STRUCTURES CLUSTER OF EXCELLENCE

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Topological Data Analysis in Python

organized by:
Michael Bleher, Maximilian Schmahl and Daniel Spitz

Heidelberg University

26th - 28th of October 2020

Contents

Contents

Homology

Algebraic Topology: Distinguishing topological spaces via algebraic invariants

Homology

Algebraic Topology: Distinguishing topological spaces via algebraic invariants

Homology: For each dimension $d \geq 0$ assigns to a topological space a vector space

$$X \mapsto H_d(X)$$

and to a continuous map a linear map

$$(f\colon X \to Y) \mapsto (H_d(f)\colon H_d(X) \to H_d(Y))$$

Homology

Algebraic Topology: Distinguishing topological spaces via algebraic invariants

Homology: For each dimension $d \geq 0$ assigns to a topological space a vector space

$$X \mapsto H_d(X)$$

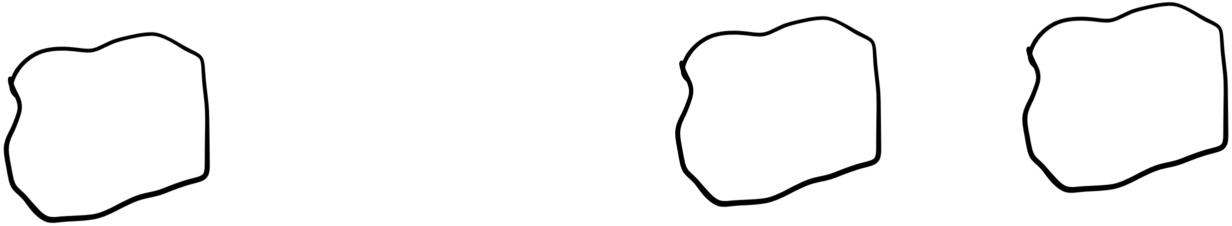
and to a continuous map a linear map

$$(f \colon X \to Y) \mapsto (H_d(f) \colon H_d(X) \to H_d(Y))$$

Definition

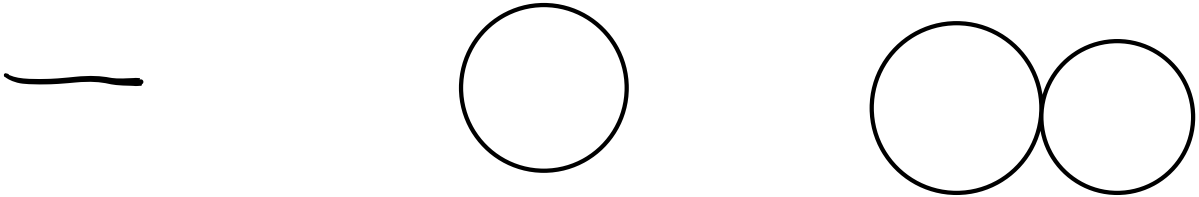
$\beta_d(X) = \dim H_d(X)$ is called the *d-th Betti number* of X.

What homology can distinguish I



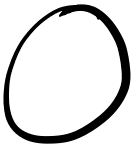
$$\beta_d(X \sqcup Y) = \beta_d(X) + \beta_d(Y).$$

What homology can distinguish II

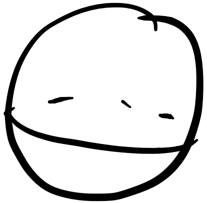


$\beta_1 = 0$ $\beta_1 = 1$ $\beta_2 = 2$

What homology can distinguish III

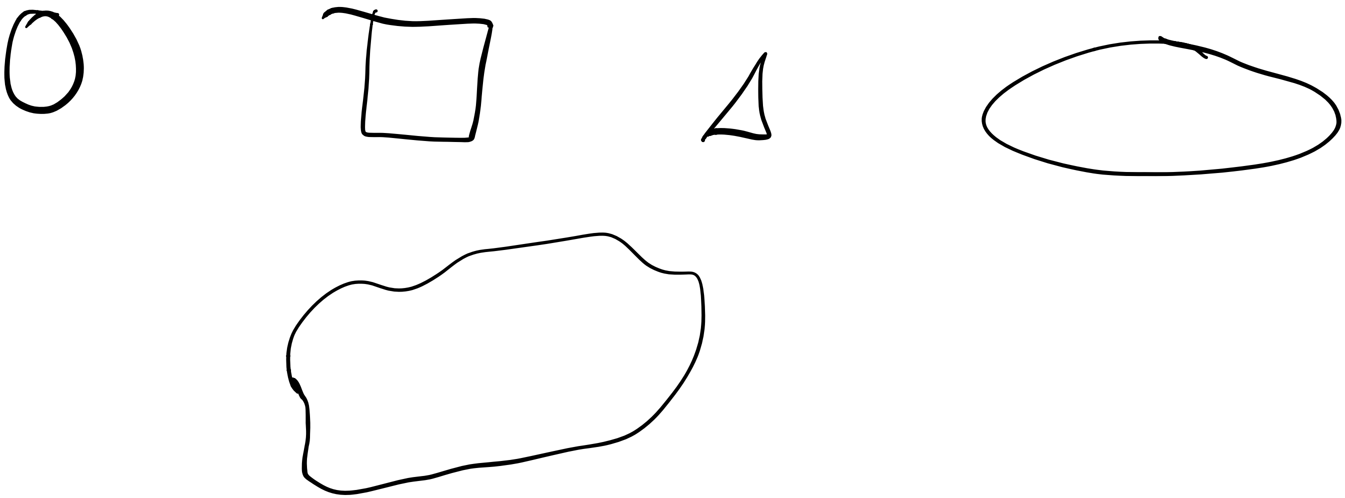


$\beta_1 = 1 \quad \beta_2 = 0$



$\beta_1 = 0 \quad \beta_2 = 1$

What homology cannot distinguish I



What homology cannot distinguish II

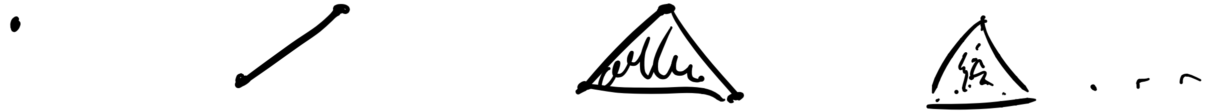
Homology in a nutshell

$\beta_d(X)$ is the number of d-dimensional holes in X.

Geometric Simplicial Complexes I

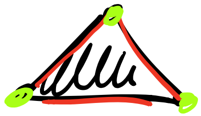
Definition

A *geometric n-simplex* is the convex hull of $n+1$ affinely independent points in $\mathbb{R}^m$.



Definition

A *geometric n-simplex* is the convex hull of $n+1$ affinely independent points in $\mathbb{R}^m$.
If $\operatorname{conv} X$ is a simplex and $Y \subseteq X$, then $\operatorname{conv} Y$ is called a *face* of $\operatorname{conv} X$.

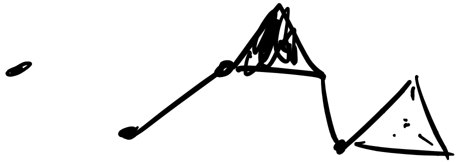


Geometric Simplicial Complexes I

Definition

A *geometric n-simplex* is the convex hull of $n+1$ affinely independent points in $\mathbb{R}^m$.
If $\operatorname{conv} X$ is a simplex and $Y \subseteq X$, then $\operatorname{conv} Y$ is called a *face* of $\operatorname{conv} X$.
A *geometric simplicial complex* is a finite set K of simplices such that

- if $\sigma \in K$ and τ is a face of K, then $\tau \in K$, and
- if $\sigma, \tau \in K$ and $\sigma \cap \tau \neq \emptyset$, then $\sigma \cap \tau$ is a face of σ and of τ.



Boundary matrix

Definition

Let $K = \{\sigma_1, \ldots, \sigma_m\}$ be a simplicial complex. We define its *boundary matrix* $D = (d_{i,j})$ as

$$d_{i,j} = \begin{cases} 1 & \text{if } \sigma_i \text{ is a } \overset{\text{face}}{\cancel{\text{boundary}}} \text{ of } \sigma_j \text{ with } \dim \sigma_i = \dim \sigma_j - 1, \\ 0 & \text{otherwise.} \end{cases}$$

Boundary matrix

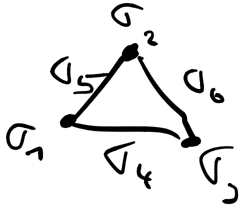
$\mathbb{F}_2 \qquad 1+1=0$

Definition

Let $K = \{\sigma_1, \dots, \sigma_m\}$ be a simplicial complex. We define its *boundary matrix* $D = (d_{i,j})$ as

$$d_{i,j} = \begin{cases} 1 & \text{if } \sigma_i \text{ is a boundary of } \sigma_j \text{ with } \dim \sigma_i = \dim \sigma_j - 1, \\ 0 & \text{otherwise.} \end{cases}$$

Example



$$\begin{array}{c} \\ \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \vdots \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccc} \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 & \sigma_5 & \sigma_6 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Definition
Let $v = (v_1, \ldots, v_m)^T$ be a column vector. We define

$$\operatorname{pivot} v = \max\{i \in \{1, \ldots, m\} \mid v_i \neq 0\}.$$

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Matrix reduction I

Definition

Let $v = (v_1, \ldots, v_m)^T$ be a column vector. We define

$$\operatorname{pivot} v = \max\{i \in \{1, \ldots, m\} \mid v_i \neq 0\}.$$

If M is a matrix, we write m_j for its j-th column and $\operatorname{cols} M$ for the set of all its non-zero columns.

Matrix reduction I

Definition
Let $v = (v_1, \dots, v_m)^T$ be a column vector. We define

$$\operatorname{pivot} v = \max\{i \in \{1, \dots, m\} \mid v_i \neq 0\}.$$

If M is a matrix, we write m_j for its j-th column and $\operatorname{cols} M$ for the set of all its non-zero columns.
We write $\operatorname{pivots} M = \{\operatorname{pivot} m_j \mid m_j \in \operatorname{cols} M\}$.

Matrix reduction I

Definition

Let $v = (v_1, \ldots, v_m)^T$ be a column vector. We define

$$\operatorname{pivot} v = \max\{i \in \{1, \ldots, m\} \mid v_i \neq 0\}.$$

If M is a matrix, we write m_j for its j-th column and $\operatorname{cols} M$ for the set of all its non-zero columns.

We write $\operatorname{pivots} M = \{\operatorname{pivot} m_j \mid m_j \in \operatorname{cols} M\}$.

We say that M is *reduced* if $\operatorname{pivot} m_j = \cancel{m_k}$ implies $m_j = m_k$ for all $m_j, m_k \in \operatorname{cols} M$.

Pivot m_k

Matrix reduction I

Definition

Let $v = (v_1, \ldots, v_m)^T$ be a column vector. We define

$$\operatorname{pivot} v = \max\{i \in \{1, \ldots, m\} \mid v_i \neq 0\}.$$

If M is a matrix, we write m_j for its j-th column and $\operatorname{cols} M$ for the set of all its non-zero columns.

We write $\operatorname{pivots} M = \{\operatorname{pivot} m_j \mid m_j \in \operatorname{cols} M\}$.

We say that M is *reduced* if $\operatorname{pivot} m_j = m_k$ implies $m_j = m_k$ for all $m_j, m_k \in \operatorname{cols} M$.

Example

Reduced: $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Not reduced: $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$

Matrix reduction II

Definition

Let M be a matrix. We say that (R, V) is a *reduction* of M if R is reduced, V is upper-triangular and invertible and we have $R = MV$.

Matrix reduction II

Definition

Let M be a matrix. We say that (R, V) is a *reduction* of M if R is reduced, V is upper-triangular and invertible and we have $R = MV$.

Input: M
Output: (R, V) reduction of M

$R \leftarrow M$
$V \leftarrow I$
while $\exists i < j$ with pivot $R_i =$ pivot R_j **do**
  $R_j \leftarrow R_j + R_i$
  $V_j \leftarrow V_j + V_i$
end while
return (R, V)

Matrix reduction III

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Computing homology I

Theorem

Let $K = \{\sigma_1, \ldots, \sigma_m\}$ be a simplicial complex with boundary matrix D. Let (R, V) be a reduction of D. Then

$$\beta_d(K) = \#\{i \mid R_i = 0, i \notin \text{pivots}\, R, \dim \sigma_i = d\}.$$

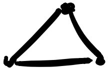
Theorem

Let $K = \{\sigma_1, \dots, \sigma_m\}$ be a simplicial complex with boundary matrix D. Let (R, V) be a reduction of D. Then

$$\beta_d(K) = \#\{i \mid R_i = 0, i \notin \operatorname{pivots} R, \dim \sigma_i = d\}.$$

Example

$$D = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & \vdots & \vdots & 0 & 1 & 1 \\ \vdots & \vdots & \vdots & 1 & 0 & 1 \\ & & & \vdots & \vdots & \vdots \end{pmatrix} \rightsquigarrow R = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ \vdots & \vdots & \vdots & 0 & 1 & 0 \\ & & & 1 & 0 & \vdots \\ & & & 0 & 0 & \\ & & & 0 & 0 & \\ & & & \vdots & \vdots & \end{pmatrix} \begin{matrix} \\ \leftarrow 2 \\ \leftarrow 3 \\ \\ \\ \\ \end{matrix}$$



i with $R_i = 0$

$\dim = 0 \rightarrow$ 1, 2, 3, 6 $\hookrightarrow \dim = 1$

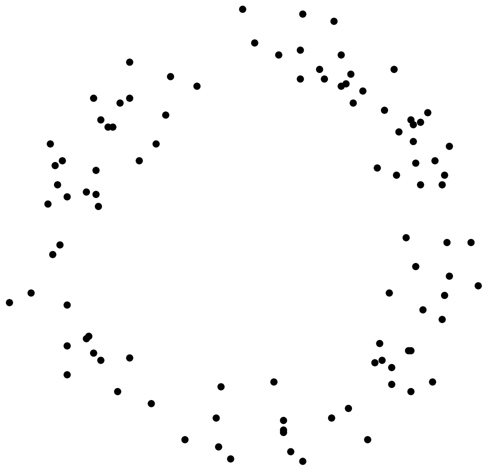
Pivots R: 2, 3

Computing homology II

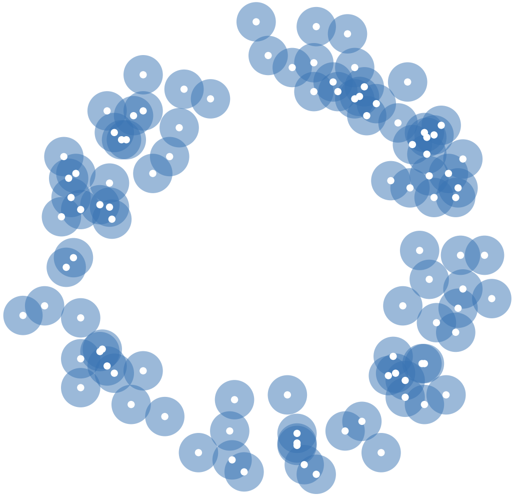
$$\begin{array}{c}\text{Topological space}\\ \Big\downarrow{\small\text{Triangulation}}\\ \text{Simplicial complex}\\ \Big\downarrow{\small\text{Matrix reduction}}\\ \text{Betti numbers}\end{array}$$

Contents

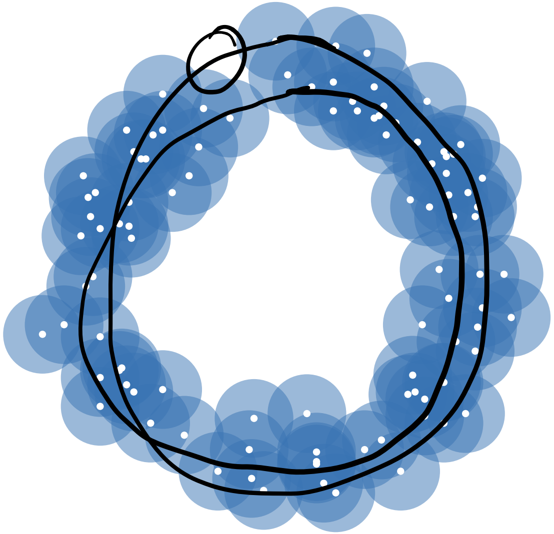
Manifold recovery I



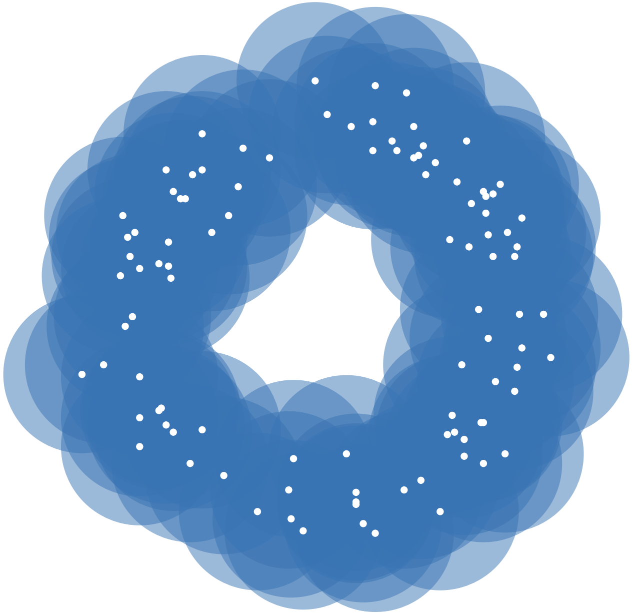
Manifold recovery I



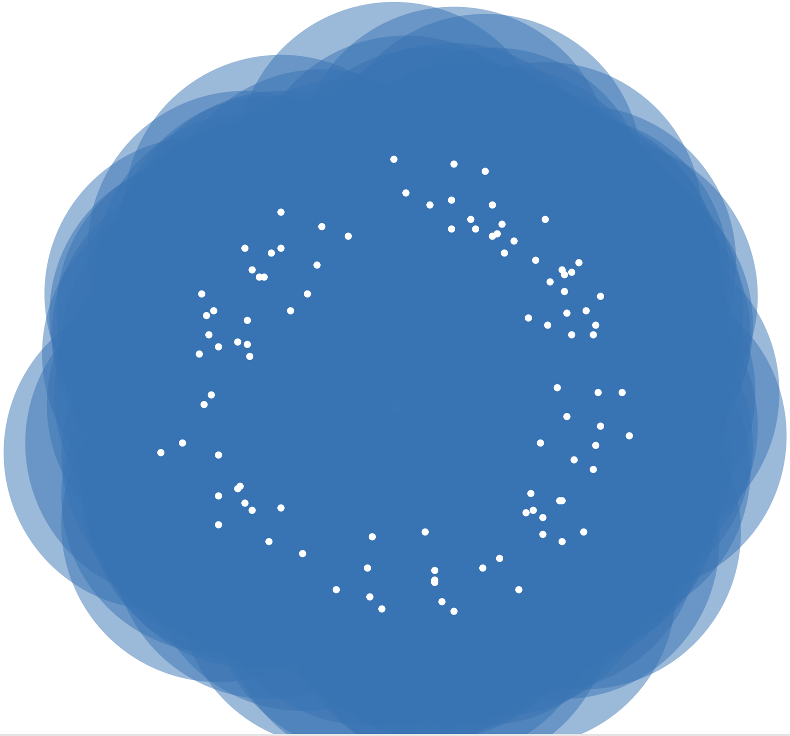
Manifold recovery I



Manifold recovery I



Manifold recovery I



Manifold recovery II

Theorem (Niyogi, Smale, Weinberger)

Let $M \subseteq \mathbb{R}^k$ be a manifold and $P \subseteq M$. There exists $c(M) > 0$ such that for every $c(M) > \delta > 0$ with $M \subseteq \bigcup_{x \in P} B_\delta(x)$ we have

$$H_*(M) \cong H_*\left(\bigcup_{x \in P} B_\delta(x)\right)$$

Manifold recovery II

Theorem (Niyogi, Smale, Weinberger)

Let $M \subseteq \mathbb{R}^k$ *be a manifold and* $P \subseteq M$*. There exists* $c(M) > 0$ *such that for every* $c(M) > \delta > 0$ *with* $M \subseteq \bigcup_{x \in P} B_\delta(x)$ *we have*

$$H_*(M) \cong H_*\left(\bigcup_{x \in P} B_\delta(x)\right)$$

How do we choose δ?

Manifold recovery II

Theorem (Niyogi, Smale, Weinberger)

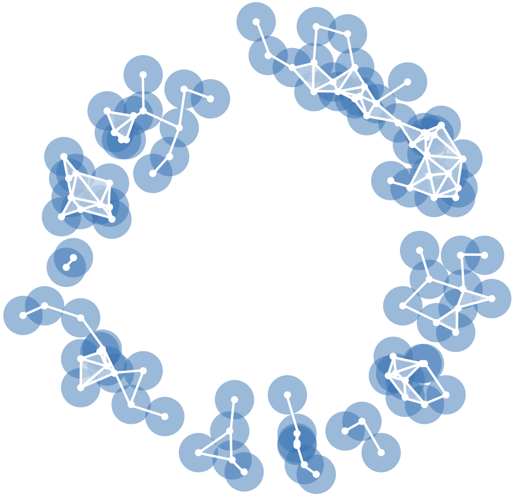
Let $M \subseteq \mathbb{R}^k$ be a manifold and $P \subseteq M$. There exists $c(M) > 0$ such that for every $c(M) > \delta > 0$ with $M \subseteq \bigcup_{x \in P} B_\delta(x)$ we have

$$H_*(M) \cong H_*\left(\bigcup_{x \in P} B_\delta(x)\right)$$

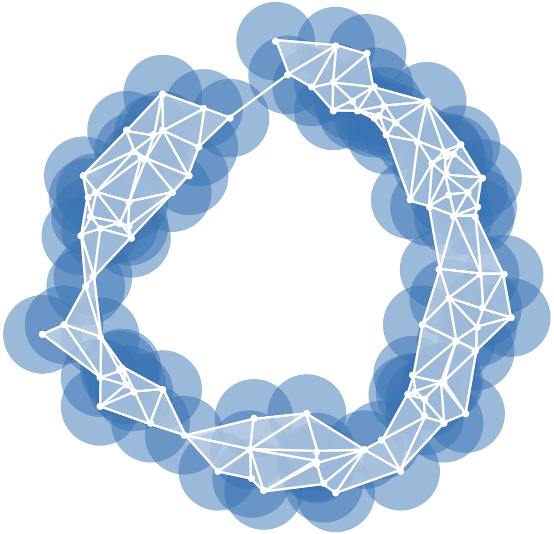
How do we choose δ?

We don't!

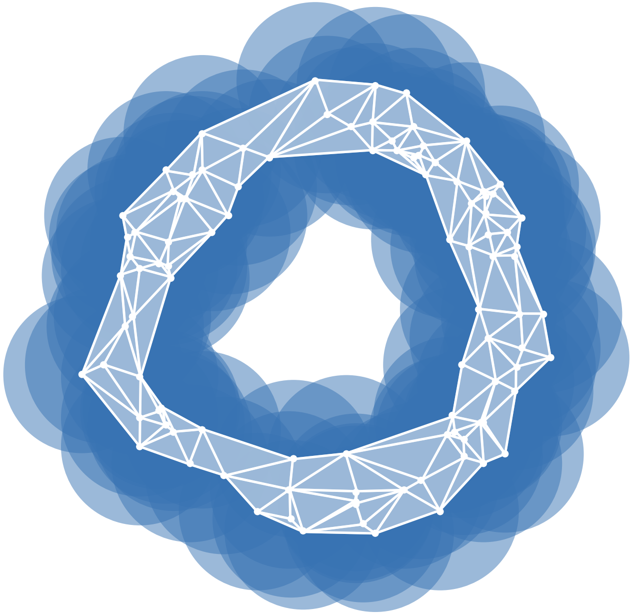
Complexes from balls I



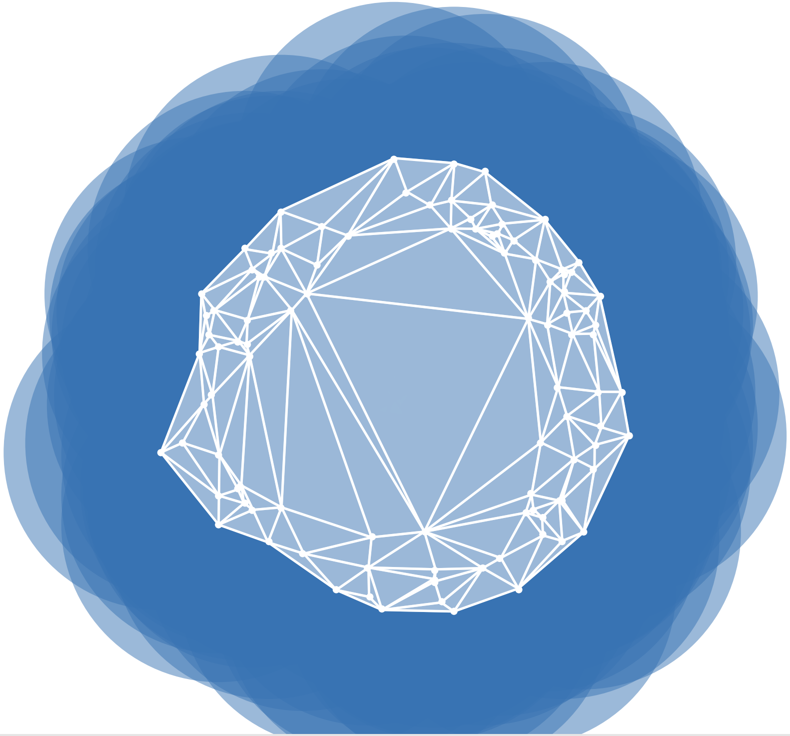
Complexes from balls I



Complexes from balls I



Complexes from balls I

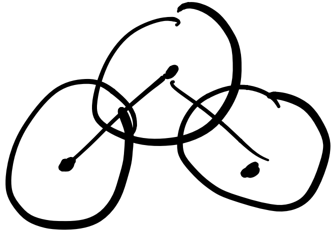


Complexes from Balls II

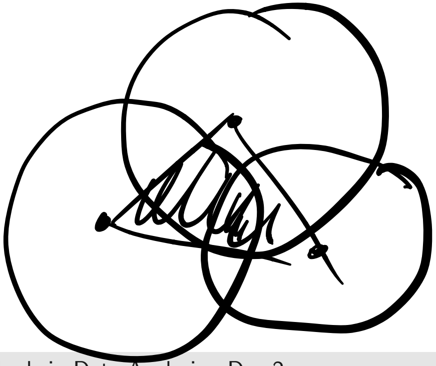
Definition

Let $P \subseteq \mathbb{R}^k$ be a finite set. For $t > 0$ define the *Čech complex*

$$\text{Čech}_t(P) = \{Q \subseteq P \mid \bigcap_{x \in Q} B_t(x) \neq \emptyset\}.$$



Note: $\text{Čech}_t(P) \subseteq \text{Čech}_u(P)$ whenever $t \leq u$.



Complexes from Balls II

Definition

Let $P \subseteq \mathbb{R}^k$ be a finite set. For $t > 0$ define the *Čech complex*

$$\text{Čech}_t(P) = \{Q \subseteq P \mid \bigcap_{x \in Q} B_t(x) \neq \emptyset\}.$$

Note: $\text{Čech}_t(P) \subseteq \text{Čech}_u(P)$ whenever $t \leq u$.

Theorem (Nerve Theorem)

Let $P \subseteq \mathbb{R}^k$ be a finite set and $t \geq 0$. If $\bigcap_{x \in Q} B_t(x)$ can be deformed to a point for all $Q \subseteq P$, then

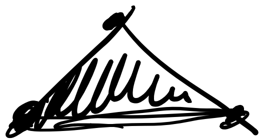
$$H_* \left(\bigcup_{x \in P} B_t(x) \right) \cong H_*(\text{Čech}_t(P))$$

Rips complexes

Definition

Let $P \subseteq \mathbb{R}^k$ be a finite set. For $t \geq 0$ we define the *Rips complex*

$$\mathrm{Rips}_t(P) = \{Q \subseteq P \mid \sup_{x,y \in Q} d(x,y) \leq t\}.$$



Note: $\mathrm{Rips}_t(P) \subseteq \mathrm{Rips}_u(P)$ whenever $t \leq u$.

Rips complexes

Definition

Let $P \subseteq \mathbb{R}^k$ be a finite set. For $t \geq 0$ we define the *Rips complex*

$$\mathrm{Rips}_t(P) = \{Q \subseteq P \mid \sup_{x,y \in Q} d(x,y) \leq t\}.$$

Note: $\mathrm{Rips}_t(P) \subseteq \mathrm{Rips}_u(P)$ whenever $t \leq u$.

Theorem

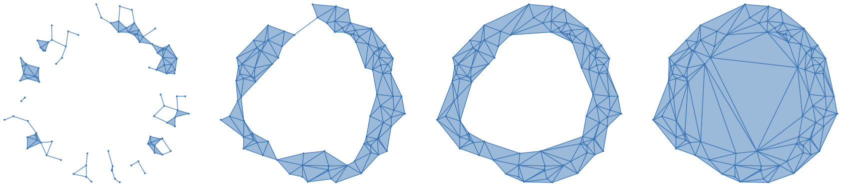
There exists $\theta > 0$ such that for all finite sets $P \subseteq \mathbb{R}^k$ and $t > 0$ we have

$$\mathrm{Rips}_t \subseteq \check{\mathrm{C}}\mathrm{ech}_{\theta t}(P)$$

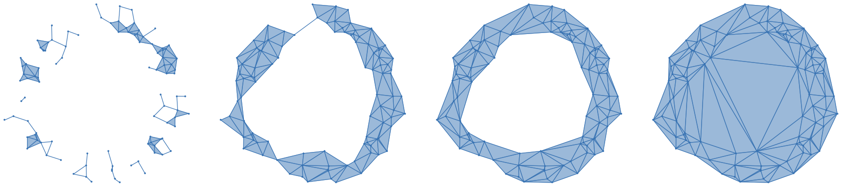
and

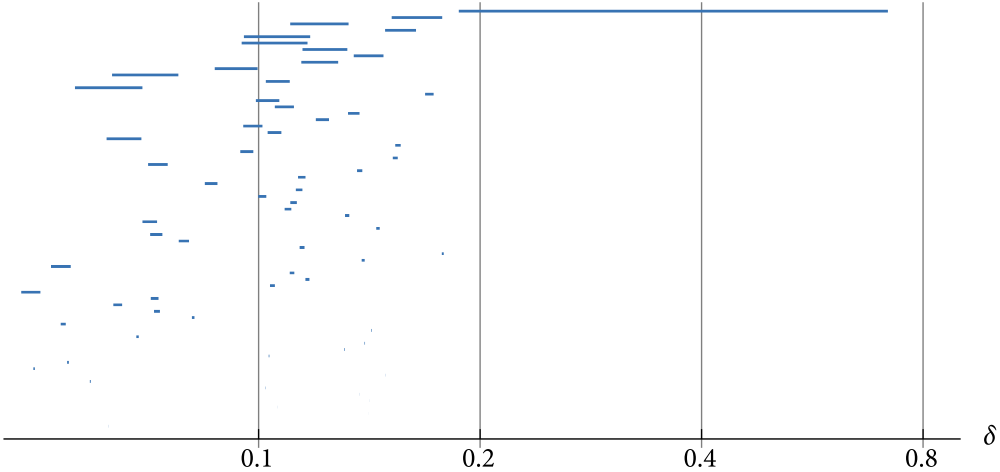
$$\check{\mathrm{C}}\mathrm{ech}_t \subseteq \mathrm{Rips}_{\theta t}(P)$$

Homology for Data

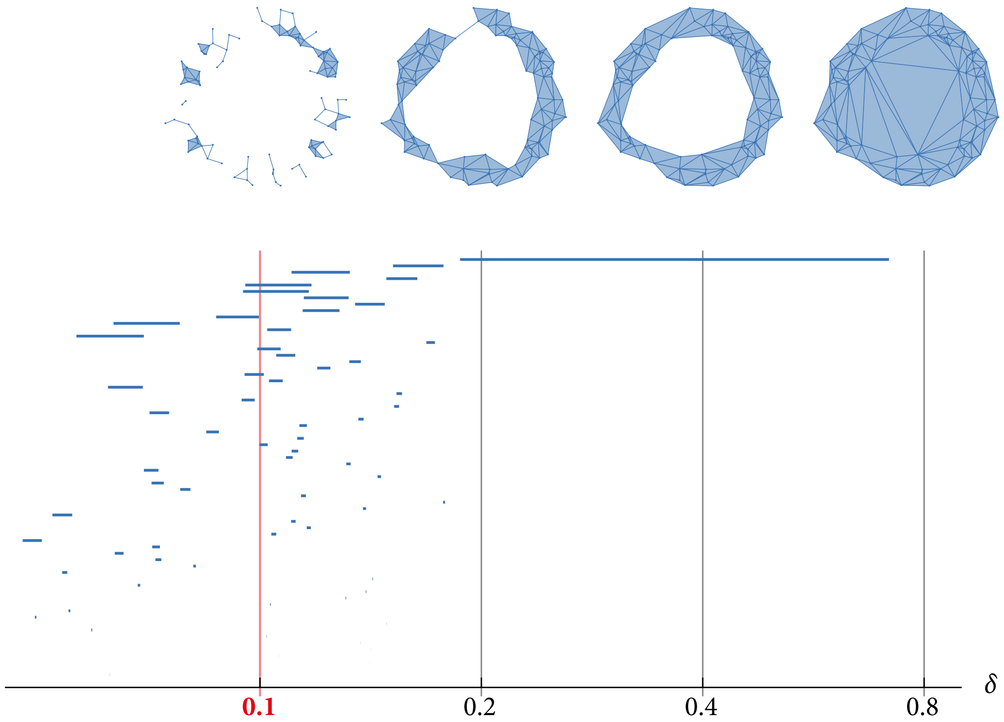


Homology for Data

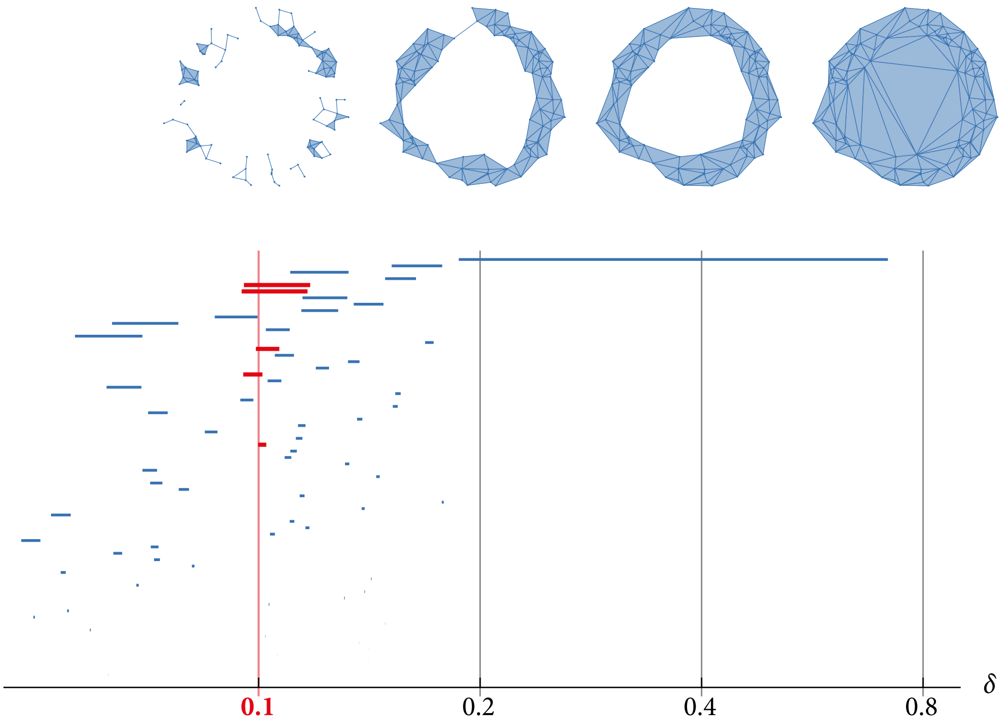




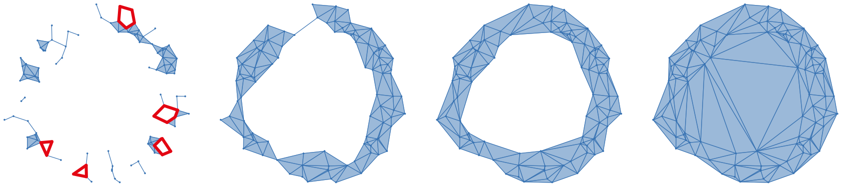
Homology for Data

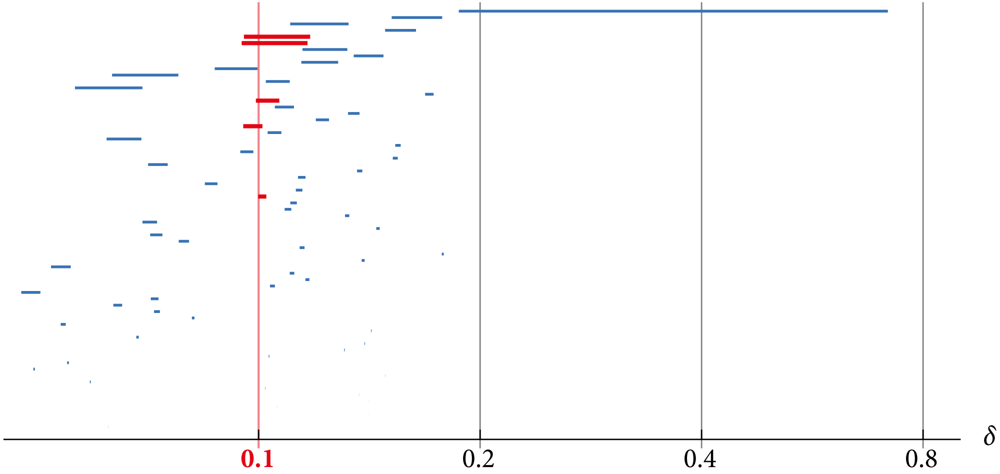


Homology for Data

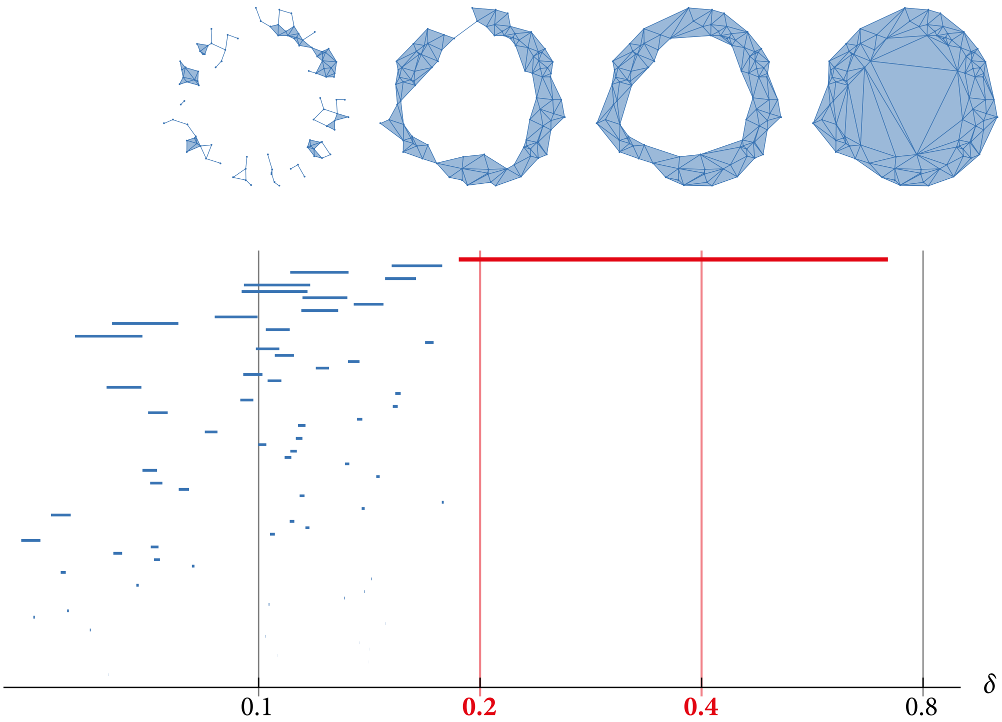


Homology for Data

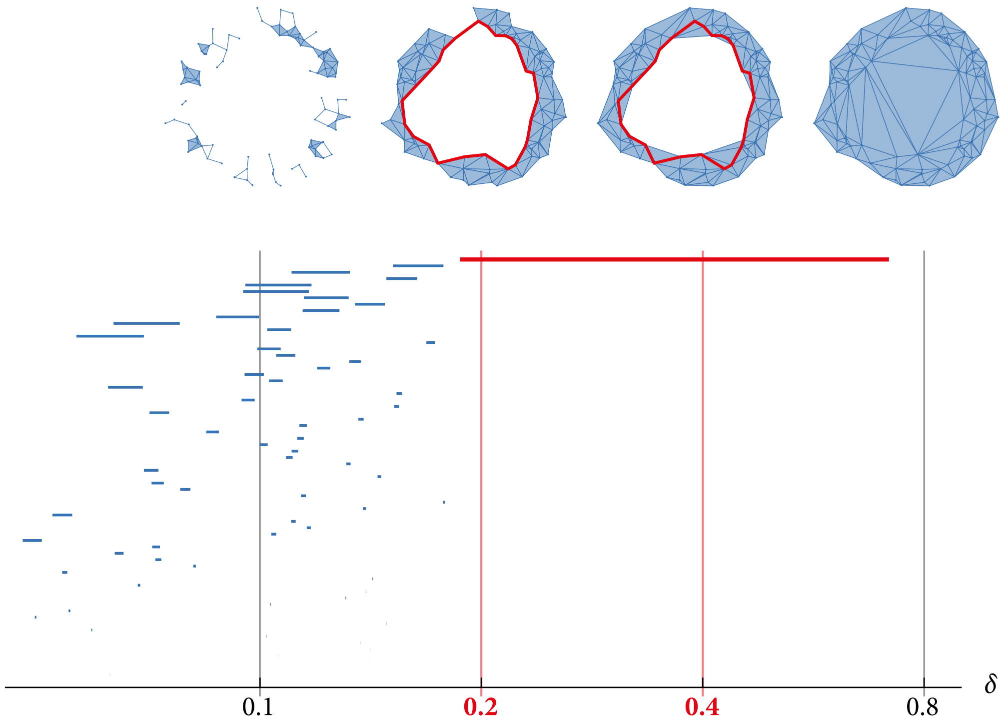




Homology for Data



Homology for Data



Barcodes

Definition

A family of finite-dimensional vector spaces V_i, $i = 1, \ldots, n$ with linear maps $V_i \to V_{i+1}$ for each i is called a *persistence module*.

$K_1 \subseteq K_2 \subseteq K_3 \subseteq \dots \subseteq K_n$

Definition $H(K_1) \to H(K_2) \to \cdots \to H(K_n)$

A family of finite-dimensional vector spaces V_i, $i = 1, \dots, n$ with linear maps $V_i \to V_{i+1}$ for each i is called a *persistence module.*

Theorem Gabriel $(V_t)_{t \in \mathbb{R}}$ $V_t \to V_s$ $s \leq t$

Let

$$V_1 \longrightarrow V_2 \longrightarrow \dots \longrightarrow V_{n-1} \longrightarrow V_n$$

be a persistence module. Then there exists a unique family of intervals $(I_k)_{k \in K}$ *with* $I_k \subseteq \{1, \dots, n\}$ *such that*

$$\dim V_i = \#\{k \in K \mid i \in I_k\}$$

and

$$\operatorname{rank}(V_i \to V_j) = \#\{k \in K \mid i, j \in I_k\}$$

Persistent homology

Definition

Let K be a simplicial complex. A *filtration* of K is a family of simplicial complexes $K_1, \ldots, K_n$ such that $K_n = K$ and $K_i \subseteq K_{i+1}$ for all n.

Persistent homology

Definition

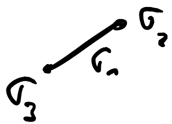
Let K be a simplicial complex. A *filtration* of K is a family of simplicial complexes $K_1, \ldots, K_n$ such that $K_n = K$ and $K_i \subseteq K_{i+1}$ for all n.

We call

$$H_*(K_1) \longrightarrow H_*(K_2) \longrightarrow \ldots \longrightarrow H_*(K_{n-1}) \longrightarrow H_*(K_n)$$

the *persistent homology* of the filtration.

Computing persistent homology



Theorem (Barannikov, Carlsson & Zomorodian, Edelsbrunner et al.)

Let $K = \{\sigma_1, \ldots, \sigma_n\}$ be a simplicial complex such that $K_i = \{\sigma_1, \ldots, \sigma_i\}$ is a subcomplex.

Computing persistent homology

Theorem (Barannikov, Carlsson & Zomorodian, Edelsbrunner et al.)

Let $K = \{\sigma_1, \ldots, \sigma_n\}$ *be a simplicial complex such that* $K_i = \{\sigma_1, \ldots, \sigma_i\}$ *is a subcomplex.*
Let D *be the corresponding boundary matrix and* (R, V) *a reduction of* D*.*

Computing persistent homology

Theorem (Barannikov, Carlsson & Zomorodian, Edelsbrunner et al.)

Let $K = \{\sigma_1, \ldots, \sigma_n\}$ be a simplicial complex such that $K_i = \{\sigma_1, \ldots, \sigma_i\}$ is a subcomplex.
Let D be the corresponding boundary matrix and (R, V) a reduction of D.
Then

$$\{[i, \infty) \mid R_i = 0, i \notin \operatorname{pivots} R\} \cup \{[i, j) \mid i = \operatorname{pivot} R_j\}$$

is the barcode of

$$H_*(K_1) \longrightarrow H_*(K_2) \longrightarrow \ldots \longrightarrow H_*(K_{n-1}) \longrightarrow H_*(K_n)$$

Persistent homology pipeline

$$
\begin{array}{c}
\text{Data} \\
\Big\downarrow{\scriptstyle \text{Rips/Čech/Delaunay/...}} \\
\text{Filtration of Complexes} \\
\Big\downarrow{\scriptstyle \text{Homology}} \\
\text{Persistence Module} \\
\Big\downarrow{\scriptstyle \text{Interval Decomposition}} \\
\text{Barcode}
\end{array}
$$

Stability I

Definition

Let $P, Q \subseteq \mathbb{R}^k$. We define their *Hausdorff distance* as

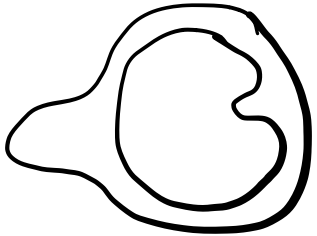
$$d_H(P, Q) = \max \left\{ \sup_{p \in P} \inf_{q \in Q} d(p, q), \sup_{q \in Q} \inf_{p \in P} d(p, q) \right\}$$

Definition

Let $P, Q \subseteq \mathbb{R}^k$. We define their *Hausdorff distance* as

$$d_H(P,Q) = \max\left\{\sup_{p\in P}\inf_{q\in Q} d(p,q), \sup_{q\in Q}\inf_{p\in P} d(p,q)\right\}$$

Example



Stability II

Definition

Let $B = (I_\alpha)_{\alpha \in A}$ and $C = (J_\beta)_{\beta \in B}$ be barcodes.
A δ-*matching* between them consists of subsets $A' \subseteq A$, $B' \subseteq B$ and a bijection $f : A' \to B'$ such that:

Stability II

Definition

Let $B = (I_\alpha)_{\alpha \in A}$ and $C = (J_\beta)_{\beta \in B}$ be barcodes.
A δ-*matching* between them consists of subsets $A' \subseteq A$, $B' \subseteq B$ and a bijection $f \colon A' \to B'$ such that:

- If $\alpha \notin A'$, $\beta \notin B'$, then $\mathrm{length}(I_\alpha), \mathrm{length}(J_\beta) < \delta$.
- If $f(I_\alpha) = J_\beta$, then the endpoints of I_α and J_β are within δ of eachother.

Stability II

Definition
Let $B = (I_\alpha)_{\alpha \in A}$ and $C = (J_\beta)_{\beta \in B}$ be barcodes.
A δ-*matching* between them consists of subsets $A' \subseteq A$, $B' \subseteq B$ and a bijection $f \colon A' \to B'$ such that:

- If $\alpha \notin A'$, $\beta \notin B'$, then $\text{length}(I_\alpha), \text{length}(J_\beta) < \delta$.
- If $f(I_\alpha) = J_\beta$, then the endpoints of I_α and J_β are within δ of eachother.

We define the *bottleneck distance* as

$$d_b(B, C) = \inf\{\delta > 0 \mid \text{there exists a } \delta\text{-matching between } B \text{ and } C\}.$$

Stability II

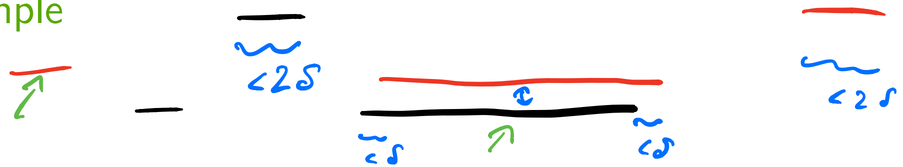
Definition
Let $B = (I_\alpha)_{\alpha \in A}$ and $C = (J_\beta)_{\beta \in B}$ be barcodes.
A δ-*matching* between them consists of subsets $A' \subseteq A$, $B' \subseteq B$ and a bijection $f : A' \to B'$ such that:

- If $\alpha \notin A'$, $\beta \notin B'$, then $\text{length}(I_\alpha), \text{length}(J_\beta) < 2\delta$.
- If $f(I_\alpha) = J_\beta$, then the endpoints of I_α and J_β are within δ of eachother.

We define the *bottleneck distance* as

$$d_b(B, C) = \inf\{\delta > 0 \mid \text{there exists a } \delta\text{-matching between } B \text{ and } C\}.$$

Example



Stability III

Theorem (Cohen-Steiner et al.)

Let $P, Q \subseteq \mathbb{R}^k$ be finite subsets and let $B(P)$ and $B(Q)$ be the barcodes of the persistent homology of their Rips filtrations. Then

$$d_b(B(P), B(Q)) \leq d_H(P, Q).$$

Persistence diagrams

Definition

If $B = (I_\alpha)_{\alpha \in A}$ is a barcode, we define its *persistence diagram* as

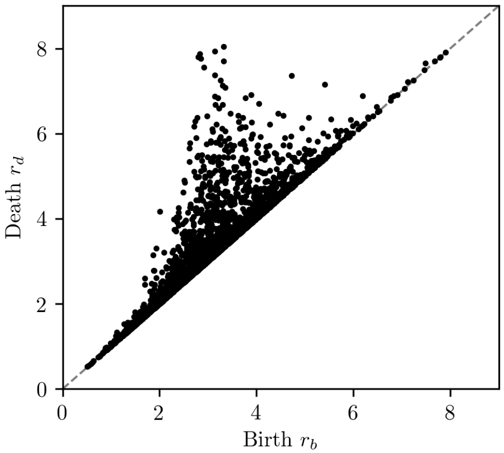
$$\operatorname{dgm}(B) = ((\inf I_\alpha, \sup I_\alpha))_{\alpha \in A} \subseteq \mathbb{R}^2.$$

Persistence diagrams

Definition

If $B = (I_\alpha)_{\alpha \in A}$ is a barcode, we define its *persistence diagram* as

$$\mathrm{dgm}(B) = ((\inf I_\alpha, \sup I_\alpha))_{\alpha \in A} \subseteq \mathbb{R}^2.$$



Contents

Ripser

C++ library by Ulrich Bauer to compute barcodes of Rips filtrations

Ripser

C++ library by Ulrich Bauer to compute barcodes of Rips filtrations

Also implemented in scikit-tda:

```
from ripser import ripser
from persim import plot_diagrams



# Compute the persistence diagram of a Rips filtration
# data is numpy array of points in euclidean space or a distance matrix
dgm = ripser(data, maxdim = 1, thresh = inf, distance_matrix = False)

# Plot the persistence diagram
plot_diagrams(dgm, show = False)
```

Contents

What can we do with persistent homology?

What can we do with persistent homology?

- Infer something about the shape of a data set (e.g. Cosmic Microwave Background, Edelsbrunner et al.)

What can we do with persistent homology?

- Infer something about the shape of a data set (e.g. Cosmic Microwave Background, Edelsbrunner et al.)
- Infer something about the complexity of a data set (e.g. lung disease detection, Brodzki et al.)

What can we do with persistent homology?

- Infer something about the shape of a data set (e.g. Cosmic Microwave Background, Edelsbrunner et al.)
- Infer something about the complexity of a data set (e.g. lung disease detection, Brodzki et al.)
- Use it as an additional layer in machine learning methods