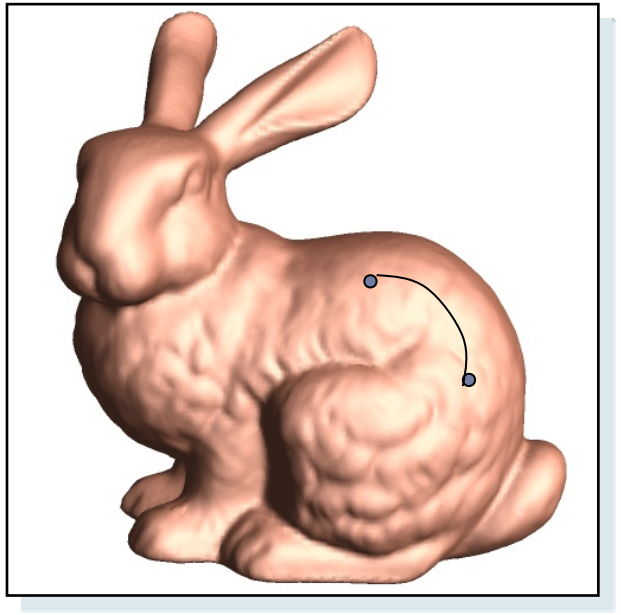
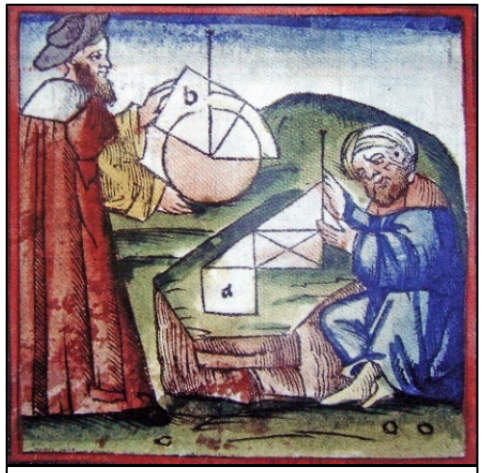
DSC 214
Topological Data Analysis
Lecture 0: Introduction

Instructor: Zhengchao Wan

Geometry = Geo (earth) + metry (measure)

- Distances and angles
  - area, volume, curvatures, etc
- Euclidean geometry
- Riemannian geometry
  - Hyperbolic geometry
  - Spherical geometry
- ...





Using geometry, by unknown artist, 15th century

Topology

- Detailed geometric information not sufficient
  - Too local
  - Or not necessary
  - Or may even be harmful
  - Wish to identify key information, “qualitative” structure

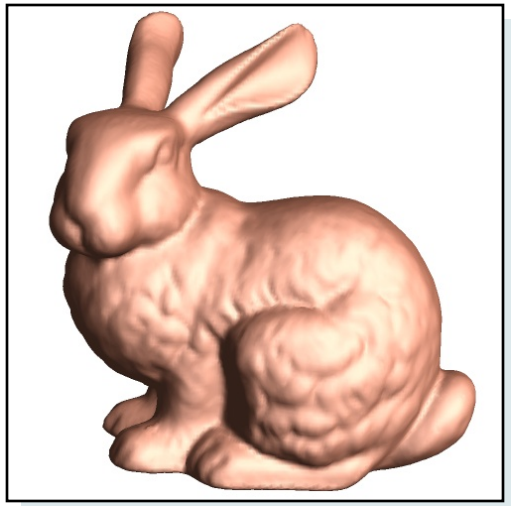
- Topology
  - Connectivity
  - Holes, Voids

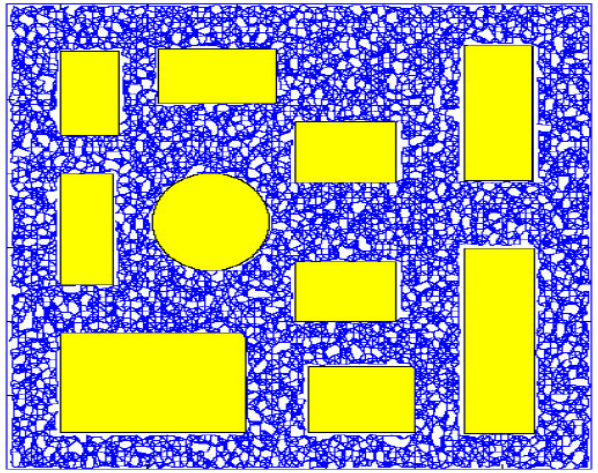
Topology

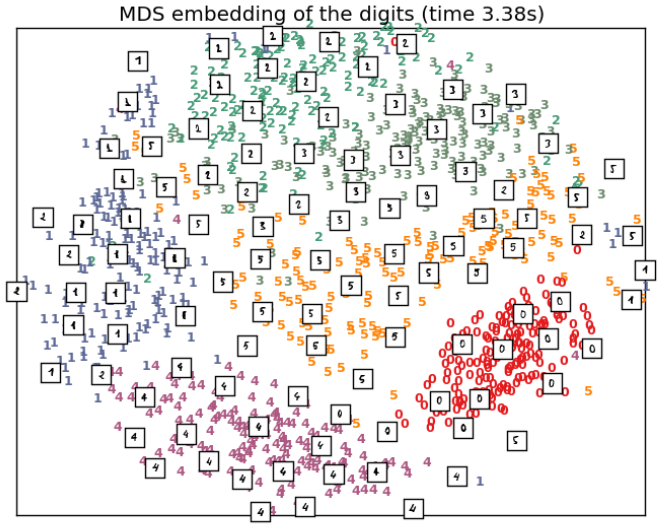
- Detailed geometric information not sufficient
  - Too local
  - Or not necessary
  - Or may even be harmful
  - Wish to identify key information, “qualitative” structure

- Topology
  - Connectivity
  - Holes, Voids

Data has shape!







Data has shape!

Data has shape!

- TGDA: Topological and geometric data analysis
  - Develop effective methodologies and algorithms for data analysis through the topological and geometric lens.

Data has shape!

- TGDA: Topological and geometric data analysis
  - Develop effective methodologies and algorithms for data analysis through the topological and geometric lens.

- This course: *Topological data analysis* (TDA)
  - Focus on topological concepts, structures, and algorithms for data analysis

TDA - a brief introduction

- Aims at extracting and analyzing topological information from data sets
  - Coarse yet essential information
  - Characterization, feature identification
  - General, powerful tools for both space and functions defined on a space
  - Elegant mathematical understanding available
- Young and active (started around 2000)
- Interdisciplinary: computer science, mathematics, statistics
- Many applications to Data Science
  - Graphics, visualization, medical image processing, computer vision, computational neuron science, computational biology, material science
  - TDA + machine learning

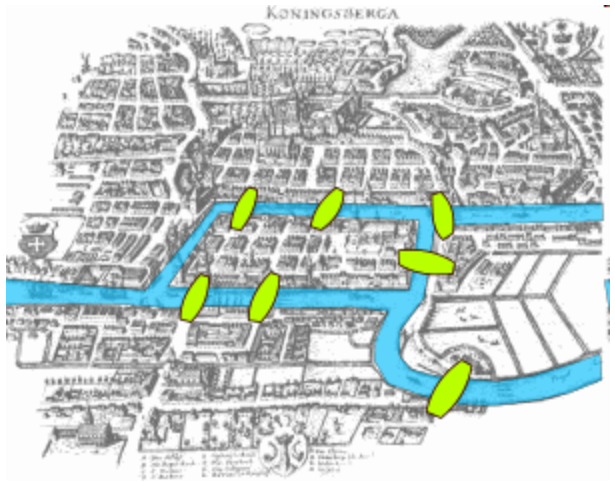
Introduction to Topology

History

- Seven Bridges of Königsberg (1736)

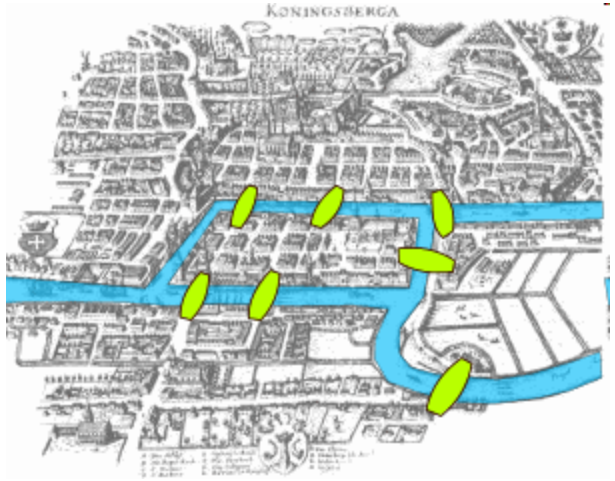
History

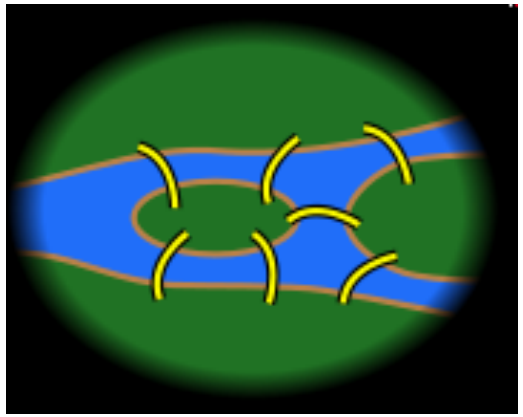
- Seven Bridges of Königsberg (1736)



History

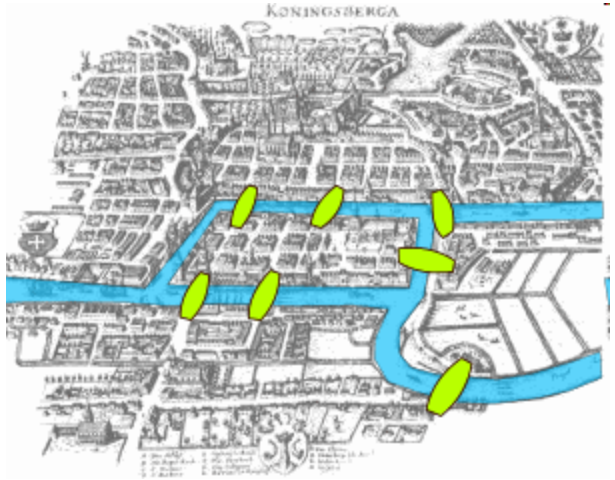
- Seven Bridges of Königsberg (1736)

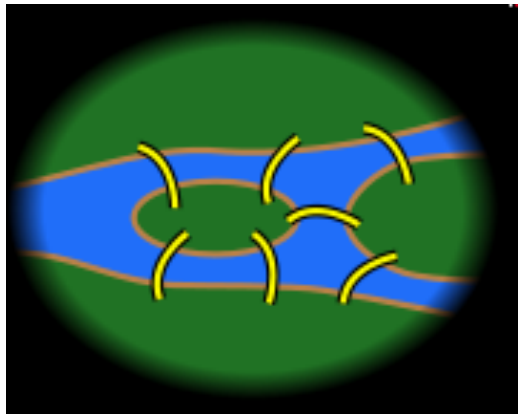


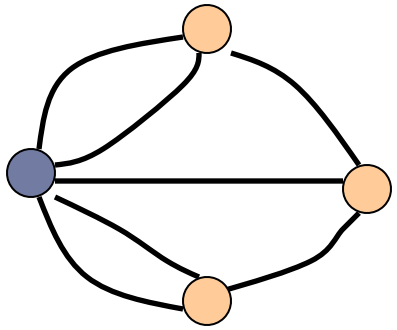


History

- Seven Bridges of Königsberg (1736)

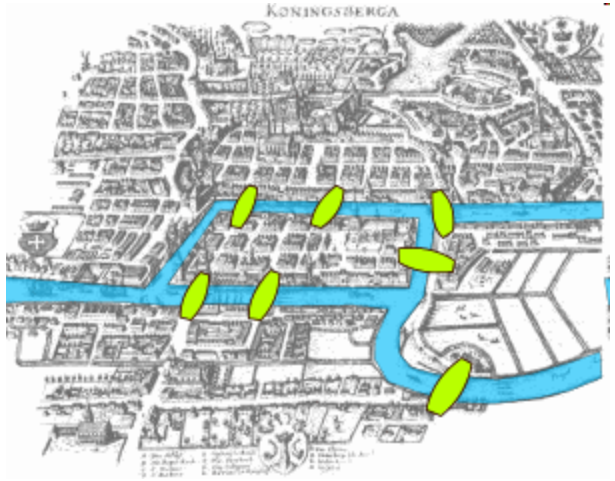


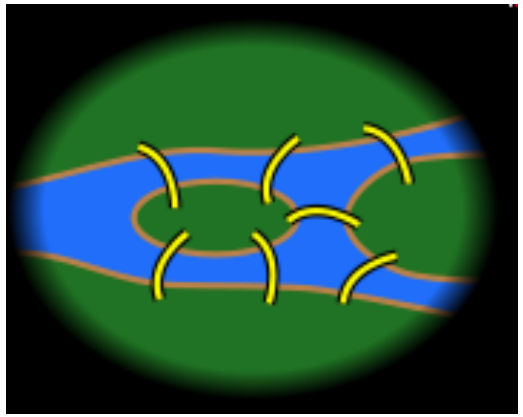


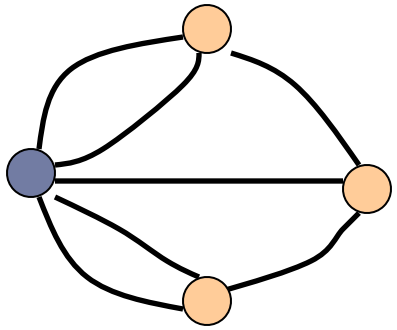


History

- Seven Bridges of Königsberg (1736)



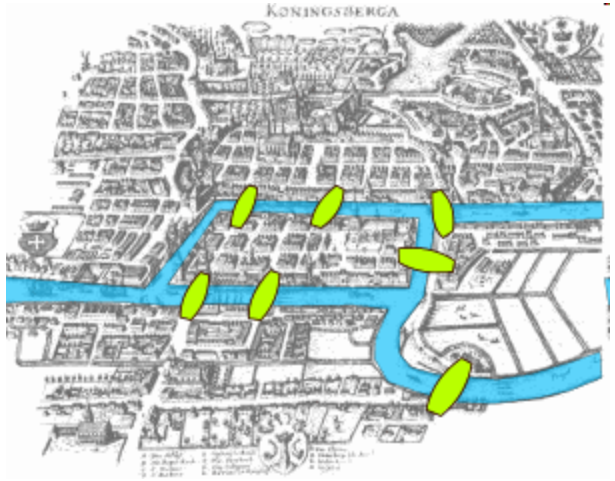


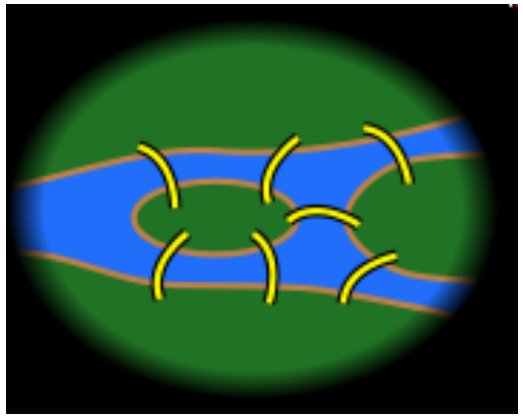


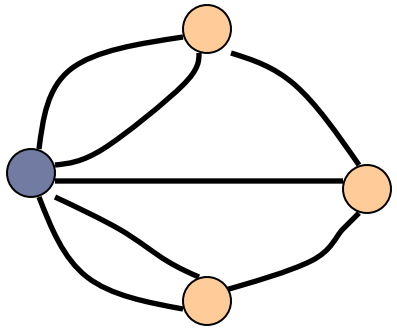
Euler cycle problem

History

- Seven Bridges of Königsberg (1736)





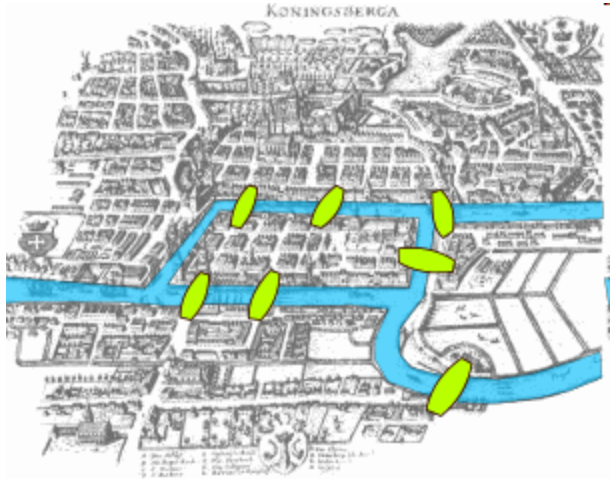


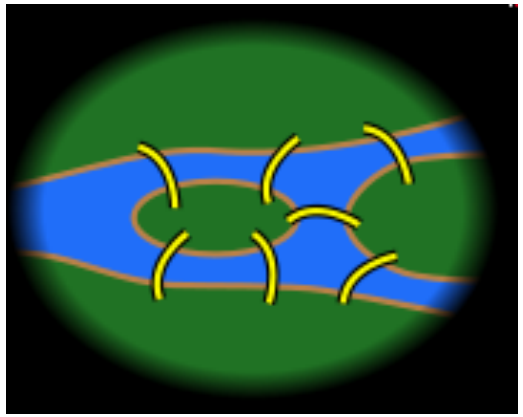
Euler cycle problem

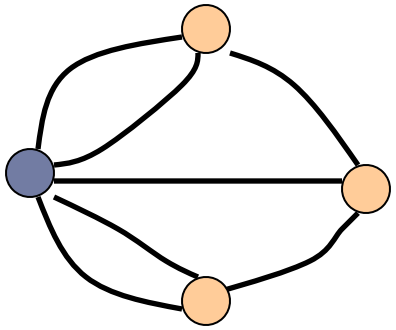
Abstraction of connectivity

History

- Seven Bridges of Königsberg (1736)







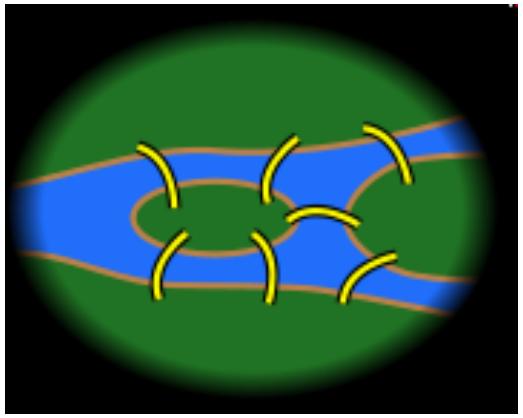
Euler cycle problem

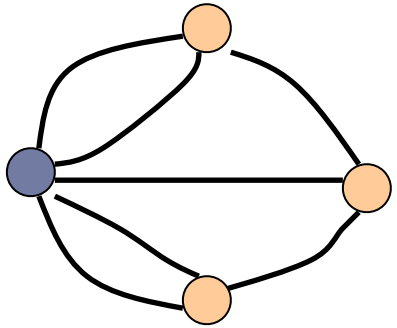
Abstraction of connectivity

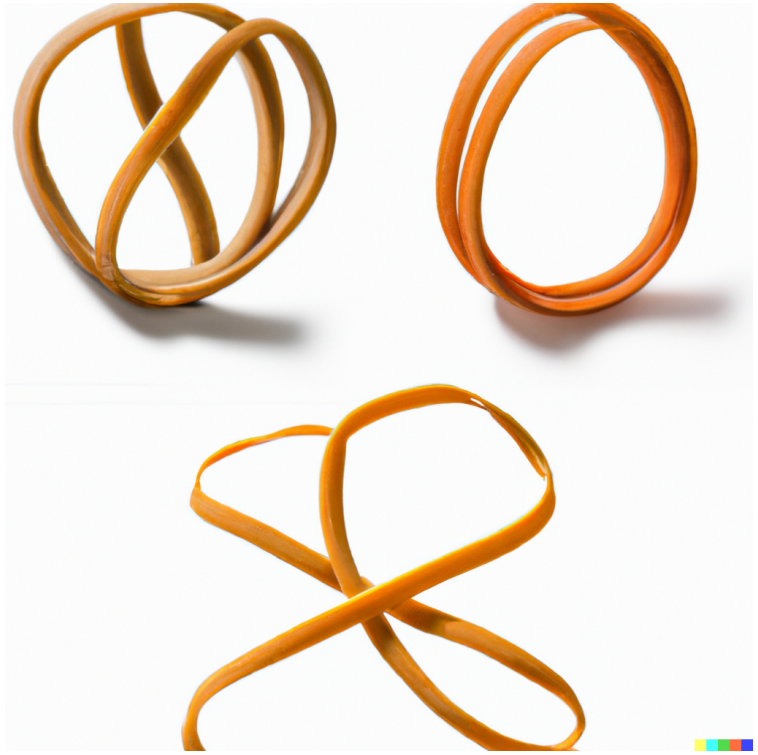
Topology: “ distinguish qualitative geometry from the ordinary geometry in which quantitative relations chiefly are treated ”

Fundamental Questions

- How do you know when the two spaces are the “same”?
- Why things are the “same” after deformation?

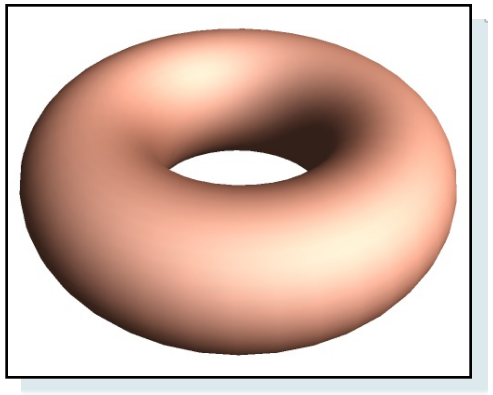






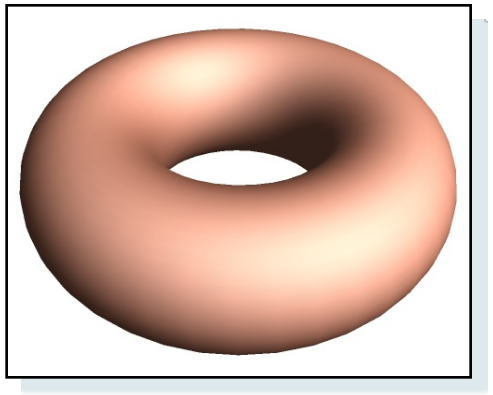
Homeomorphism

- Intuitively, two spaces have the same topology if one can continuously deform one to the other without breaking, gluing, and inserting new things



Homeomorphism

- Intuitively, two spaces have the same topology if one can continuously deform one to the other without breaking, gluing, and inserting new things

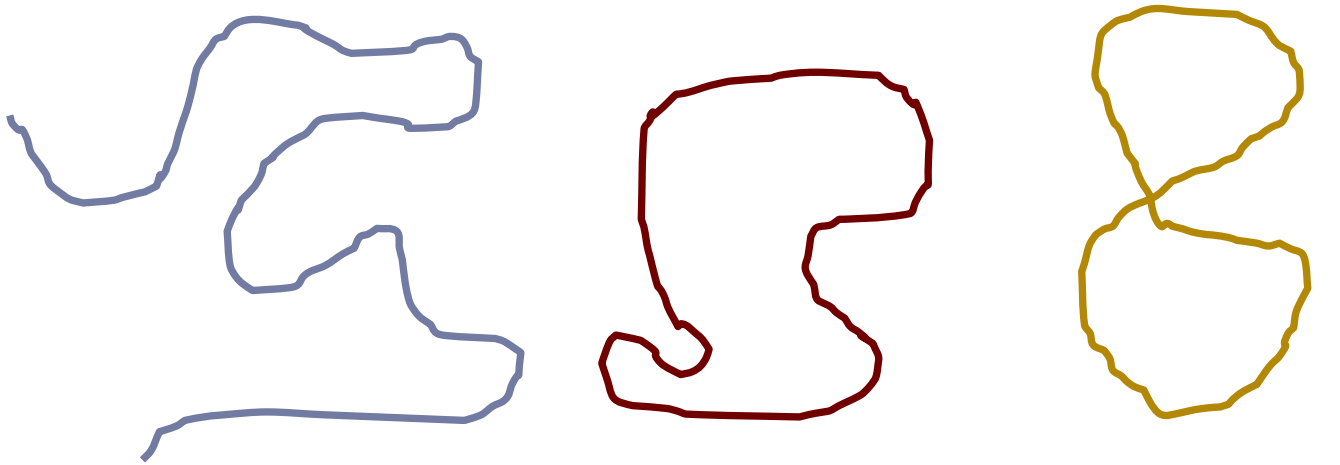


Homeomorphism

- Intuitively, two spaces have the same topology if one can continuously deform one to the other without breaking, gluing, and inserting new things

Homeomorphism

- Intuitively, two spaces have the same topology if one can continuously deform one to the other without breaking, gluing, and inserting new things



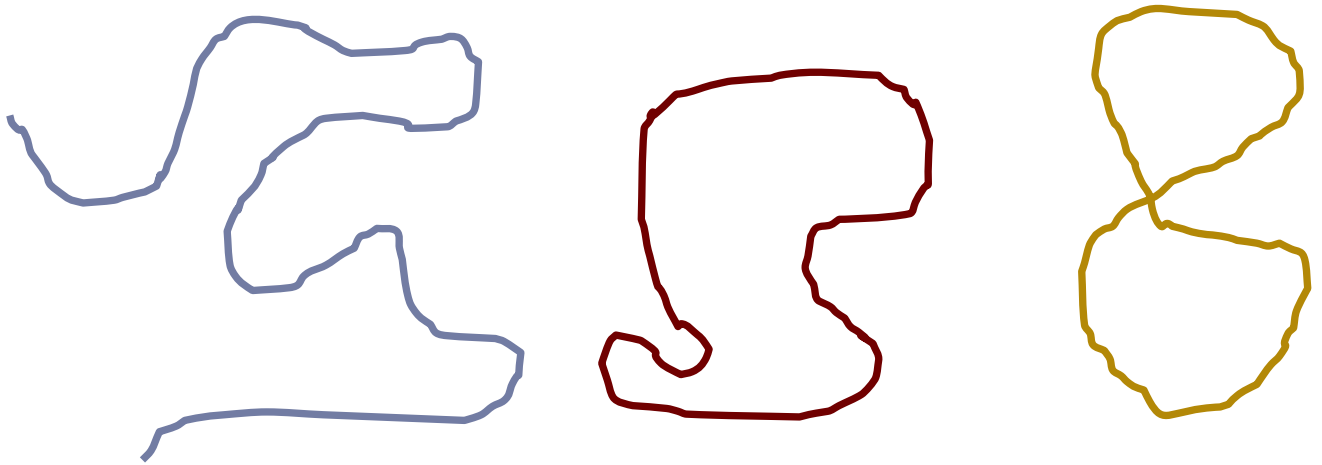
open curve

closed curve

self-intersecting curve

Homeomorphism

- Intuitively, two spaces have the same topology if one can continuously deform one to the other without breaking, gluing, and inserting new things



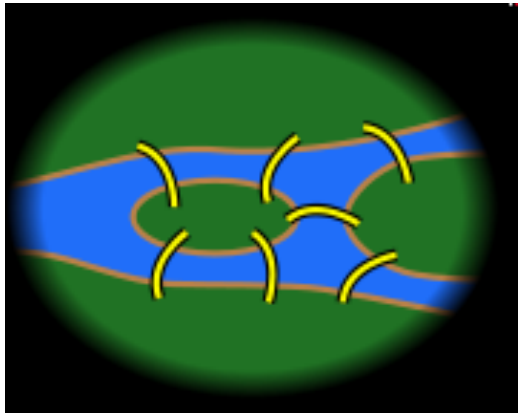
open curve

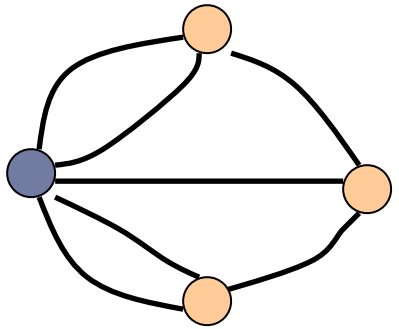
closed curve

self-intersecting curve

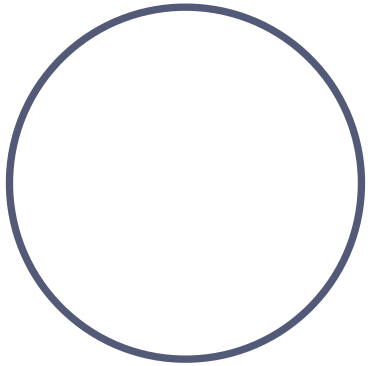
Two spaces with the same topology are ***homeomorphic***

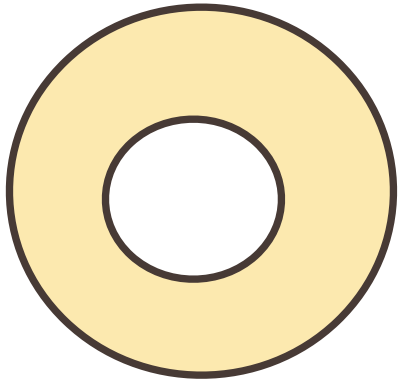
Homotopy Equivalent - A Relaxation of Homeomorphism

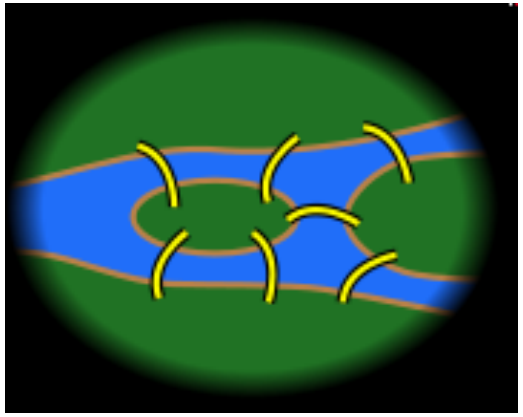


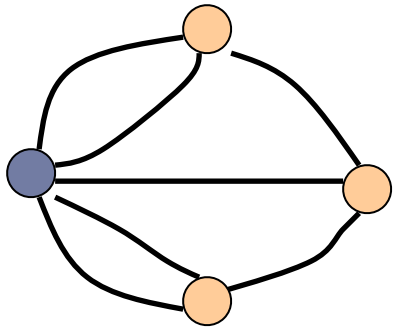


Homotopy Equivalent - A Relaxation of Homeomorphism

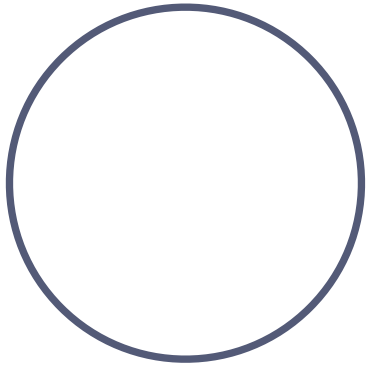


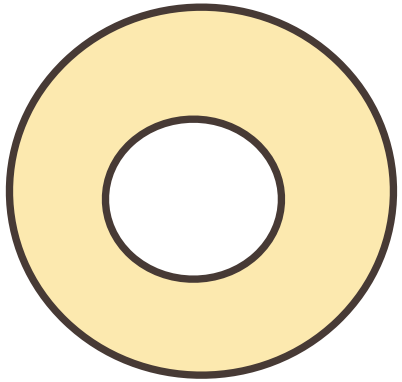


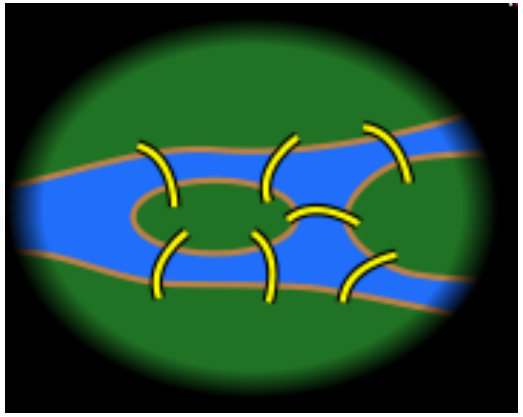


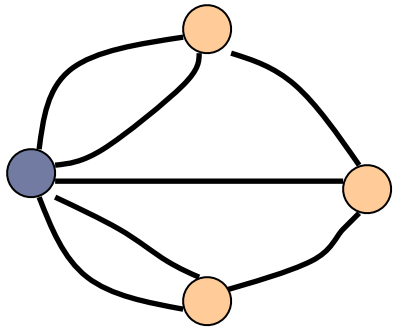


Homotopy Equivalent - A Relaxation of Homeomorphism





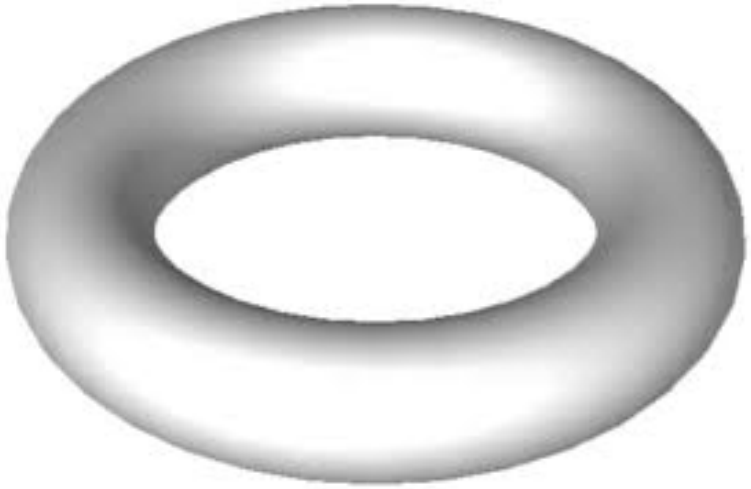




Homotopy equivalent - “same” in a topologist’s perspective

Homology - Algebraic Tools for Determining Homotopy

- Homology can count # connected components, # holes, # voids.



- The above two spaces are not homotopy equivalent

Topological Quantities

Topological Quantities

- Homeomorphism $\rightarrow$ homotopy equivalence
  - Different levels of similarity between spaces

Topological Quantities

- Homeomorphism ⟶ homotopy equivalence
  - Different levels of similarity between spaces

- Quantities invariant under them (topological quantities)

Topological Quantities

- Homeomorphism —› homotopy equivalence
  - Different levels of similarity between spaces

- Quantities invariant under them (topological quantities)

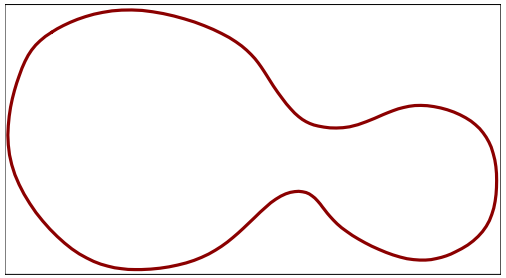
  ⇒ *(Essential) features*

  - Make topologic objects (such homology) powerful for feature identification and characterization
  - This course will give
    - definitions, intuition, and their computation
    - also examples of applications

TDA - Principle

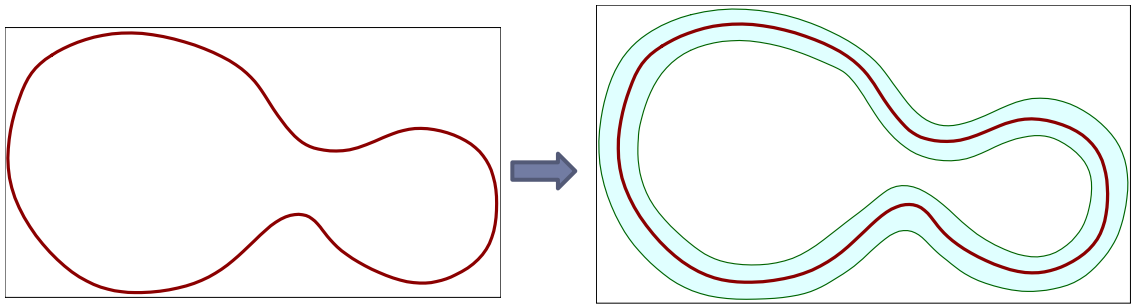
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



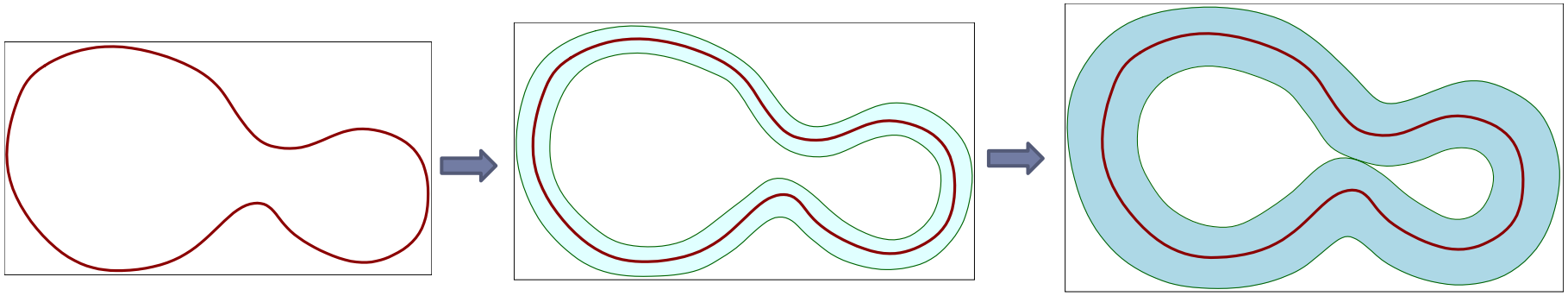
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



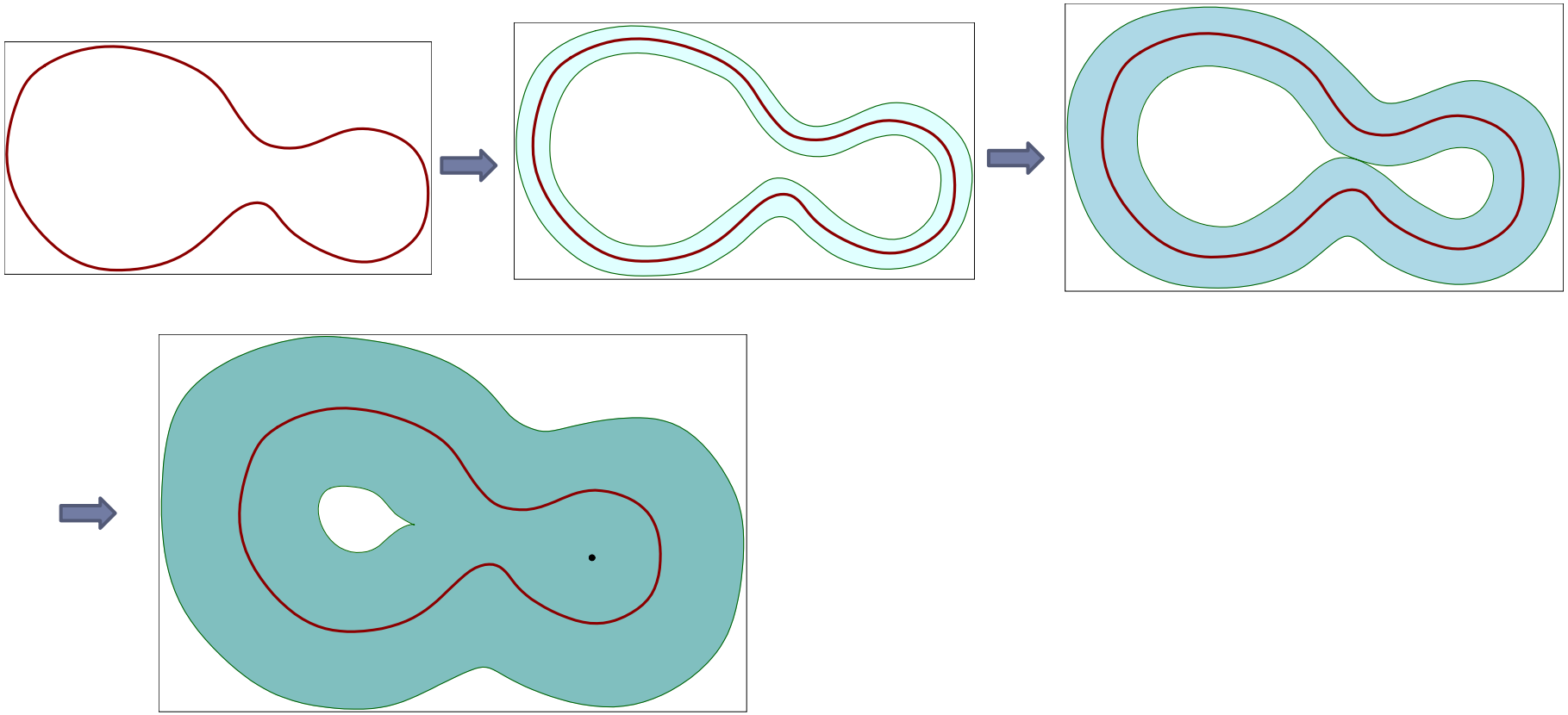
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



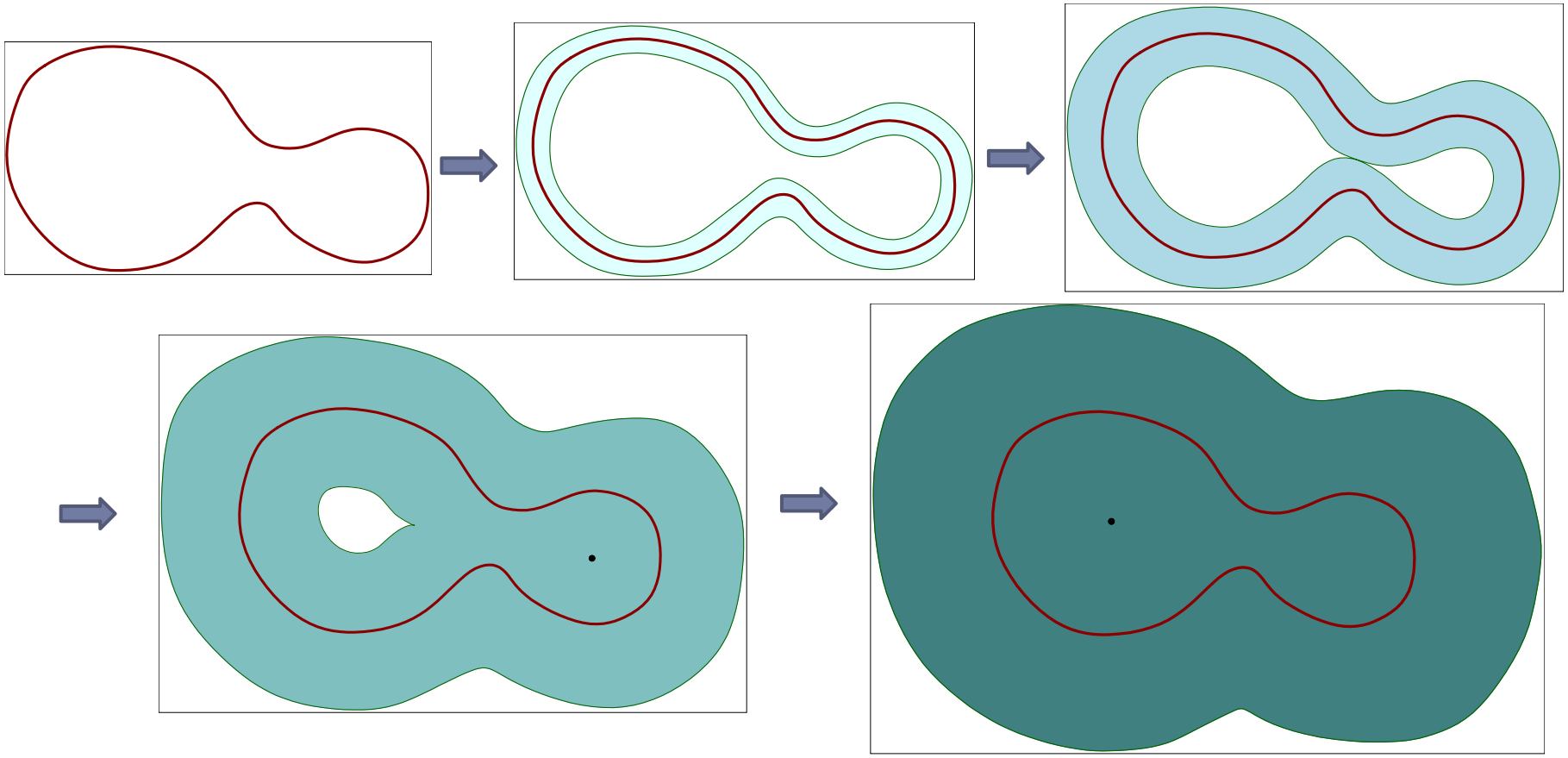
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



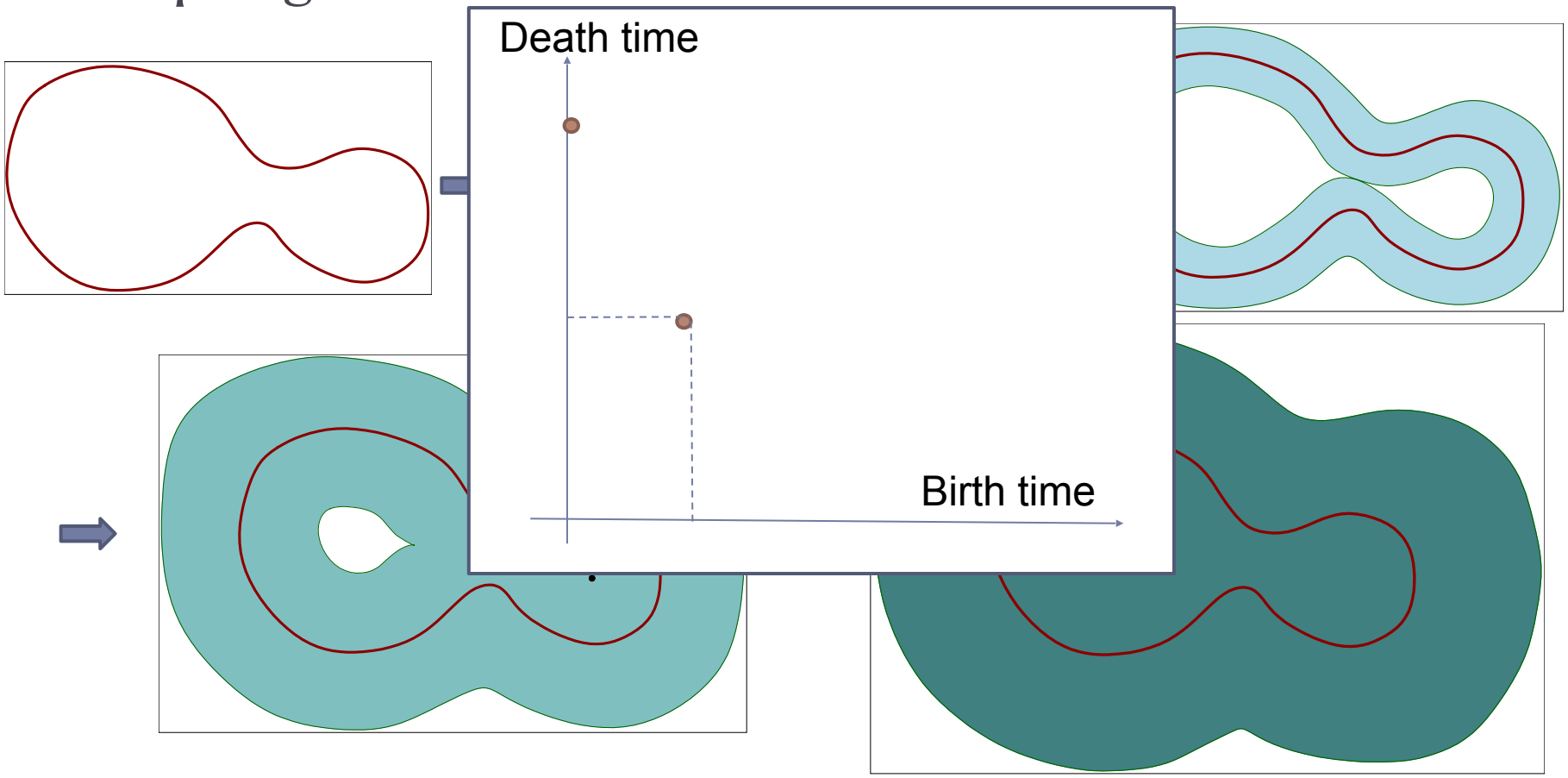
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



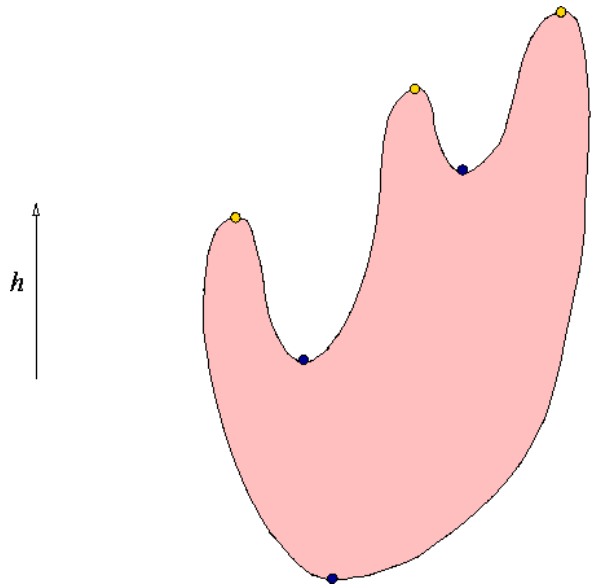
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features

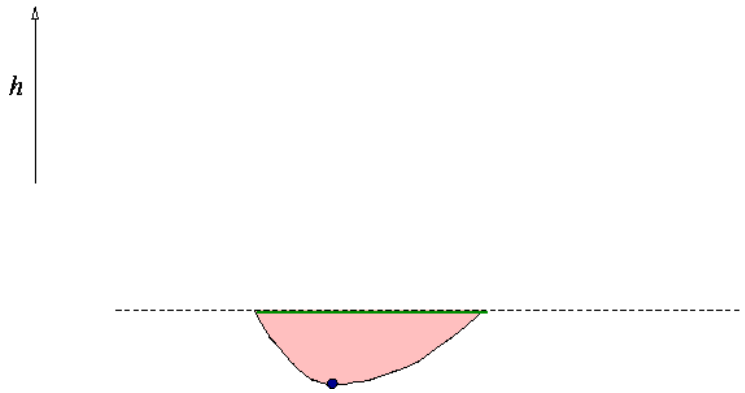


TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features

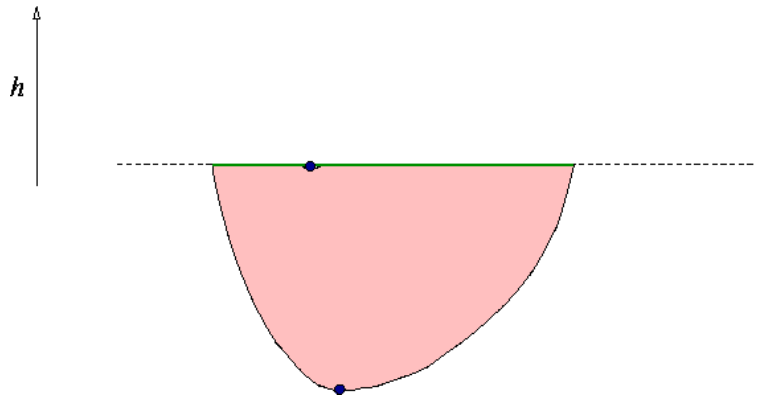
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



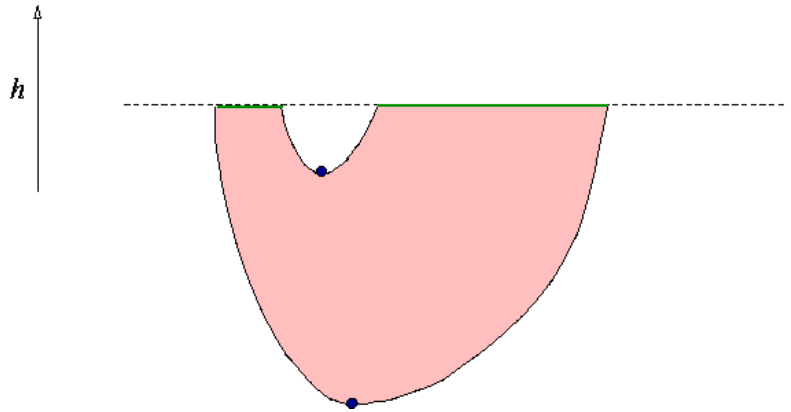
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



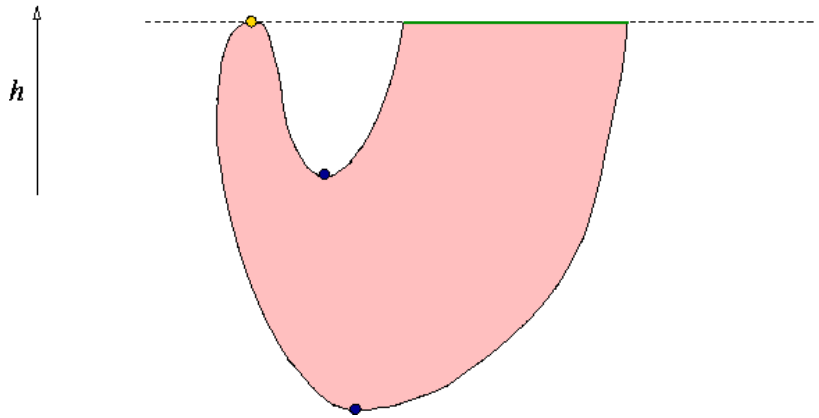
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



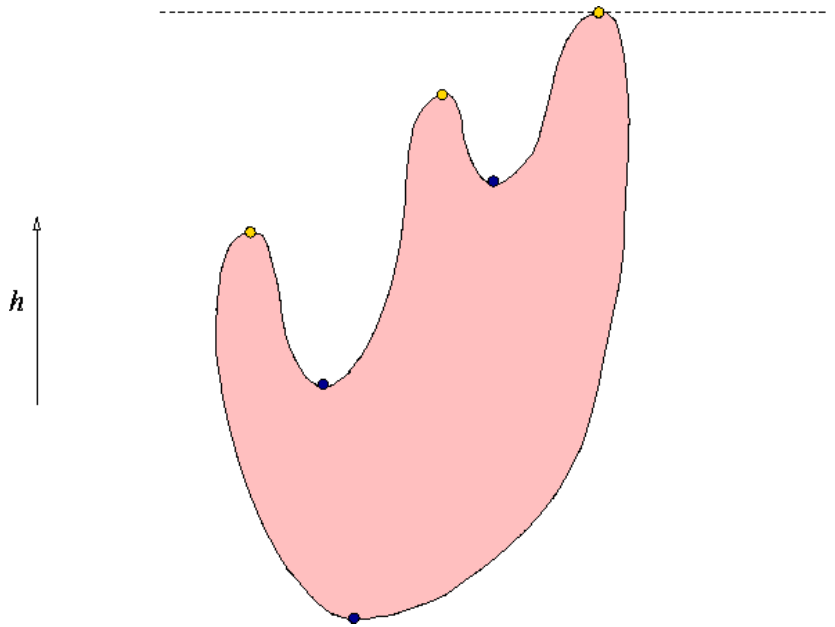
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



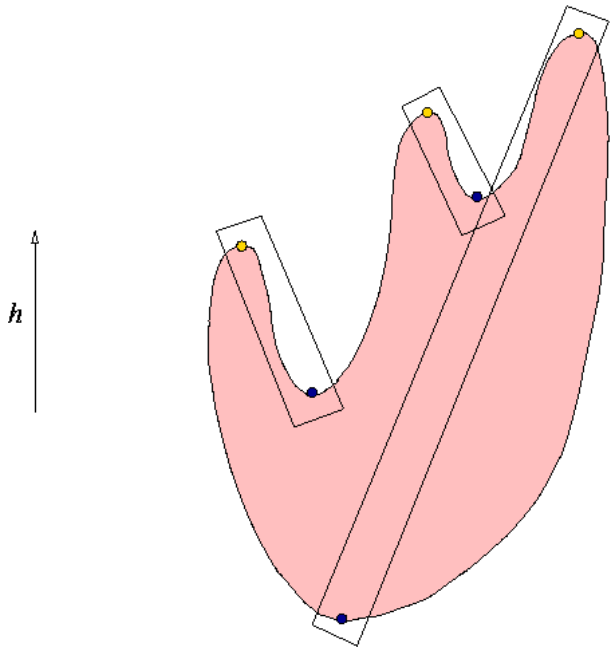
TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



TDA Principle - Tracking Multiscale Topology

- Incorporate geometry, functions or maps of a space to create and capture multiscale topological features



Topics

- Basics in Topology
- Common complexes (*aka. how we model space of interests*)
- (Simplicial) homology (*aka. how do we quantify topological features*)
- Persistent homology (*aka. powerful modern extension of homological features*)
- Analysis of point cloud data (PCD) and graph data
- Analysis of functions on data
- Discrete Morse theory (*higher order skeletonal structure behind data*) and applications
- TDA and machine learning

Topics

- Basics in Topology
- Common complexes (*aka. how we model space of interests*)
- (Simplicial) homology (*aka. how do we quantify topological features*)
- Persistent homology (*aka. powerful modern extension of homological features*)
- Analysis of point cloud data (PCD) and graph data
- Analysis of functions on data
- Discrete Morse theory (*higher order skeletonal structure behind data*) and applications
- TDA

Will focus on concepts, definitions, algorithms, also intuition why they work, and how they can be used.

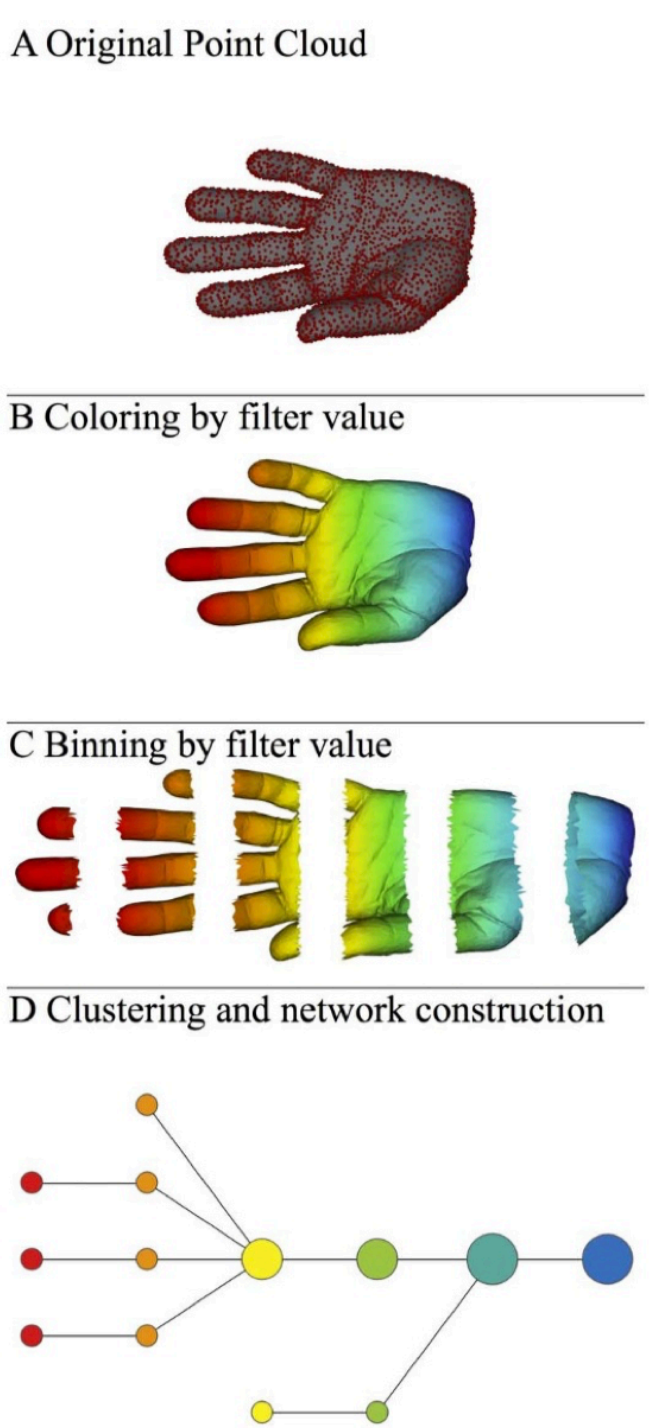
TDA in Data Science

Motivating Examples I

- Shape of data
  - Topological summary

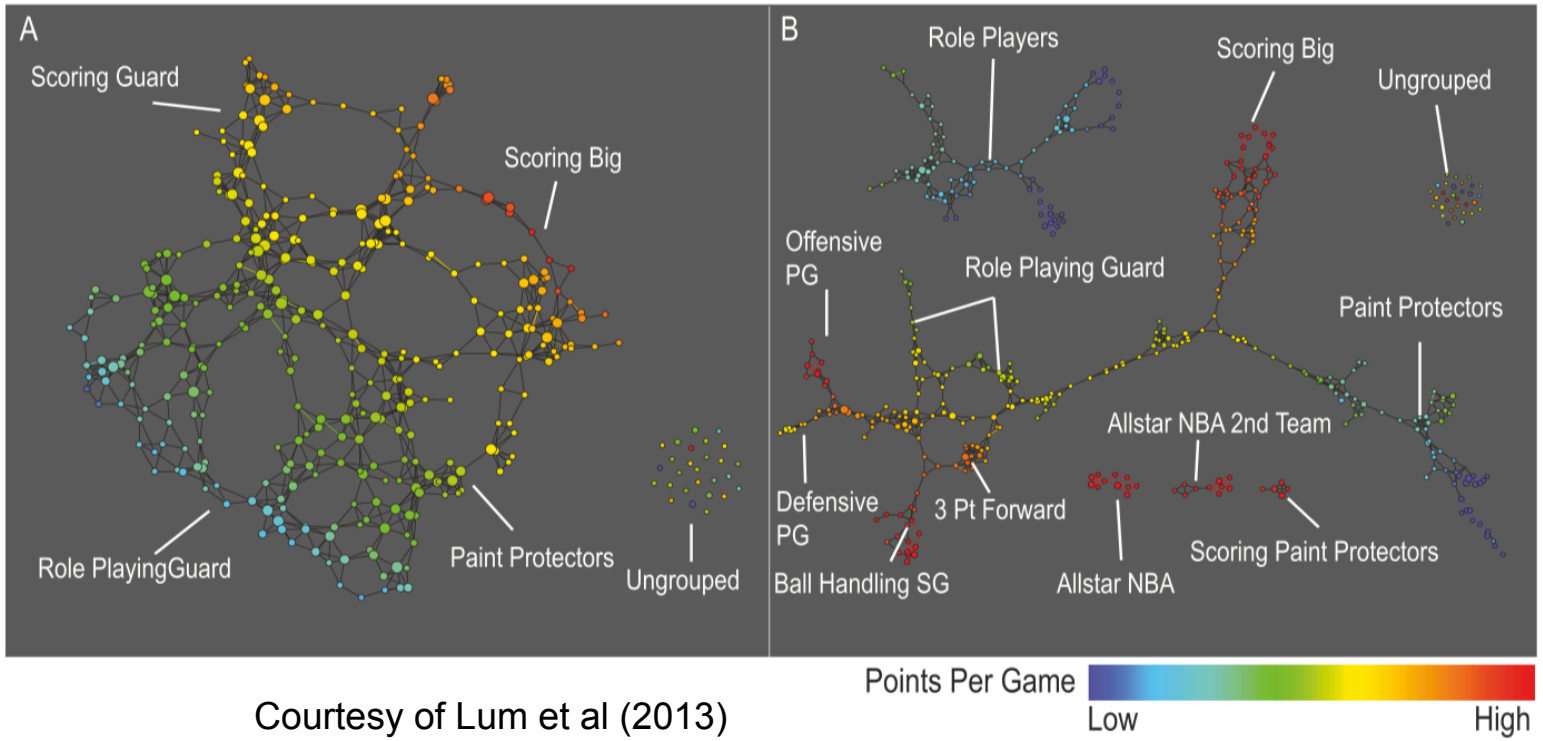
Motivating Examples I

- Shape of data
  - Topological summary

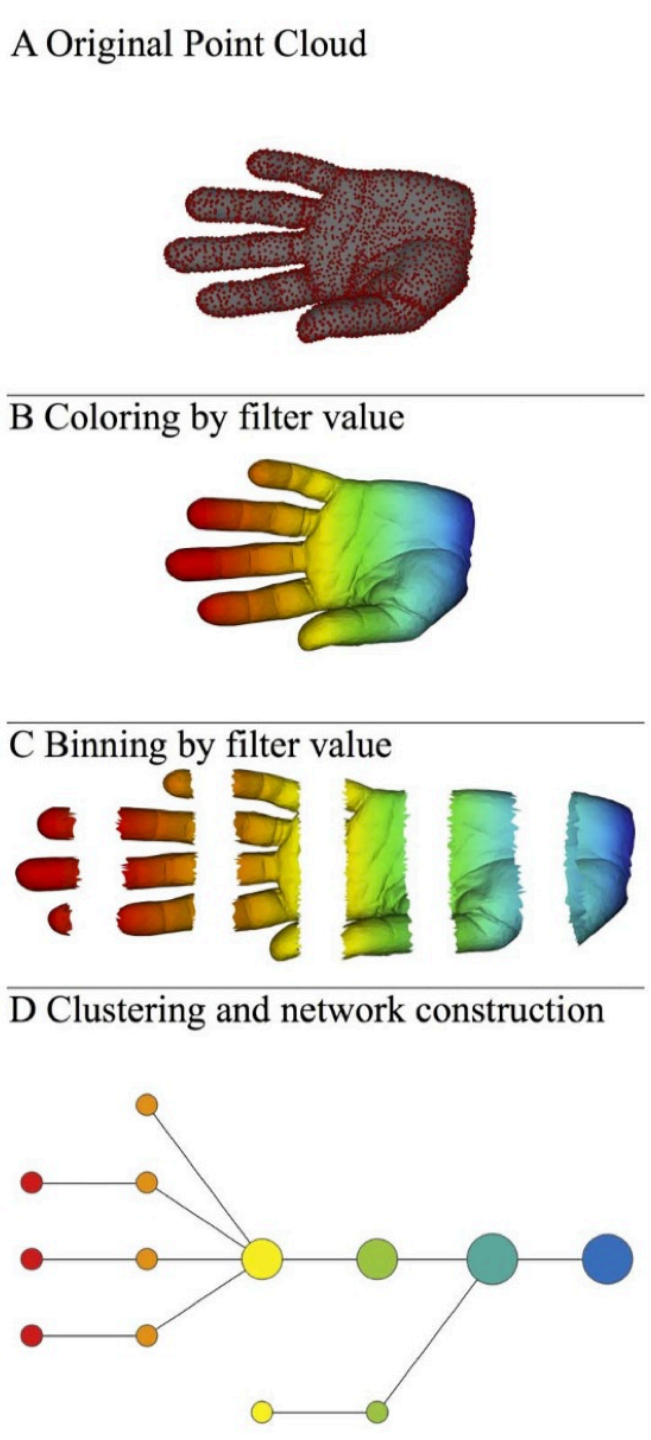


Motivating Examples I

- Shape of data
  - Topological summary



Courtesy of Lum et al (2013)

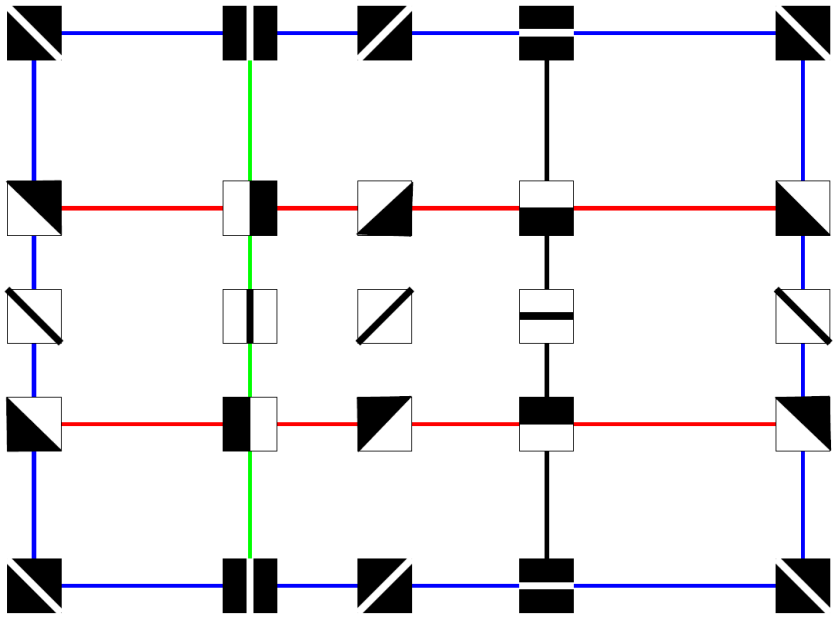


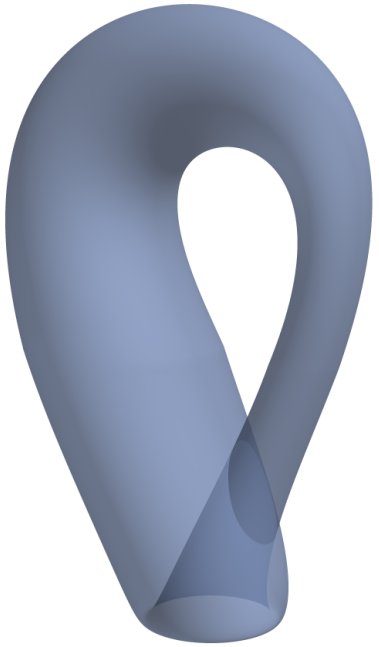
Motivating Examples II

- Computer Vision
  - Clustering
  - Shape space

Motivating Examples II

- Computer Vision
  - Clustering
  - Shape space





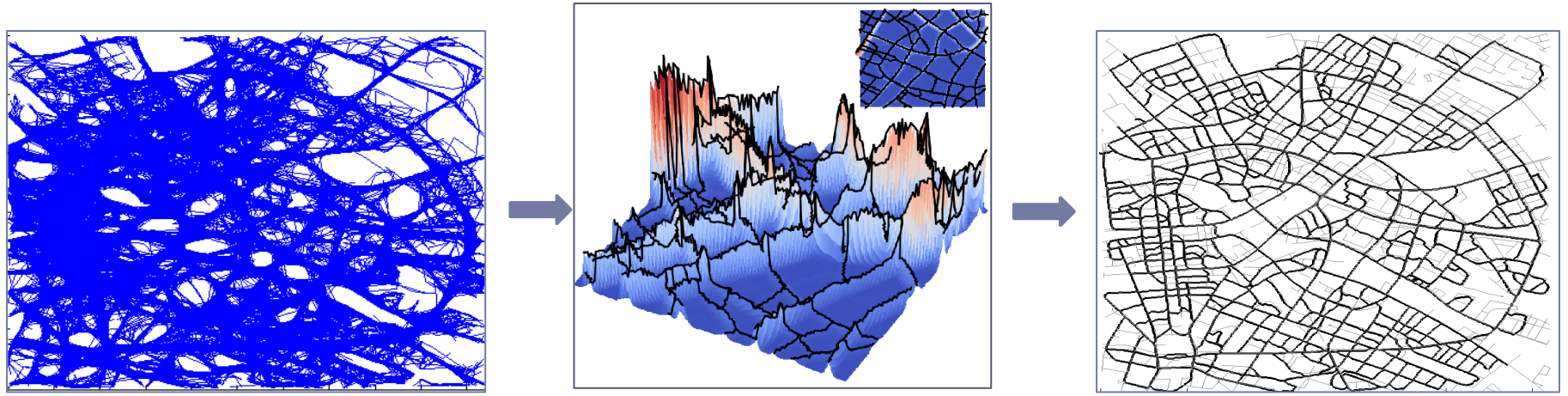
Courtesy of Carlsson et al, *(2008)*

Motivating Examples IV

- Graph reconstruction
  - Road network reconstruction, neuron skeletonization, etc

Motivating Examples IV

- Graph reconstruction
  - Road network reconstruction, neuron skeletonization, etc



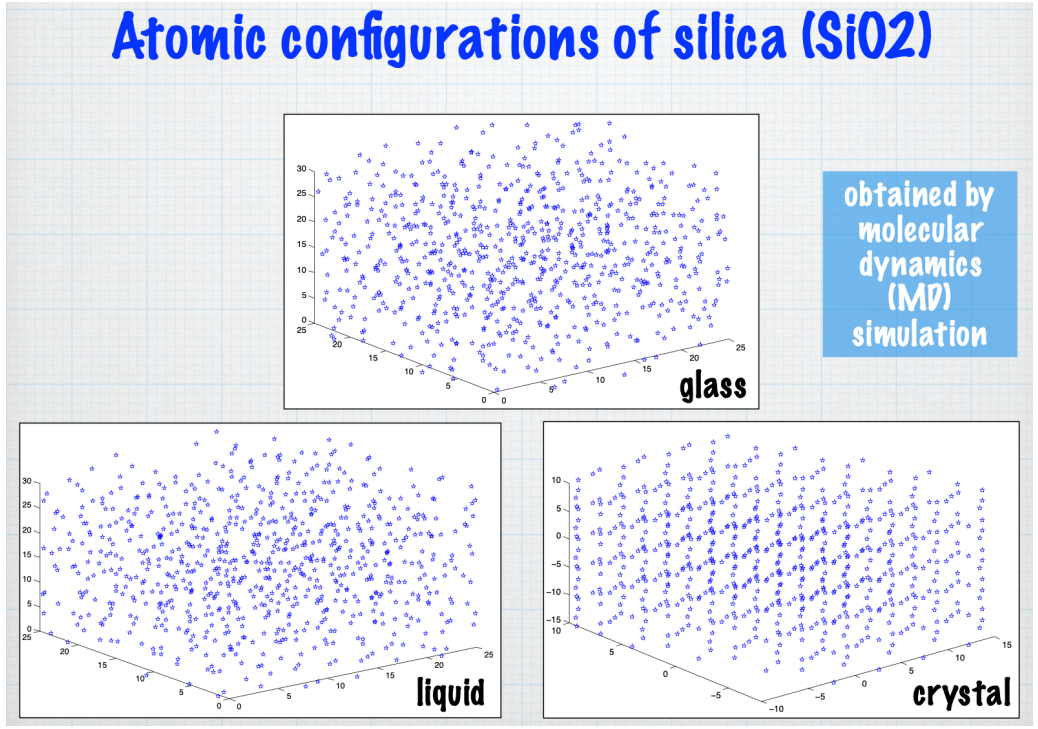
Input: GPS trajectories

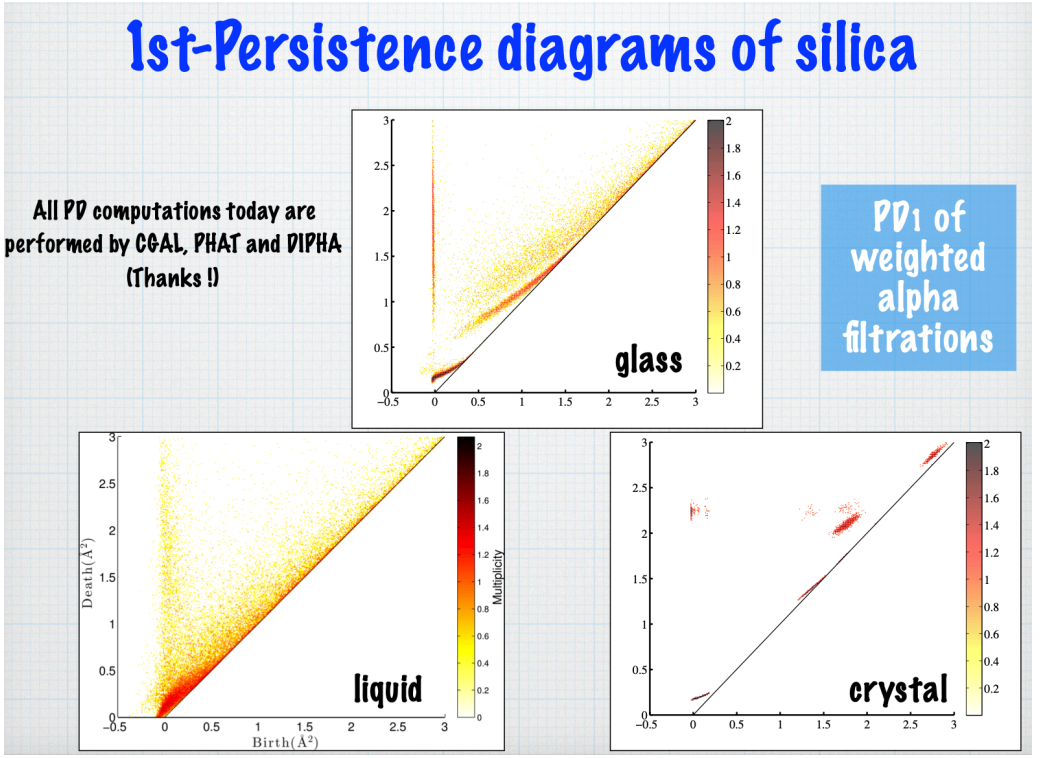
density field + discrete Morse

Output: Road network

Motivating Examples V

- Material Science



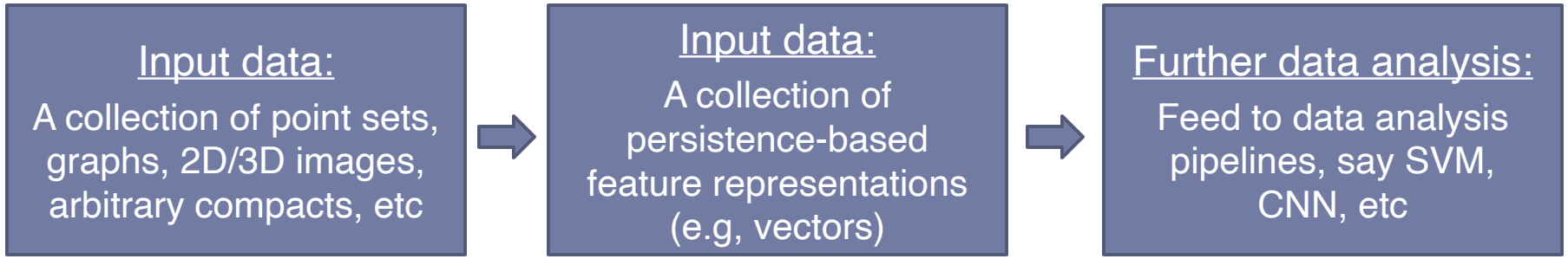


Nakamura et al. 2015

TDA in Machine Learning

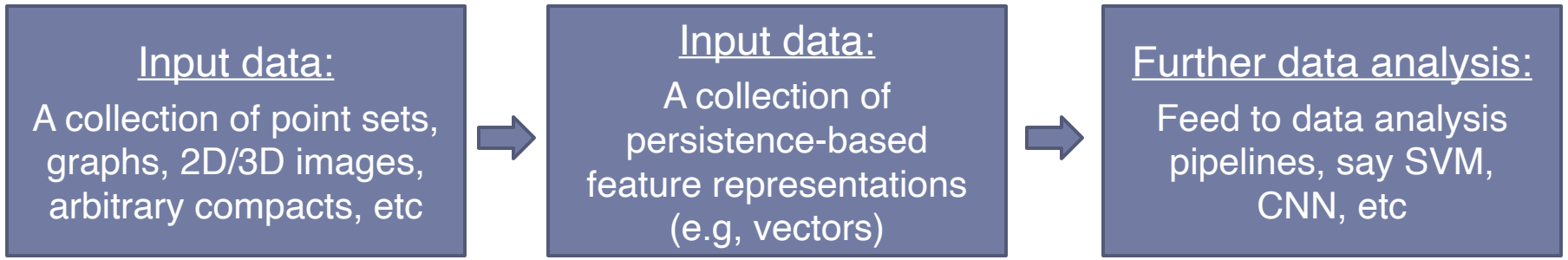
Motivating Examples VI

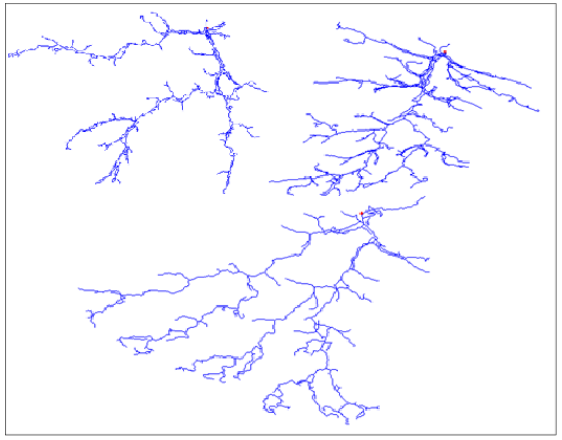
- Persistence-based feature vectorization + machine learning
  - Persistence images, kernels for persistence diagrams, etc



Motivating Examples VI

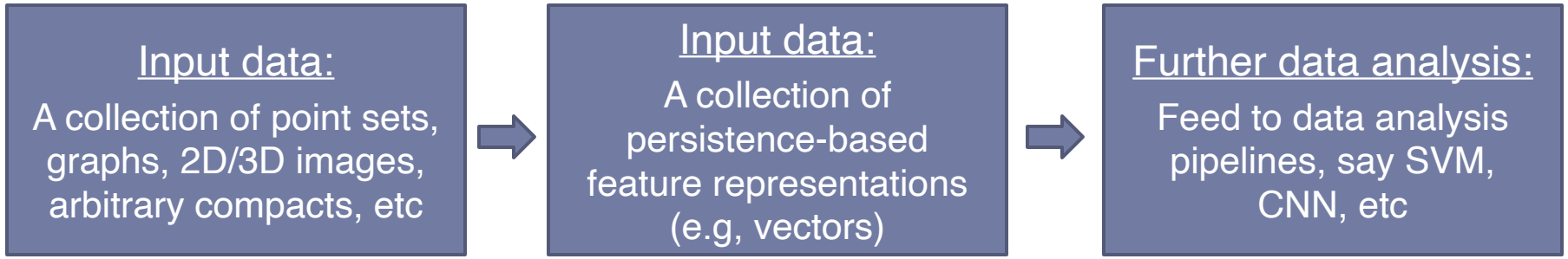
- Persistence-based feature vectorization + machine learning
  - Persistence images, kernels for persistence diagrams, etc

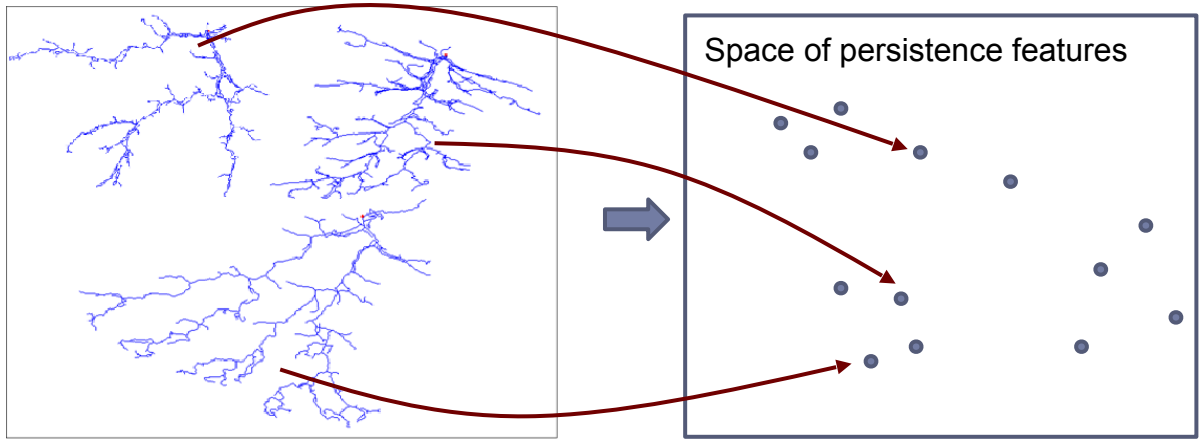




Motivating Examples VI

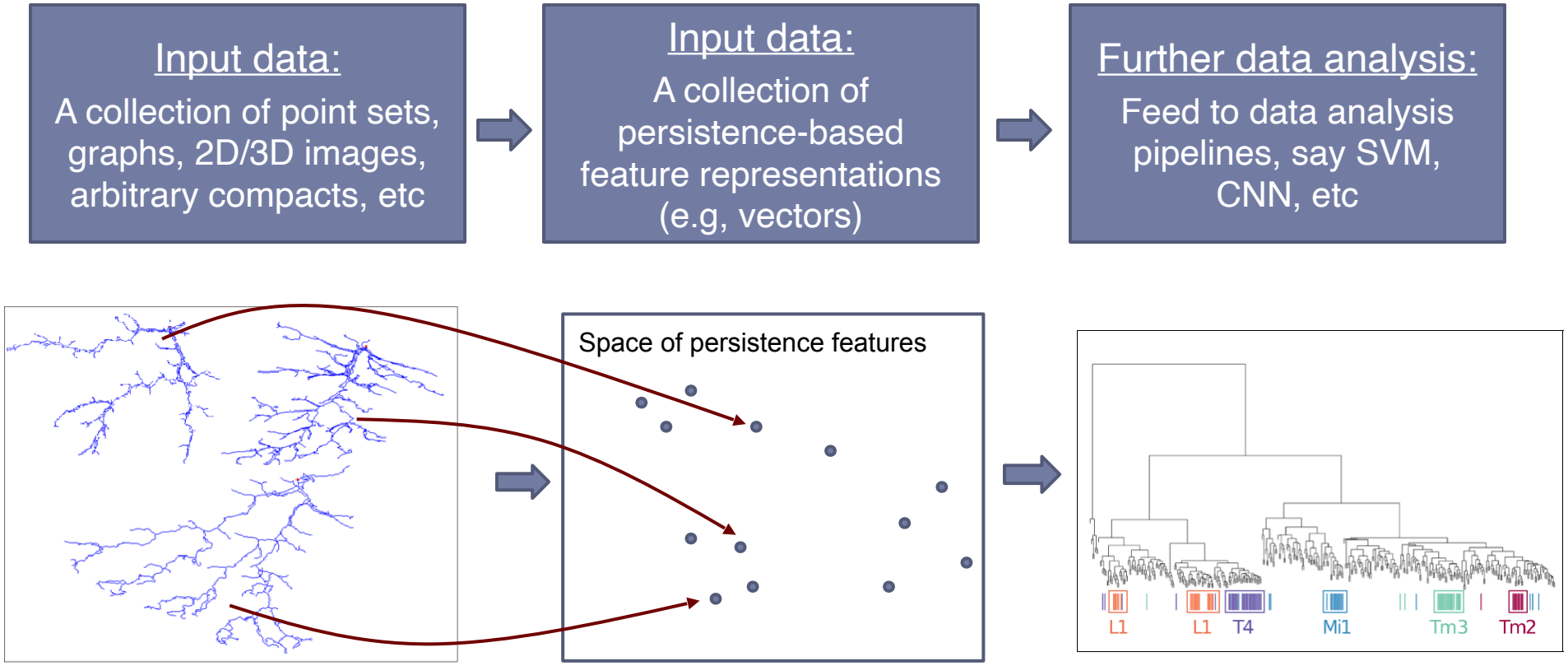
- Persistence-based feature vectorization + machine learning
  - Persistence images, kernels for persistence diagrams, etc





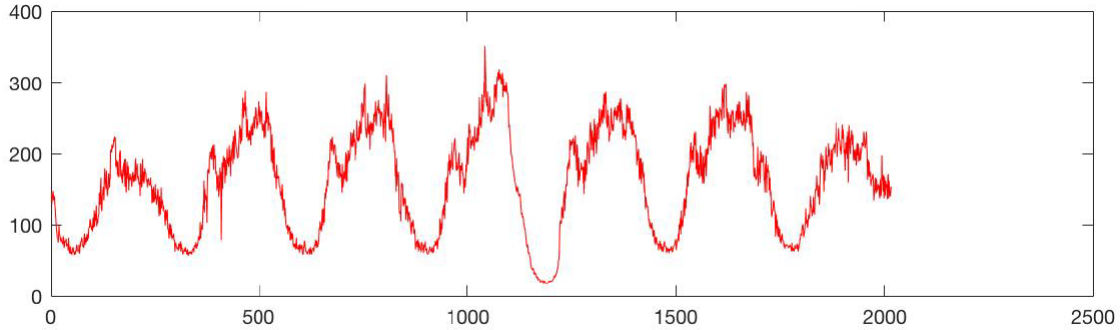
Motivating Examples VI

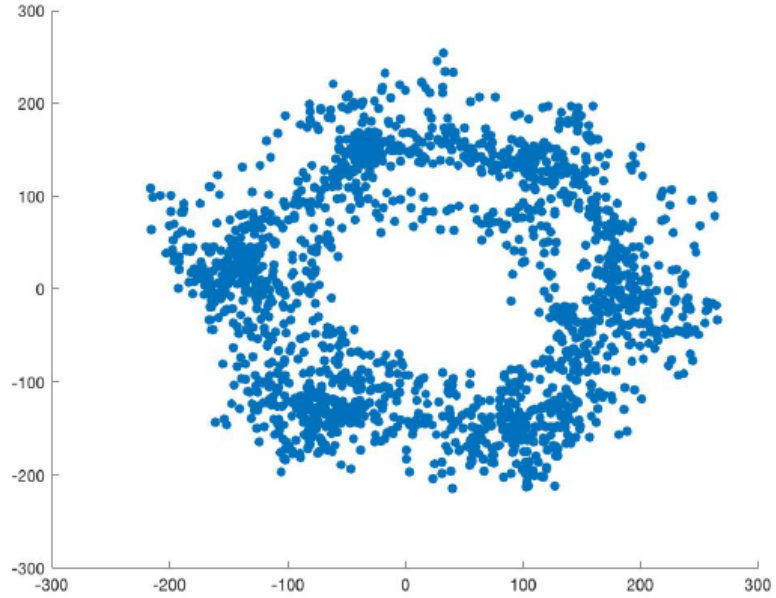
- Persistence-based feature vectorization + machine learning
  - Persistence images, kernels for persistence diagrams, etc



Motivating Examples VII

- Time series analysis
  - Finance - investment decisions

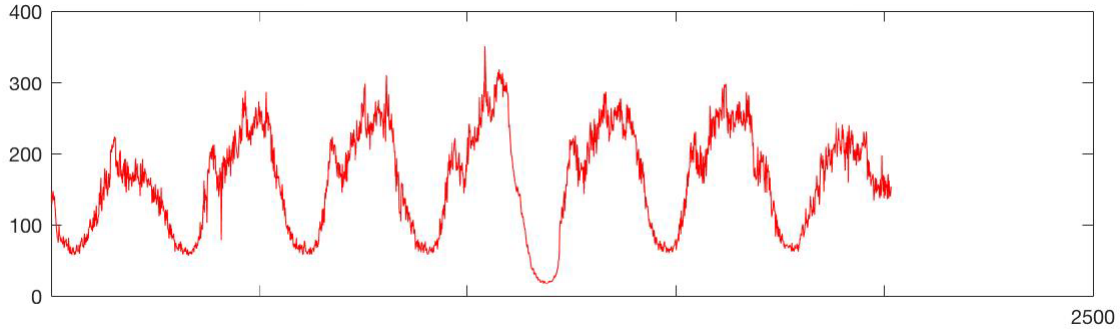


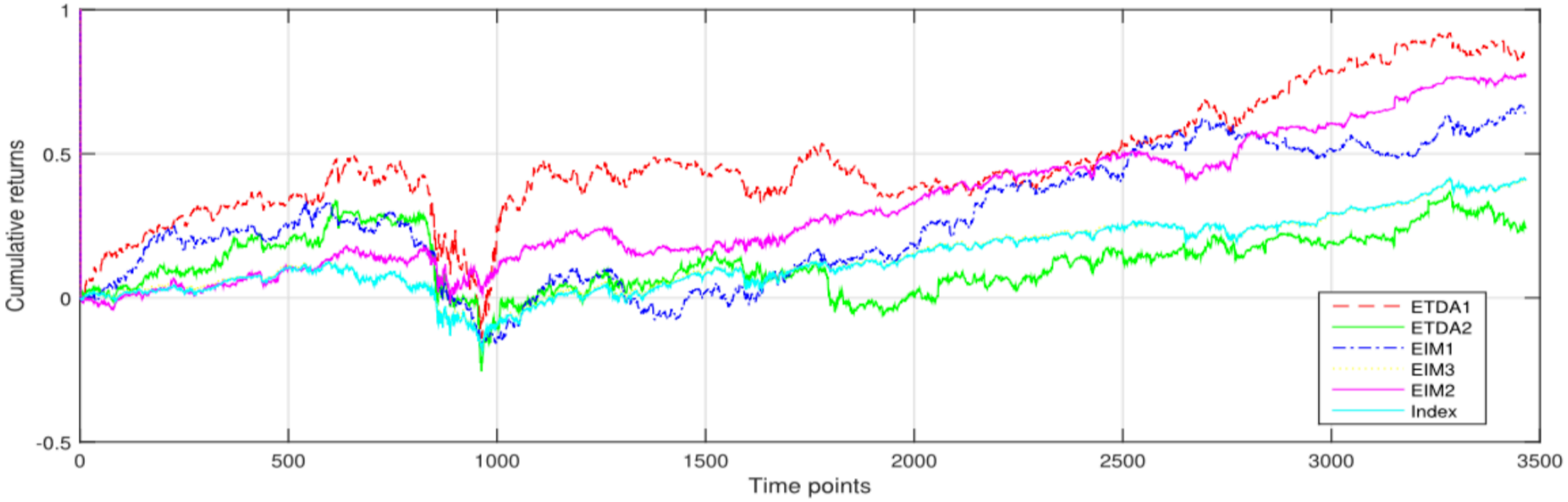


(a) Week 1

Motivating Examples VII

- Time series analysis
  - Finance - investment decisions



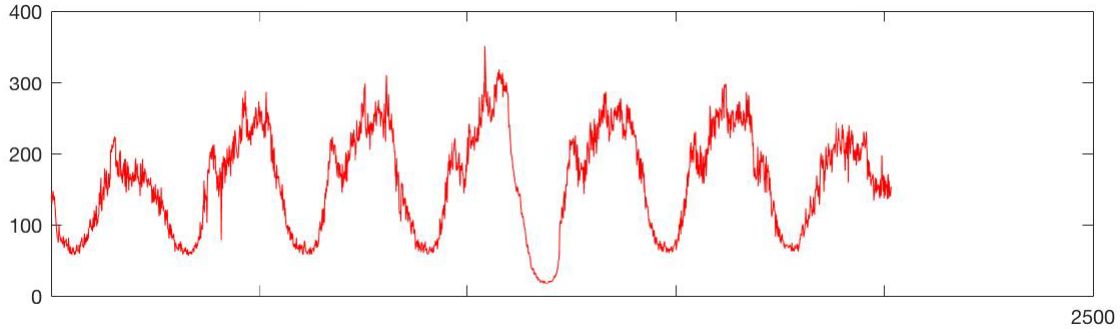


(d) Dow Jones Index

Goel et al. (2020)

Motivating Examples VII

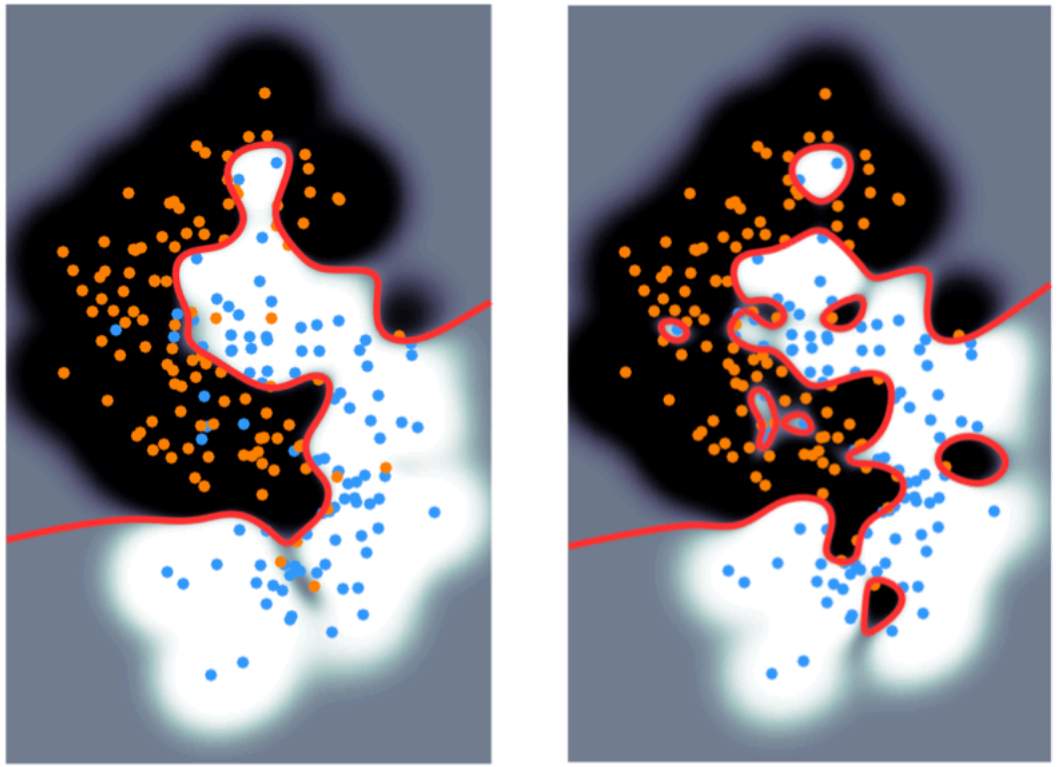
- Time series analysis
  - Finance - investment decisions



This is not financial advise!

Motivating Examples VIII

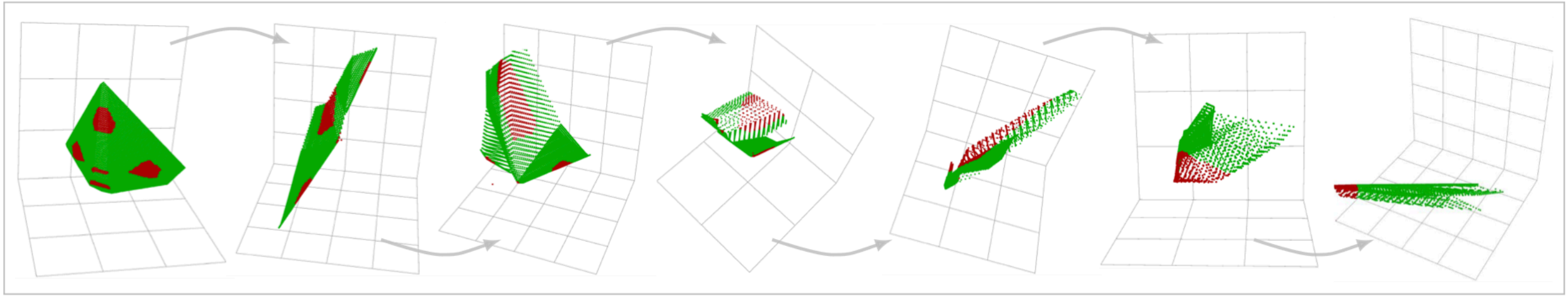
- Topological enhanced classifiers



Courtesy of Chen et al., A Topological Regularizer for Classifiers via Persistent Homology

Motivating Examples IX

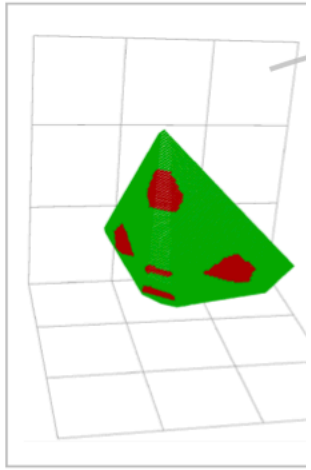
- Topology of neural networks

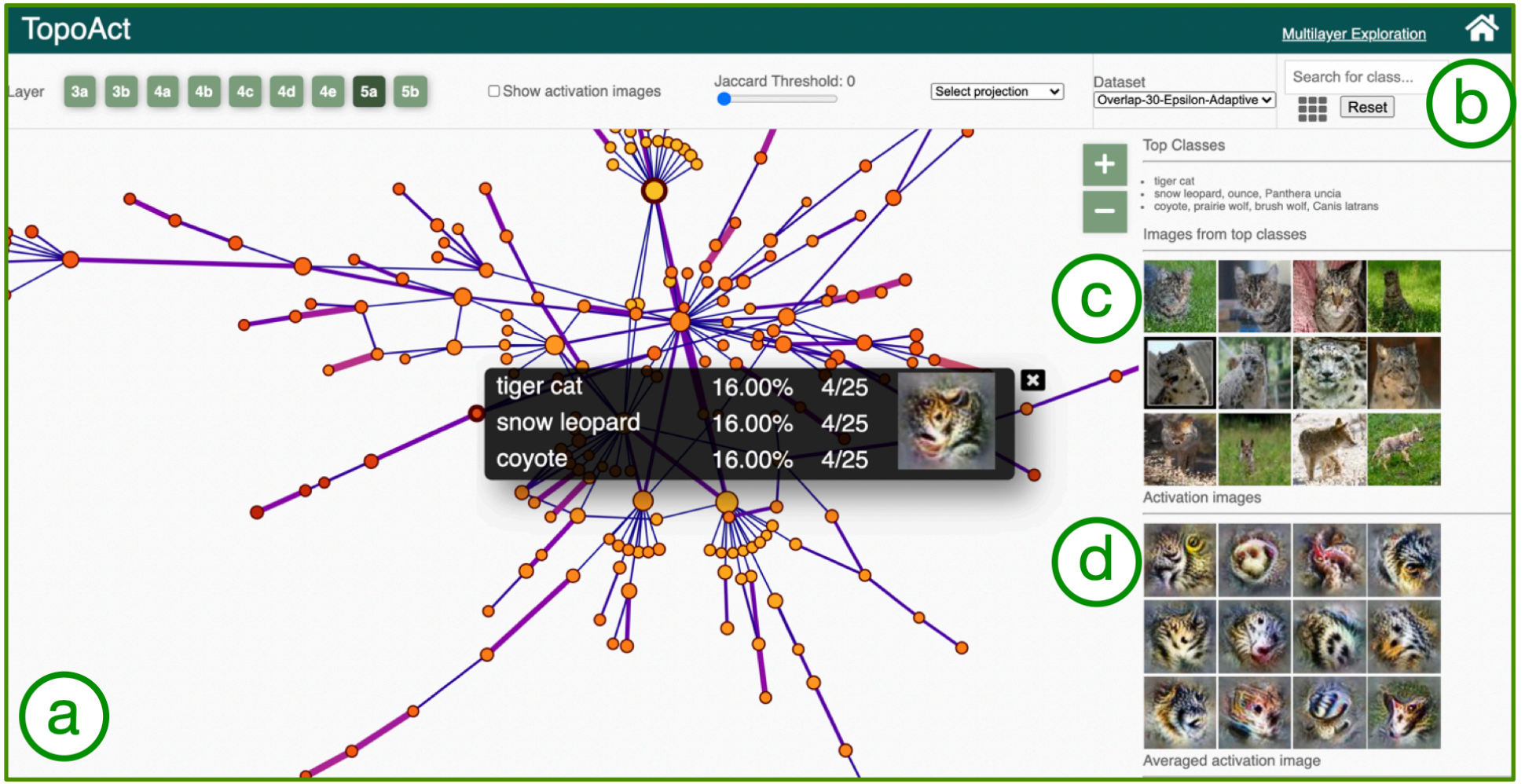


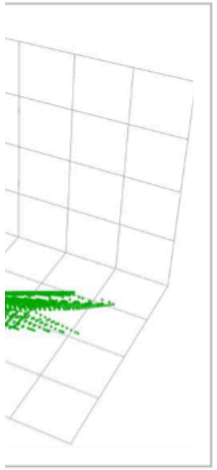
Courtesy of Naitzat et al., Topology of Deep Neural Networks

Motivating Examples IX

- Topo







Courtesy of Rathore et al (2019)

A Nice DataBase for TDA - Applications

- [Zotero Groups: TDA-Applications]

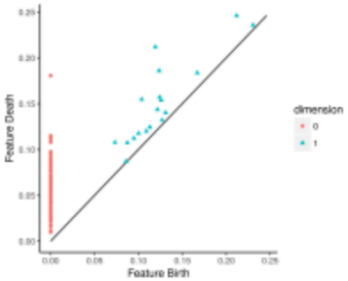
Home > Groups > TDA-Applications

TDA-Applications

Group Library

Recently Added Items

Title	Added By	Date Modified
Chatter Classification in Turning using Machine Learning and...	BarbaraGiunti	3/29/2023, 04:31:38
A Novel Quality Clustering Methodology on Fab-Wide Wafer Map...	BarbaraGiunti	2/4/2023, 05:44:16
Dynamic State Analysis of a Driven Magnetic Pendulum Using O...	BarbaraGiunti	2/4/2023, 05:12:07
Topological data analysis for true step detection in periodi...	BarbaraGiunti	2/4/2023, 04:55:32
Shape Terra: mechanical feature recognition based on a persi...	BarbaraGiunti	2/4/2023, 04:40:14



A database for applications of TDA outside maths. The scope of this database is to provide an as exhaustive as possible collection of applications of TDA to real data. Therefore, works pertaining (improving of) algorithms, new mathematical techniques or other improvements of the existing methods but not containing applications to real data sets will not be added.
For further questions, please contact the owner.

FIN