

Topological Data Analysis

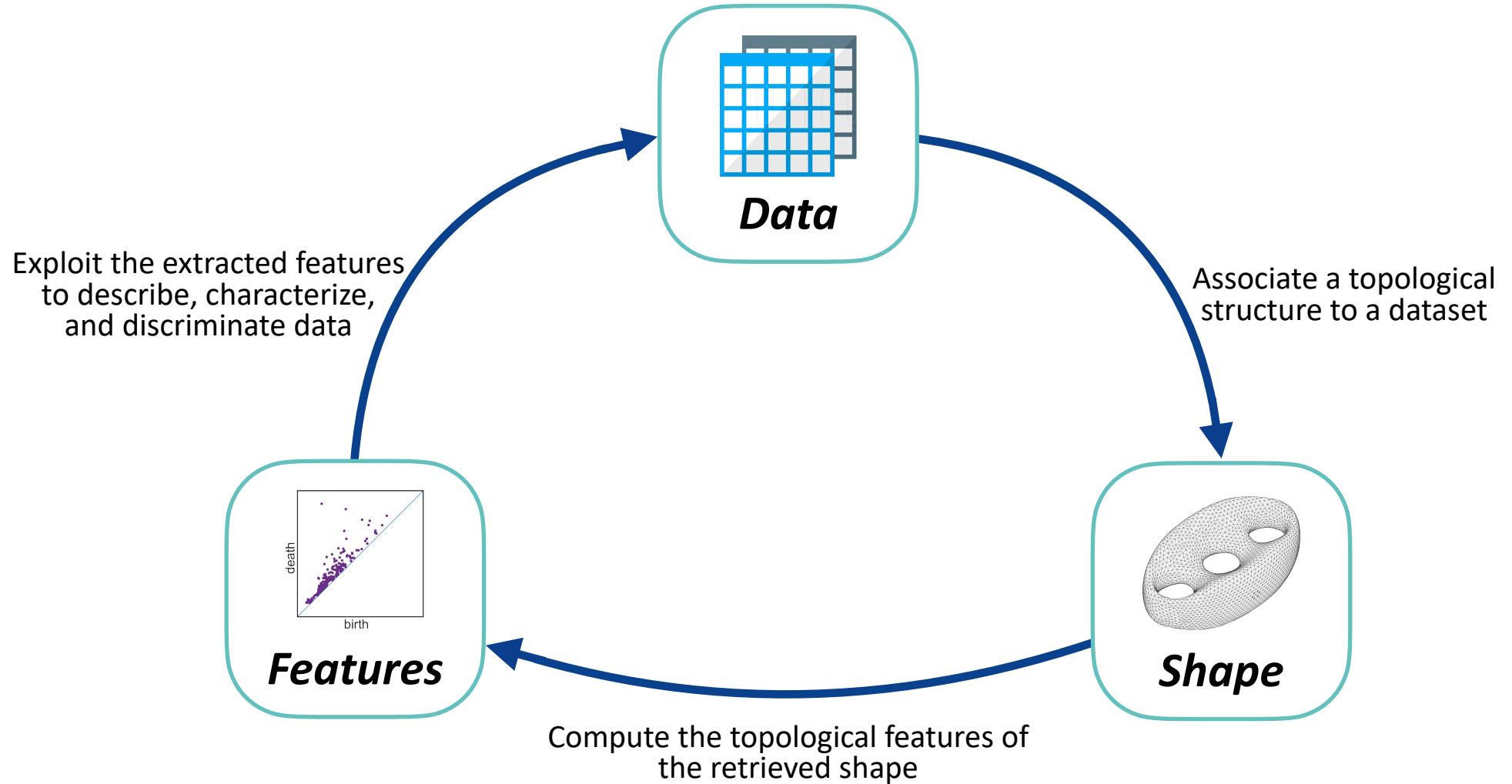
Persistence & Stability

Ulderico Fugacci

CNR - IMATI

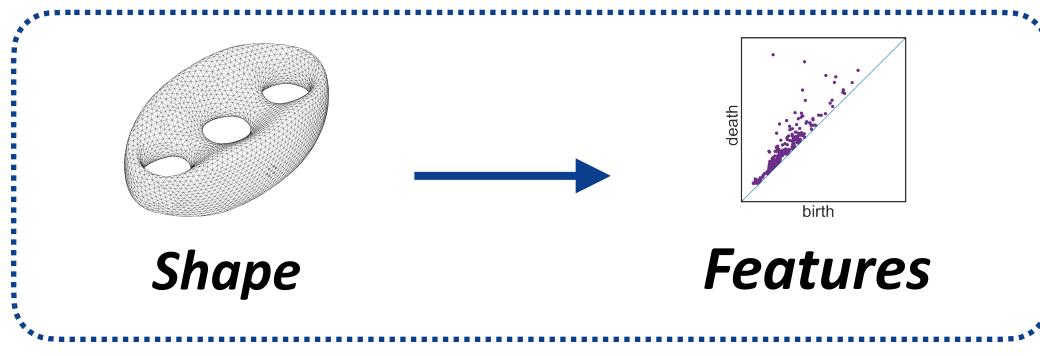


Topological Data Analysis



Persistence & Stability

(Persistent) Homology allows for assigning to any (filtered) simplicial complex
topological information expressed in terms of algebraic structures



Goal:

Today, we address two main questions:

- ◆ *Can this topological information be characterized in a simpler and “more visualizable” way?*
- ◆ *Is this information stable under small perturbations of the input data?*

Persistence & Stability

- ◆ *Persistence Pairs and their Visualizations*
- ◆ *Stability Results for Persistent Homology*

Persistence & Stability

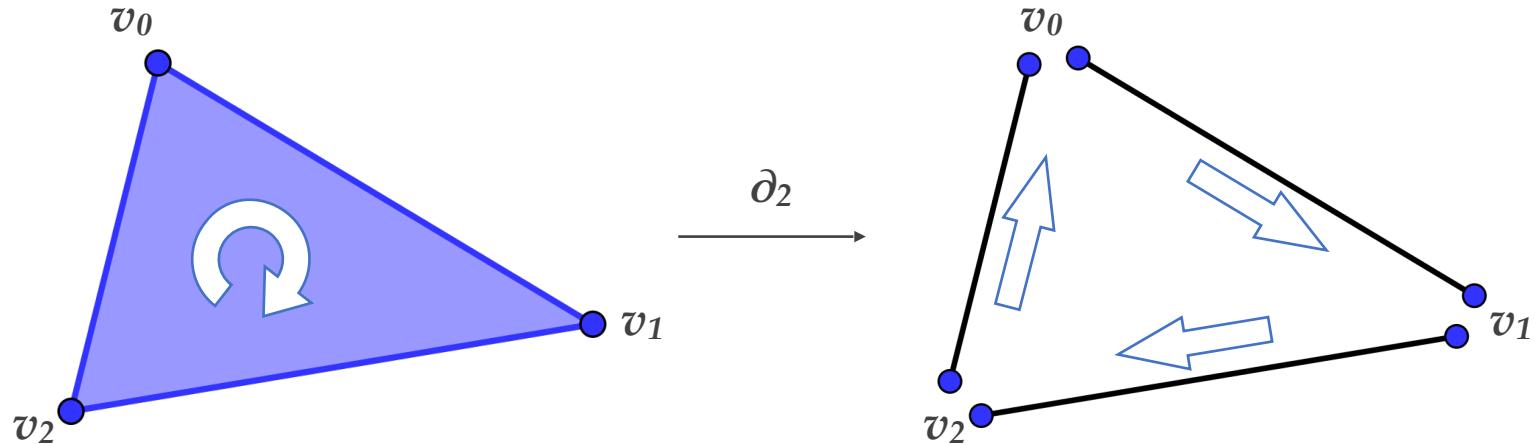
- ◆ *Persistence Pairs and their Visualizations*
- ◆ *Stability Results for Persistent Homology*

Homology

Given a (finite) simplicial complex K ,

- ◆ a ***k-chain*** is a formal sum (with \mathbb{Z} coefficients) of (oriented) k -simplices of K
- ◆ $C_k(K)$ is the ***group of the k -chains of K***
- ◆ the ***boundary map*** $\partial_k : C_k(K) \longrightarrow C_{k-1}(K)$ is defined as

$$\partial_k([v_0, \dots, v_k]) := \sum_{i=0}^k (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_k]$$

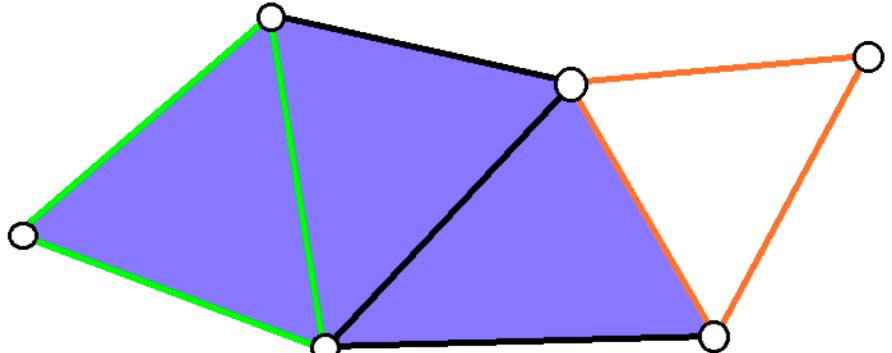


Homology

Given a (finite) simplicial complex K ,

- ◆ a ***k-chain*** is a formal sum (with \mathbb{Z} coefficients) of (oriented) k -simplices of K
- ◆ $C_k(K)$ is the ***group of the k-chains of K***
- ◆ the ***boundary map*** $\partial_k : C_k(K) \longrightarrow C_{k-1}(K)$ is defined as

$$\partial_k([v_0, \dots, v_k]) := \sum_{i=0}^k (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_k]$$



A k -simplex σ is called:

- ◆ ***k-cycle*** if $\sigma \in \text{Ker}(\partial_k)$
- ◆ ***k-boundary*** if $\sigma \in \text{Im}(\partial_{k+1})$

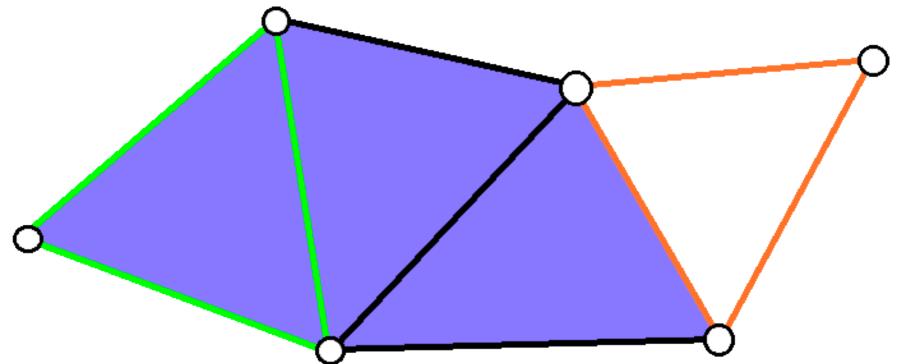
Homology

Given a (finite) simplicial complex K , the **k -homology group** $H_k(K)$ of K is defined as

$$H_k(K) := Z_k(K)/B_k(K)$$

where:

- ◆ $Z_k(K)$ is the **group of k -cycles** of K
- ◆ $B_k(K)$ is the **group of k -boundaries** of K



Homology

The ***theorem of structure for finitely generated Abelian groups*** ensures us that

Theorem:

Each homology group can be expressed as

$$H_k(K) \cong \mathbb{Z}^{\beta_k} \langle c_1, \dots, c_{\beta_k} \rangle \oplus \mathbb{Z}_{\lambda_1} \langle c'_1 \rangle \oplus \dots \oplus \mathbb{Z}_{\lambda_{p_k}} \langle c'_{p_k} \rangle$$

with $\lambda_{i+1} \mid \lambda_i$

We call:

- ◆ β_k , the ***kth Betti number*** of K
- ◆ $\lambda_1, \dots, \lambda_{p_k}$, the ***torsion coefficients*** of K
- ◆ $c_1, \dots, c_{\beta_k}, c'_1, \dots, c'_{p_k}$, the ***homology generators*** of K

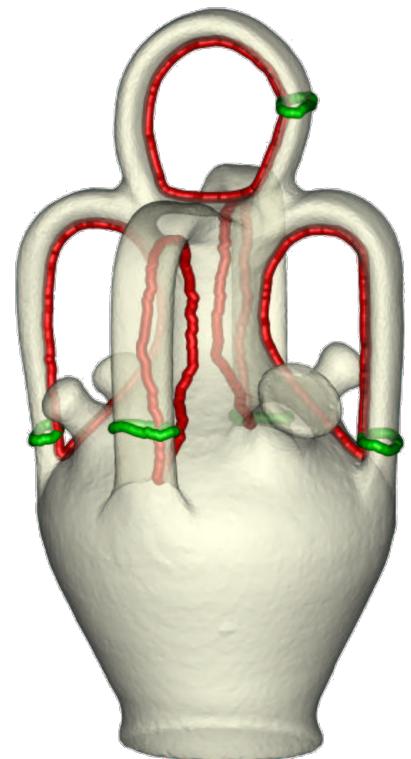


Image from [Dey et al. 2008]

Homology

Hence:

Up to isomorphism, the **Betti numbers** and the **torsion coefficients** of K
completely characterize the **homology groups** of K

Working with coefficients in a field \mathbb{F} :

Up to isomorphism, the **Betti numbers** of K
completely characterize the **homology groups** of K

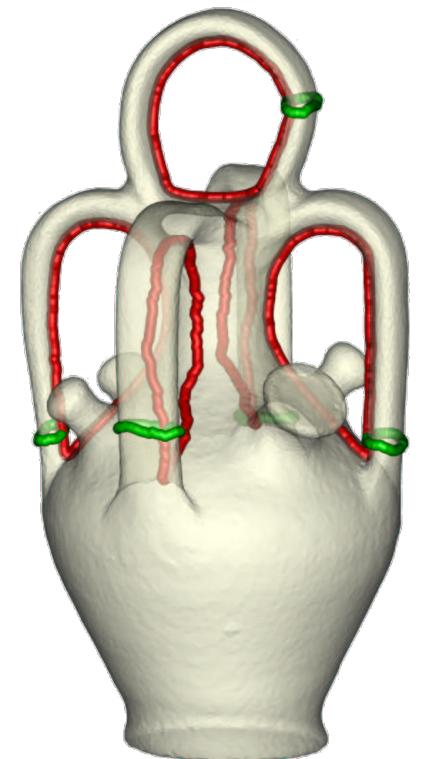
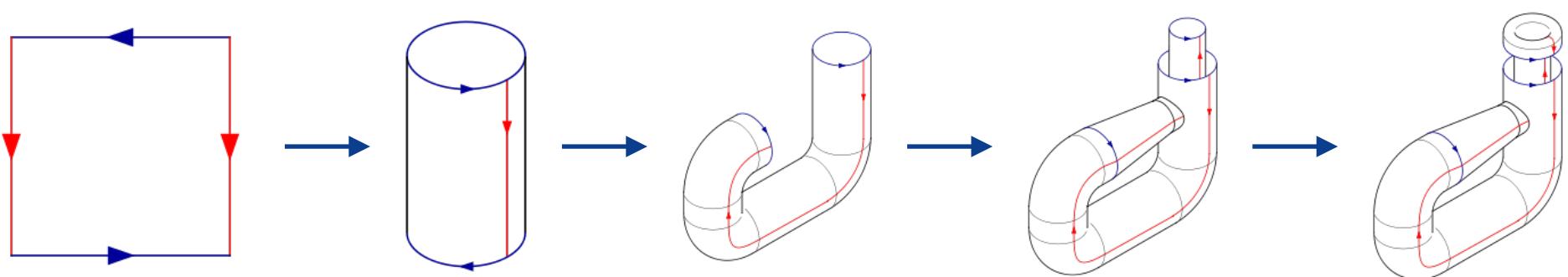
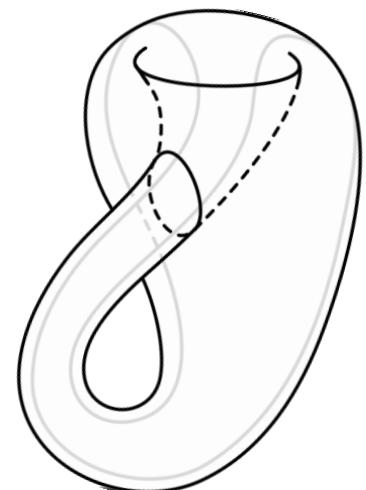


Image from [Dey et al. 2008]

Homology

Example:

The **Klein bottle K** is a non-orientable 2-dimensional manifold embeddable in \mathbb{R}^4 which can be built from a unit square by the following construction

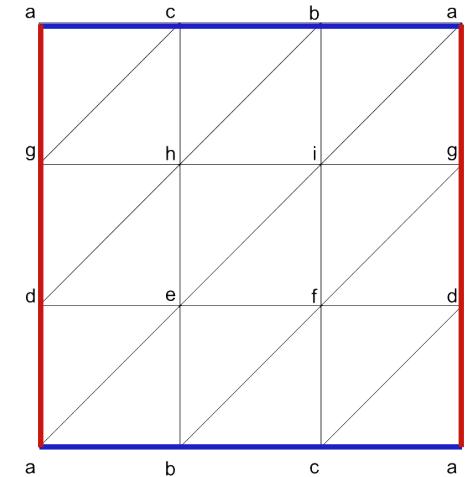
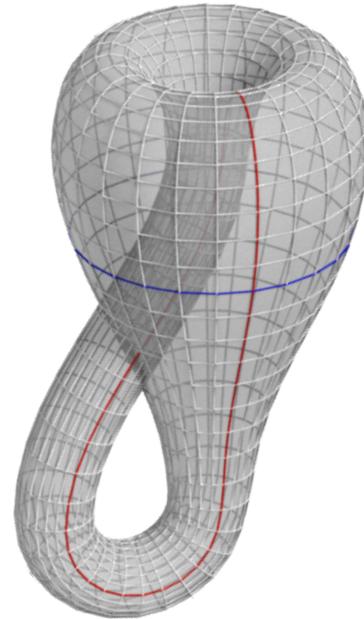


Homology

Example:

K has the following homology groups

$$H_k(K) \cong \begin{cases} \mathbb{Z} & \text{for } k = 0 \\ \mathbb{Z} \oplus \mathbb{Z}_2 & \text{for } k = 1 \\ 0 & \text{for } k \geq 2 \end{cases}$$



Homology

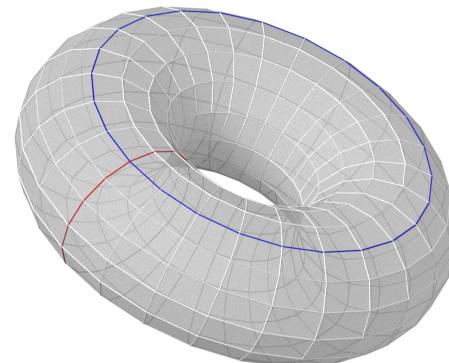
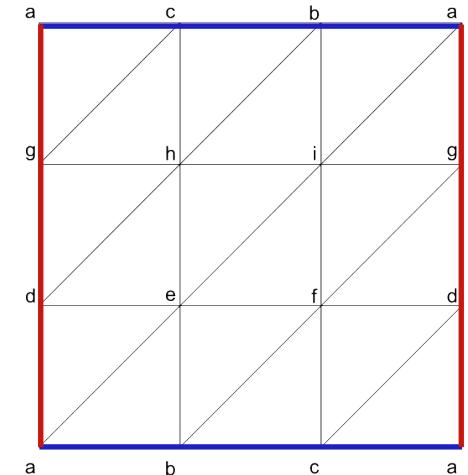
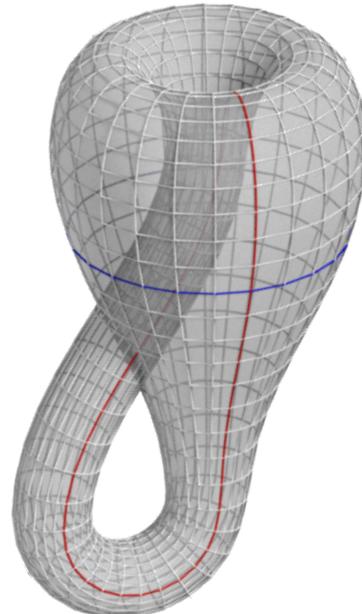
Example:

K has the following homology groups

$$H_k(K) \cong \begin{cases} \mathbb{Z} & \text{for } k = 0 \\ \mathbb{Z} \oplus \mathbb{Z}_2 & \text{for } k = 1 \\ 0 & \text{for } k \geq 2 \end{cases}$$

So, it can be distinguished from a torus T

$$H_k(T) \cong \begin{cases} \mathbb{Z} & \text{for } k = 0 \\ \mathbb{Z}^2 & \text{for } k = 1 \\ \mathbb{Z} & \text{for } k = 2 \\ 0 & \text{for } k > 2 \end{cases}$$

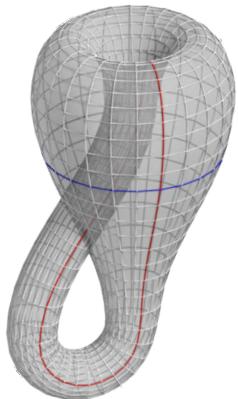


Homology

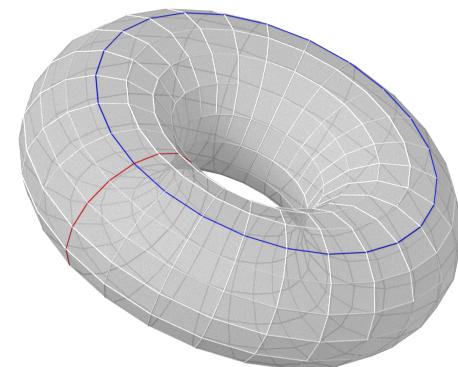
Example:

By considering \mathbb{Z}_2 as coefficient group,

the Klein bottle K and the torus T have isomorphic homology groups



$$H_k(K; \mathbb{Z}_2) \cong \begin{cases} \mathbb{Z}_2 & \text{for } k = 0 \\ \mathbb{Z}_2 \oplus \mathbb{Z}_2 & \text{for } k = 1 \\ \mathbb{Z}_2 & \text{for } k = 2 \\ 0 & \text{for } k > 2 \end{cases} \cong H_k(T; \mathbb{Z}_2)$$



Persistent Homology

In a Nutshell:

Persistent homology allows for
describing the changes in the shape of an evolving object

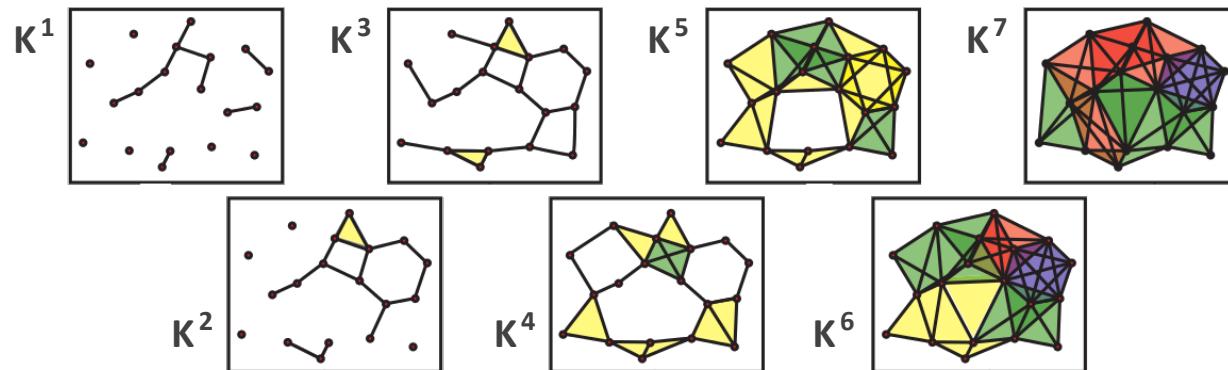


Image from [Ghrist 2008]

Persistent Homology

An Evolving Notion:

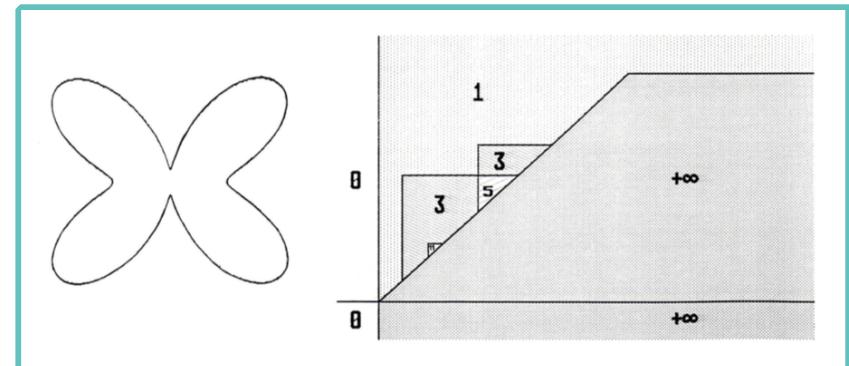
1990



Frosini

Size Functions:

- ◆ *Estimation of natural pseudo-distance* between shapes endowed with a function f
- ◆ Tracking of the ***connected components*** of a shape along its evolution induced by f



Actually, this coincides with ***persistent homology in degree 0***

Image from [Frosini 1992]

Persistent Homology

An Evolving Notion:



Incremental Algorithm for Betti Numbers:

- ◆ Introduction of the notion of ***filtration***
- ◆ De facto computation of ***persistence pairs***

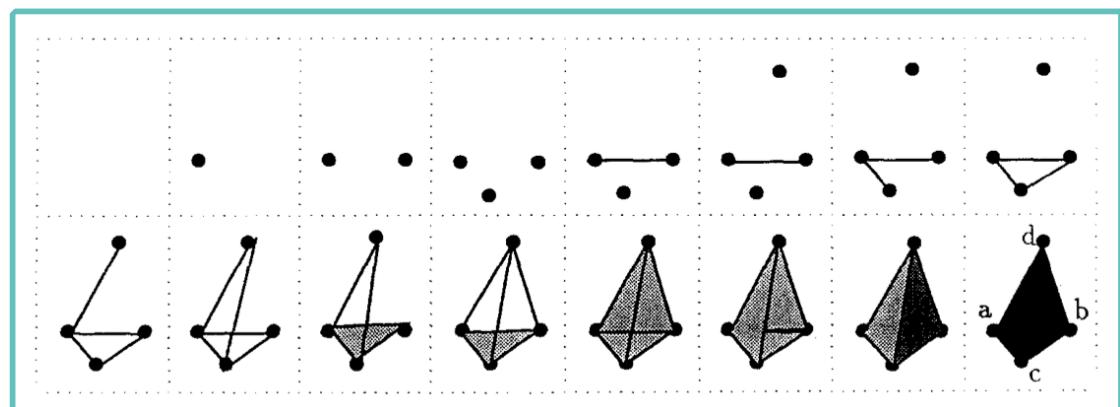


Image from [Delfinado, Edelsbrunner 1995]

Persistent Homology

An Evolving Notion:



Homology from Finite Approximations:

- ◆ *Extrapolation of the homology* of a metric space from a *finite point-set approximation*
- ◆ Introduction of *persistent Betti numbers*

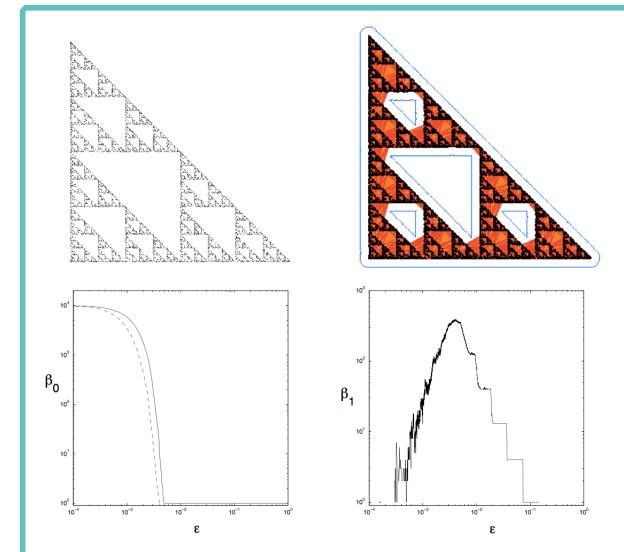
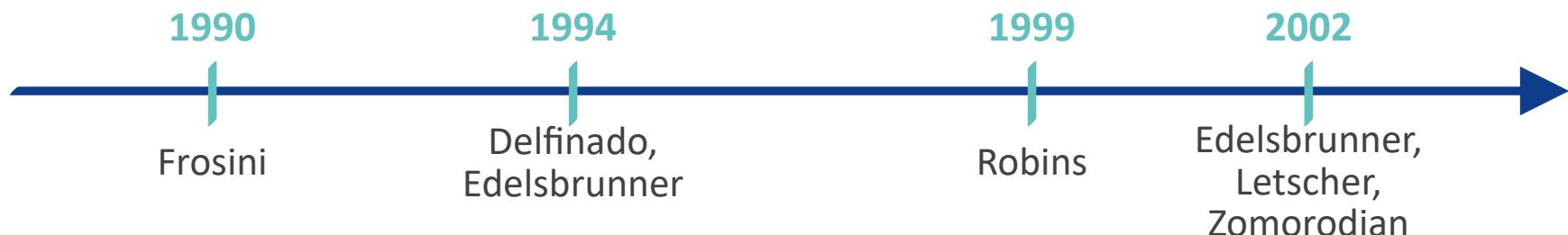


Image from [Robins 1999]

Persistent Homology

An Evolving Notion:



Topological Persistence:

- ◆ Introduction and algebraic formulation of the notion of ***persistent homology***
- ◆ ***Description of an algorithm*** for computing persistent homology

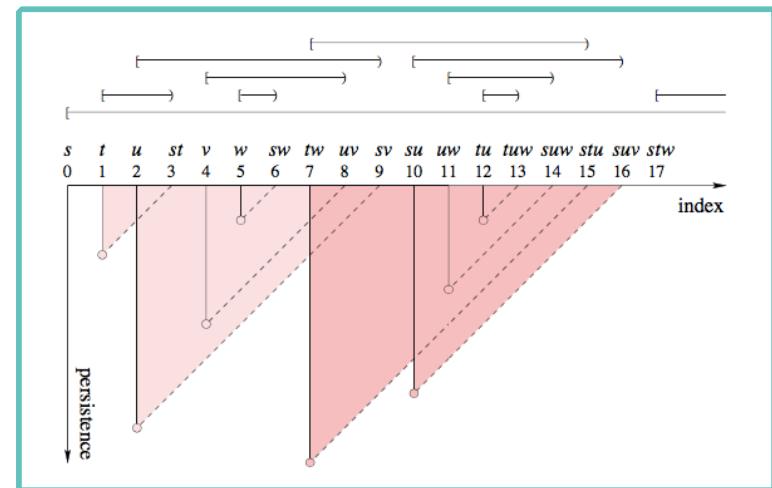


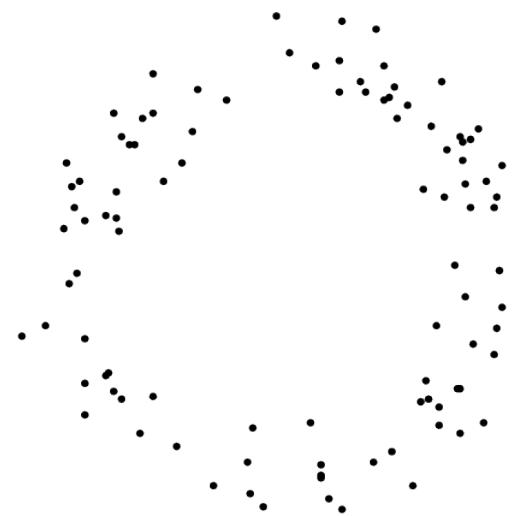
Image from [Edelsbrunner et al. 2002]

Persistent Homology

A Twofold Purpose:

Shape Description

- ◆ *Which is the shape of a given data?*

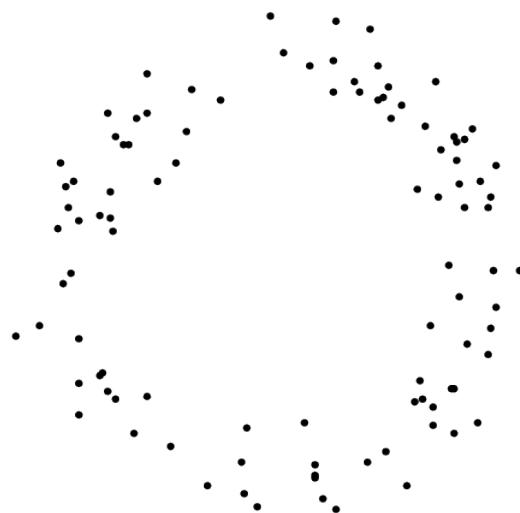
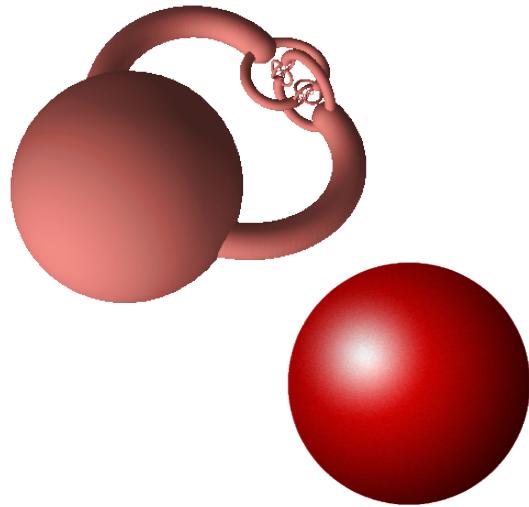


Persistent Homology

A Twofold Purpose:

Shape Description

- ◆ *Which is the shape of a given data?*



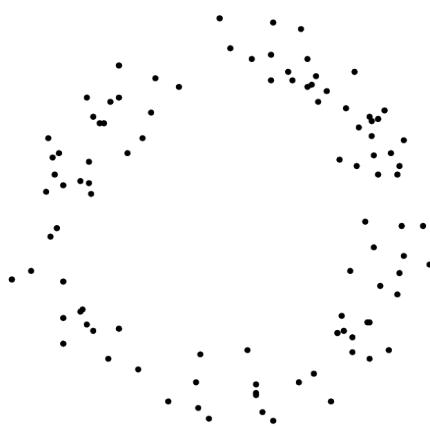
Shape Comparison

- ◆ *Given two data, do they have the same shape?*

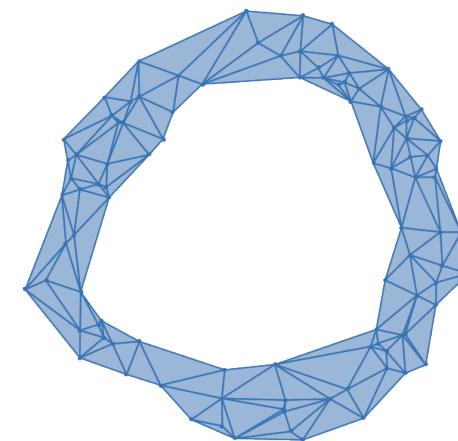
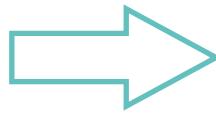
Persistent Homology

- ◆ *Which is the shape of a given data?*

Persistent homology allows for the retrieval of the “*actual*” homological information of a data



Point Cloud Dataset



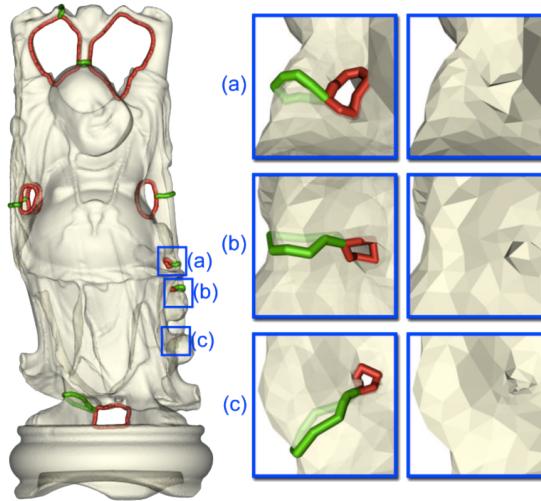
*Topological Nature of
the “Underlying” Shape*

Image from [Bauer 2015]

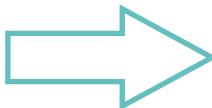
Persistent Homology

◆ *Which is the shape of a given data?*

Persistent homology allows for the retrieval of the “*actual*” homological information of a data



Noisy Dataset



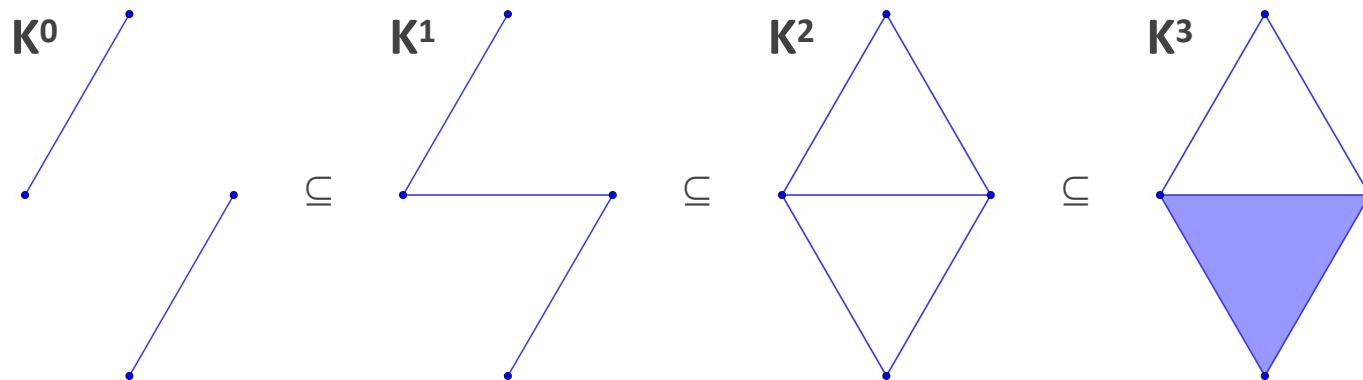
*Relevant Homological
Information*

Image from [Dey et al. 2008]

Persistence Pairs

The *core information* of persistent homology is given by the *persistence pairs*

Given a filtration $K^0 \subseteq K^1 \subseteq \dots \subseteq K^m$,

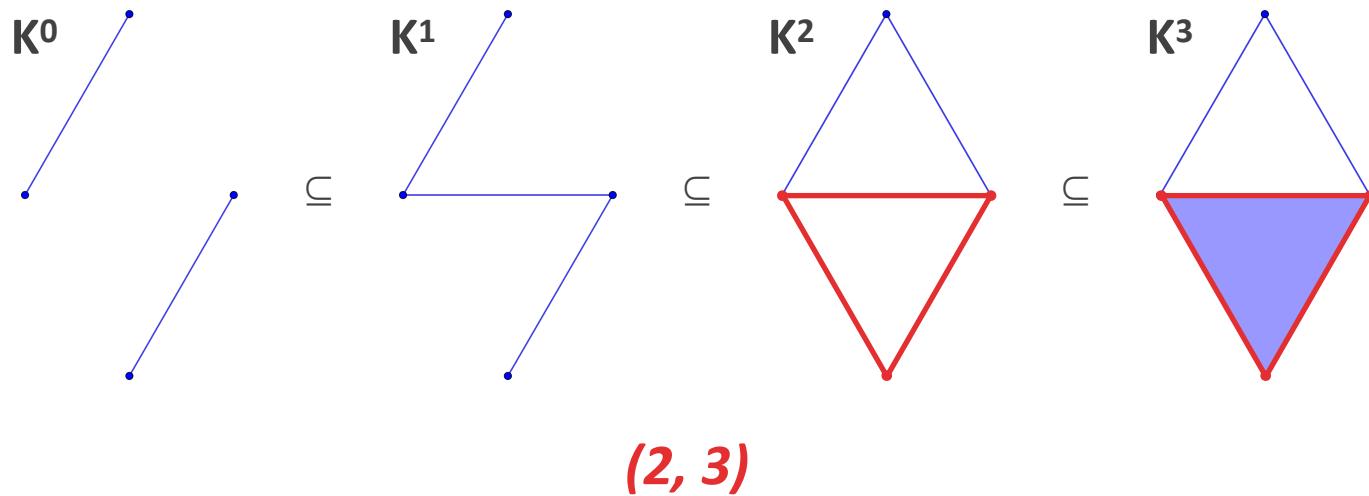


A **persistence pair** (p, q) is an element in $\{0, \dots, m\} \times (\{0, \dots, m\} \cup \{\infty\})$ such that $p < q$ representing a **homological class** that is **born at step p** and **dies at step q**

Persistence Pairs

The *core information* of persistent homology is given by the *persistence pairs*

Given a filtration $K^0 \subseteq K^1 \subseteq \dots \subseteq K^m$,

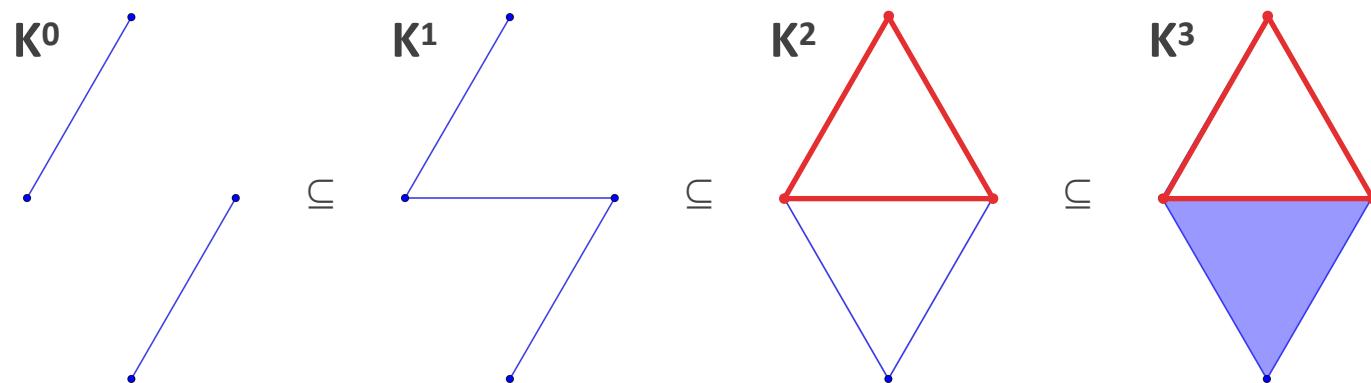


A **persistence pair** (p, q) is an element in $\{0, \dots, m\} \times (\{0, \dots, m\} \cup \{\infty\})$ such that $p < q$ representing a **homological class** that is **born at step p** and **dies at step q**

Persistence Pairs

The *core information* of persistent homology is given by the *persistence pairs*

Given a filtration $K^0 \subseteq K^1 \subseteq \dots \subseteq K^m$,



(2, ∞) essential pair

A **persistence pair** (p, q) is an element in $\{0, \dots, m\} \times (\{0, \dots, m\} \cup \{\infty\})$ such that $p < q$ representing a **homological class** that is **born at step p** and **dies at step q**

Persistence Pairs

Given a filtered simplicial complex K , a field \mathbb{F} , and $k \in \mathbb{N}$,

its ***persistence module*** $M := \bigoplus_{t \in \mathbb{N}} H_k(K^t; \mathbb{F})$ is a ***finitely generated $\mathbb{F}[x]$ -module***

The corresponding structure theorem ensures us that

Theorem:

The persistence module M can be expressed as

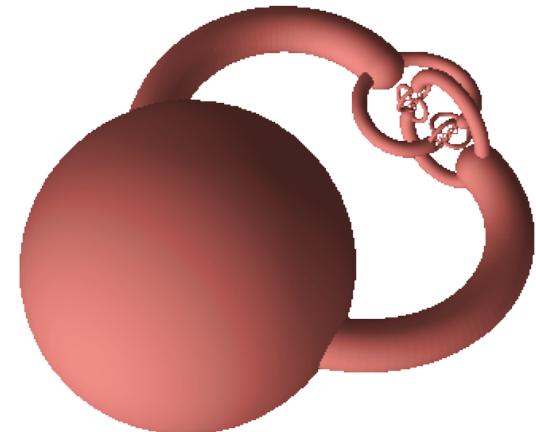
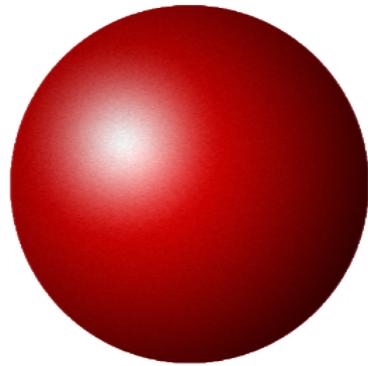
$$M \cong \bigoplus_{j=1}^n \mathbb{F}[x](-a_j) \bigoplus_{i=1}^m \mathbb{F}[x](-c_i)/x^{d_i}$$

*So, the ***persistence module M*** is completely determined by its ***persistence pairs****

i.e., the collection of the pairs $(c_i, d_i), (a_j, \infty)$

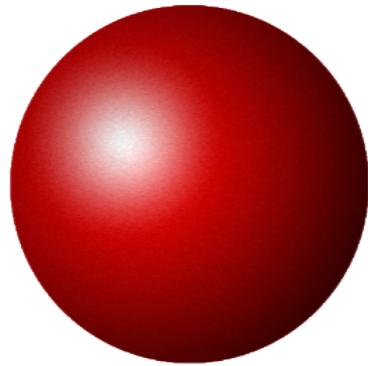
Persistence Pairs

◆ *Do they have the same shape?*

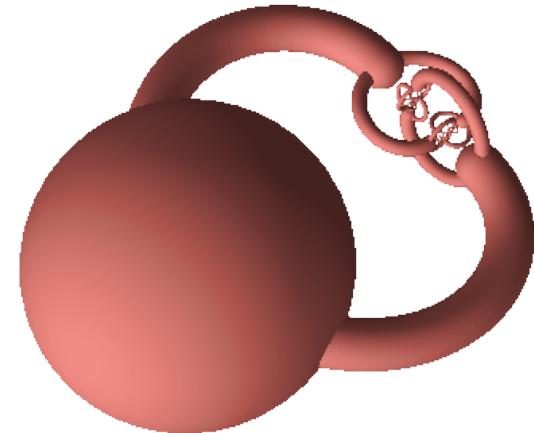


Persistence Pairs

◆ *Do they have the same shape?*



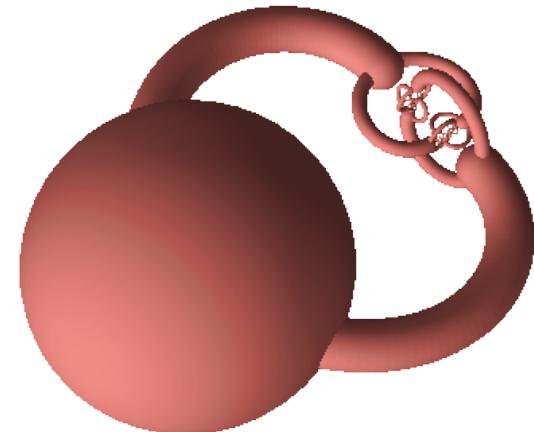
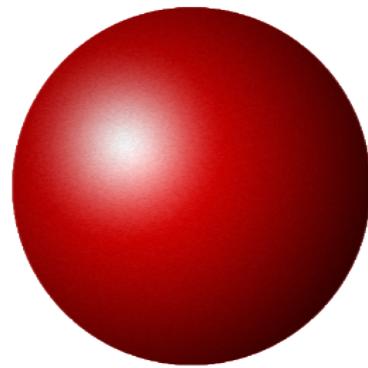
In Practice?



In Theory?

Persistence Pairs

◆ *Do they have the same shape?*



In Practice?



In Theory?



They are homeomorphic

Persistence Pairs

◆ *Do they have the same shape?*



Persistence Pairs

◆ *Do they have the same shape?*



In Practice?

In Theory?

Persistence Pairs

◆ *Do they have the same shape?*



In Practice?



In Theory?



They are not homeomorphic

Persistence Pairs

*Differently from homology, persistent homology provides
a notion of “shape” closer to our everyday perception*

It is possible to *compare two shapes* by comparing their *homology groups*

Persistence Pairs

*Differently from homology, persistent homology provides
a notion of “shape” closer to our everyday perception*

It is possible to **compare two shapes** by comparing their **homology groups**



Persistence Pairs

*Differently from homology, persistent homology provides
a notion of “shape” closer to our everyday perception*

It is possible to **compare two shapes** by comparing their **homology groups**

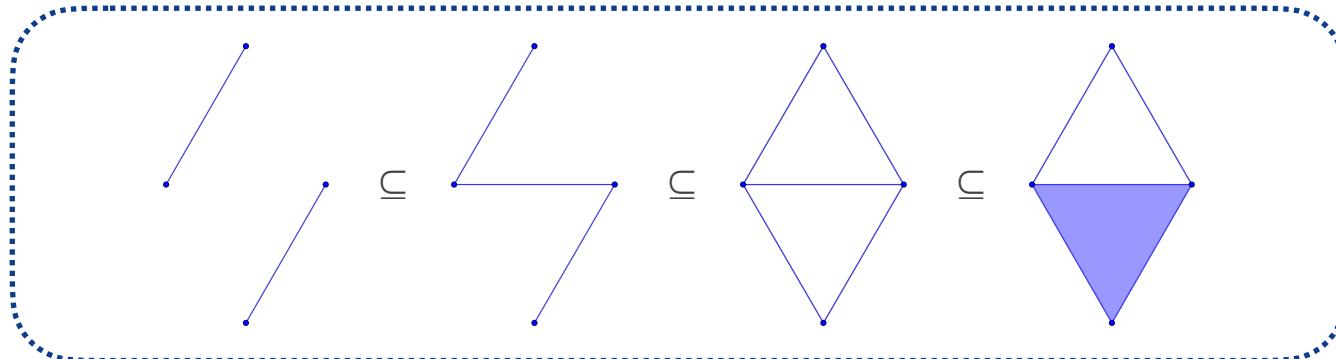


In order to better perform the above task, we need:

- ◆ **Visual** and **descriptive representations** for persistence pairs
- ◆ Notions of **distance** between sets of persistence pairs and **stability results**

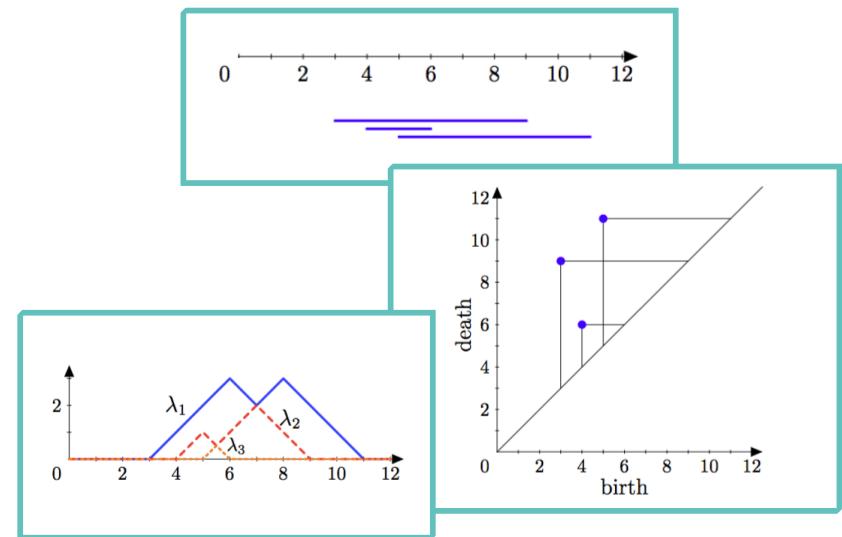
Visualizing Persistence Pairs

Given a filtered simplicial complex K ,



Persistent pairs of K can be visualized through:

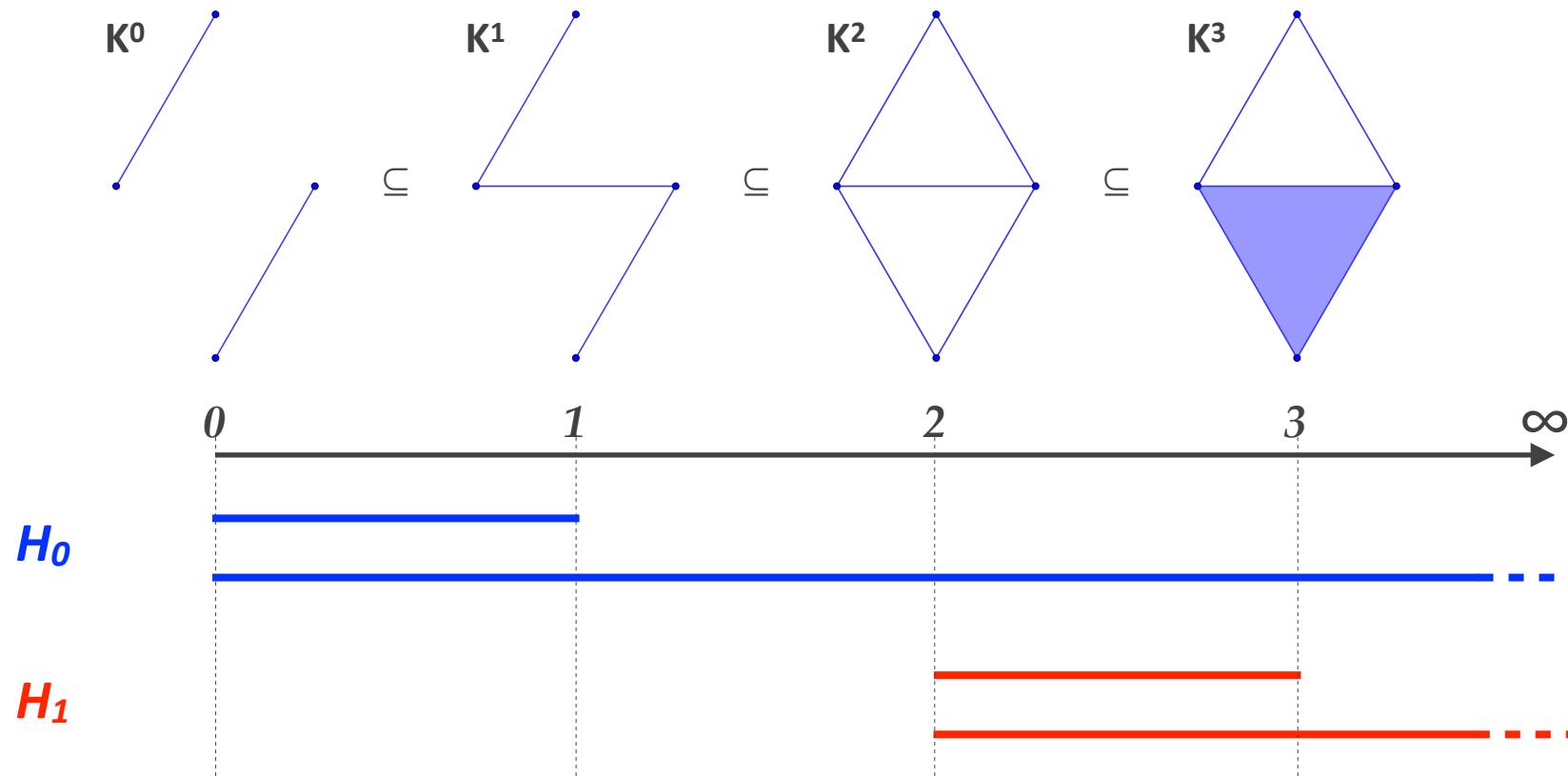
- ◆ **Barcodes** [Carlsson et al. 2005; Ghrist 2008]
- ◆ **Persistence diagrams** [Edelsbrunner, Harer 2008]
- ◆ **Persistence landscapes** [Bubenik 2015]
- ◆ **Corner points and lines** [Frosini, Landi 2001]
- ◆ **Half-open intervals** [Edelsbrunner et al. 2002]
- ◆ **k -triangles** [Edelsbrunner et al. 2002]



Visualizing Persistence Pairs

Barcodes:

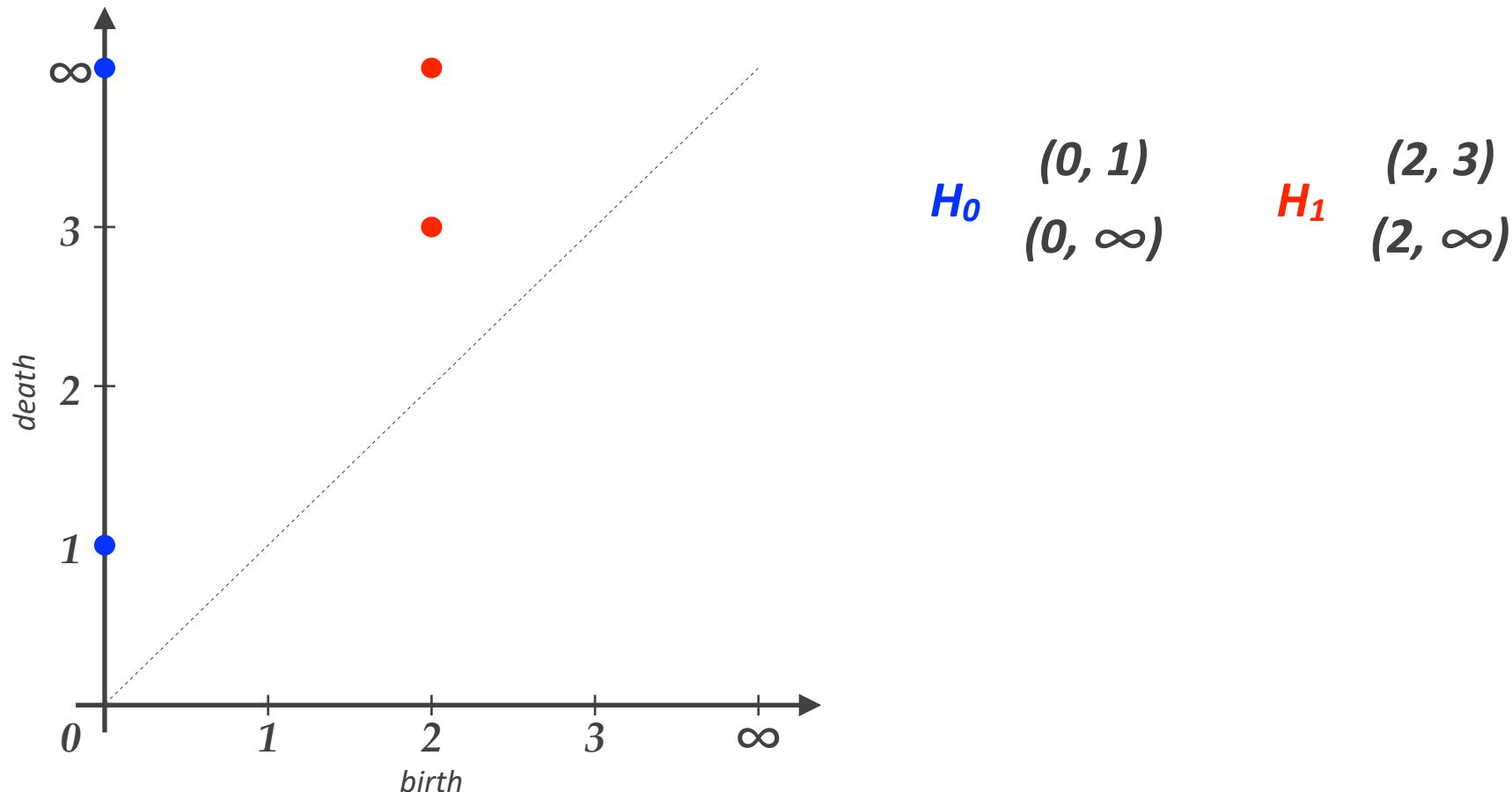
Persistence pairs are represented as **intervals in \mathbb{R}**



Visualizing Persistence Pairs

Persistence Diagrams:

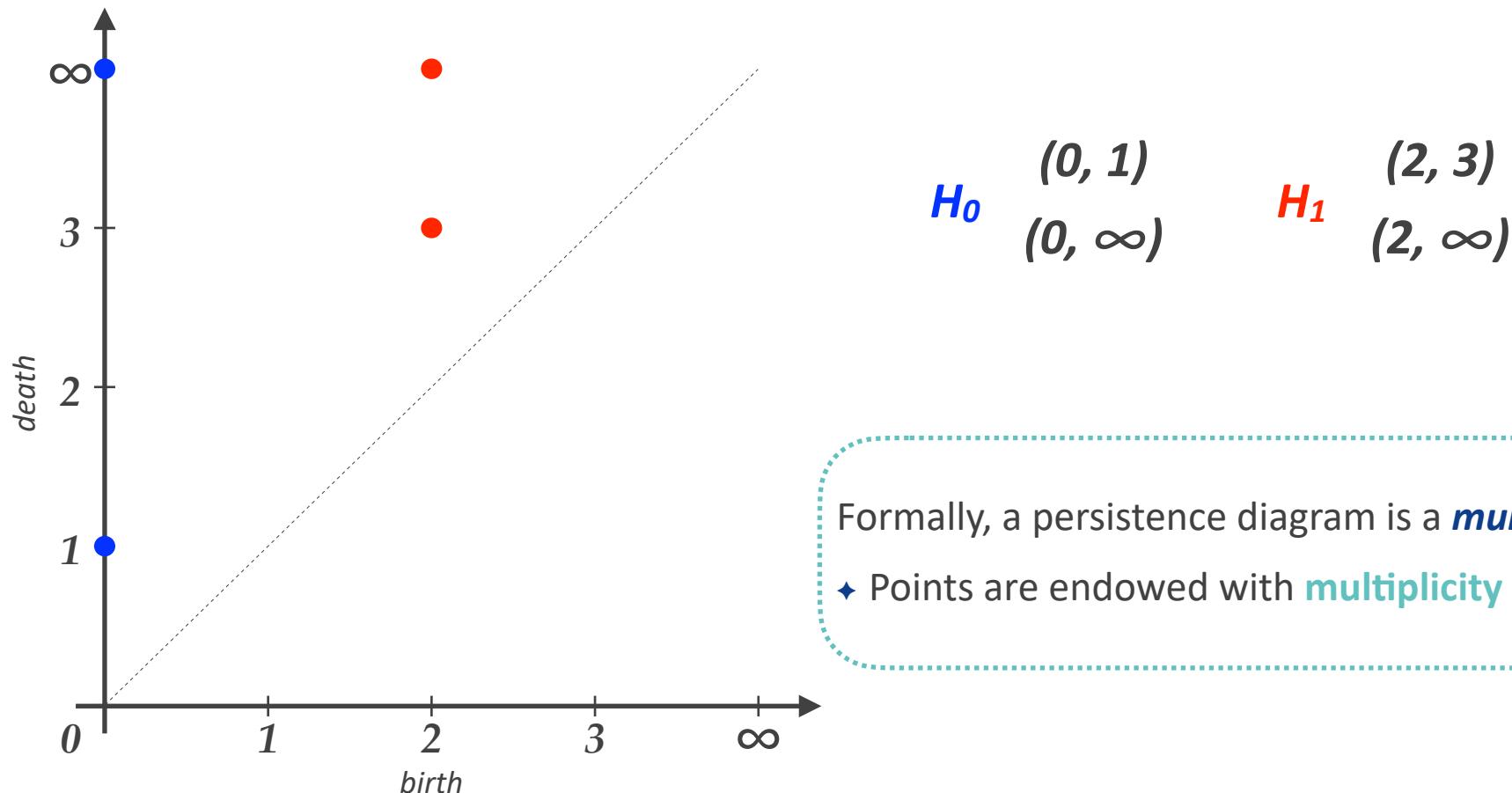
Persistence pairs are represented as *points* in \mathbb{R}^2



Visualizing Persistence Pairs

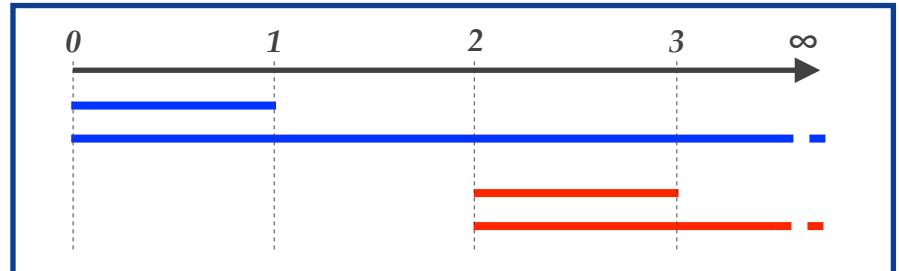
Persistence Diagrams:

Persistence pairs are represented as **points in $\mathbb{R} \times (\mathbb{R} \cup \{\infty\})$**



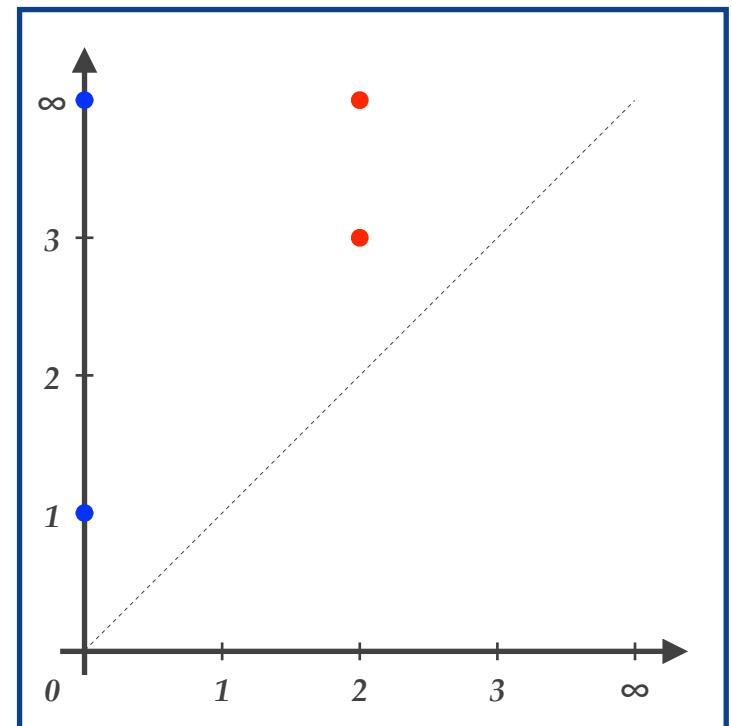
Visualizing Persistence Pairs

Both tools **visually represent** the **lifespan** of the homology classes:



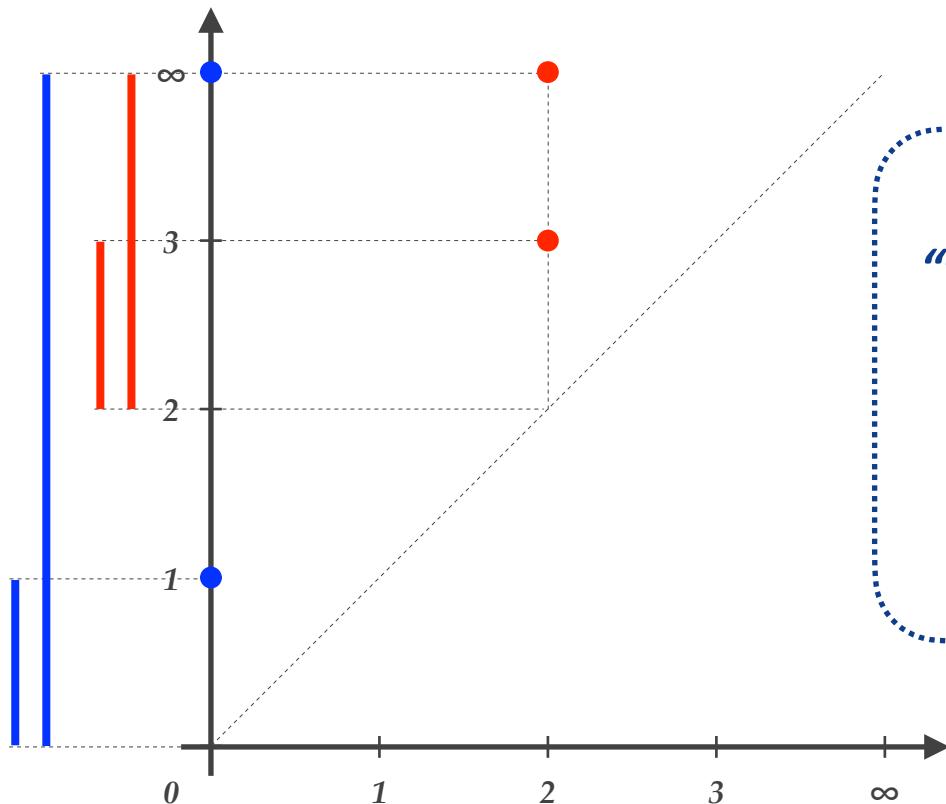
- ◆ Barcode: **length of the intervals**
- ◆ Persistence Diagram: **distance from the diagonal**

Barcodes and Persistence Diagrams
encode equivalent information



Visualizing Persistence Pairs

Barcodes and Persistence Diagrams *encode equivalent information*



A visualization can be easily
“*translated*” into the other one:

$$\begin{array}{ccc} [p, q] & \longleftrightarrow & (p, q) \\ [p, \infty) & & (p, \infty) \end{array}$$

Visualizing Persistence Pairs

Example:

Brain functional networks from fMRI (functional magnetic resonance imaging)

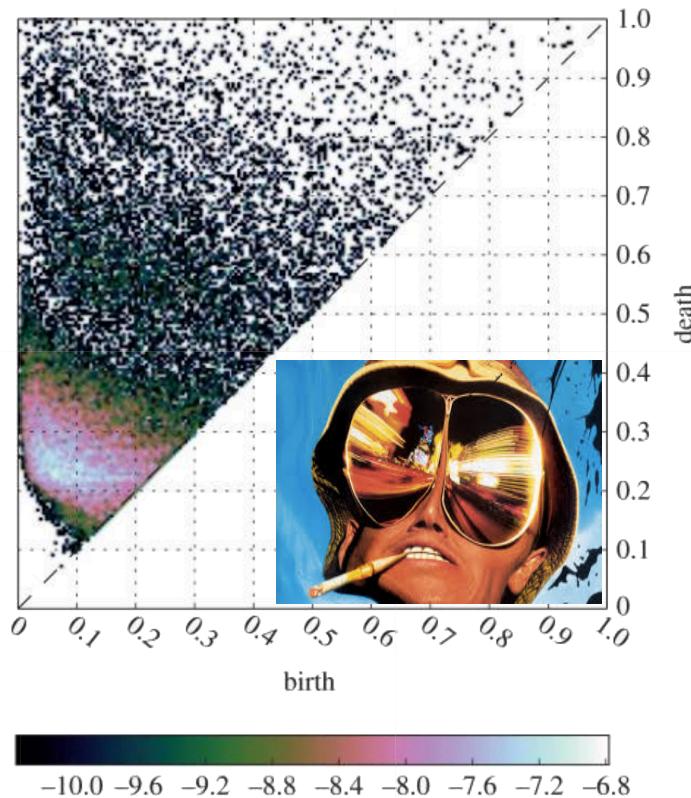
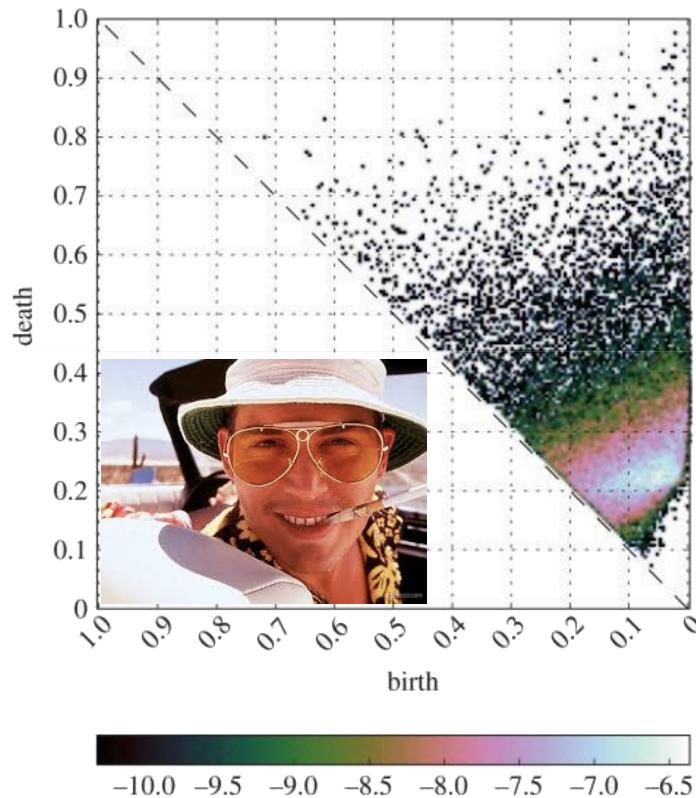


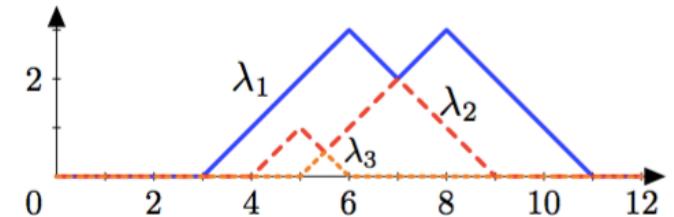
Image from
[Petri et al. 2014]

(log-)Probability densities of H_1 for the placebo (left) and the psilocybin (right) groups

Visualizing Persistence Pairs

Persistence Landscapes:

Persistence landscapes are statistics-friendly representations of persistence pairs

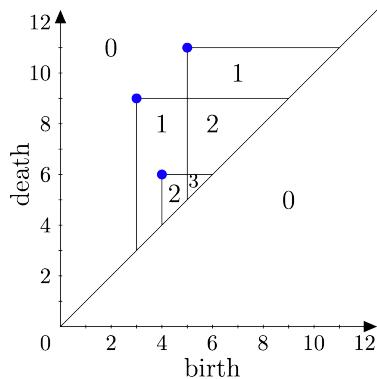


Given a persistence module M , persistence landscapes

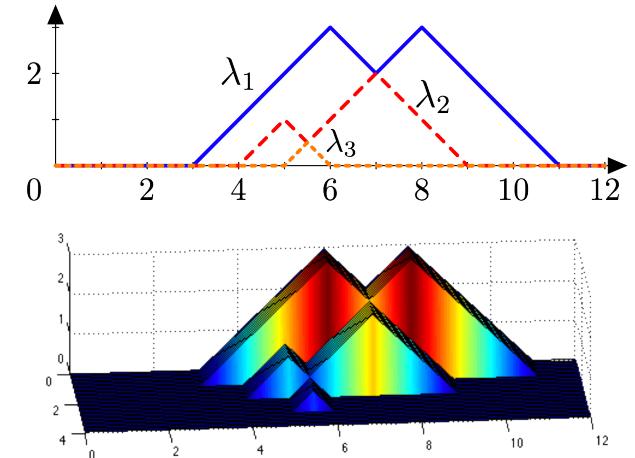
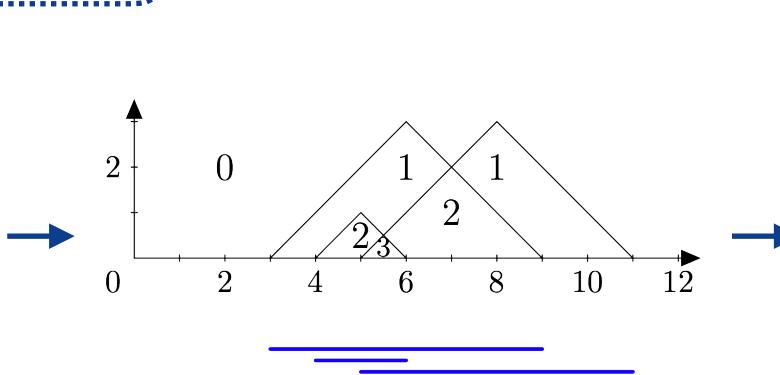
- ◆ Consist of a collection of **1-Lipschitz functions**
- ◆ Lie in a **vector space**
- ◆ Are **stable** (under small perturbations of the input filtration)

Visualizing Persistence Pairs

Persistence Landscapes:



Given a persistence module M ,



Formally,

Images from [Bubenik 2015]

$$\lambda_i(x) := \sup\{m \geq 0 \mid \beta^{x-m, x+m} \geq i\}$$

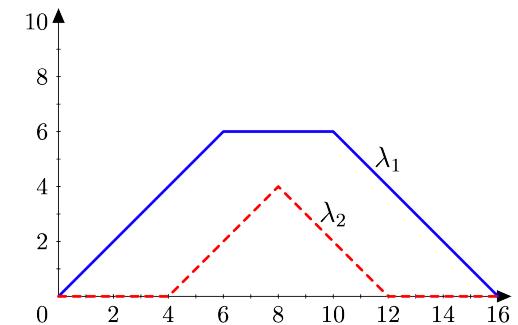
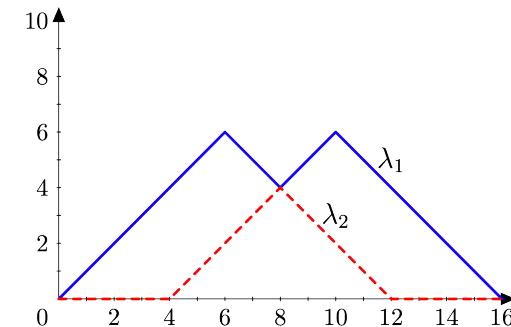
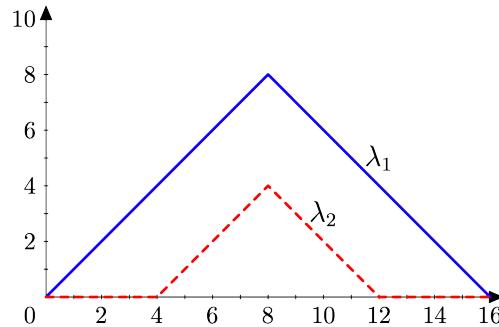
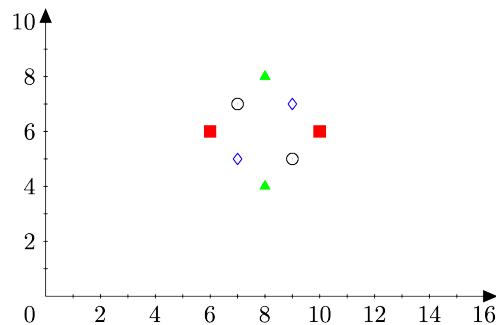
where $\beta^{a,b} := \dim(\text{im}(\iota_{a,b} : M_a \rightarrow M_b))$

Visualizing Persistence Pairs

Persistence Landscapes:

Mean of persistence diagrams is *not unique*, but ...

Mean of persistence landscapes is **well-defined**



Images from [Bubenik 2015]

Persistence & Stability

- ◆ *Persistence Pairs and their Visualizations*
- ◆ ***Stability Results for Persistent Homology***

Stability of Persistence Pairs

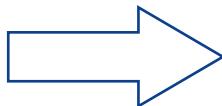
In order to be adopted in real applicative domains, it is crucial that

persistent homology is not affected by noisy data and small perturbations

Stability Result:

*By defining **distances*** for both domains,*

Similar Data



***Similar
Persistent Homology***

*The term “distance” is intended in a broad sense, including pseudo-metrics and dissimilarity measures

Stability of Persistence Pairs

Distances:

- ◆ **For the Data in Input:**
 - ❖ *Natural pseudo-distance* of shapes
 - ❖ *L_∞ -distance* of filtering functions
 - ❖ *Gromov-Hausdorff distance* of metric spaces/point clouds
- ◆ **For the retrieved Persistent Homology Information:**
 - ❖ *Interleaving distance* of persistence modules
 - ❖ *Bottleneck distance* of persistence diagrams
 - ❖ *Hausdorff distance* of persistence diagrams
 - ❖ *Matching distance* of persistence diagrams
 - ❖ *Wasserstein distances* of persistence diagrams

Stability of Persistence Pairs

Distances for Input Data:

Let (X, f) be a **pair** such that:

- ◆ X is a **(triangulable) topological space**
- ◆ $f: X \rightarrow \mathbb{R}$ is a **continuous function**

A pair (X, f) induces a **filtration**:

- ◆ $X^t := f^{-1}((-\infty, t])$

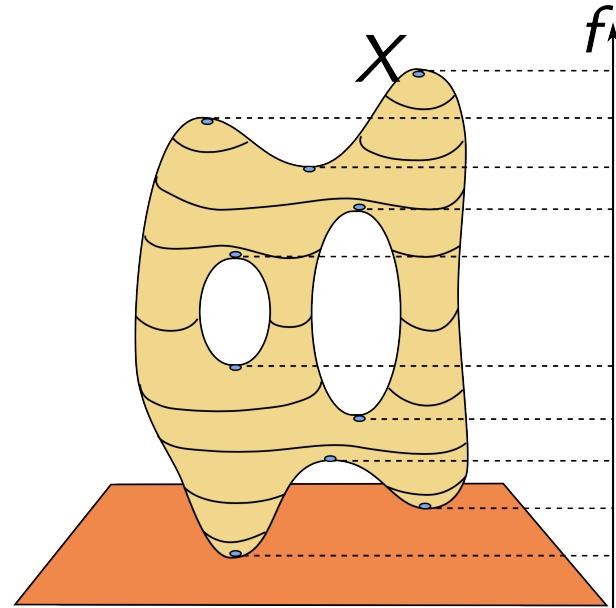


Image from [Ferri et al. 2015]

Definition:

The function f is called **tame** if:

- ◆ f has a **finite number of homological critical values** (i.e. the “time” steps in which homology changes)
- ◆ For any $k \in \mathbb{N}$ and $t \in \mathbb{R}$, the **homology group $H_k(X^t, \mathbb{F})$ has finite dimension**

Stability of Persistence Pairs

Distances for Input Data:

Definition:

Given two pairs (X, f) and (Y, g) , their **natural pseudo-distance d_N** is defined as:

$$d_N((X, f), (Y, g)) := \begin{cases} \inf_{h \in H(X, Y)} \{\max_{x \in X} \{|f(x) - g \circ h(x)|\}\} & \\ +\infty & \text{if } H(X, Y) = \emptyset \end{cases}$$

where **$H(X, Y)$** is the set of all the **homeomorphisms between X and Y**

Stability of Persistence Pairs

Distances for Input Data:

Working with two functions $f, g: X \rightarrow \mathbb{R}$ defined on the same topological space X , one can simply consider the L_∞ -distance between f and g

$$\|f - g\|_\infty := \sup_{x \in X} \{|f(x) - g(x)|\}$$

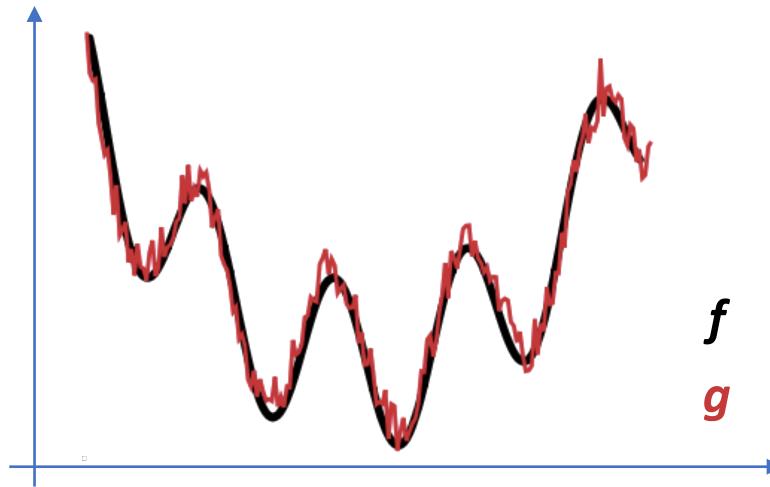


Image from [Rieck 2016]

Stability of Persistence Pairs

Distances for Input Data:

Given two **finite metric spaces** (X, d_X) , (Y, d_Y) (e.g. two finite point clouds in \mathbb{R}^n),

Definitions:

A **correspondence** $C: X \rightrightarrows Y$ from X to Y is a subset of $X \times Y$ such that
the **canonical projections** $\pi_X: C \rightarrow X$ and $\pi_Y: C \rightarrow Y$ are both **surjective**

The **distortion $dis(C)$** of a correspondence $C: X \rightrightarrows Y$ is defined as:

$$dis(C) := \sup \left\{ |d_X(x, x') - d_Y(y, y')| : (x, y), (x', y') \in C \right\}$$

The **Gromov-Hausdorff distance d_{GH}** between (X, d_X) and (Y, d_Y) is defined as:

$$d_{GH}(X, Y) := \frac{1}{2} \inf \{ dis(C) \mid C: X \rightrightarrows Y \text{ is a correspondence} \}$$

Stability of Persistence Pairs

Distances for Persistent Homology Information:

Two persistence modules M and N are called ε -interleaved with $\varepsilon \geq 0$ if there exist f and g such that, for any $p, q \in \mathbb{R}$ with $p \leq q$, the following **diagrams commute**

$$\begin{array}{ccc}
 & M_p & \\
 g_{p-\varepsilon} \nearrow & \searrow f_p & \\
 N_{p-\varepsilon} & \xrightarrow{\quad} & N_{p+\varepsilon} \\
 & M_p \longrightarrow & M_q \\
 & f_p \searrow & \swarrow f_q \\
 & N_{p+\varepsilon} & \xrightarrow{\quad} N_{q+\varepsilon} \\
 \\
 M_{p-\varepsilon} & \longrightarrow & M_{p+\varepsilon} \\
 f_{p-\varepsilon} \searrow & \nearrow g_p & \\
 & N_p & \\
 \\
 & M_{p+\varepsilon} & \longrightarrow M_{q+\varepsilon} \\
 g_p \nearrow & \nearrow g_q & \\
 N_p & \xrightarrow{\quad} & N_q
 \end{array}$$

Definition:

Given two persistence modules M and N , their **interleaving distance d_I** is defined as:

$$d_I(M, N) := \inf\{\varepsilon \geq 0 \mid M \text{ and } N \text{ are } \varepsilon\text{-interleaved}\}$$

Stability of Persistence Pairs

Distances for Persistent Homology Information:

Definitions:

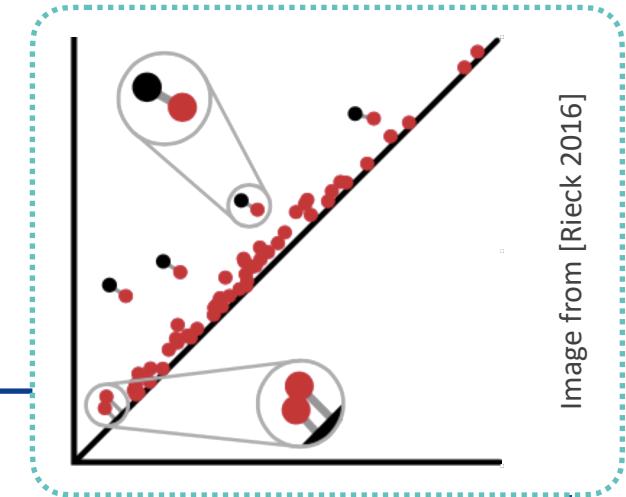
Given two persistence diagrams D_1 and D_2 ,

their **bottleneck distance** d_B and **Hausdorff distance** d_H are defined as:

$$d_B(D_1, D_2) := \inf_{\gamma} \left\{ \sup_{x \in D_1} \{ \|x - \gamma(x)\|_{\infty} \} \right\}$$

$$d_H(D_1, D_2) := \max \left\{ \sup_{x \in D_1} \left\{ \inf_{y \in D_2} \{ \|x - y\|_{\infty} \} \right\}, \sup_{y \in D_2} \left\{ \inf_{x \in D_1} \{ \|y - x\|_{\infty} \} \right\} \right\}$$

where γ ranges over all bijections from D_1 to D_2



Stability of Persistence Pairs

Distances for Persistent Homology Information:

Definitions:

Given two persistence diagrams D_1 and D_2 ,

their **bottleneck distance** d_B and **Hausdorff distance** d_H are defined as:

$$d_B(D_1, D_2) := \inf_{\gamma} \left\{ \sup_{x \in D_1} \{ \|x - \gamma(x)\|_{\infty}\} \right\}$$

$$d_H(D_1, D_2) := \max \left\{ \sup_{x \in D_1} \left\{ \inf_{y \in D_2} \{ \|x - y\|_{\infty}\} \right\}, \sup_{y \in D_2} \left\{ \inf_{x \in D_1} \{ \|y - x\|_{\infty}\} \right\} \right\}$$

where γ ranges over all bijections from D_1 to D_2

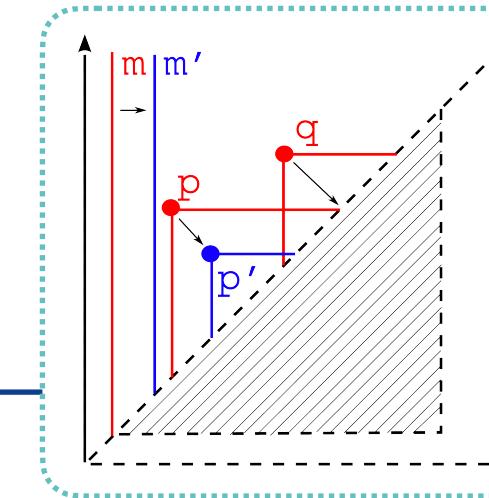


Image from [Ferri et al. 2015]

Stability of Persistence Pairs

Stability Results:

Given two pairs $(X, f), (Y, g)$ of topological spaces and **tame** functions and $k \in \mathbb{N}$, let M, N be the induced k^{th} persistence modules and let D_1, D_2 be the corresponding persistence diagrams

- ◆ $d_H(D_1, D_2) \leq d_B(D_1, D_2)$
- ◆ $d_I(M, N) = d_B(D_1, D_2)$

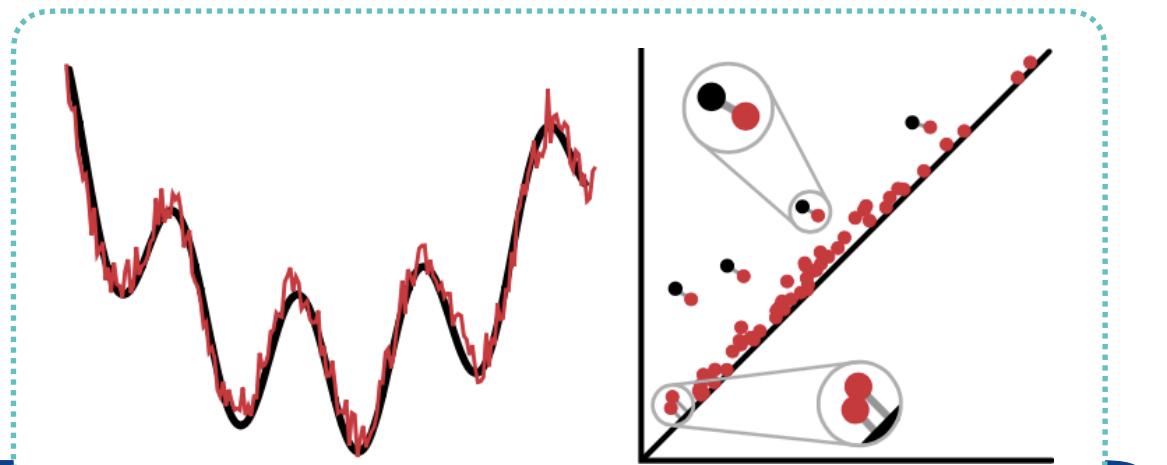
Theorem:

Under the above hypothesis, the following **optimal lower bound** holds

$$d_I(M, N) \leq d_N((X, f), (Y, g))$$

Stability of Persistence Pairs

Stability Results:



Theorem:

Given two **tame** continuous functions $f, g: X \rightarrow \mathbb{R}$
on a topological space X , $k \in \mathbb{N}$, and D_f, D_g the induced k^{th} persistence diagrams,

$$d_B(D_f, D_g) \leq \|f - g\|_\infty$$

Stability of Persistence Pairs

Stability Results:

Theorem:

Given two finite metric spaces (X, d_X) , (Y, d_Y) , $k \in \mathbb{N}$, and D_X, D_Y the k^{th} persistence diagrams of the **filtrations of the Vietoris-Rips complexes generated by X and Y** ,

$$d_B(D_X, D_Y) \leq d_{GH}(X, Y)$$

Bibliography

General References:

- ◆ **Books on TDA:**
 - ❖ H. Edelsbrunner, J. Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.
 - ❖ R. W. Ghrist. *Elementary applied topology*. Seattle: Createspace, 2014.
- ◆ **Papers on TDA:**
 - ❖ G. Carlsson. *Topology and data*. Bulletin of the American Mathematical Society 46.2, pages 255-308, 2009.

Today's References:

- ◆ **Persistent Homology:**
 - ❖ U. Fugacci, S. Scaramuccia, F. Iuricich, L. De Floriani. *Persistent homology: a step-by-step introduction for newcomers*. Eurographics Italian Chapter Conference, pages 1-10, 2016.
- ◆ **Stability Results:**
 - ❖ D. Cohen-Steiner, H. Edelsbrunner, J. Harer. *Stability of persistence diagrams*. Discrete & Computational Geometry 37.1, pages 103-120, 2007.
 - ❖ F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, S. Y. Oudot. *Proximity of persistence modules and their diagrams*. Proc. of the 35 annual symposium on Computational Geometry, pages 237-246, 2009.
 - ❖ F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, S. Y. Oudot. *Gromov-Hausdorff stable signatures for shapes using persistence*. Computer Graphics Forum 28.5, pages 1393-1403, 2009.