

Topological Data Analysis

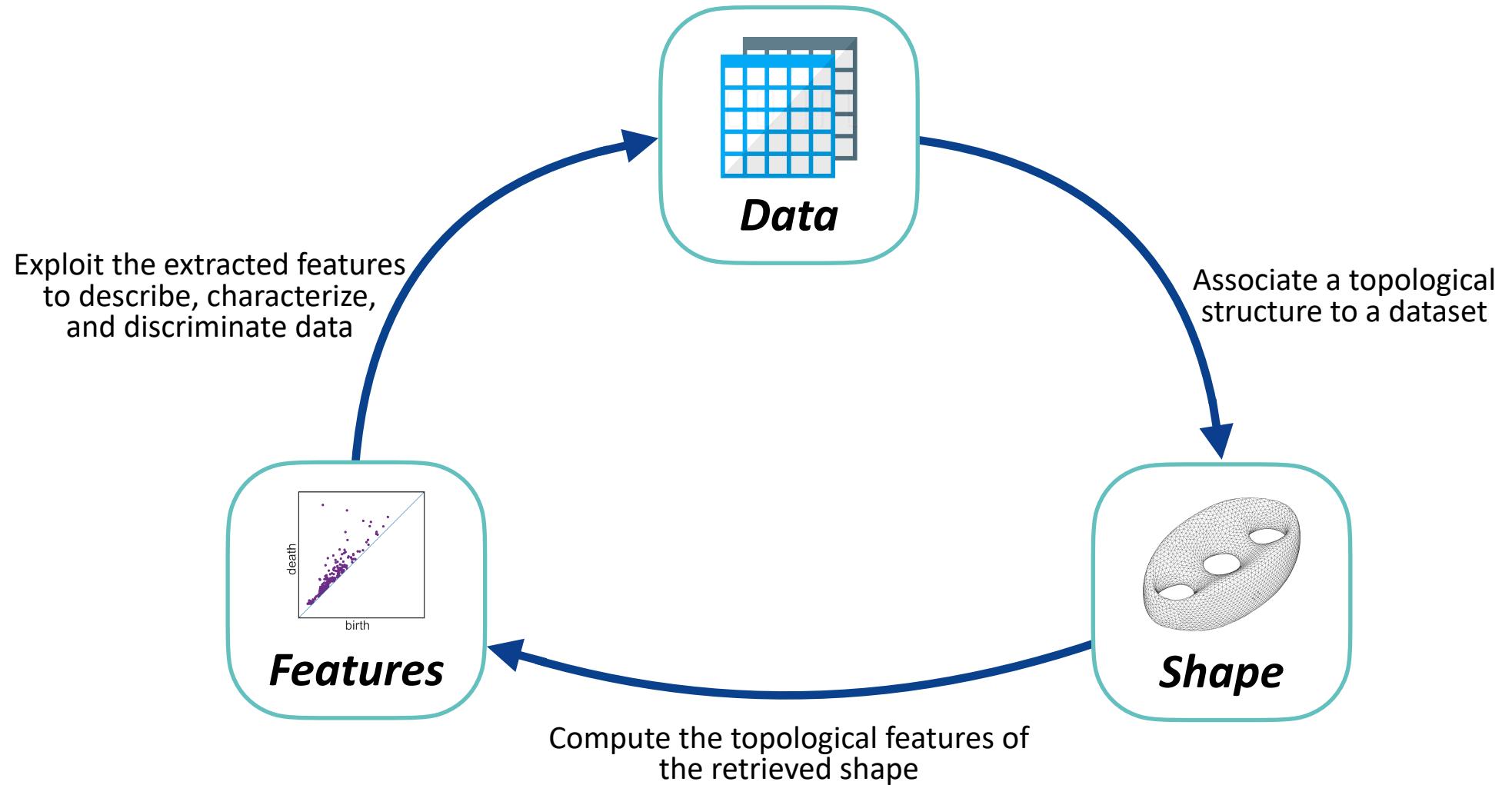
Homology & Persistence

Ulderico Fugacci

CNR - IMATI

international production
computer science
excellent research
innovation engineering
mathematics
education
scientific projects
leadership participation
knowledge
productive
applied
local
societal
tools
development
mission challenges
system spreading initiatives
high level infrastructure

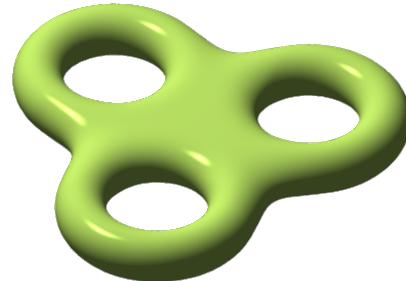
Topological Data Analysis



Homology & Persistence

Homology:

A topological invariant detecting
the “*holes*” of a shape

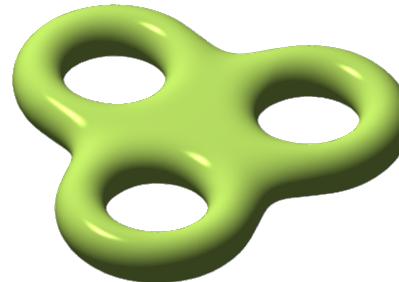


$$H_k(K) \cong \begin{cases} \mathbb{Z} & \text{for } k = 0 \\ \mathbb{Z}^6 & \text{for } k = 1 \\ \mathbb{Z} & \text{for } k = 2 \end{cases}$$

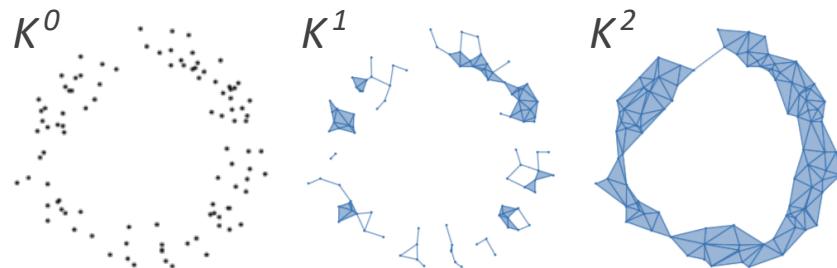
Homology & Persistence

Homology:

A topological invariant detecting the “holes” of a shape



$$H_k(K) \cong \begin{cases} \mathbb{Z} & \text{for } k = 0 \\ \mathbb{Z}^6 & \text{for } k = 1 \\ \mathbb{Z} & \text{for } k = 2 \end{cases}$$



Persistent Homology:

Captures the *changes in homology* during a filtration

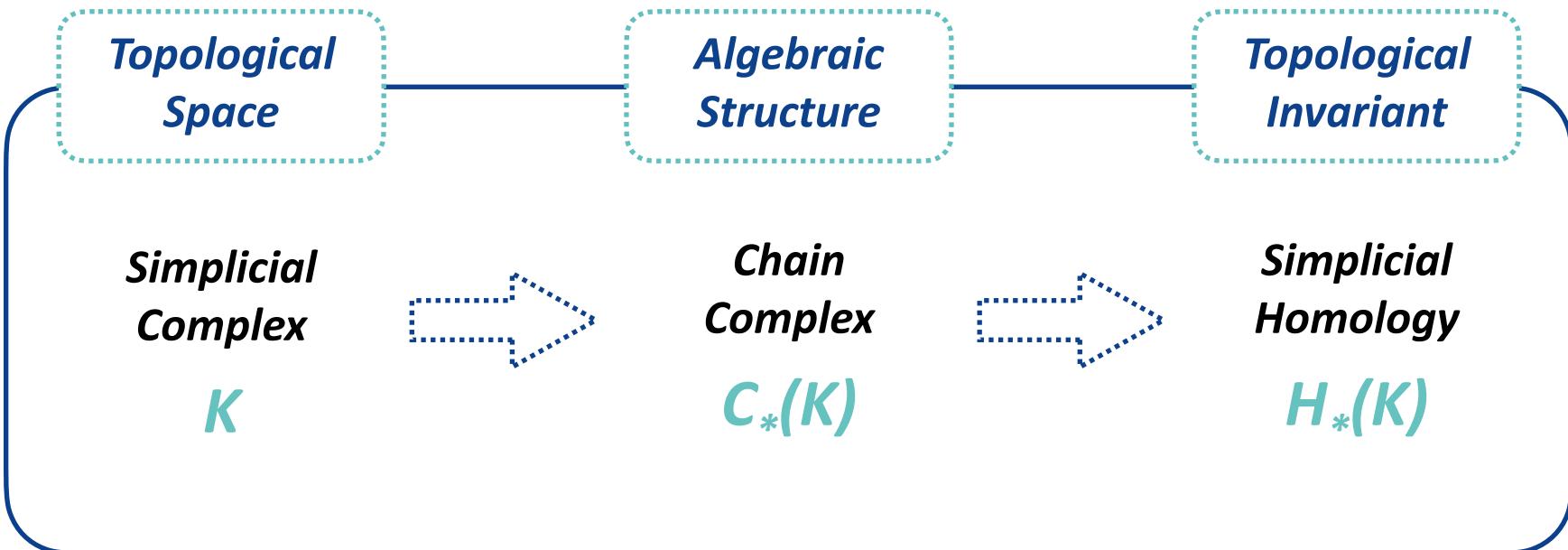
Homology & Persistence

- ◆ *Chain Complexes and Simplicial Homology*
- ◆ *Filtrations Persistent Homology*

Homology & Persistence

- ◆ *Chain Complexes and Simplicial Homology*
- ◆ *Filtrations and Persistent Homology*

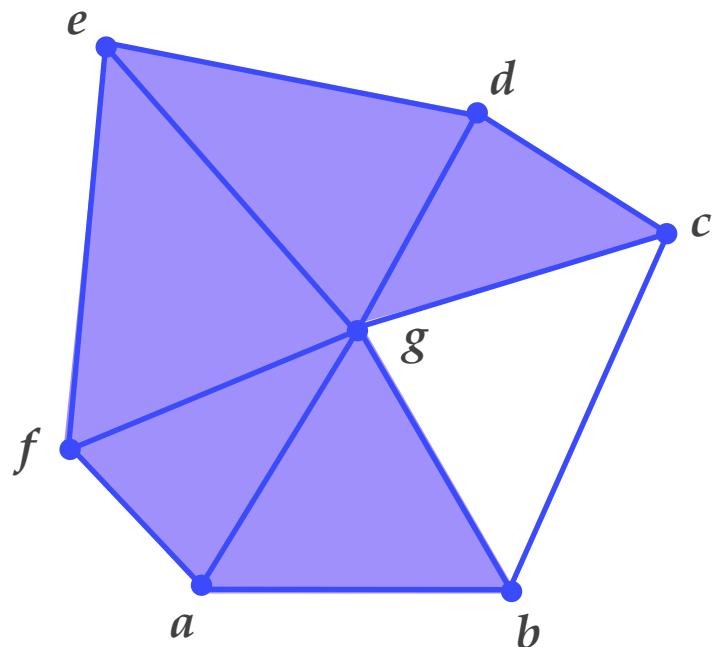
Simplicial Homology



Simplicial Homology

Given a simplicial complex K ,

- ◆ a ***k-chain*** is a formal sum (*with \mathbb{Z}_2 coefficients*) of k -simplices of K



Examples:

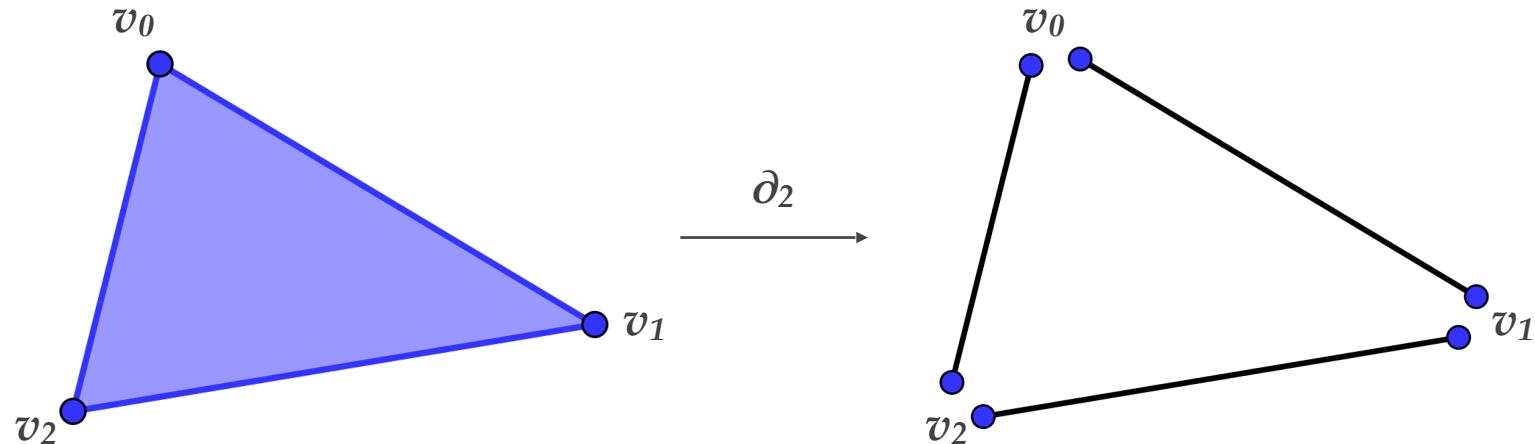
- ◆ $a + b + e$ is a 0-chain
- ◆ $fg + dg + de + eg$ is a 1-chain
- ◆ $abg + afg$ is a 2-chain

Simplicial Homology

The **chain complex** $C_*(K)$ associated with K consists of:

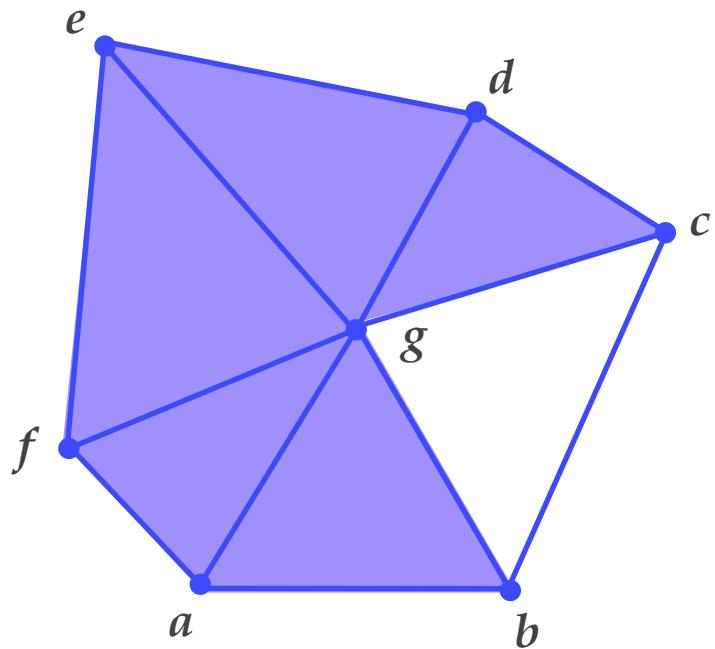
- ◆ a collection $\{C_k(K)\}_{k \in \mathbb{Z}}$ of vector spaces where $C_k(K)$ is the **group of the k -chains** of K
- ◆ a collection $\{\partial_k\}_{k \in \mathbb{Z}}$ of linear maps where the **boundary map** $\partial_k : C_k(K) \rightarrow C_{k-1}(K)$ is defined by

$$\partial_k(v_0 \cdots v_k) := \sum_{i=0}^k v_0 \cdots \hat{v}_i \cdots v_k$$



Simplicial Homology

Examples:



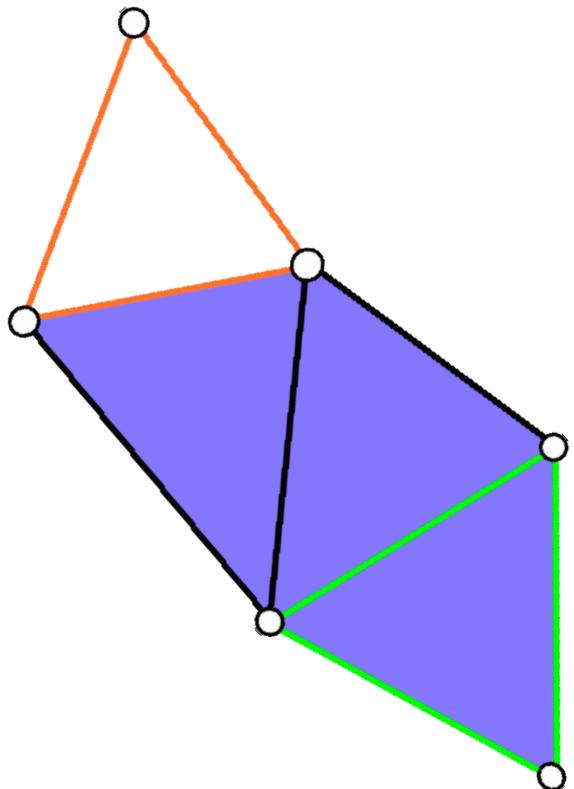
- ◆ $\partial_1(ab) = a + b$
- ◆ $\partial_1(ab + bc) = a + 2b + c = a + c$
- ◆ $\partial_2(afg + efg) = af + ag + 2fg + ef + eg = af + ag + ef + eg$
- ◆ $\partial_1(af + ag + ef + eg) = 2a + 2f + 2g + 2e = 0$

Simplicial Homology

Properties:

- ◆ For $k < 0$ or $k > \dim(K)$, $C_k(K)$ is the **null group**
- ◆ For $k \leq 0$ or $k > \dim(K)$, ∂_k is the **null map**
- ◆ For any $k \in \mathbb{Z}$, $\partial_k \circ \partial_{k+1} = 0$
- ◆ For any $k \in \mathbb{Z}$, $Im(\partial_{k+1}) \subseteq Ker(\partial_k)$

Simplicial Homology



Definition:

A k -chain c is called:

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

Each k -boundary is a k -cycle

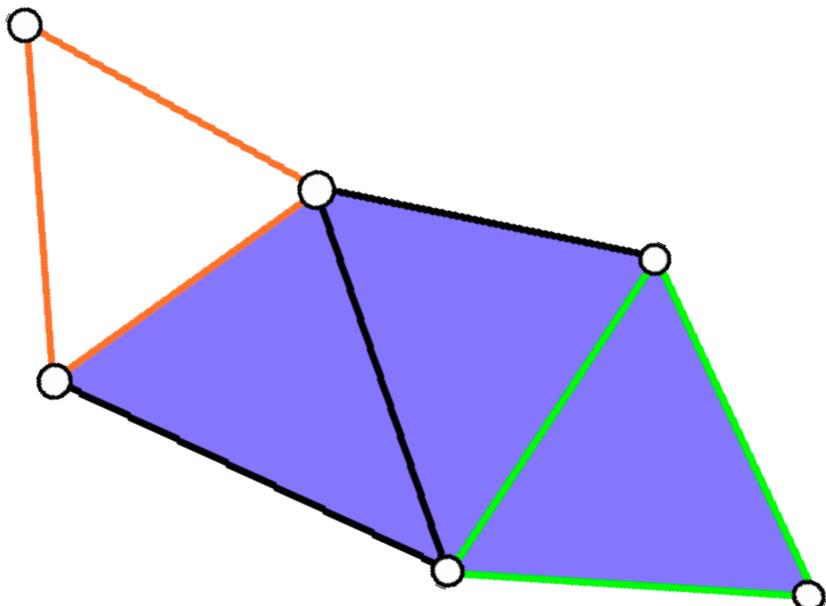
Simplicial Homology

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

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

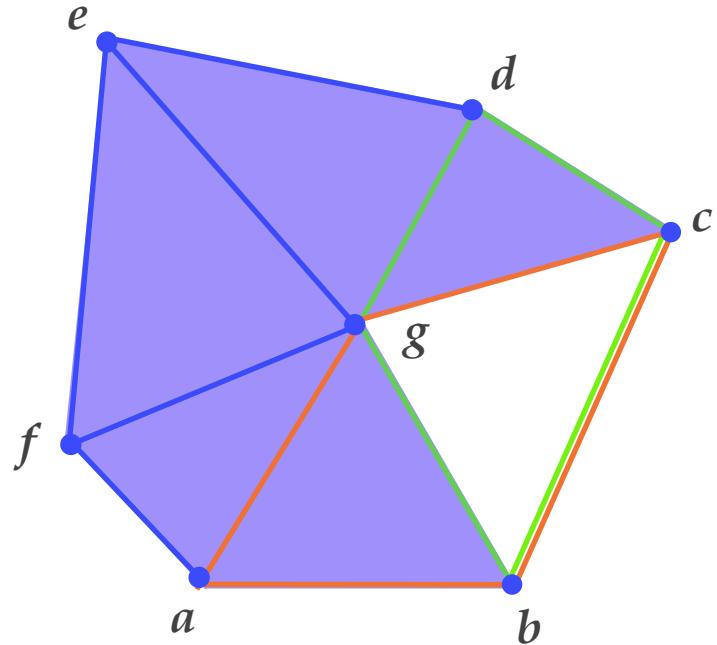
where:

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



Simplicial Homology

$H_k(K)$ partitions the k -cycles into equivalence classes called *homology classes*



Definition:

Two k -cycles are said *homologous* if they belong to the same homology class or, equivalently, *if their difference is a k -boundary*

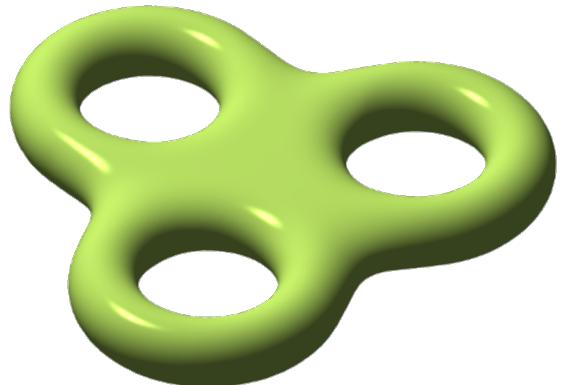
$ab+ag+bc+cg$ is homologous to $bc+bg+cd+dg$

Simplicial Homology

Theorem:

Each homology group can be expressed as

$$H_k(K) \cong (\mathbb{Z}_2)^{\beta_k}$$



$$H_k(K) \cong \begin{cases} \mathbb{Z}_2 & \text{for } k = 0 \\ (\mathbb{Z}_2)^6 & \text{for } k = 1 \\ \mathbb{Z}_2 & \text{for } k = 2 \end{cases}$$

β_k is called the *kth Betti number* of K

Simplicial Homology

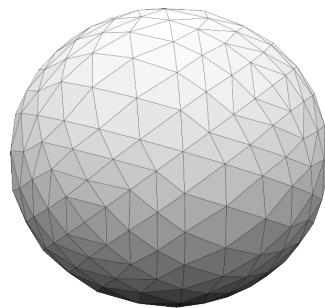
Examples:

◆ **point P**



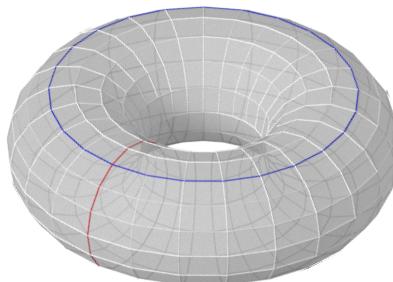
$$\beta_k(P) = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k > 0 \end{cases}$$

◆ **n -dimensional sphere S^n**



$$\beta_k(S^n) = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } 0 < k < n \\ 1 & \text{for } k = n \\ 0 & \text{for } k > n \end{cases}$$

◆ **torus T**



$$\beta_k(T) = \begin{cases} 1 & \text{for } k = 0 \\ 2 & \text{for } k = 1 \\ 1 & \text{for } k = 2 \\ 0 & \text{for } k > 2 \end{cases}$$

Simplicial Homology

Homology groups can be defined ***in a more general way*** by choosing coefficients in \mathbb{Z}

Theorem:

Each homology group can be expressed as

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

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

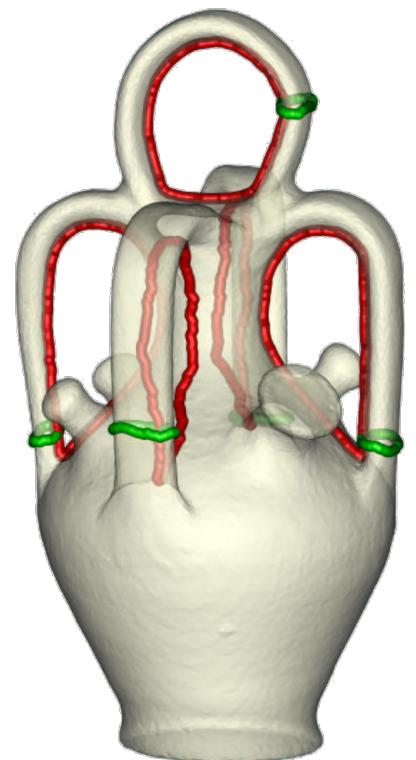


Image from [Dey et al. 2008]

Simplicial Homology

Working with coefficients in \mathbb{Z} :

*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*

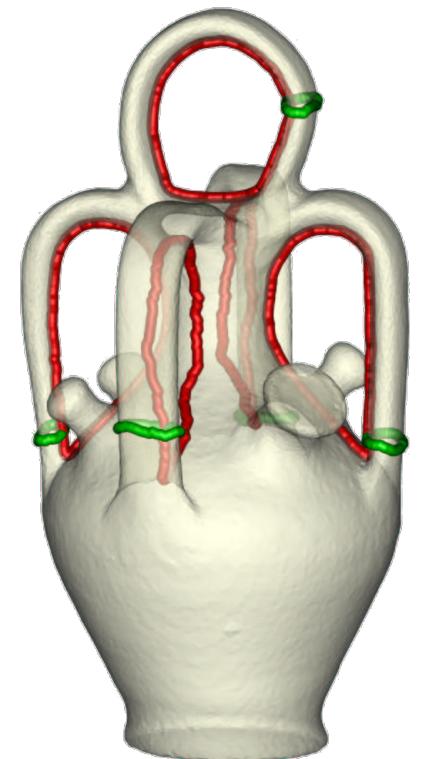
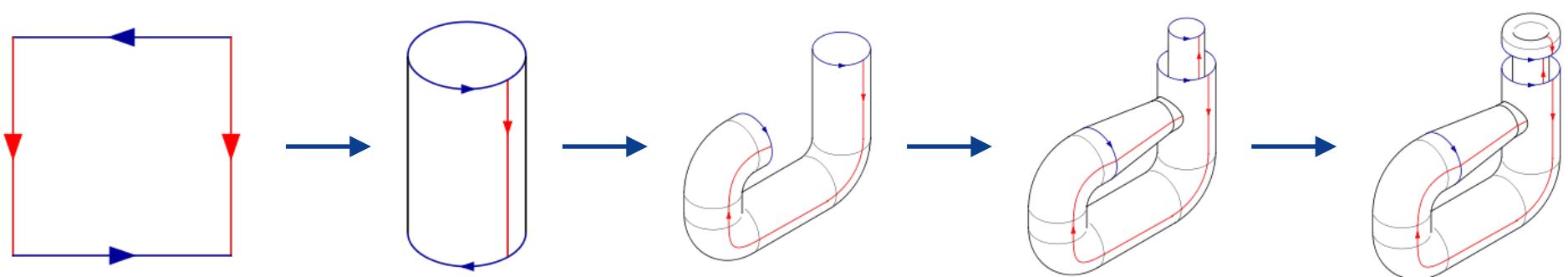
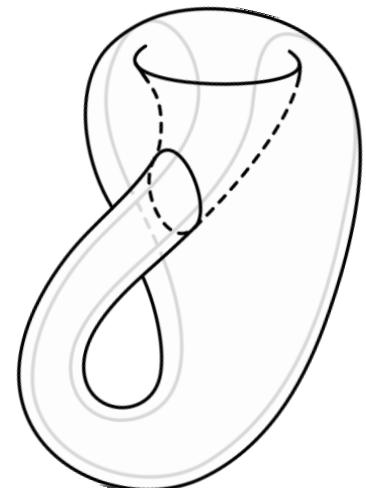


Image from [Dey et al. 2008]

Simplicial 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



Simplicial Homology

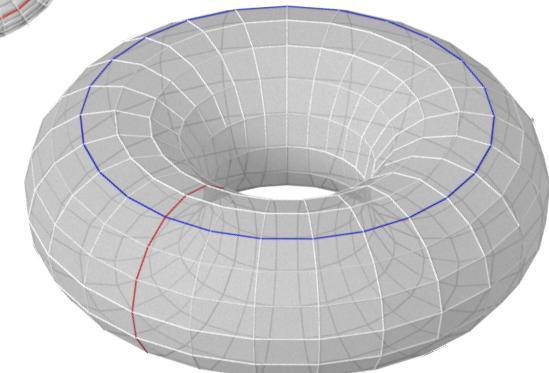
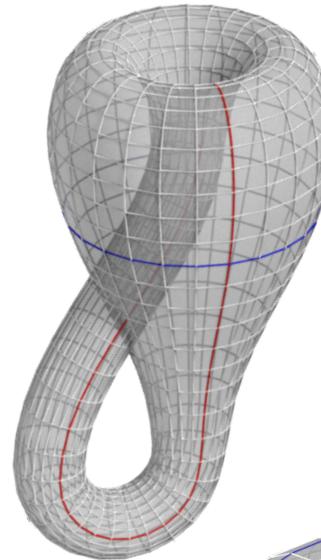
Example:

K has the following homology groups

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

So, it can be distinguished from a torus T

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

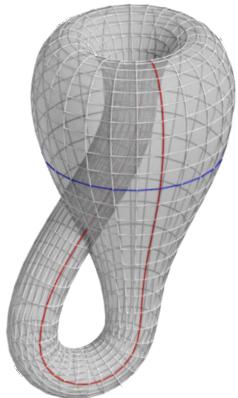


Simplicial Homology

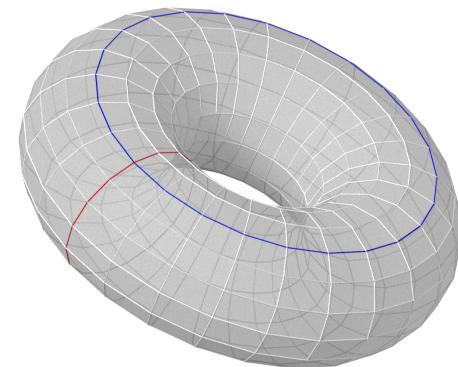
Example:

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

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



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

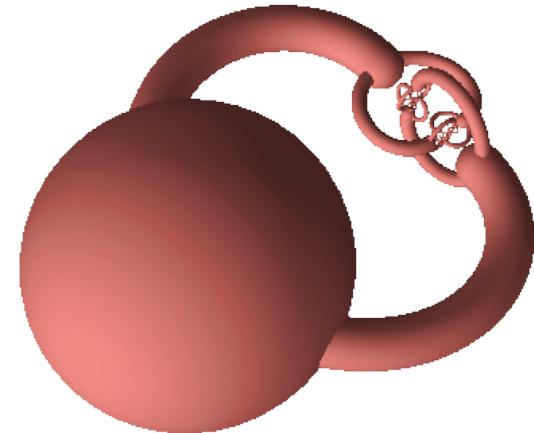
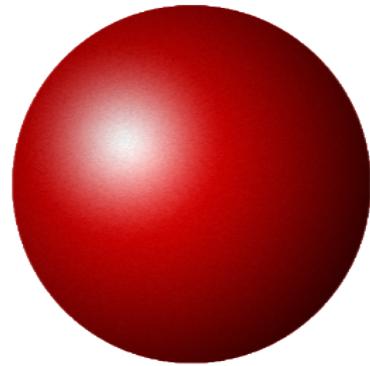


Homology & Persistence

- ◆ *Chain Complexes and Simplicial Homology*
- ◆ ***Filtrations and Persistent Homology***

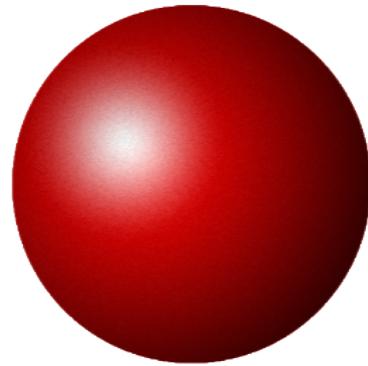
Persistent Homology

◆ *Do they have the same shape?*

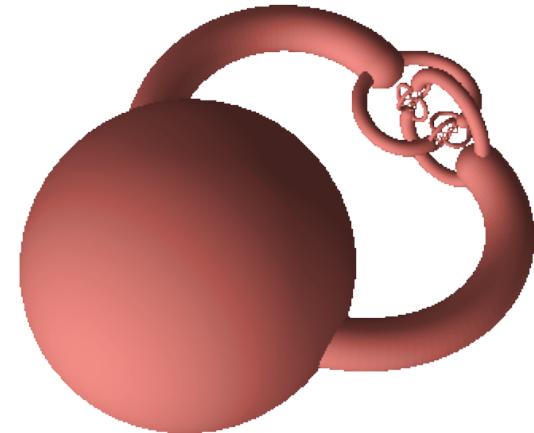


Persistent Homology

◆ *Do they have the same shape?*



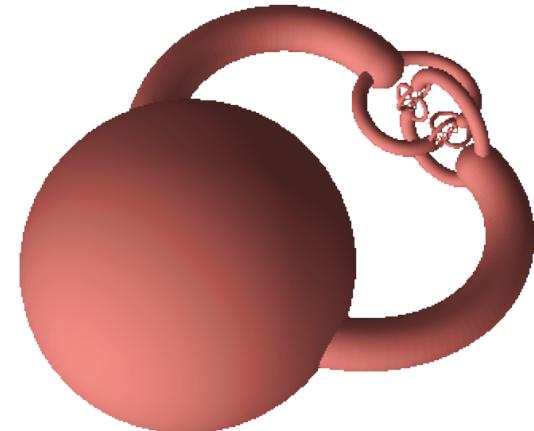
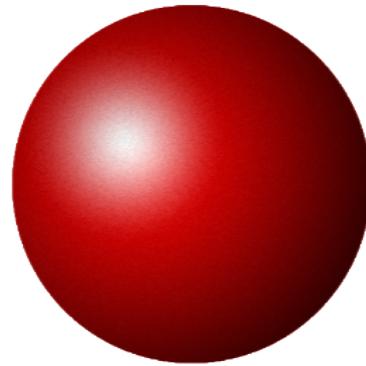
In Practice?



In Theory?

Persistent Homology

◆ *Do they have the same shape?*



In Practice?



In Theory?



*They are **homeomorphic***

Persistent Homology

◆ *Do they have the same shape?*



Persistent Homology

◆ *Do they have the same shape?*



In Practice?

In Theory?

Persistent Homology

◆ *Do they have the same shape?*



In Practice?



In Theory?



They are not homeomorphic

Persistent Homology

In a Nutshell:

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

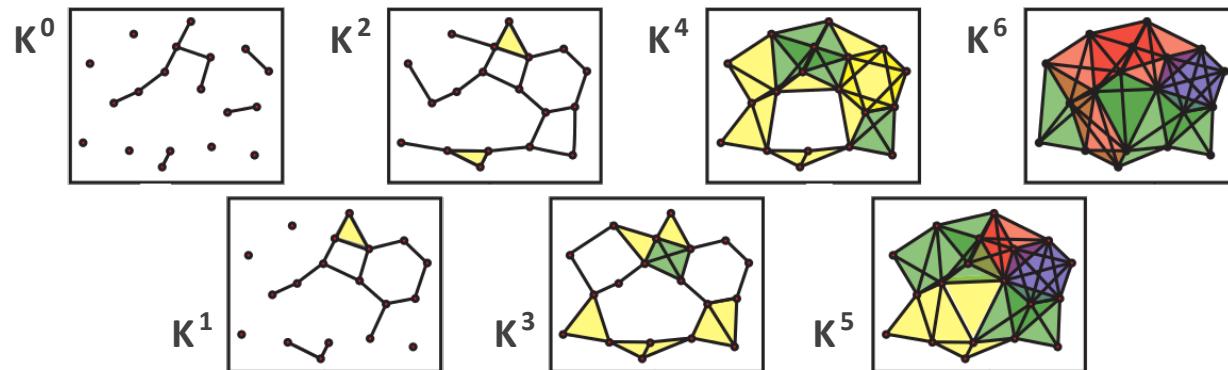


Image from [Ghrist 2008]

Persistent Homology

An Evolving Notion:

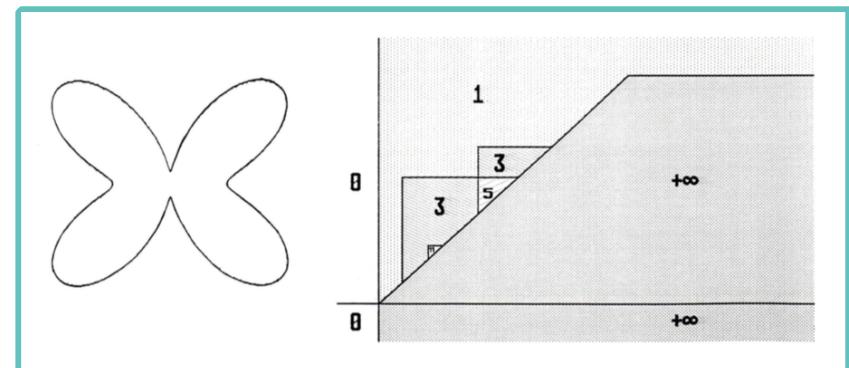
1990



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***

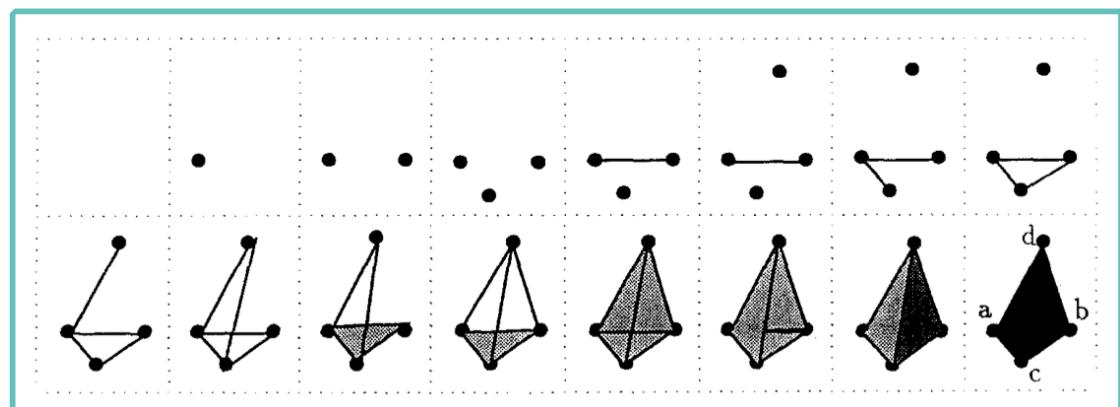


Image from [Delfinado, Edelsbrunner 1995]

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**

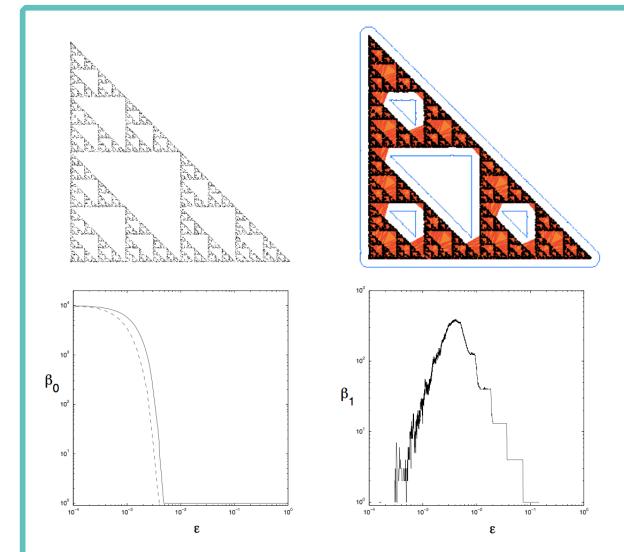


Image from [Robins 1999]

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

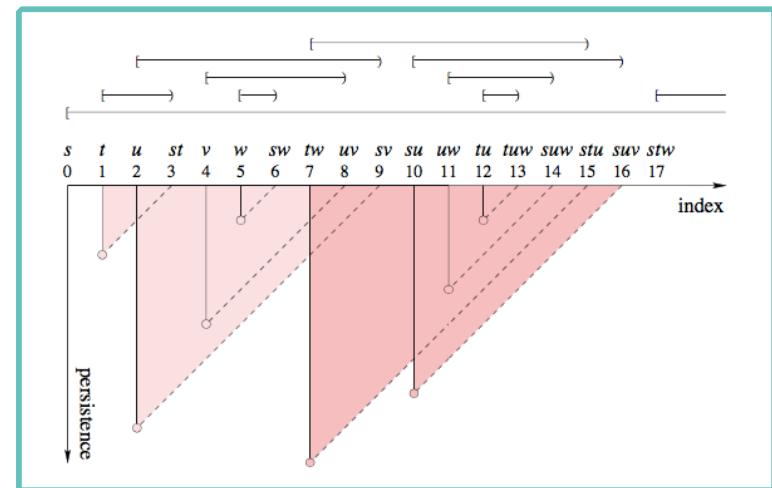


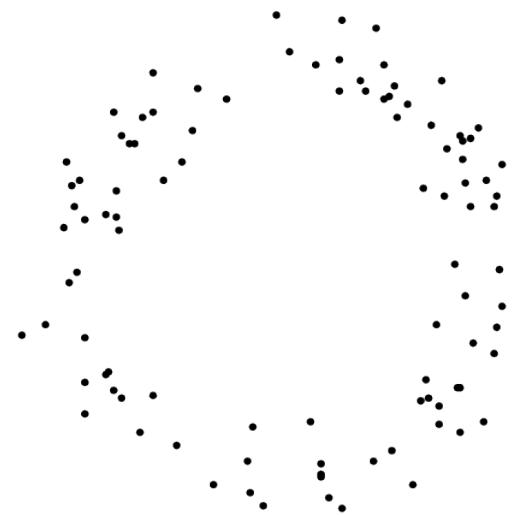
Image from [Edelsbrunner et al. 2002]

Persistent Homology

A Twofold Purpose:

Shape Description

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

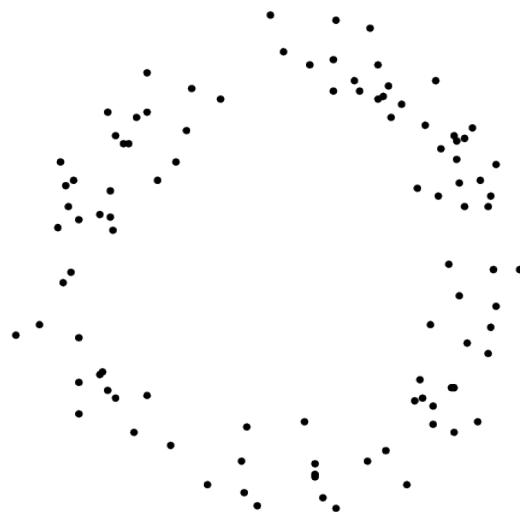
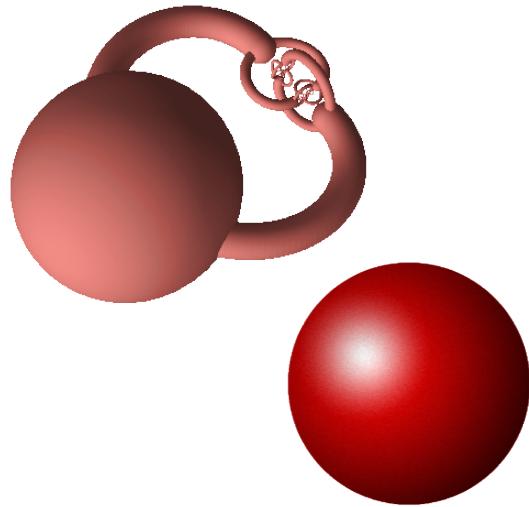


Persistent Homology

A Twofold Purpose:

Shape Description

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



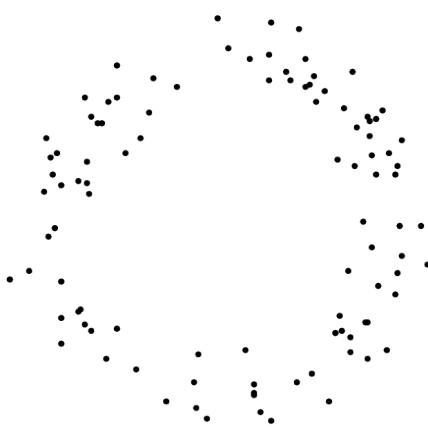
Shape Comparison

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

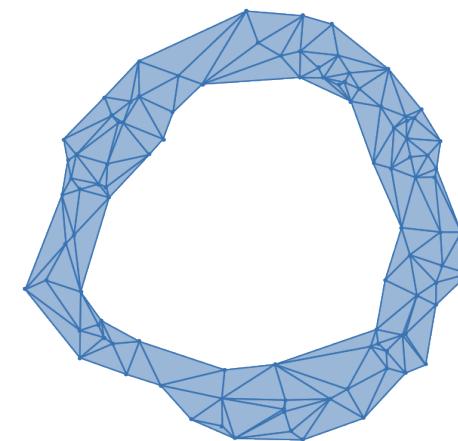
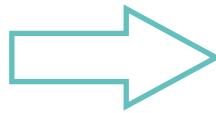
Persistent Homology

- ◆ *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



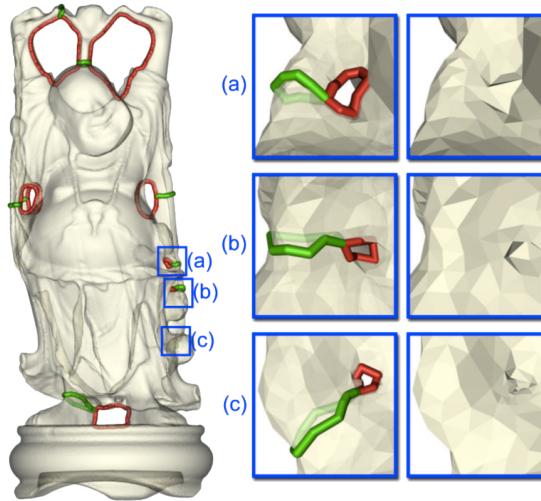
*Topological Nature of
the “Underlying” Shape*

Image from [Bauer 2015]

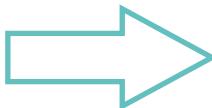
Persistent Homology

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

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



Noisy Dataset



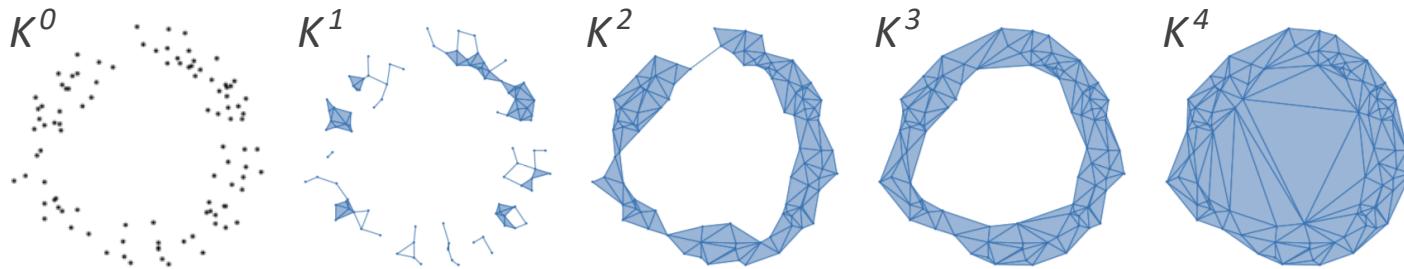
Relevant Homological Information

Image from [Dey et al. 2008]

Persistent Homology

Definition:

A ***filtration*** \mathcal{F} is a finite “growing” sequence of simplicial complexes



$$K^0 \subseteq K^1 \subseteq \dots \subseteq K^m$$

Most of the techniques transforming a dataset into a simplicial complex depending on the choice of a parameter actually produce a filtration

Persistent Homology

Definitions:

Given a filtration \mathcal{F} and $p, q \in \mathbb{N}$ such that $0 \leq p \leq q \leq m$,

- the inclusion $K^p \subseteq K^q$ induces a *linear map* $i_k^{p,q} : H_k(K^p) \rightarrow H_k(K^q)$
- the *(p,q) -persistent k -homology group* $H_k^{p,q}(\mathcal{F})$ of \mathcal{F} is defined as

$$H_k^{p,q}(\mathcal{F}) := \text{Im}(i_k^{p,q})$$

and consists of the *k -cycles of K^p that will not turn into k -boundaries in K^q*

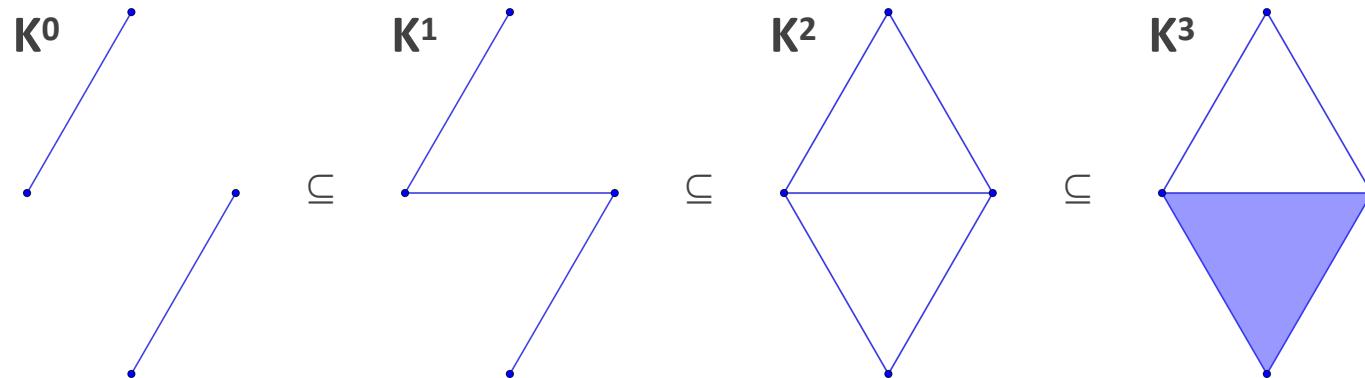
(in other terms, it identifies the *homology classes that “persist” from K^p to K^q*)

- the *(p,q) -persistent k^{th} Betti number* $\beta_k^{p,q}$ of \mathcal{F} is defined as the *rank of $i_k^{p,q}$*

Persistent Homology

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

Given a filtration \mathcal{F} : $K^0 \subseteq K^1 \subseteq \dots \subseteq K^m$,

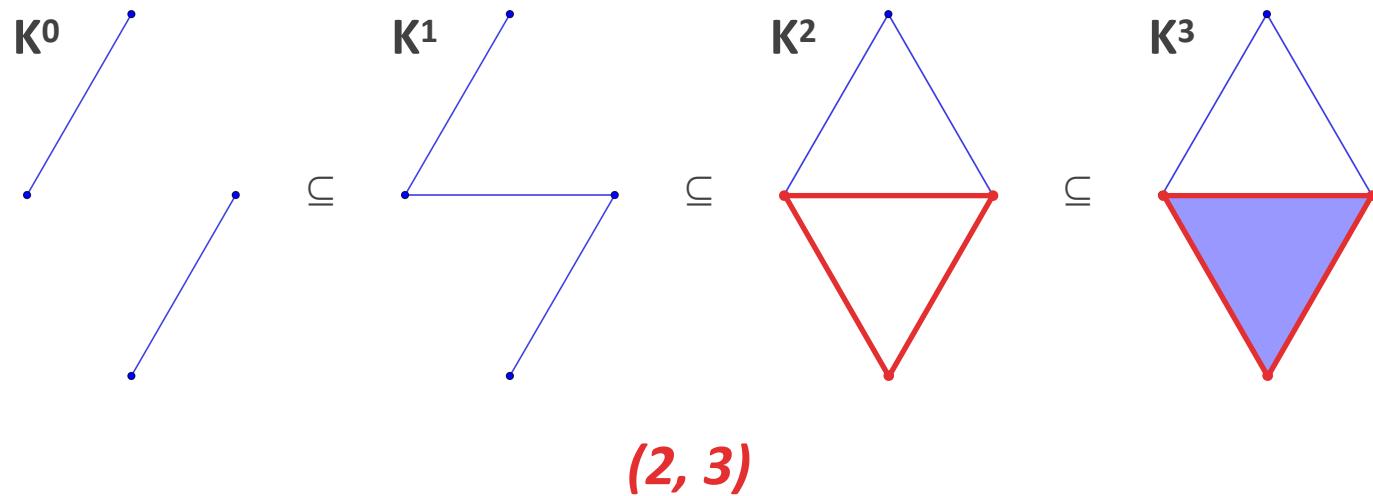


A *persistence pair* (p, q) is an element in $\{0, \dots, m\} \times (\{0, \dots, m\} \cup \{\infty\})$ such that $p < q$ representing a **homological class** that is **born at step p** and **dies at step q**

Persistent Homology

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

Given a filtration \mathcal{F} : $K^0 \subseteq K^1 \subseteq \dots \subseteq K^m$,

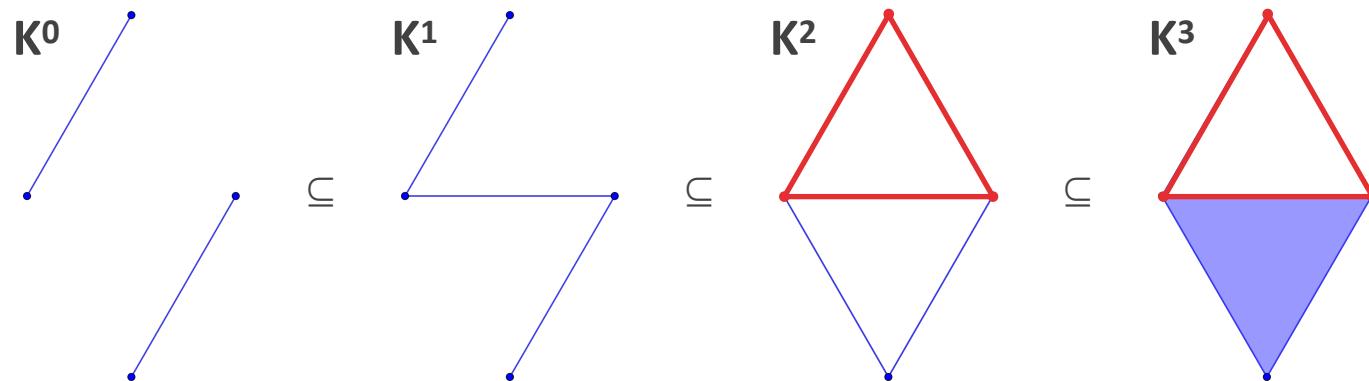


A *persistence pair* (p, q) is an element in $\{0, \dots, m\} \times (\{0, \dots, m\} \cup \{\infty\})$ such that $p < q$ representing a **homological class** that is **born at step p** and **dies at step q**

Persistent Homology

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

Given a filtration \mathcal{F} : $K^0 \subseteq K^1 \subseteq \dots \subseteq K^m$,



(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**

Persistent Homology

Given a filtration \mathcal{F} : $K^0 \subseteq K^1 \subseteq \dots \subseteq K^m$, $k \in \mathbb{N}$, and a field \mathbb{F} ,

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

The corresponding structure theorem ensures us that

Theorem:

The persistence module M can be expressed as

$$M \cong \bigoplus_{j=1}^N \mathbb{F}[x](-p'_j) \oplus \bigoplus_{i=1}^M \mathbb{F}[x](-p_i)/(x^{q_i})$$

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

i.e., the collection of the pairs $(p_i, q_i), (p'_j, \infty)$

Bibliography

General References:

- ◆ **Books on TDA:**
 - ❖ A. J. Zomorodian. *Topology for computing*. Cambridge University Press, 2005.
 - ❖ H. Edelsbrunner, J. Harer. *Computational topology: an introduction*. American Mathematical Society, 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:

- ◆ **Simplicial Homology:**
 - ❖ J. R. Munkres. *Elements of algebraic topology*. CRC Press, 1984.
- ◆ **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.