Advanced Data Visualization

CS 6965

Spring 2018

Prof. Bei Wang Phillips

University of Utah

Lecture 13

Announcement

- A few students might be able to form groups: survey of teams, who is going solo?
- In the process of getting virus project presentation
- Project 2: penalty free 2-day extension (due Thursday 2/22 9:10 a.m.)

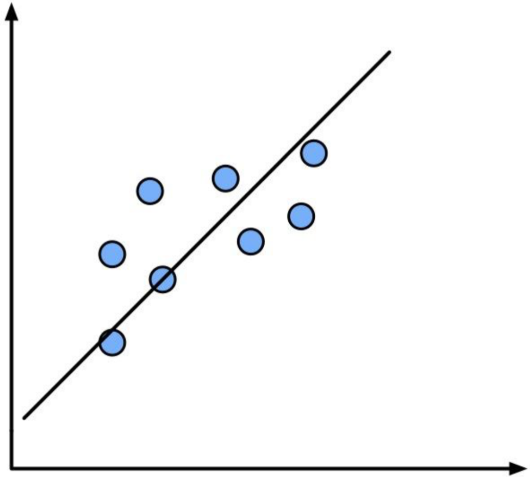
Topological Abstractions

For Scalar Field Data

TOPO

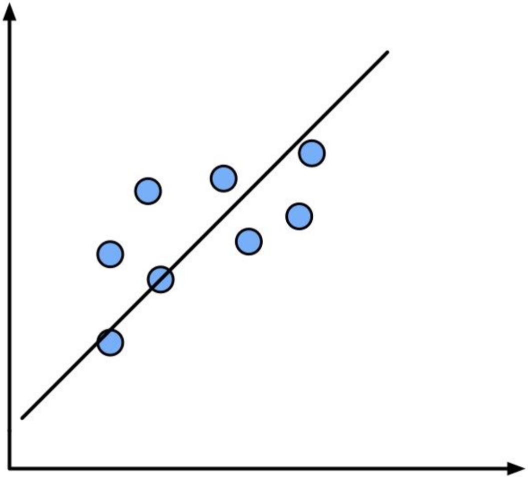
"Data has shape, and shape matters."
– Gunnar Carlsson

Linear regression



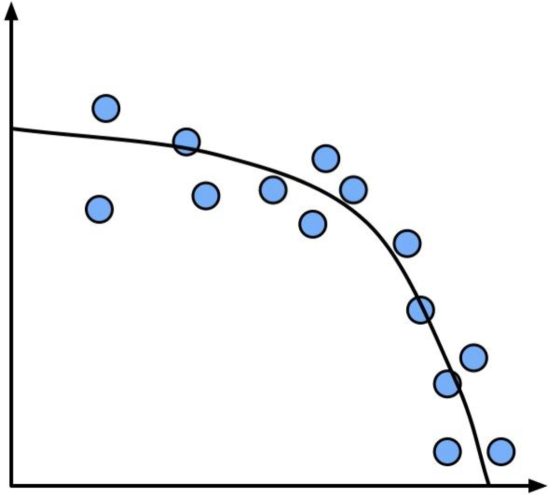
- Shape of data?
- A line.

Linear regression



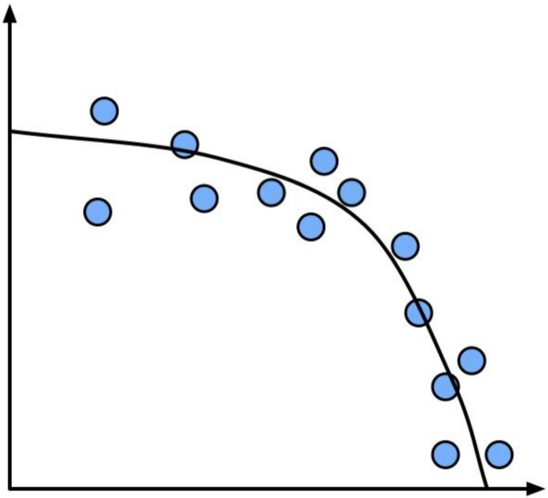
- Shape of data?
- A line.

Cubic polynomial regression



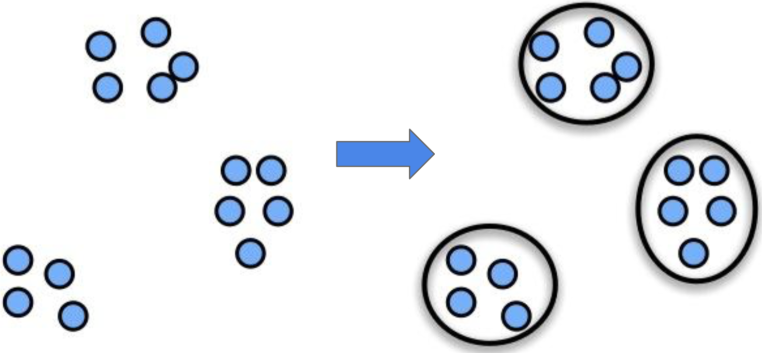
- Shape of data?
- A curve.

Cubic polynomial regression



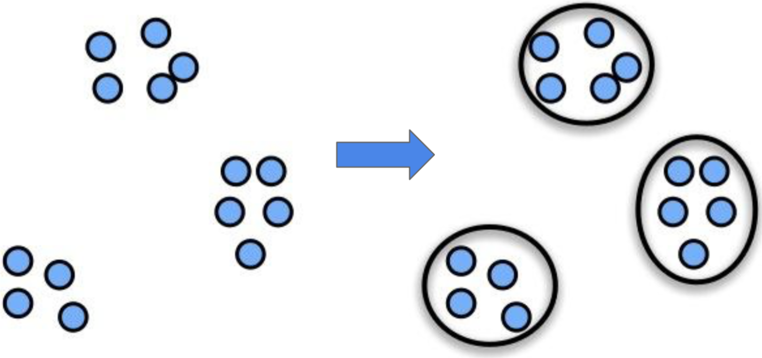
- Shape of data?
- A curve.

Clustering



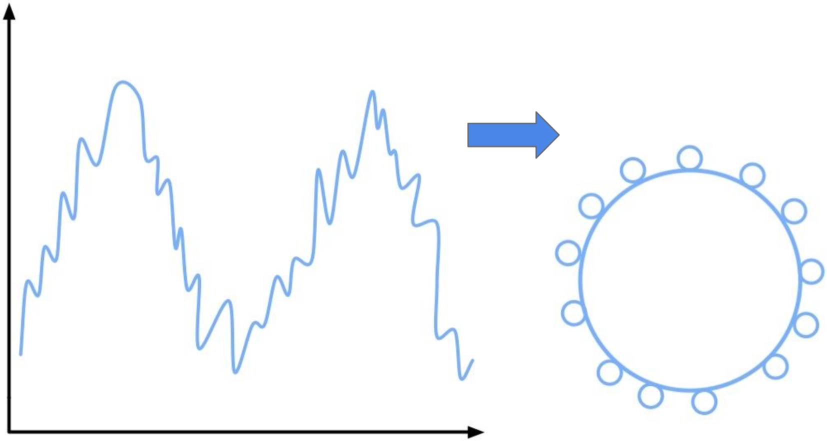
- Shape of data?
- Clusters.

Clustering



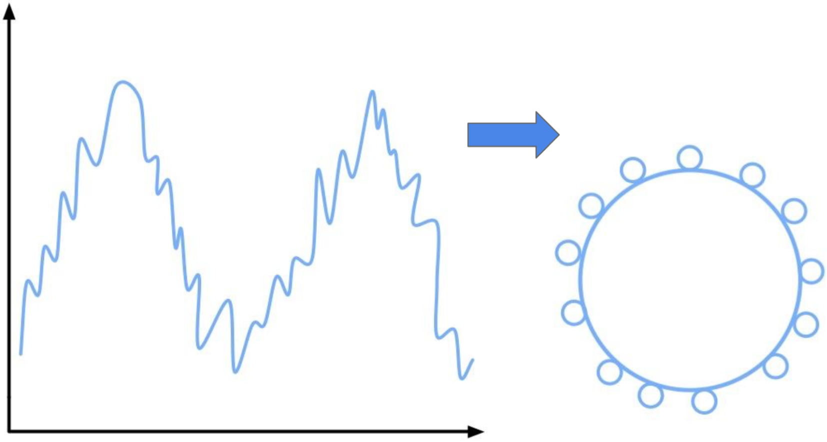
- Shape of data?
- Clusters.

Time series analysis



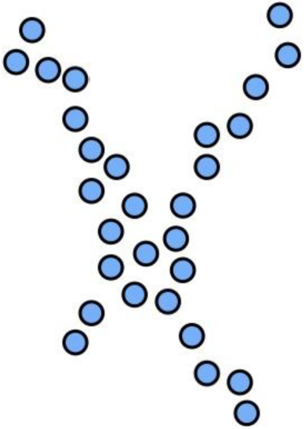
- Shape of data?
- Depends on the mapping.

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).

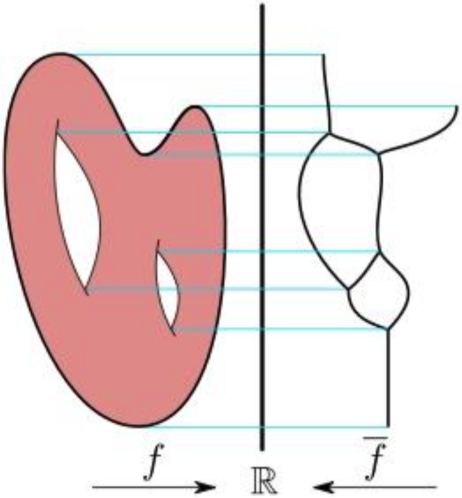
Discrete samples: a point cloud

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

Discrete samples: a point cloud

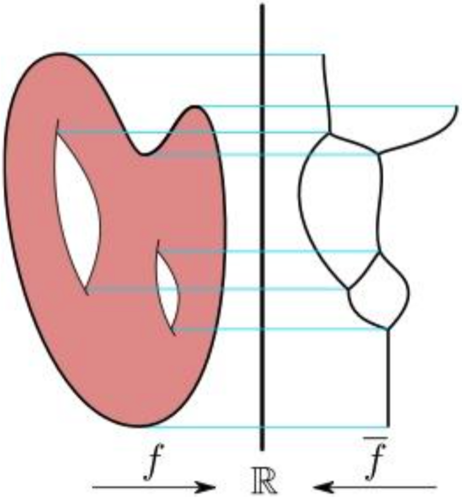
- Shape of data?
- Depending on the scale (or the resolution).

A scalar function defined on a manifold



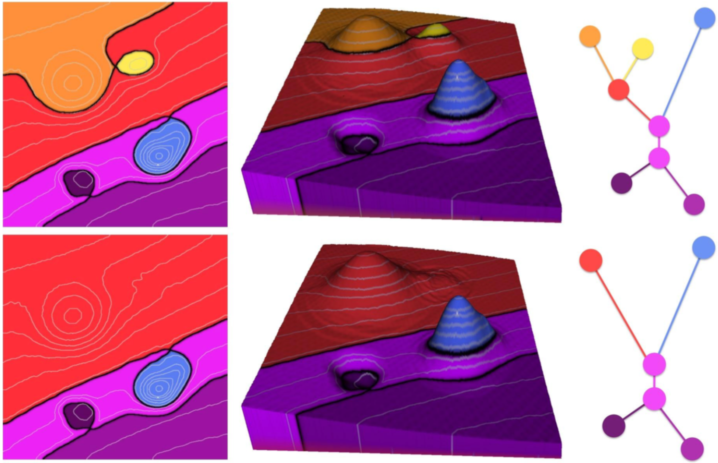
- Shape of data?
- A Reeb graph.

A scalar function defined on a manifold



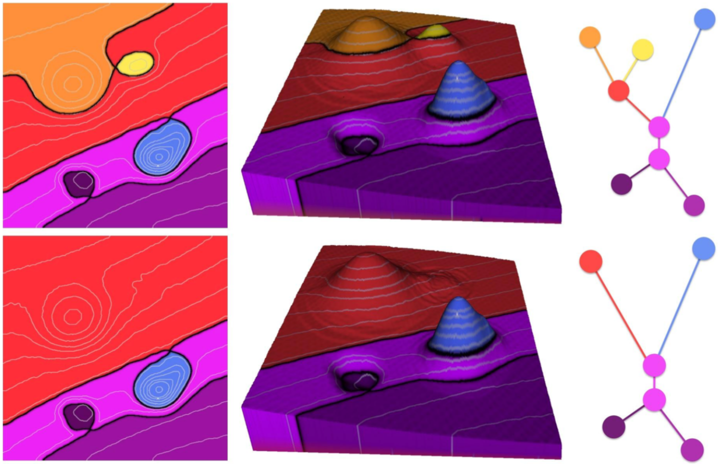
- Shape of data?
- A Reeb graph.

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



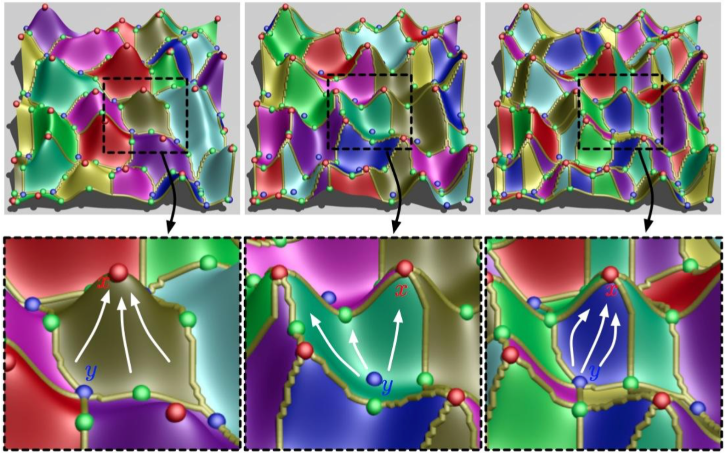
- Shape of data?
- A contour tree.

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



- Shape of data?
- A contour tree.

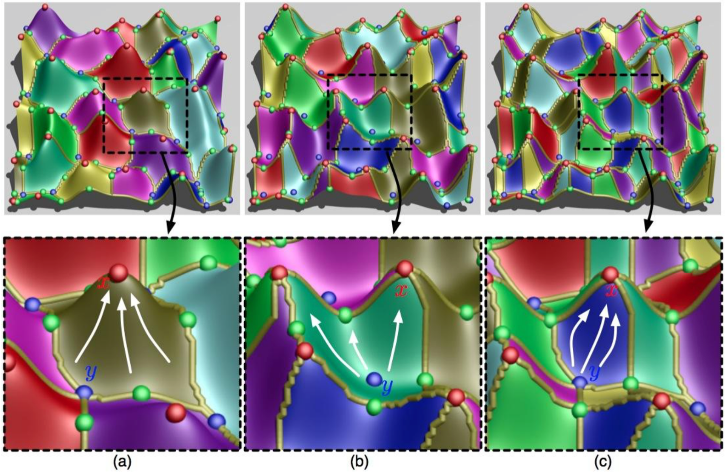
Elevation on terrain: a scalar function on a 2D domain



(a) (b) (c)

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

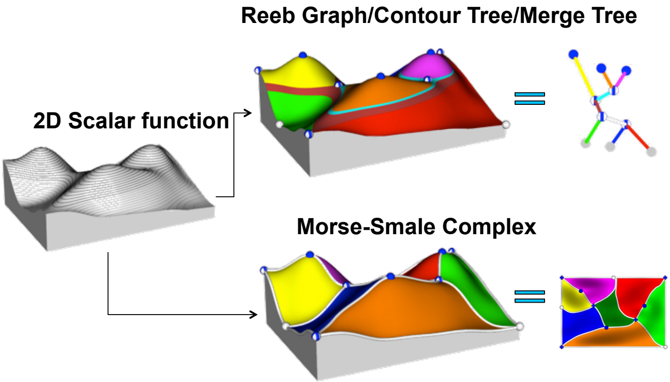
Elevation on terrain: a scalar function on a 2D domain



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

Some basic tools in topological data analysis (TDA)

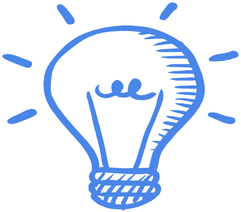
- Abstraction of the data: topological structures and their combinatorial representations
- Separate features from noise: persistent homology



Fundamental Tasks in Topological Data Analysis

Topology + Point Cloud = Magic Happens!

Task 1



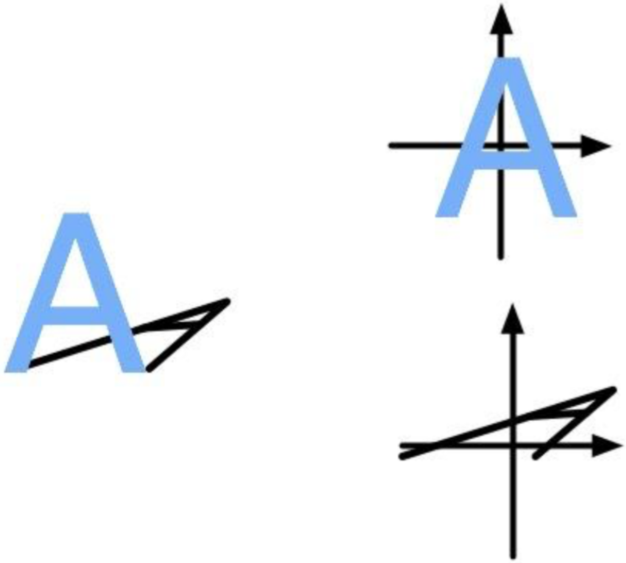
RECONSTRUCTION

How to assemble discrete point samples
into a global structure?

INFERENCE

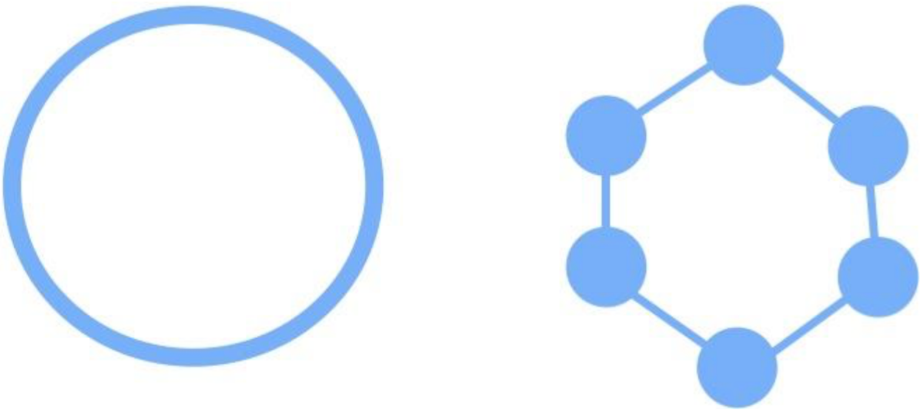
How to infer high-dimensional structures
from low-dimensional representations?

Key idea 1: coordinate free

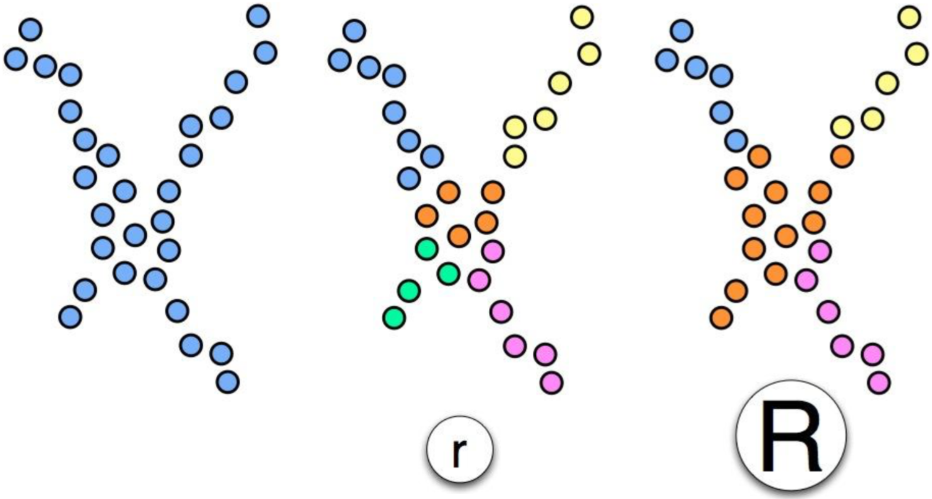


Key idea 2: deformation invariant

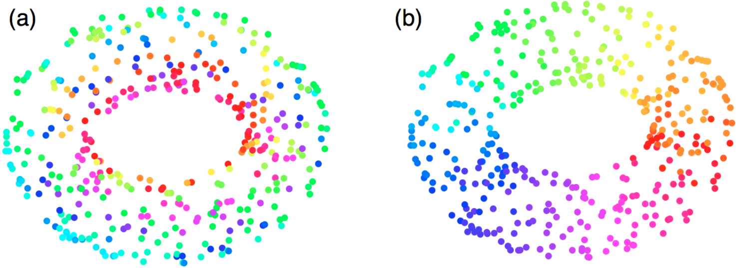
Key idea 3: compressed representation



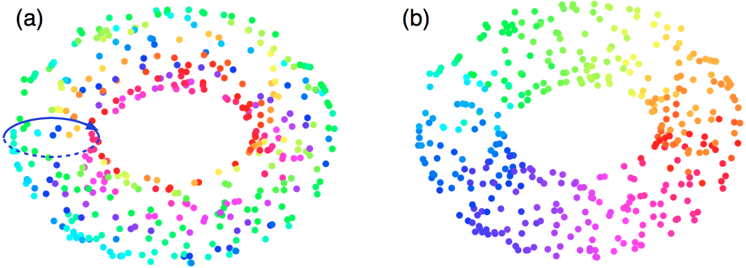
Inference: stratification learning



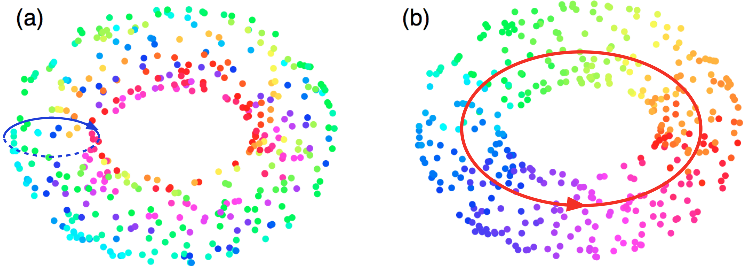
Inferring circular structures in high dimensions



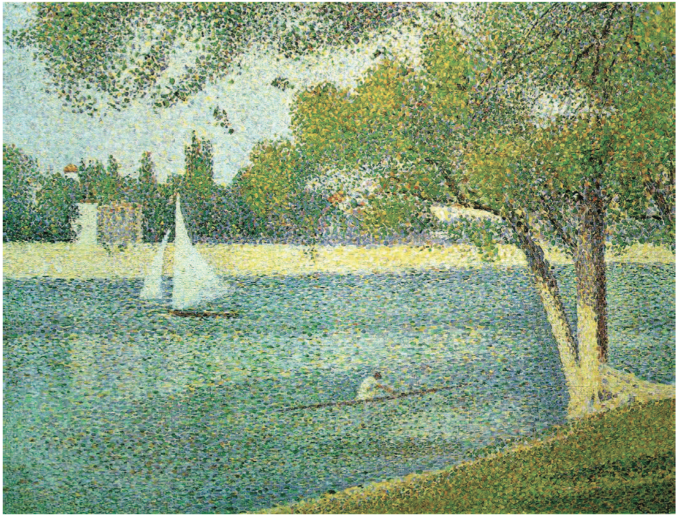
Inferring circular structures in high dimensions



Inferring circular structures in high dimensions

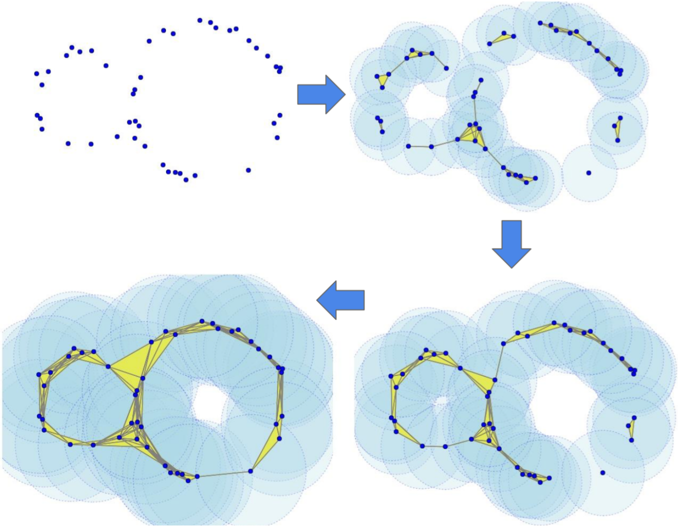


Persistent homology: an artistic view point



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...

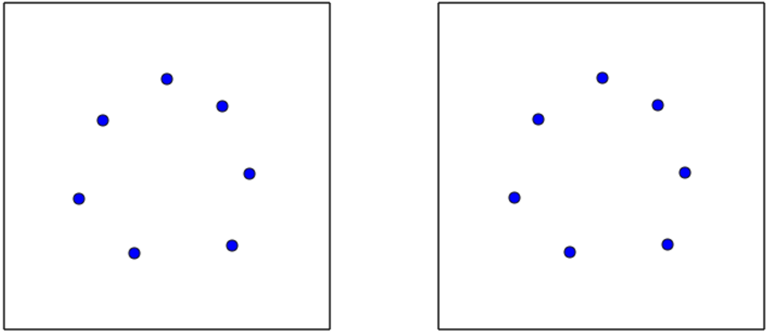
Who thinks the coffee mug and a donut is the same? Topologist!

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

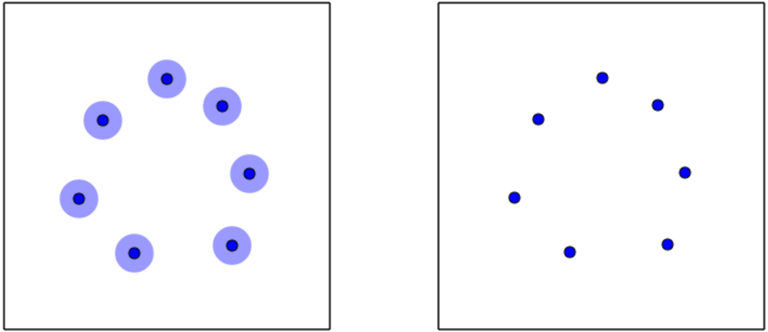
Persistent homology

- 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

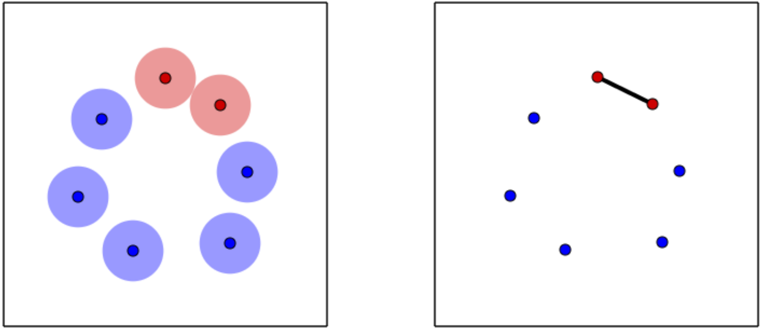
Persistent homology



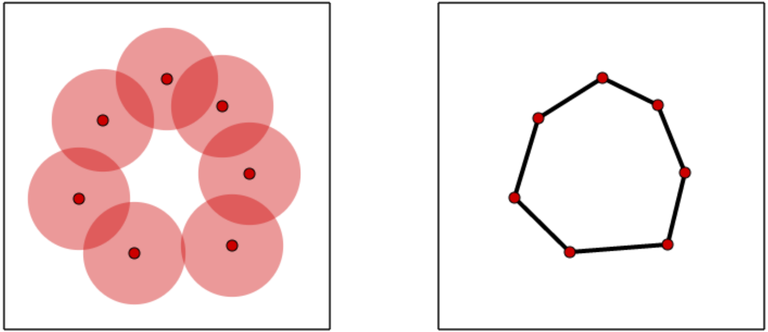
Persistent homology



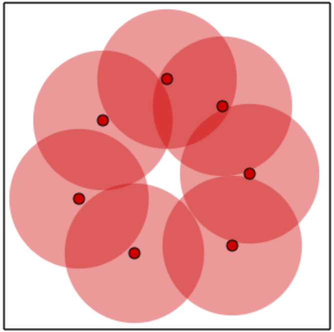
Persistent homology

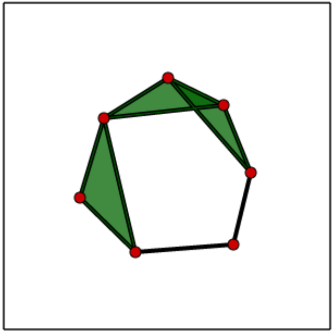


Persistent homology

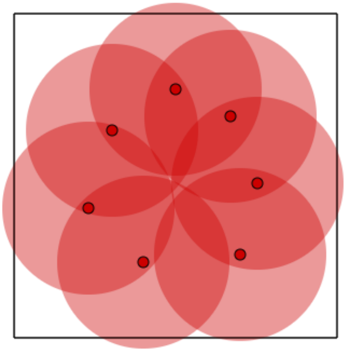


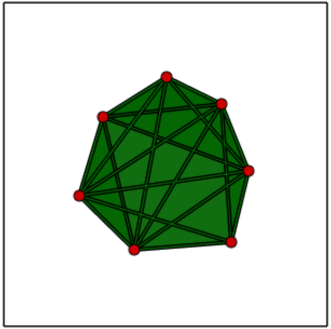
Persistent homology





Persistent homology





Persistence diagrams

Homological features encoded as barcodes or persistent diagrams

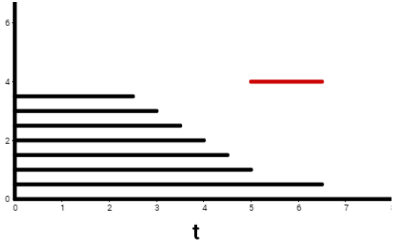


Figure: Barcode

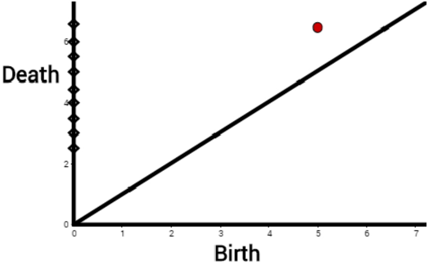
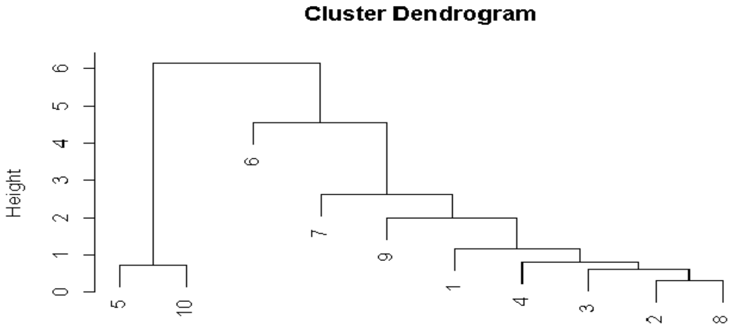


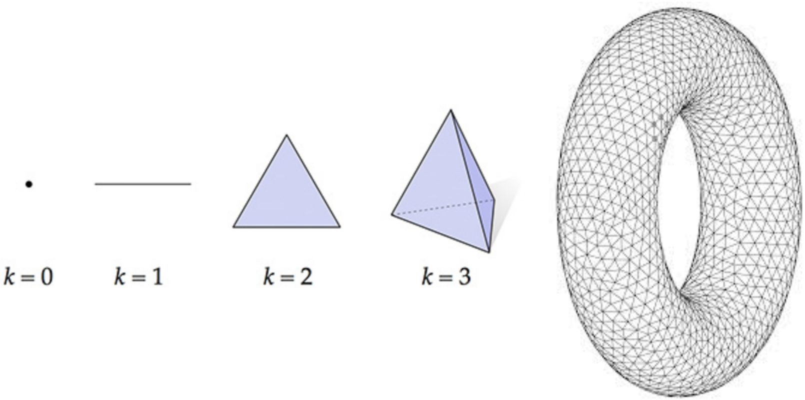
Figure: Persistence Diagram

Interpretation of connected components

Dim 0 features: hierarchical clustering



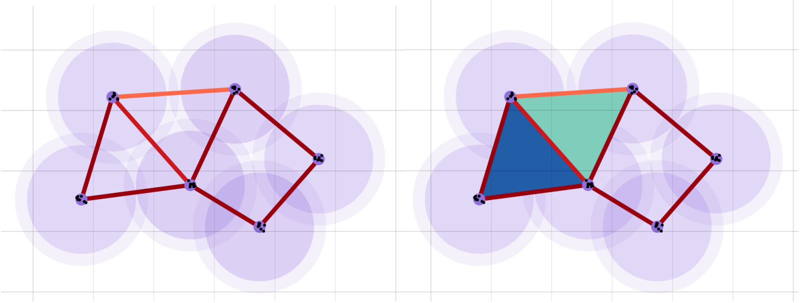
Simplicial complex



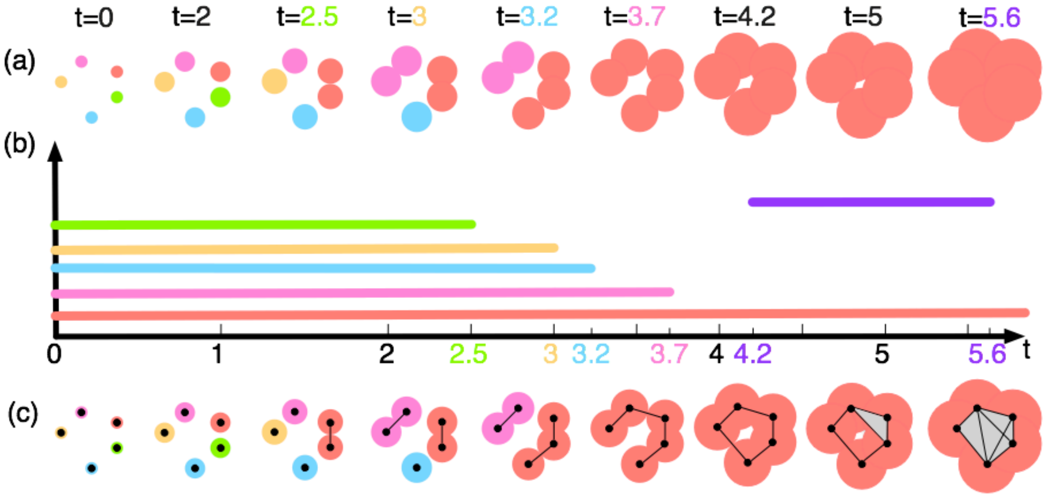
Different types of simplicial complexes

- 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



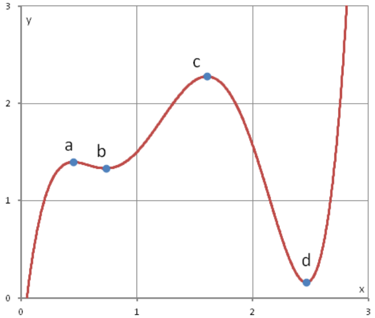
Persistent homology with Čech complex



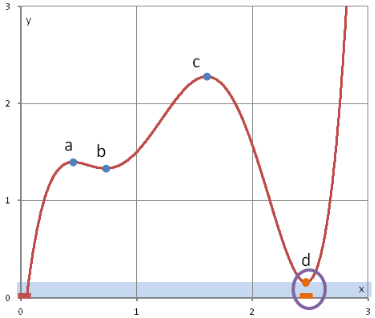
To apply persistent homology

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

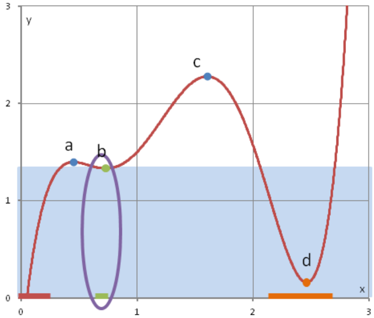
Sublevel set filtration



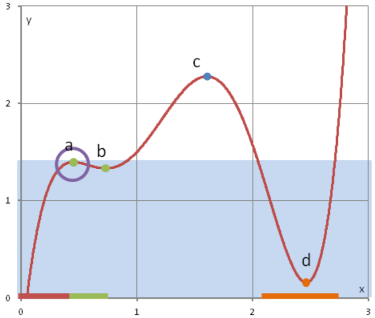
Sublevel set filtration



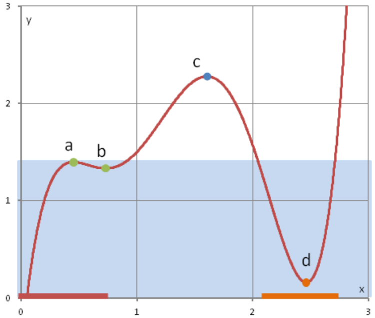
Sublevel set filtration



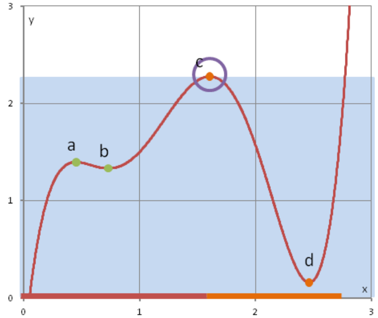
Sublevel set filtration



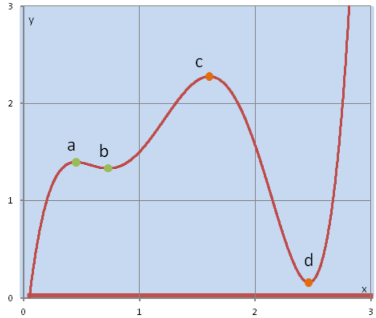
Sublevel set filtration



Sublevel set filtration



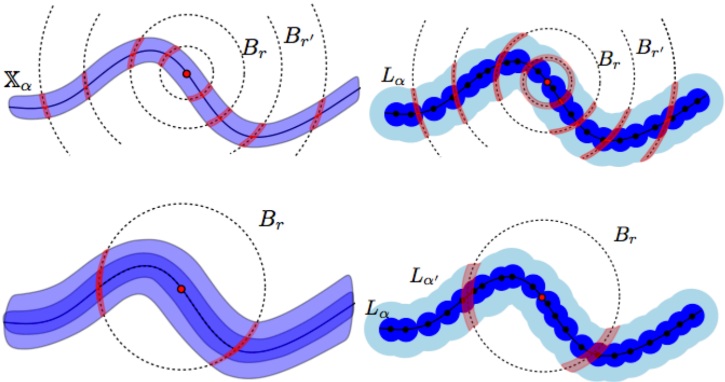
Sublevel set filtration



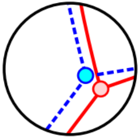
Distance between persistence diagrams

- Stability of persistence diagrams
- Bottleneck distance
- Wasserstein distance

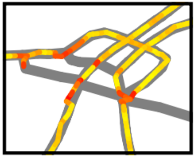
Beyond Persistent Homology


$\mathbb{X}_\alpha$
B_r
$B_{r'}$
L_α
B_r
$B_{r'}$
B_r
$L_{\alpha'}$
L_α
B_r

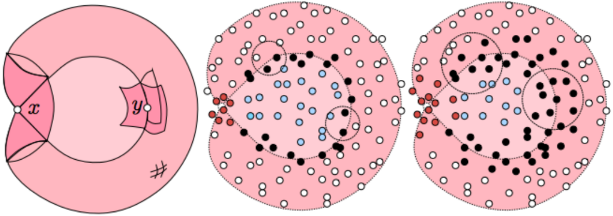
Persistent local homology: applications



(a) Local Structures



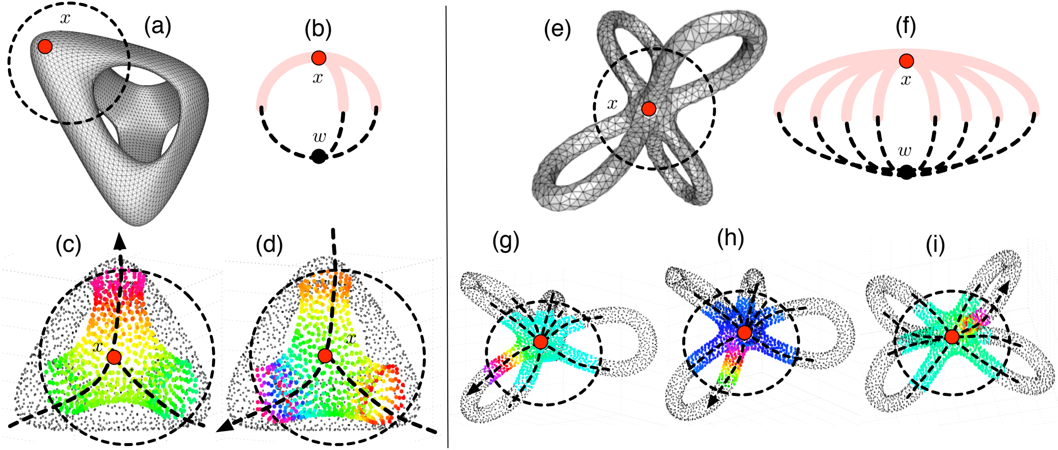
(b) Signature restricted to $\mathbb{X}$



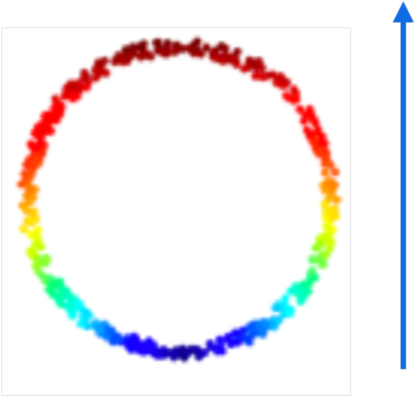
Road network comparison; stratification learning...

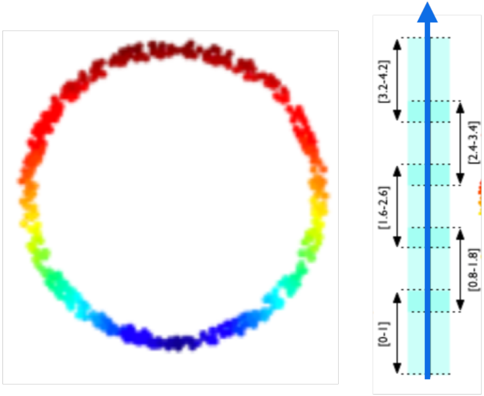
TDA and dimensionality reduction (DR)

Detecting circular and branching structures for DR



Mapper algorithm



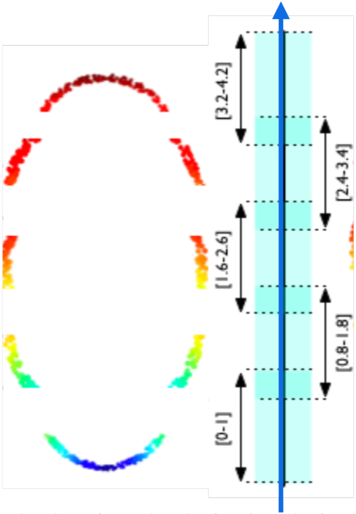


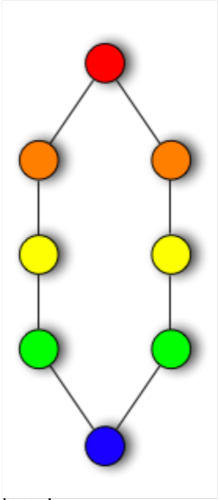
1. Input: a point cloud with a filter function e.g., a height function. Also assume that there is a distance (metric) defined between any two points in the point cloud.

2. Cover the range of the function with intervals: using # of intervals, and amount % of overlap as parameters.
E.g., # of intervals = 5, overlap = 25%.

[SinghMemoliCarlsson2007]

Mapper algorithm



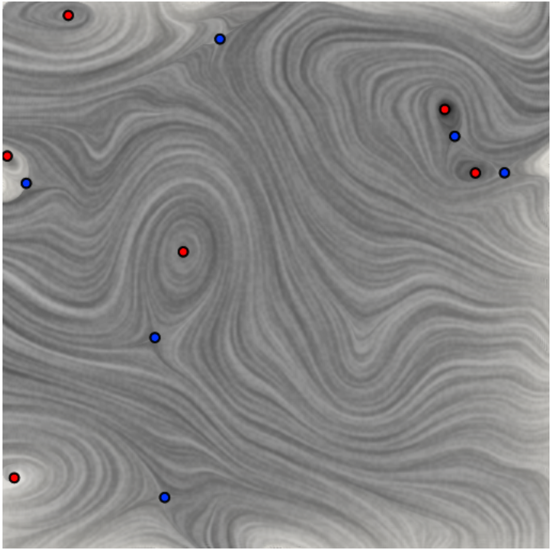


4.Obtaining the nerve of all clusters (a covering) in the domain. E.g., here it is a graph representation that summarizes the data.

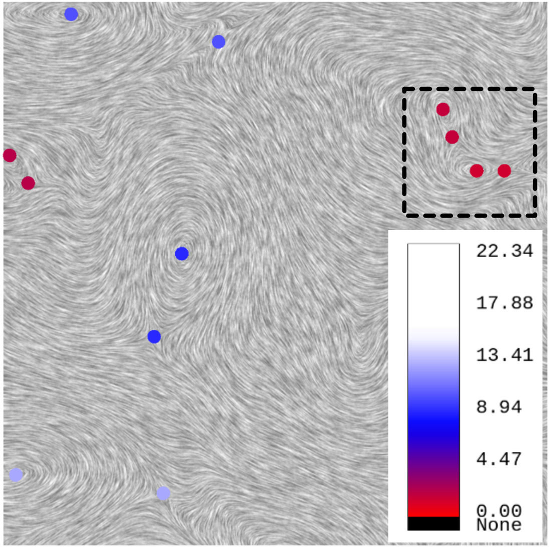
Such a graph can interface with machine learning and interactive visualization…

3. Look at the points in the domain that falls into each interval, and apply clustering to these points.
E.g., following the inverse map.

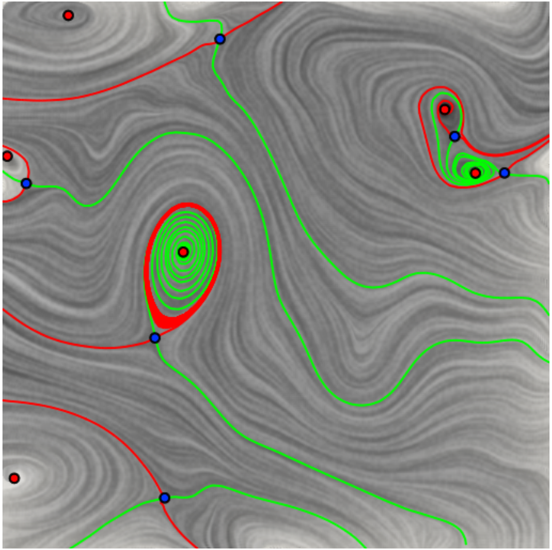
Robust feature extraction and simplification



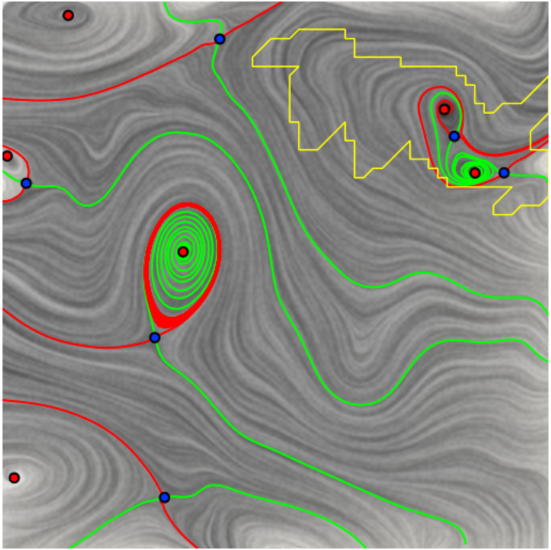
Robust feature extraction and simplification



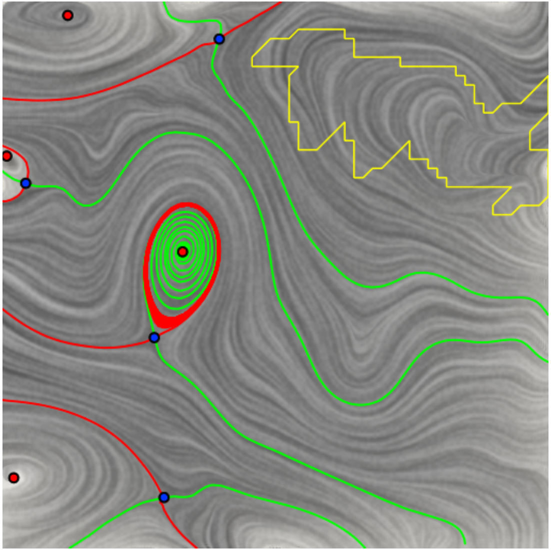
Robust feature extraction and simplification



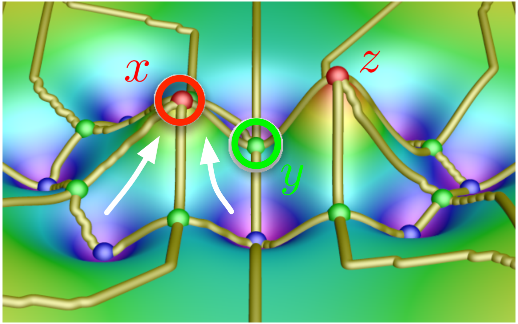
Robust feature extraction and simplification



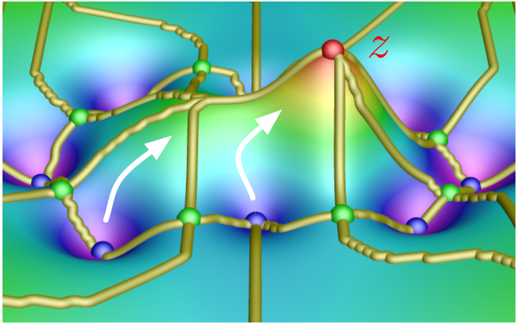
Robust feature extraction and simplification



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]