TDA lecture

Priyavrat Deshpande

Chennai Mathematical Institute

February 15, 2022

Outline

Simplicial Complexes - I

Definition (Affine independence)

Points $v_0, \ldots, v_k \in \mathbb{R}^N$ are affinely independent if

$$\left(\sum_{i=0}^{k} t_i v_i = 0 \text{ and } \sum_{i=0}^{k} t_i = 0 \right) \Rightarrow t_0 = t_1 = \cdots = t_k = 0.$$

Definition (A k-simplex and its faces)

A k-simplex σ is the convex hull of $k+1$ affinely independent points $\{v_0, \ldots, v_k\}$. Denoted

$$\sigma = [v_0, \ldots, v_k].$$

A face of σ is a subset of v_i's. In particular, singletons are called vertices.

Simplicial Complexes - II

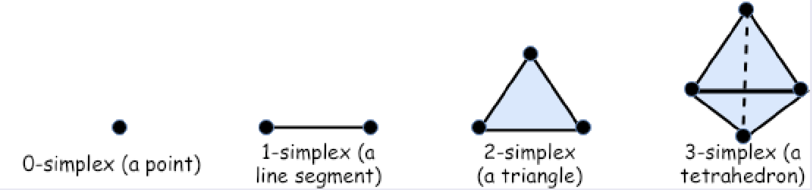
Building blocks

- 0-simplex: single point.
- 1-simplex: line segment.
- 2-simplex: filled triangle.
- 3-simplex: filled tetrahedron.

Simplicial Complexes - II

Building blocks

- 0-simplex: single point.
- 1-simplex: line segment.
- 2-simplex: filled triangle.
- 3-simplex: filled tetrahedron.



Simplicial Complexes - III

Definition (Simplicial complex K)

It is (finite) union of simplices such that

i for all $\sigma \in K$ all the faces of σ are also in K;

ii the intersection of any two simplices is either empty or a common face.

Simplicial chains

- Let $\{\sigma_1, \ldots, \sigma_p\}$ be the set of k-simplices of K.
- A simplicial k-chain is a linear combination $c := \sum_{i=1}^{p} \epsilon_i \sigma_i$ where $\epsilon_i \in \mathbb{F}_2$.
- The set of all simplicial chains, $C_k(K)$, form a vector space over $\mathbb{F}_2$.
- We have $\dim C_k(K) = p$.

Definition (The boundary operator)

The linear map $\partial_k : C_k(K) \to C_{k-1}(K)$ is given by

$$\partial_k([v_{i_0} \ldots, v_{i_k}]) = \sum_j [v_{i_0}, \ldots, \widehat{v_{i_j}}, \ldots, v_{i_k}]$$

Important Property

$$\partial_{k+1} \circ \partial_k = 0.$$

The homology

Definition (The simplicial chain complex)

$$0 \to C_n(K) \xrightarrow{\partial_n} C_{n-1}(K) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_2} C_1(K) \xrightarrow{\partial_1} C_0(K) \to 0.$$

Definition

The simplicial homology groups

$$H_k(K, \mathbb{F}_2) := \frac{\ker \partial_k}{\mathrm{im}\ \partial_{k+1}}.$$

Definition (Betti numbers)

For $1 \leq i \leq n$

$$\beta_i(K) := \dim H_i(K, \mathbb{F}_2).$$

Topological invariance

- The simplicial homology is a functor from simplicial complexes to abelian groups.
- A homeomorphism of simplicial complexes induces an isomorphism of homology groups in all dimensions.

Outline

Data to space: the idea

The most common type of data is a point cloud — a set of vectors $\mathbb{X} = \{x_1, \ldots, x_N\}$ in $\mathbb{R}^d$.

A PCD is a 0-dimensional simplicial complex, no interesting topology

Is there a nice topological space (say a manifold) M such that $\mathbb{X} \subset M$?

Approximating manifolds I: offsets

Definition (Offset)

Given a compact set Y in $\mathbb{R}^d$ and a real number $r > 0$ the r-offset of Y is defined as

$$Y^r := \bigcup_{y \in Y} B(y; r).$$

Definition (Reach)

Let Y be as above.

$$\begin{aligned}\operatorname{reach}(Y) := \sup\{r \in \mathbb{R} \mid \forall y \notin Y, \text{ and } d(y, Y) < r, \\ \exists! z \in Y \text{ such that } d(y, z) = d(y, Y)\}.\end{aligned}$$

Definition

Hasudorff distance Let X, Y be two compact subsets of $\mathbb{R}^d$. Then

$$d_H(X, Y) := \inf\{\alpha > 0 \mid X \subset Y^\alpha \text{ and } Y \subset X^\alpha\}.$$

Approximating manifolds II: reconstruction theorems

Theorem (Chazal and Lieutier 2007)

Let $\mathbb{X}$ be a PCD and M be closed manifold in $\mathbb{R}^d$ and let $\epsilon > 0$ be such that $d_H(\mathbb{X}, M) < \epsilon$. Further assume that reach of both the sets is at least 2ϵ. Then there is an r with $0 < r < 2\epsilon$ such that $\mathbb{X}^r$ and M^r are homotopy equivalent.

Theorem (Niyogi et al. 2008)

The Betti numbers of Riemanninan manifolds with positive reach can be recovered with high probability from offsets of a sample on (or close to) the manifold.

How to code offset?

Definition (Cech Complex)

Given a PCD $\mathbb{X}$ and $r > 0$ the Cech complex $Ch_r(\mathbb{X})$ is an (abstract) simplicial complex whose simplices are those subsets $\sigma \in \mathbb{X}$ such that

$$\bigcap_{x \in \sigma} B(x; r) \neq \emptyset.$$

Theorem (Nerve lemma)

The offset $\mathbb{X}^r$ of a PCD $\mathbb{X}$ is homotopy equivalent to the Cech complex $Ch_r(\mathbb{X})$.

Vietoris-Rips complex

Definition

Let $X \subset \mathbb{R}^N$ be a finite point cloud and $\epsilon > 0$. The Vietoris-Rips complex, $VR_\epsilon(X)$, has as k-simplices those $(k+1)$-subsets $\{x_{i_0}, \dots, x_{i_k}\}$ of X for which

$$d(x_{i_j}, x_{i_l}) \leq \epsilon.$$

- In general $VR_\epsilon(X)$ does not embed in $\mathbb{R}^N$.
- If $\epsilon < \epsilon'$ then $VR_\epsilon(X) \subseteq VR_{\epsilon'}(X)$.
- It is a clique complex, i.e., completely determined by its 1-skeleton.

Data to space: an example

Filtered complexes

An increasing sequence $\epsilon_{i_1} < \cdots < \epsilon_{i_n}$ induces a filtration

$$\emptyset \subset VR_{\epsilon_{i_1}}(X) \subseteq \cdots \subseteq VR_{\epsilon_{i_n}}(X).$$

For every $p \geq 0$ we have:

$$H_p(VR_1(X)) \xrightarrow{f_p^{0,1}} H_p(VR_2(X)) \xrightarrow{f_p^{0,2}} \cdots \xrightarrow{f_p^{n-1,n}} H_p(VR_n(X)).$$

In general for $i < j$

$$f_p^{i,j} : H_p(VR_i(X)) \to H_p(VR_j(X)).$$

Persistent homology

$$f_p^{i,j} : H_p(VR_i(X)) \to H_p(VR_j(X)).$$

Definitions

- p-**th persistent homology group**: $\mathcal{H}_p^{i,j} := \mathrm{Im}(f_p^{i,j})$.
- p-**th persistence Betti number**: $\beta_p^{i,j} := \mathrm{rank}(\mathcal{H}_p^{i,j})$.
- **Birth at** i-**th stage**: A class c such that

$$c \in H_p(VR_i(X)) \text{ but } c \notin \mathcal{H}_p^{i-1,i}.$$

- **Death of a class at** j-**th stage**: A class c such that

$$f_p^{i,j-1}(c) \notin \mathcal{H}_p^{i-1,j-1} \text{ and } f_p^{i,j}(c) \in \mathcal{H}_p^{i-1,j}.$$

- The number of cycles that are born at ϵ_i and are dead at ϵ_j is:

$$\mu_p^{i,j} := (\beta_p^{i,j-1} - \beta_p^{i-1,j-1}) - (\beta_p^{i,j} - \beta_p^{i-1,j})$$

Barcodes and diagrams

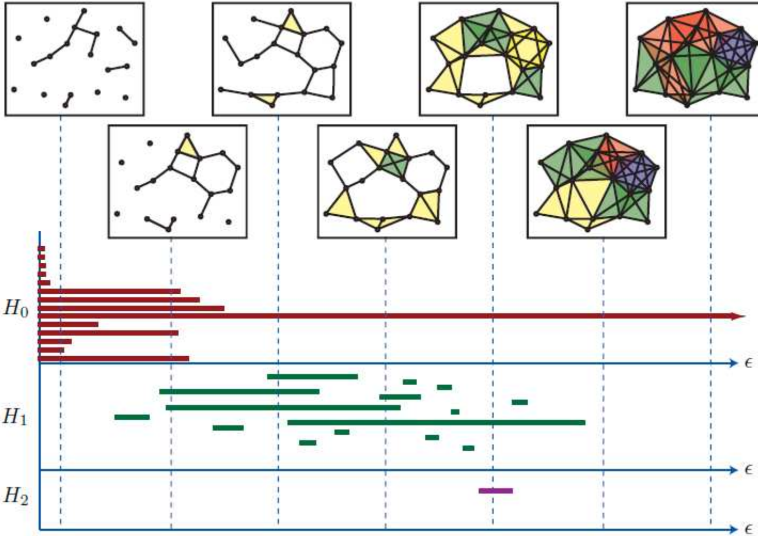
Definition (Persistence barcodes)

For every $p \geq 0$ we draw a graph whose verticle axis corresponds to all possible p-homology generators and the horizontal axis is the time parameter.

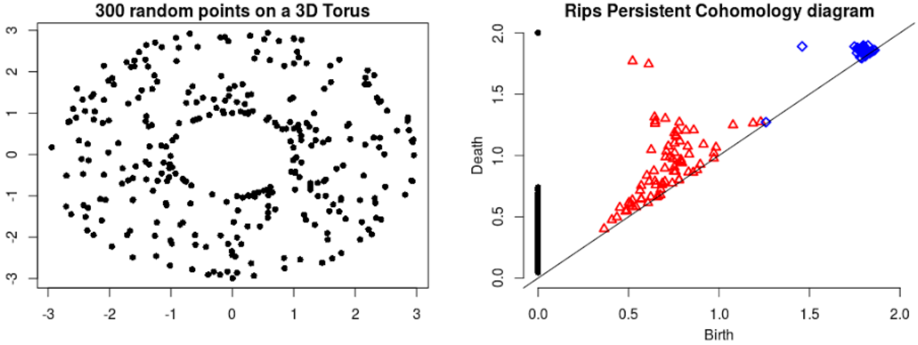
Definition (Persistence diagram)

The p-persistence diagram is a 2-d coordinate system where x is the birth coordinate and y is the death coordinate. For every p-homology class there is a point (i, j) representing its birth and death time.

Example



Persistence diagrmas



How to interprete?

1. High persistence implies existence of **robust** features.
2. Spurious topological features are short-lived, i.e., **noise**.
3. The *summary description* is always 2-dimensional.
4. Persistent diagrams are a similarity metirc.
5. β_0: number of connected components (clusters?).
6. β_1: number of cycles (periodic features?).
7. β_2: number of hollow spaces (?).

Outline

R TDA

- Install `TDA` library in R.

PH using Rips complex

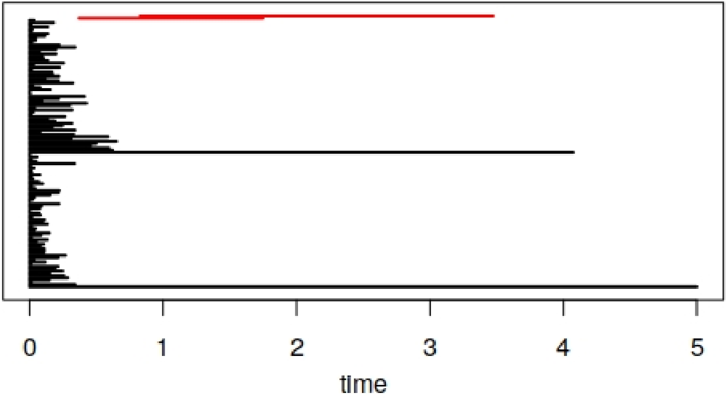
ripsDiag(X, maxdimension, maxscale, dist = "euclidean", library = "GUDHI", location = FALSE, printProgress = FALSE)

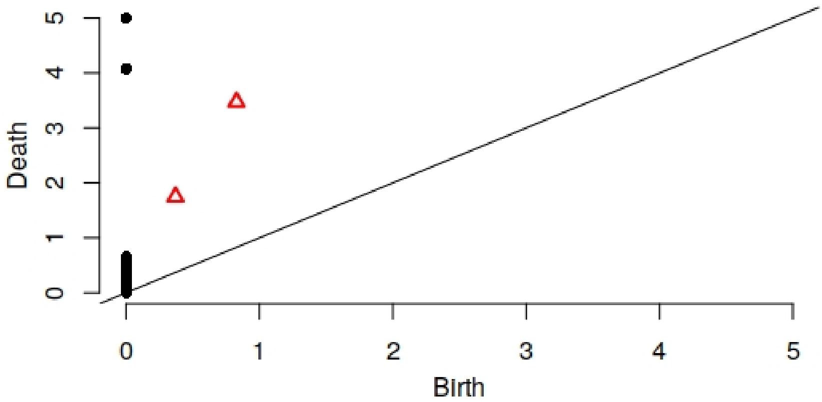
- X is and $n \times d$ matrix of coordinates; $n =$ number of points and $d =$ dimension of the ambient space.
- *maxdimension*: max dimension of the topological feature.
- *maxscale*: maximum value of the filtration.

Plot the diagram

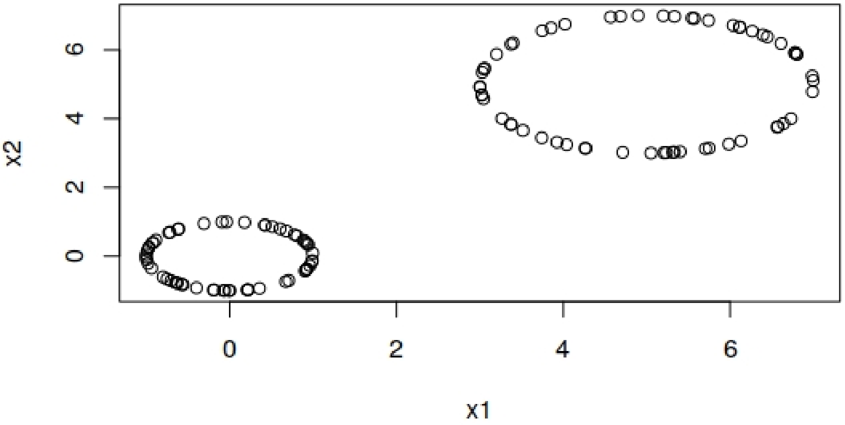
plot(x, diagLim = NULL, dimension = NULL, col = NULL, rotated = FALSE, barcode = FALSE, band = NULL, lab.line = 2.2, colorBand = "pink", colorBorder = NA, add = FALSE)

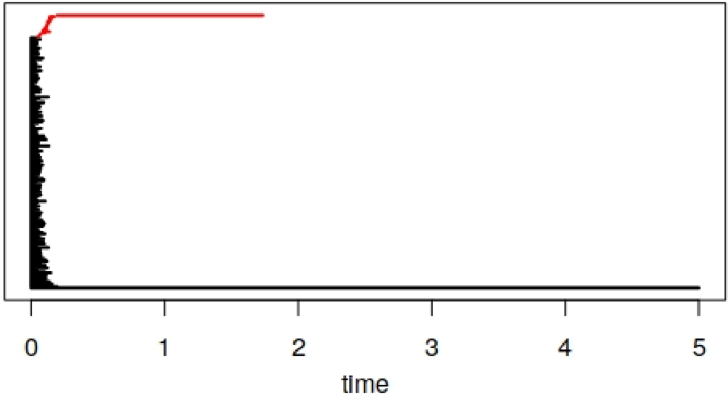
Examples



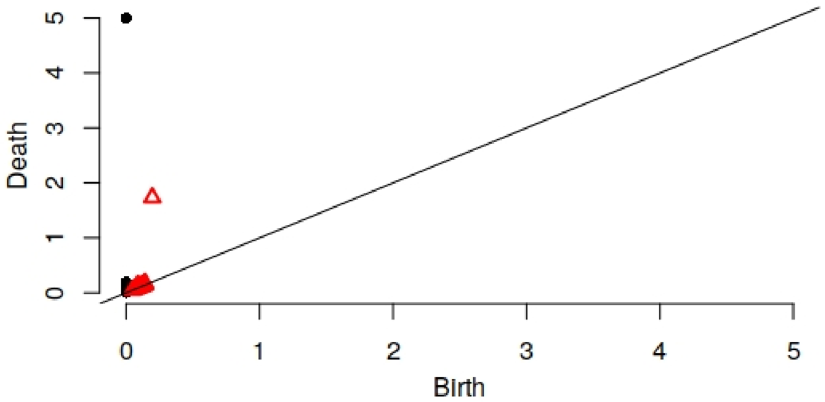


Death
Birth
0
1
2
3
4
5

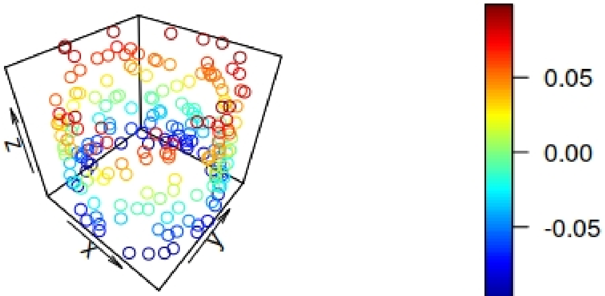

x2
x1
0
2
4
6
0
2
4
6



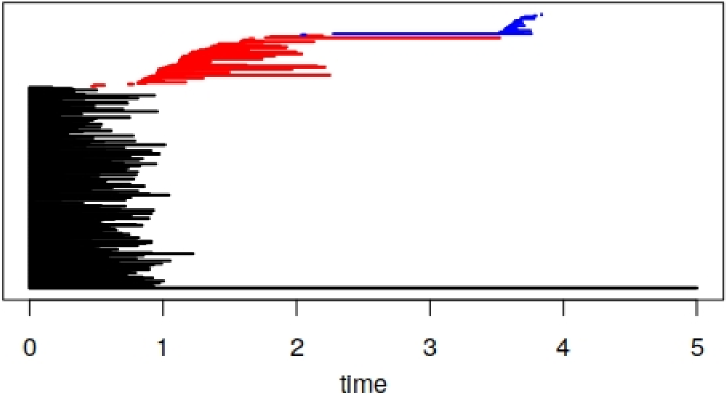
0
1
2
3
4
5
time

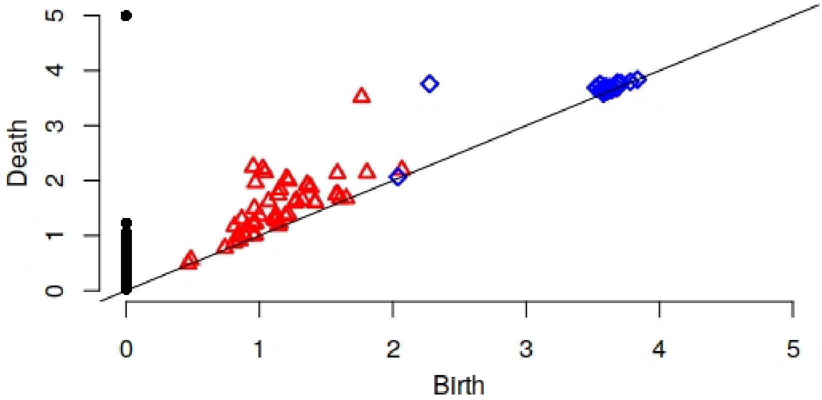


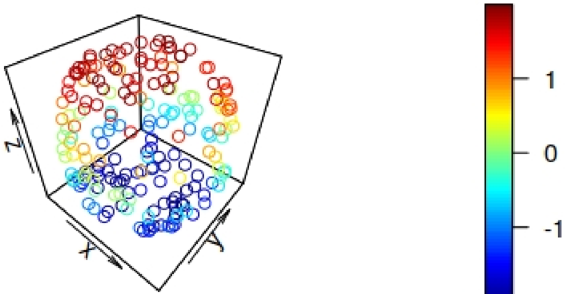
Death
Birth
0
1
2
3
4
5



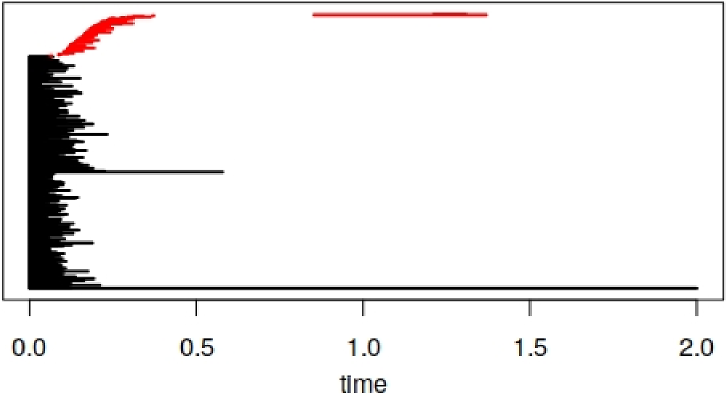
Z
X
Y
0.05
0.00
-0.05

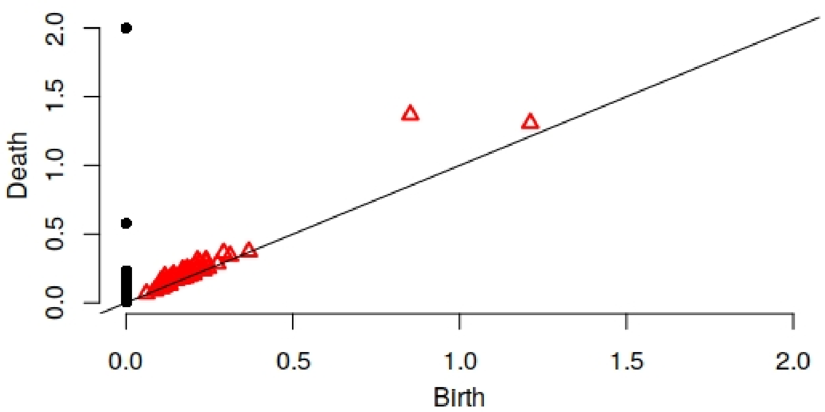

0
1
2
3
4
5
time


Death
Birth
0
1
2
3
4
5

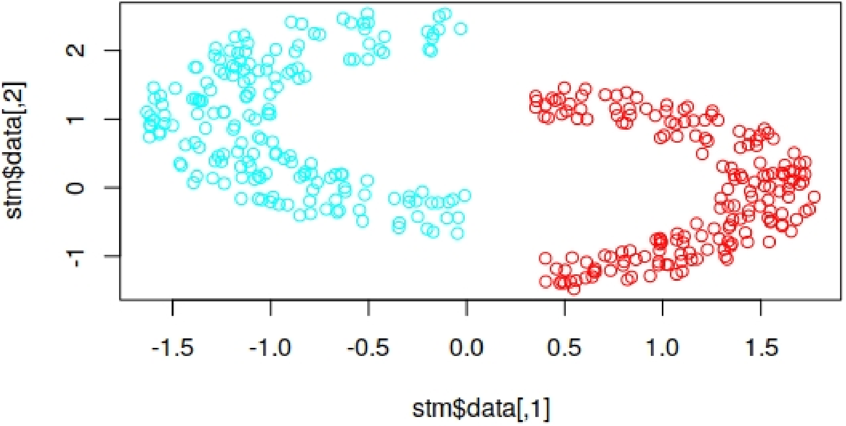


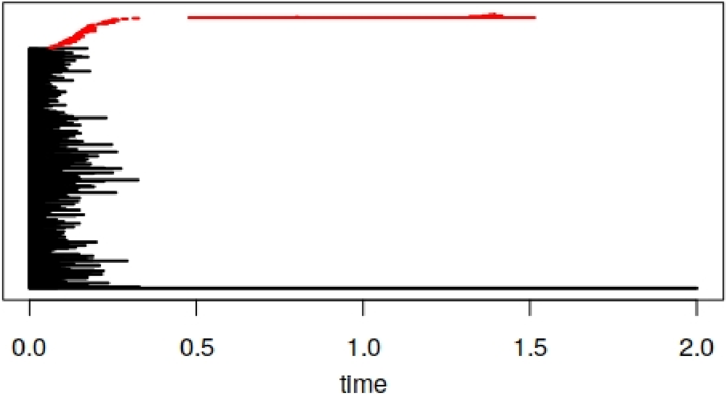
z
x
y
1
0
-1

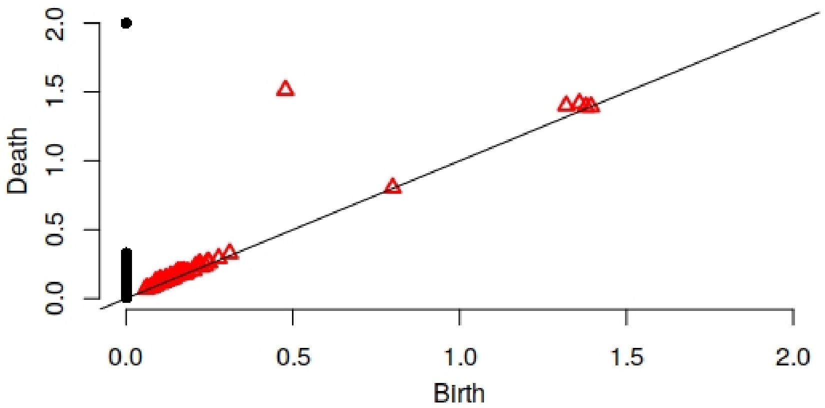

0.0
0.5
1.0
1.5
2.0
time

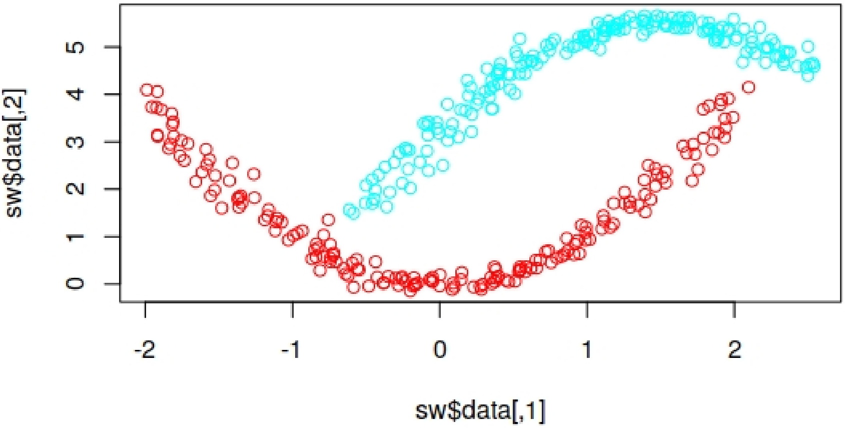


0.0
0.5
1.0
1.5
2.0
Death
0.0
0.5
1.0
1.5
2.0
Birth

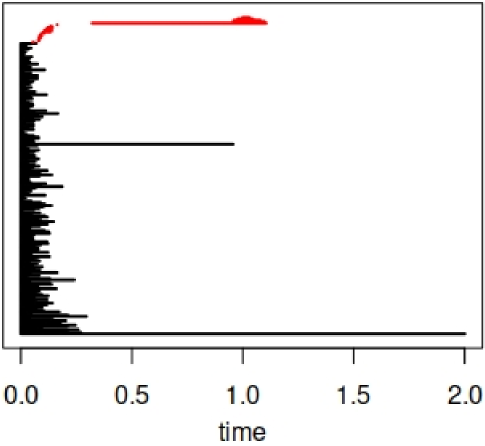

stm$data[,2]
2
1
0
-1
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
stm$data[,1]


0.0
0.5
1.0
1.5
2.0
time

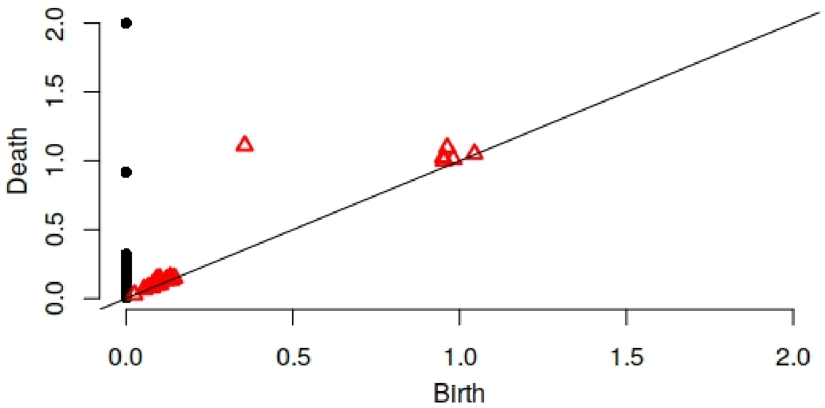

Death
0.0
0.5
1.0
1.5
2.0
Birth

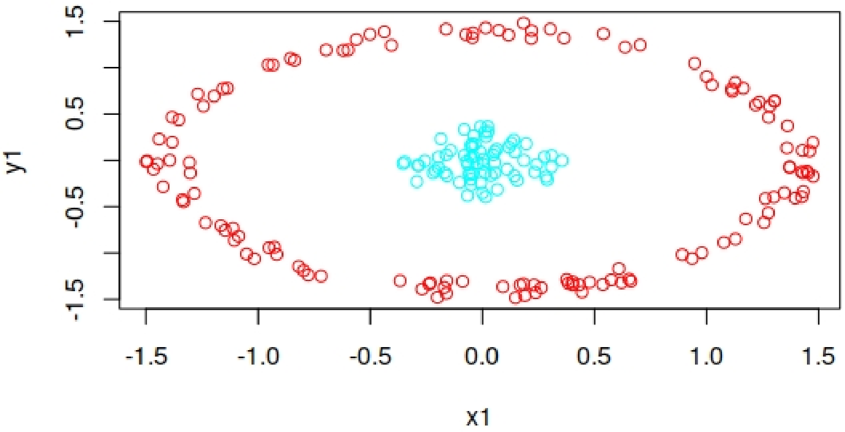


sw$data[,2]
0
1
2
3
4
5
-2
-1
0
1
2
sw$data[,1]



0.0
0.5
1.0
1.5
2.0
time


Death
Birth
0.0
0.5
1.0
1.5
2.0



1.5
0.5
-0.5
-1.5
y1
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
x1