# Algebraic topology

...

# Persistence of point clouds

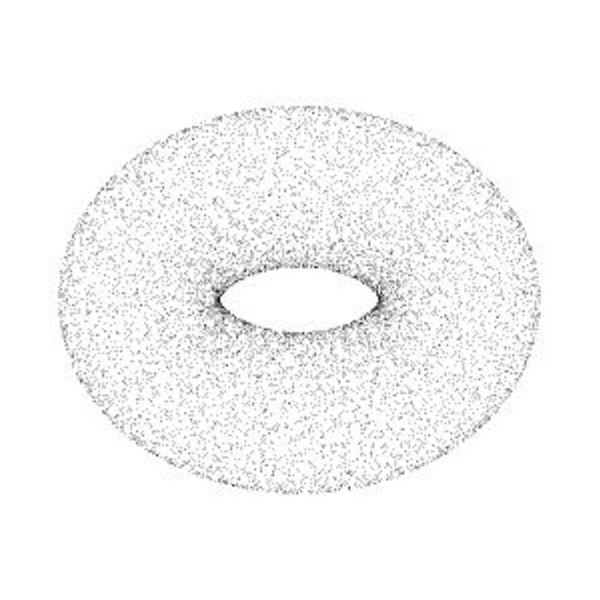
# Persistence of point clouds

Remember the definition of persistence on a point cloud:

- A collection of points that are unorderly distributed in n-dimensional space is a point cloud

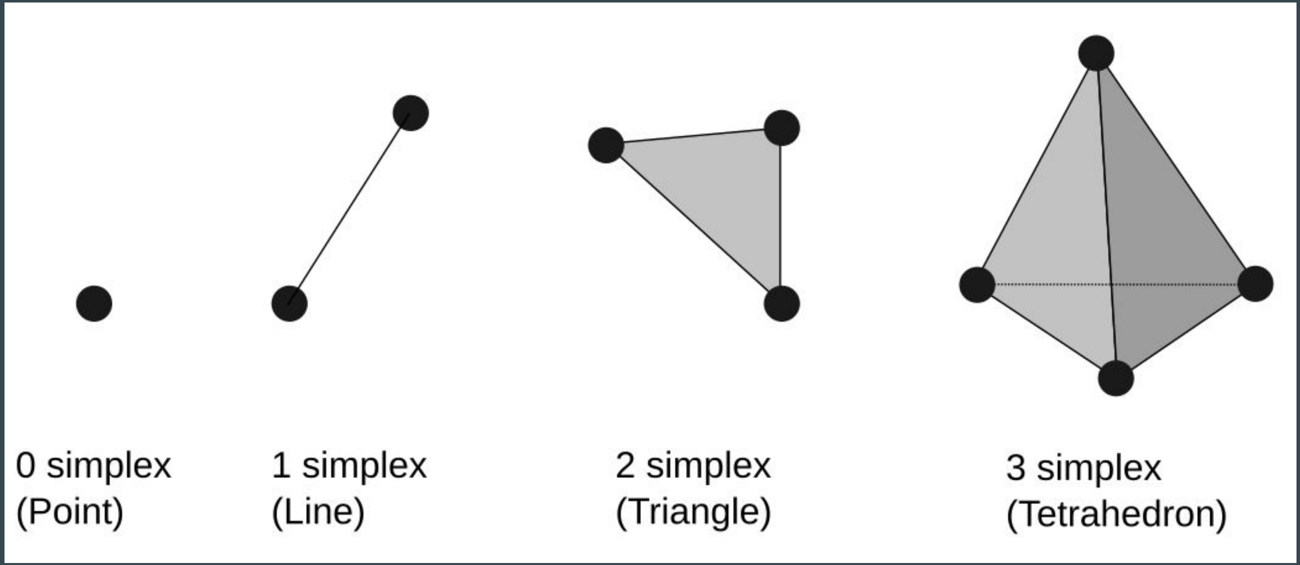
# Point cloud

An example of a point cloud is a set of points uniformly distributed on any geometric shape, such as the torus:



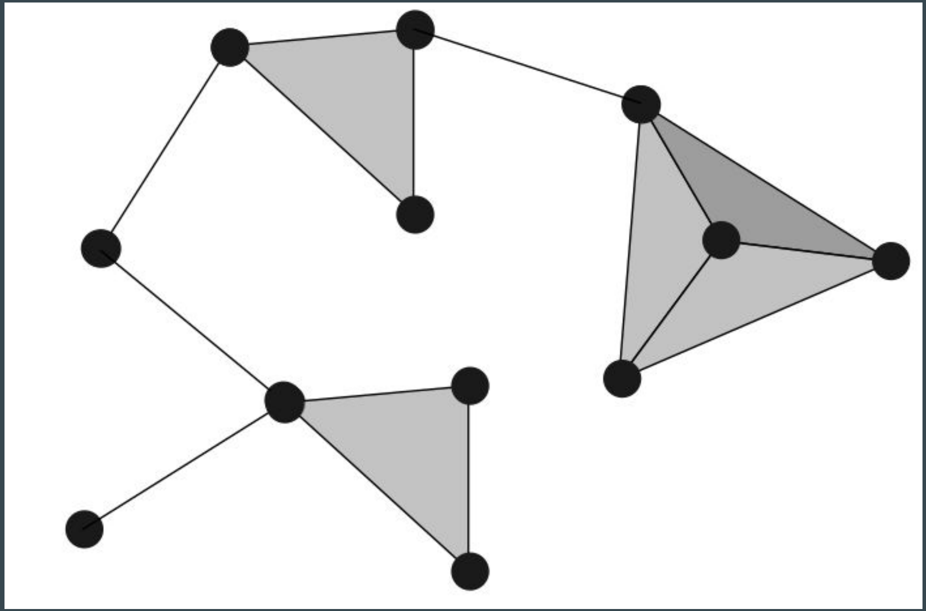
# Simplices

In Topological Data Analysis (TDA), a simplex is a geometric object that generalizes the notion of a triangle to higher dimensions. A simplex of dimension k is defined as the convex hull of (k+1) affinely independent points in Euclidean space.



# Simplicial complexes

Intuitively, a simplicial complex is a collection of simple building blocks (the simplices) that are glued together in a way that preserves their combinatorial structure.



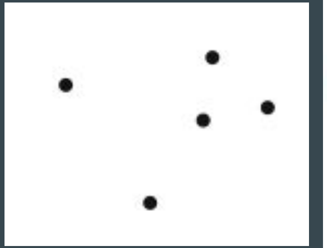
# Persistence of point clouds

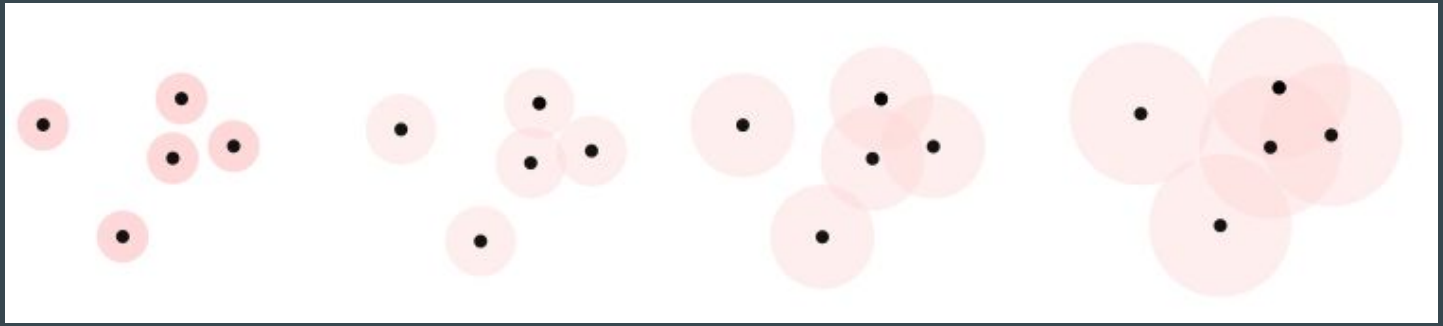
Remember the definition of persistence on a point cloud:

- A collection of points that are unorderly distributed in n-dimensional space is a point cloud
- A filtration is a sequence of simplicial complexes that is used to track the evolution of the topological features of a data set.

# Filtration of circles with increasing radii

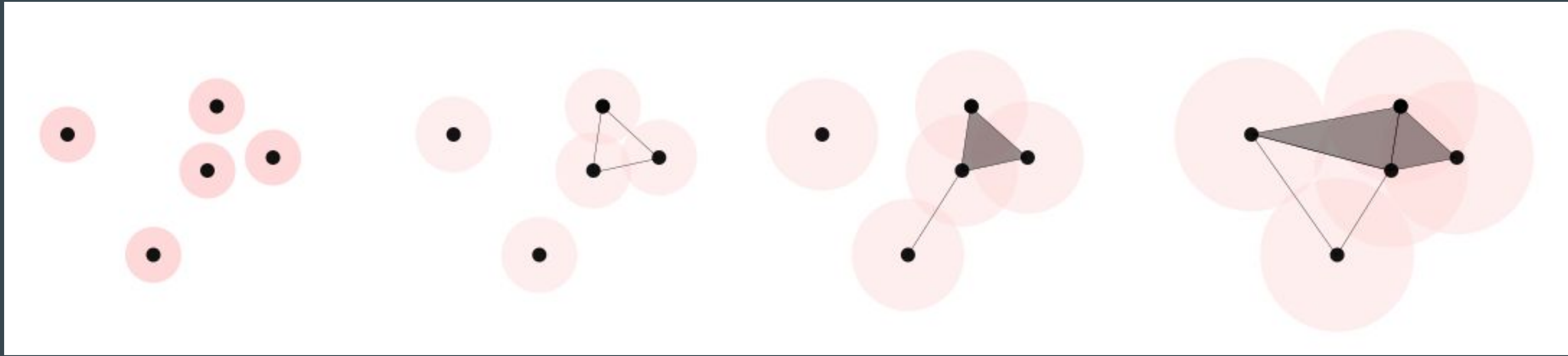
For a collection of points (in 2D), we add balls around each point of radius ε, for increasing values of ε.





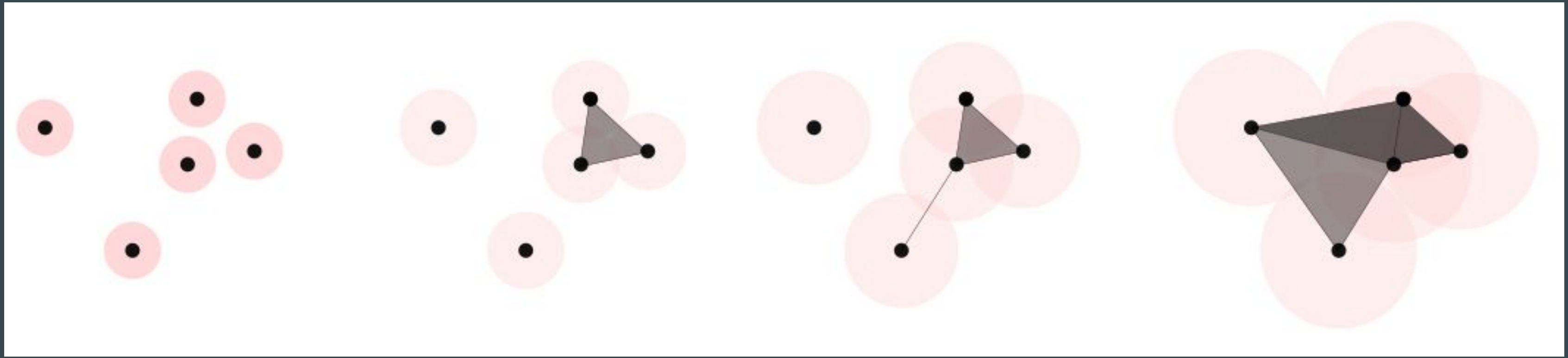
# Čech complexes

In Čech complex, two points are connected if their pairwise distance is less than epsilon. Similarly, for three points to be connected and form a triangle, all epsilon circles should intersect.



# Vietoris - Rips complexes

In Vietoris-Rips complex, two points are connected if their pairwise distance is less than epsilon. Each higher dimensional simplex is added to the simplicial complex, as long as all the points are already connected by lines.



# Persistence of point clouds

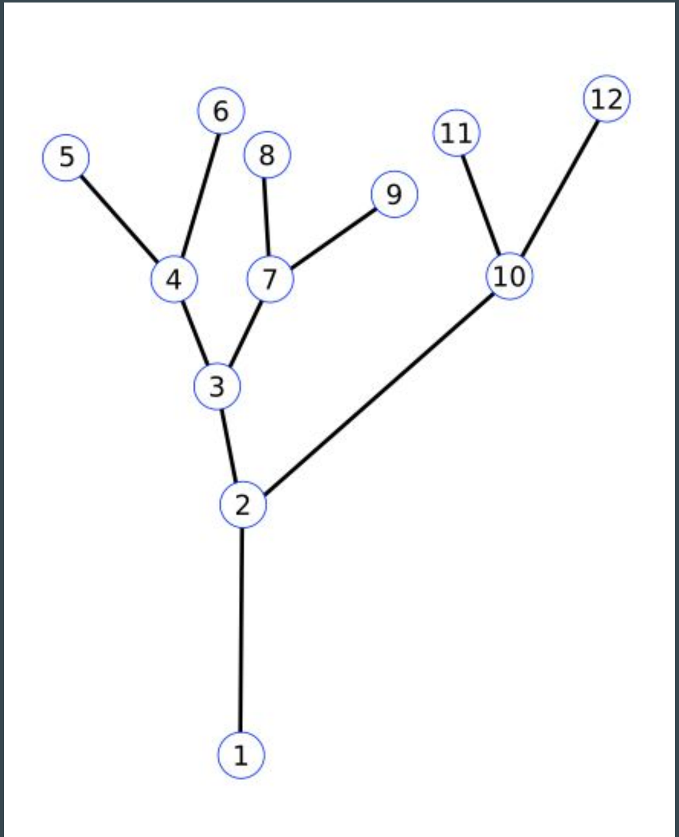
Remember the definition of persistence on a point cloud:

- A collection of points that are unorderly distributed in n-dimensional space is a point cloud
- A filtration is a sequence of simplicial complexes that is used to track the evolution of the topological features of a data set.
- A way to keep track of the evolution of topological features is the persistence barcode. The persistence barcode keeps track of the first time (birth time) a simplex is observed in a simplicial complex, and the time (death time) it merges with a larger component, according to a reference filtration parameter

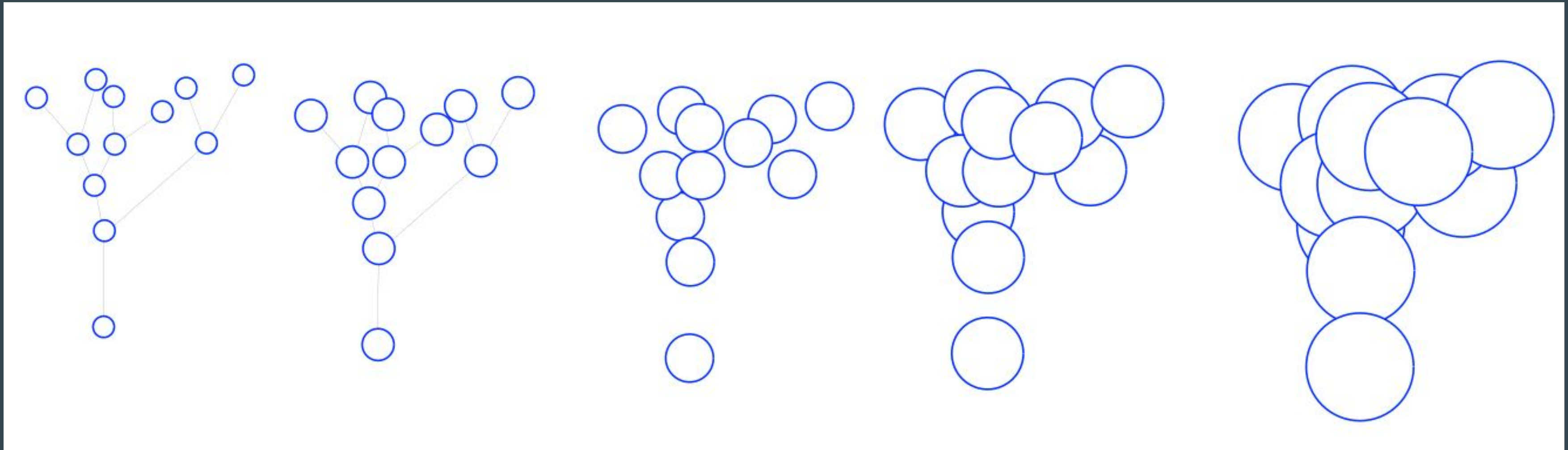
# Tree filtrations

# Tree filtrations

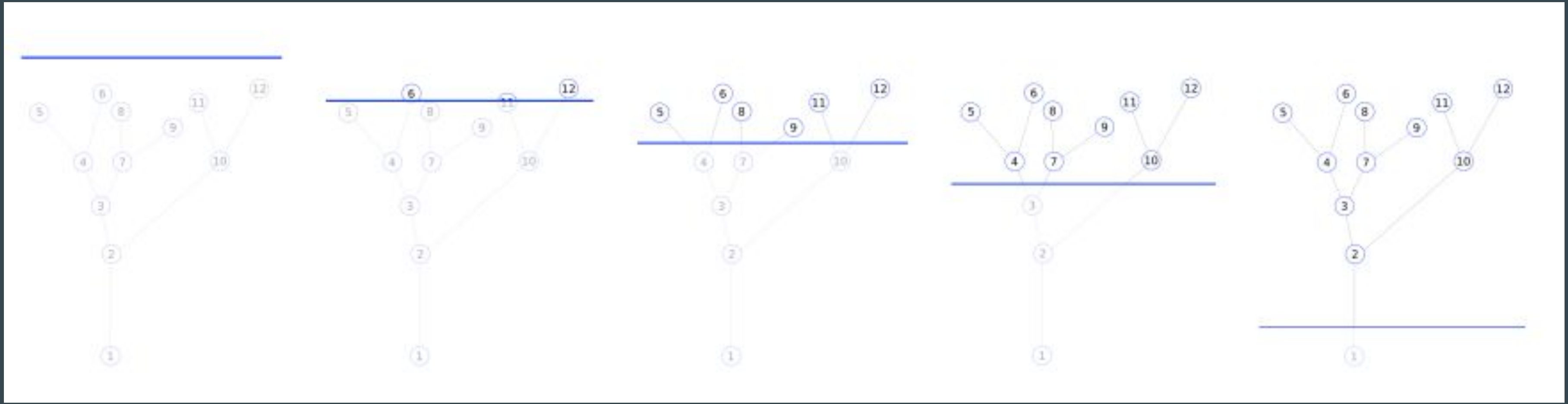
There is a variety of filtration functions that can be performed on a tree structure



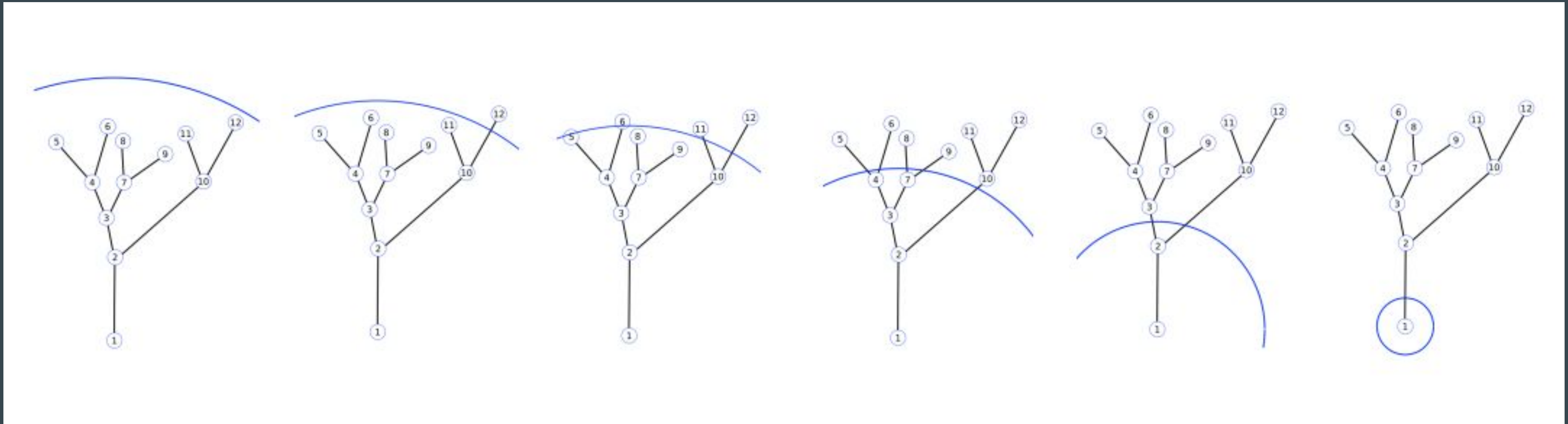
# Tree filtrations



# Tree filtrations



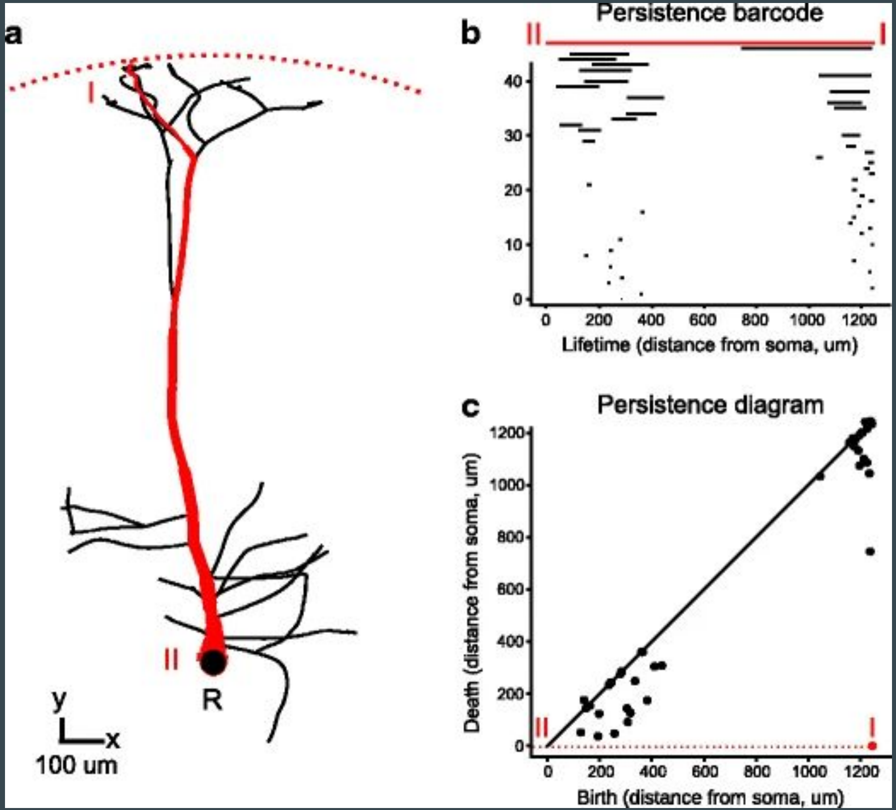
# Tree filtrations



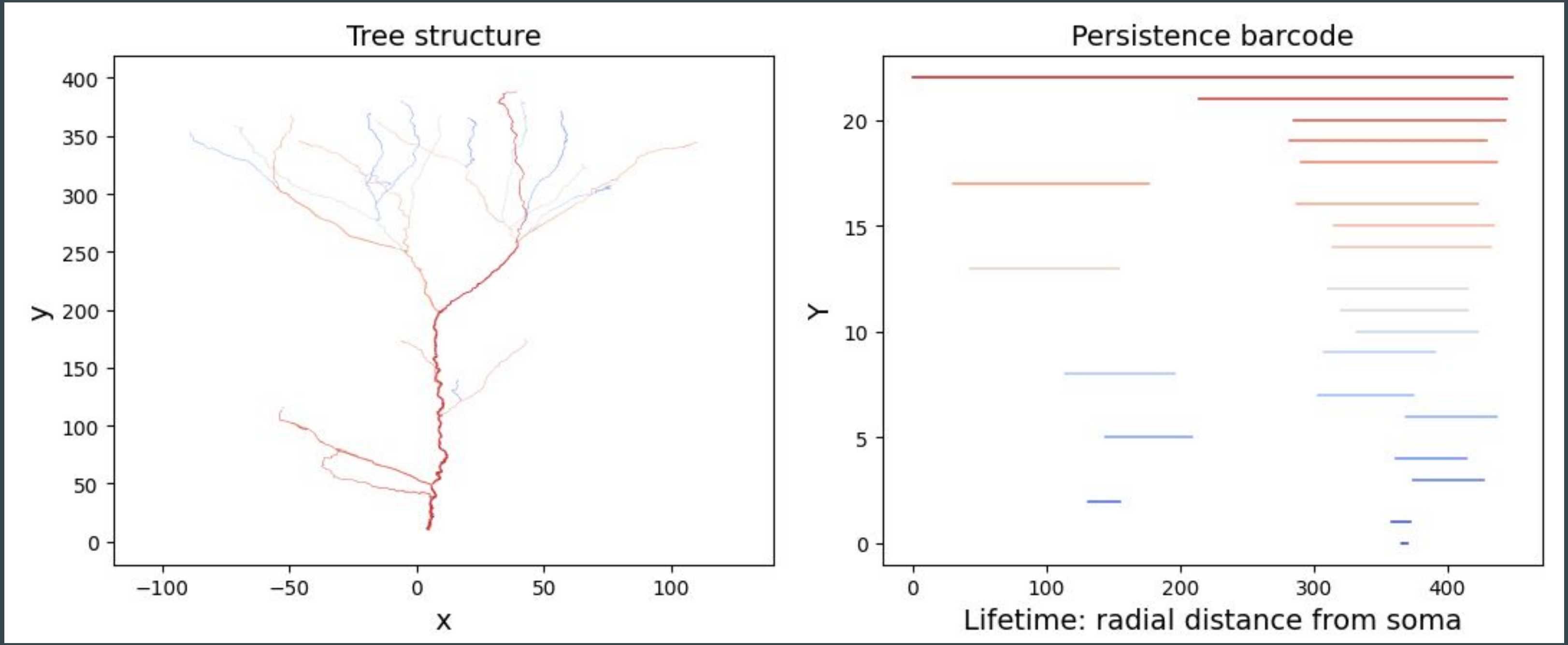
# Topological Morphology Descriptor

# Topological Morphology Descriptor

The persistence barcode (B) of a tree (A) represents each component as a horizontal line whose endpoints mark its birth and death depending on the choice of the function f used for the ordering of the nodes of the tree. For example, f is the radial distance of the nodes from the root (R). The largest component is shown in red together with its birth (I) and death (II). The persistence barcode can be equivalently represented as points in a persistence diagram (C) where the birth (I) and death (II) of a component are the X and Y coordinates of a point respectively (in red).



# Topological Morphology Descriptor




Tree decomposition into a barcode from longer (red) to shorter (blue) components

# Topological Morphology Descriptor

Definitions

Given a tree T with vertices $v_i, i \in [0, n]$ with leaves $l_j, j \in [0, L]$

And a function f applied on the vertices: $f(v_i)$

The TMD persistence of a tree is a collection of intervals that represent the topological components of the tree

$$B(T) = \{(b_j, d_j), j \leq L)\}$$

# Topological Morphology Descriptor

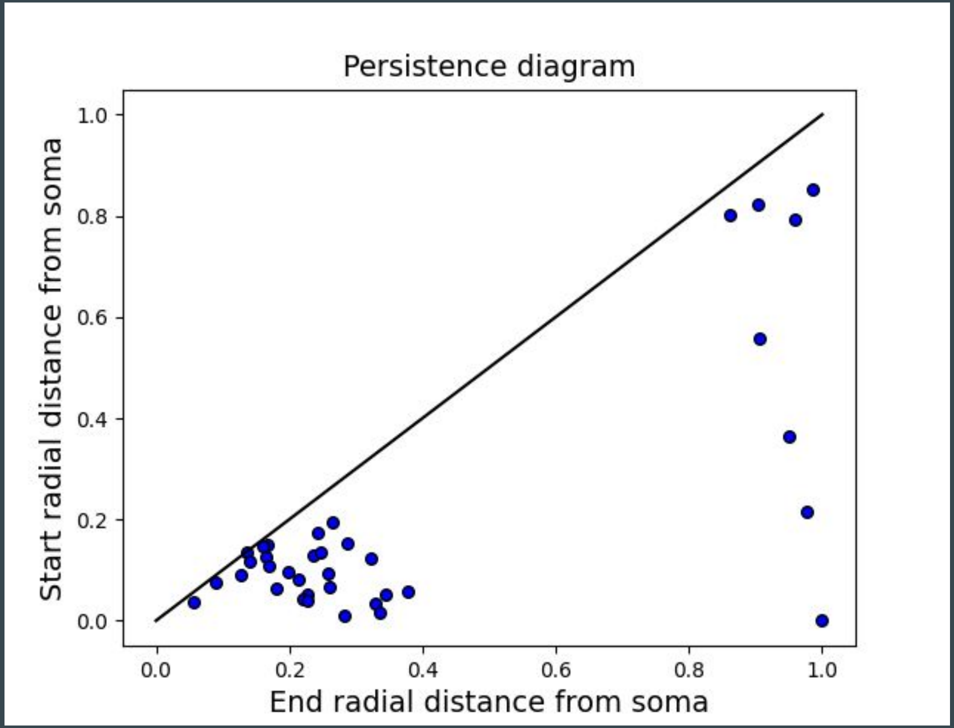
Algorithm

Given a tree T with vertices $v_i$ and leaves $l_j$, and a function f applied in all vertices $f(v_i)$ the TDM of the tree is given by the following algorithm:

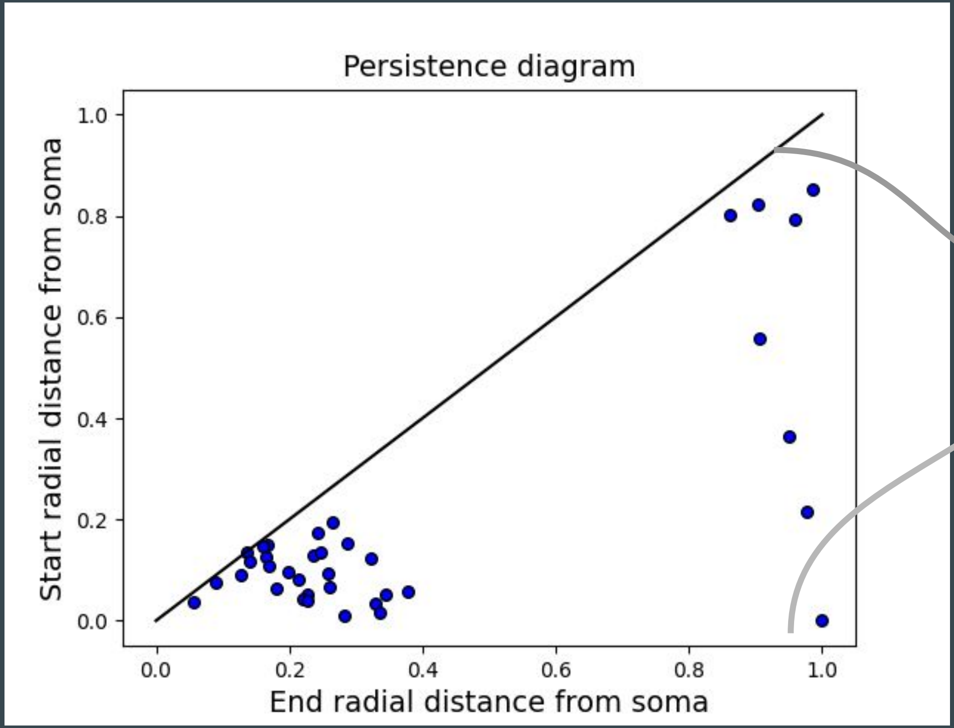
1. Collect all leaves
2. Find all parents of leaves (parent is a vertex one step towards the root)
3. Find all siblings (vertices that share the same parent)
4. Compare their values f, the larger value persists according to Elder rule
5. Repeat the process until the root is reached

# Topological Morphology Descriptor (intuition)

# Topological Morphology Descriptor - Number of branches

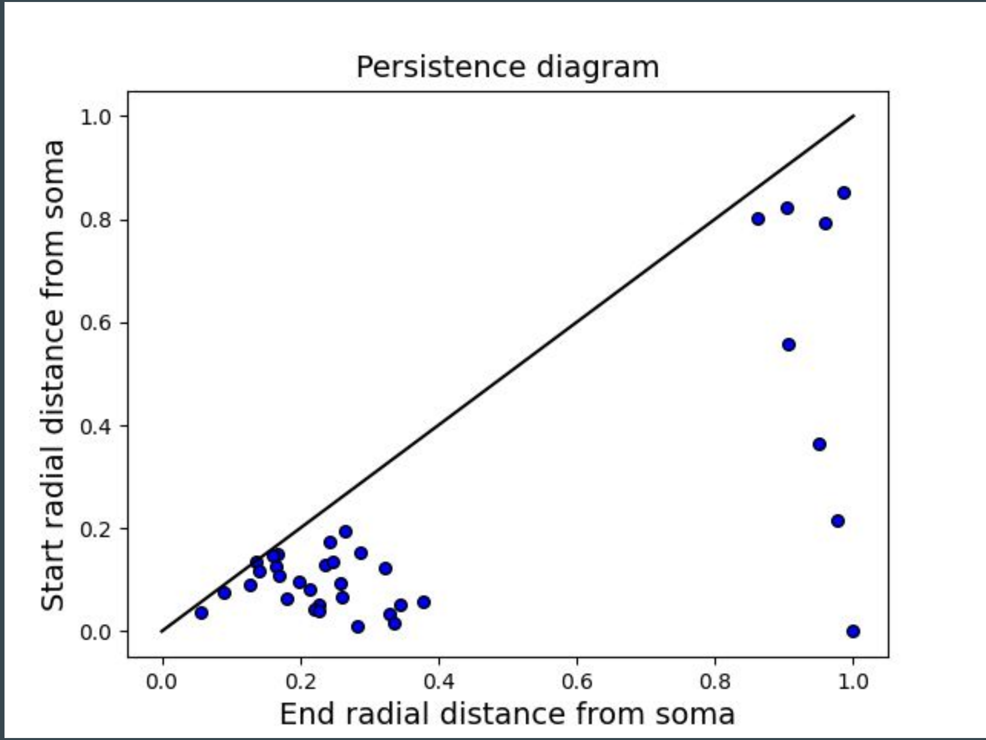


# Topological Morphology Descriptor - Number of branches

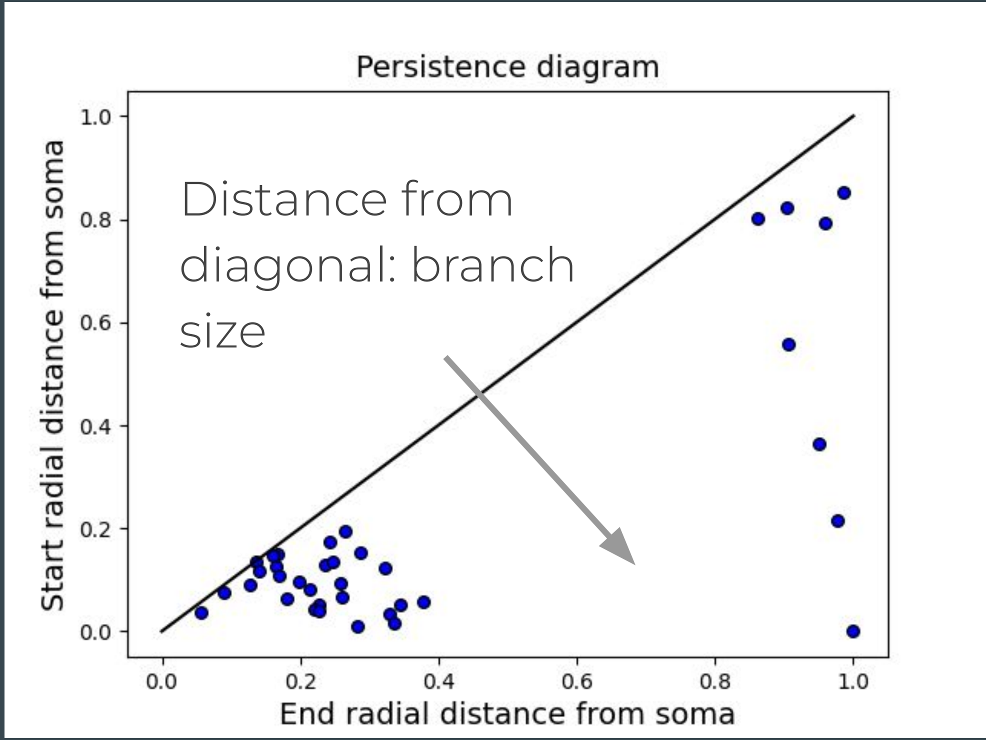



Number of points -> number of branches in the tree

# Topological Morphology Descriptor - Branch length

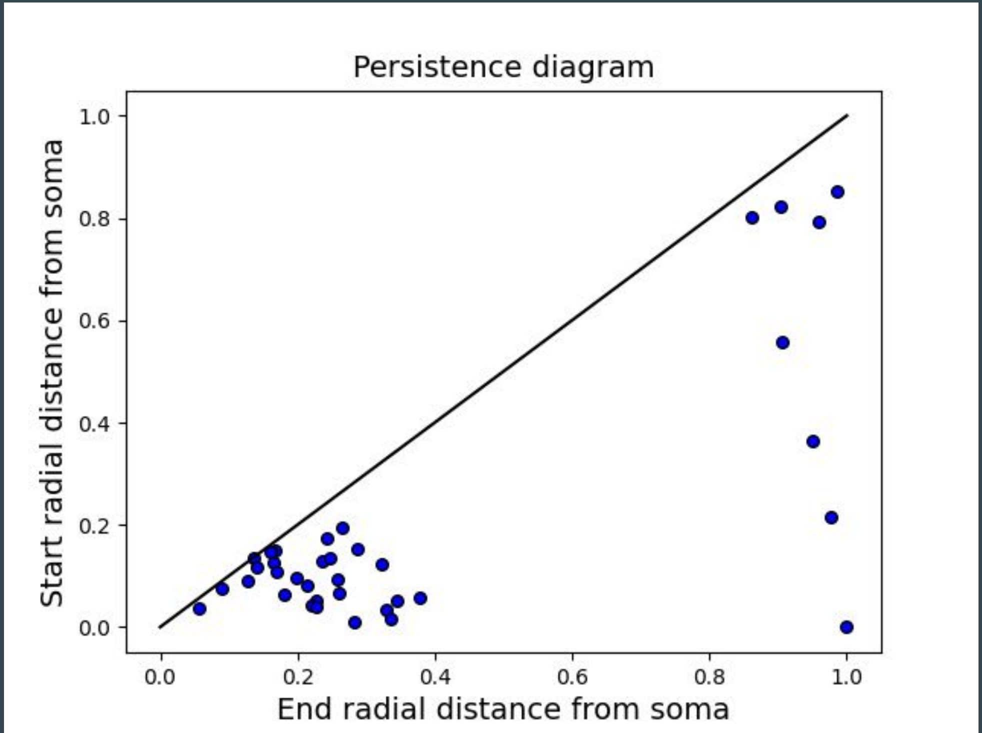


# Topological Morphology Descriptor - Branch length

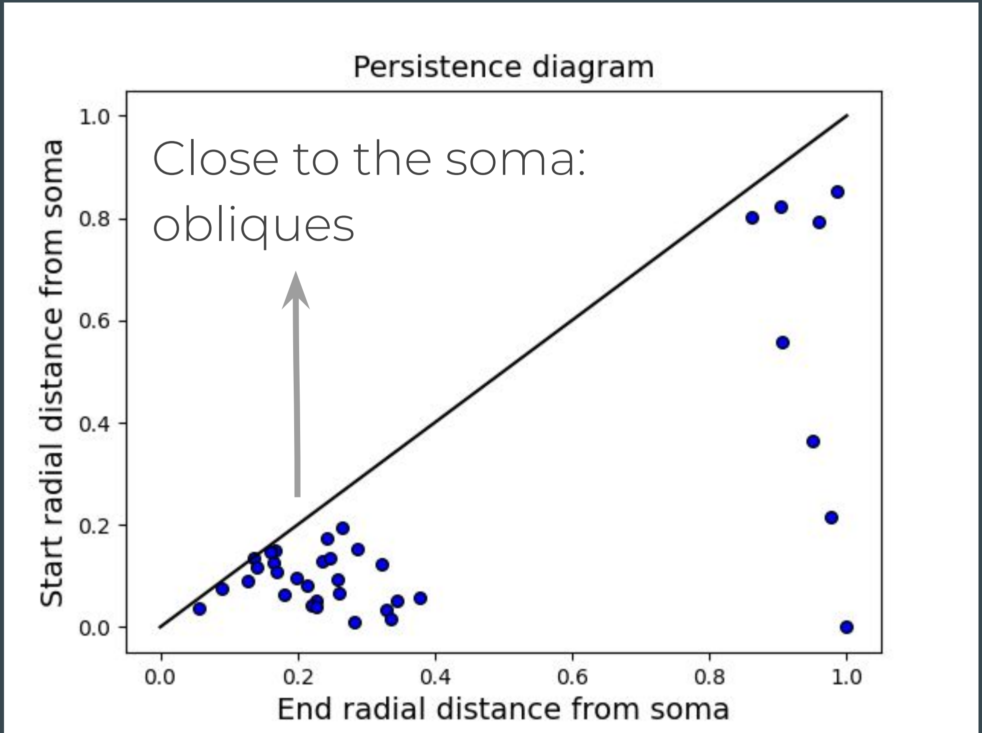



Far from diagonal: largest branch, corresponds to apical main trunk

# Topological Morphology Descriptor - obliques from tuft



# Topological Morphology Descriptor - obliques from tuft




Far from the soma:
apical tufts

## How can we compute the total length?

# Representations of persistence

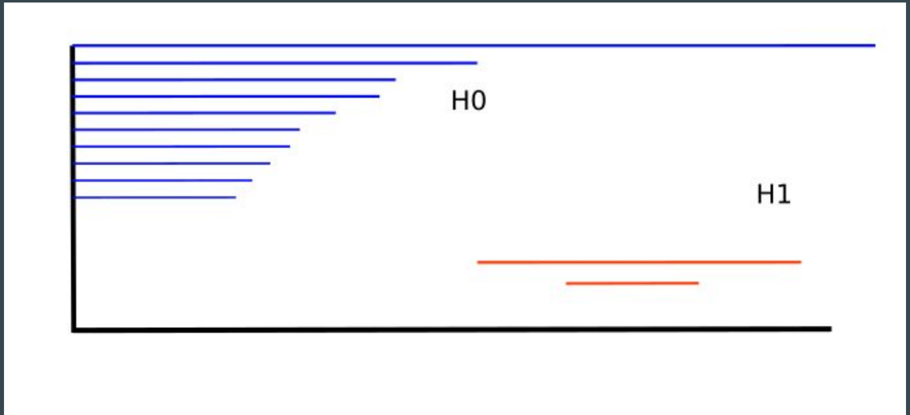
# Topological summaries

Many data sets can be viewed as a noisy sampling of an underlying space, and tools from topological data analysis can characterize this structure for the purpose of knowledge discovery. One such tool is persistent homology, which provides a multiscale description of the homological features within a data set. The topology of an object can be summarized in the information encoded in the barcode.

However this information can be rearranged and presented in a variety of methods. The use of TDA has been limited by the difficulty of combining the main tools, such as the barcode or persistence diagram with statistics and machine learning. There are many summaries that allow the combination of TDA with machine learning and statistics.

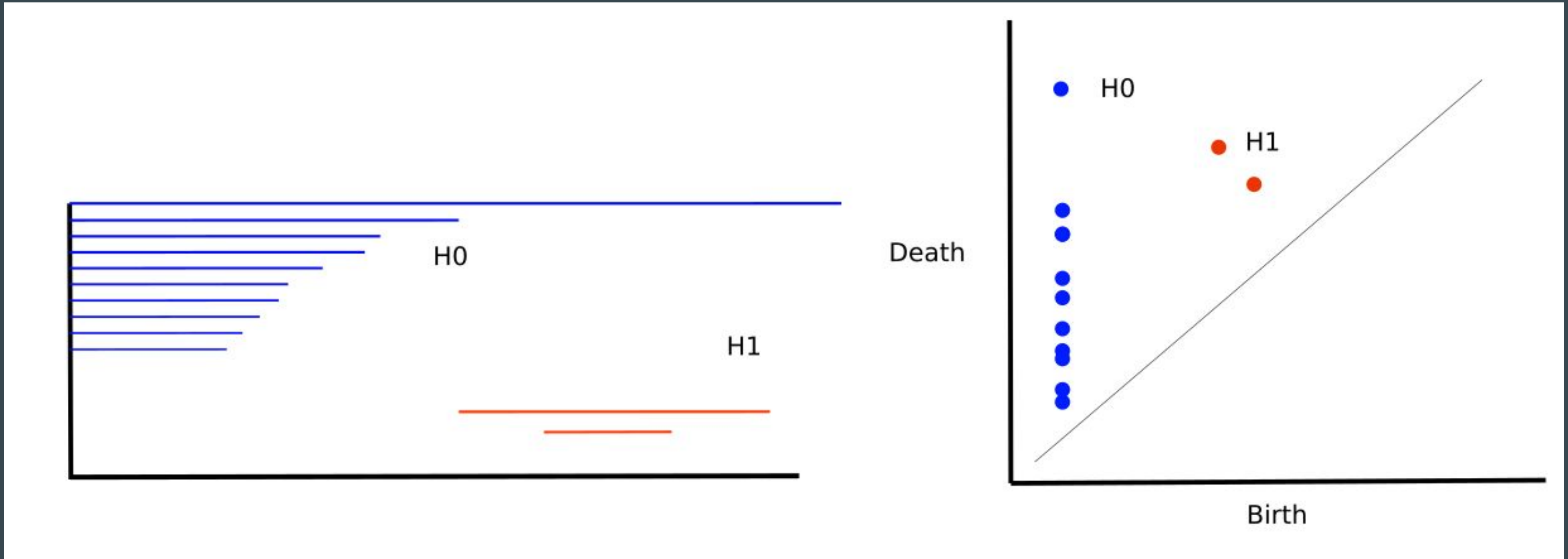
# Persistence barcodes

The two standard topological summaries of data are the barcode and the persistence diagram. They represent almost the same dataset (orientation of the lifetime is lost in the barcode). A barcode illustrates the lifetime of each component in the underlying topological structure.



# Persistence diagrams

The persistence diagram encodes the start and end time of each component in the 2D plane



# Persistence diagrams / barcodes

Both of these representations are powerful visual tools to investigate the underlying topological properties:

- How many components are present at each stage
- How many higher structures (holes) are observed
- If there are infinite persistent features

However…

# Persistence diagrams / barcodes

The problem with this representations is:

- They cannot be used as input to machine learning as they are not vectorized
- They cannot be used directly to compute averages for a group of objects
- It is not straightforward to define a distance between them

That's why different vectorization techniques have been proposed.

# Betti curves

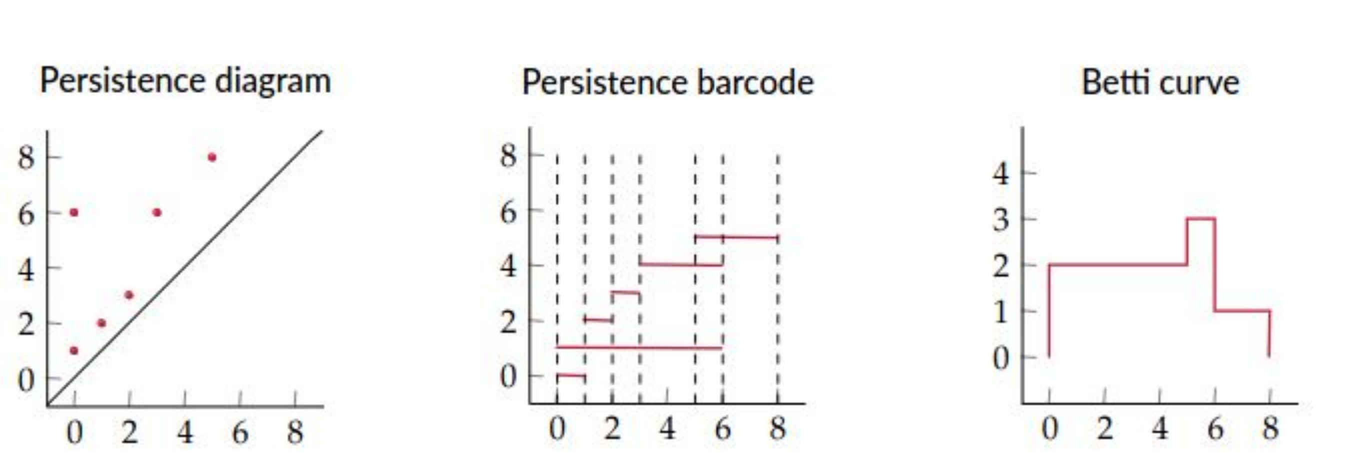
The Betti curve is a function mapping a persistence diagram to an integer-valued curve, i.e. each Betti curve is a function $B : \mathbb{R} \to \mathbb{N}$.

It sums the number of bars at each parameter level. This representation is:

- Easy to calculate
- Simple representation, ‘living’ in the space of piecewise linear functions
- Vector space operations are possible (addition, scalar multiplication)
- Distances and kernels can be defined

# Betti curves

A persistence diagram (a), its persistence barcode (b), and its corresponding persistence indicator function (c)




Rieck et al. 2019

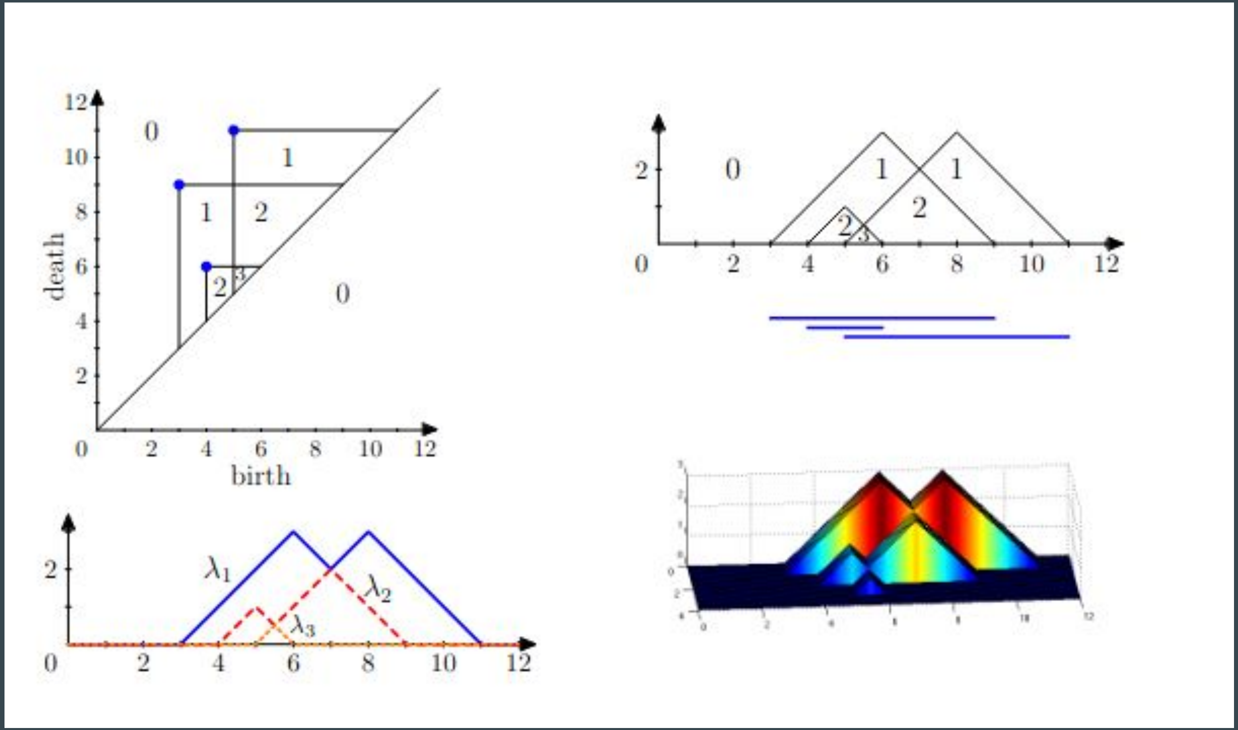
# Persistence landscapes (Bubenik 2016)

A topological summary for data introduced in 2016 by Bubenik is the persistence landscape. Since this summary lies in a vector space, it is easy to combine with tools from statistics and machine learning, in contrast to the standard topological summaries. Viewed as a random variable with values in a Banach space, this summary obeys a strong law of large numbers and a central limit theorem.

# Persistence landscapes

Statistical Topological Data Analysis using Persistence Landscapes

Peter Bubenik 2016

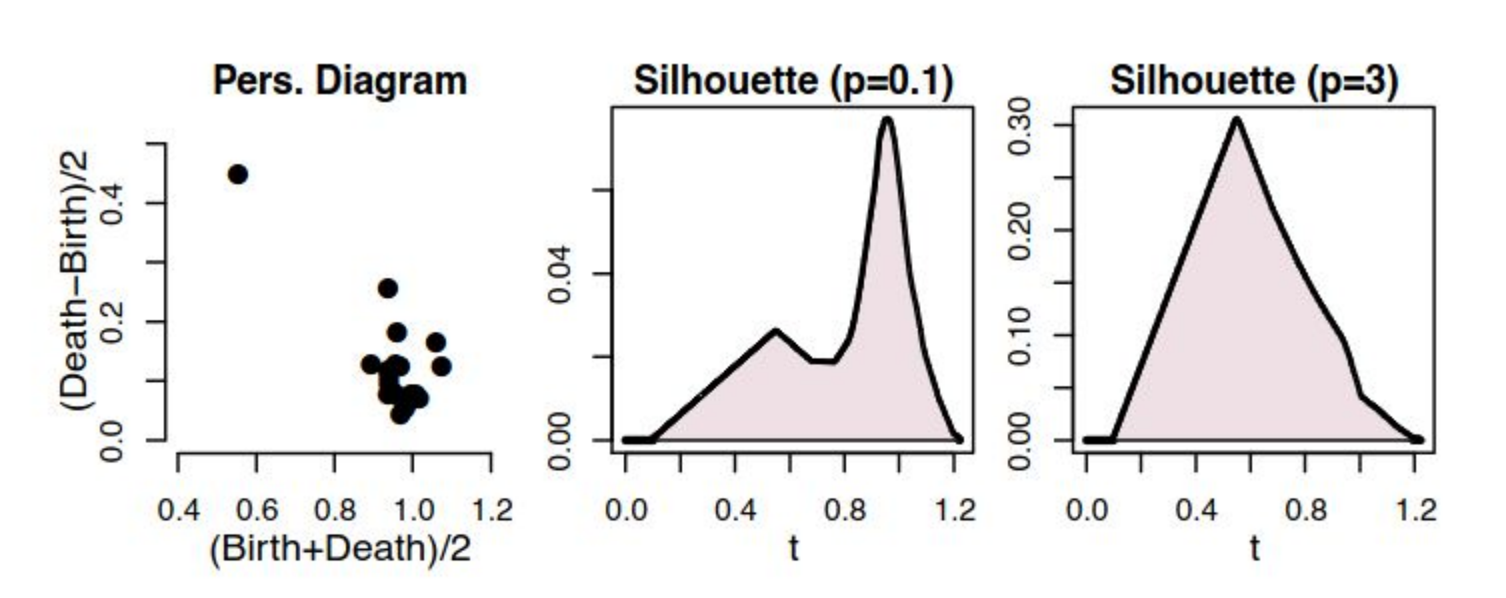


# Persistence silhouette (Chazal et al. 2014)

Persistent homology is a widely used tool in Topological Data Analysis that encodes multiscale topological information as a multiset of points in the plane called a persistence diagram. It is difficult to apply statistical theory directly to a random sample of diagrams. Instead, we need to summarize persistent homology with a vectorized version, which converts a diagram into a well-behaved real-valued function. In 2014, Chazal et al. introduced an alternate functional summary of persistent homology, the persistent silhouette.

# Persistence silhouette

The persistence silhouette, similar to the landscape transforms the diagram to a vectorized version that can be used as input for statistics and machine learning

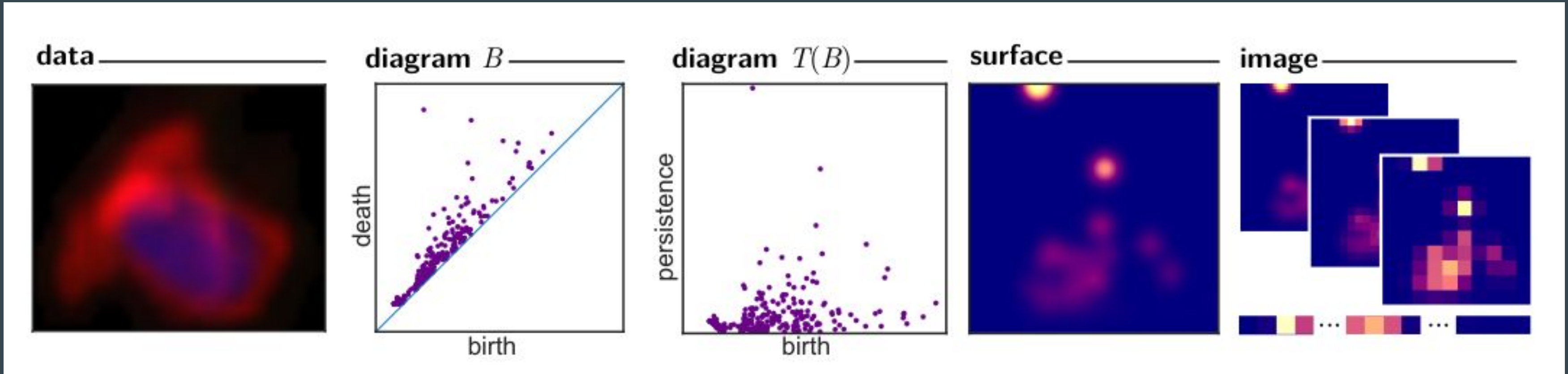


# Persistence images (Adams et al. 2017)

Another useful representation of homological information is a persistence diagram (PD). Efforts have been made to map PDs into spaces with additional structure valuable to machine learning tasks. Adams et al. (2017) convert a PD to a finite-dimensional vector representation, the persistence image (PI). The discriminatory power of PIs is compared against existing methods, showing significant performance gains.

# Persistence images

Algorithm pipeline to transform data into a persistence image. From the data, to the diagram, to surfaces that can be encoded into vector images.



# Projects

# Projects

1. General idea, participants, title
2. Objectives
3. Methods
4. Datasets