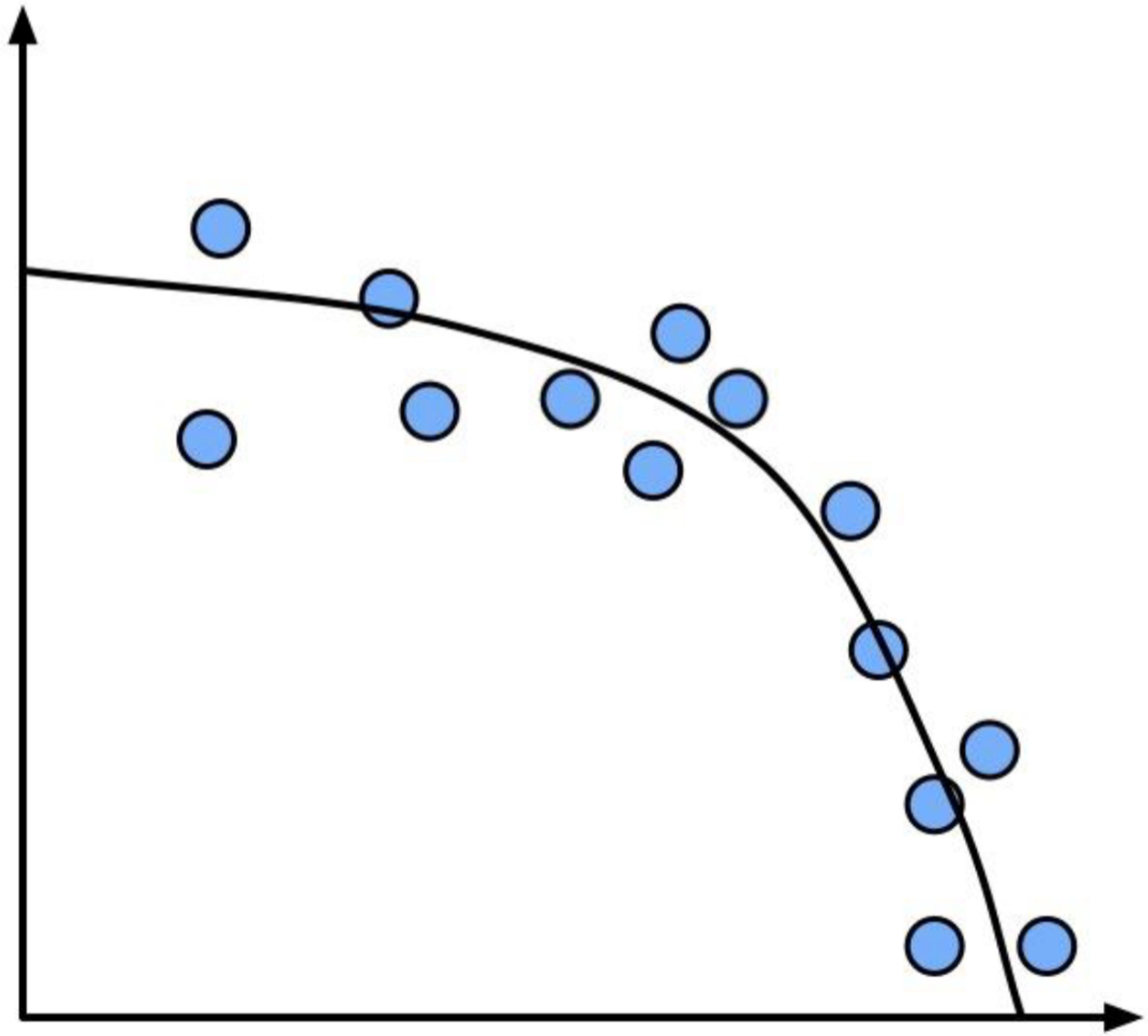
- Shape of data?
- A line.

Linear regression



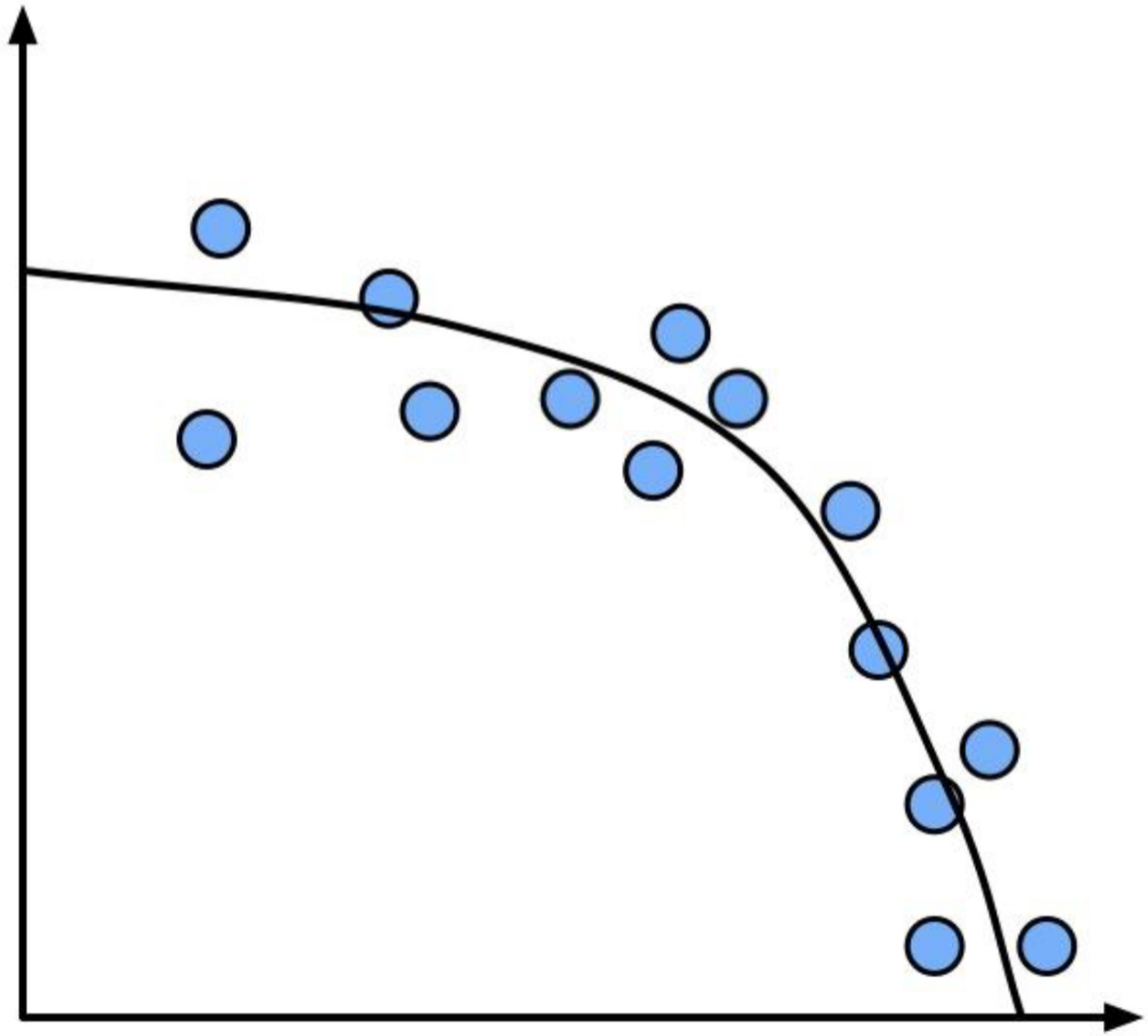
- Shape of data?
- A line.

Cubic polynomial regression



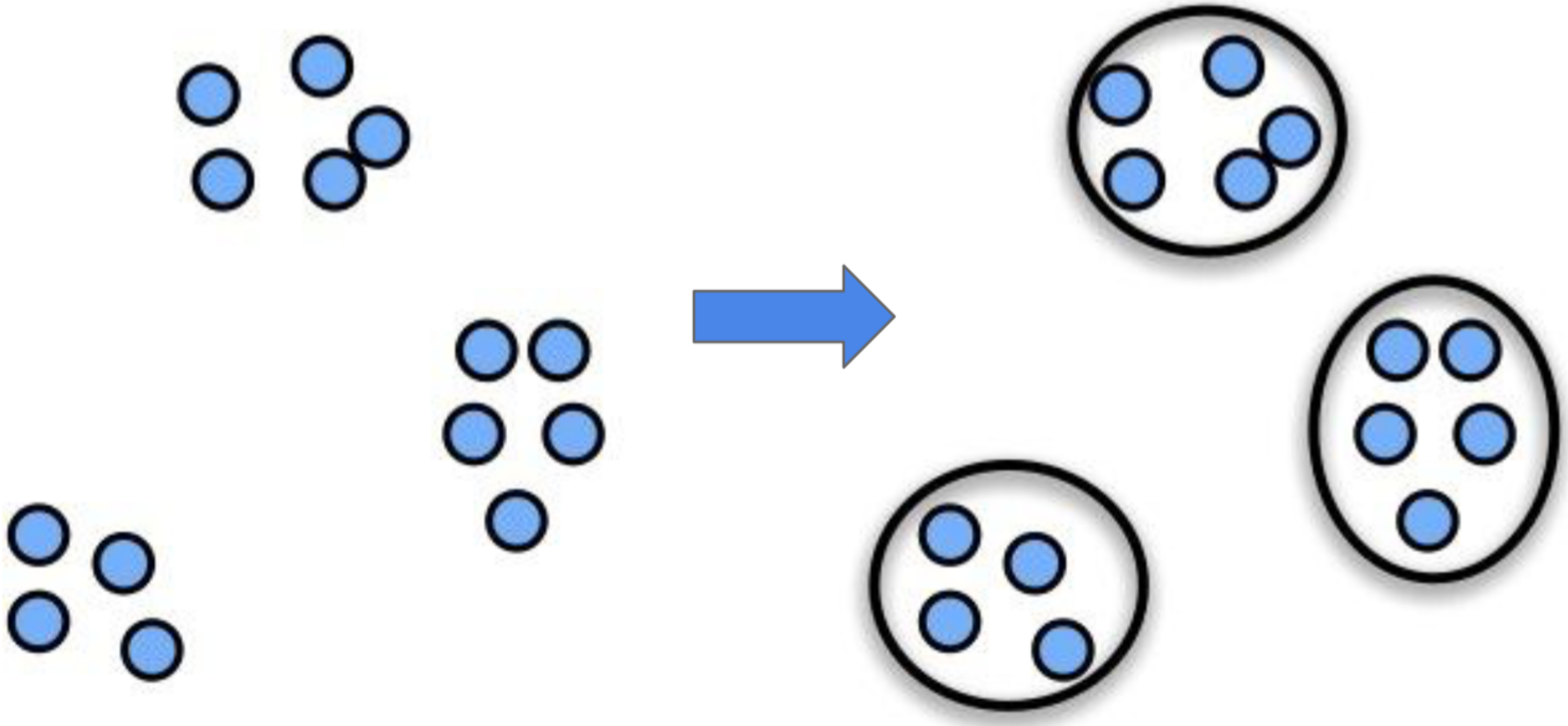
- Shape of data?
- A curve.

Cubic polynomial regression



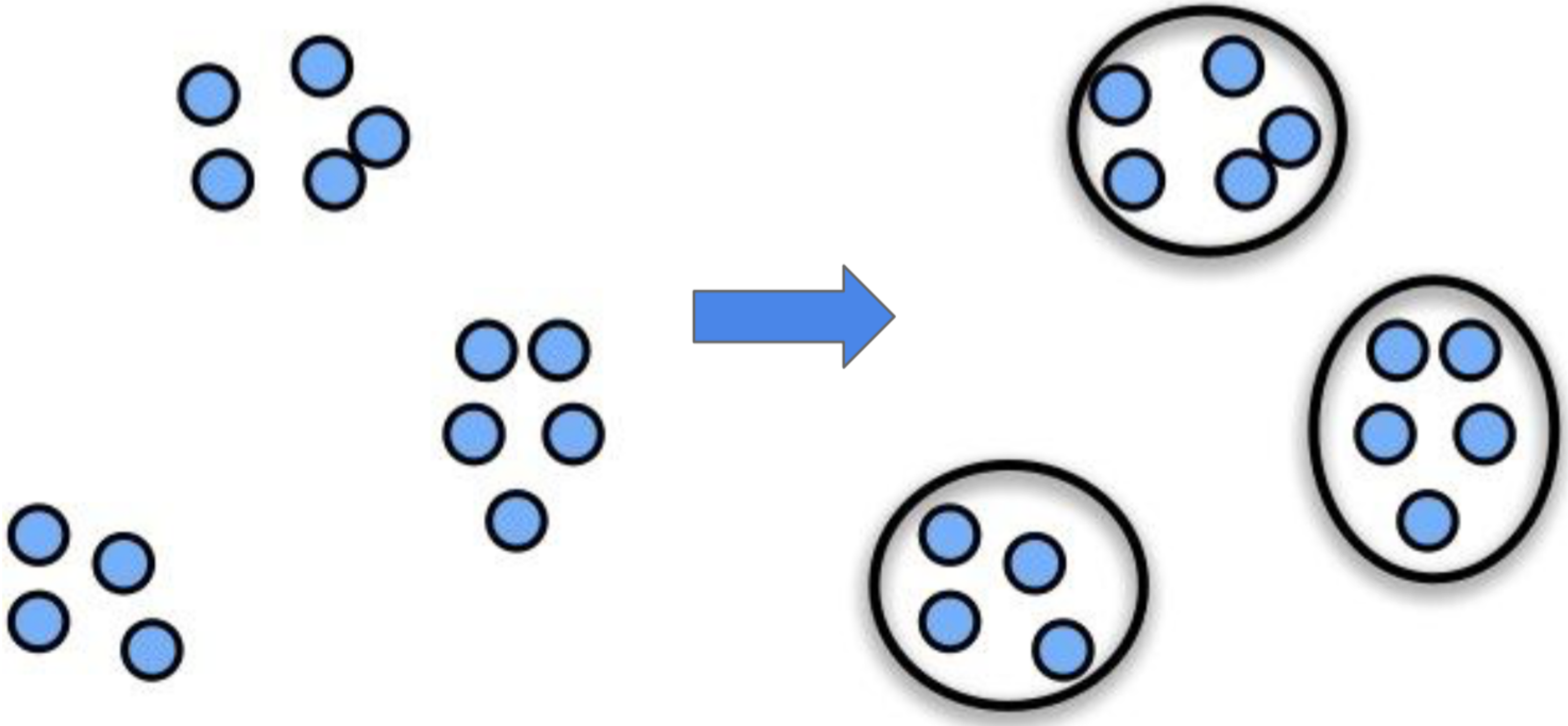
- Shape of data?
- A curve.

Clustering



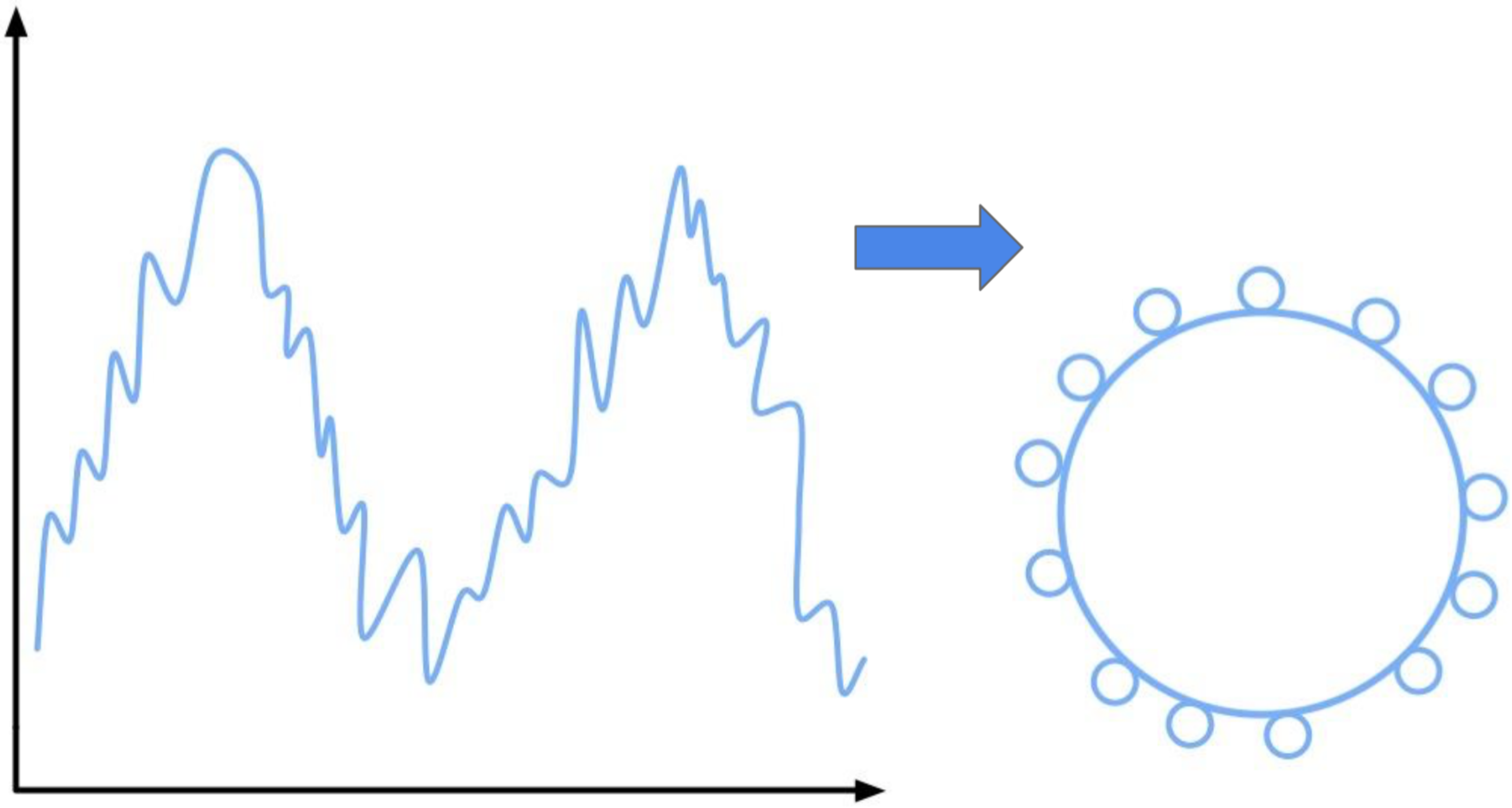
- Shape of data?
- Clusters.

Clustering



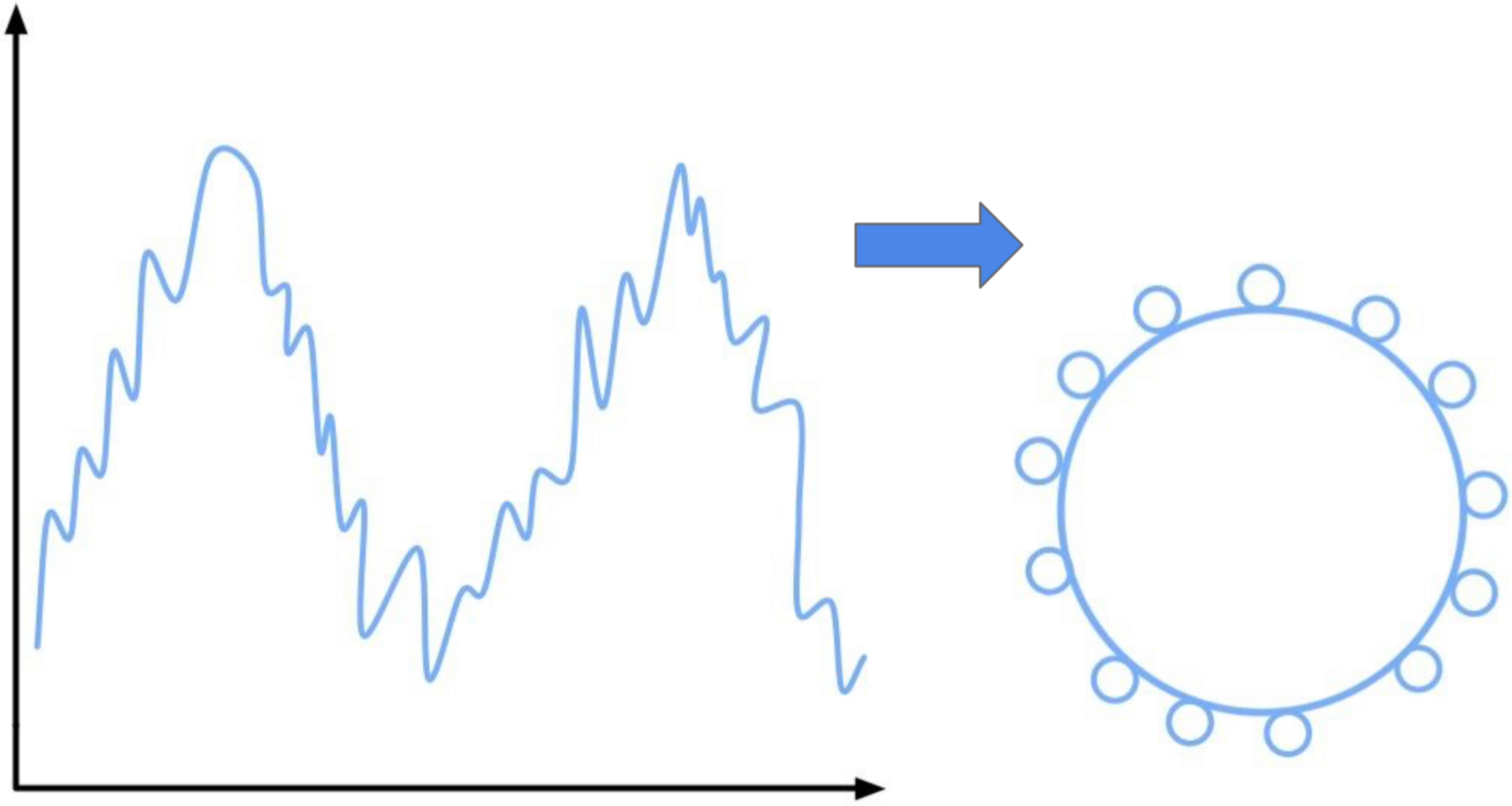
- Shape of data?
- Clusters.

Time series analysis



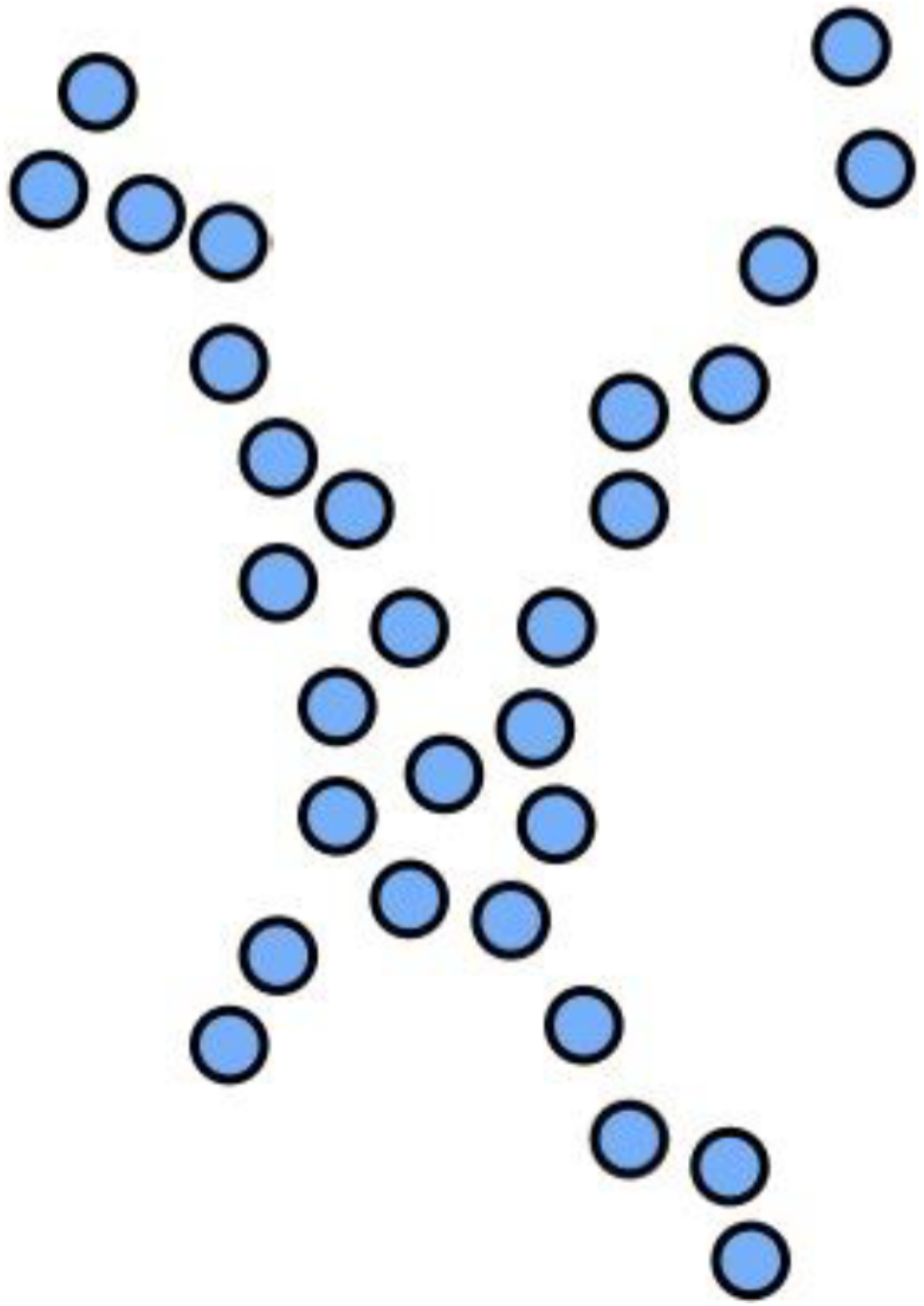
- Shape of data?
- Depends on the mapping.

Time series analysis



- Shape of data?
- Depends on the mapping.

Discrete samples: a point cloud



- Shape of data?
- Depending on the scale (or the resolution).

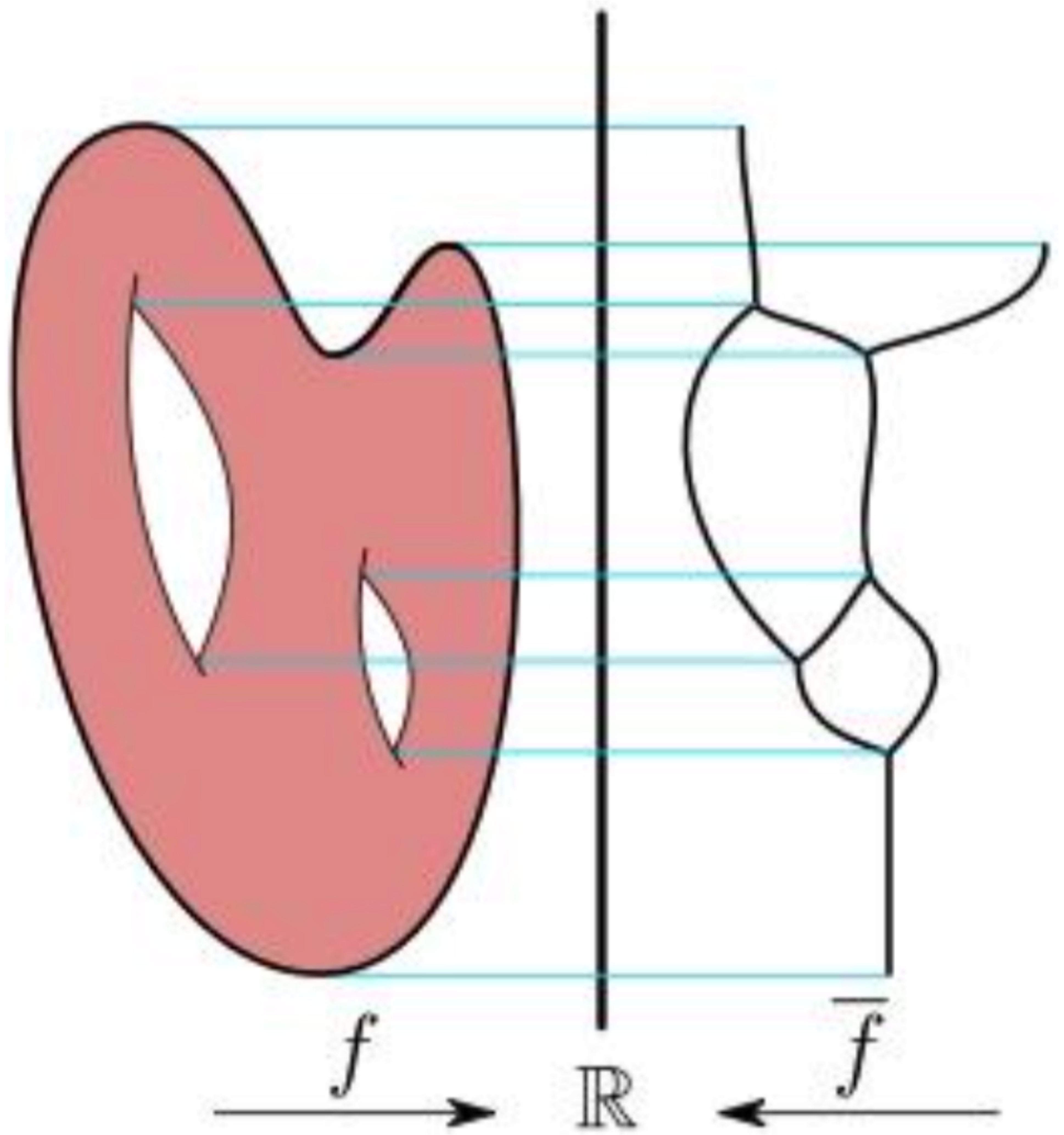
Discrete samples: a point cloud

- Shape of data?
- Depending on the scale (or the resolution).

Discrete samples: a point cloud

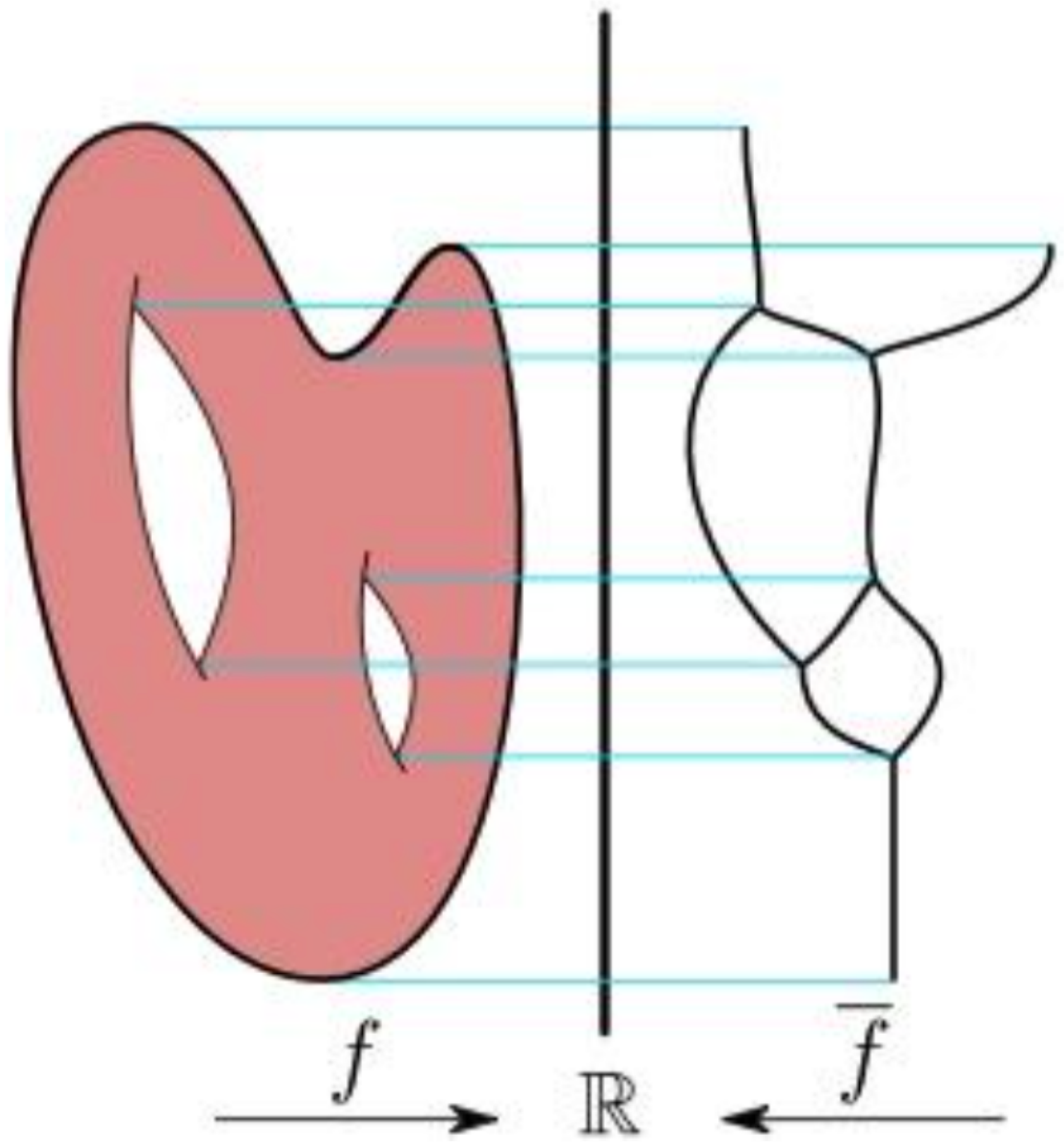
- Shape of data?
- Depending on the scale (or the resolution).

A scalar function defined on a manifold



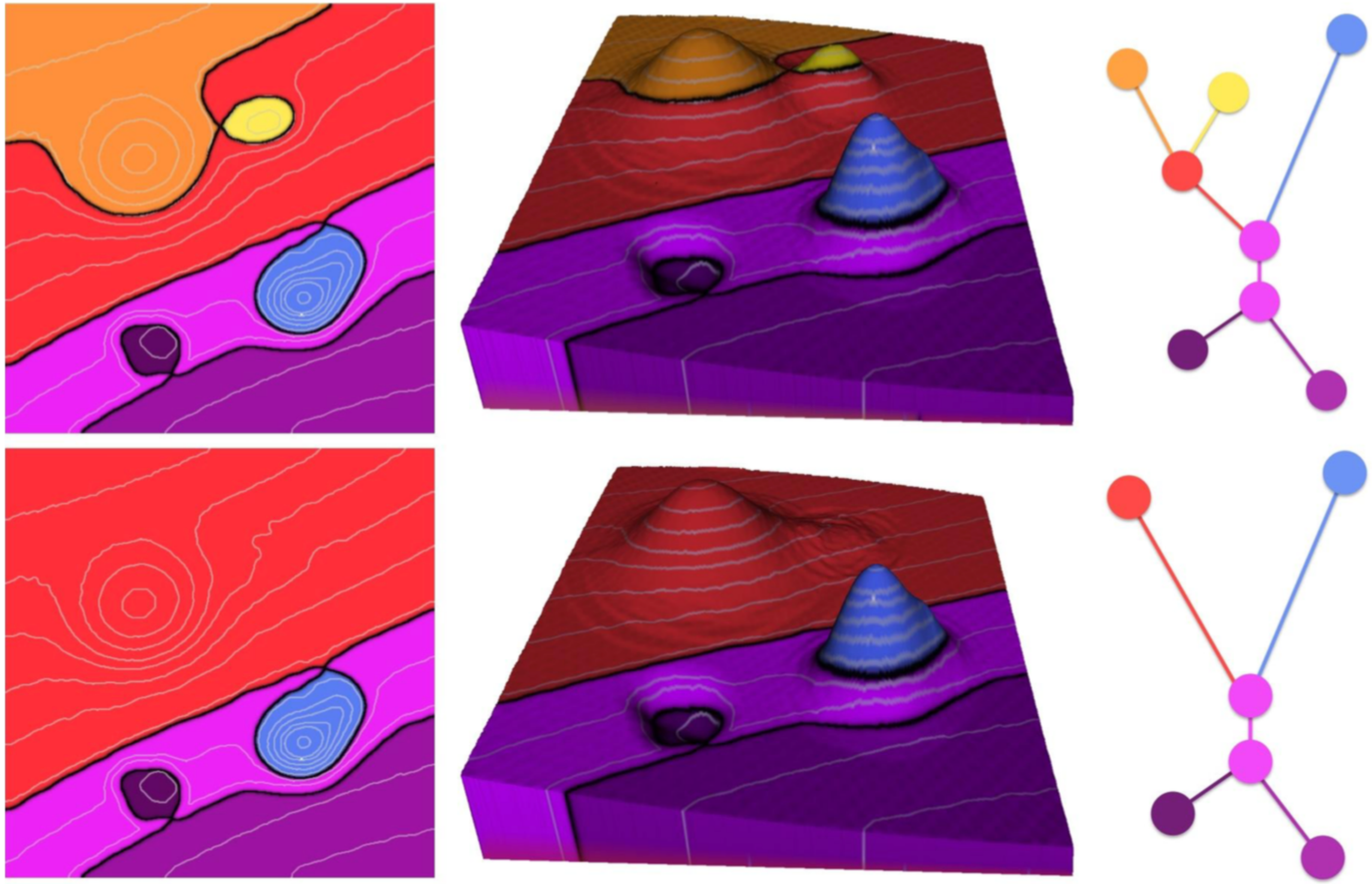
- Shape of data?
- A Reeb graph.

A scalar function defined on a manifold



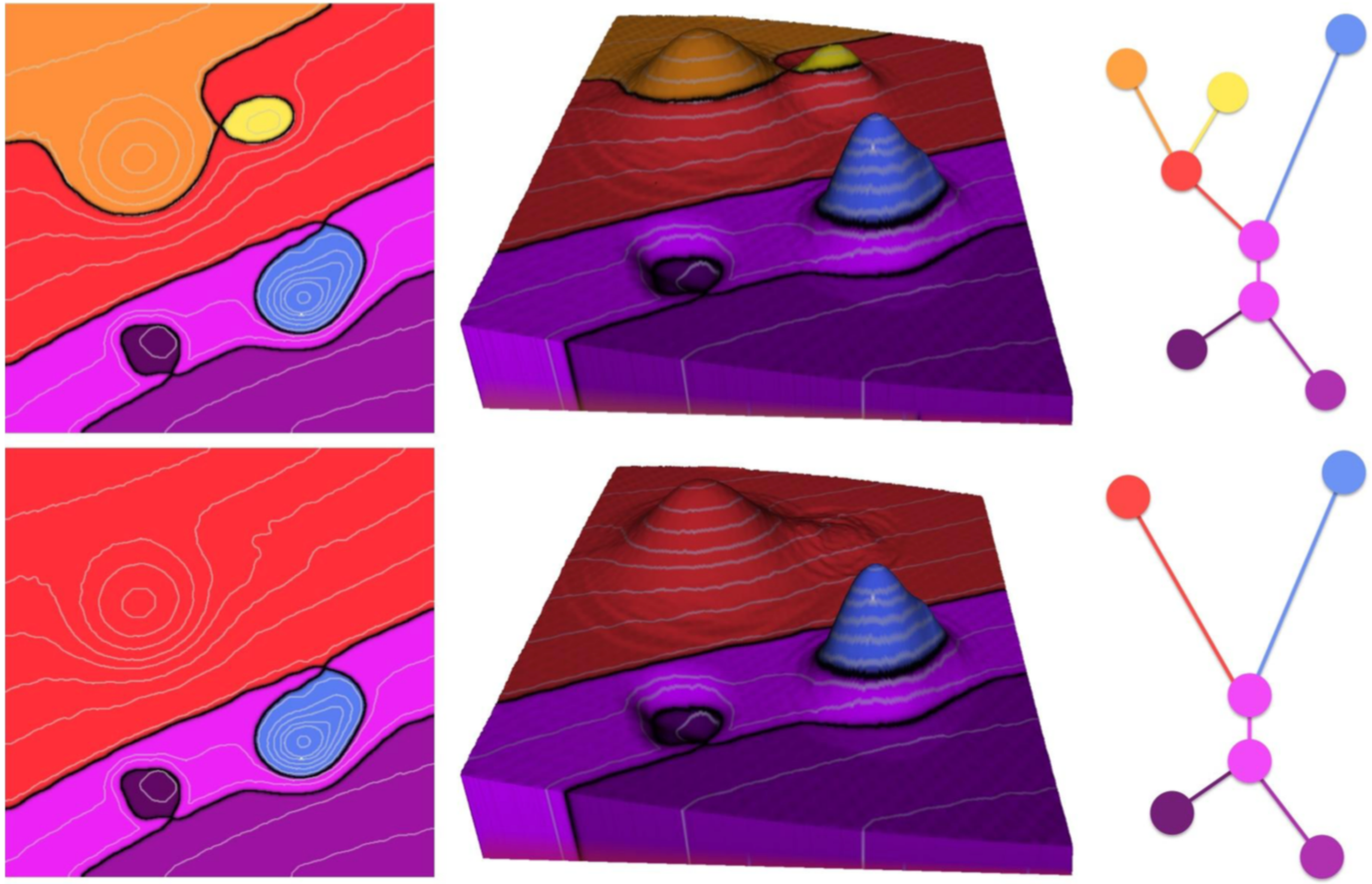
- Shape of data?
- A Reeb graph.

Elevation on terrain: a scalar function on a 2D domain



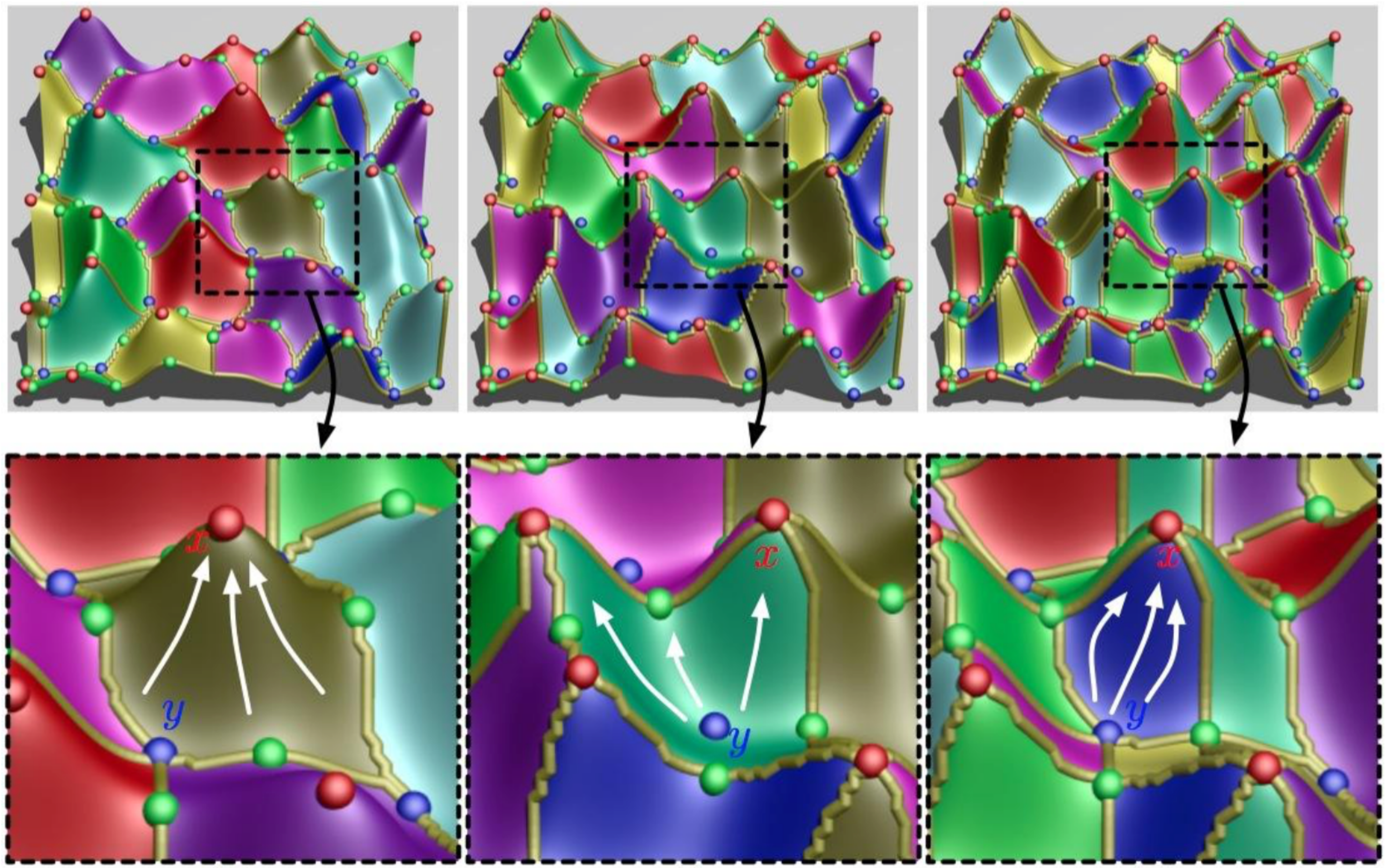
- Shape of data?
- A contour tree.

Elevation on terrain: a scalar function on a 2D domain



- Shape of data?
- A contour tree.

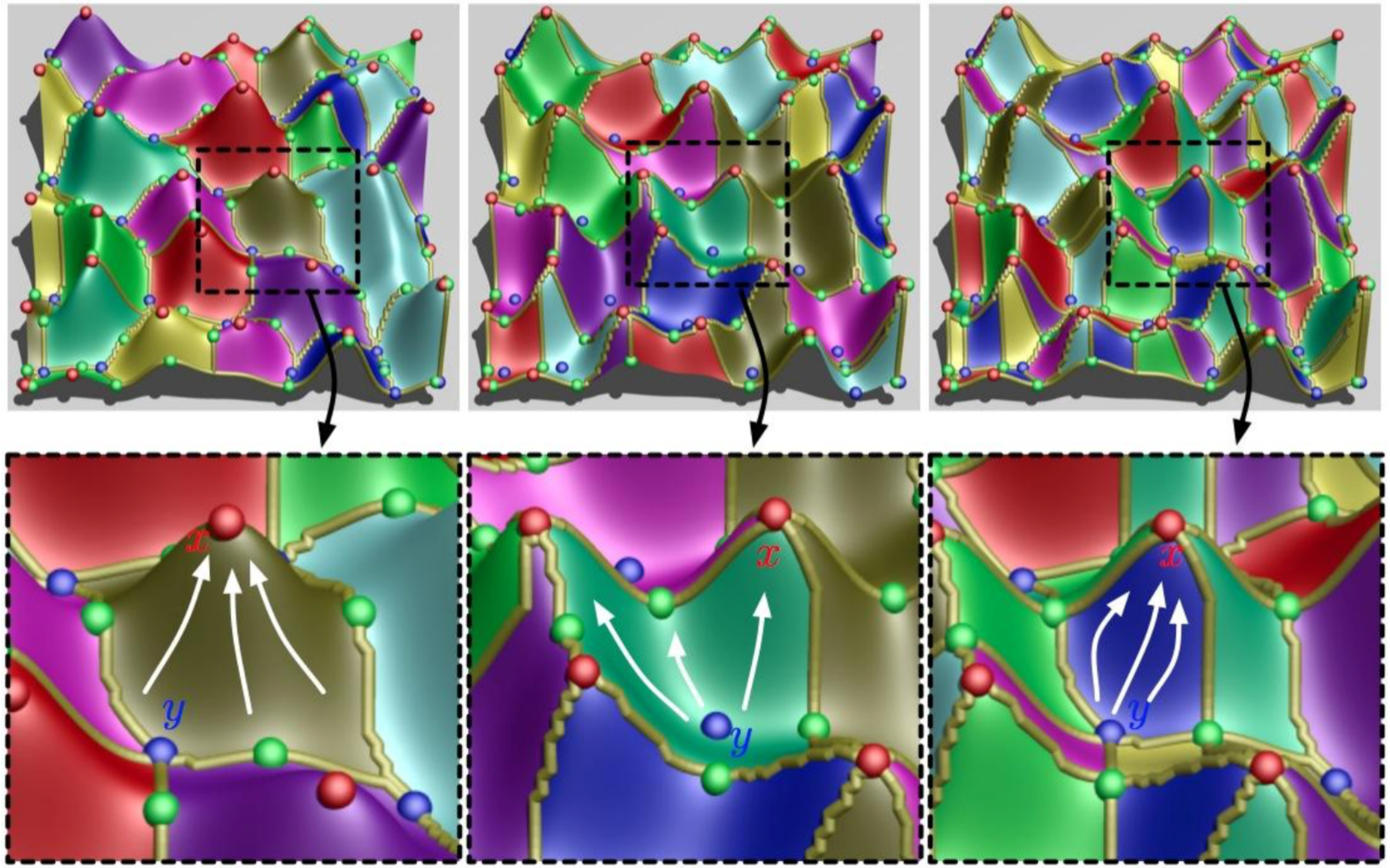
Elevation on terrain: a scalar function on a 2D domain



(a) (b) (c)

- Shape of data?
- A Morse-Smale complex.

Elevation on terrain: a scalar function on a 2D domain

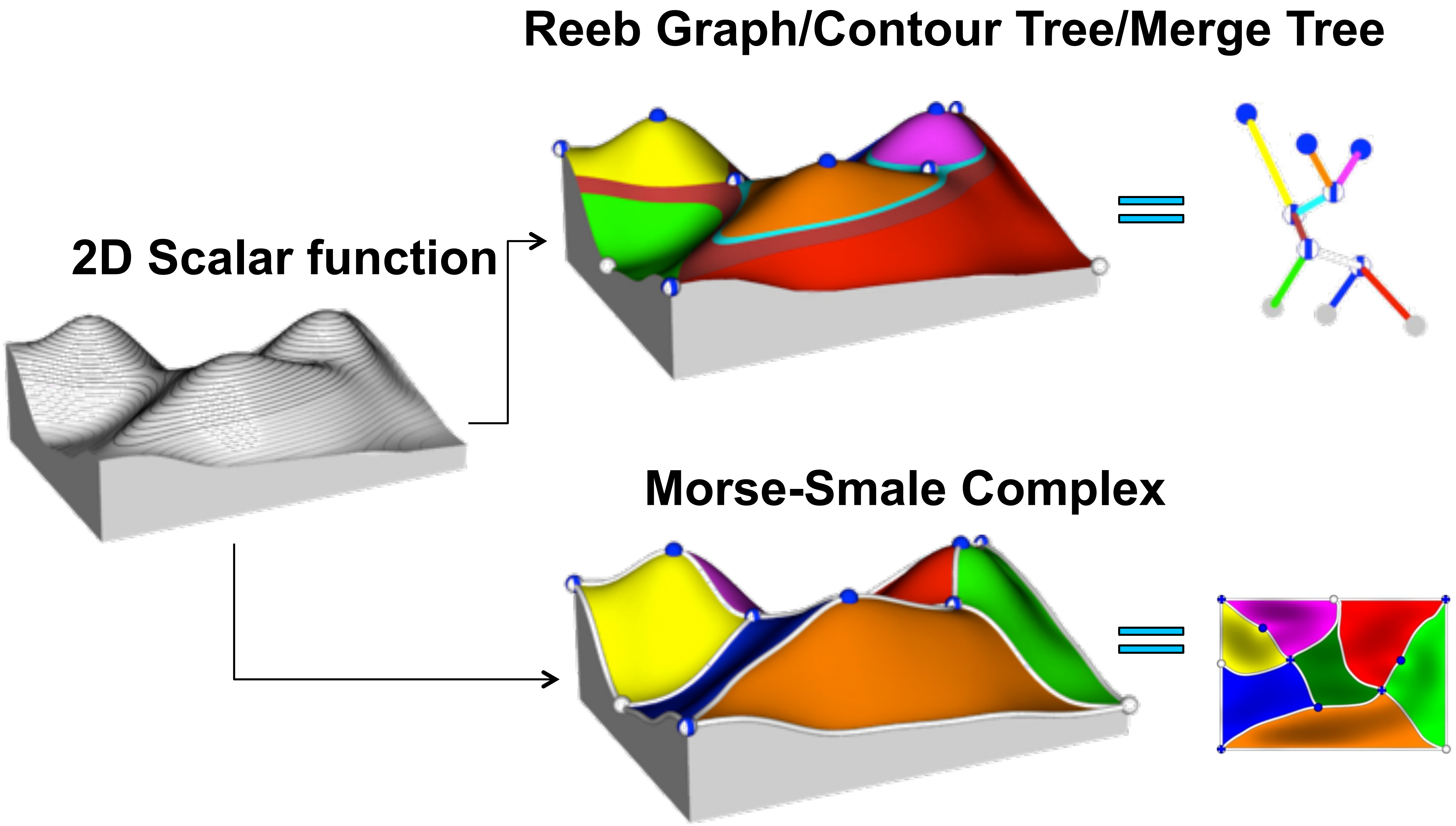


(a) (b) (c)

- Shape of data?
- A Morse-Smale complex.

Some basic tools in topological data analysis (TDA)

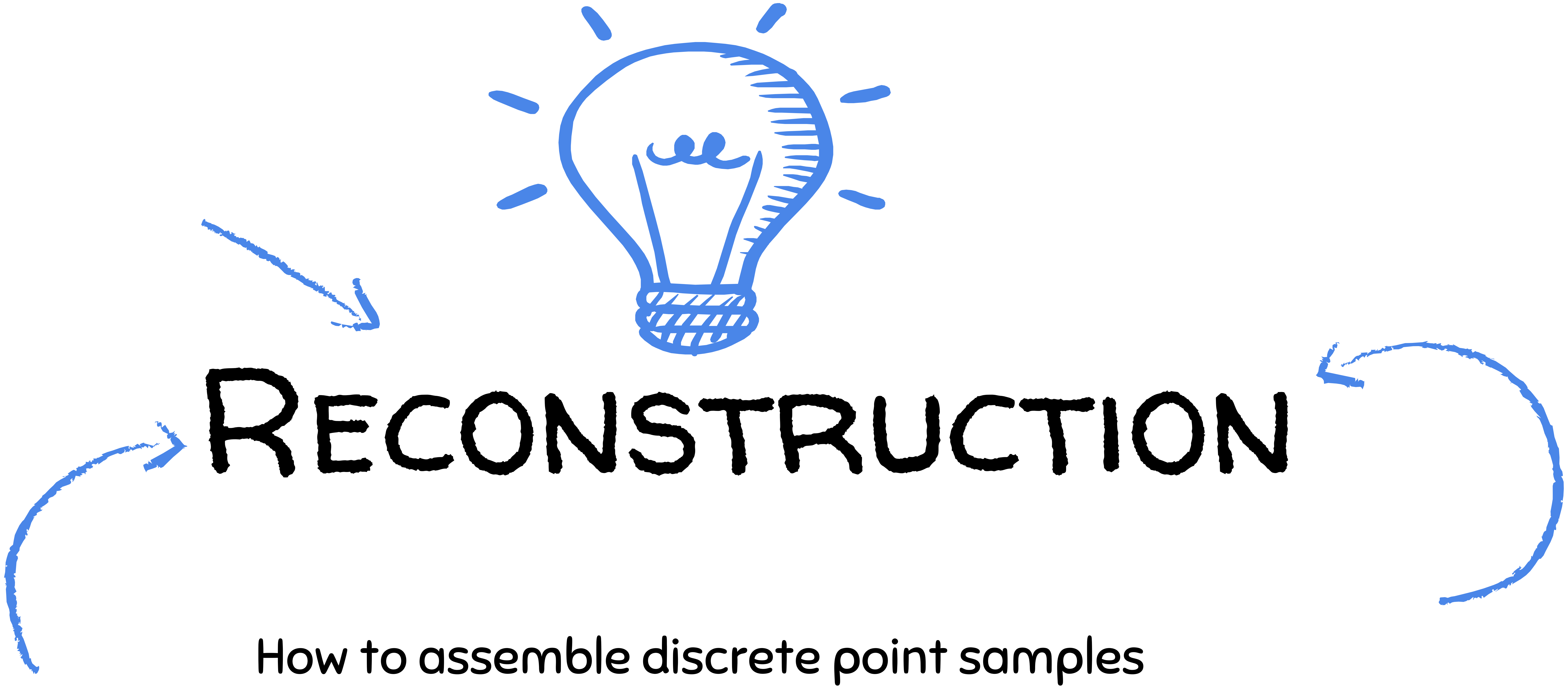
- Abstraction of the data: topological structures and their combinatorial representations
- Separate features from noise: persistent homology



Fundamental Tasks in Topological Data Analysis

Topology + Point Cloud = Magic Happens!

Task 1



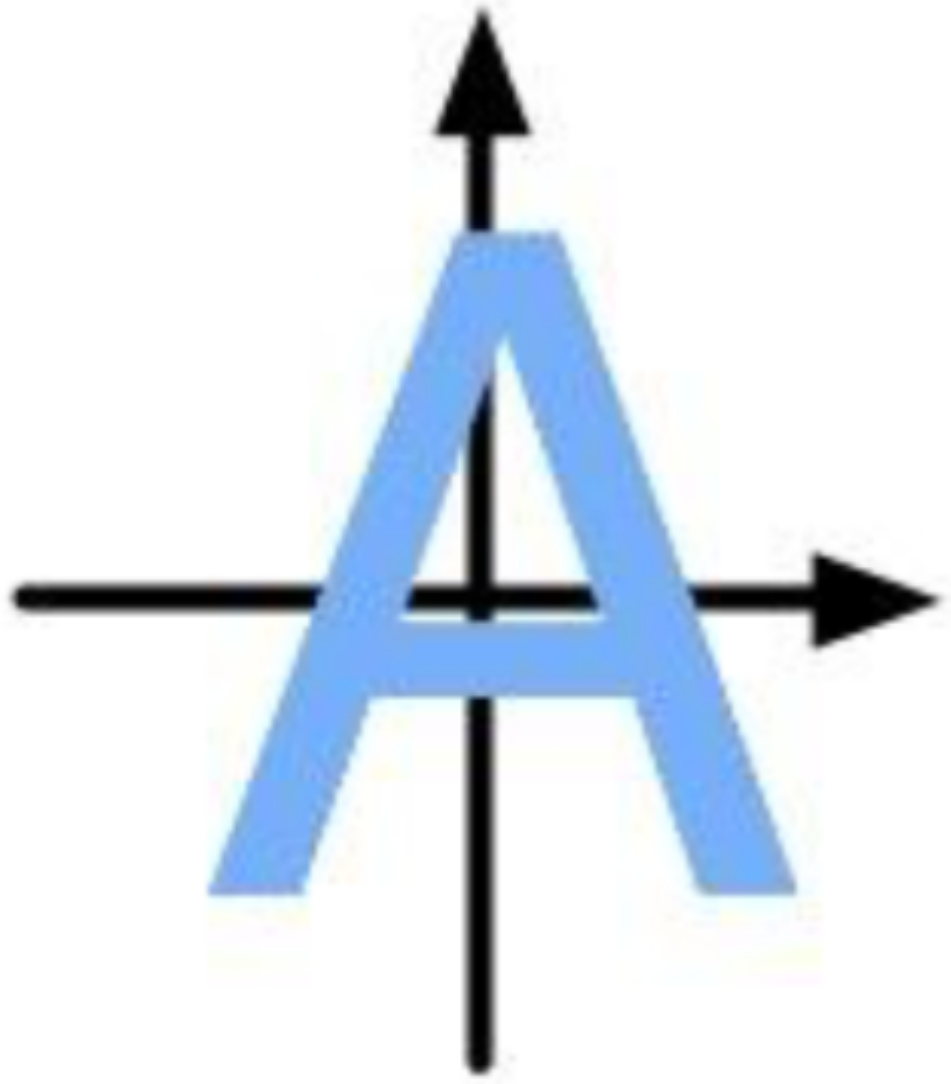
How to assemble discrete point samples
into a global structure?

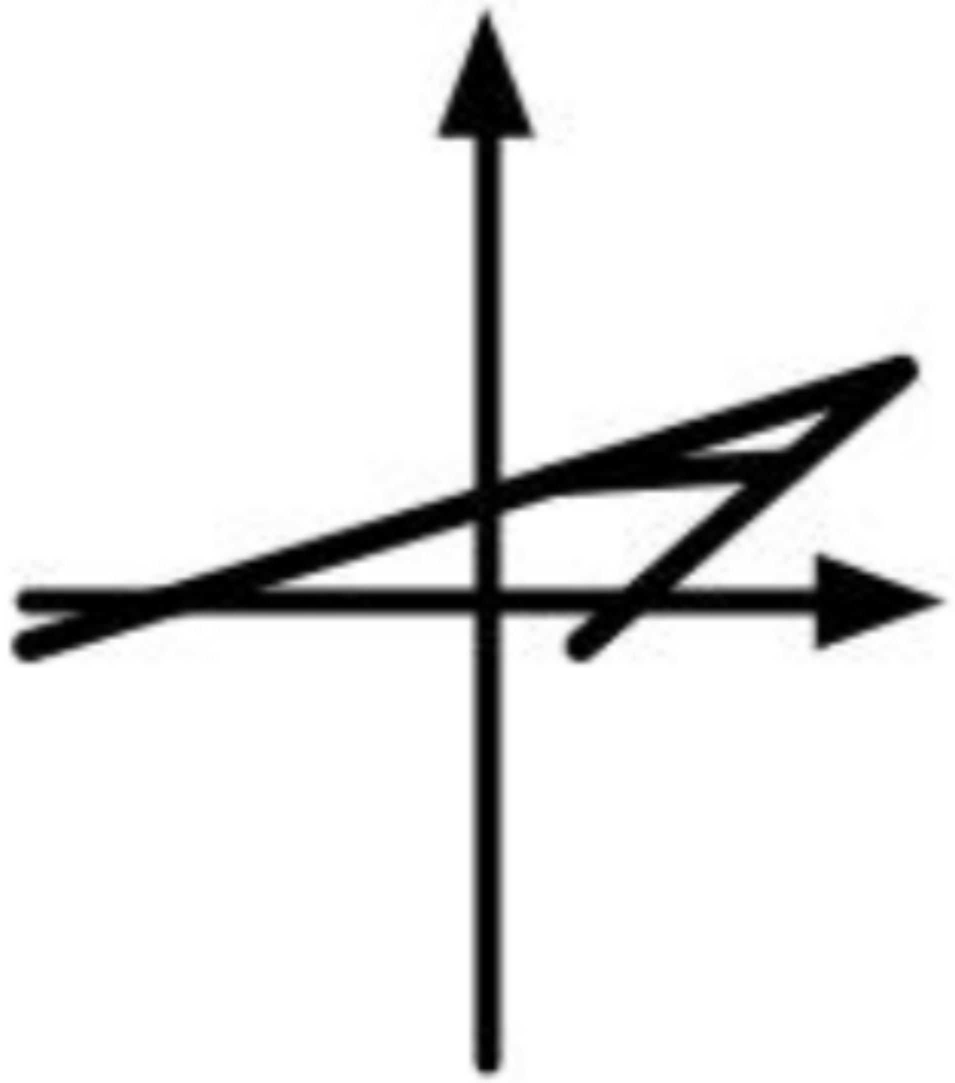
Task 2

INFERENCE

How to infer high-dimensional structures
from low-dimensional representations?

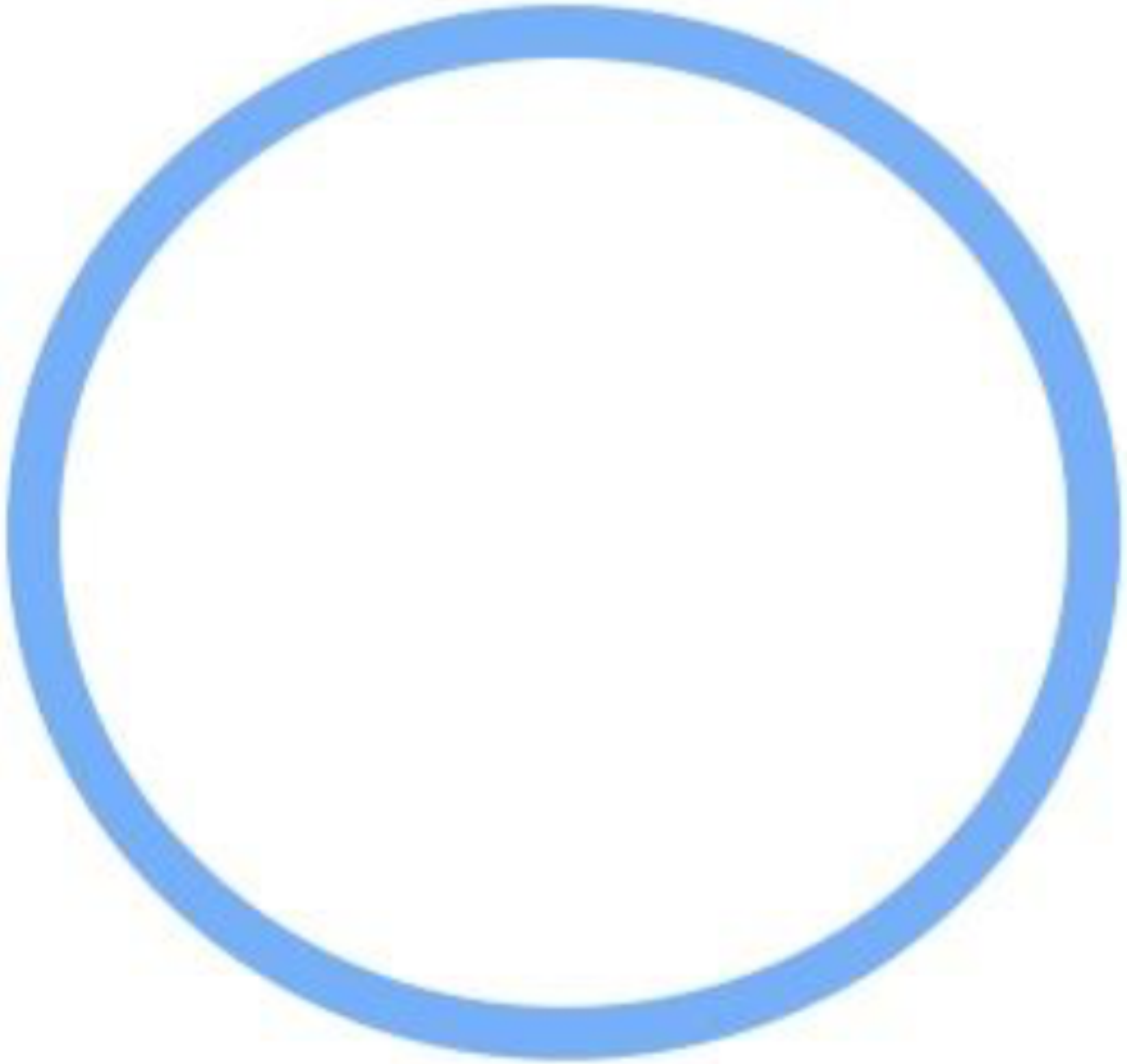
Key idea 1: coordinate free

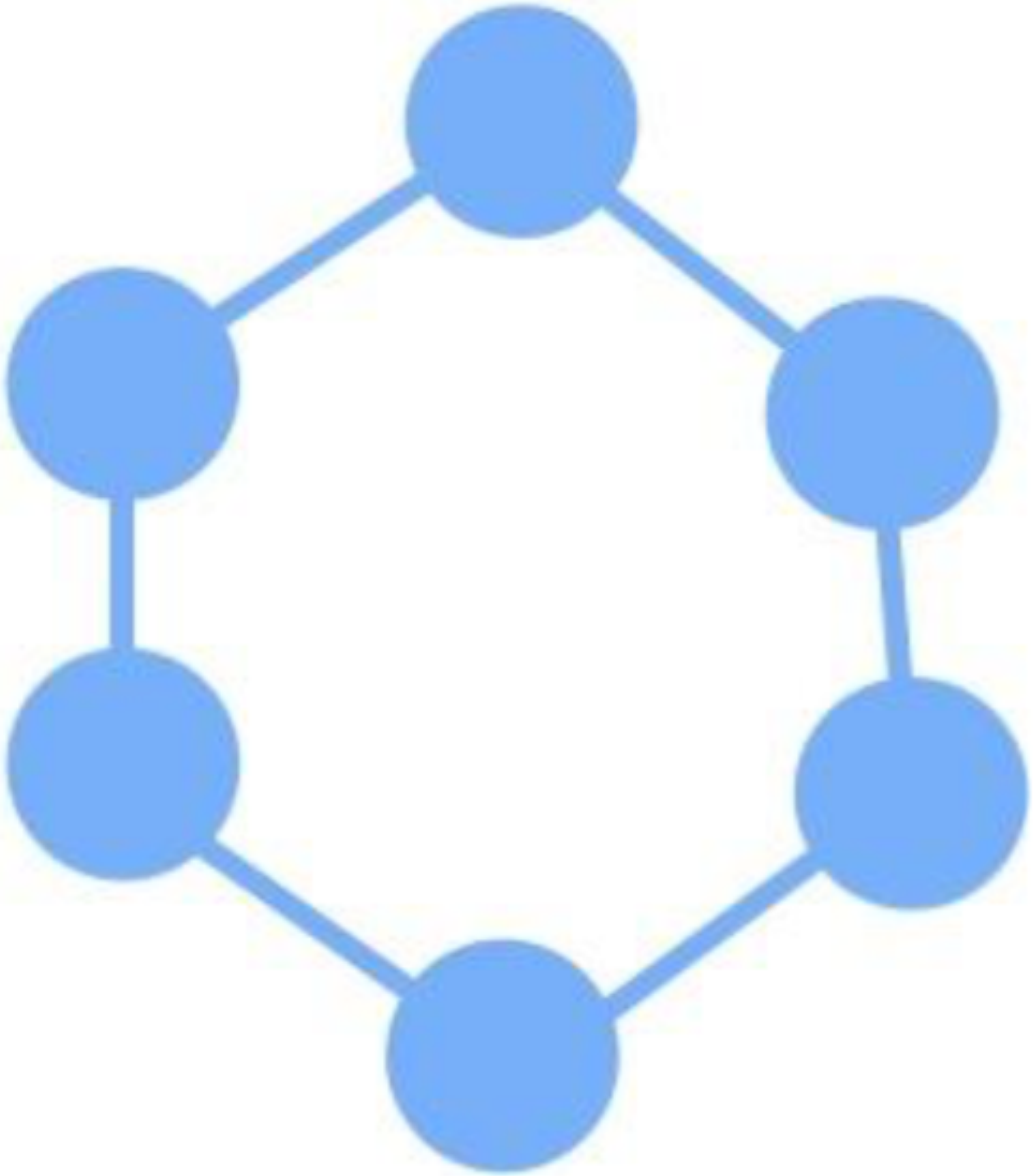




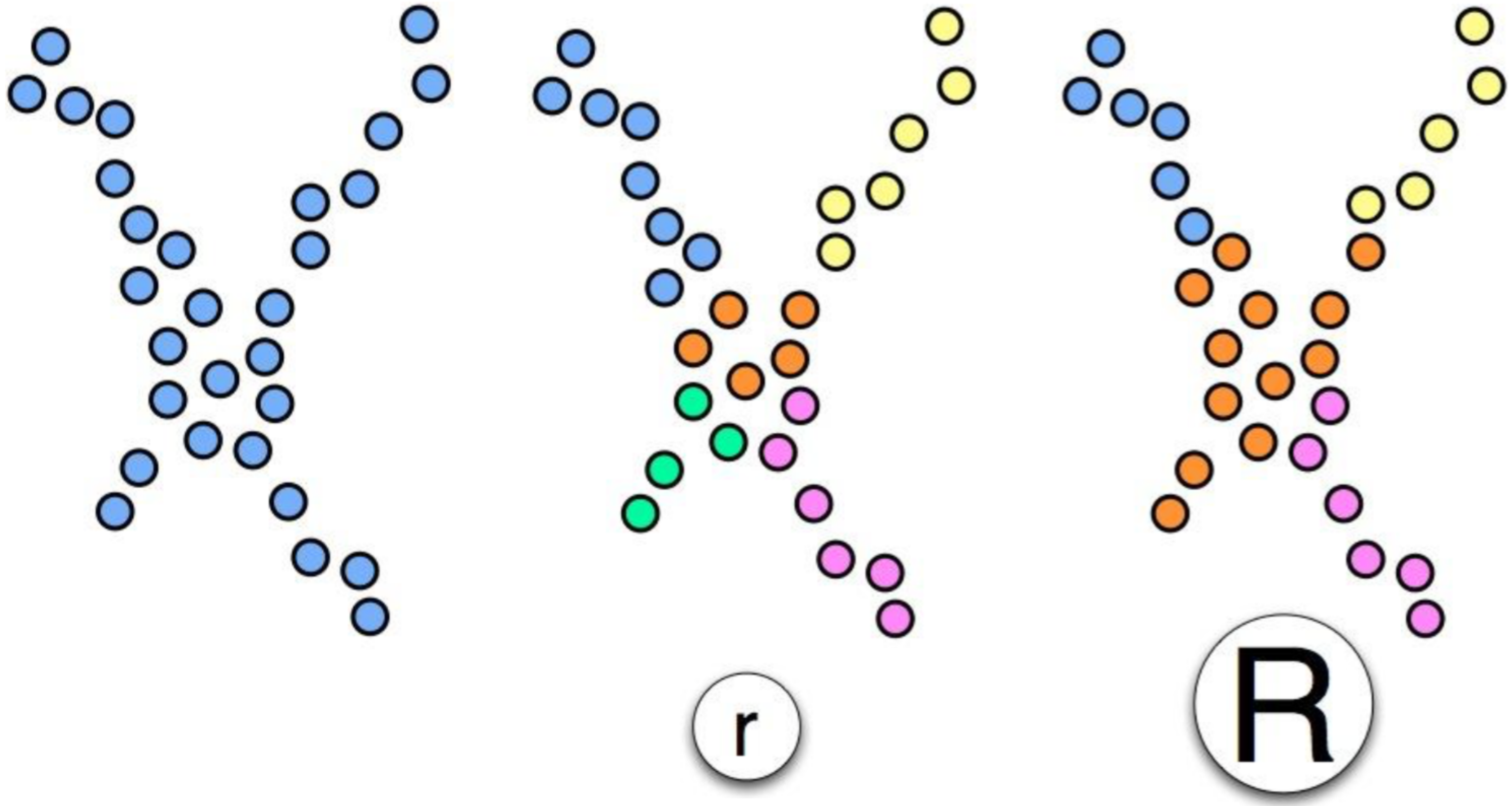
Key idea 2: deformation invariant

Key idea 3: compressed representation

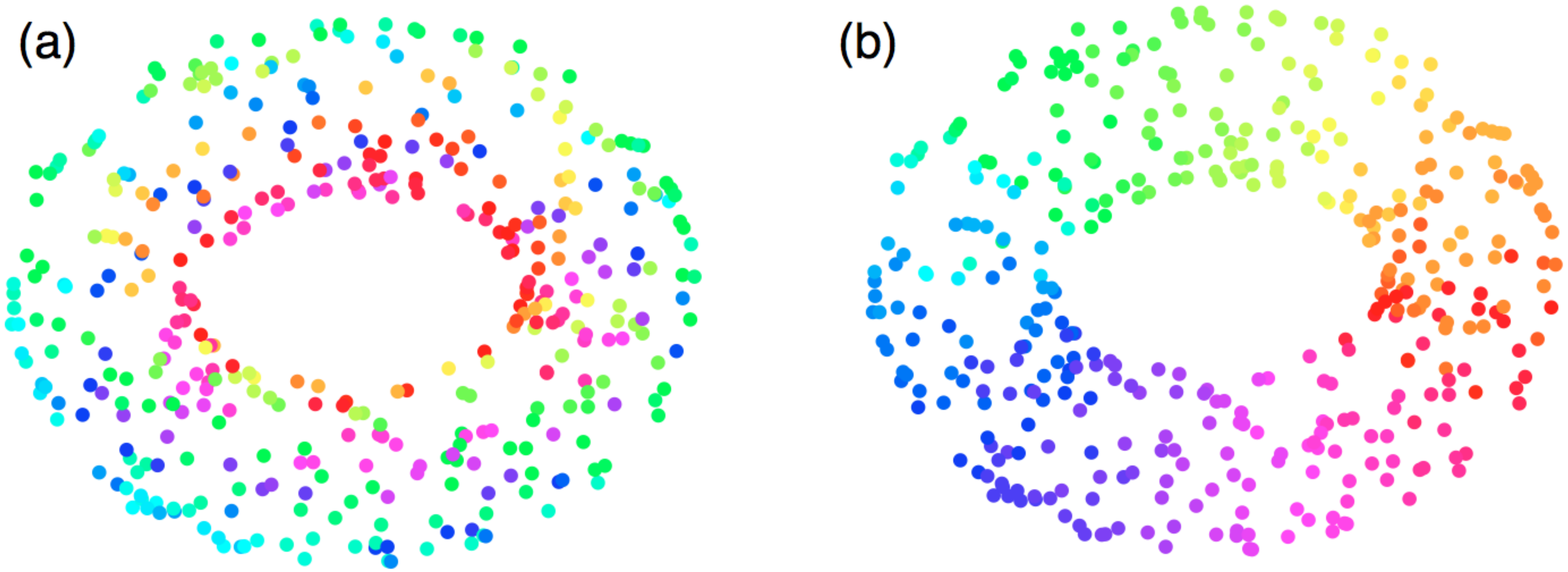




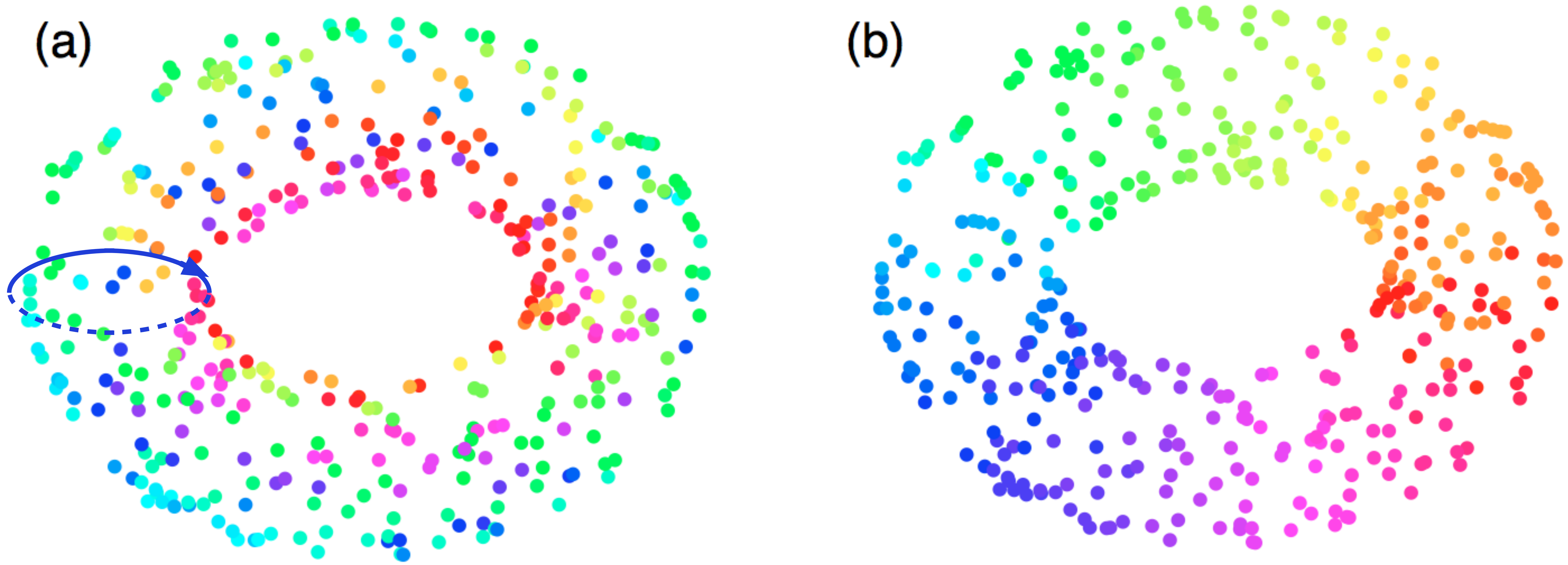
Inference: stratification learning



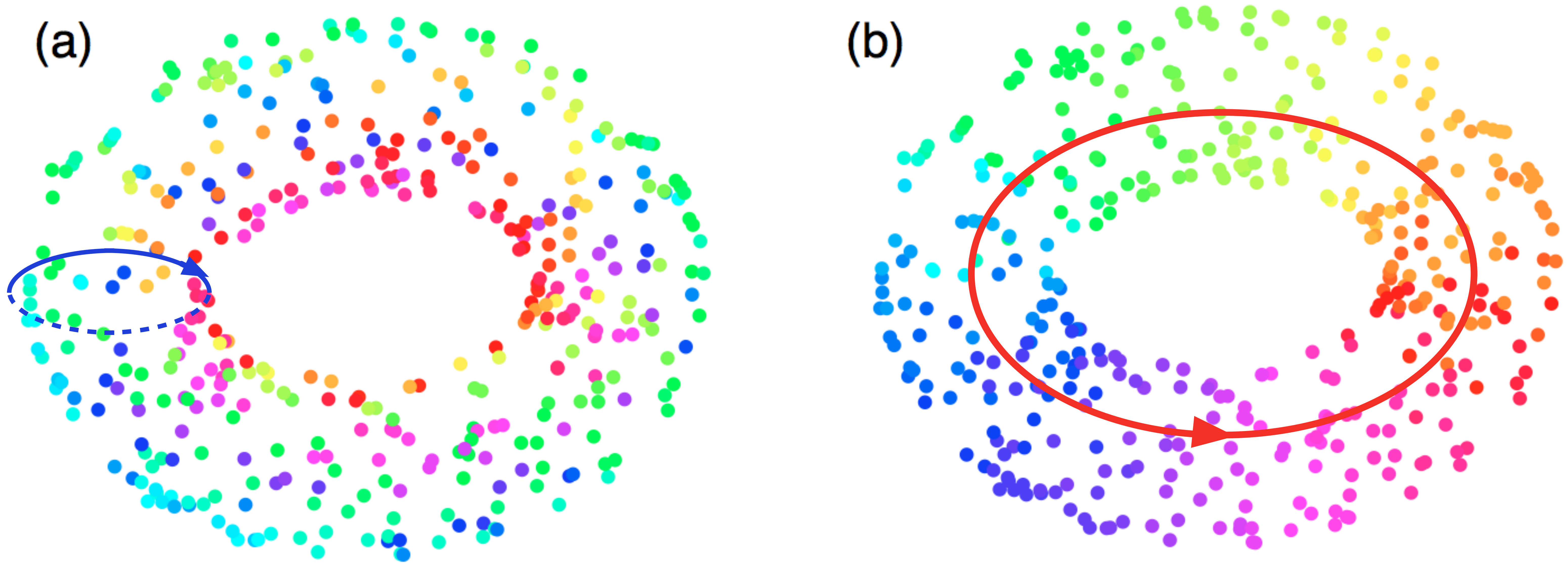
Inferring circular structures in high dimensions



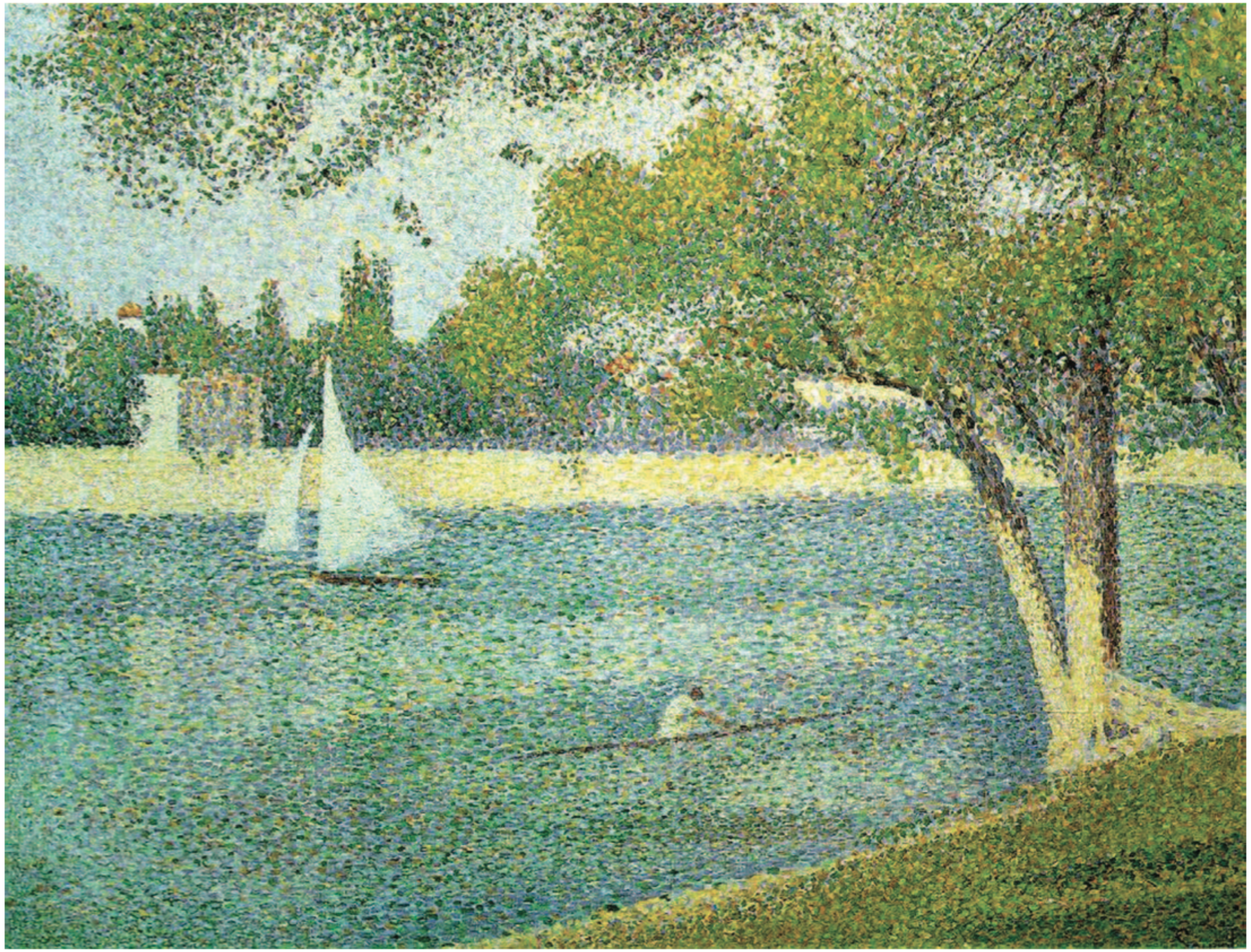
Inferring circular structures in high dimensions



Inferring circular structures in high dimensions

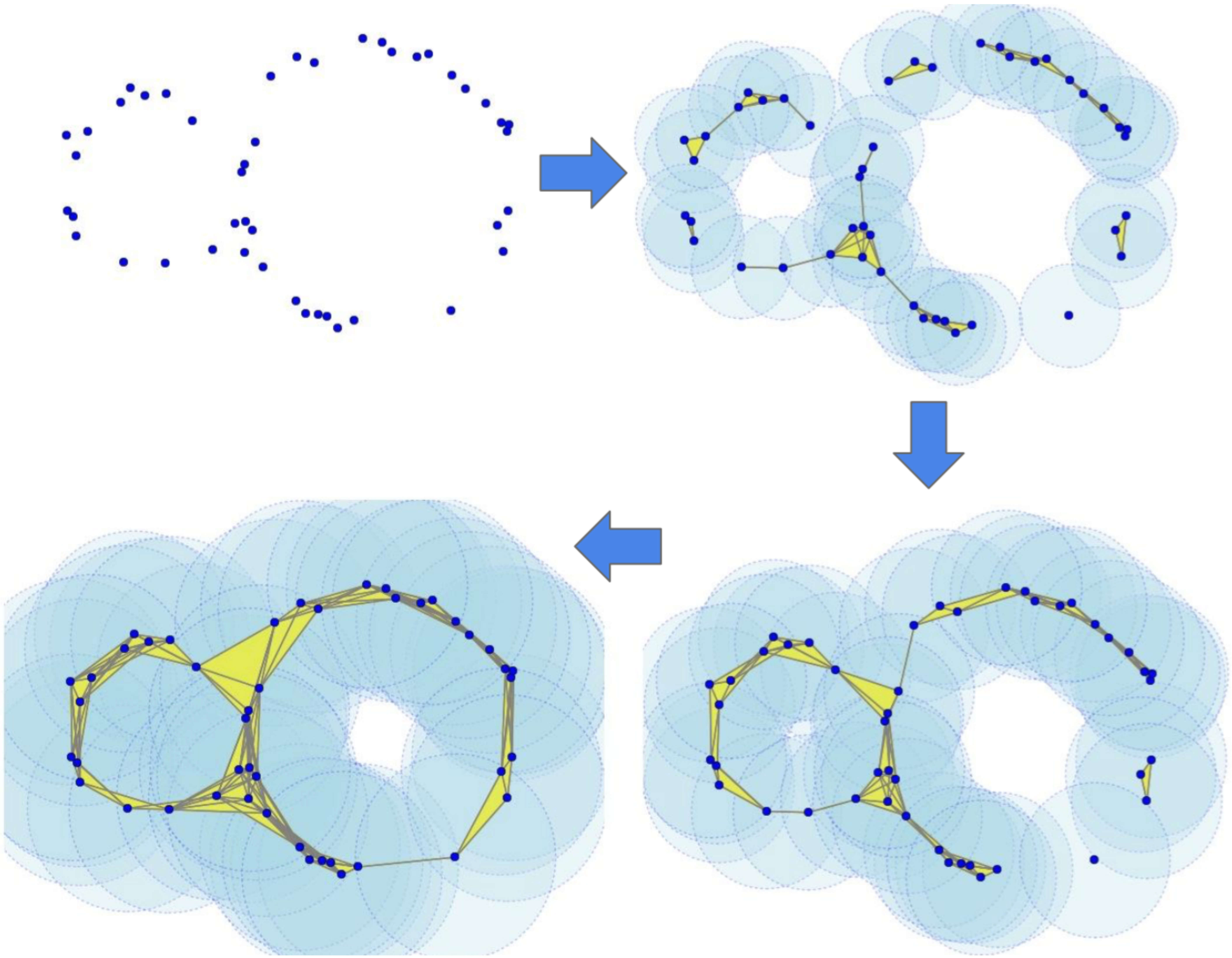


Persistent homology: an artistic view point



Persistent homology: inferring the continuous from the discrete.

Persistent homology: a multi-scale view of data



Persistent homology: quantifying the shape of data.

Persistent Homology

A really old joke...

Who thinks the coffee mug and a donut is the same? Topologist!

- Topologists care about topological structures of a space: connected components, tunnels, voids, etc.
- Formally, these correspond to the notions of homology.

A really old joke...

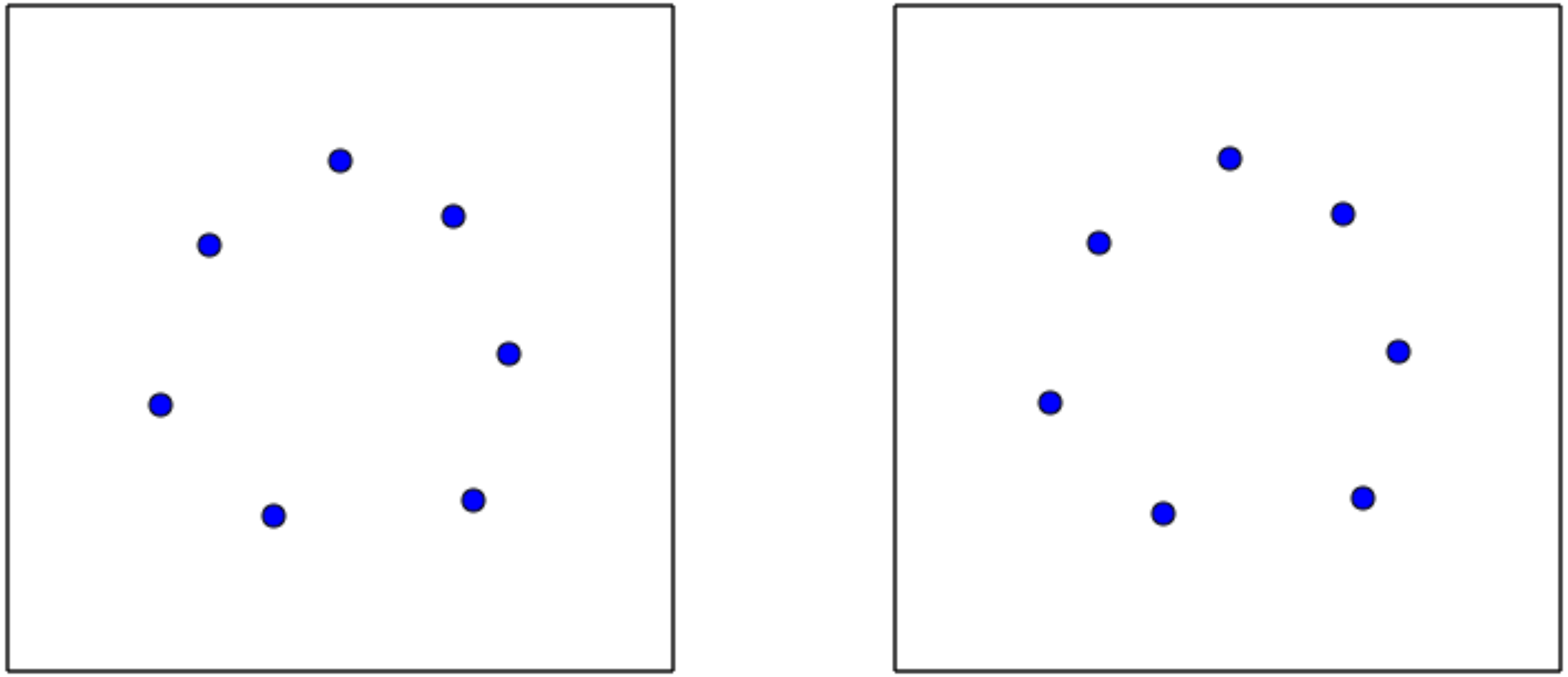
Who thinks the coffee mug and a donut is the same? Topologist!

- Topologists care about topological structures of a space: connected components, tunnels, voids, etc.
- Formally, these correspond to the notions of homology.

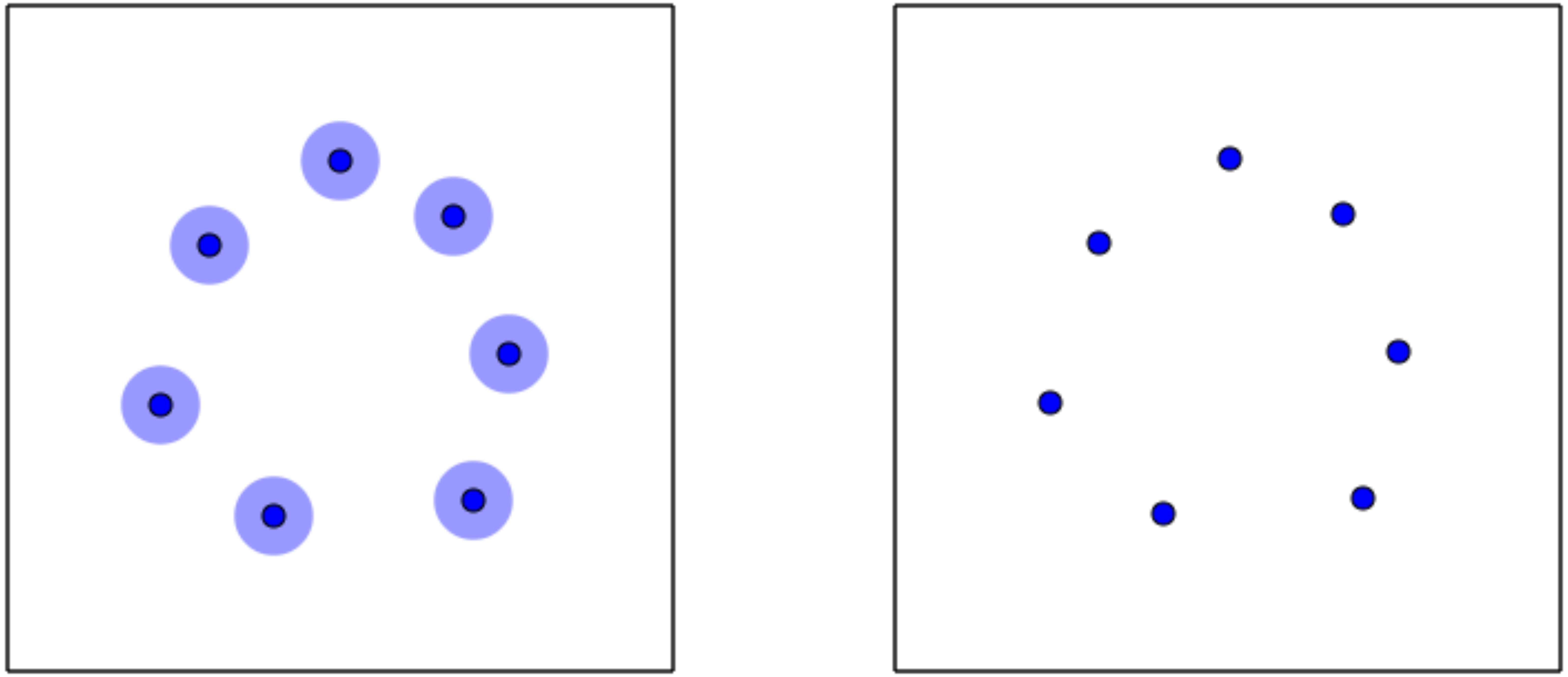
Persistent homology

- What are topological features? Homological features:
  - Dim 0 - Connected Components
  - Dim 1 - Tunnels / Loops
  - Dim 2 - Voids
- How to compute them (in a nutshell)?
  - Begin with a point cloud
  - Grow balls of diameter t around each point
  - Track features of the union of balls as t increases

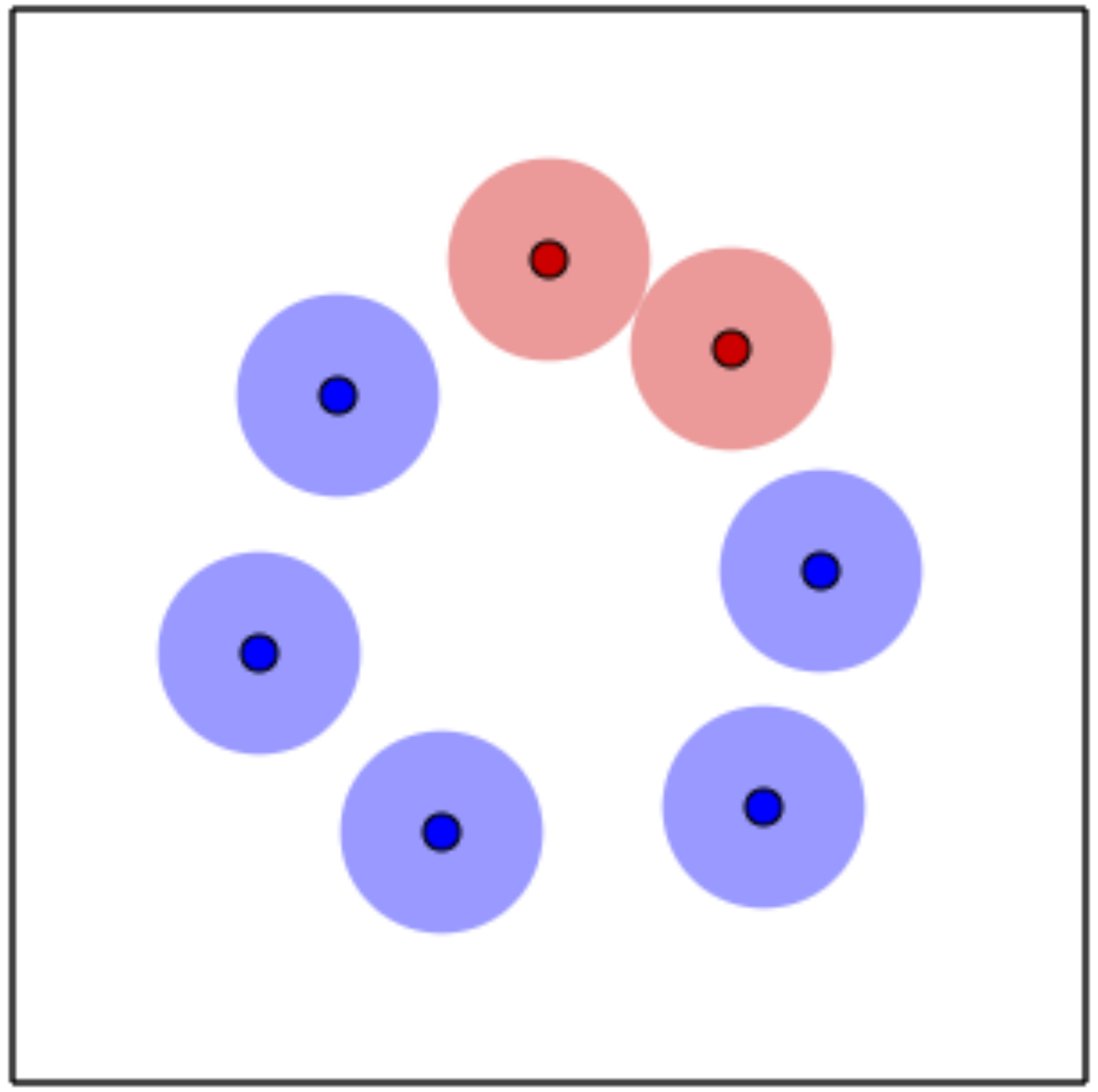
Persistent homology

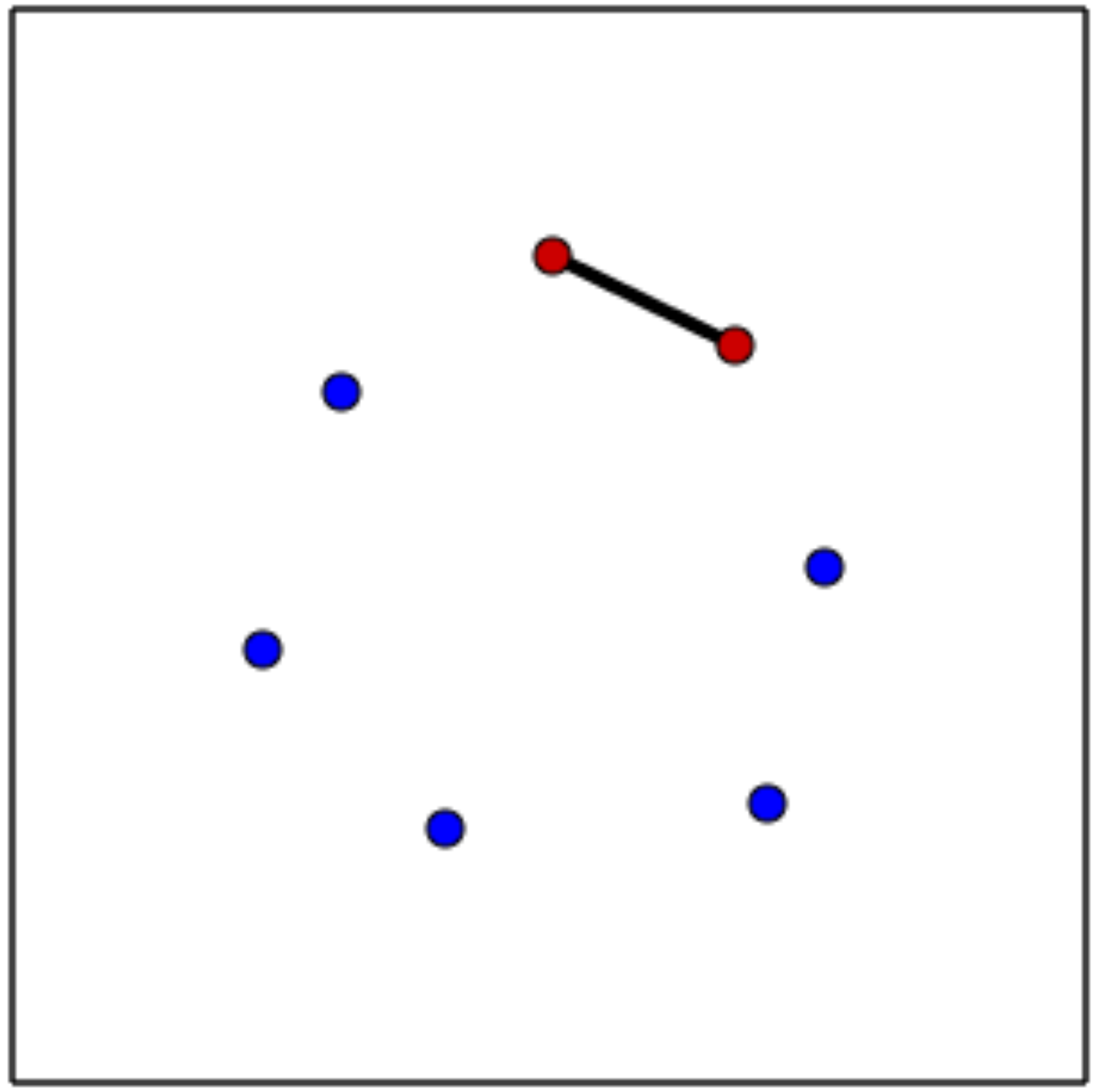


Persistent homology

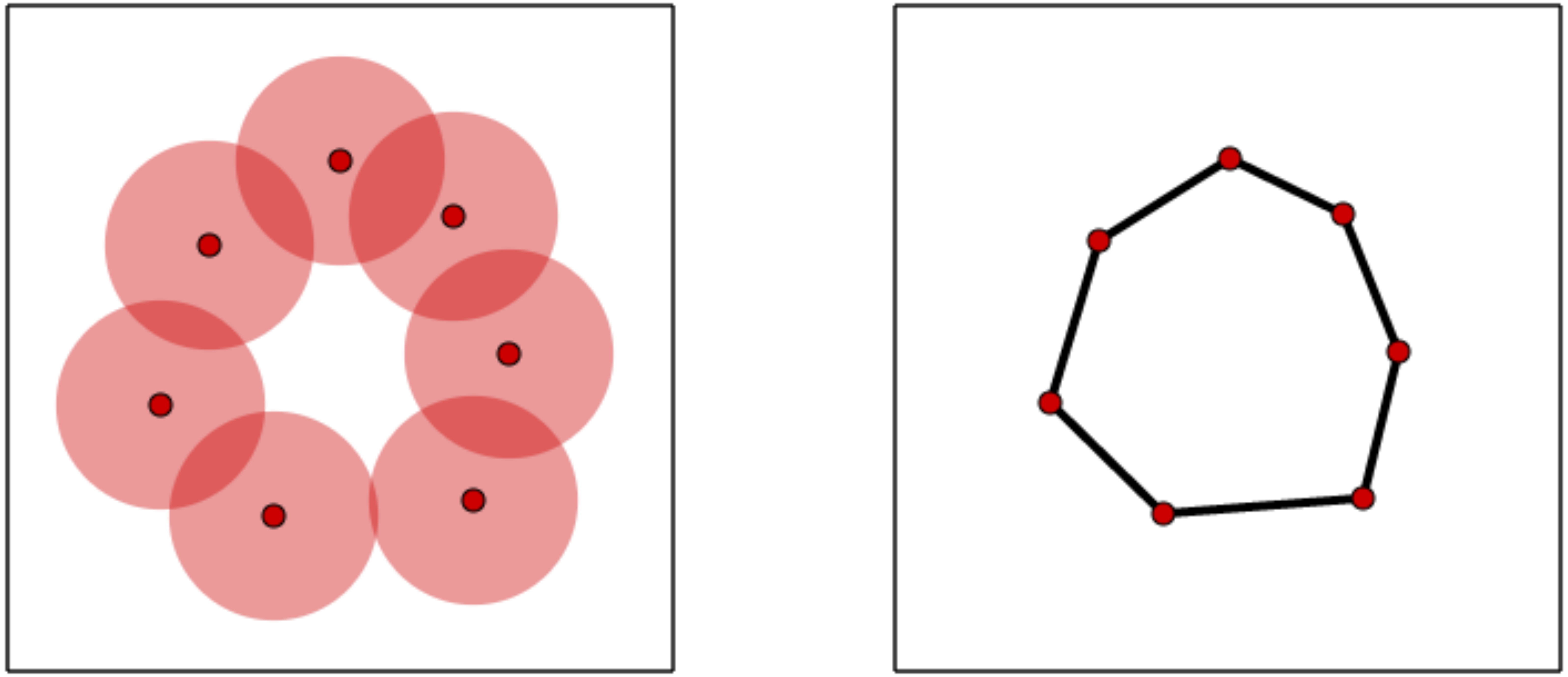


Persistent homology

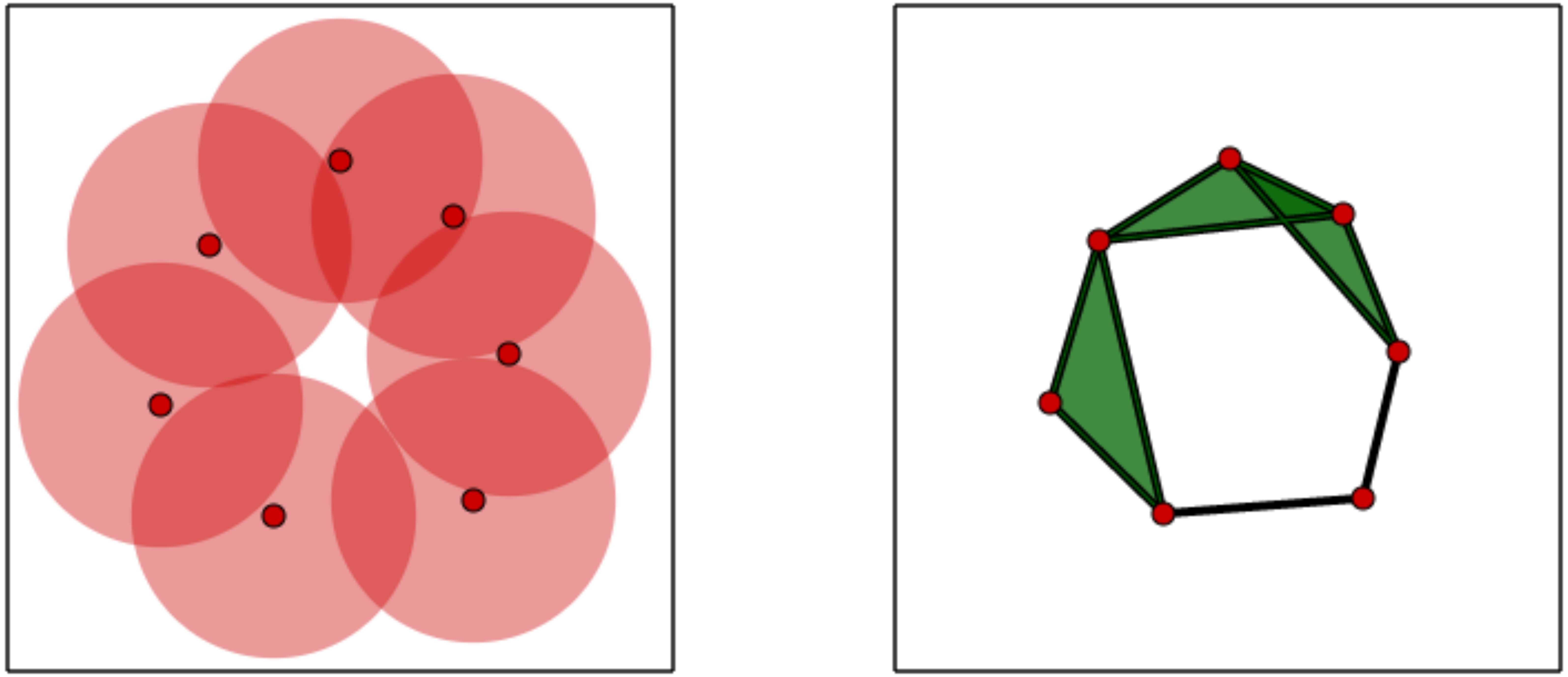




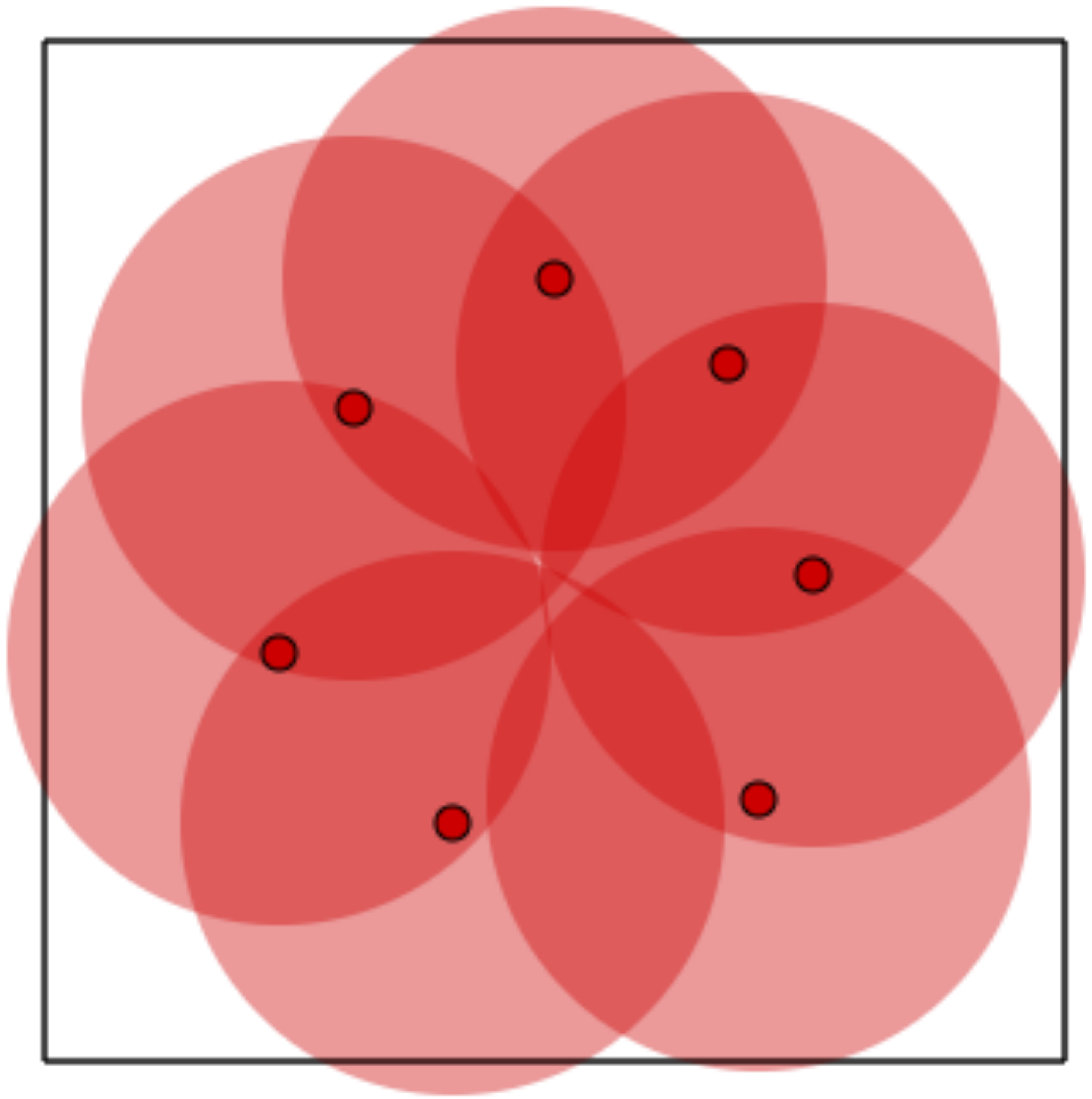
Persistent homology

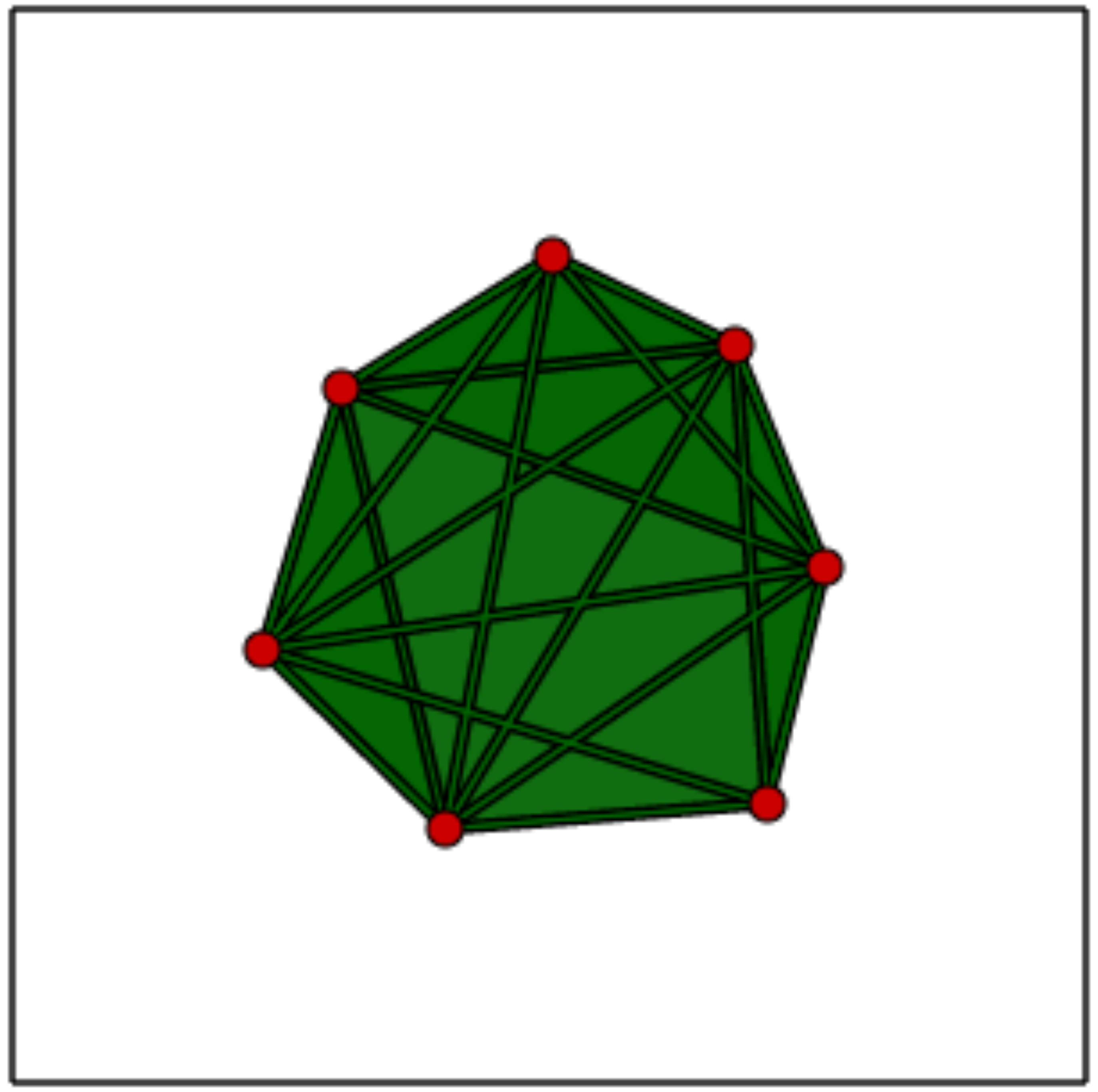


Persistent homology



Persistent homology





Persistence diagrams

Homological features encoded as barcodes or persistent diagrams

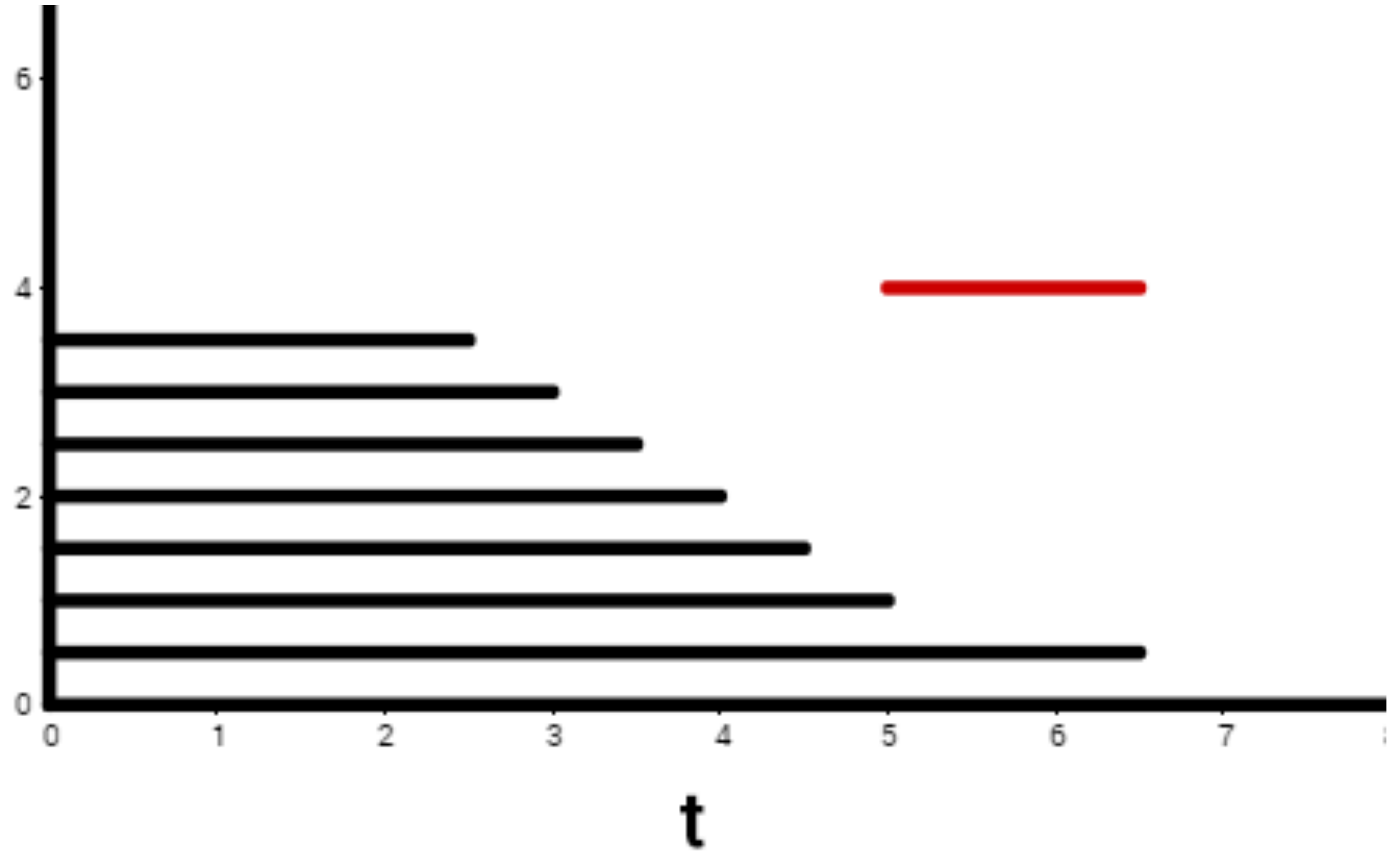


Figure: Barcode

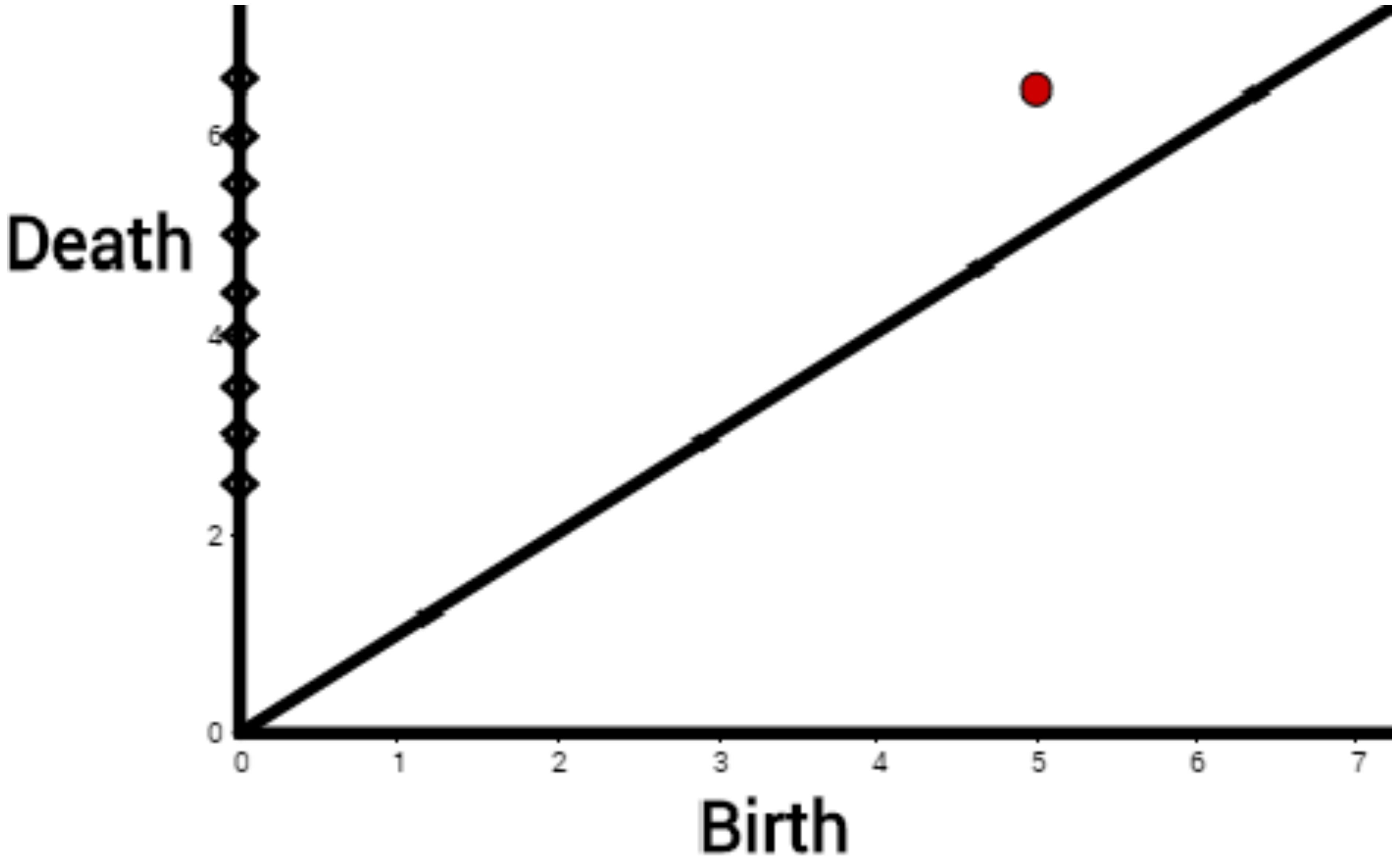
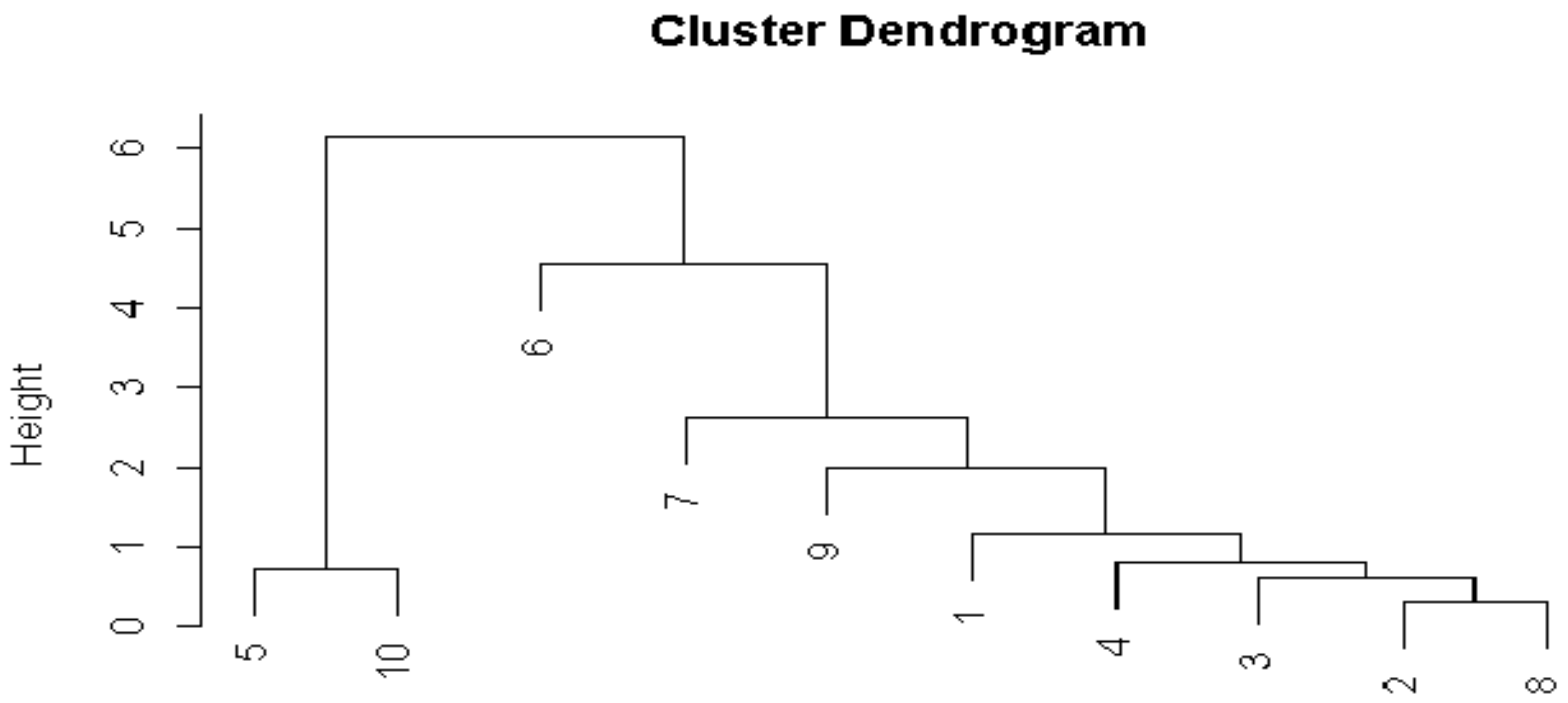


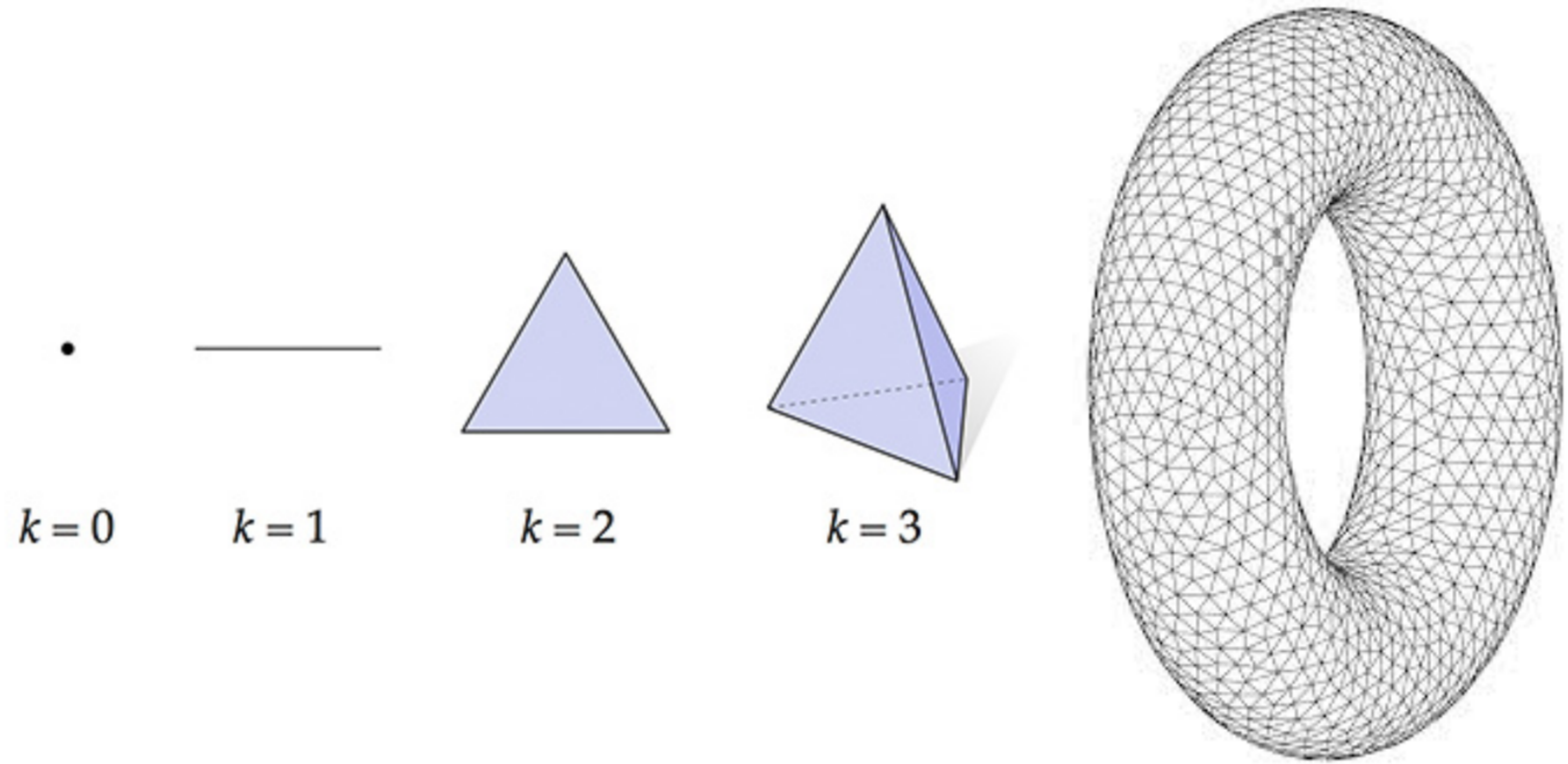
Figure: Persistence Diagram

Interpretation of connected components

Dim 0 features: hierarchical clustering



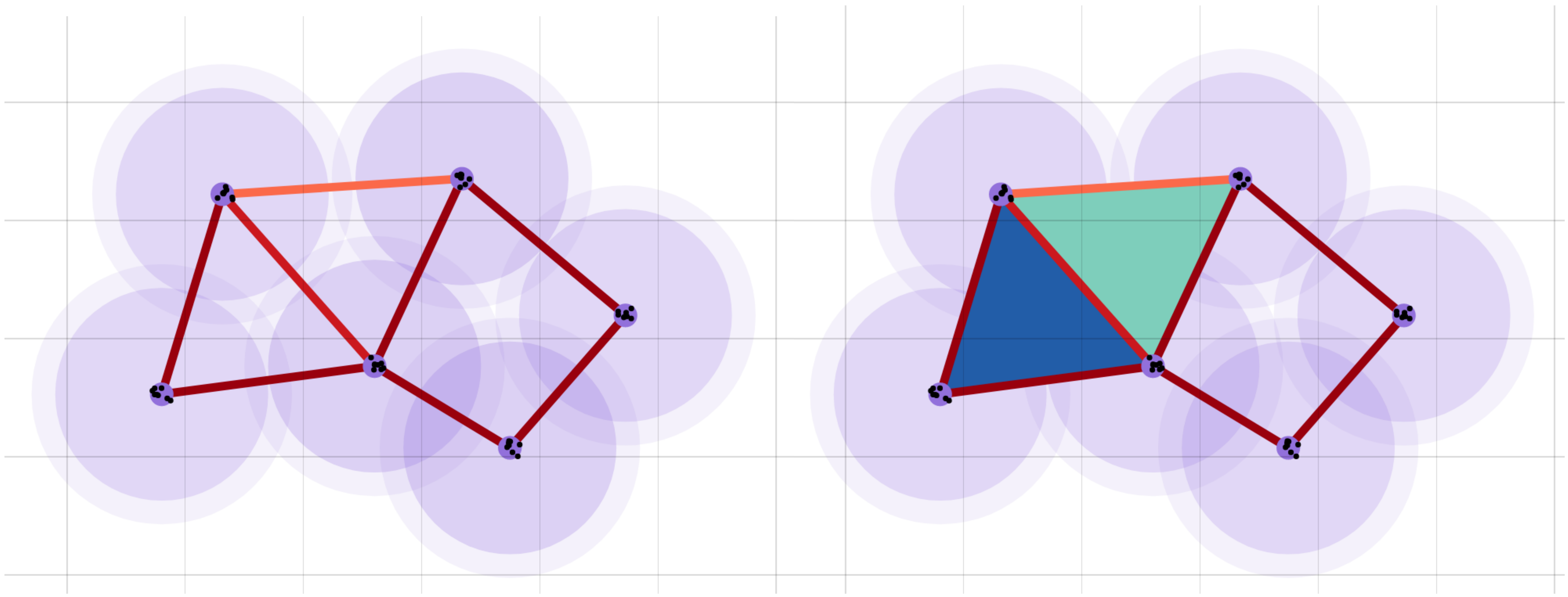
Simplicial complex



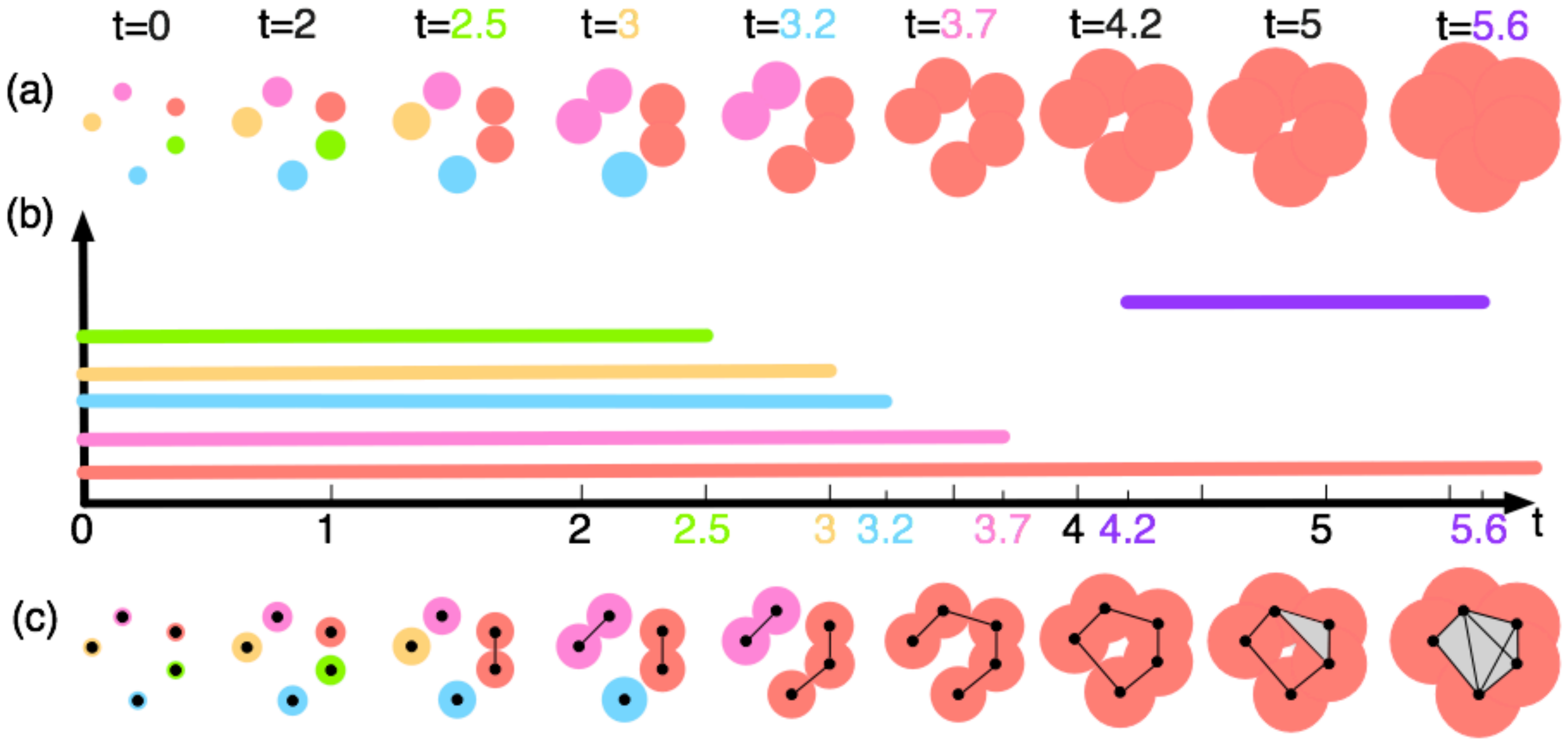
Different types of simplicial complexes

- Abstract simplicial complex
- Čech complex
- Vietoris-Rips complex
- Delaunay triangulation (related to Voronoi diagram)
- Alpha complex
- Sparsified versions:
  - Witness complex
  - Graph induced complex

Čech complex vs. Vietoris-Rips complex



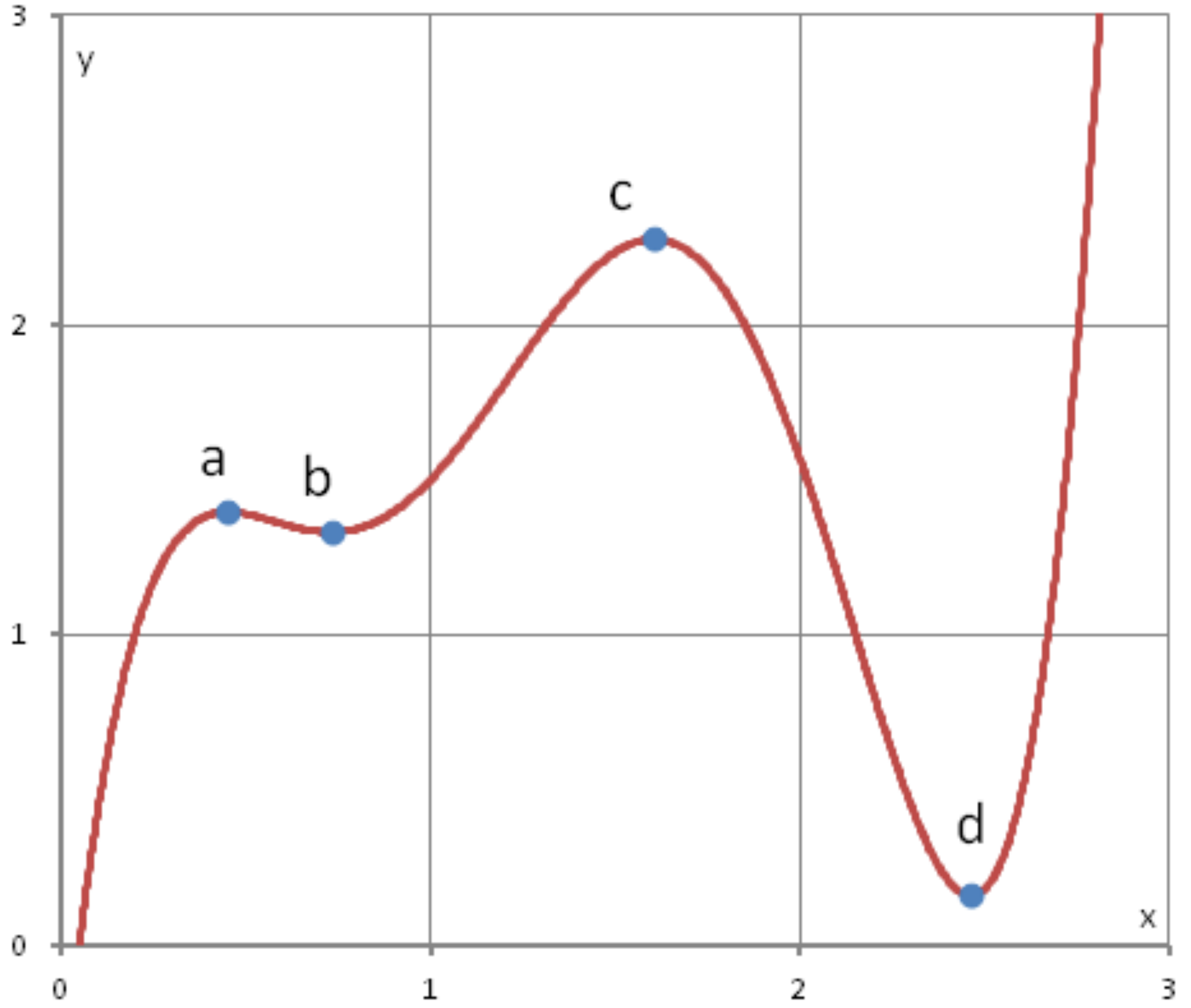
Persistent homology with Čech complex



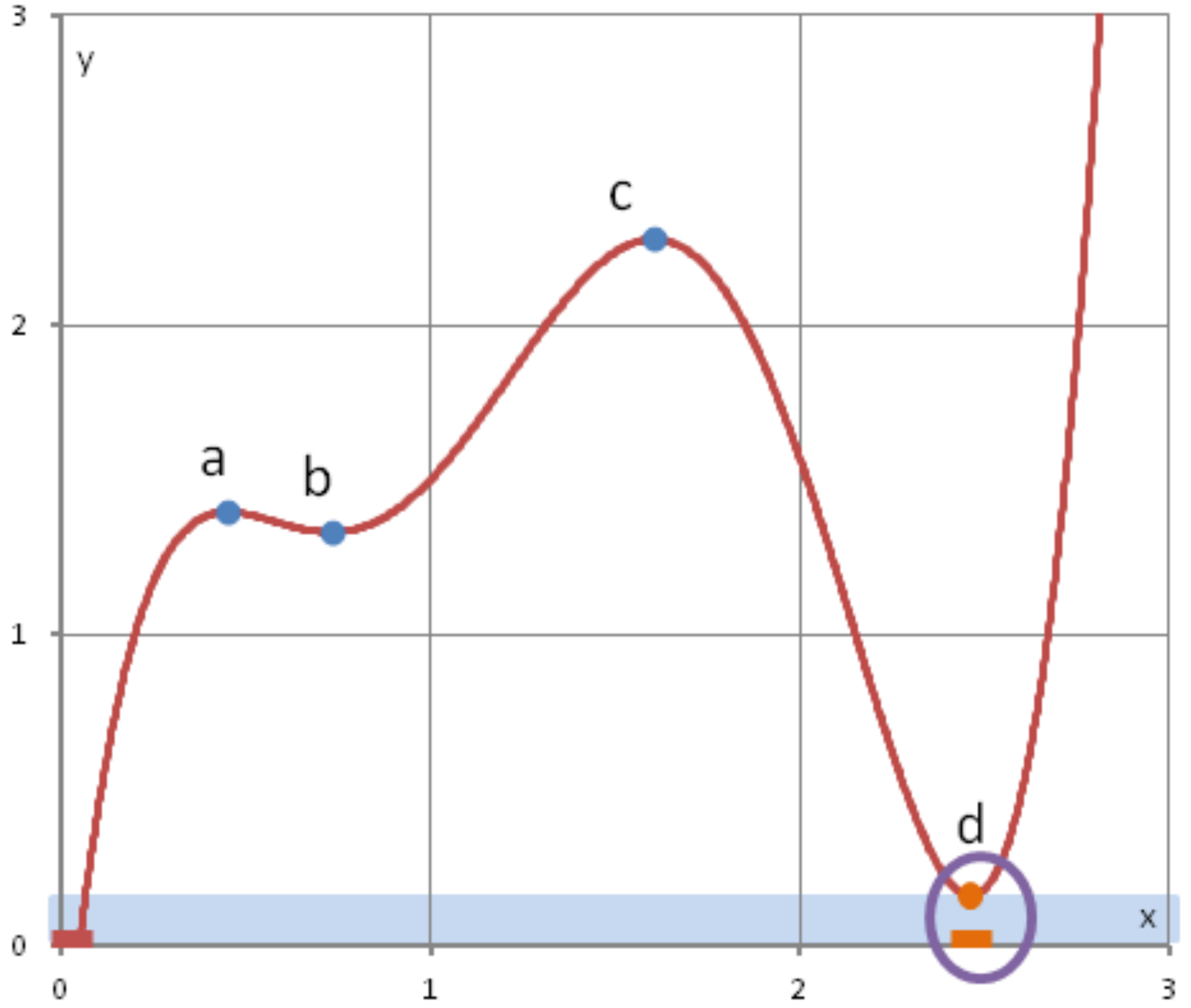
To apply persistent homology

- A filtration of spaces with maps between them
- A scale parameter

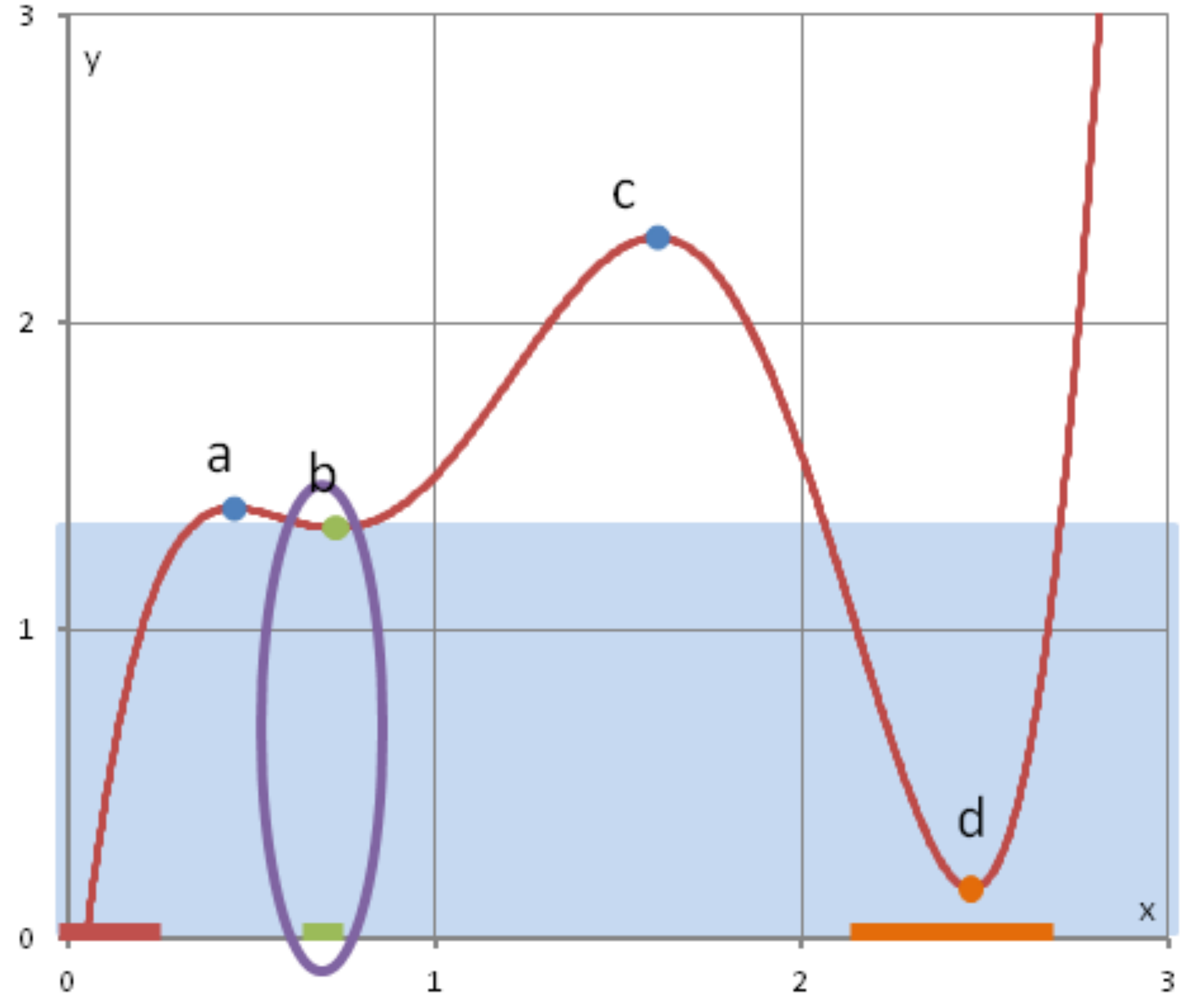
Sublevel set filtration



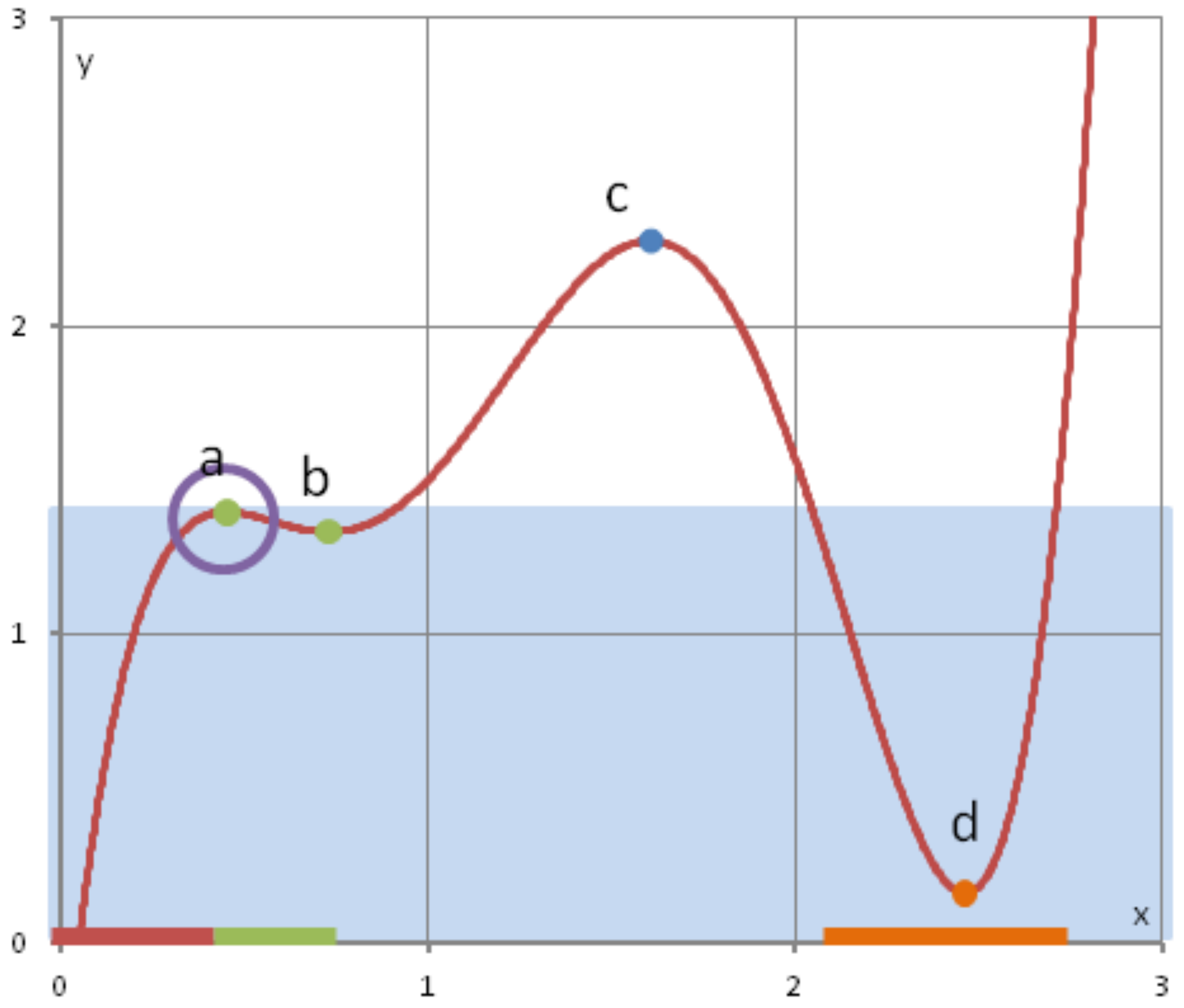
Sublevel set filtration



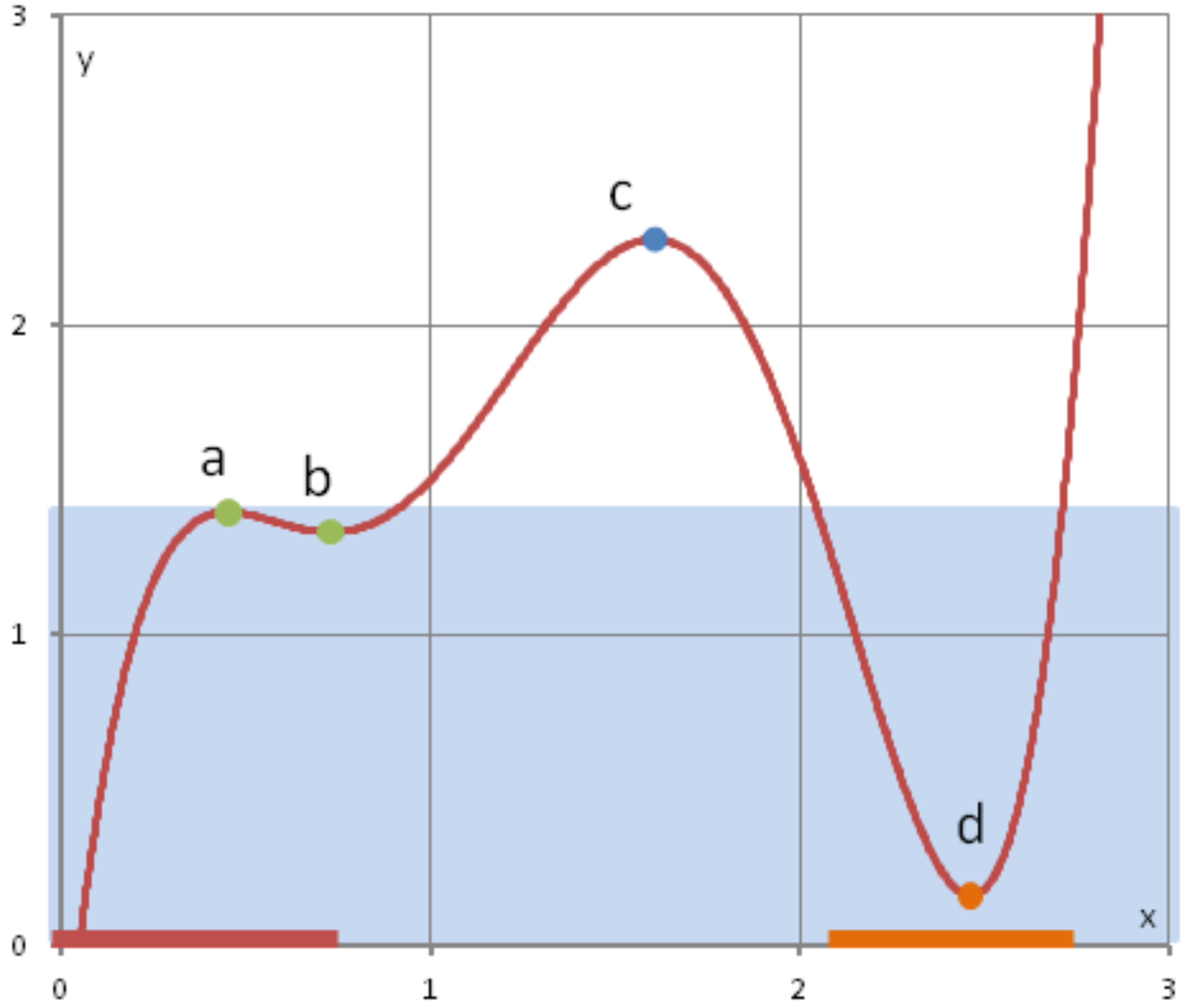
Sublevel set filtration



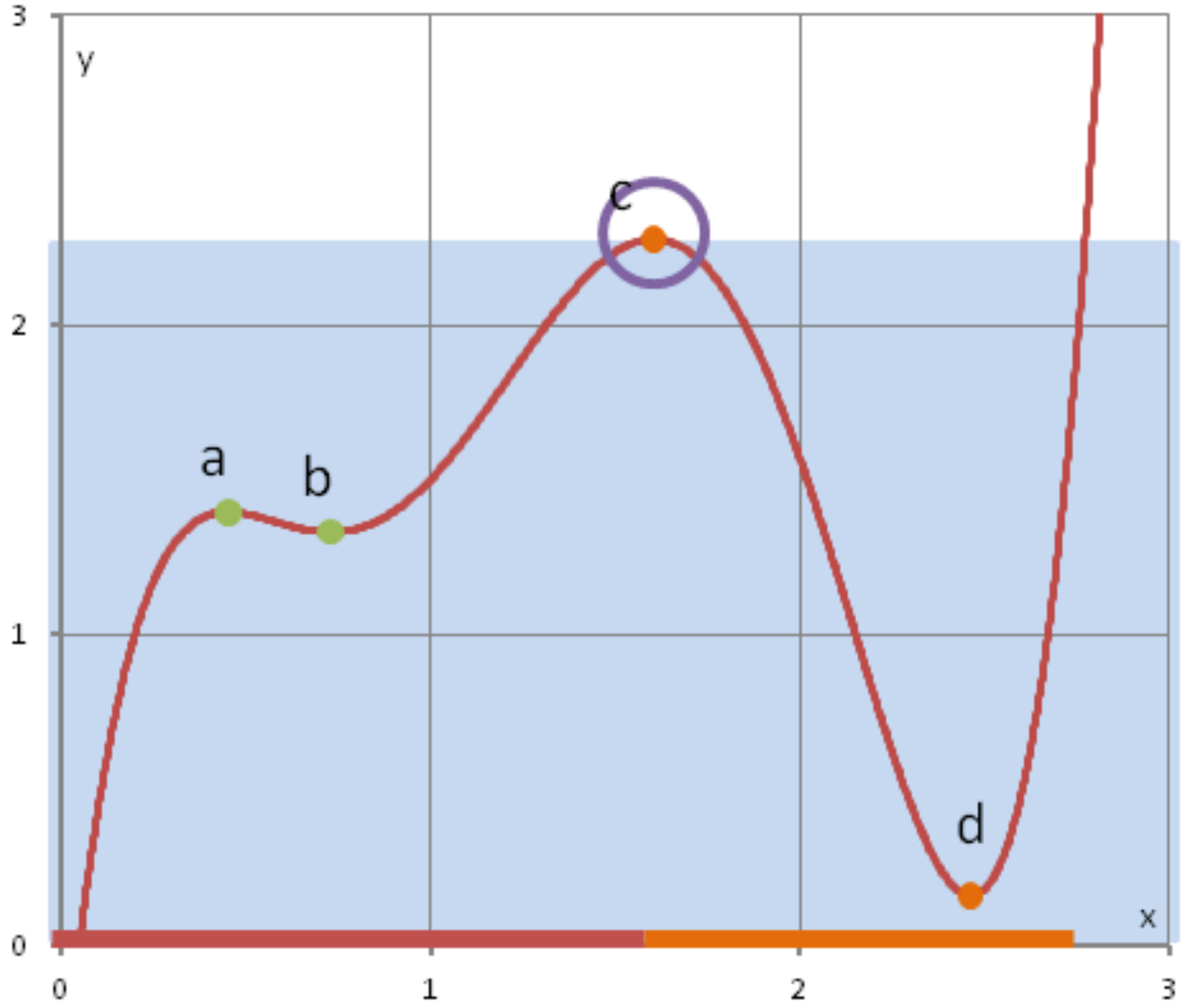
Sublevel set filtration



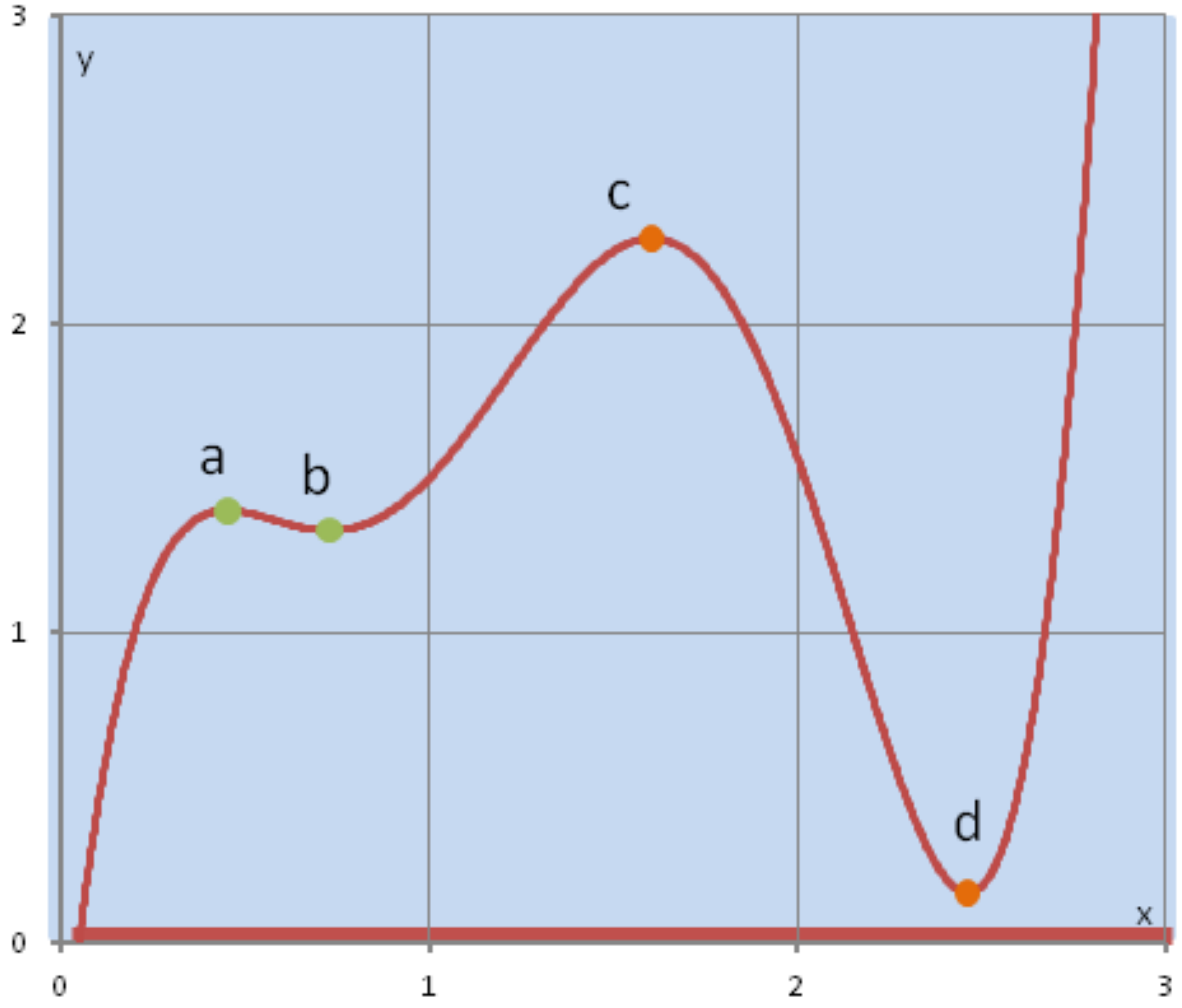
Sublevel set filtration



Sublevel set filtration



Sublevel set filtration

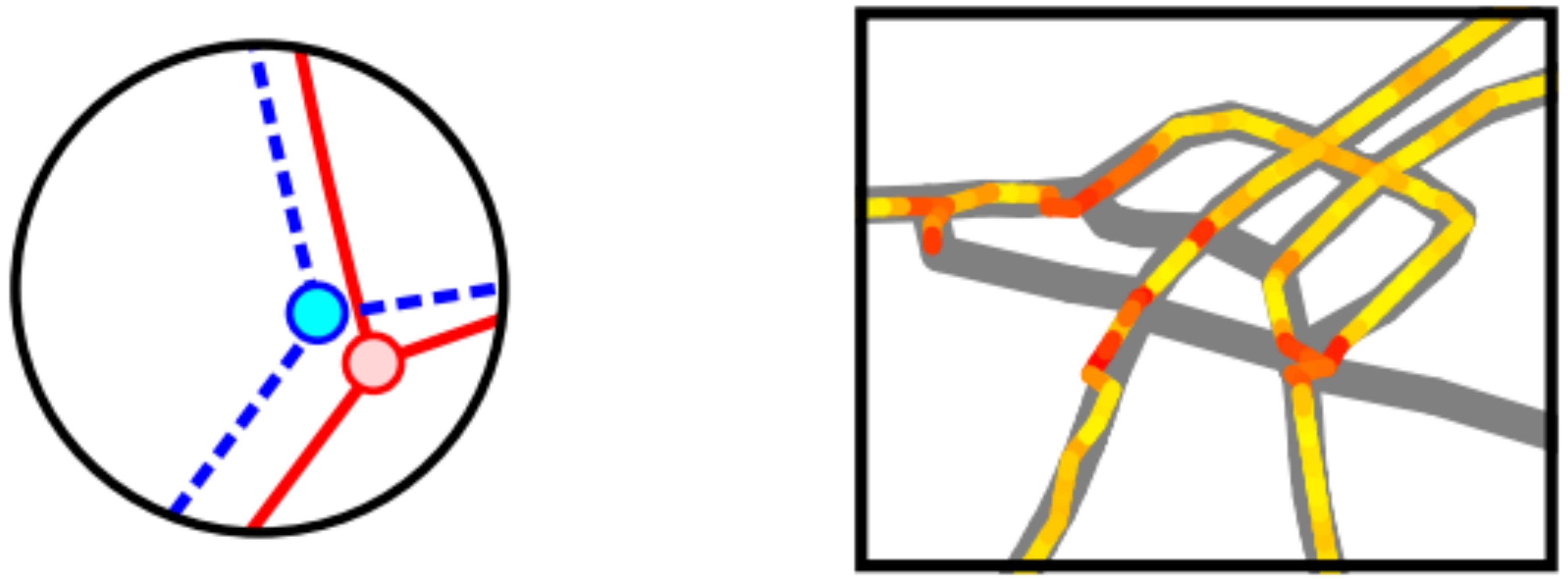


Distance between persistence diagrams

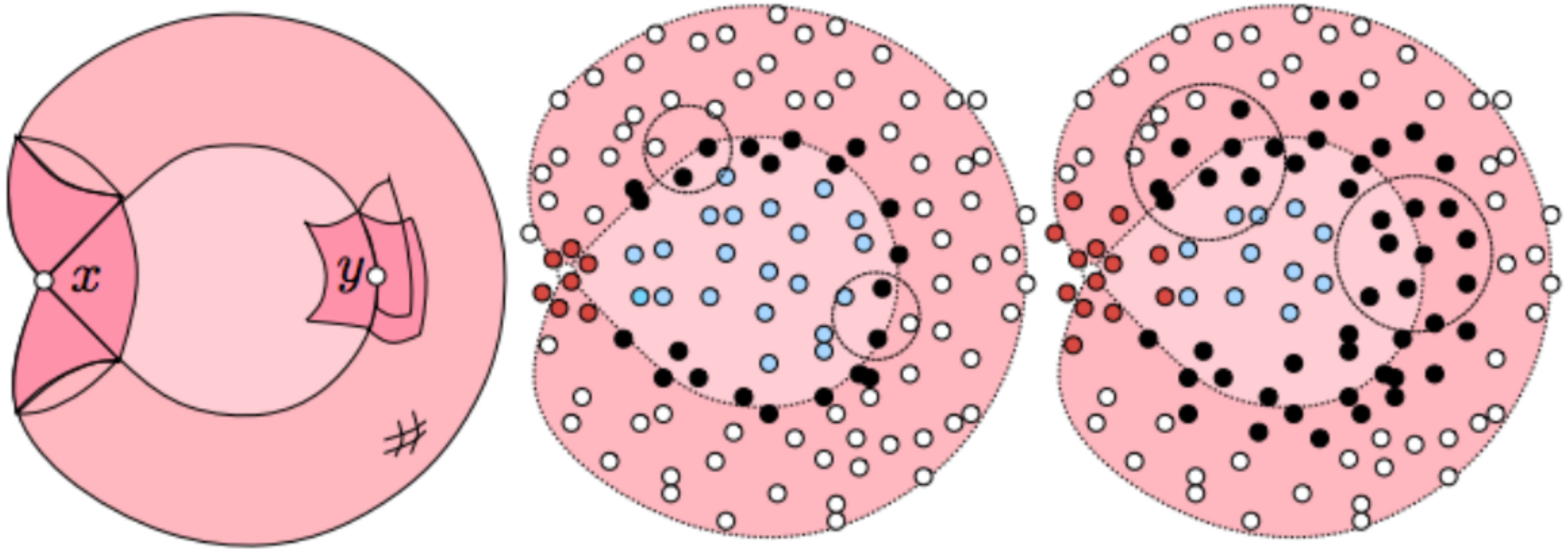
- Stability of persistence diagrams
- Bottleneck distance
- Wasserstein distance

Beyond Persistent Homology

Persistent local homology: applications



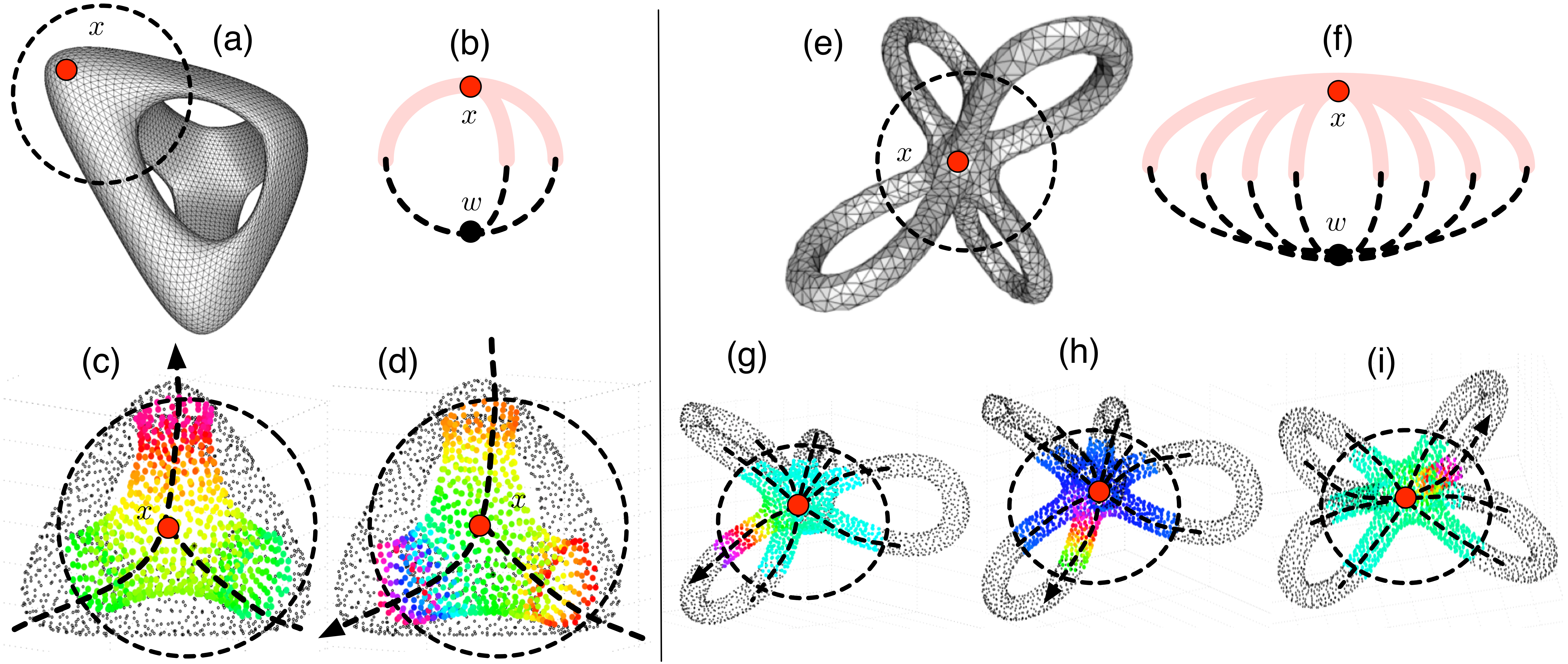
(a) Local Structures (b) Signature restricted to $\mathbb{X}$



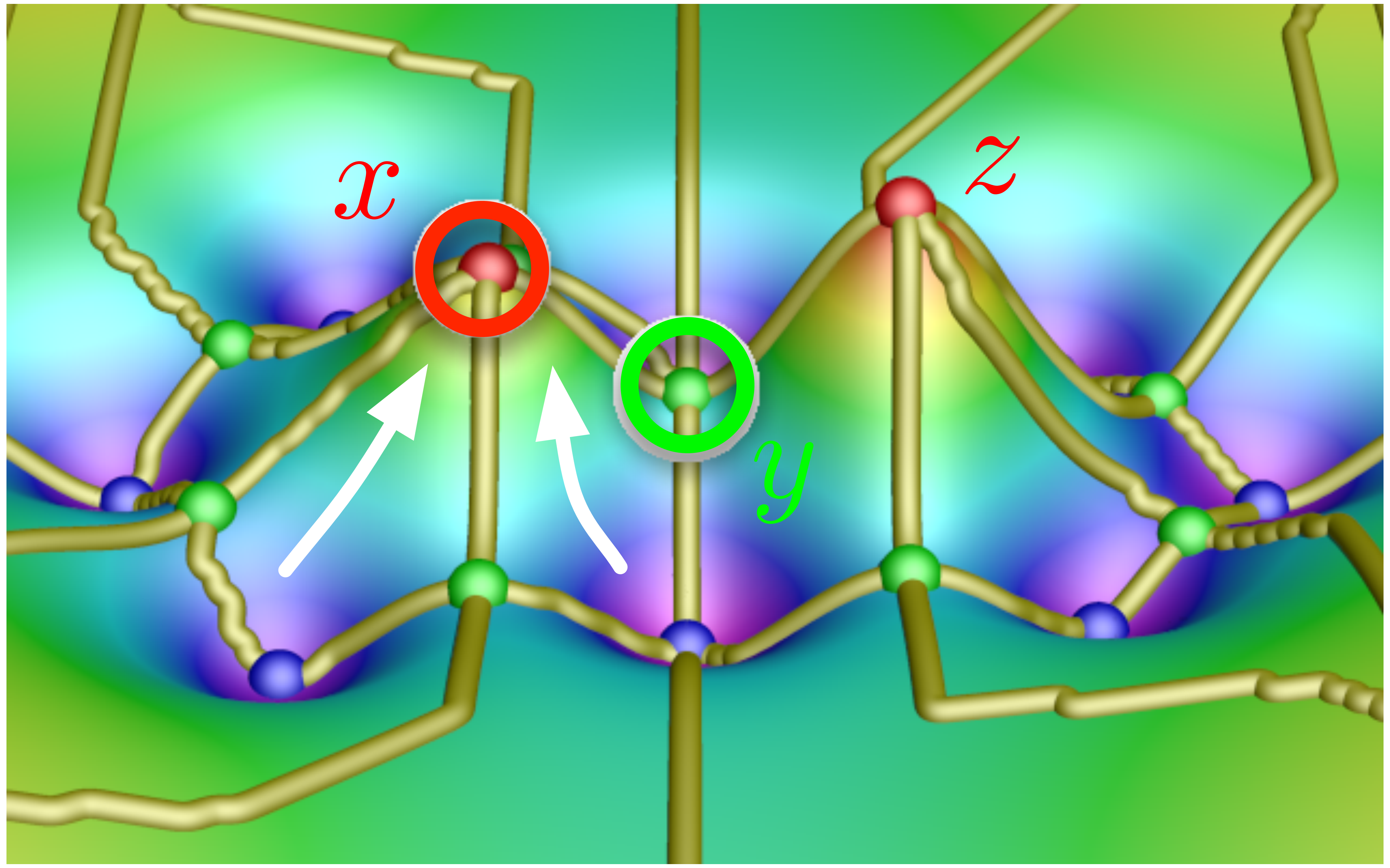
Road network comparison; stratification learning...

TDA and dimensionality reduction (DR)

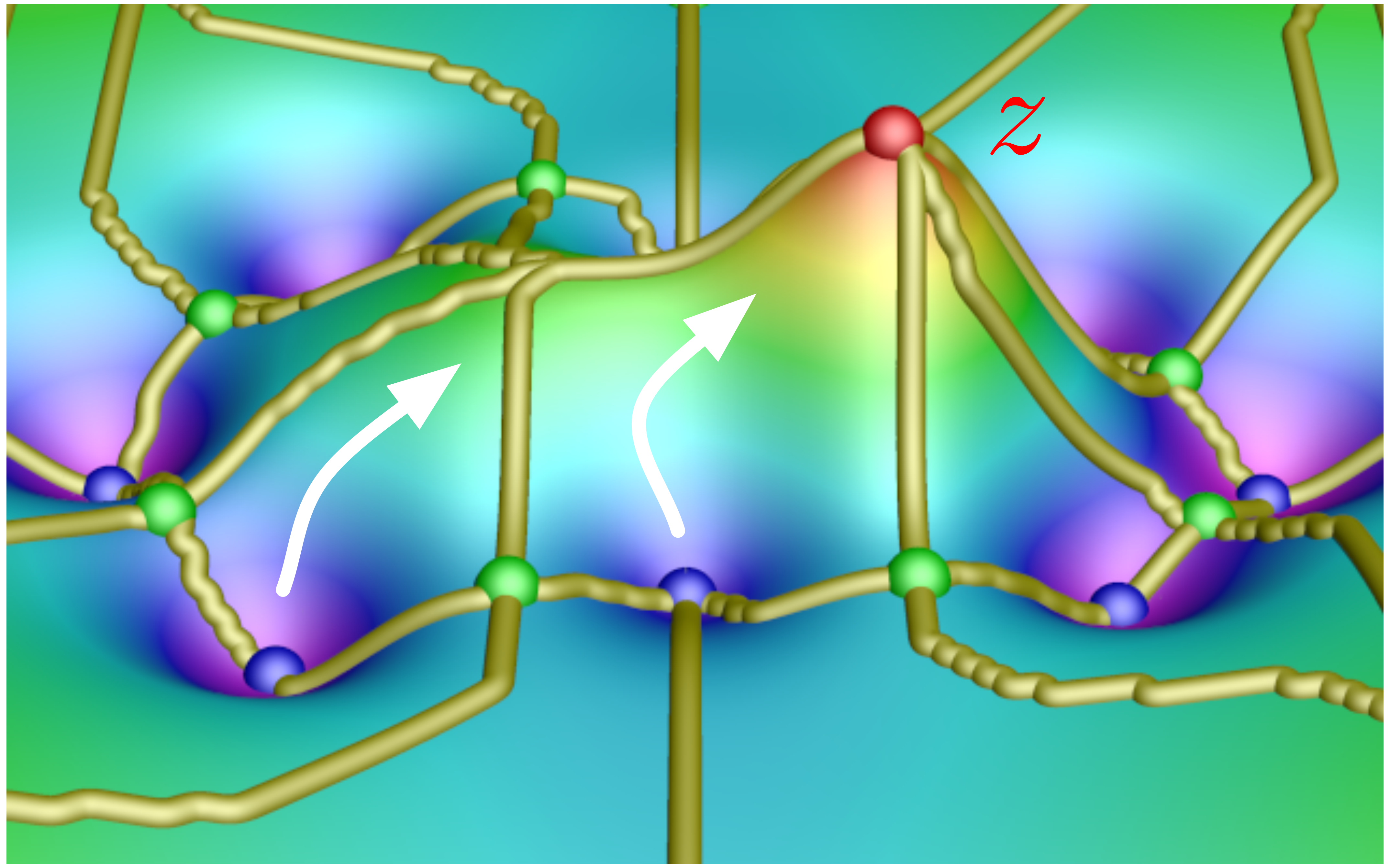
Detecting circular and branching structures for DR



Persistence simplification of Morse-Smale complex



(a)



(b)

Discussions

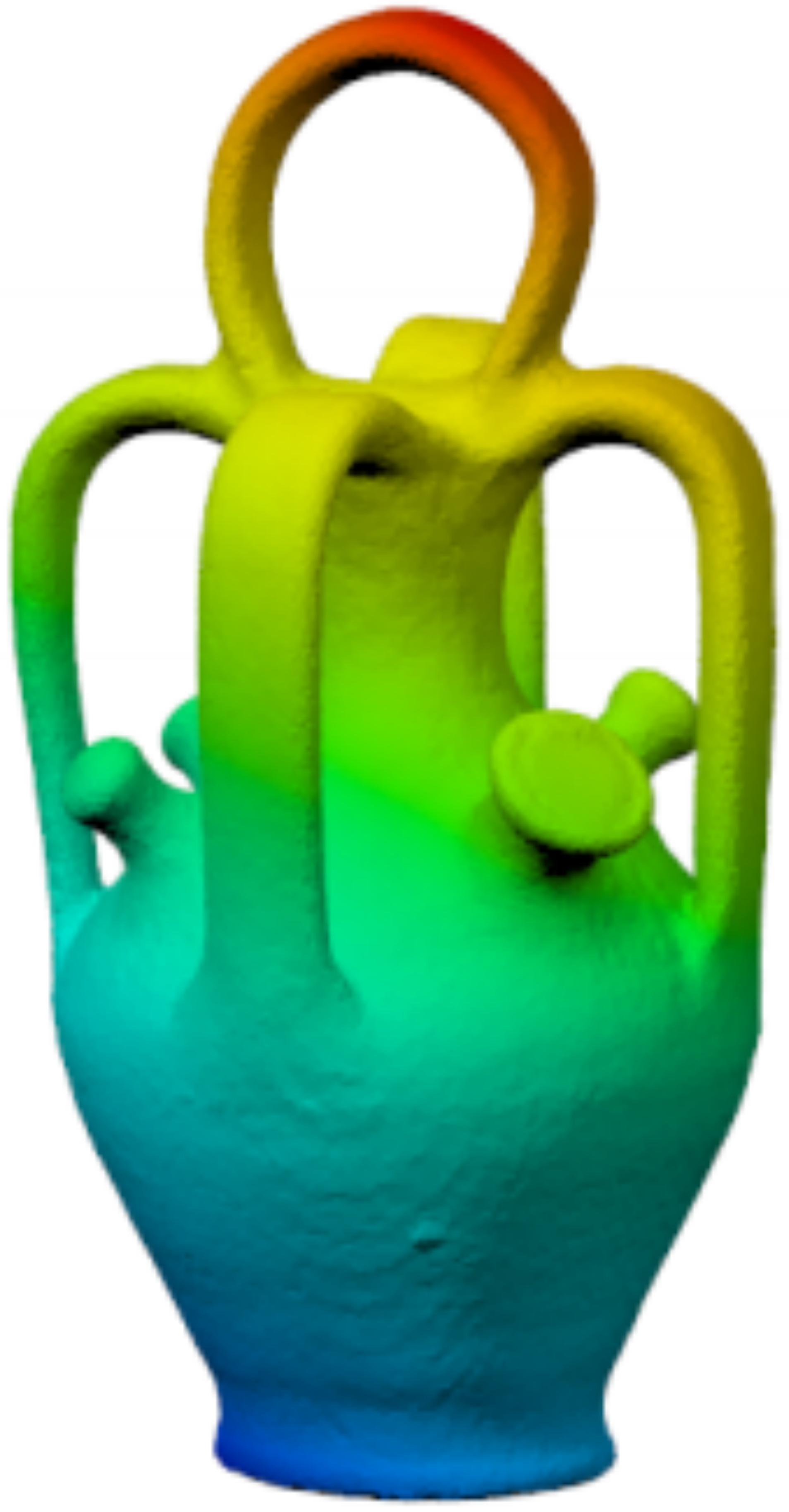
Research directions in TDA and visualization

- Reeb graphs, Reeb Spaces, and Mappers.
- Topological analysis and visualization of multivariate data.
- New opportunities for vector field topology.
- Category theory: theory and applications.
- Multidimensional persistent homology.
- Singularity theory and fiber topology in multivariate data analysis.
- Scalable computation.
- Software tools and libraries.

[Dagstuhl Seminar 17292 Report 2017]

More case studies...

- Study of low-dimensional data inspires techniques for high-dimensional data



Handles of 3D models

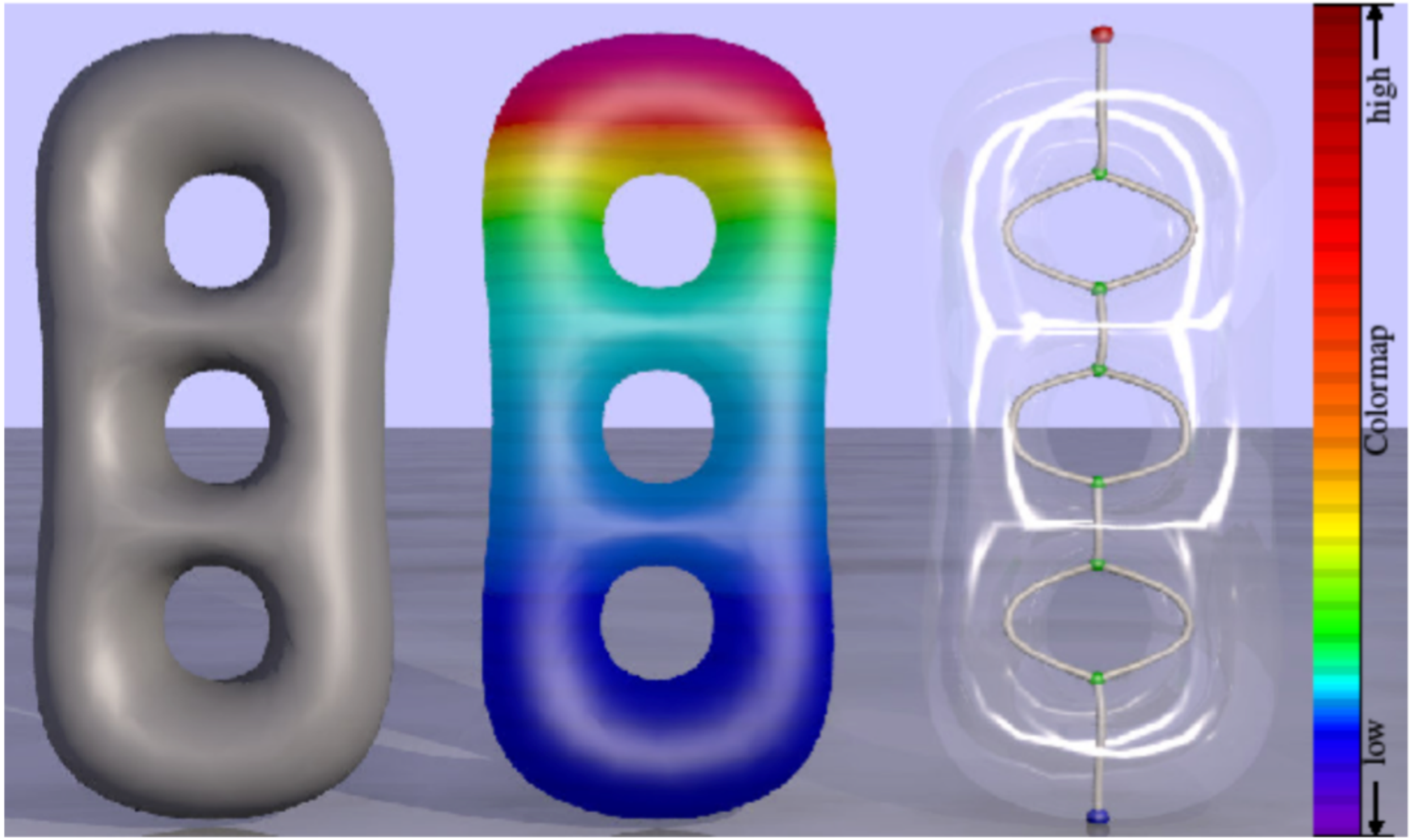
[DeyFanWang2013]

http://web.cse.ohio-state.edu/~wang.1016/papers/sig2013-loops.pdf

Review: Reeb Graph

A generalization of
contour tree

Graph obtained by continuos contraction of all the contours in a scalar field, where each contour is collapsed to a distinct point.



[K. Cole-McLaughlin, H. Edelsbrunner, J. Harer, V. Natarajan and V. Pascucci, 2004]

High-level techniques

- Using Reeb Graph to find initial nontrivial loops/tunnels/handles
- Using optimization to find the ideal ones

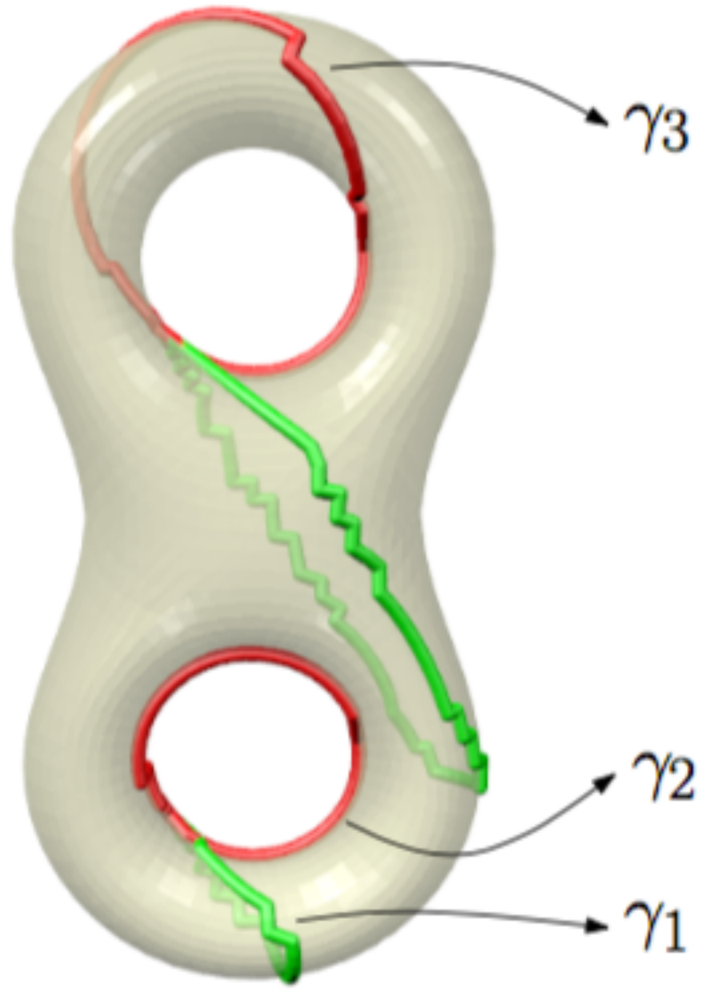


Figure 2: *γ_1 is a handle loop and γ_2 a tunnel loop. γ_3 is neither.*

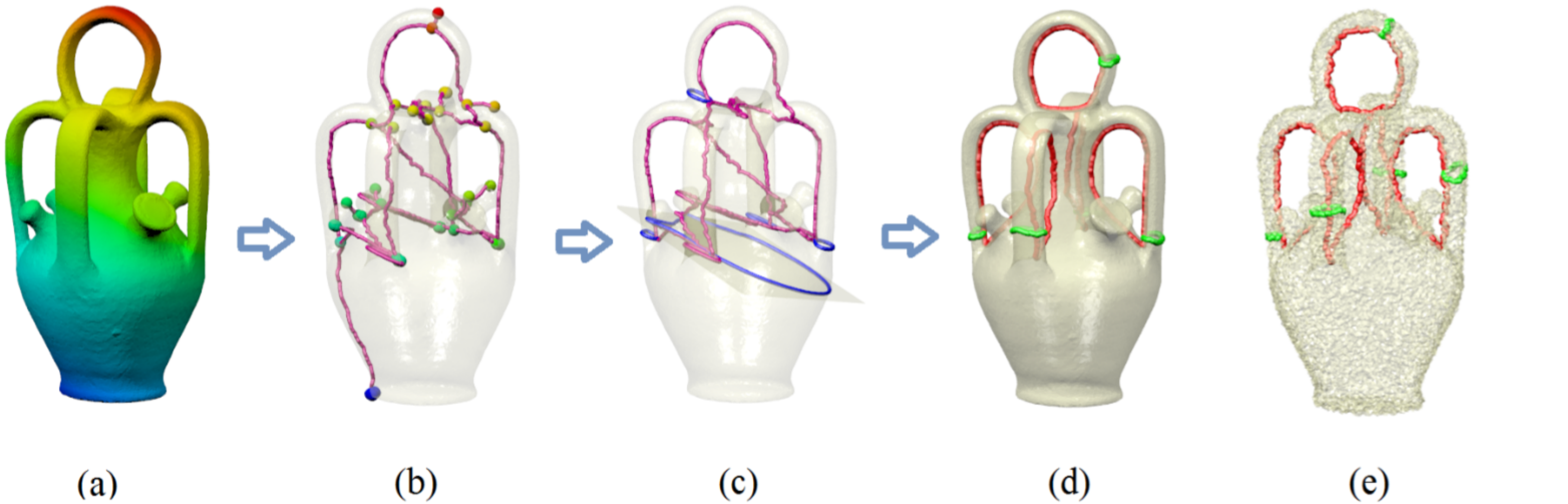


Figure 1: *(a) – (d) shows the pipeline of our algorithm: (a) The height function on the input surface. (b) Reeb graph w.r.t. the height function. (c) Initial handle and tunnel loops. (d) Final handle / tunnel loops after geometric optimization. (e) The output is stable under noise.*

Fast processing with original mesh

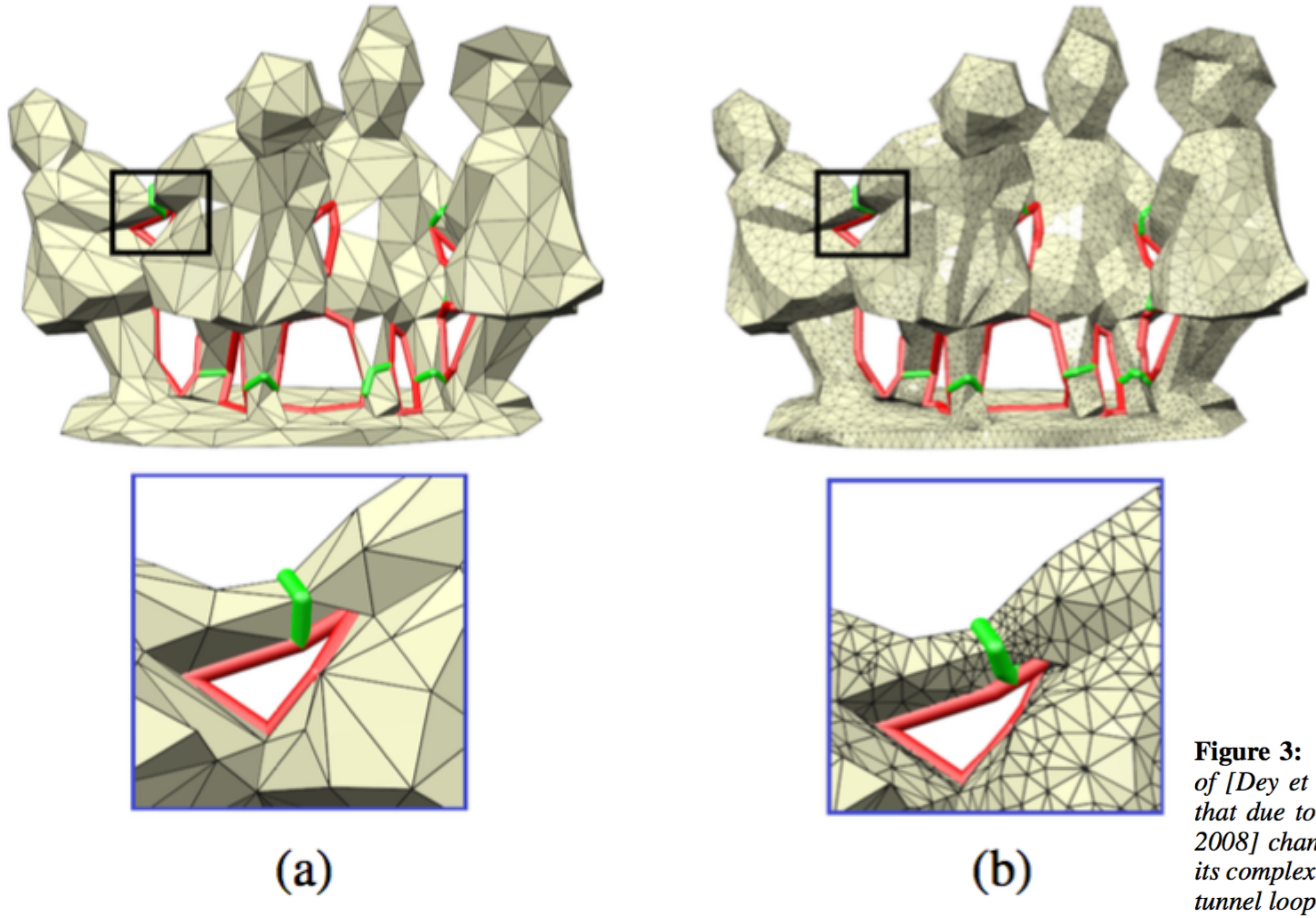


Figure 3: *The output of (a) our algorithm and (b) the algorithm of [Dey et al. 2008] for an input mesh with 449 vertices. Note that due to the tetrahedral meshing, the algorithm of [Dey et al. 2008] changes the input surface mesh and significantly increases its complexity to 7943 vertices. Our algorithm obtained handle and tunnel loops of good quality from the original sparse mesh.*

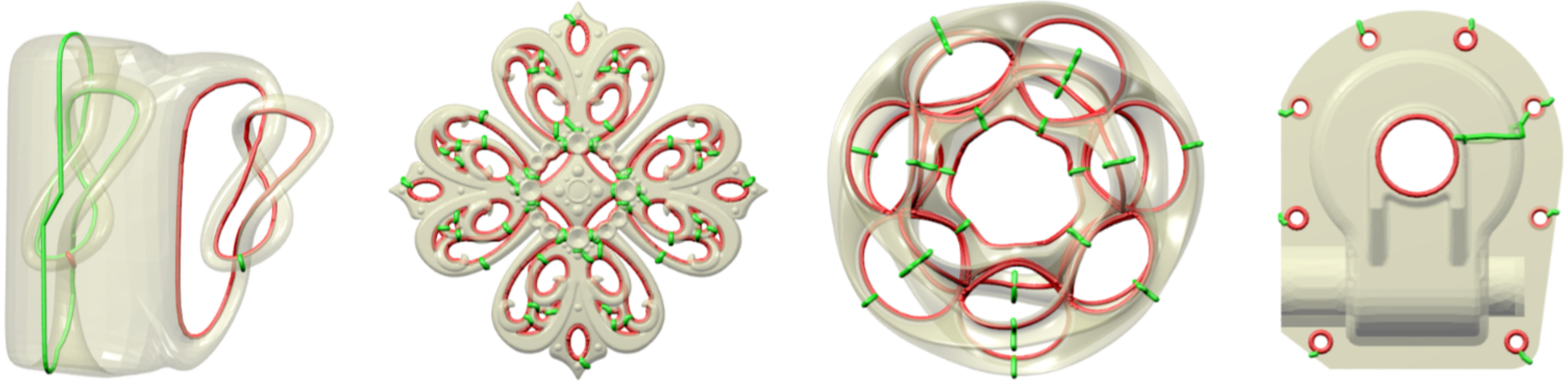
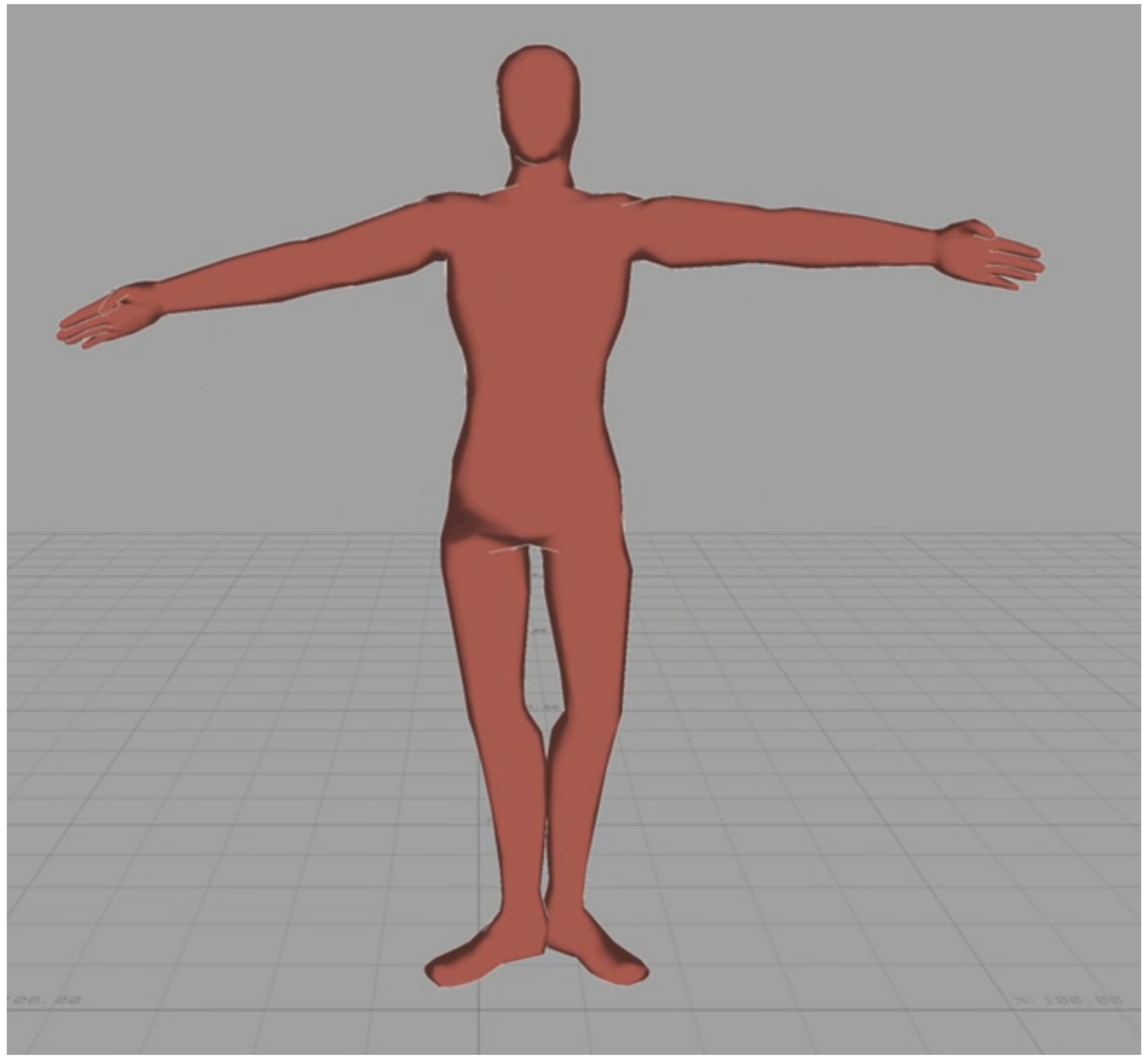


Figure 6: *Various examples. From left to right:* KNOTTY-CUP, FILIGREE, HEPTOROID *and* CASTING.

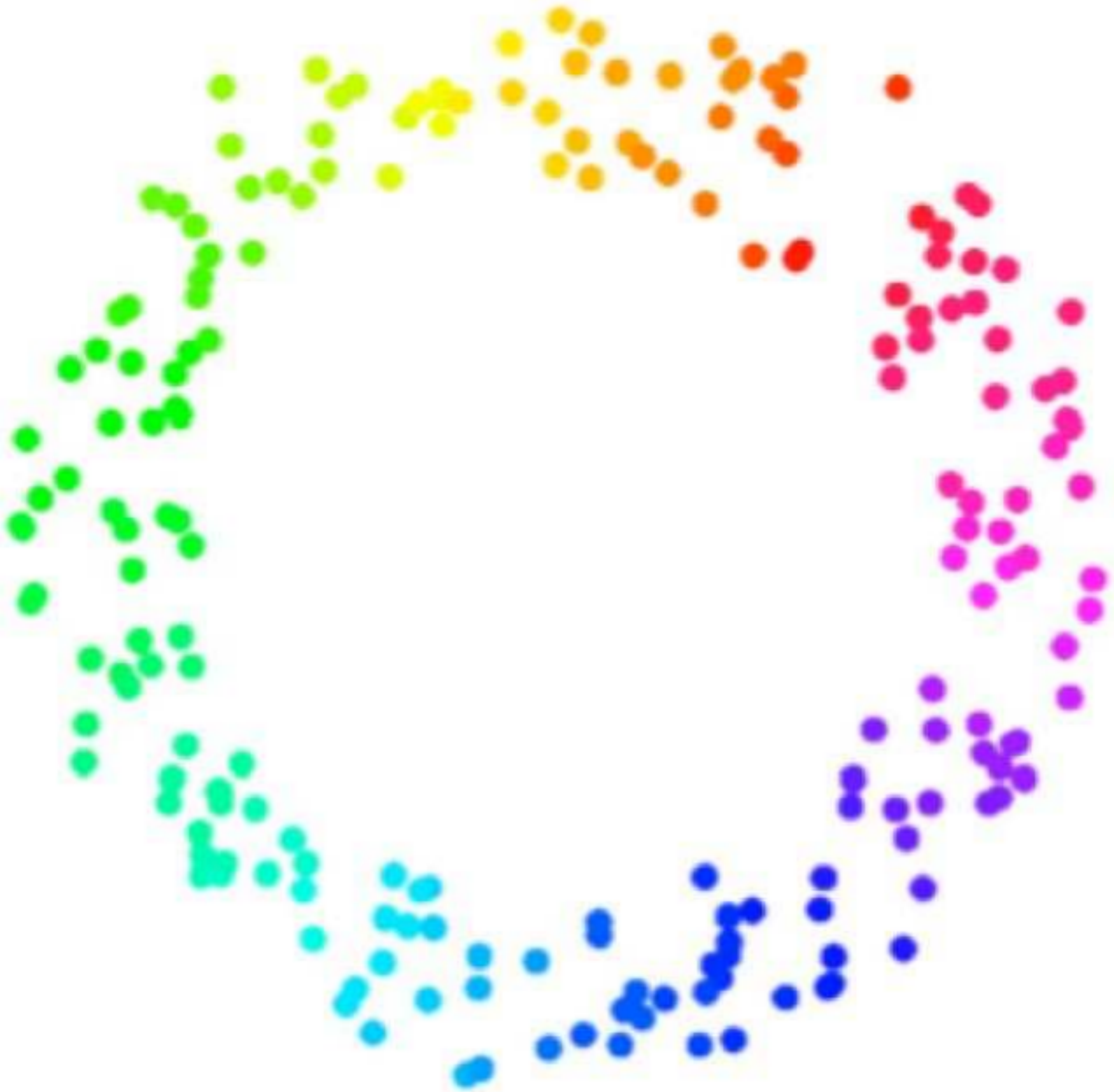


Circular and Branching Structures in High-dim

[WangSummaPascucci2011]

http://www.sci.utah.edu/~beiwang/publications/Branching_BeiWang_2011.pdf

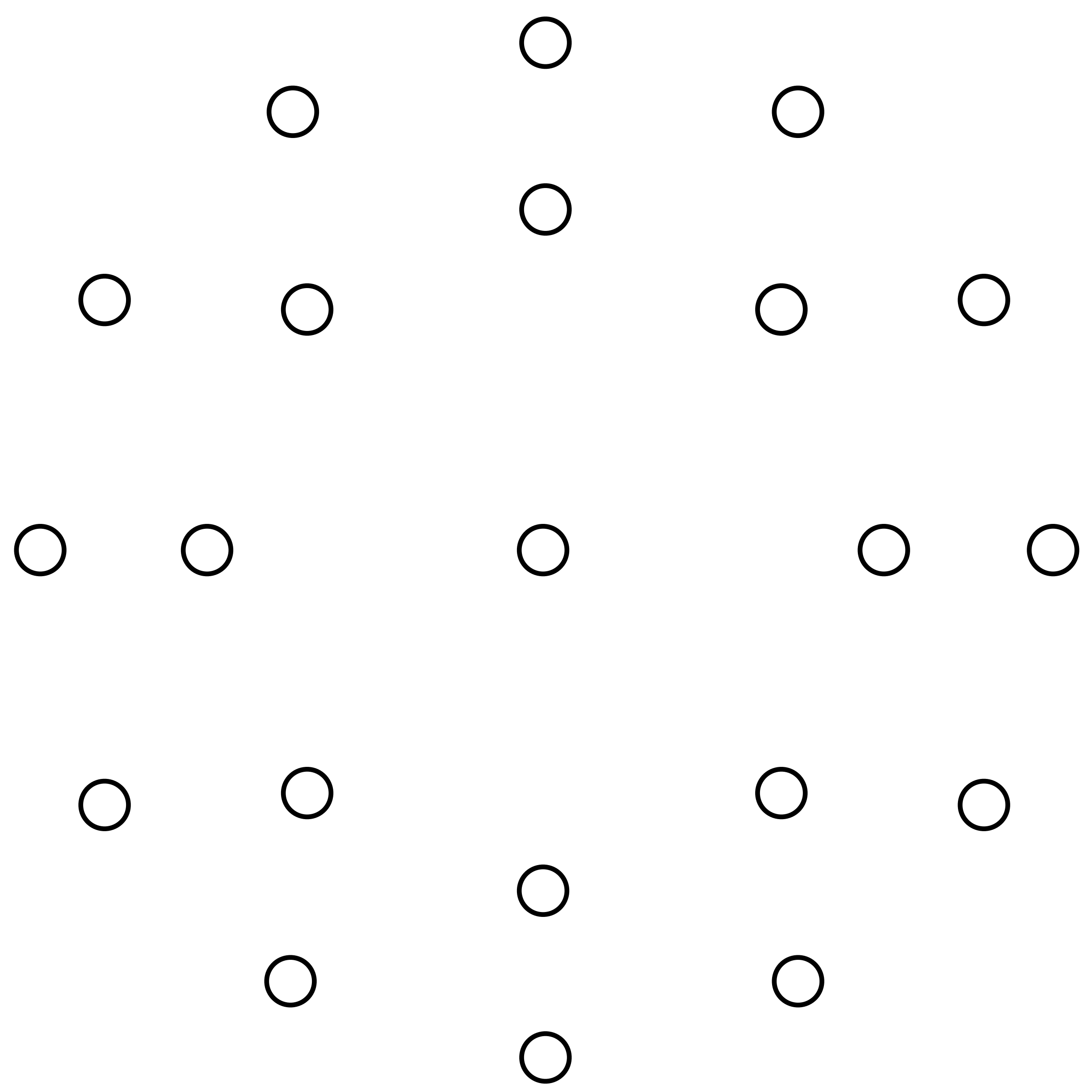
Inferring circular structure



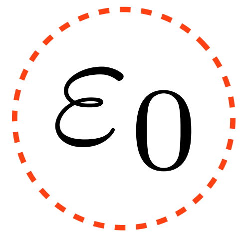
High-level techniques

- Persistent homology (PH), persistent cohomology (dual version)
- Circular parametrization

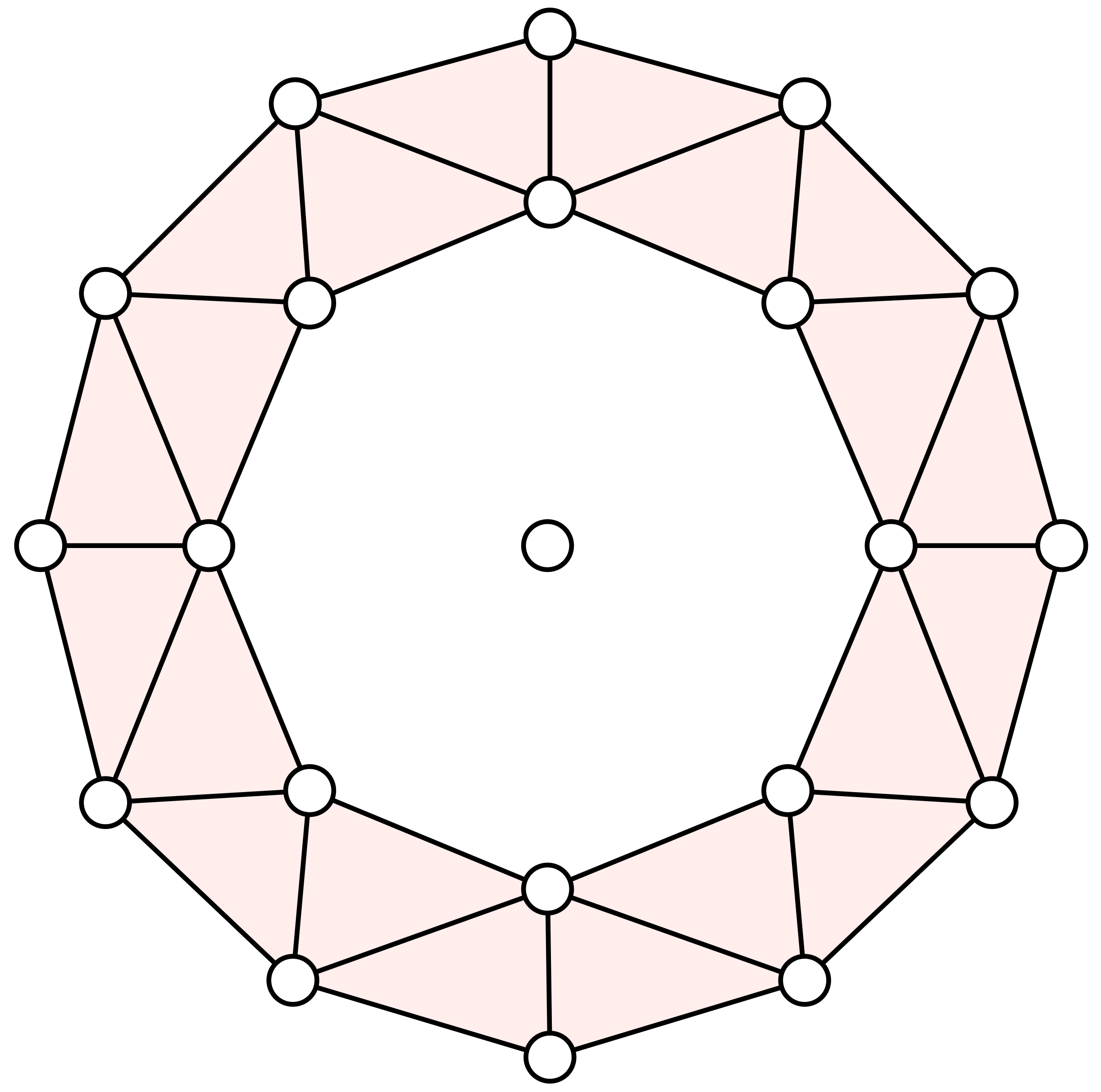
PH and parametrization



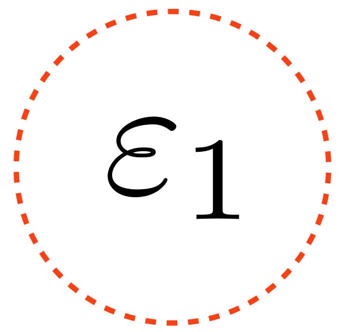
$Rips(X, \varepsilon_0)$



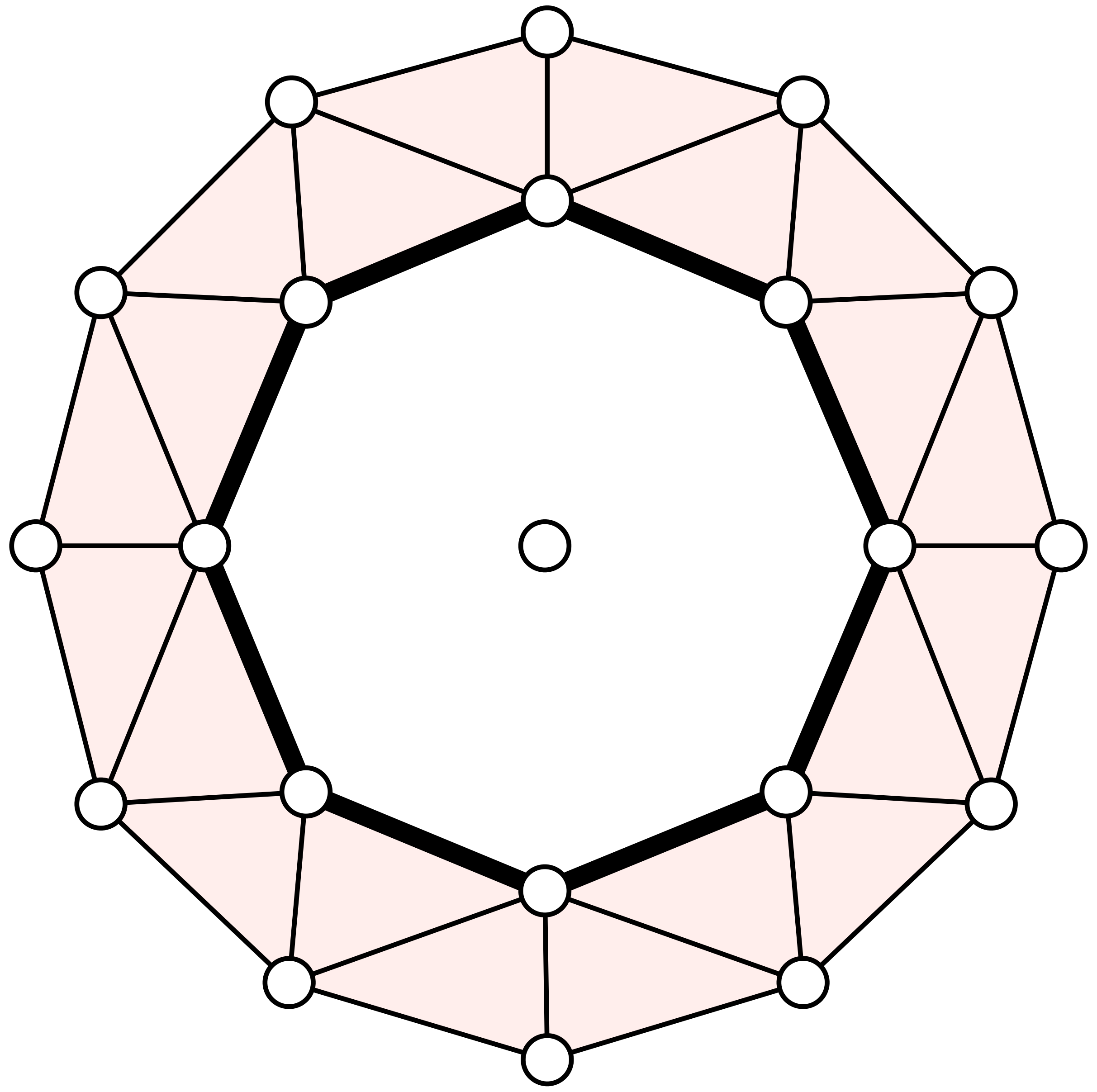
PH and parametrization

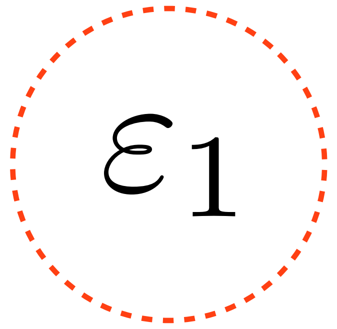


$$Rips(X, \varepsilon_0) \subseteq Rips(X, \varepsilon_1)$$



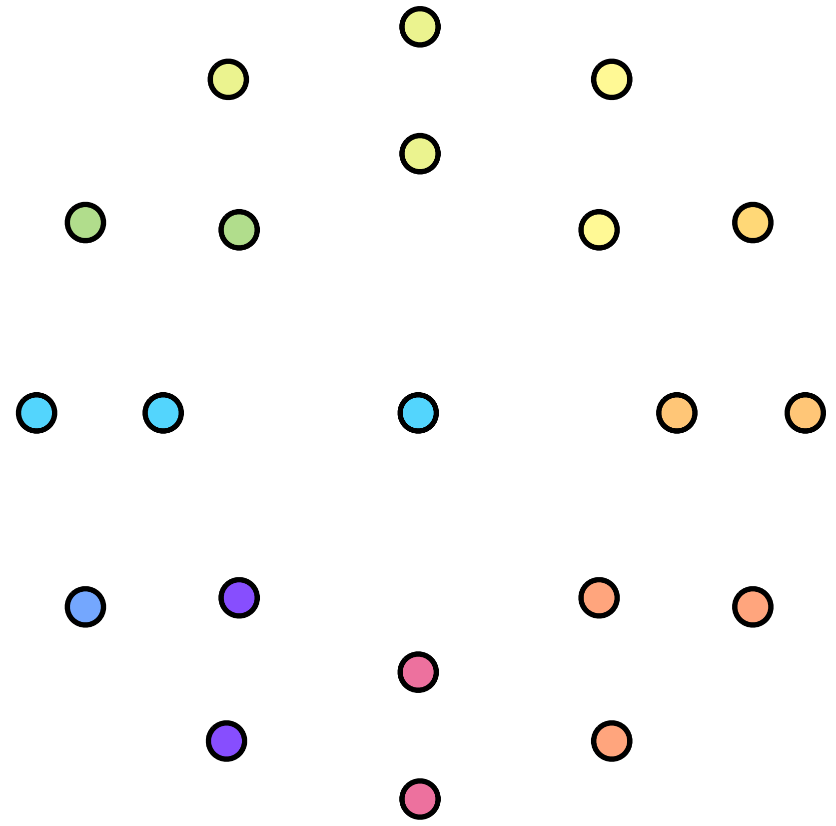
PH and parametrization



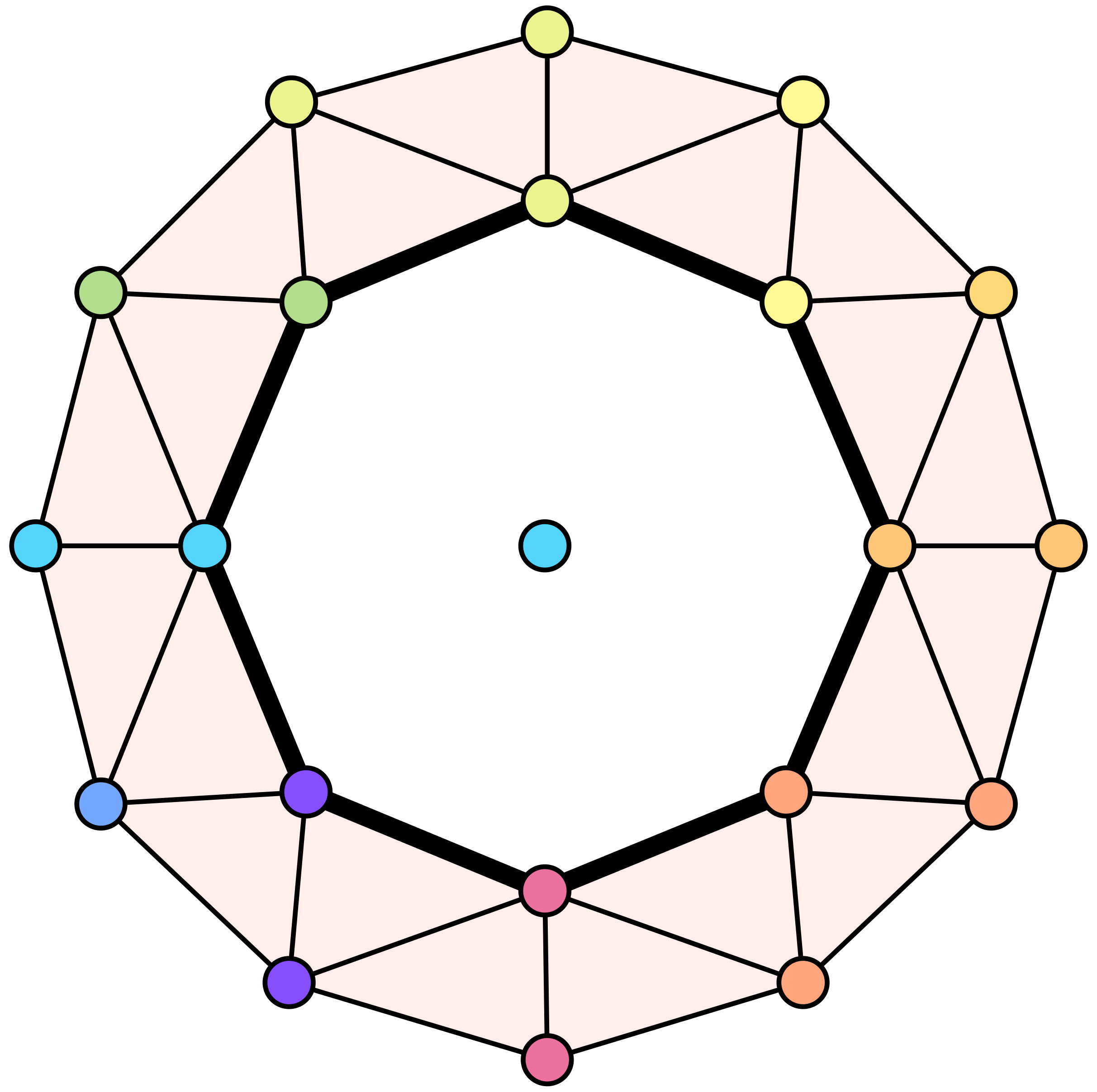


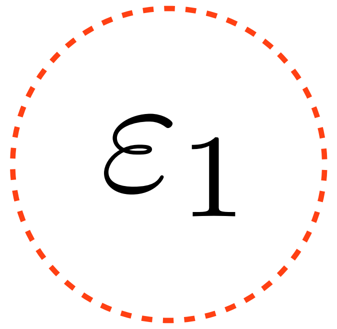
$Rips(X, \varepsilon_0) \subseteq Rips(X, \varepsilon_1)$

PH and parametrization



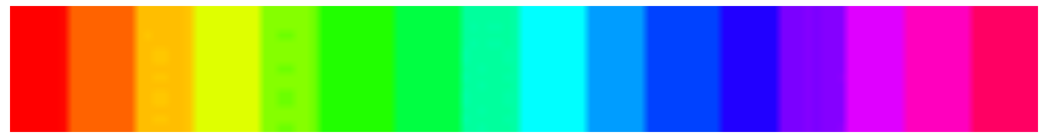
Born: ε_1



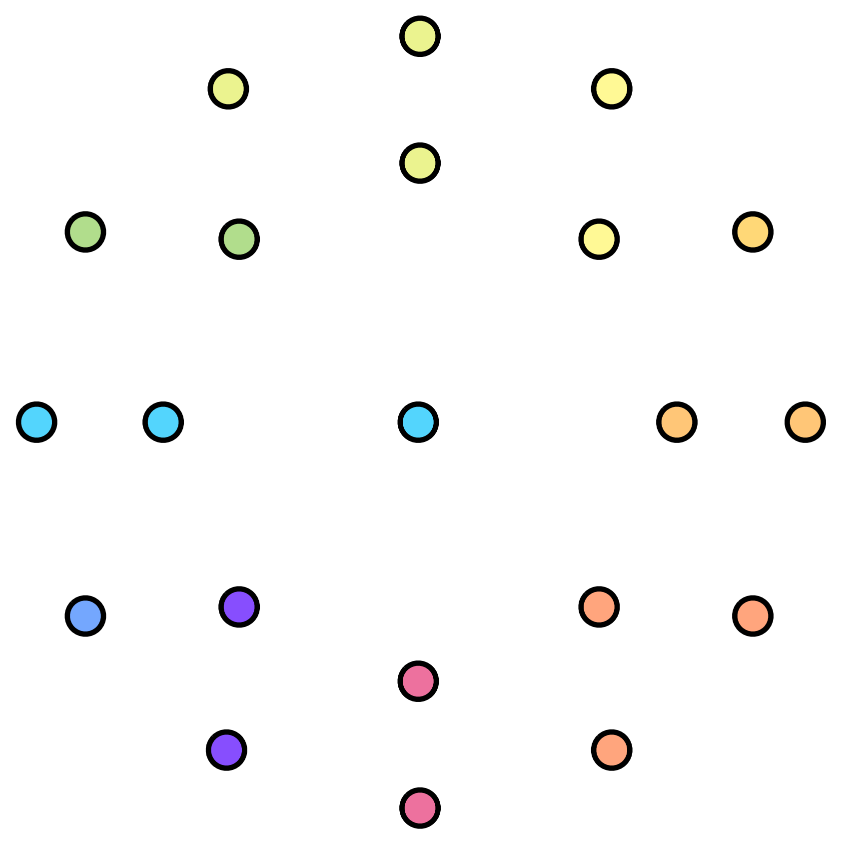


$$Rips(X, \varepsilon_0) \subseteq Rips(X, \varepsilon_1)$$

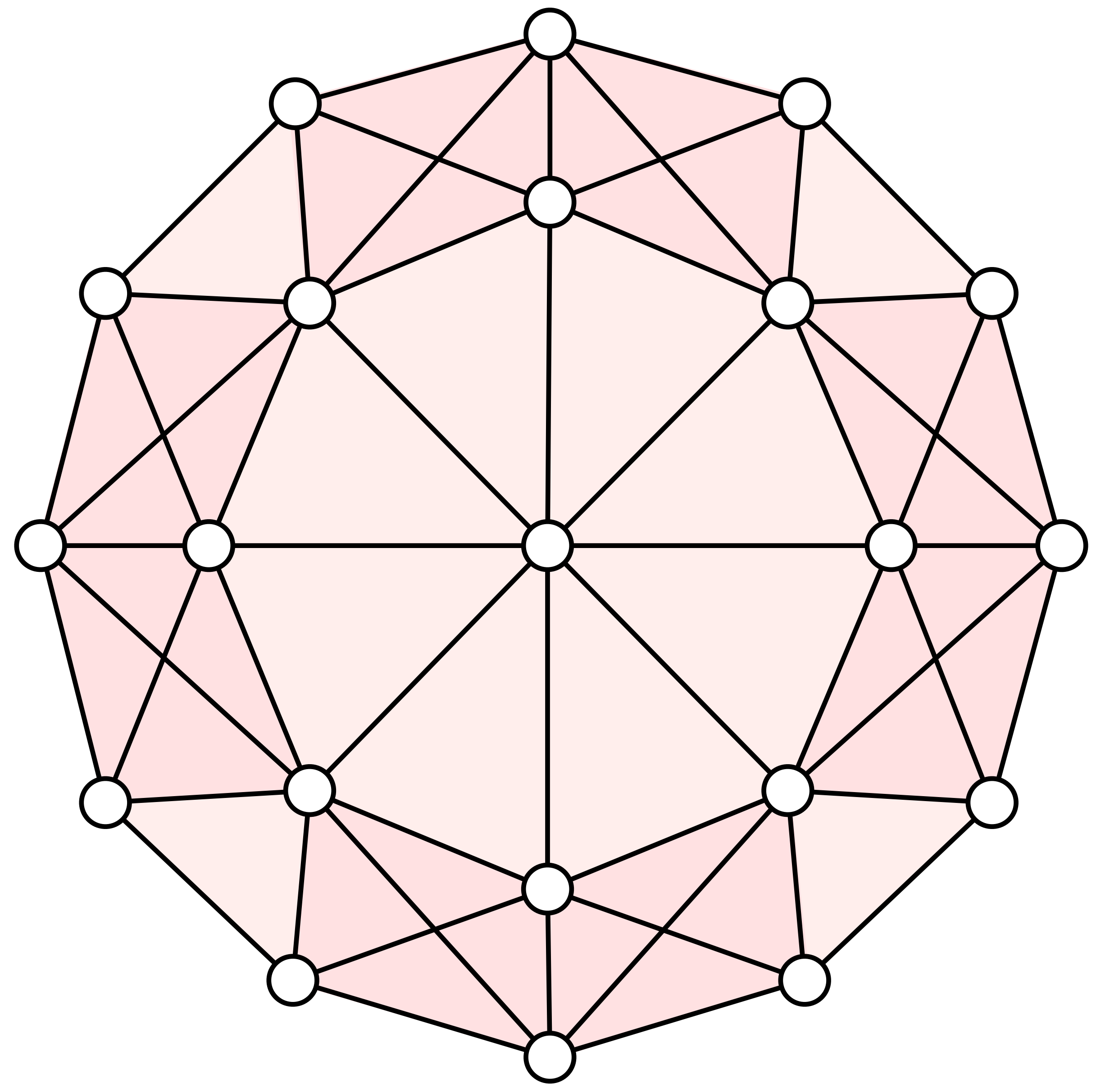
Parameter Space:

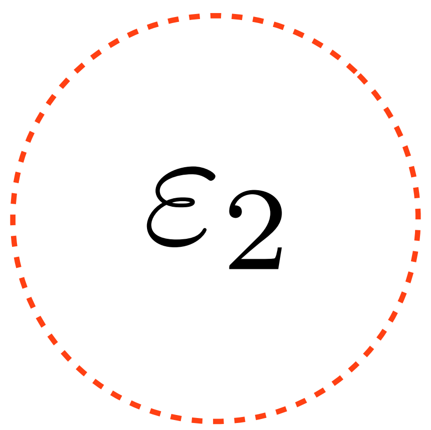


PH and parametrization



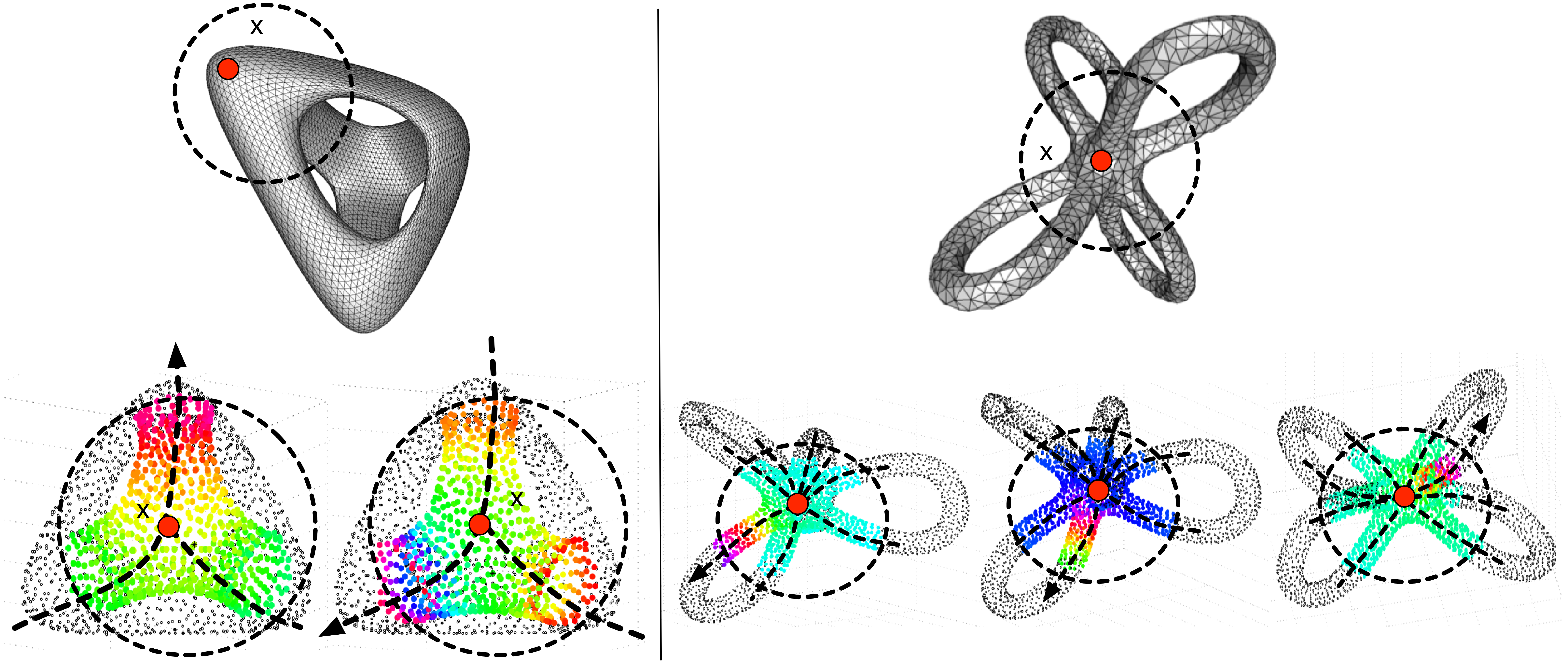
Born: ε_1 Died: ε_2
Persistence: $\varepsilon_2 - \varepsilon_1$



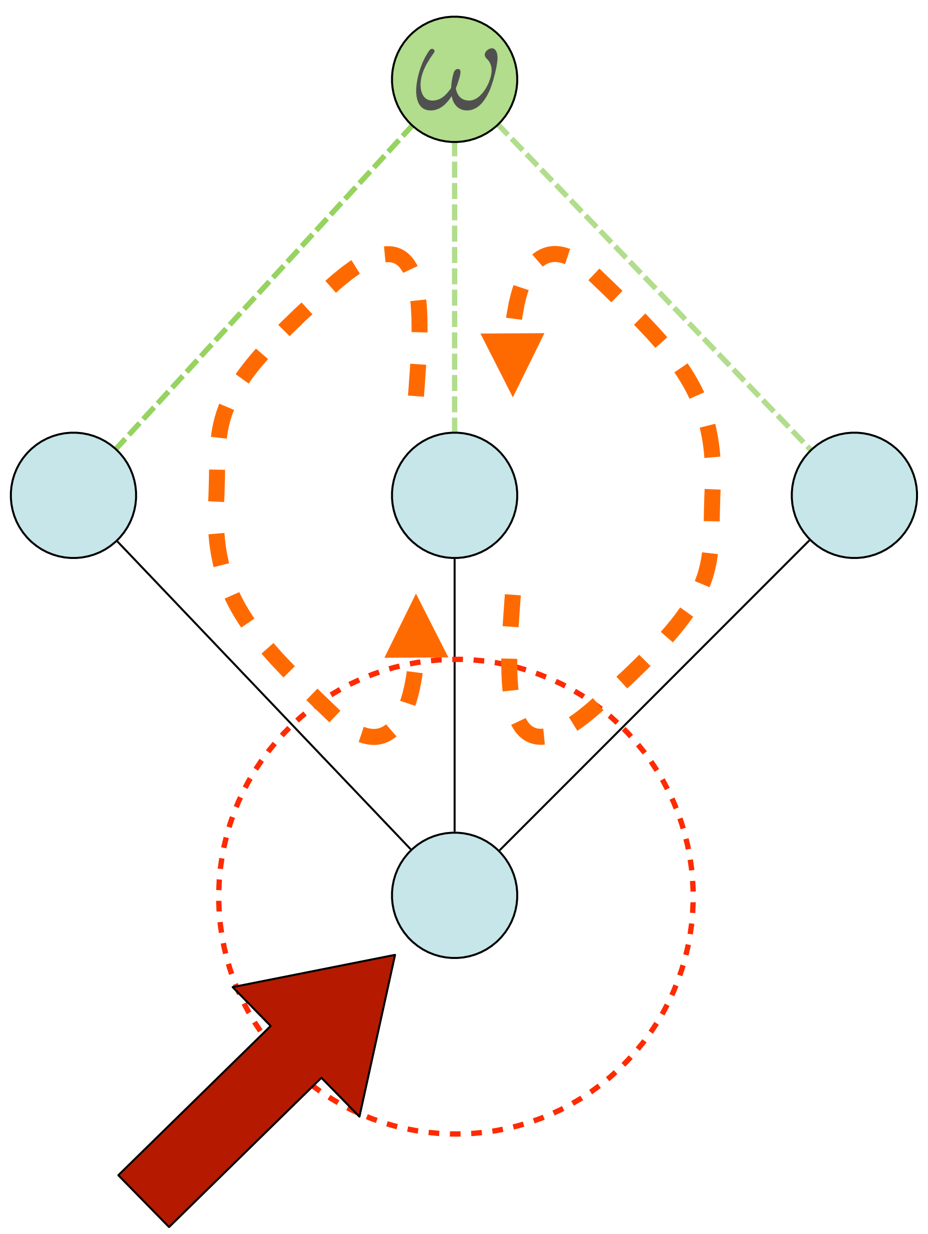


$$Rips(X, \varepsilon_0) \subseteq Rips(X, \varepsilon_1) \subseteq Rips(X, \varepsilon_2)$$

Inferring branching structure



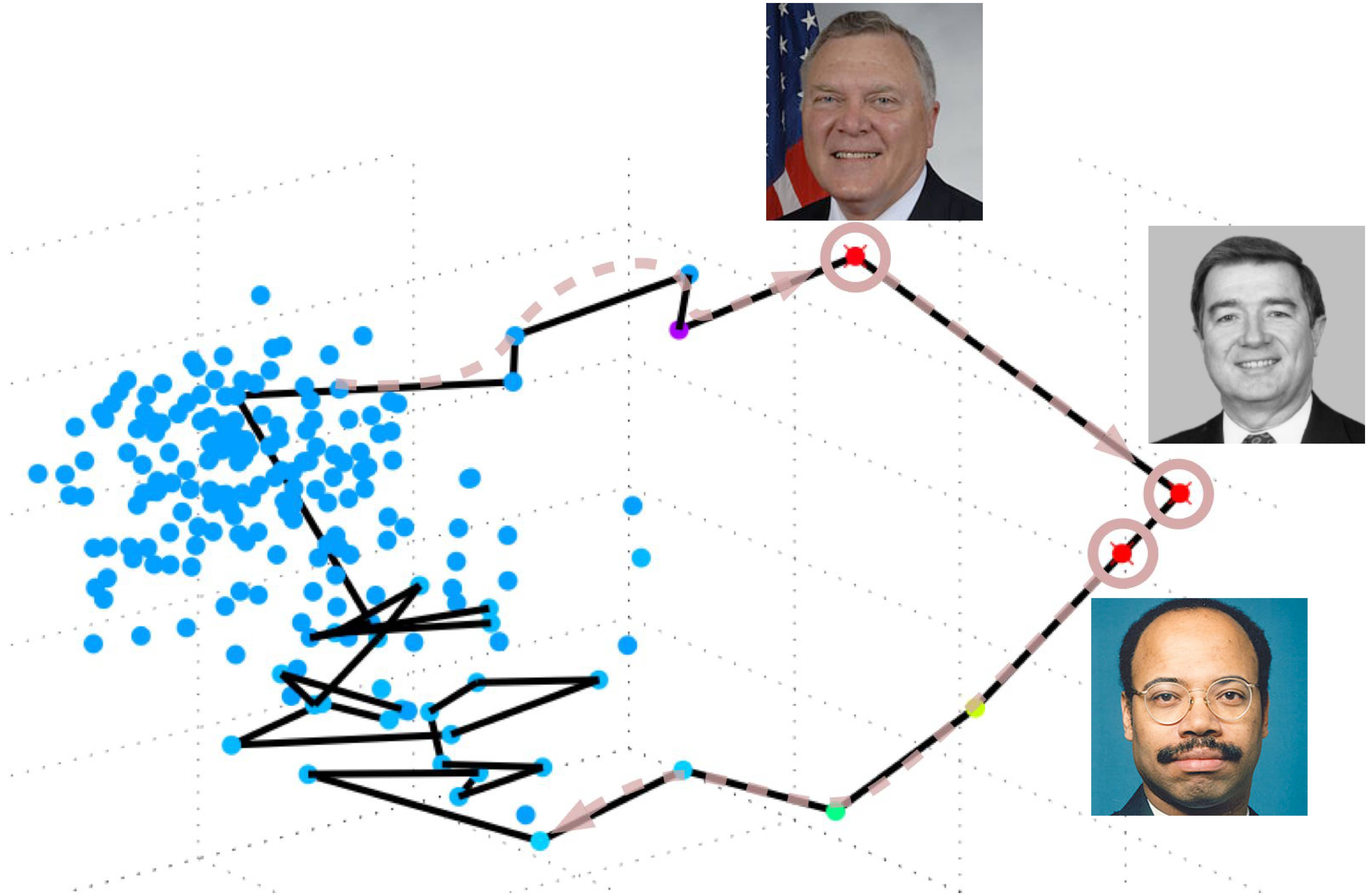
Branching and parametrization



Given a neighborhood around a point, attach simplicies which cross the neighborhood threshold to a dummy vertex ω.

In this way, we turn local branching features into circular structures.

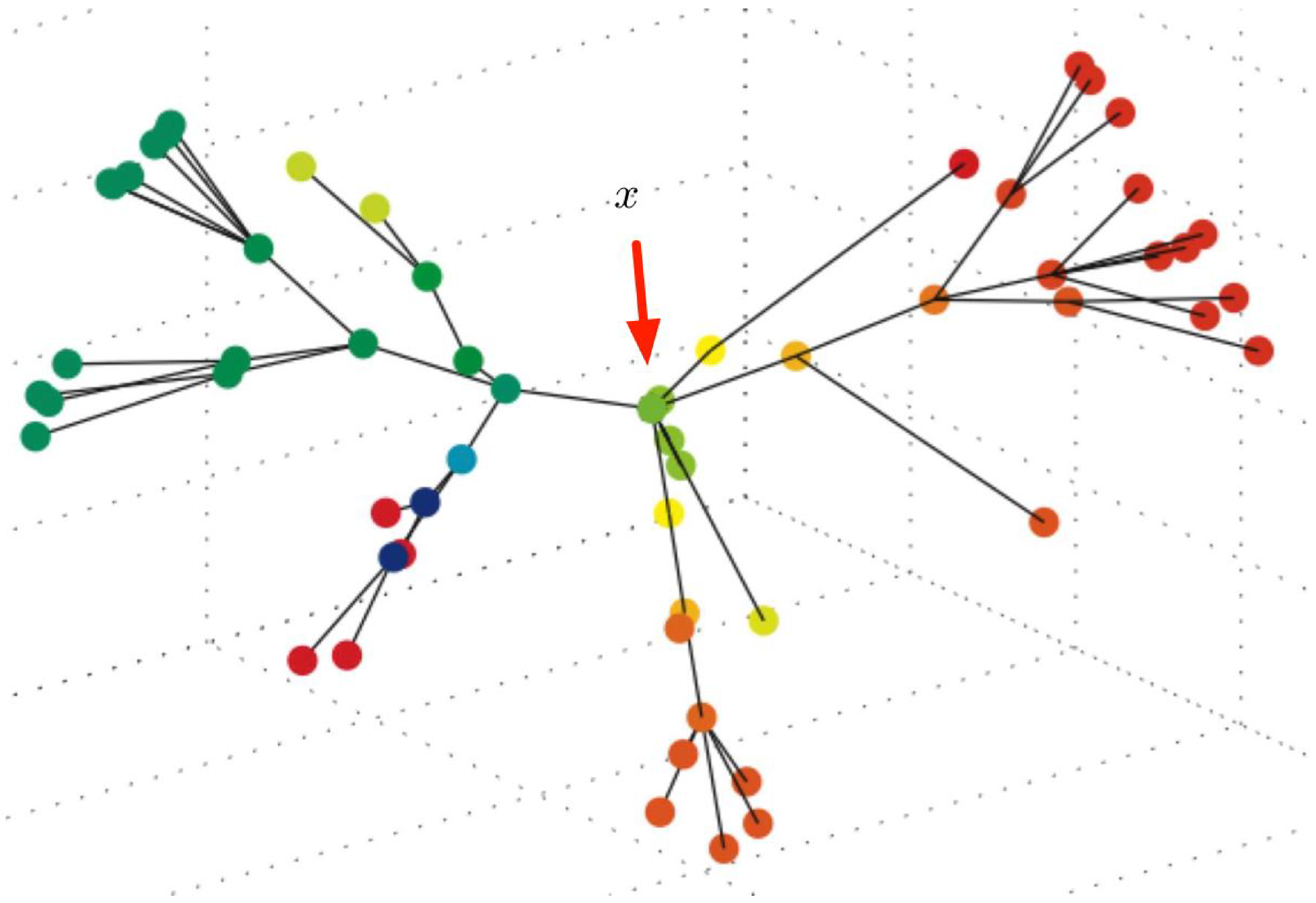
Voting Data



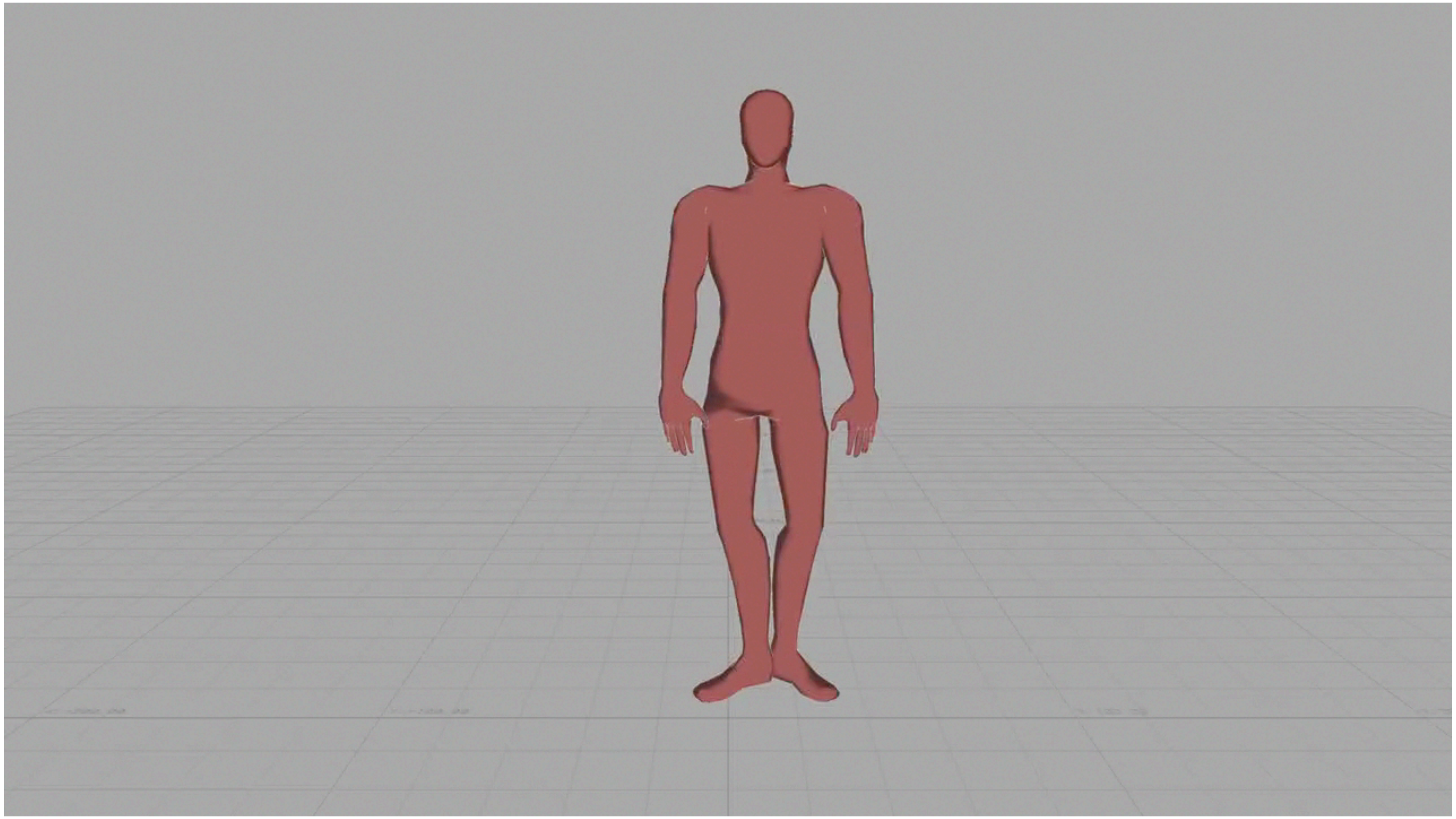
1995 House of Representatives Voting Record
885 votes (dimension)
205 Democratic congresspeople (points)
Record: (Yea/Nay/Absent)
94.27 seconds to compute
(92.15 Rips, 1.76 Persistence)

Outliers: switched party or resigned

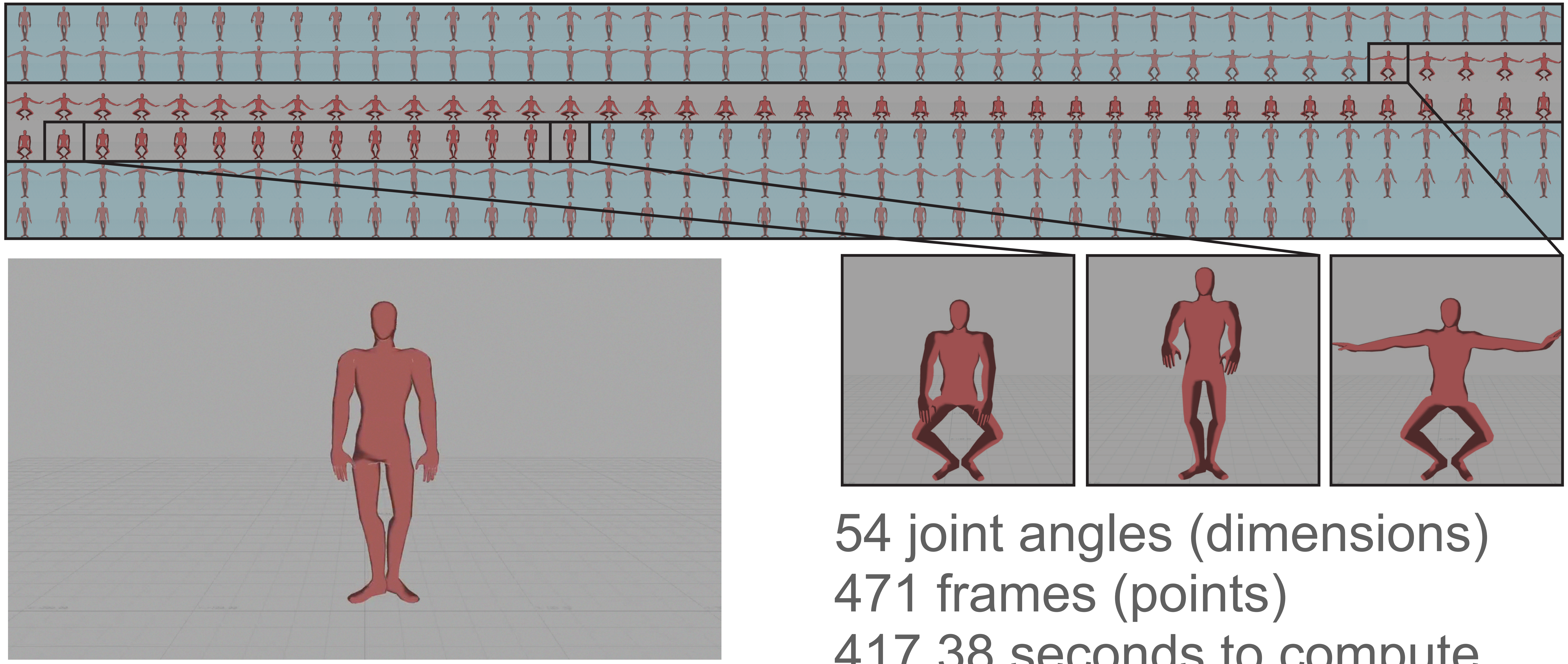
Virus Data



1045 nucleotides (dimensions)
58 mutated genetic sequences (points)
0.09 seconds to compute
(0.05 Rips, 0.02 Persistence)

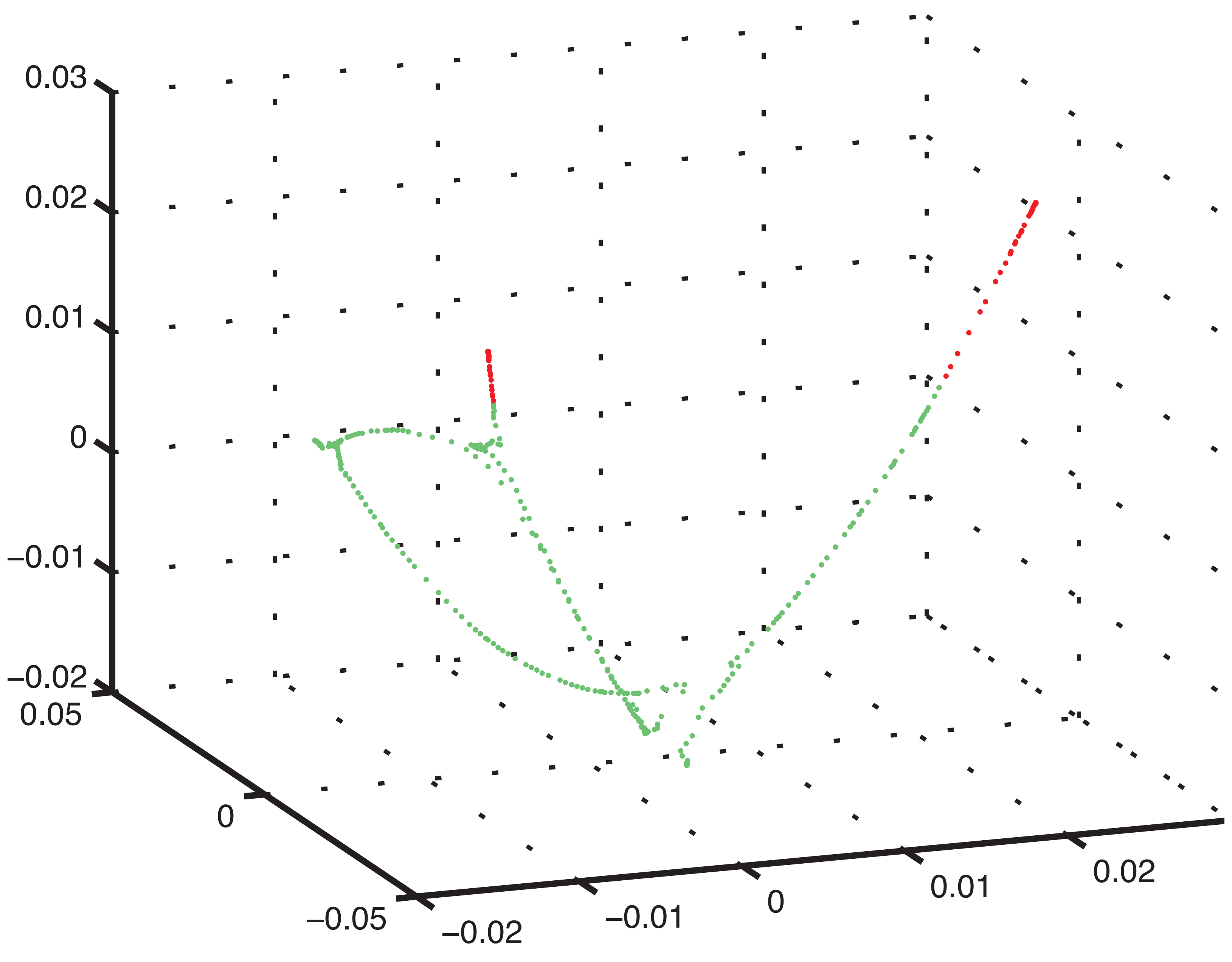


Motion Capture: Ballet

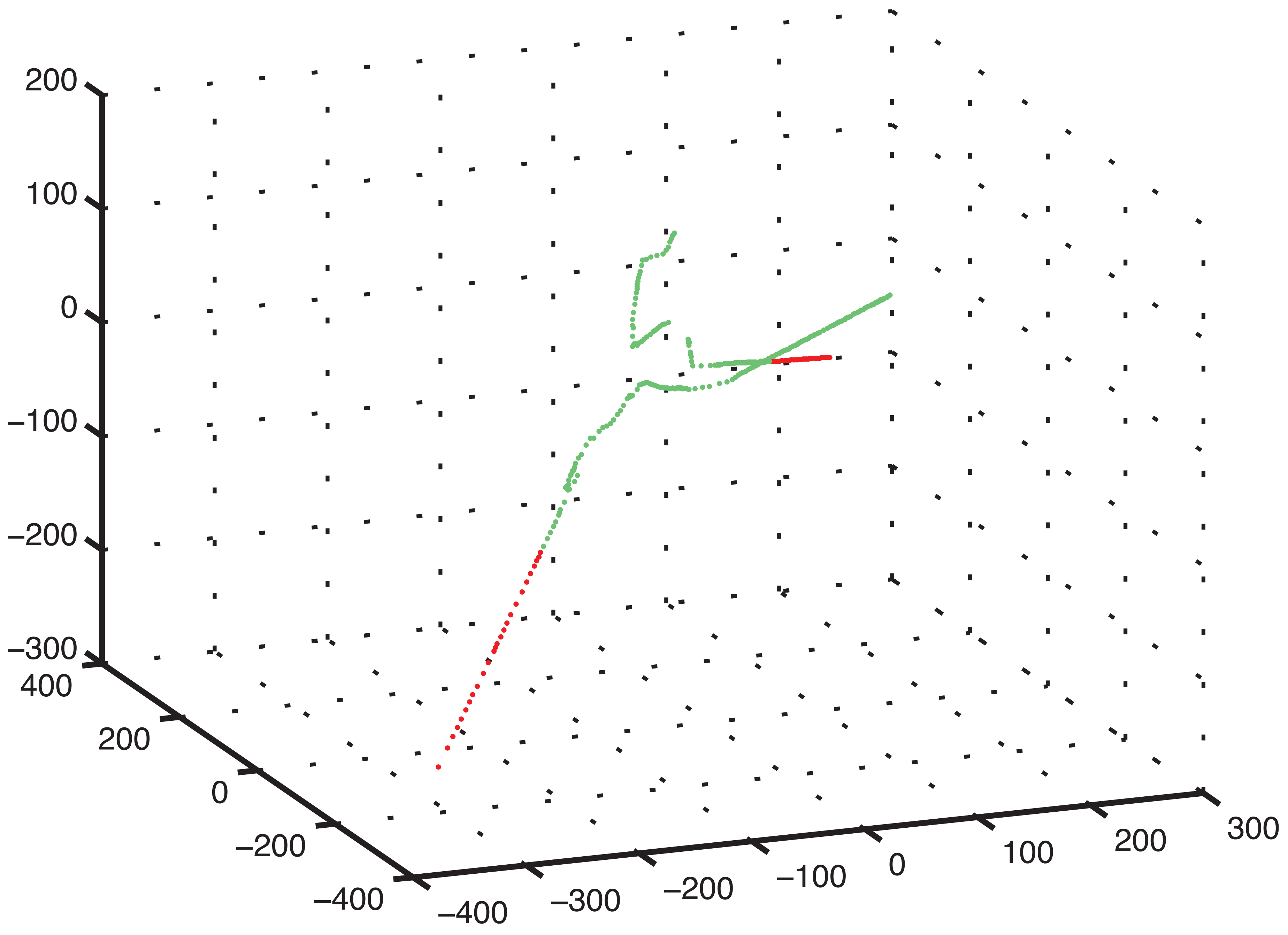


54 joint angles (dimensions)
471 frames (points)
417.38 seconds to compute
(363.67 Rips, 30.47 Persistence)

Motion Capture: Ballet

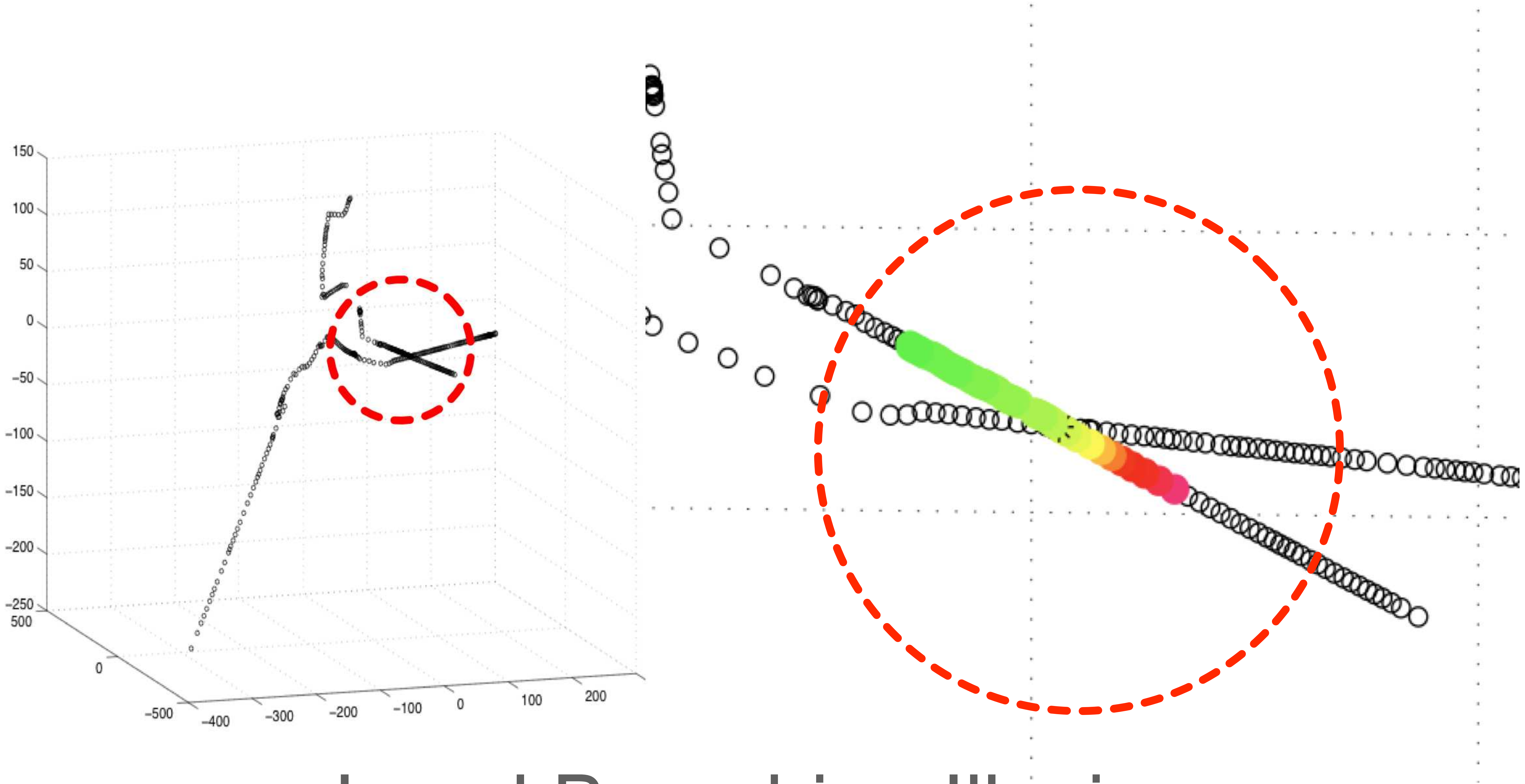


Laplacian Eigenmaps



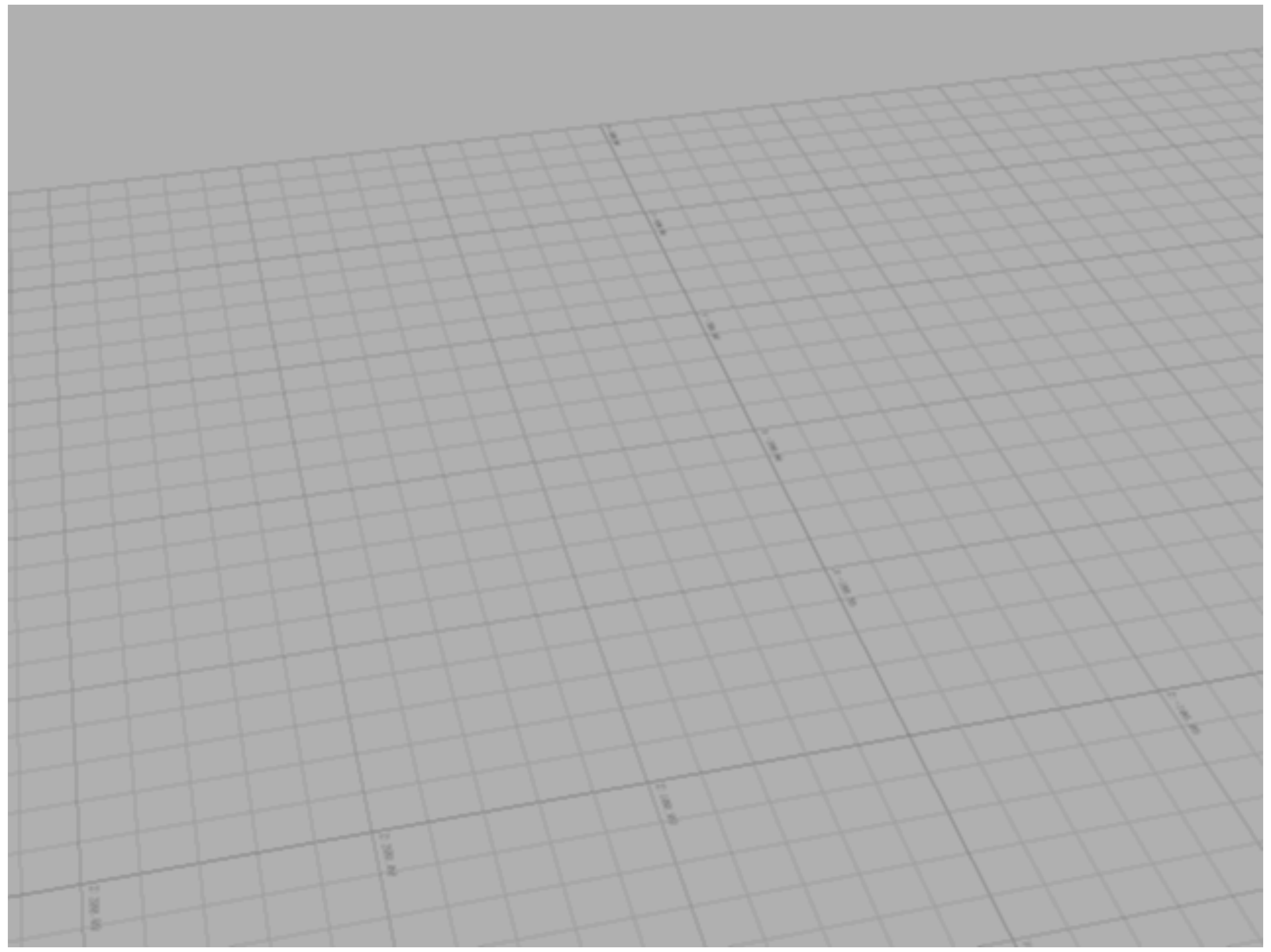
3D Isomap

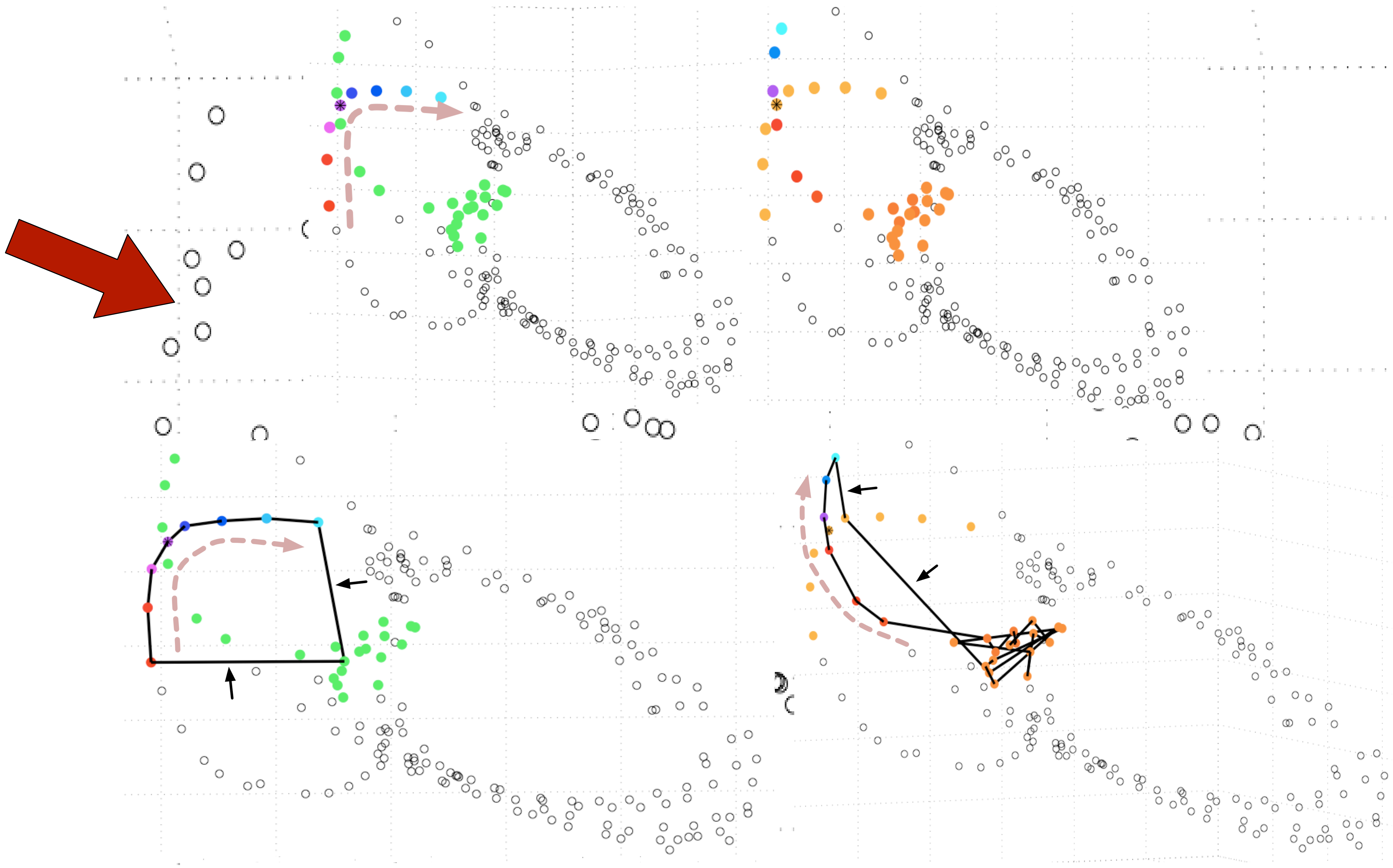
Motion Capture: Ballet



Local Branching Illusion

Motion Capture - Walk/Hop/Walk





66 joint angles (dimensions)
189 frames (points)
0.08 seconds to compute
(0.08 Rips)

Thanks!

Any questions?

You can find me at: beiwang@sci.utah.edu

CREDITS

Special thanks to all people who made and share these awesome resources for free:

- Presentation template designed by Slidesmash
- Photographs by unsplash.com and pexels.com
- Vector Icons by Matthew Skiles

Presentation Design

This presentation uses the following typographies and colors:

Free Fonts used:

http://www.1001fonts.com/oswald-font.html

https://www.fontsquirrel.com/fonts/open-sans

Colors used