Persistent Homology

Análisis Topológico de Datos (2264)

Mauricio Toledo-Acosta

January 29, 2026

Topological Data Analysis

Topological Data Analysis (TDA) is an approach to the analysis of datasets using techniques from topology.

Topological Data Analysis

Topological Data Analysis (TDA) is an approach to the analysis of datasets using techniques from topology. Extraction of information from high-dimensional, noisy datasets is challenging. TDA provides a general framework to analyze such data in a manner that is insensitive to the underlying metric, and provides robustness to noise.

Topological Data Analysis

Topological Data Analysis (TDA) is an approach to the analysis of datasets using techniques from topology. Extraction of information from high-dimensional, noisy datasets is challenging. TDA provides a general framework to analyze such data in a manner that is insensitive to the underlying metric, and provides robustness to noise.

The initial motivation is to study the shape of data. TDA combines algebraic topology and other tools from pure mathematics to allow mathematically rigorous study of "shape". One of its main tools is persistent homology, which has been applied to many types of data across many fields.

Introduction

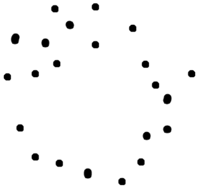
- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

Topological Persistence and Simplification

Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

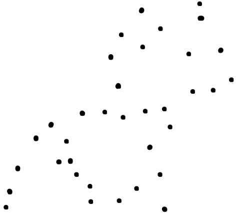
Topological Persistence and Simplification



Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

Topological Persistence and Simplification



Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

Topological Persistence and Simplification

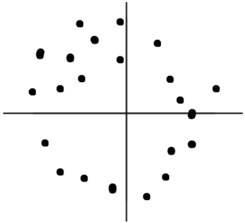
- It is independent of dimensions and coordinates

Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

Topological Persistence and Simplification

- It is independent of dimensions and coordinates

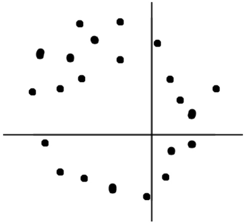


Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

  Topological Persistence and Simplification

- It is independent of dimensions and coordinates

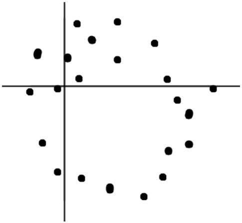


Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

  Topological Persistence and Simplification

- It is independent of dimensions and coordinates



Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

Topological Persistence and Simplification

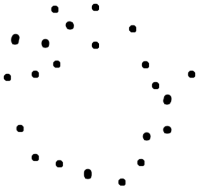
- It is independent of dimensions and coordinates, robust to perturbations of input data.

Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

  Topological Persistence and Simplification

- It is independent of dimensions and coordinates, robust to perturbations of input data.

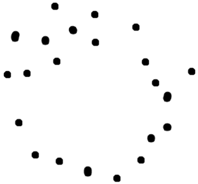


Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

Topological Persistence and Simplification

- It is independent of dimensions and coordinates, robust to perturbations of input data.



Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

  Topological Persistence and Simplification

- It is independent of dimensions and coordinates, robust to perturbations of input data.
- Persistent Homology accounts for local geometry and global topology of a dataset.

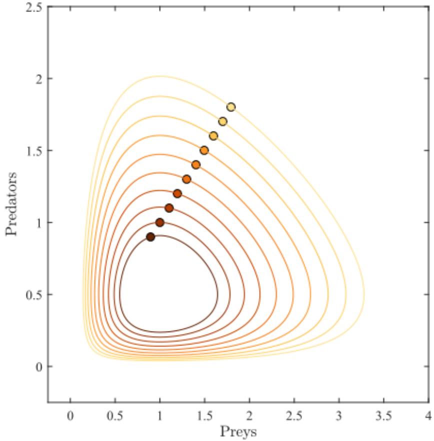
Introduction

- Persistent Homology is a method used in topological data analysis to study qualitative features of data that persist across multiple scales. It was introduced in 2002.

  Topological Persistence and Simplification

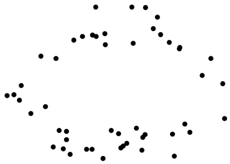
- It is independent of dimensions and coordinates, robust to perturbations of input data.
- Persistent Homology accounts for local geometry and global topology of a dataset.
- It has many applications: Image analysis, NLP, Flows, input features for machine learning algorithms, mathematics results.

An example: The Lotka-Volterra equations



The shape of a cloud of points

We have a set of points $X = \{x_1, ..., x_n\} \subset \mathbb{R}^D$, we want to find out the *shape* of X.

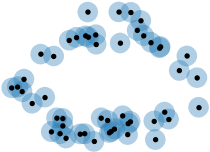


The shape of a cloud of points

We have a set of points $X = \{x_1, ..., x_n\} \subset \mathbb{R}^D$, we want to find out the *shape* of X.

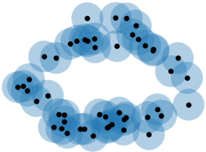
The shape of a cloud of points

We have a set of points $X = \{x_1, ..., x_n\} \subset \mathbb{R}^D$, we want to find out the *shape* of X.



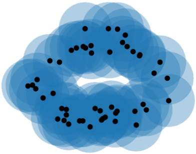
The shape of a cloud of points

We have a set of points $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^D$, we want to find out the *shape* of X.



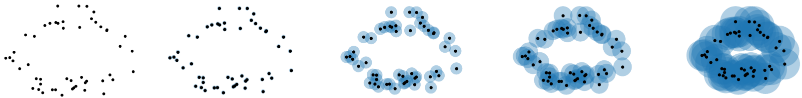
The shape of a cloud of points

We have a set of points $X = \{x_1, ..., x_n\} \subset \mathbb{R}^D$, we want to find out the *shape* of X.



The shape of a cloud of points

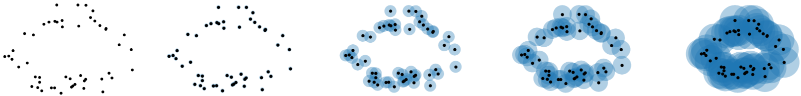
We have a set of points $X = \{x_1, ..., x_n\} \subset \mathbb{R}^D$, we want to find out the *shape* of X.



What is the *right* radius?

The shape of a cloud of points

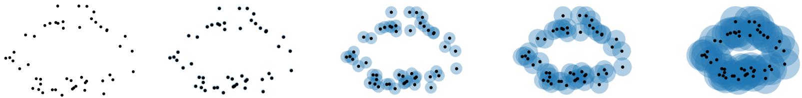
We have a set of points $X = \{x_1, ..., x_n\} \subset \mathbb{R}^D$, we want to find out the *shape* of X.



What is the *right* radius? We consider all the possible radii.

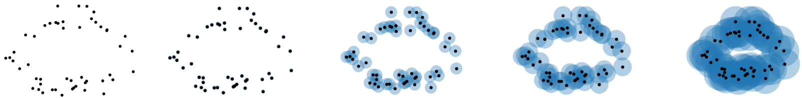
Filtrations

- We consider the family of these spaces for each radius $r > 0$.



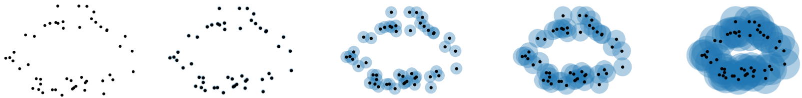
Filtrations

- We consider the family of these spaces for each radius $r > 0$.
- Each of these spaces represent the figure hiding behind the set of points.



Filtrations

- We consider the family of these spaces for each radius $r > 0$.
- Each of these spaces represent the figure hiding behind the set of points.
- Each space is contained in the next one. This is called a **filtration**.

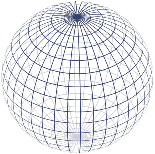


Counting holes

The shape we are trying to determine is given by the number of k-dimensional holes for $k < D$.

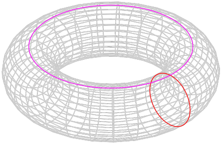
Counting holes

The shape we are trying to determine is given by the number of k-dimensional holes for $k < D$.



Counting holes

The shape we are trying to determine is given by the number of k-dimensional holes for $k < D$.



Counting holes

The shape we are trying to determine is given by the number of k-dimensional holes for $k < D$.

Counting holes

The shape we are trying to determine is given by the number of k-dimensional holes for $k < D$.

This sequence of numbers are the **Betti numbers**.

Counting holes

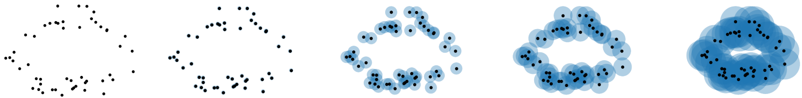
The shape we are trying to determine is given by the number of k-dimensional holes for $k < D$.

This sequence of numbers are the **Betti numbers**. They can be regarded as:

- b_0: number of connected components.
- b_1: number of one-dimensional or "circular" holes.
- b_2: number of two-dimensional "voids" or "cavities.

Tracking holes

The goal is to keep track of the appearing and disappearing holes as we go through the filtration.



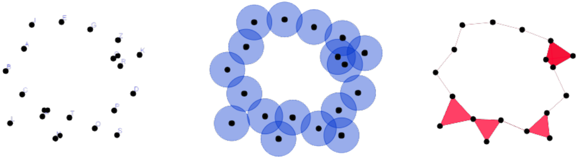
The time each hole (cycle) persists is an indicative of the presence of said cycle in the figure.

Čech Complex and Vietoris-Rips Complex

- Computing the Betti numbers (rank of the homology groups) in the previous filtration is hard.

Čech Complex and Vietoris-Rips Complex

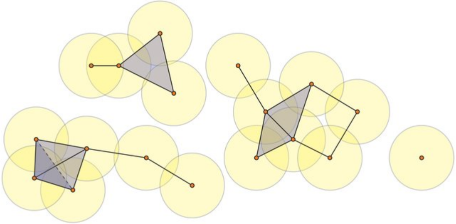
- Computing the Betti numbers (rank of the homology groups) in the previous filtration is hard.
- We construct a simplicial complex:



Čech Complex and Vietoris-Rips Complex

There are two ways of constructing the simplicial complex:

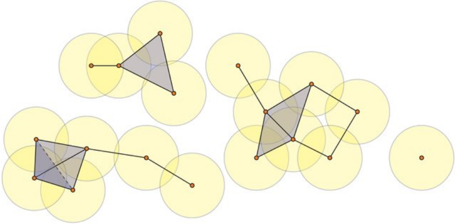
- Vietoris-Rips Complex:



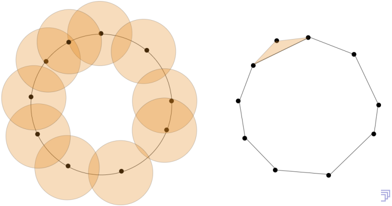
Čech Complex and Vietoris-Rips Complex

There are two ways of constructing the simplicial complex:

- Vietoris-Rips Complex:



- Čech Complex:



Čech Complex and Vietoris-Rips Complex: Differences

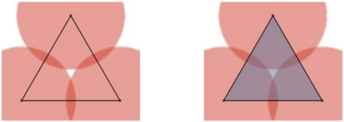


Figure: Čech (left), Vietoris-Rips (right)

Čech complex	Vietoris-Rips complex
Smaller	Bigger
Computationally expensive	Less expensive
More accurate	Less accurate

The new filtration

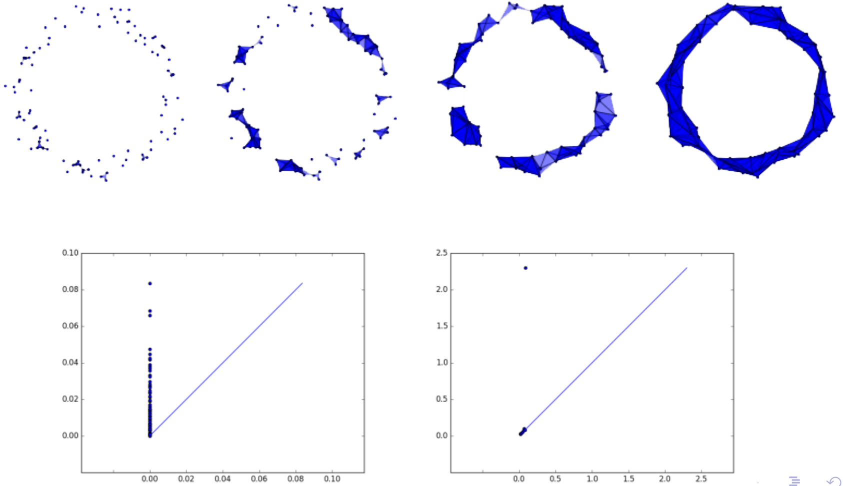
The filtration is infinite, but there is only a finite number of different complexes.

The new filtration

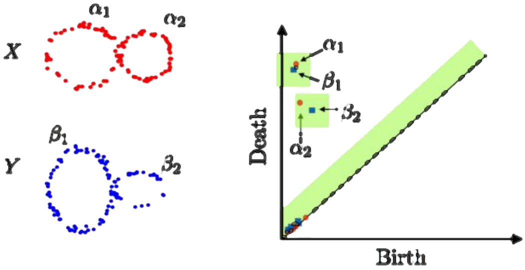
The filtration is infinite, but there is only a finite number of different complexes.

We track the appearing and disappearing holes as we go through each complex of the filtration. The time each cycle persists is an indicative of the presence of said cycle in the figure.

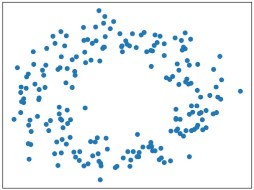
Representing Results: Persistence Diagrams

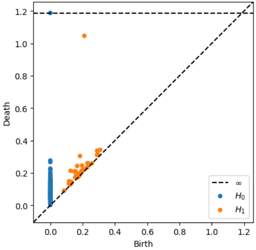


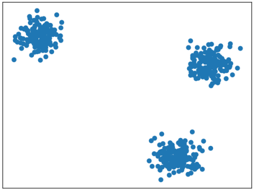
Visualizing Results: Persistence Diagrams

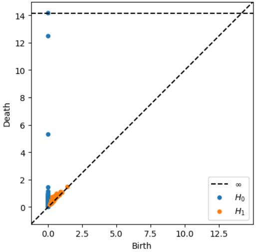


Examples

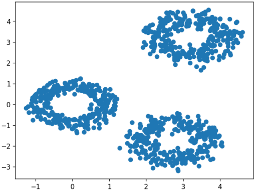


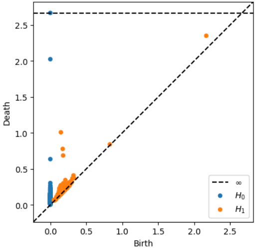

1.2
1.0
0.8
0.6
0.4
0.2
0.0
Death
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Birth
∞
H_0
H_1

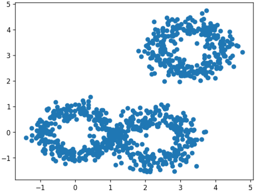


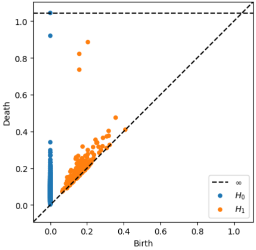

14
12
10
8
6
4
2
0
Death
0.0
2.5
5.0
7.5
10.0
12.5
Birth
∞
H_0
H_1

Examples



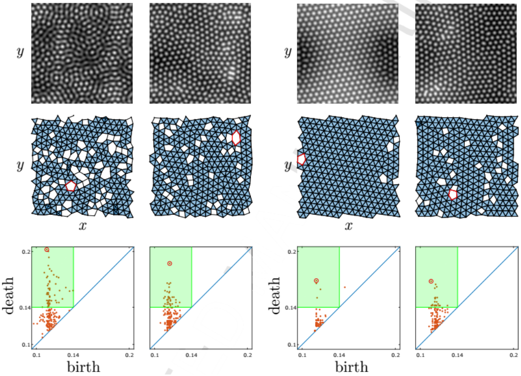






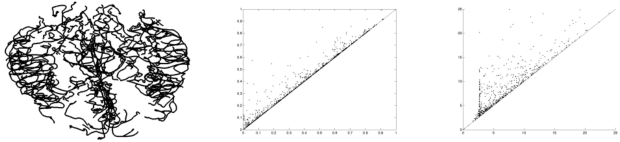
Local Geometry

In this paper, authors compare quantitative measures of order for nearly hexagonal, planar lattices using Persistent Homology.



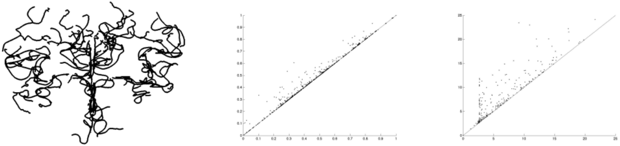
Motta, F. C., Neville, R., Shipman, P. D., Pearson, D. A., Bradley, R. M. (2018). **Measures of order for nearly hexagonal lattices**. Physica D: Nonlinear Phenomena, 380, 17-30.

When *noise* is the important feature



(a) Brain tree (b) Dgm_0 (c) Dgm_1

Figure 13: Persistent homology data objects from a 24-year old. Left: brain tree. Middle: zero-dimensional diagram. Right: one-dimensional diagram.



(a) Brain tree (b) Dgm_0 (c) Dgm_1

Figure 14: Persistent homology data objects from a 68-year old. Left: brain tree. Middle: zero-dimensional diagram. Right: one-dimensional diagram.

Bendich, P., Marron, J. S., Miller, E., Pieloch, A., Skwerer, S. (2016). **Persistent homology analysis of brain artery trees**. The annals of applied statistics, 10(1), 198

Identifying Semantic *Tie-Backs*

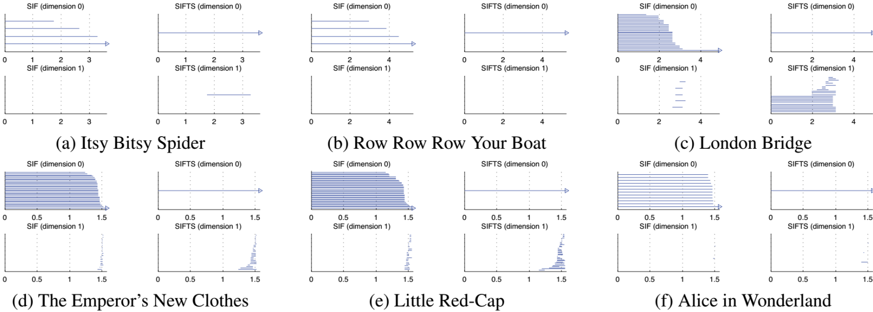
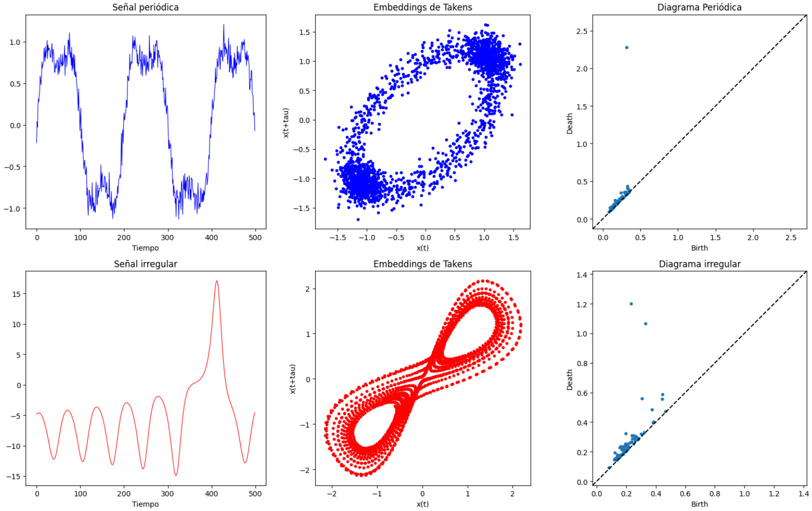


Figure 1: Persistent homology on nursery rhymes and other stories

Zhu, X. (2013, August). **Persistent Homology: An Introduction and a New Text Representation for Natural Language Processing**. In IJCAI (pp. 1953-1959).

Quantifying periodicity in signals



More examples

- Hamilton, W., Borgert, J. E., Hamelryck, T., & Marron, J. S. (2022). **Persistent topology of protein space**. In Research in computational topology 2 (pp. 223-244). Cham: Springer International Publishing.
- Kovacev-Nikolic, V., Bubenik, P., Nikolić, D., & Heo, G. (2016). **Using persistent homology and dynamical distances to analyze protein binding**. Statistical applications in genetics and molecular biology, 15(1), 19-38.
- Gidea, M., & Katz, Y. (2018). **Topological data analysis of financial time series: Landscapes of crashes**. Physica A: Statistical mechanics and its applications, 491, 820-834.