7th Italian-Japanese summer school in Mathematics

Persistent Homology:

A roadmap in Topological Data Analysis

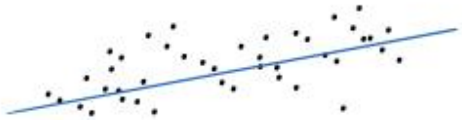
Claudia Landi

University of Modena and Reggio Emilia

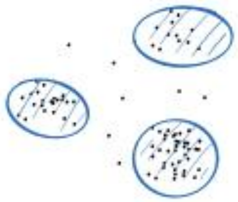
claudia.landi@unimore.it

Why topology for data analysis?

Topological data analysis (TDA) is a branch of data science that applies topology to extract hidden structures (*shape*) in data.



Correlation Methods



Clustering Methods

Premise of Topological Data Analysis

Data is noisy sample of some subspace in $\mathbb{R}^n$.
The "topology" of this subspace captures abstract (cor)relations in the data.

Loops
(and higher dim. generalizations)

Bifurcations

When is topology useful?

Many real datasets lie on or near nonlinear manifolds:

- periodic signals → circles
- motion trajectories → tori
- cell development → branching trees
- image and pose datasets → low-dimensional manifolds

Traditional methods (PCA, clustering, regression) typically assume a linear structure.

TDA instead asks: How many connected pieces? How many loops? Any voids?

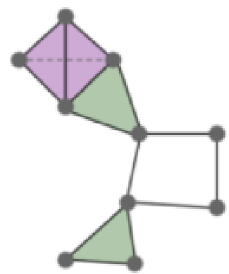
The goal is “seeing shape in data”

What will the course cover?

The whole **Persistent Homology** pipeline …

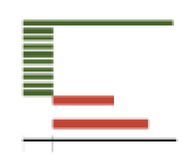
Data

Topology



$$0 \to \mathbb{Z}_2 \xrightarrow{\begin{bmatrix}1\\0\end{bmatrix}} \mathbb{Z}_2 \oplus \mathbb{Z}_2 \xrightarrow{\begin{bmatrix}1 & 0\end{bmatrix}} \mathbb{Z}_2 \to 0.$$

Algebra

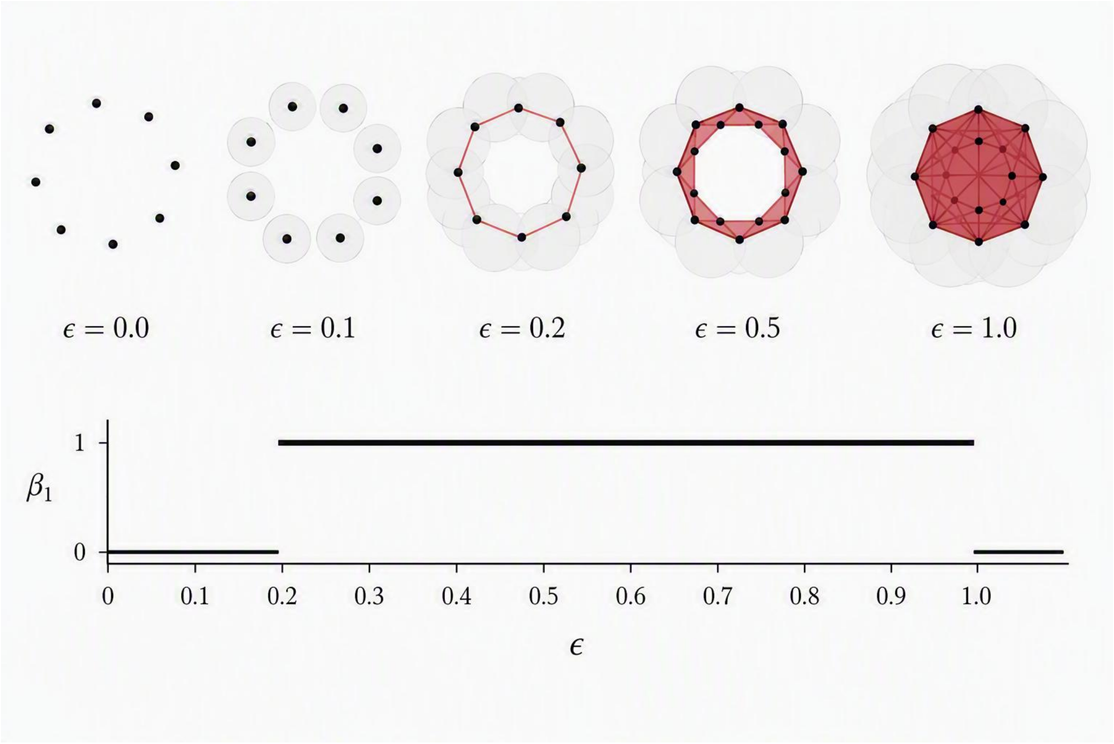


Invariants

What will the course cover?

… and its properties:

- Stability with respect to input data perturbation
- Topological inference



Syllabus and textbook

17/9: From data to simplicial complexes

- simplicial complexes, Čech and Vietoris–Rips complexes

18/9: From simplicial complexes to algebra:

- simplicial homology, filtrations, persistence modules

24/9: From persistence modules to invariants

- Barcodes, persistence modules

25/9: persistent homology stability, topological inference

Reference:

M.B. Botnan: Topological Data Analysis Mastermath, lecture_notes.pdf