Topological Data Analysis for data analysis and AI in robotics

Víctor Toscano Durán
Department of Applied Mathematics I, University of Seville
16 October, 2024

About me

- Graduate in Statistics.
- Specialised in Artificial Intelligence, with a master’s degree.
- 2 years of experience working as a data scientist, focusing on the development of AI models, as well as data análisis including debugging and data processing.
- Currently, researcher at the USE in the REXASI-PRO European Project and doing a PhD in mathematics, which is focused on the intersection between Topological Data Analysis and AI.

Talk objective

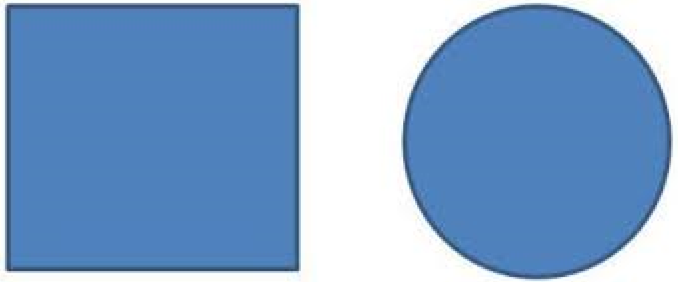
- Introduction to Topological Data Analysis (TDA). Why it is relevant?
  - Topology
    - Simplicial complex
    - Homology
  - TDA
  - TDA pipeline
  - Persistence homology

- Topological data analysis connection with Artificial Intelligence.
  - Vectorizations of PH

- Topological Data Analysis applications to REXASI-PRO robots simulations.
  - REXASI-PRO
  - TDA applications

What is Topology?

Topology, is a branch of Mathematics, which explores the properties of space that remain unchanged under continuous transformations, like stretching or vending, without tearing.



Topological Key Concepts

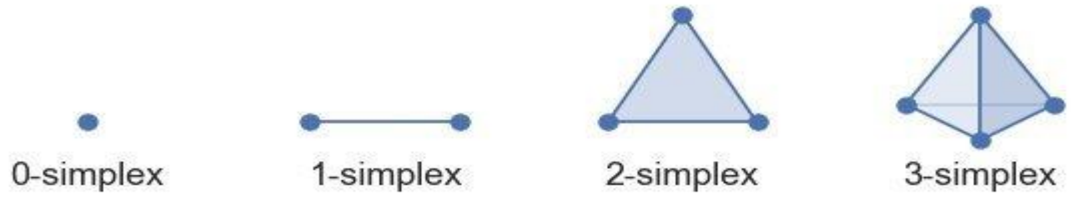
- Simplicial Complexes

- Homology

Simplicial complex

Simplicial complexes are a data structure used to represent topological spaces.

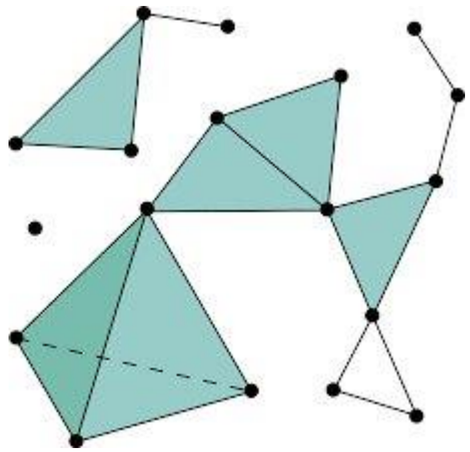
- A simplex is a general geometric object that have dimension:
  - 0-simplex: a point (called a vertex)
  - 1-simplex: a line segment (called an edge)
  - 2-simplex: a triangle (filled)
  - 3-simplex: a tetrahedron



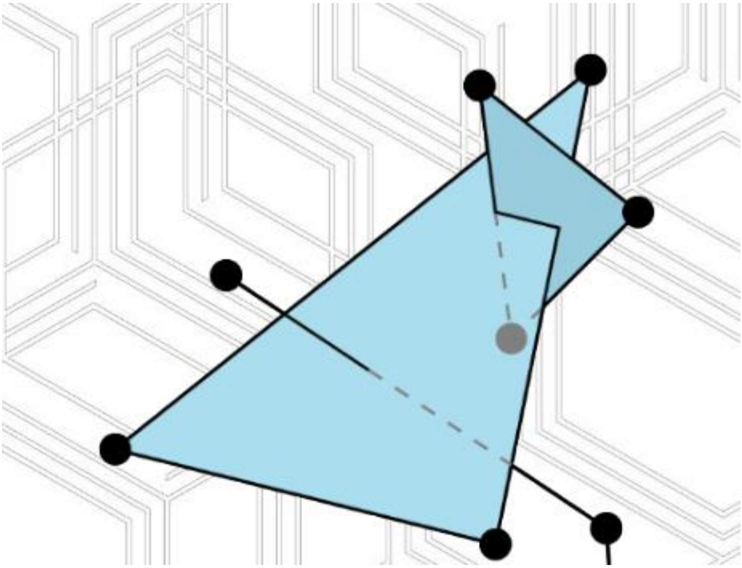
Simplicial complex

Simplicial complexes are a data structure used to represent topological spaces.

- A simplicial complex is obtained by a nested family of simplices.



Simplicial Complex example



This is not a simplicial complex!

Simplicial complex types

- Cech complex

- Vietoris-Rips complex

Cech complex

- The most elementary simplicial complex.

- It is defined in terms of the intersection of balls centered at the point.

Cech complex

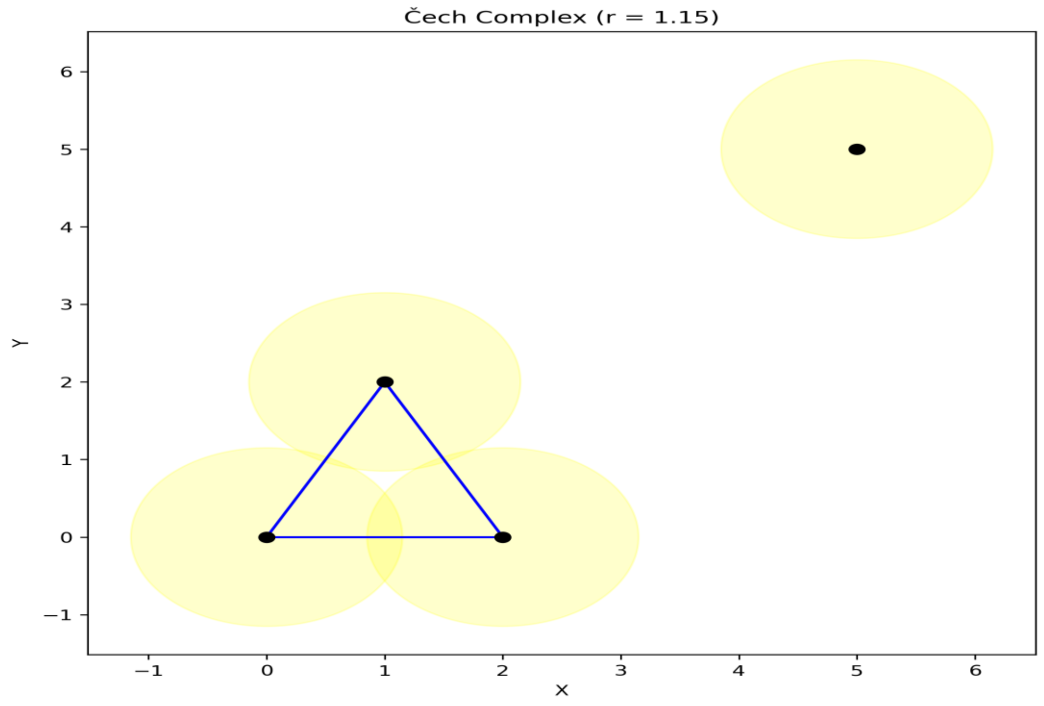
P is a set of points and r a positive radius

The Čech complex $C_r(P)$ is defined as:

$$C_r(P) = \{\sigma \subseteq P \mid \bigcap_{p_i \in \sigma} B_r(p_i) \neq \emptyset\},$$

where $B_r(p_i)$ is the ball of radius r centered at p_i:

$$B_r(p_i) = \{x \in \mathbf{R}^d \mid \|x - p_i\| \leq r\}.$$



Vietoris-Rips complex

The Vietoris-Rips complex is related to the Čech complex, but instead of using the intersection of balls to define connectivity, it relies on distances between points. (However, this can also be visualized using balls.)

$$\boldsymbol{R_{r(P)} \subseteq C_{r(P)} \subseteq R_{\{2r\}(P)}}$$

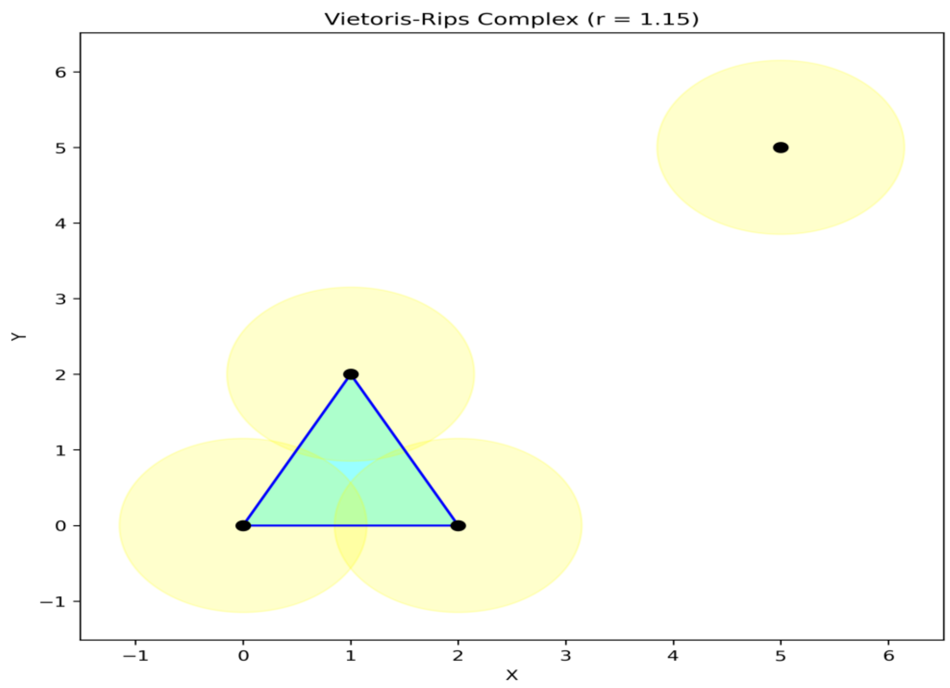
Vietoris-Rips complex

Let $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ be a set of points in a metric space $\mathbb{R}^d$ and r a radius.

The Vietoris-Rips complex $\mathcal{R}_r(\mathcal{P})$ is defined as:

$$\mathcal{R}_r(\mathcal{P}) = \{\sigma \subseteq \mathcal{P} \mid \|p_i - p_j\| \leq r \text{ for all } p_i, p_j \in \sigma\}$$

where $\|p_i - p_j\|$ denotes the distance between points p_i and p_j.



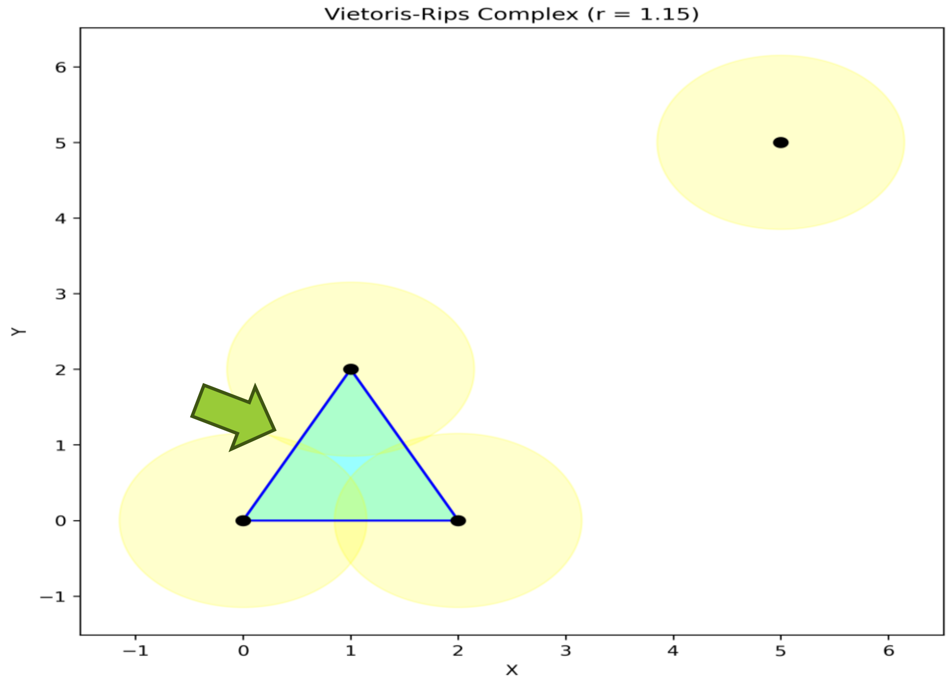
Vietoris-Rips complex

Let $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ be a set of points in a metric space $\mathbb{R}^d$ and r a radius.

The Vietoris-Rips complex $\mathcal{R}_r(\mathcal{P})$ is defined as:

$$\mathcal{R}_r(\mathcal{P}) = \{\sigma \subseteq \mathcal{P} \mid \|p_i - p_j\| \leq r \text{ for all } p_i, p_j \in \sigma\}$$

where $\|p_i - p_j\|$ denotes the distance between points p_i and p_j.



Homology

- What does homology (and Betti numbers) measure?

- **Homology** quantify how many structures, such as connected components, loops, or cavities, exist in a space. → **Betti numbers**

  - β_0 : number of connected components
  - β_1 : number of cycles
  - β_2 : number of voids
  - β_k : number of k-dimensional holes

Homology

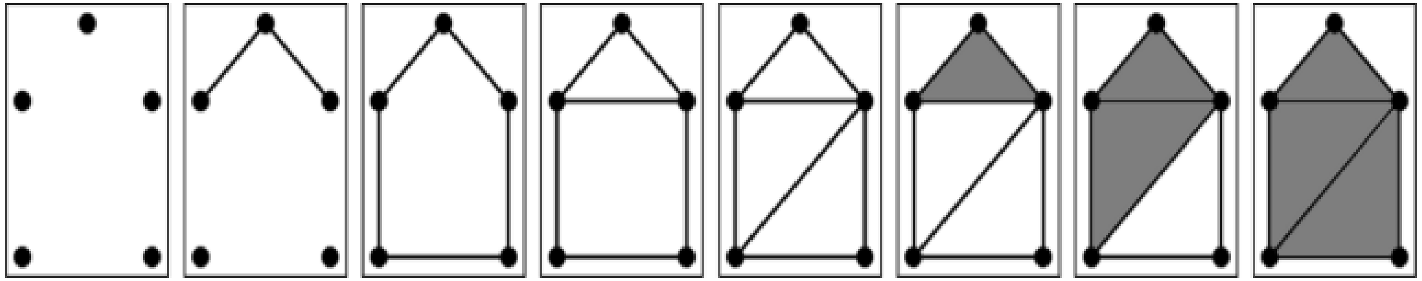


Figure *Filtration.*
A filtration for $t = 0, 1, 2, 3, 4, 5, 6, 7$ (from left to right)

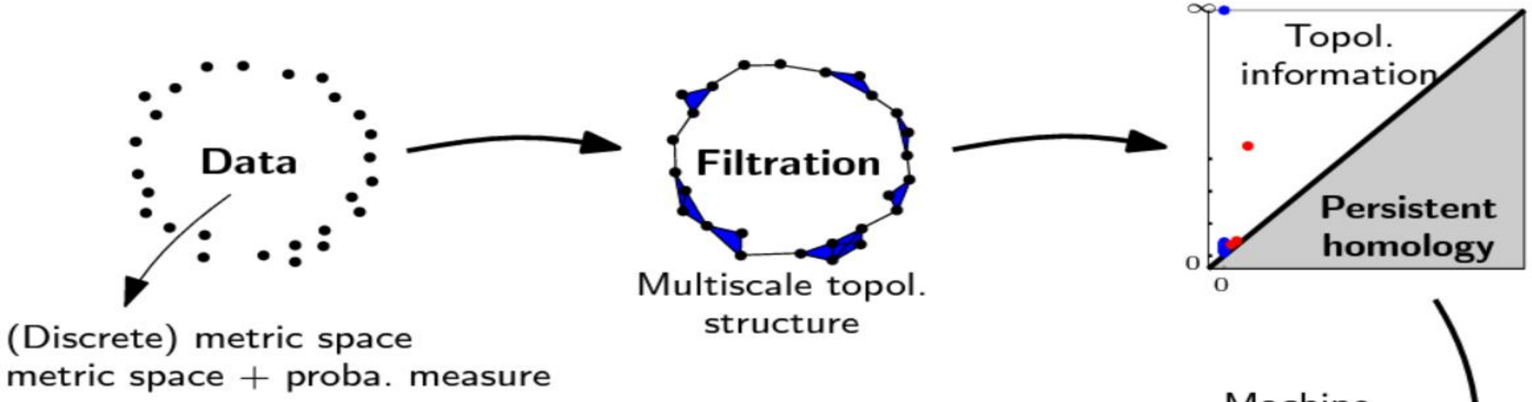
- $t = 0 \rightarrow \beta_0 = 5, \beta_1 = 0, \beta_2 = 0$
- $t = 2 \rightarrow \beta_0 = 1, \beta_1 = 1, \beta_2 = 0$
- $t = 5 \rightarrow \beta_0 = 1, \beta_1 = 2, \beta_2 = 1$

What is Topological Data Analysis?

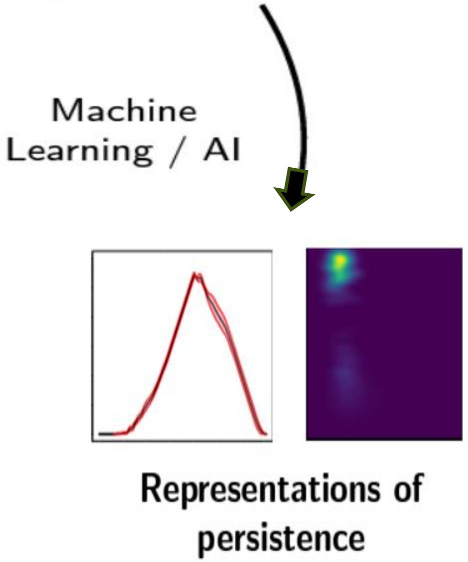
Topological Data Analysis (TDA) consists in applying techniques from topology and algebra to the analysis of data, stuyind the shape of the data.

The data we use often has a complex geometric/topological structure, which can be very useful to know and use in tasks such as data analysis and AI modelling.

Typical TDA pipeline



1. Build a multiscale topological structure on top of data (filtrations)
2. Compute persistence homology.
3. Use vectorizations of topological information in ML/AI.



Persistence Homology

- Persistence homology is a mathematical framework to encode the evolution of the topology of a collection of simplicial complex (filtration) from a topological space(data)

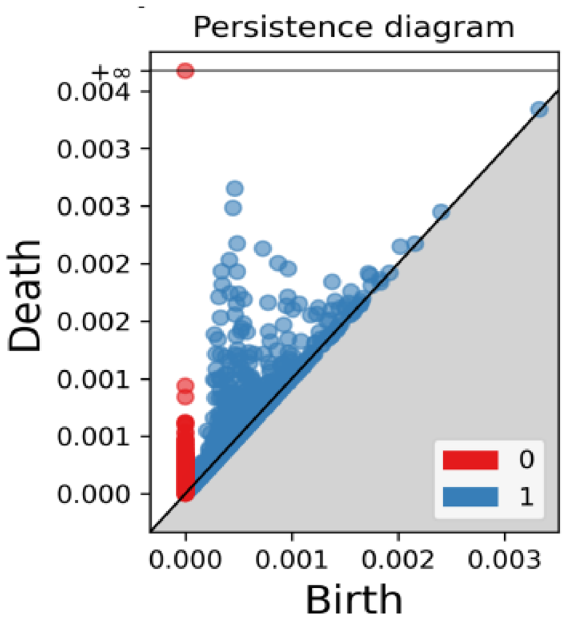
1. Point cloud data
2. Construct the simplicial complex at different spatial resolutions
3. Persistence Homology describes how the homology of the filtration changes varying the resolution of the space.

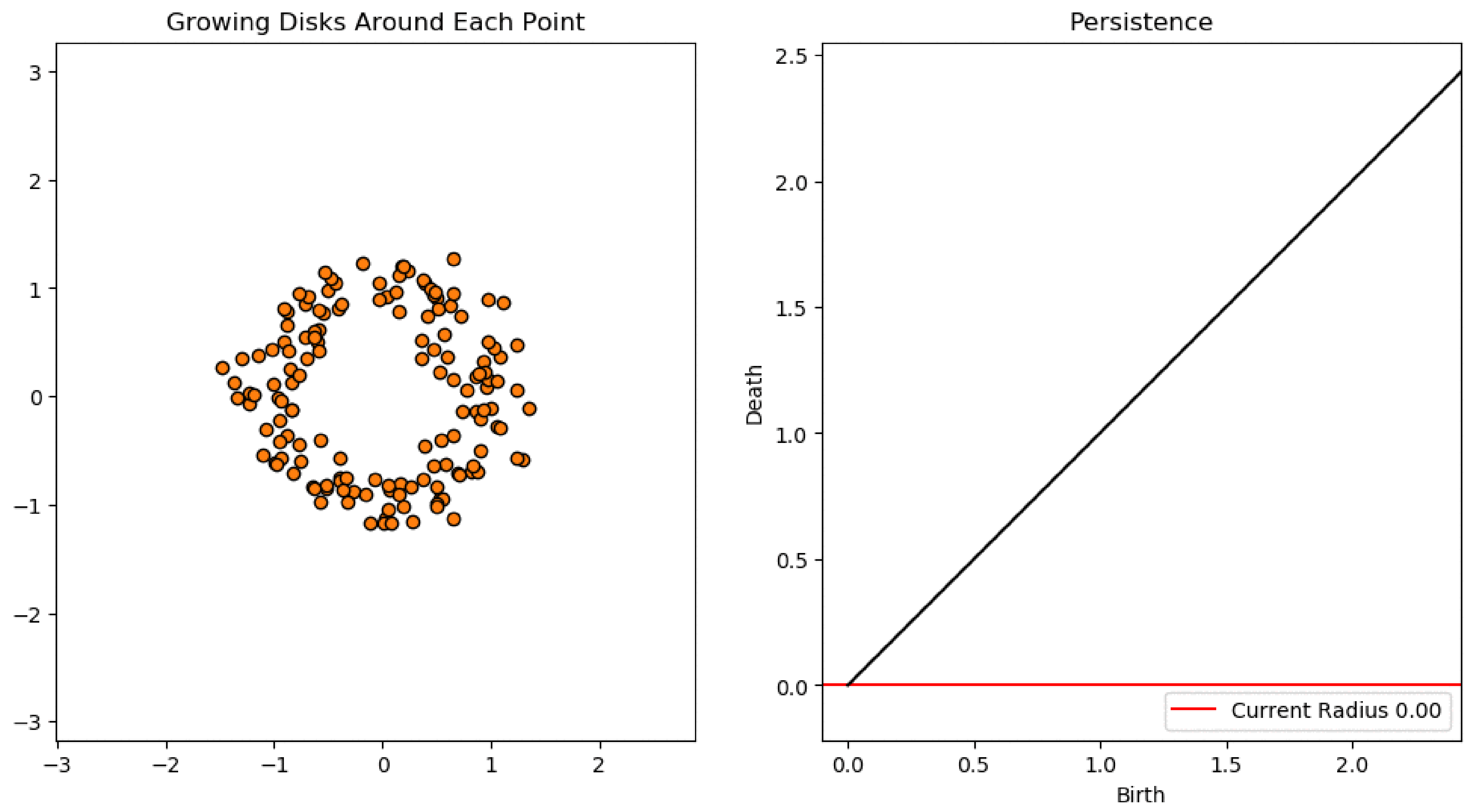
Persistence diagram

- Use for visualize the persistence of topological features.

- Consisting in a set of point, representing the birth and deatc of topological features

- Points near the diagonal indicate short-lived features,where points far away from the diagonal represent more persistent.

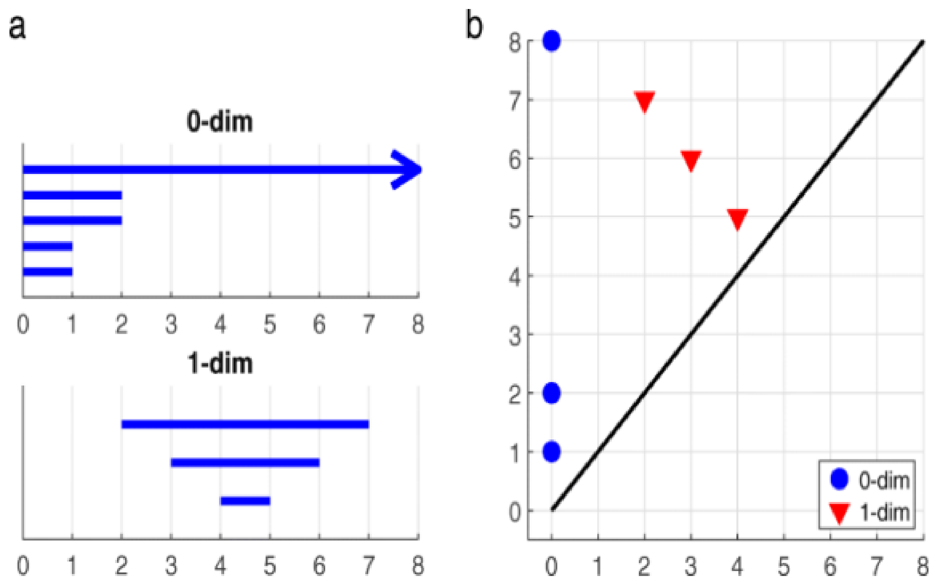




Growing Disks Around Each Point
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
Persistence
2.5
2.0
1.5
1.0
0.5
0.0
Death
0.0
0.5
1.0
1.5
2.0
Birth
Current Radius 0.00

Persistence Barcode

Alternative to persistence diagram, drawing intervals representing bith and death of topological features. Each bar represent a point in the persistence diagram.



Vectorization of PH

- Persistence entropy

- Persistence landscapes

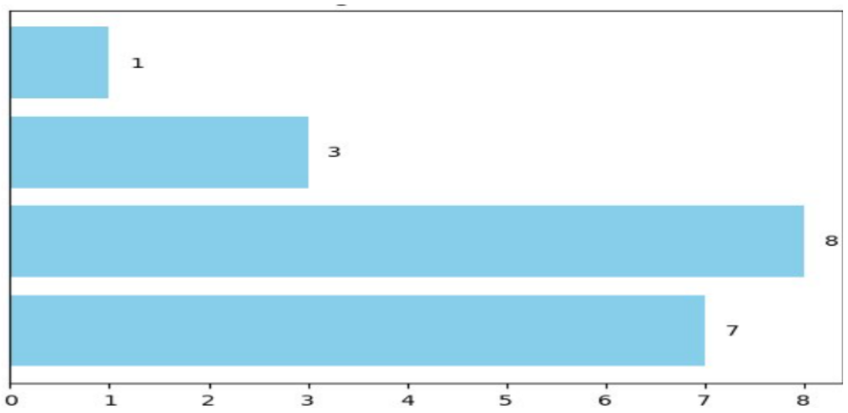
- Persistence images

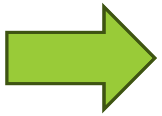
Persistence entropy

Persistence entropy is a measure of the complexity of a topological space based on its persistence diagram, measuring how different bars are in length.

$$H = -\sum_{i \in I} p_i \ln p_i$$

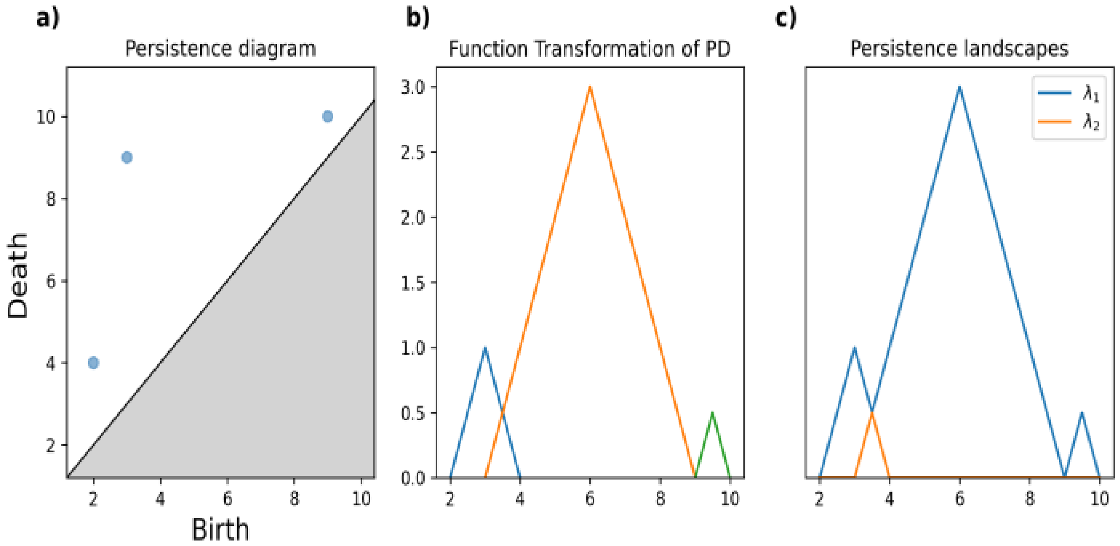
Maximum persistence entropy corresponds to the situation in which all the intervals in the barcode are of equal length.





Persistence entropy = 1.18

Persistence landscapes

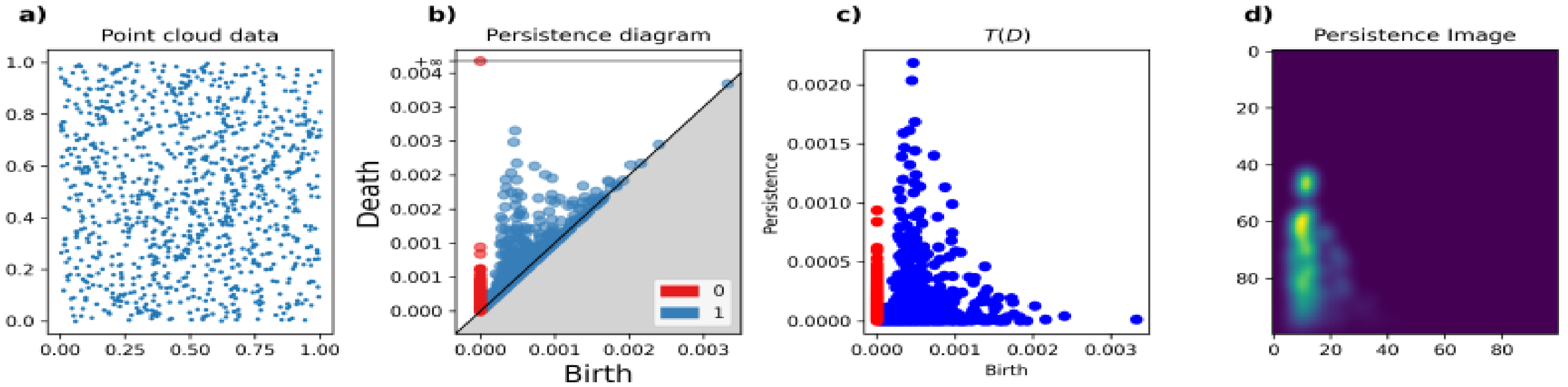


a) Persistence diagram containing three points. b) Transforming points in the persistence diagram into a mathematical function. Each point is represented by different colors. c) The corresponding persistence landscape. Notice that λ_1 gives a measure of the dominant homological feature at each point of the filtration.

Persistence image

Persistence image is an alternative way to represent information from persistence diagrams, transforming them into images using image analysis techniques.

- Persistence Diagram
- Transformation of PD
- Calculating persistence Surface
- Discretize and integrate the Surface, producing the image.

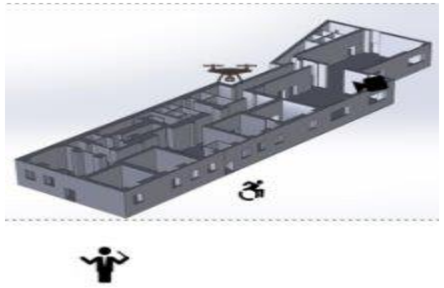


REXASI-PRO

Wheelchair

Drone

Orchestrator



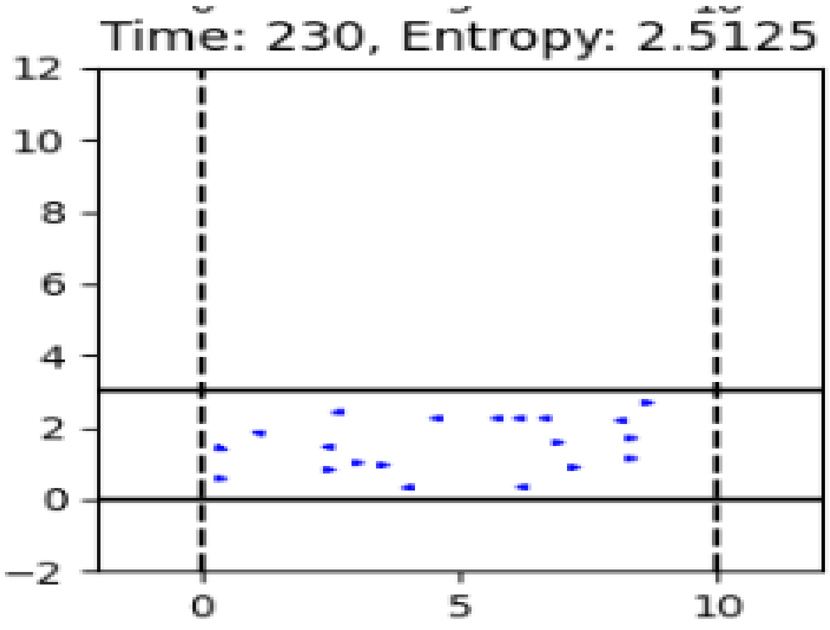
Navground simulation

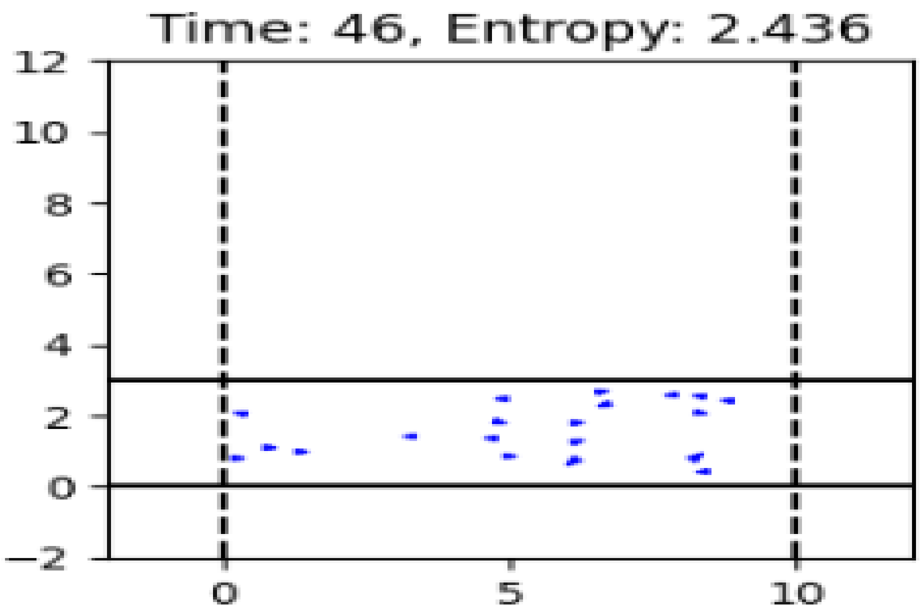
REXASI-PRO

- Using TDA for analyse the discrete model of the fleet behavior (persons and wheelchairs in movement).
  - Comparing simulations using their persistent entropy time series.
  - Explain how the robots are distributed on the space using persistence entropy.
  - Relationing negative events like collisions and deadlock with persistence entropy.
  - We want to extend this work using other vectorizations as persistence landscapes

REXASI-PRO

In a single simulation, what means to have a high entropy or low entropy at a specific moment?

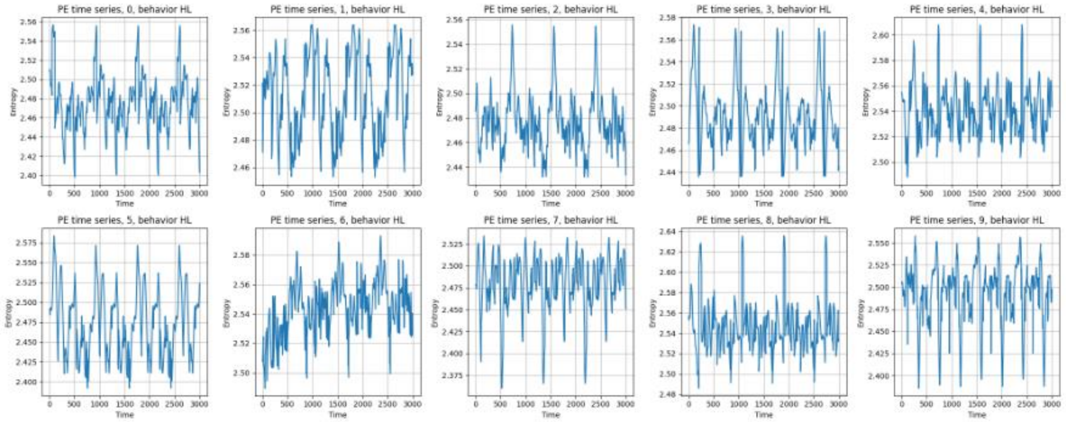




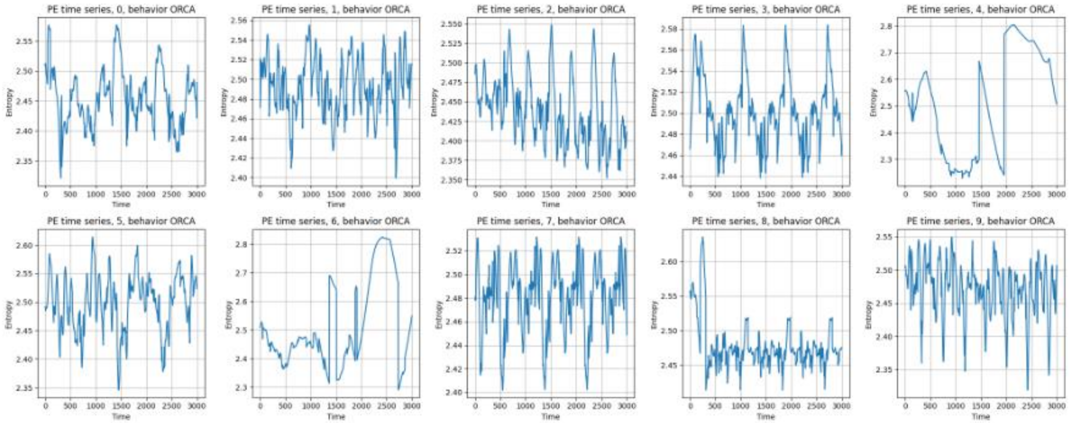
REXASI-PRO

Comparing PE time series of simulaions with different behaviors

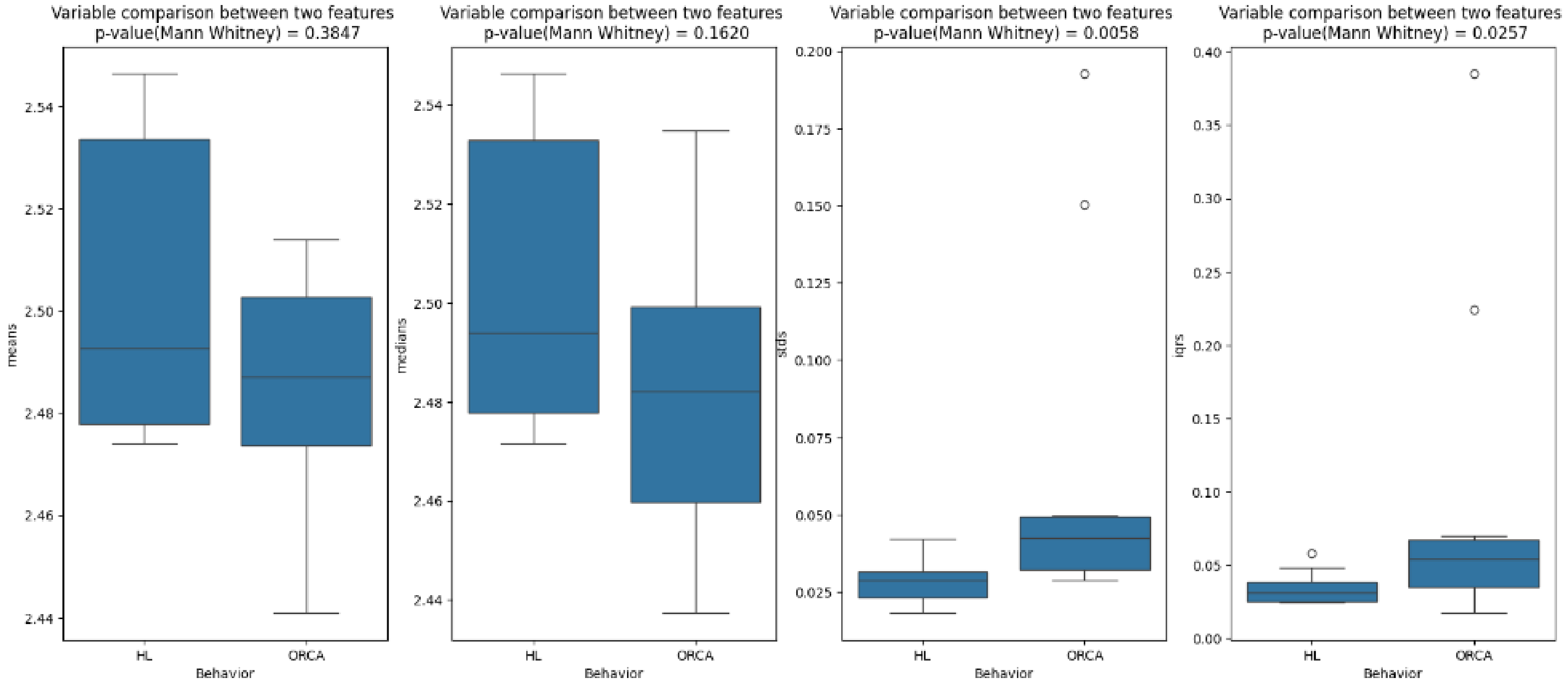
PE time series of simulations with "HL"



PE time series of simulations with "ORCA"



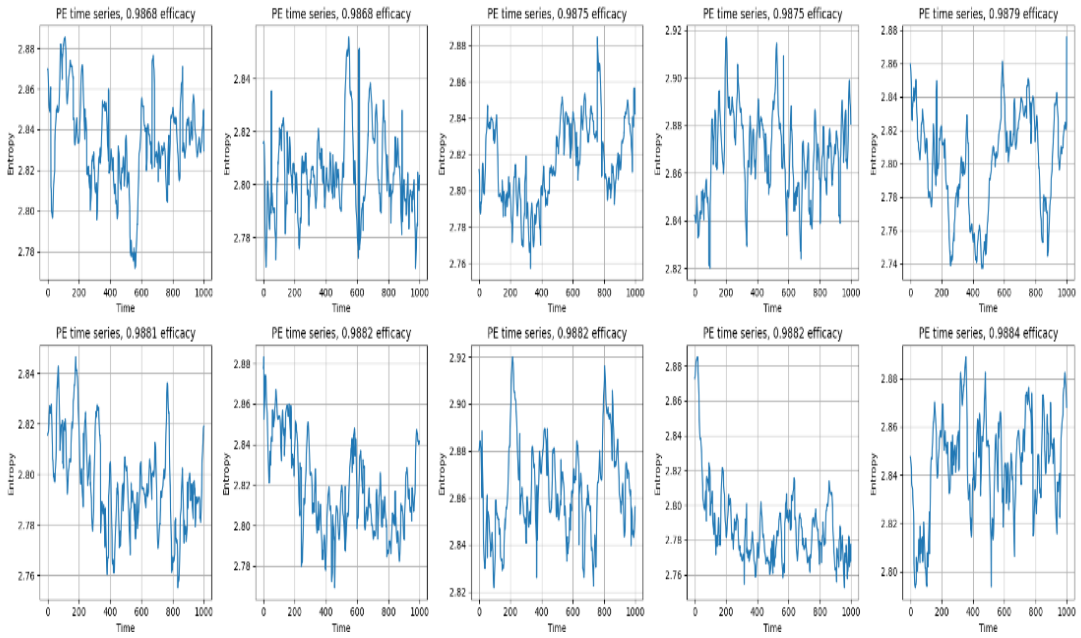
REXASI-PRO



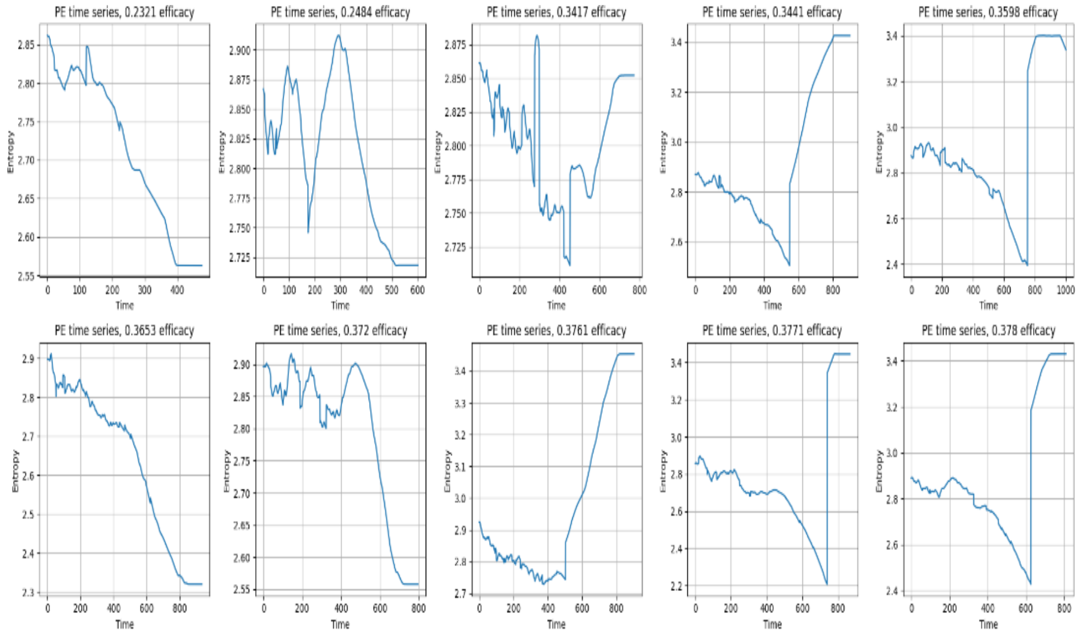
Variable comparison between two features
p-value(Mann Whitney) = 0.3847
means
HL
ORCA
Behavior
Variable comparison between two features
p-value(Mann Whitney) = 0.1620
medians
HL
ORCA
Behavior
Variable comparison between two features
p-value(Mann Whitney) = 0.0058
stds
HL
ORCA
Behavior
Variable comparison between two features
p-value(Mann Whitney) = 0.0257
iqrs
HL
ORCA
Behavior

REXASI-PRO

PE time series of simulations with high efficacy



PE time series of simulations less efficacy



Take home message

Topological approaches and methods (TDA) in ML and AI:

- New methods based on topology are emerging, offering strong mathematical guarantees and providing robust, global, and often interpretable information or features.
- While TDA remains an active área of research, it already has real-world applications.
- There are clear advantages when TDA is combined with or integrated into ML and AI techniques, as enhancing performance and model understanding.
- High quality open-source software is available, such as GUDHI or Ripser, which are accesible and easy to use, with good SOAT.

Thank you for your attention! ☺

If you have any question, please shoot!