FGV EMAp — Seminário — 22/04/21

Topological inference in Topological Data Analysis

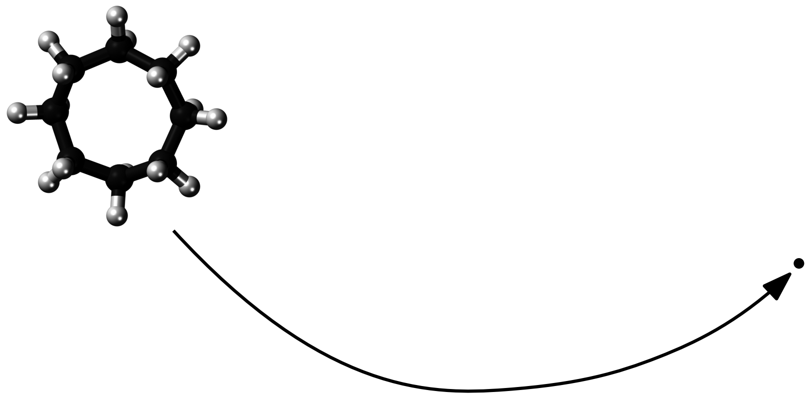
Talk I (/II): Topology in datasets

https://raphaeltinarrage.github.io

Last update: April 22, 2021

[S. Martin, A. Thompson, E. A. Coutsias, and J-P. Watson, Topology of cyclo-octane energy landscape, 2010]

The cyclo-octane molecule $\mathrm{C}_8\mathrm{H}_{16}$ contains 24 atoms.
Each atom has 3 spatial coordinates.
Hence a conformation of a molecule can be summarized by **a point** in $\mathbb{R}^{72}$ $(3 \times 24 = 72)$.

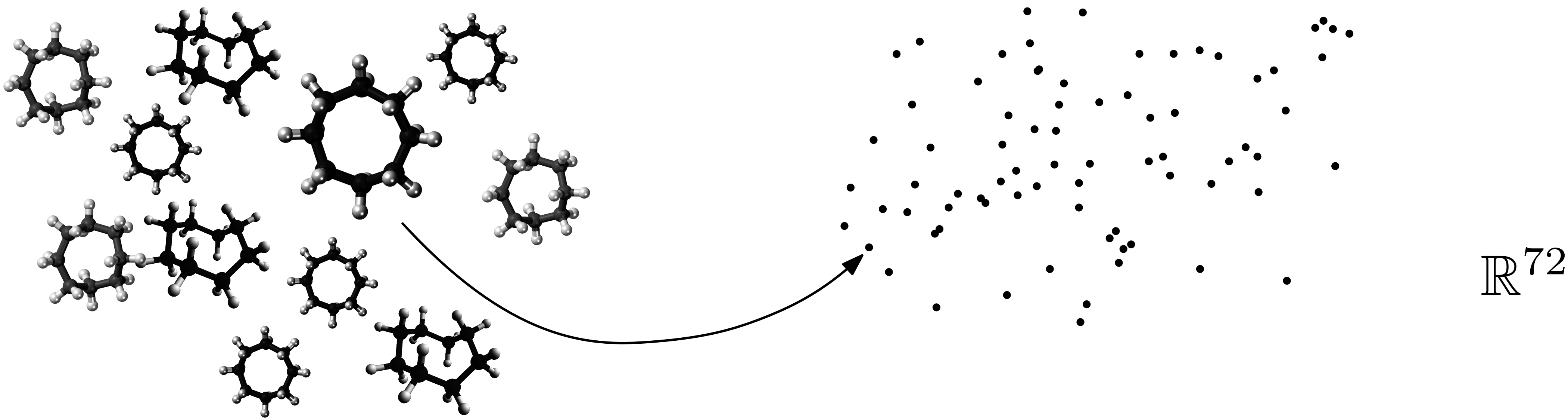


$\mathbb{R}^{72}$

Aperitivo topológico

[S. Martin, A. Thompson, E. A. Coutsias, and J-P. Watson, Topology of cyclo-octane energy landscape, 2010]

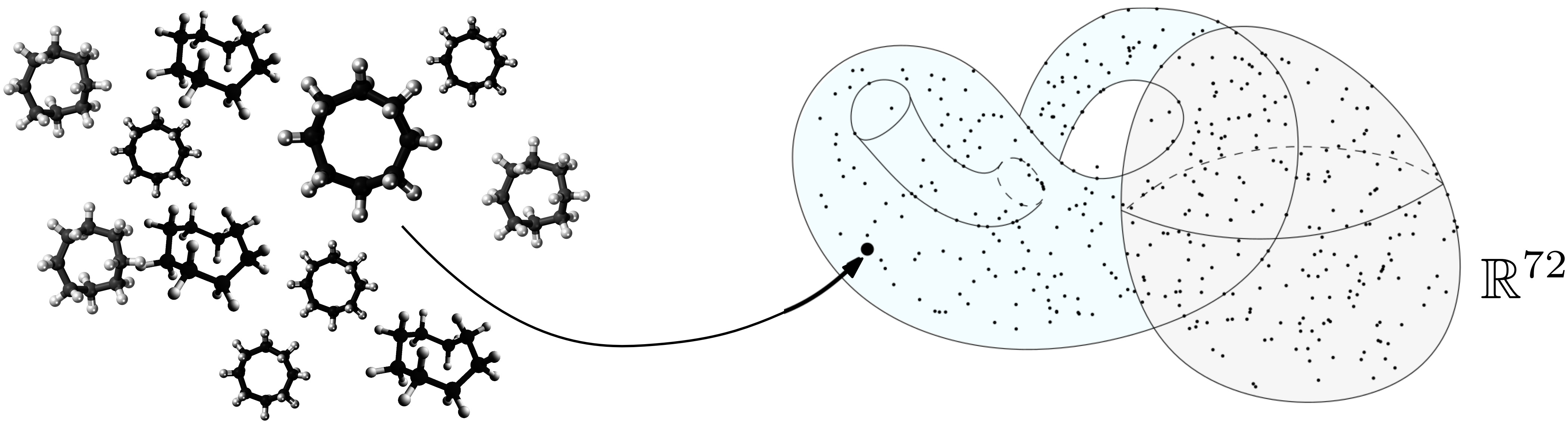
The cyclo-octane molecule $\mathrm{C_8H_{16}}$ contains 24 atoms.
Each atom has 3 spatial coordinates.
Hence a conformation of a molecule can be summarized by **a point** in $\mathbb{R}^{72}$ ($3 \times 24 = 72$).



By considering a lot of such molecules, we obtain a **point cloud** in $\mathbb{R}^{72}$.

[S. Martin, A. Thompson, E. A. Coutsias, and J-P. Watson, Topology of cyclo-octane energy landscape, 2010]

The cyclo-octane molecule $\mathrm{C_8H_{16}}$ contains 24 atoms.
Each atom has 3 spatial coordinates.
Hence a conformation of a molecule can be summarized by **a point** in $\mathbb{R}^{72}$ $(3 \times 24 = 72)$.

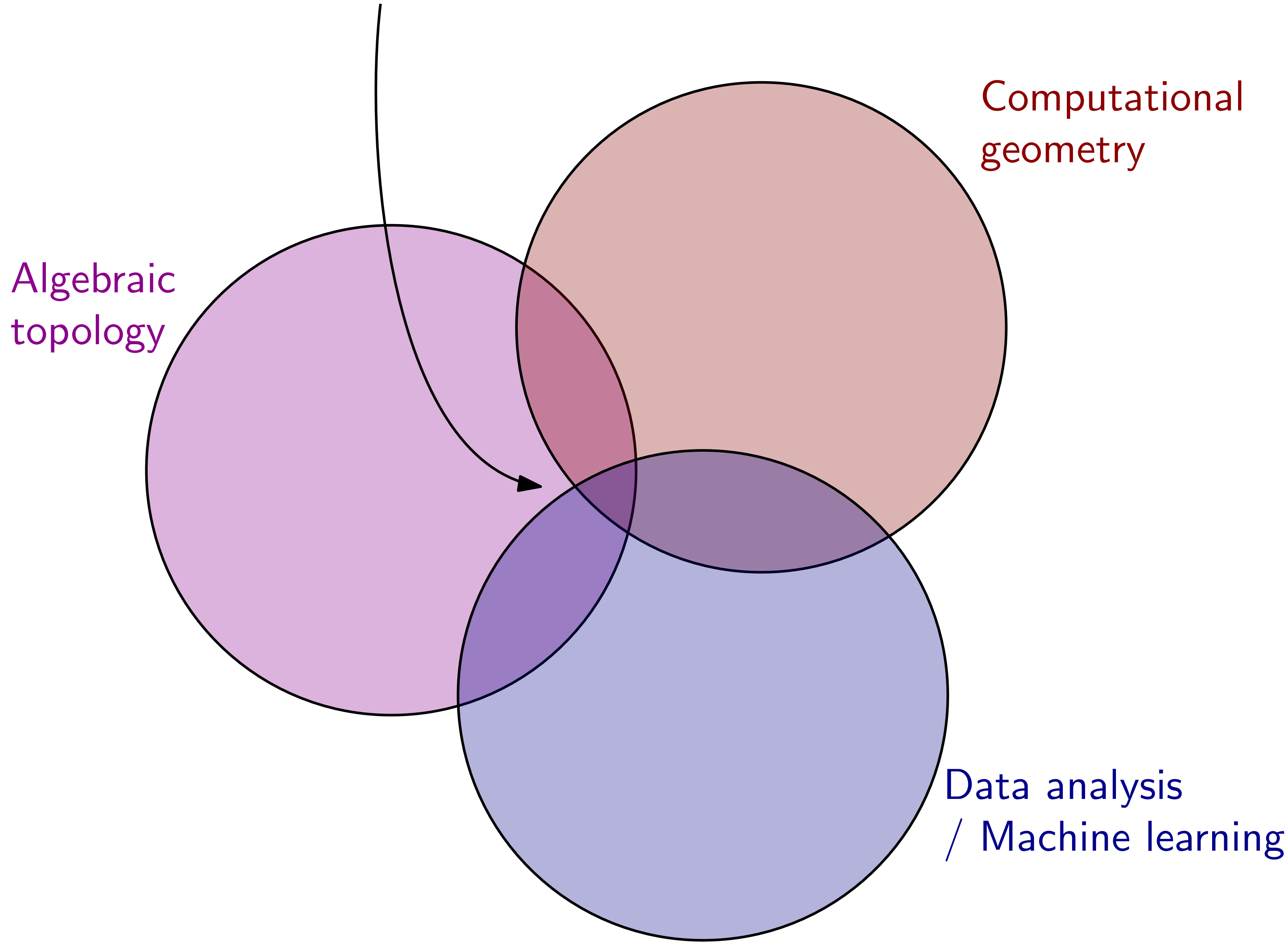


By considering a lot of such molecules, we obtain a **point cloud** in $\mathbb{R}^{72}$.

The authors show that this point cloud lies close to a small dimensional object: **the union of a sphere and a Klein bottle**.

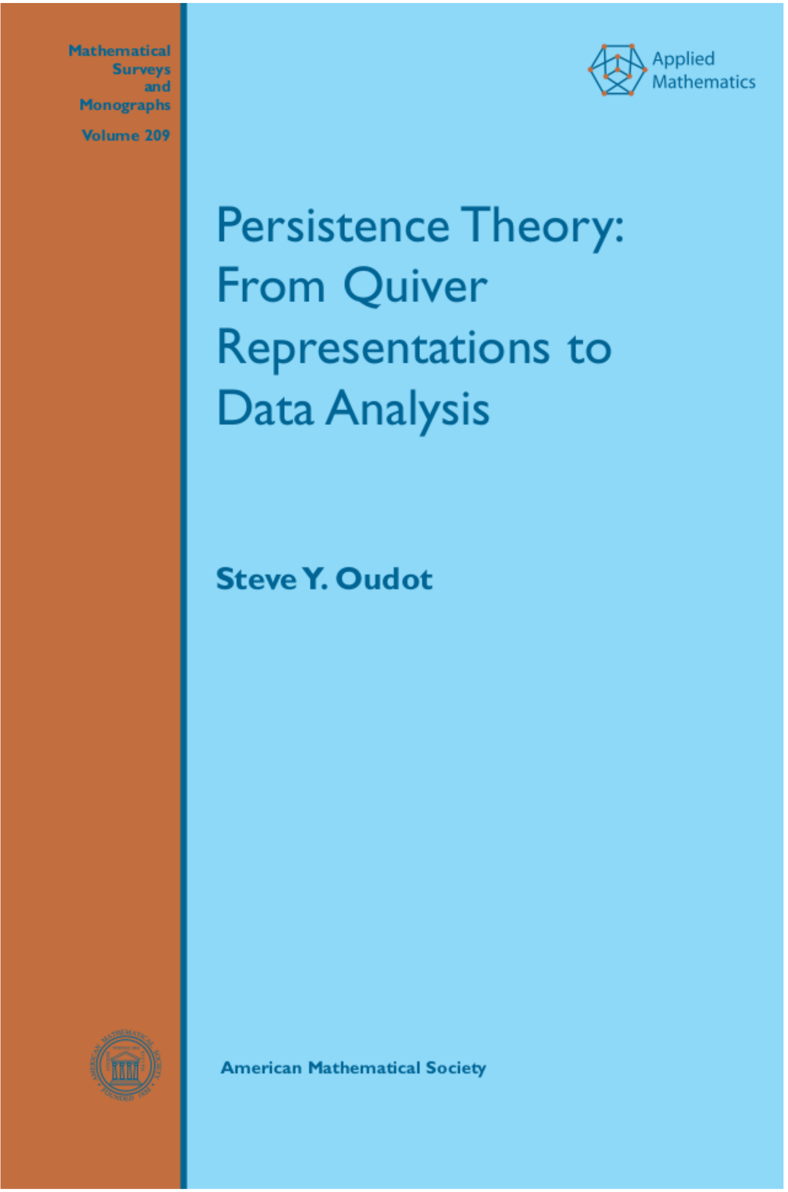
Introduction

Topological Data Analysis (TDA) allows to explore and understand the topology of datasets.



Introduction

Steve Oudot in 2015



Applications. This richness is also reflected in the diversity of the applications, whose list has been ever growing since the early developments of the theory. The following excerpt[5] illustrates the variety of the topics addressed:

- analysis of random, modular and non-modular scale-free networks and networks with exponential connectivity distribution [158],
- analysis of social and spatial networks, including neurons, genes, online messages, air passengers, Twitter, face-to-face contact, co-authorship [210],
- coverage and hole detection in wireless sensor fields [98, 136],
- multiple hypothesis tracking on urban vehicular data [23],
- analysis of the statistics of high-contrast image patches [54],
- image segmentation [70, 209],
- 1d signal denoising [212],
- 3d shape classification [58],
- clustering of protein conformations [70],
- measurement of protein compressibility [135],
- classification of hepatic lesions [1],
- identification of breast cancer subtypes [205],
- analysis of activity patterns in the primary visual cortex [224],
- discrimination of electroencephalogram signals recorded before and during epileptic seizures [237],
- analysis of 2d cortical thickness data [82],
- statistical analysis of orthodontic data [134, 155],
- measurement of structural changes during lipid vesicle fusion [169],
- characterization of the frequency and scale of lateral gene transfer in pathogenic bacteria [125],
- pattern detection in gene expression data [105],
- study of plant root systems [115, §IX.4],
- study of the cosmic web and its filamentary structure [226, 227],
- analysis of force networks in granular matter [171],
- analysis of regimes in dynamical systems [25].

In most of these applications, the use of persistence resulted in the definition of new descriptors for the considered data, which revealed previously hidden structural information and allowed the authors to draw original conclusions.

I - Comparing topological spaces

1 - Homeomorphic equivalence

2 - Homotopy equivalence

II - Topological invariants

1 - Number of connected components

2 - Euler characteristic

3 - Betti numbers

(Next week - Persistent homology)

Protagonists in topology

In topology we study **topological spaces**.

Definition: a topological space is a set X endowed with a collection of **open sets** $\{O_\alpha, \alpha \in A\}$, with $O_\alpha \subset X$, such that

- $\emptyset$ and X are open sets,
- an infinite union of open sets is an open set,
- a finite intersection of open sets is an open set.

Definition: Given two topological spaces X and Y, a map $f\colon X \to Y$ is **continuous** if for every open set $O \subset Y$, the preimage $f^{-1}(O)$ is an open set of X.

$$X \longrightarrow Y$$

Protagonists in topology

In topology we study **topological spaces**.

Definition: a topological space is a set X endowed with a collection of **open sets** $\{O_\alpha, \alpha \in A\}$, with $O_\alpha \subset X$, such that

- $\emptyset$ and X are open sets,
- an infinite union of open sets is an open set,
- a finite intersection of open sets is an open set.

Definition: Given two topological spaces X and Y, a map $f\colon X \to Y$ is **continuous** if for every open set $O \subset Y$, the preimage $f^{-1}(O)$ is an open set of X.

$$X \longrightarrow Y$$

translation in ϵ-δ calculus

One can think of **subsets** $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$,

and maps $f\colon X \to Y$ **continuous** in the following sense:

$$\forall x \in X, \forall \epsilon > 0, \exists \eta > 0, \forall y \in X, \|x - y\| < \eta \implies \|f(x) - f(y)\| < \epsilon.$$

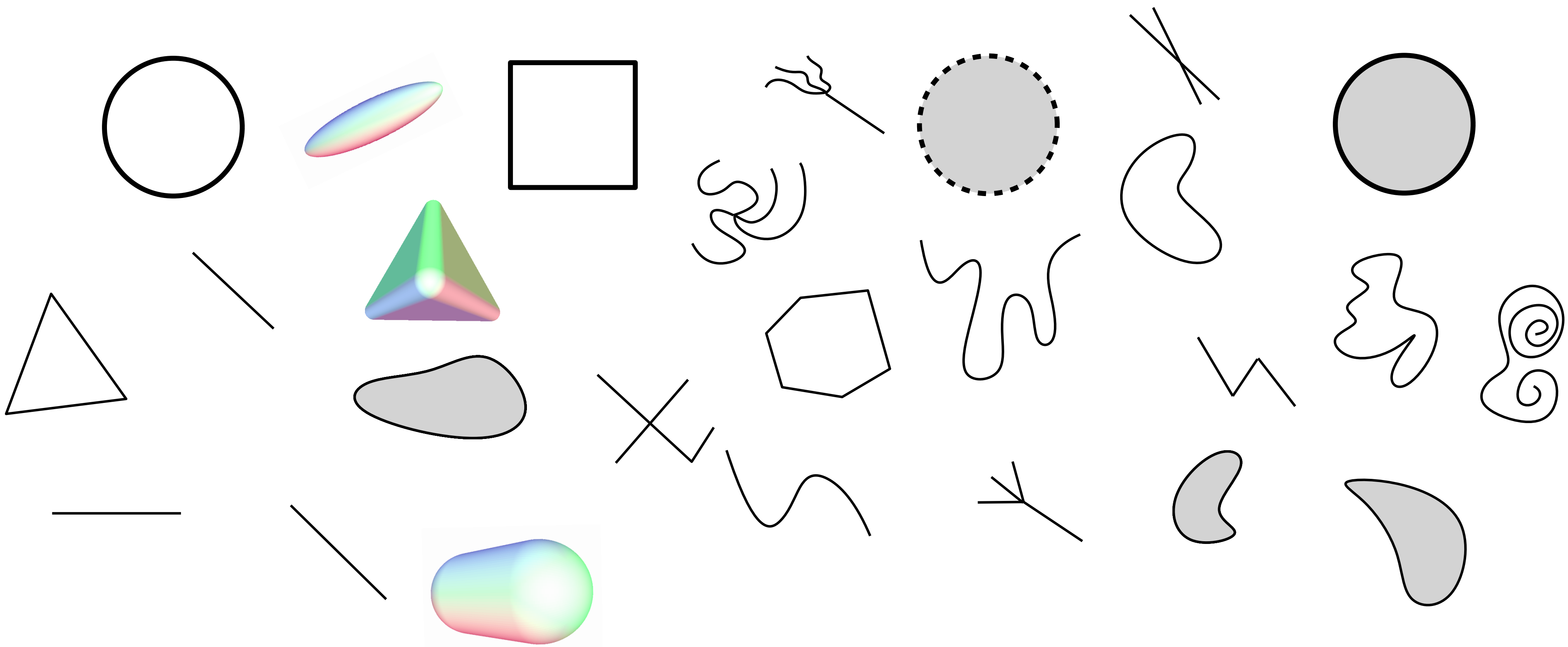
Examples of topological spaces

In $\mathbb{R}^n$, we can define:

- the unit sphere $\mathbb{S}_{n-1} = \{x \in \mathbb{R}^n, \|x\| = 1\}$
- the unit cube $\mathcal{C}_{n-1} = \{x = (x_1, ..., x_n) \in \mathbb{R}^n, \max(|x_1|, ..., |x_n|) = 1\}$
- the open balls $\mathcal{B}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$
- the closed balls $\overline{\mathcal{B}}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| \leq r\}$

In $\mathbb{R}^n$, we can define:

- the unit sphere $\mathbb{S}_{n-1} = \{x \in \mathbb{R}^n, \|x\| = 1\}$
- the unit cube $\mathcal{C}_{n-1} = \{x = (x_1, ..., x_n) \in \mathbb{R}^n, \max(|x_1|, ..., |x_n|) = 1\}$
- the open balls $\mathcal{B}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$
- the closed balls $\overline{\mathcal{B}}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| \leq r\}$

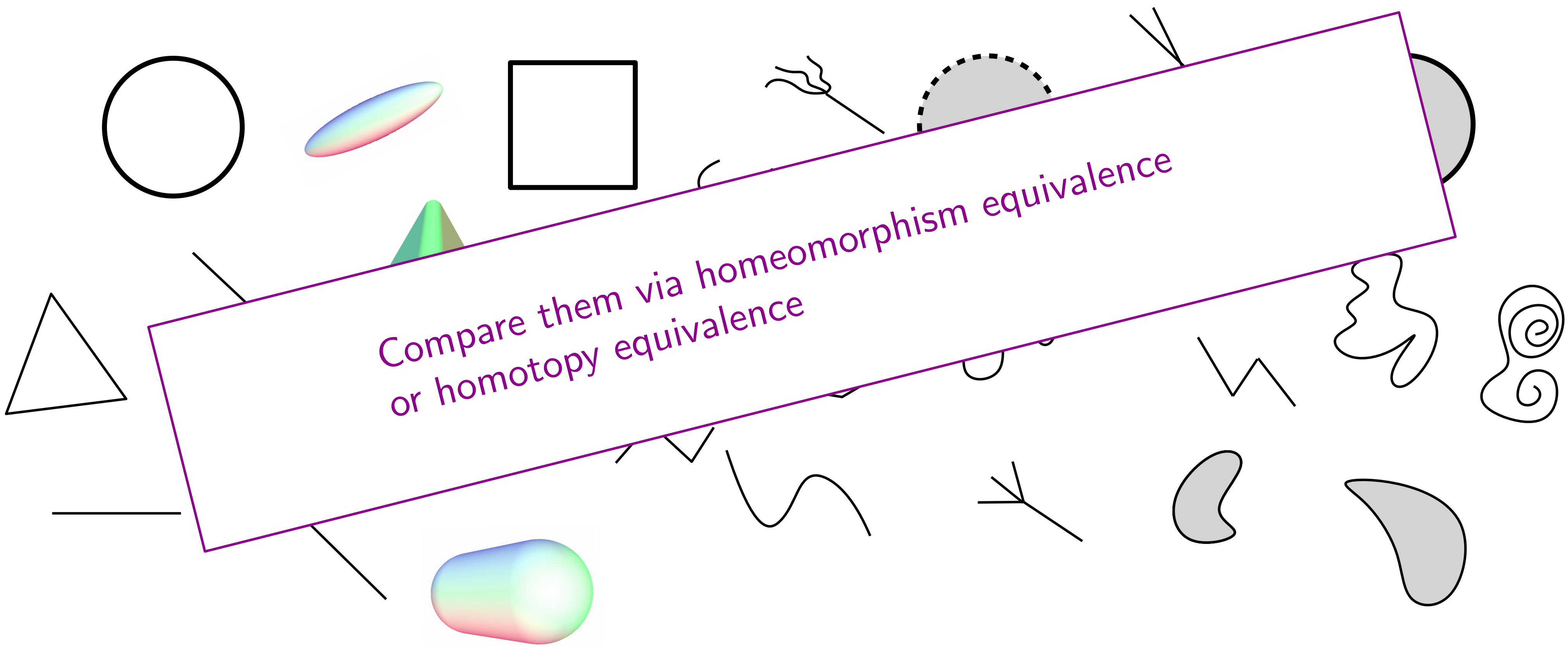


Most of the time, we do not have a nice algebraic definition...

Examples of topological spaces

In $\mathbb{R}^n$, we can define:

- the unit sphere $\mathbb{S}_{n-1} = \{x \in \mathbb{R}^n, \|x\| = 1\}$
- the unit cube $\mathcal{C}_{n-1} = \{x = (x_1, ..., x_n) \in \mathbb{R}^n, \max(|x_1|, ..., |x_n|) = 1\}$
- the open balls $\mathcal{B}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$
- the closed balls $\overline{\mathcal{B}}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| \leq r\}$



Most of the time, we do not have a nice algebraic definition...

I - Comparing topological spaces

1 - Homeomorphic equivalence

2 - Homotopy equivalence

II - Topological invariants

1 - Number of connected components

2 - Euler characteristic

3 - Betti numbers

(Next week - Persistent homology)

Homeomorphisms

Definition: Let X and Y be two topological spaces, and $f\colon X \to Y$ a map.
We say that f is a **homeomorphism** if

- f is a bijection,
- $f\colon X \to Y$ is continuous,
- $f^{-1}\colon Y \to X$ is continuous.

If there exist such a homeomorphism, we say that the two topological spaces are **homeomorphic**.

Homeomorphisms

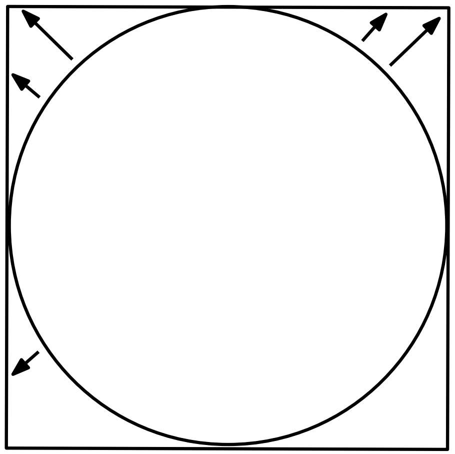
Definition: Let X and Y be two topological spaces, and $f\colon X \to Y$ a map.
We say that f is a **homeomorphism** if

- f is a bijection,
- $f\colon X \to Y$ is continuous,
- $f^{-1}\colon Y \to X$ is continuous.

If there exist such a homeomorphism, we say that the two topological spaces are **homeomorphic**.

Example: The unit circle and the unit square are homeomorphic via

$$\begin{aligned} f\colon \mathbb{S}_1 &\longrightarrow \mathcal{C} \\ (x_1, x_2) &\longmapsto \frac{1}{\max(|x_1|, |x_2|)}(x_1, x_2) \end{aligned}$$



Interpretation: Homeomorphisms allow 'continuous deformations'.

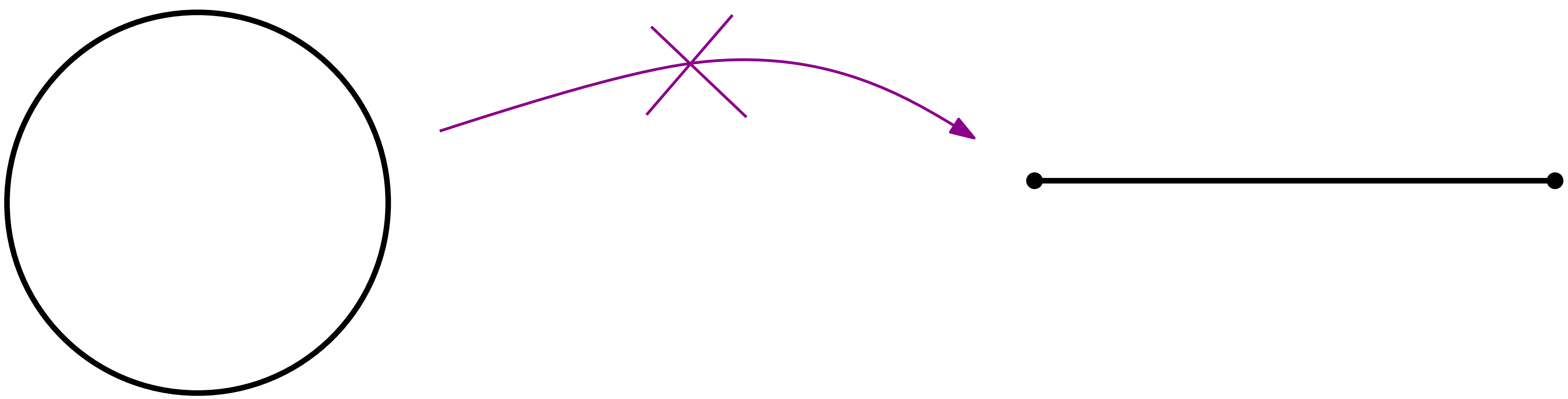
Homeomorphisms

Definition: Let X and Y be two topological spaces, and $f\colon X \to Y$ a map.
We say that f is a **homeomorphism** if

- f is a bijection,
- $f\colon X \to Y$ is continuous,
- $f^{-1}\colon Y \to X$ is continuous.

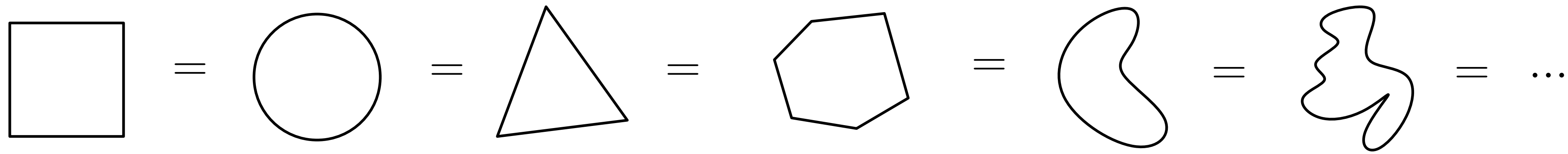
If there exist such a homeomorphism, we say that the two topological spaces are **homeomorphic**.

Example: The unit circle and the interval $[0, 1]$ are not homeomorphic.



Homeomorphism classes

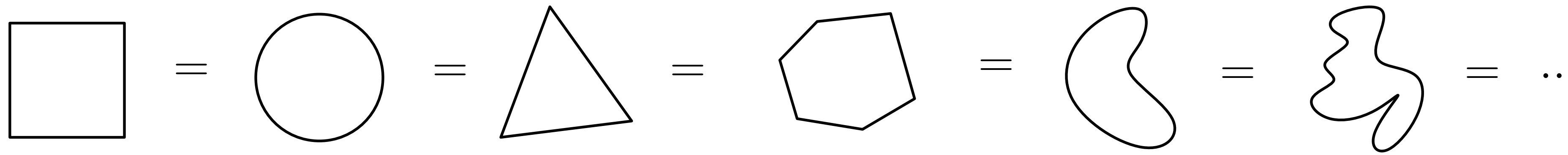
We can gather topological spaces that are homeomorphic



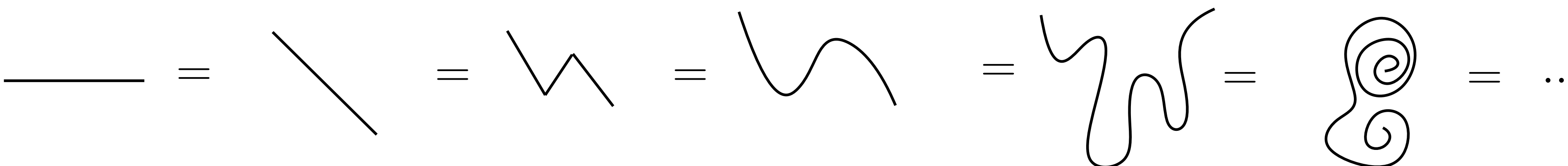
the class of circles

Homeomorphism classes

We can gather topological spaces that are homeomorphic



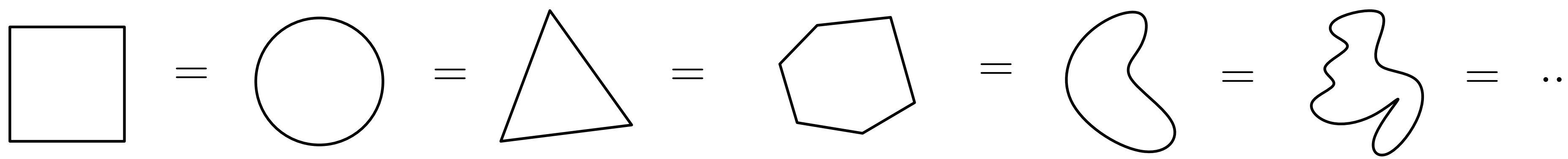
the class of circles



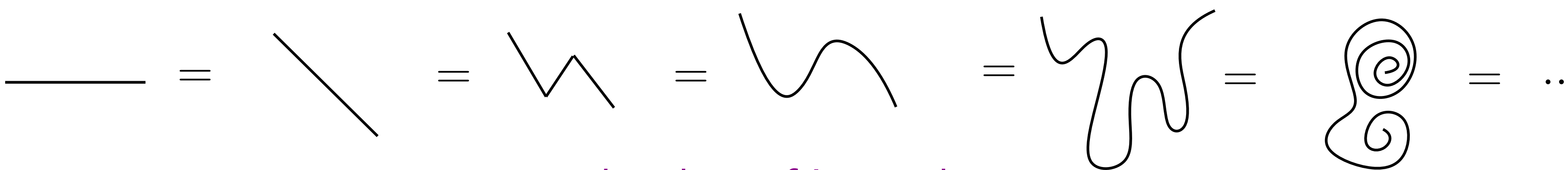
the class of intervals

Homeomorphism classes

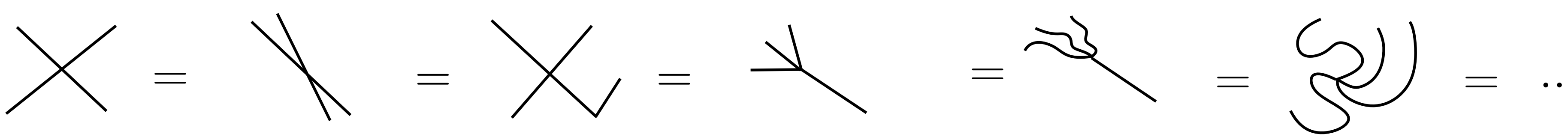
We can gather topological spaces that are homeomorphic



the class of circles



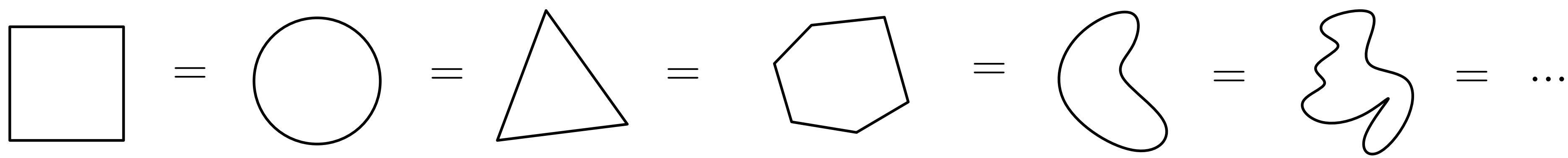
the class of intervals



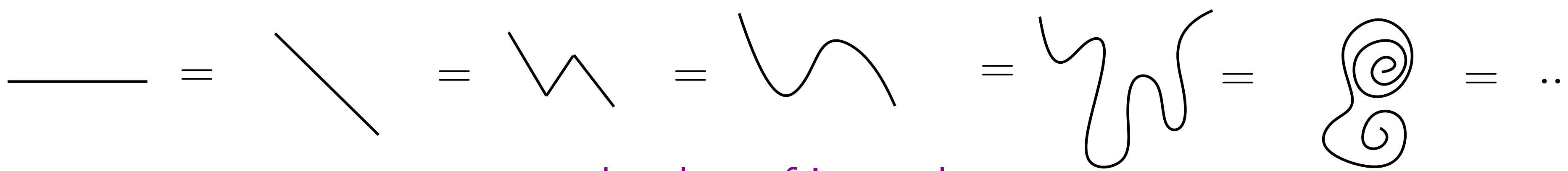
the class of crosses

Homeomorphism classes

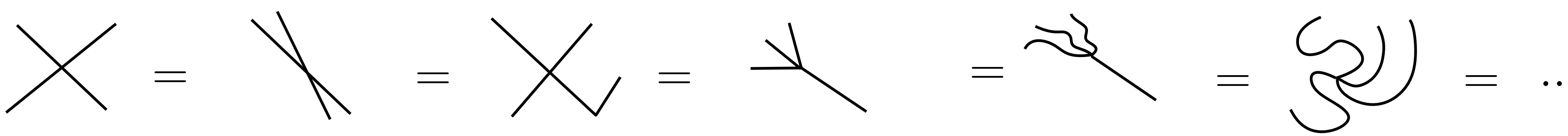
We can gather topological spaces that are homeomorphic



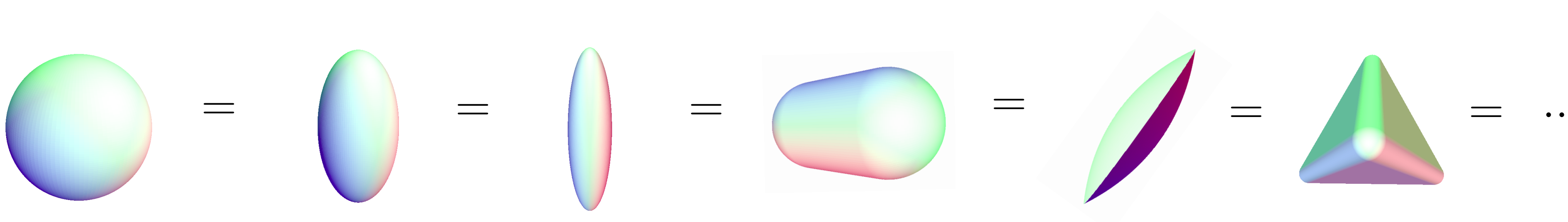
the class of circles



the class of intervals

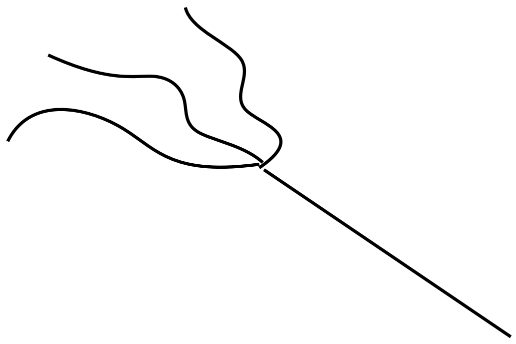


the class of crosses

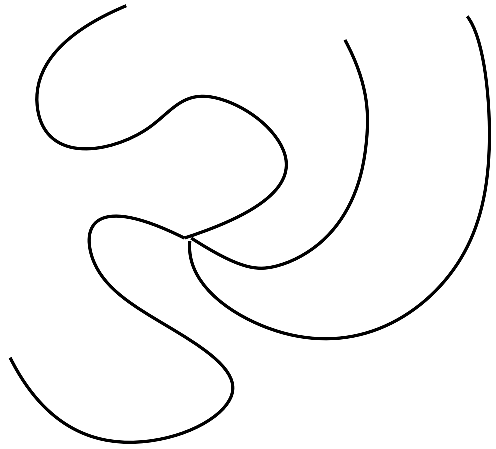


the class of spheres

In general, it may be complicated to determine whether two spaces are homeomorphic.



$$\stackrel{?}{=}$$



To answer this problem, we will use the notion of **invariant**.

I - Comparing topological spaces

1 - Homeomorphic equivalence

2 - Homotopy equivalence

II - Topological invariants

1 - Number of connected components

2 - Euler characteristic

3 - Betti numbers

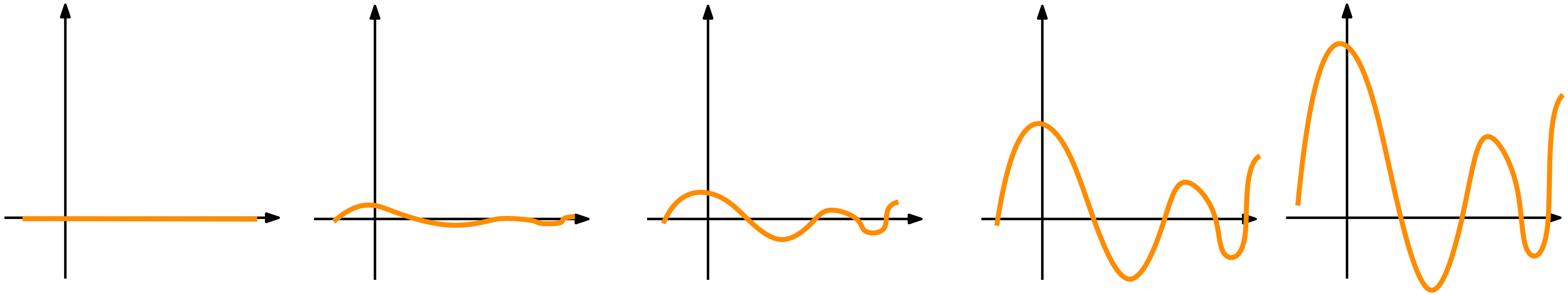
(Next week - Persistent homology)

Definition: Let X, Y be two topological spaces, and $f, g\colon X \to Y$ two continuous maps. A **homotopy** between f and g is a map $F\colon X \times [0,1] \to Y$ such that:

- $x \mapsto F(x,0)$ is equal to f,
- $x \mapsto F(x,1)$ is equal to g,
- $F\colon X \times [0,1] \to Y$ is continuous.

If such a homotopy exists, we say that the maps f and g are **homotopic**.

Example: Homotopy between $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$.

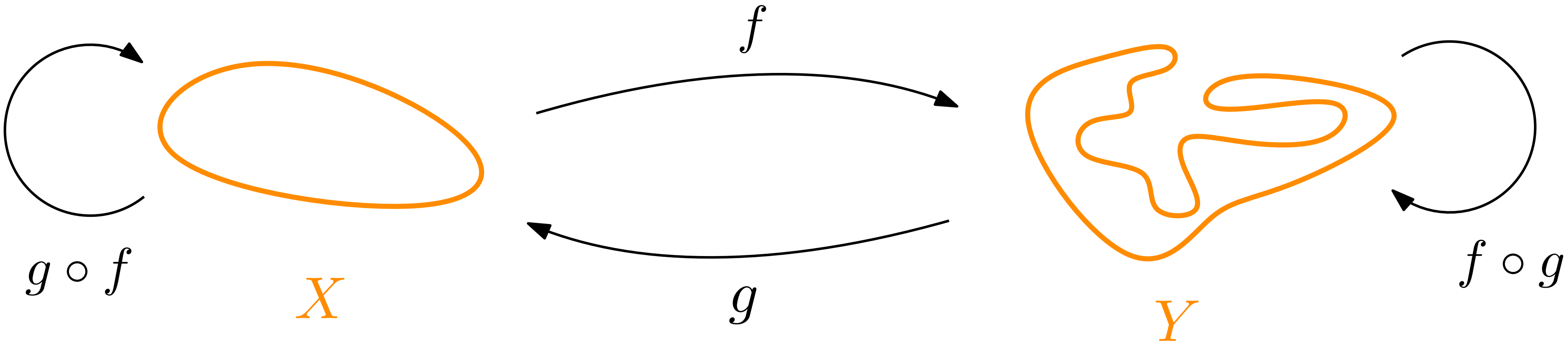


$F(\cdot,0) = f$ $\quad$ $F(\cdot,0.2)$ $\quad$ $F(\cdot,0.5)$ $\quad$ $F(\cdot,0.6)$ $\quad$ $F(\cdot,1) = g$

Defintion: Let X and Y be two topological spaces. A **homotopy equivalence** between X and Y is a pair of continuous maps $f\colon X \to Y$ and $g\colon Y \to X$ such that:

- $g \circ f\colon X \to X$ is homotopic to the identity map $\mathrm{id}\colon X \to X$,
- $f \circ g\colon Y \to Y$ is homotopic to the identity map $\mathrm{id}\colon Y \to Y$.

If such a homotopy equivalence exists, we say that X and Y are **homotopy equivalent**.

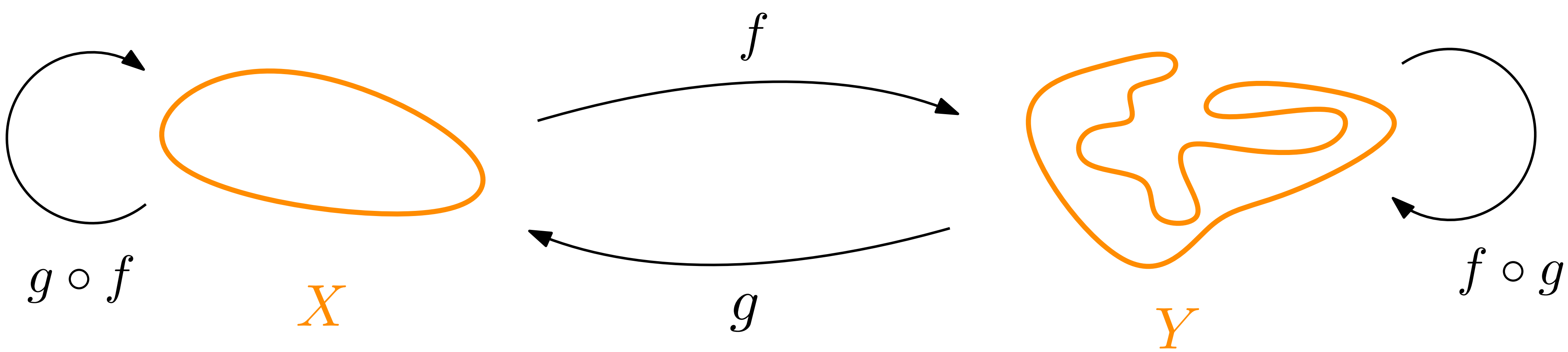


Homotopy equivalence

Defintion: Let X and Y be two topological spaces. A **homotopy equivalence** between X and Y is a pair of continuous maps $f\colon X \to Y$ and $g\colon Y \to X$ such that:

- $g \circ f\colon X \to X$ is homotopic to the identity map $\mathrm{id}\colon X \to X$,
- $f \circ g\colon Y \to Y$ is homotopic to the identity map $\mathrm{id}\colon Y \to Y$.

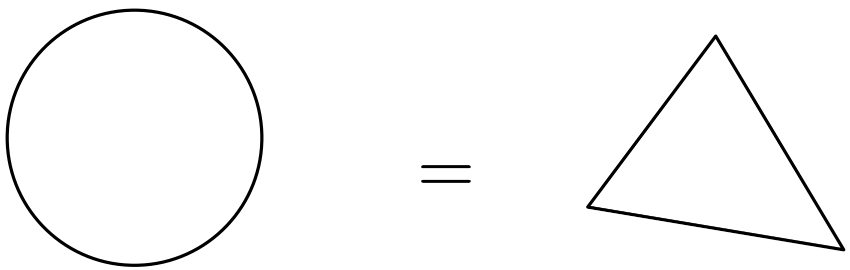
If such a homotopy equivalence exists, we say that X and Y are **homotopy equivalent**.



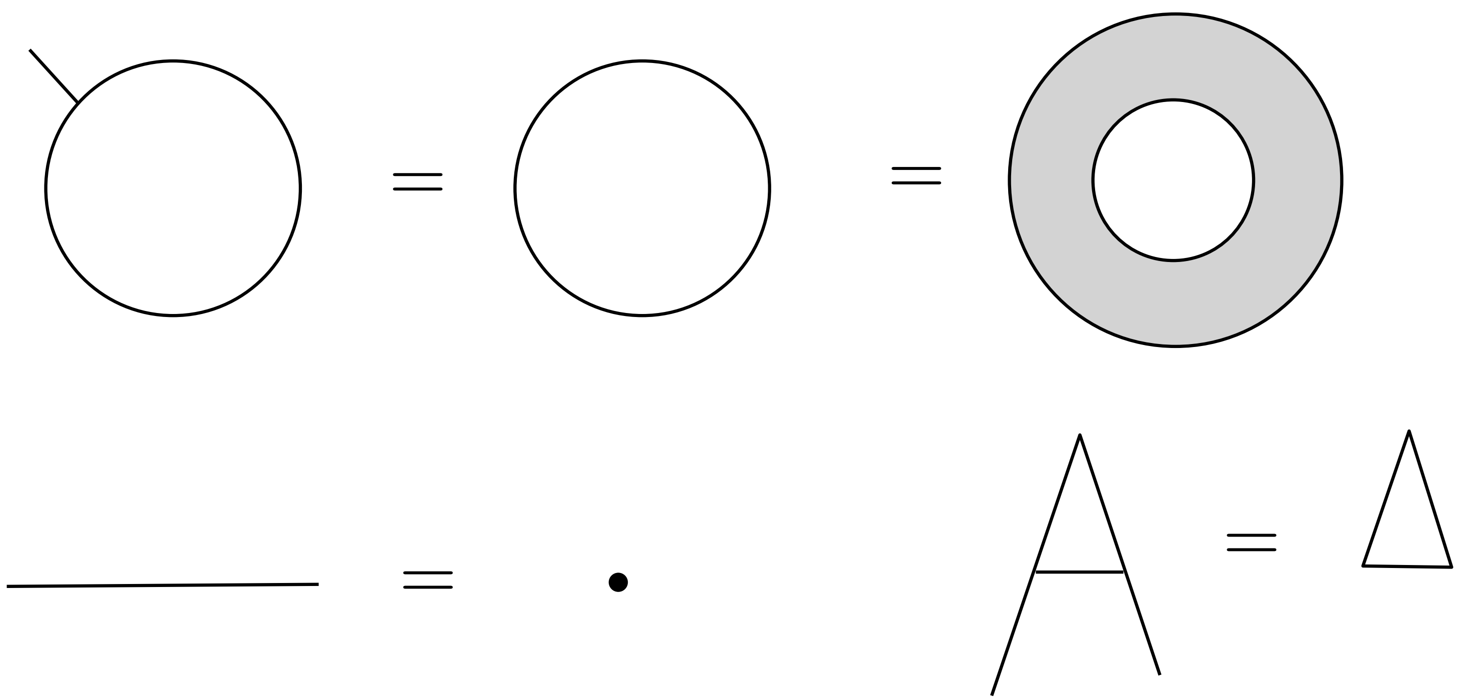
Proposition: If two topological spaces are homeomorphic, then they are homotopy equivalent.

Homotopy equivalence

Homotopy equivalence allows to continuously **deform** the space

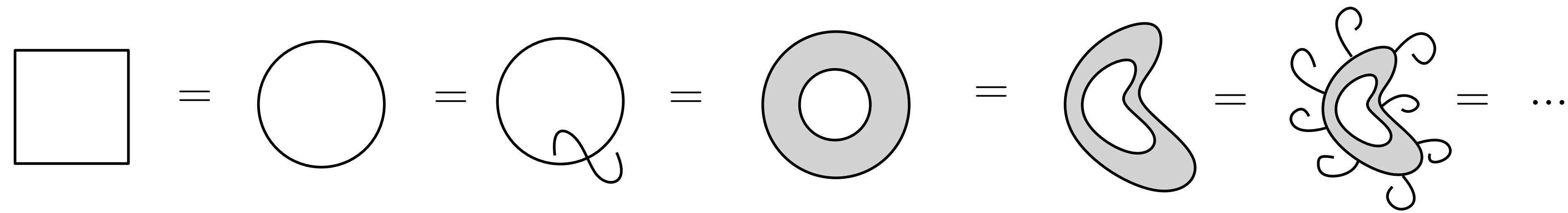


and to **retract** it.

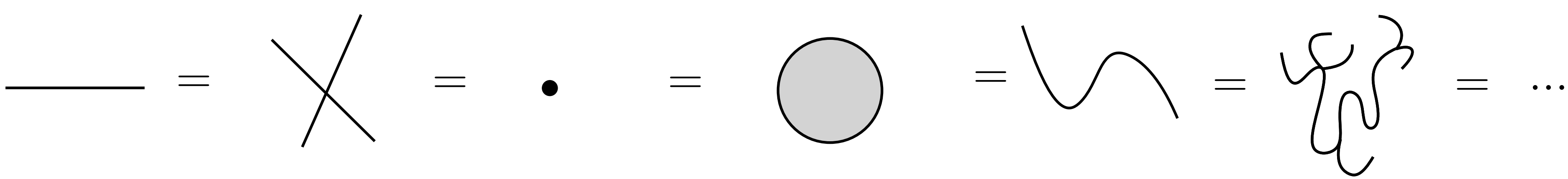


Homotopy classes

Just as before, we can classify topological spaces according to this relation, and obtain **classes of homotopy equivalence**:



the class of circles



the class of points

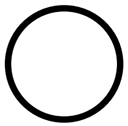
the class of spheres, the class of torii, the class of Klein bottles, ...

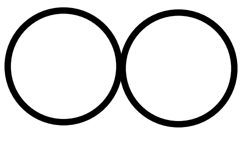
Homotopy classes

Example: Classification, up to homotopy equivalence, of the alphabet.

A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z				

Example: Classification, up to homotopy equivalence, of the alphabet.

A D O P Q R $\approx$ 

B $\approx$ 

C E F G H I J K L
M N S T U V W X Y Z $\approx \bullet$

I - Comparing topological spaces

1 - Homeomorphic equivalence

2 - Homotopy equivalence

II - Topological invariants

1 - Number of connected components

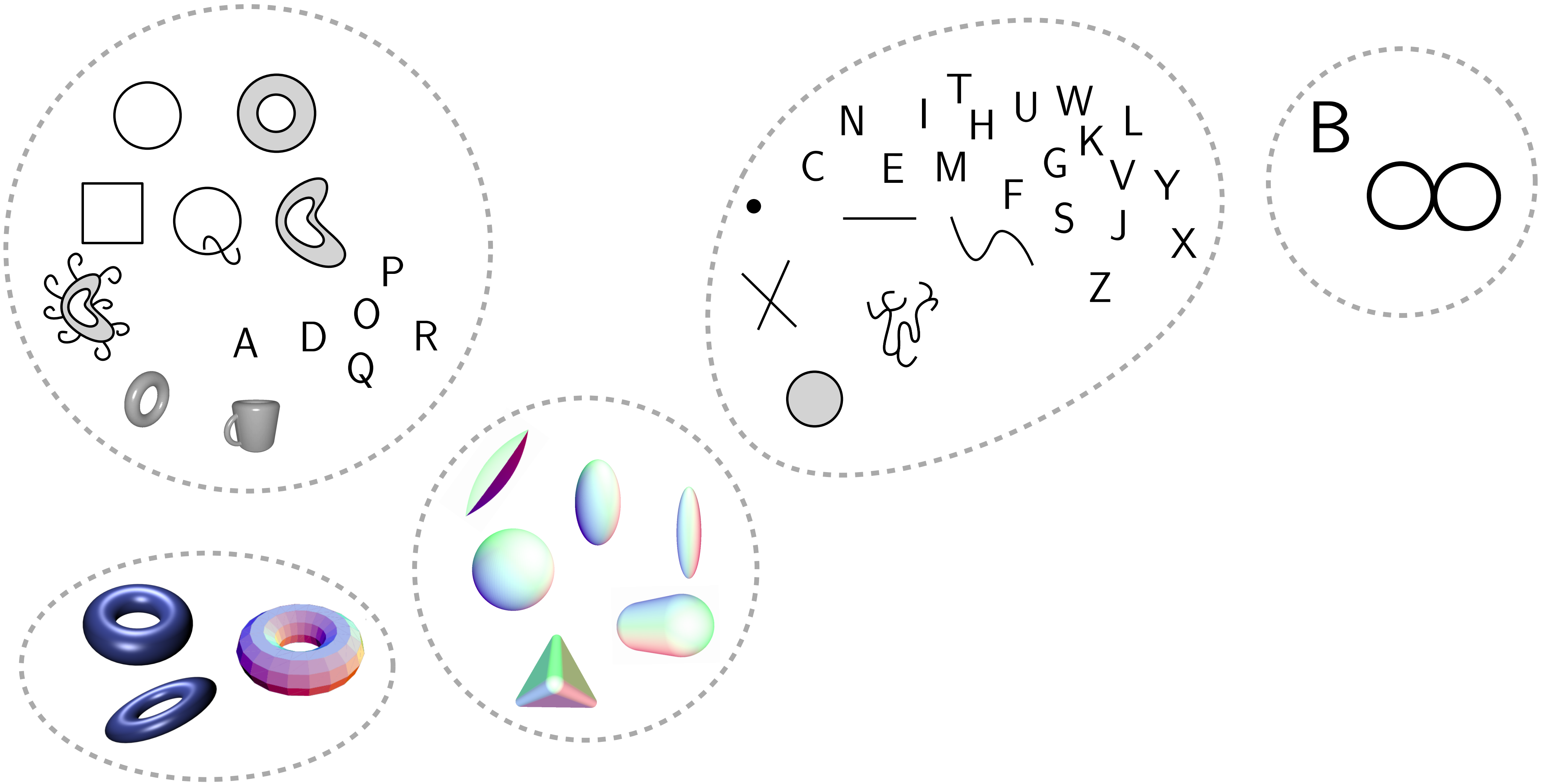
2 - Euler characteristic

3 - Betti numbers

(Next week - Persistent homology)

Invariants as features

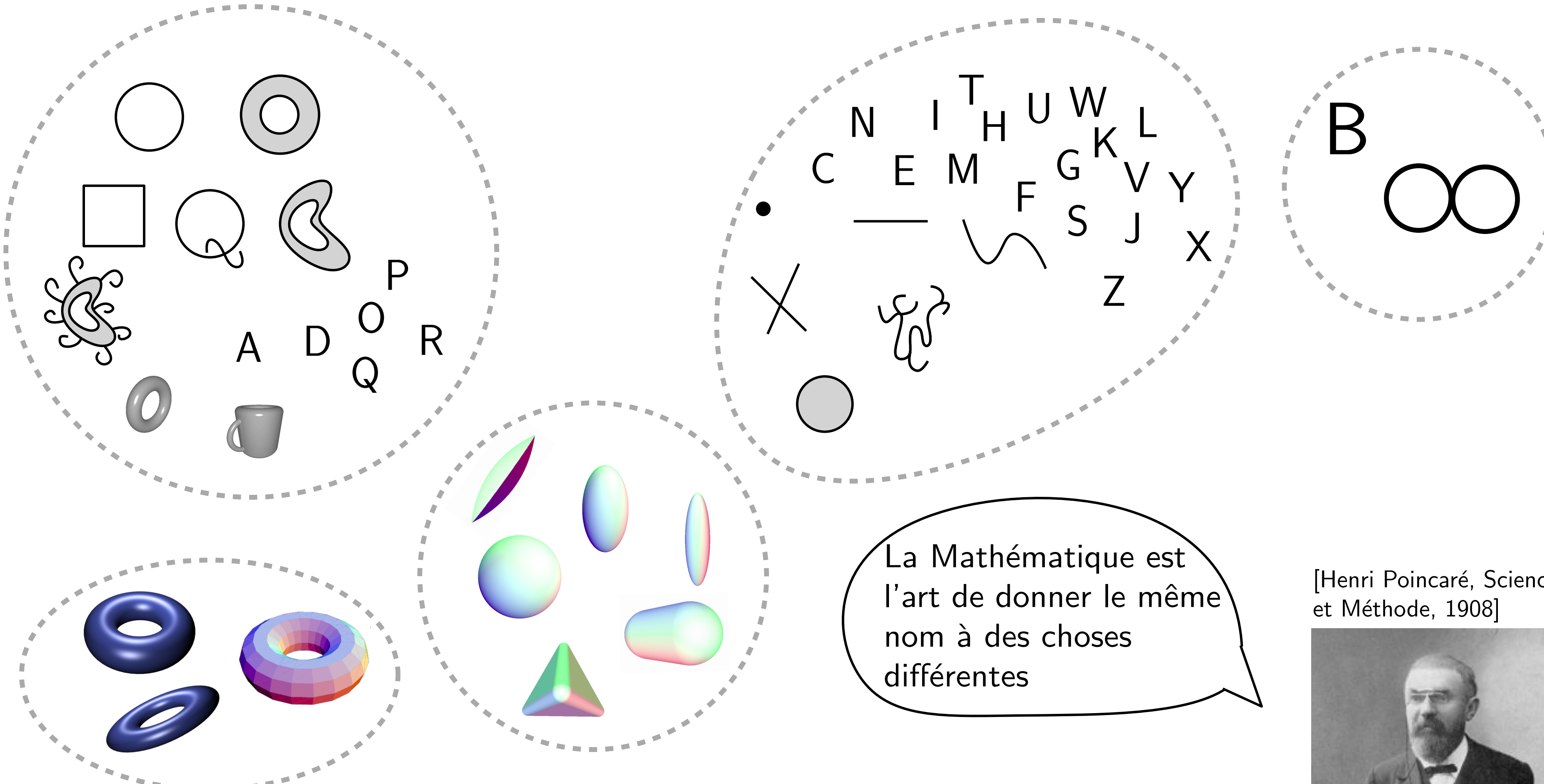
We gathered topological spaces into homotopy classes.



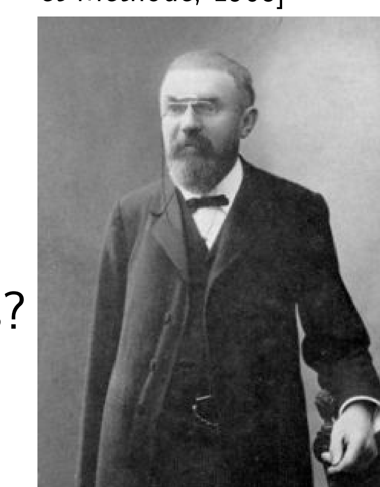
- Given a topological space X, how to recognize in which class it belongs?
- What are the common **features** of spaces in a same class?

Invariants as features

We gathered topological spaces into homotopy classes.



[Henri Poincaré, Science et Méthode, 1908]



- Given a topological space X, how to recognize in which class it belongs?
- What are the common **features** of spaces in a same class?

I - Comparing topological spaces

1 - Homeomorphic equivalence

2 - Homotopy equivalence

II - Topological invariants

1 - Number of connected components

2 - Euler characteristic

3 - Betti numbers

(Next week - Persistent homology)

Connected components

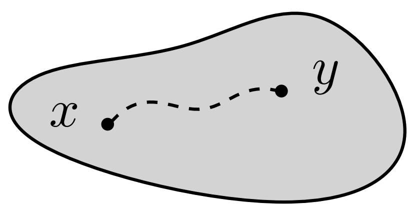
Definition: A subset $X \subset \mathbb{R}^n$ is (path-) **connected** if for every $x, y \in X$, there exists a continuous map $f\colon [0,1] \to X$ such that $f(0) = x$ and $f(1) = y$.

connected space

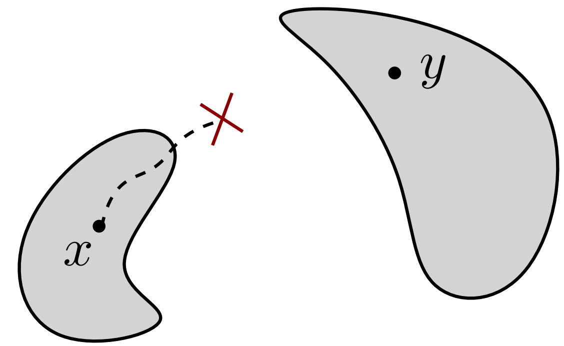
non-connected space

Connected components

Definition: A subset $X \subset \mathbb{R}^n$ is (path-) **connected** if for every $x, y \in X$, there exists a continuous map $f\colon [0,1] \to X$ such that $f(0) = x$ and $f(1) = y$.



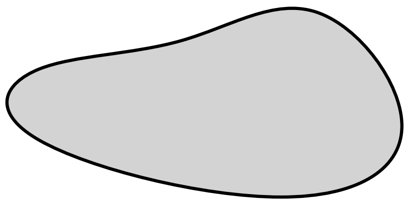
connected space



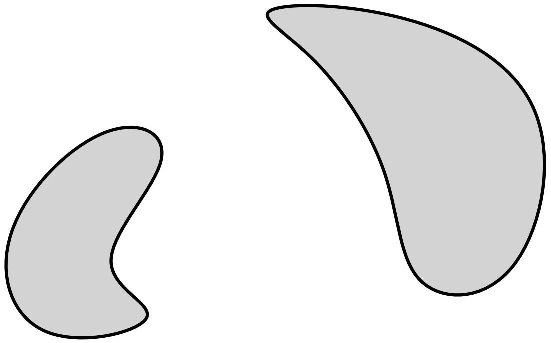
non-connected space

Connected components

Definition: A subset $X \subset \mathbb{R}^n$ is (path-) **connected** if for every $x, y \in X$, there exists a continuous map $f\colon [0,1] \to X$ such that $f(0) = x$ and $f(1) = y$.

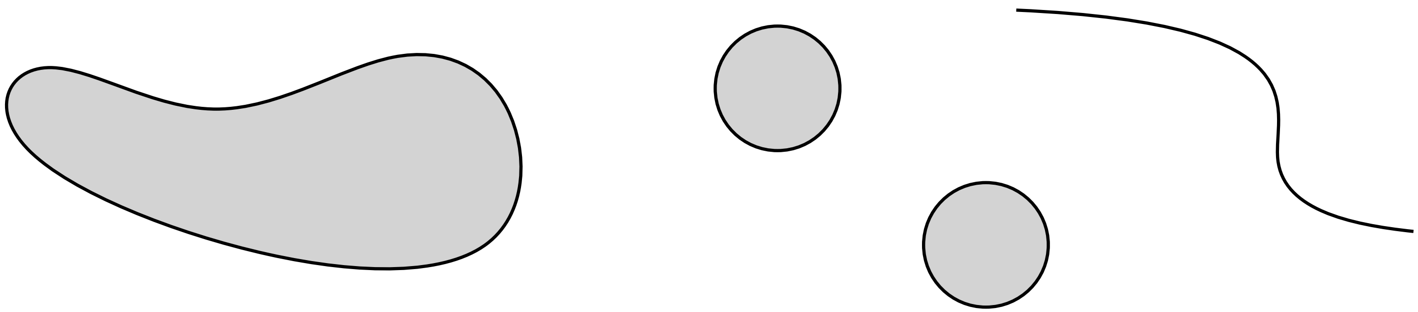


connected space



non-connected space

More generally, any topological space X can be partitioned into **connected components**.



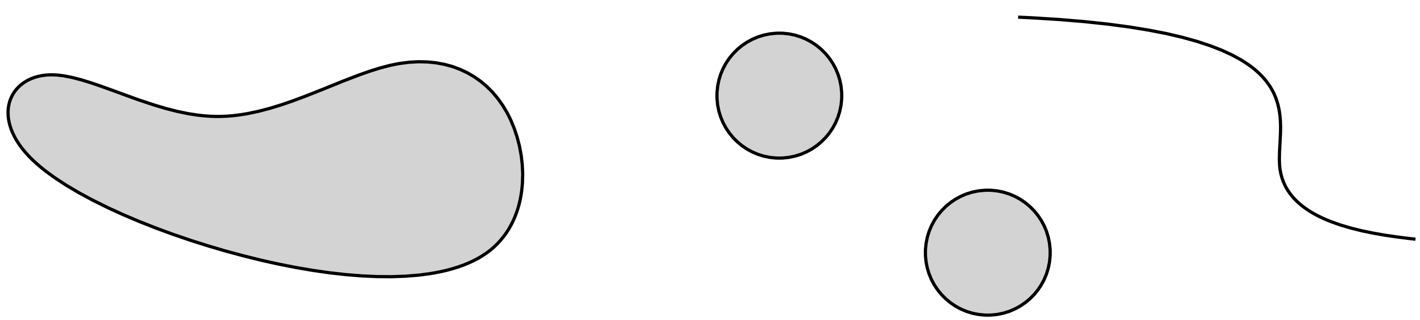
Connected components

Definition: A subset $X \subset \mathbb{R}^n$ is (path-) **connected** if for every $x, y \in X$, there exists a continuous map $f\colon [0,1] \to X$ such that $f(0) = x$ and $f(1) = y$.

connected space

non-connected space

More generally, any topological space X can be partitioned into **connected components**.



Proposition: If two spaces X and Y are homotopy equivalent, then they have the same number of connected components.

Invariance property - in theory

Proposition: If two spaces X and Y are homotopy equivalent, then they have the same number of connected components.

Consequence: If two spaces X and Y are homeomorphic, then they have the same number of connected components.

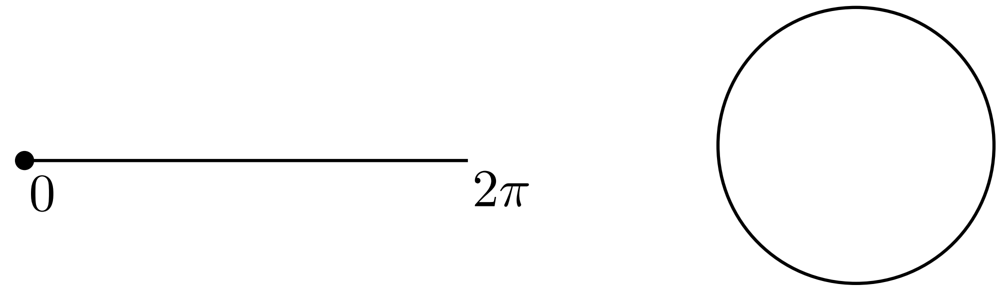
Example: The subsets $[0,1]$ and $[0,1] \cup [2,3]$ of $\mathbb{R}$ are not homeomorphic, neither homotopy equivalent.
Indeed, the first one has one connected component, and the second one two.

Invariance property - in theory

Proposition: If two spaces X and Y are homotopy equivalent, then they have the same number of connected components.

Example: The interval $[0, 2\pi)$ and the circle $\mathbb{S}_1 \subset \mathbb{R}^2$ are not homeomorphic.

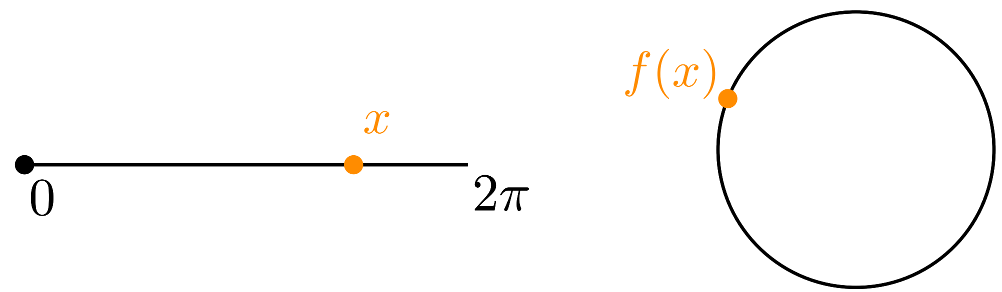
We will prove this by contradiction. Suppose that they are homeomorphic. By definition, this means that there exists a map $f\colon [0, 2\pi) \to \mathbb{S}_1$ which is continuous, invertible, and with continuous inverse.



Proposition: If two spaces X and Y are homotopy equivalent, then they have the same number of connected components.

Example: The interval $[0, 2\pi)$ and the circle $\mathbb{S}_1 \subset \mathbb{R}^2$ are not homeomorphic.

We will prove this by contradiction. Suppose that they are homeomorphic. By definition, this means that there exists a map $f\colon [0, 2\pi) \to \mathbb{S}_1$ which is continuous, invertible, and with continuous inverse.



Let $x \in [0, 2\pi)$ such that $x \neq 0$. Consider the subsets $[0, 2\pi) \setminus \{x\} \subset [0, 2\pi)$ and $\mathbb{S}_1 \setminus \{f(x)\} \subset \mathbb{S}_1$, and the induced map

$$g\colon [0, 2\pi) \setminus \{x\} \to \mathbb{S}_1 \setminus \{f(x)\}.$$

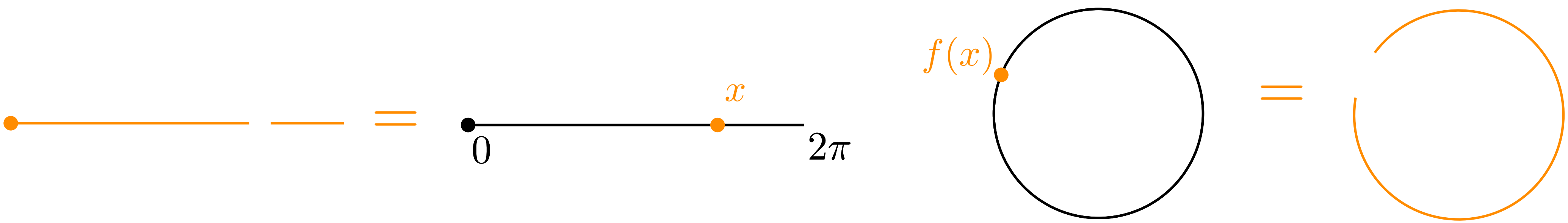
The map g is a homeomorphism.

Invariance property - in theory

Proposition: If two spaces X and Y are homotopy equivalent, then they have the same number of connected components.

Example: The interval $[0, 2\pi)$ and the circle $\mathbb{S}_1 \subset \mathbb{R}^2$ are not homeomorphic.

We will prove this by contradiction. Suppose that they are homeomorphic. By definition, this means that there exists a map $f\colon [0, 2\pi) \to \mathbb{S}_1$ which is continuous, invertible, and with continuous inverse.



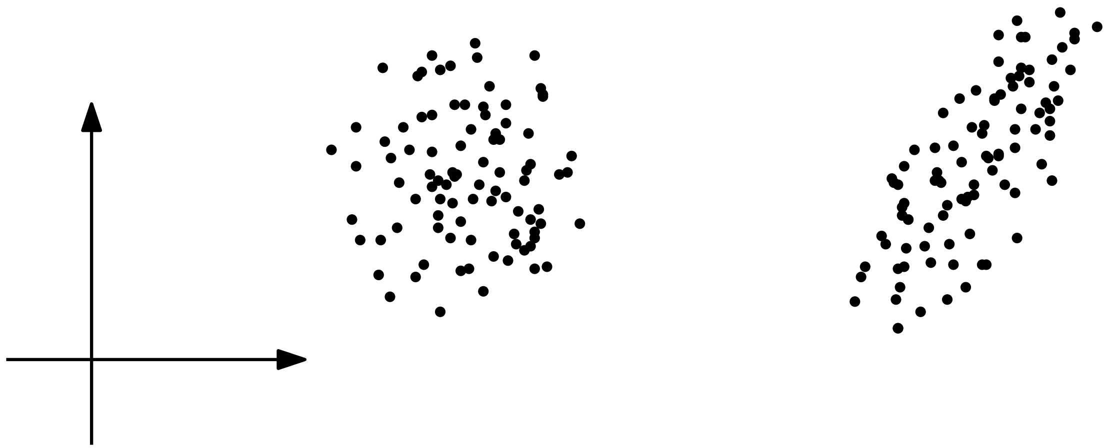
Let $x \in [0, 2\pi)$ such that $x \neq 0$. Consider the subsets $[0, 2\pi) \setminus \{x\} \subset [0, 2\pi)$ and $\mathbb{S}_1 \setminus \{f(x)\} \subset \mathbb{S}_1$, and the induced map

$$g\colon [0, 2\pi) \setminus \{x\} \to \mathbb{S}_1 \setminus \{f(x)\}.$$

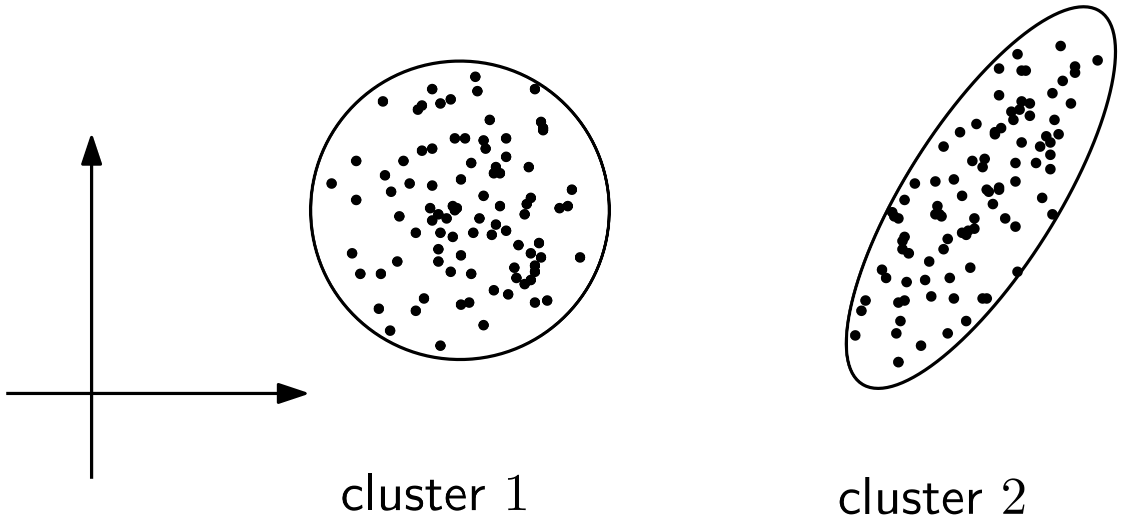
The map g is a homeomorphism.

Moreover, $[0, 2\pi) \setminus \{x\}$ has two connected components, and $t\mathbb{S}_1 \setminus \{f(x)\}$ only one. This is absurd.

In applications, finding connected components corresponds to a **classification** task.

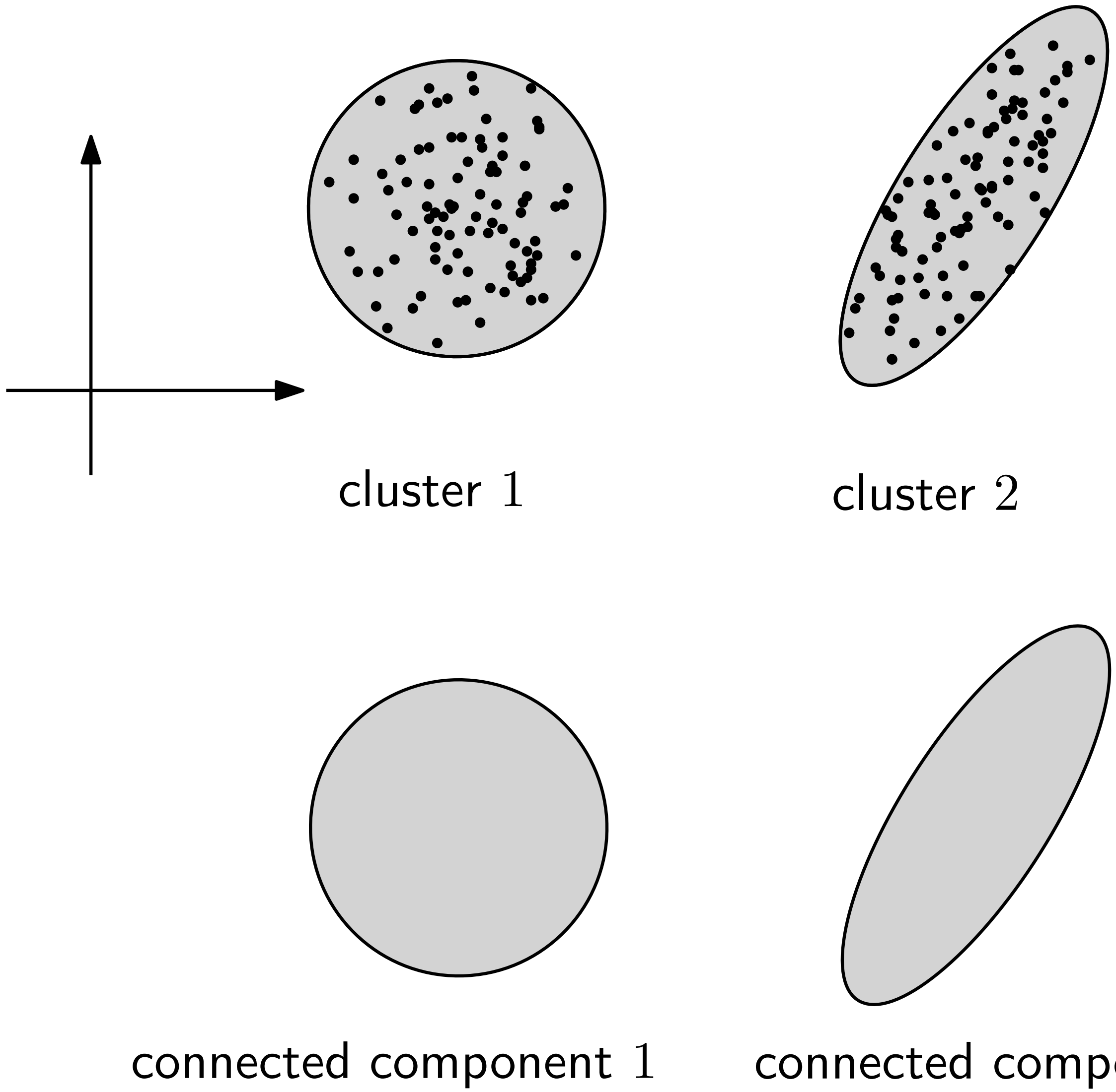


In applications, finding connected components corresponds to a **classification** task.



Invariance property - in applications

In applications, finding connected components corresponds to a **classification** task.



We can think of these sets as an **underlying topological space**, on which the points are sampled.

I - Comparing topological spaces

1 - Homeomorphic equivalence

2 - Homotopy equivalence

II - Topological invariants

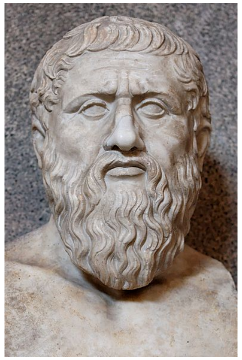
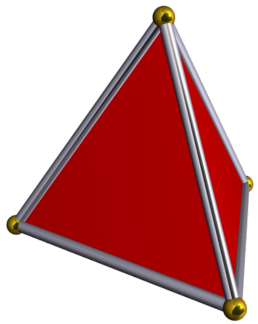
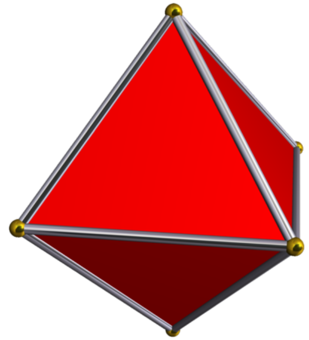
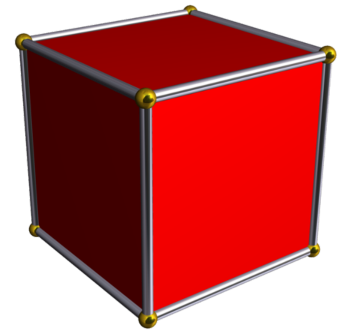
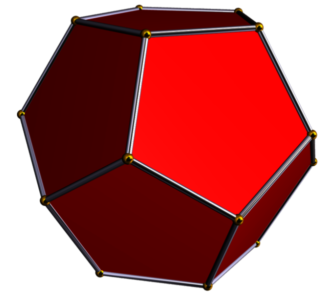
1 - Number of connected components

2 - Euler characteristic

3 - Betti numbers

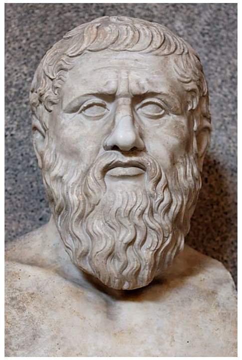
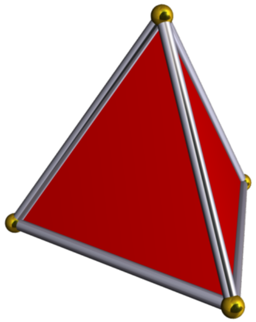
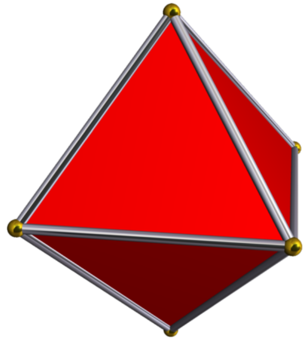
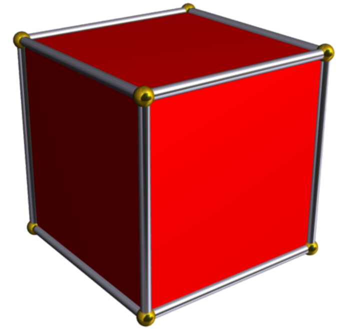
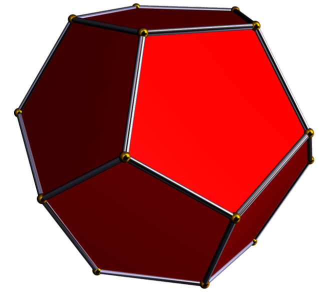
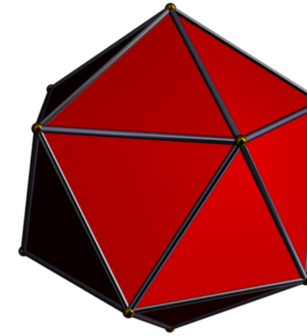
(Next week - Persistent homology)

Euler characteristic


number of faces	4	8	6	12	20
number of edges	6	12	12	30	30
number of vertices	4	6	8	20	12

Euler characteristic

					
number of faces	4	8	6	12	20
number of edges	6	12	12	30	30
number of vertices	4	6	8	20	12
χ	2	2	2	2	2

Proposition [Euler, 1758]: In any convex polyhedron, we have
number of faces $-$ number of edges $+$ number of vertices $= 2$

Euler characteristic

Definition: Let V be a set (called the set of *vertices*). A **simplicial complex** over V is a set K of subsets of V (called the *simplices*) such that, for every $\sigma \in K$ and every non-empty $\tau \subset \sigma$, we have $\tau \in K$.

The **dimension** of a simplex $\sigma \in K$ is defined as $|\sigma| - 1$.

Euler characteristic

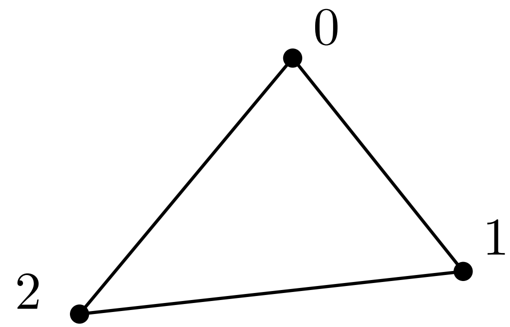
Definition: Let V be a set (called the set of *vertices*). A **simplicial complex** over V is a set K of subsets of V (called the *simplices*) such that, for every $\sigma \in K$ and every non-empty $\tau \subset \sigma$, we have $\tau \in K$.

The **dimension** of a simplex $\sigma \in K$ is defined as $|\sigma| - 1$.

Example: Let $V = \{0, 1, 2\}$ and

$$K = \{[0], [1], [2], [0, 1], [1, 2], [0, 2]\}.$$

This is a simplicial complex.



It contains three simplices of dimension 0 ($[0]$, $[1]$ and $[2]$) and three simplices of dimension 1 ($[0, 1]$, $[1, 2]$ and $[0, 2]$).

Euler characteristic

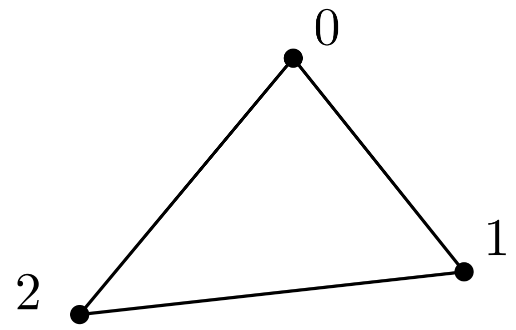
Definition: Let V be a set (called the set of *vertices*). A **simplicial complex** over V is a set K of subsets of V (called the *simplices*) such that, for every $\sigma \in K$ and every non-empty $\tau \subset \sigma$, we have $\tau \in K$.

The **dimension** of a simplex $\sigma \in K$ is defined as $|\sigma| - 1$.

Example: Let $V = \{0, 1, 2\}$ and

$$K = \{[0], [1], [2], [0, 1], [1, 2], [0, 2]\}.$$

This is a simplicial complex.



(this is a circle)

It contains three simplices of dimension 0 ($[0]$, $[1]$ and $[2]$) and three simplices of dimension 1 ($[0, 1]$, $[1, 2]$ and $[0, 2]$).

Euler characteristic

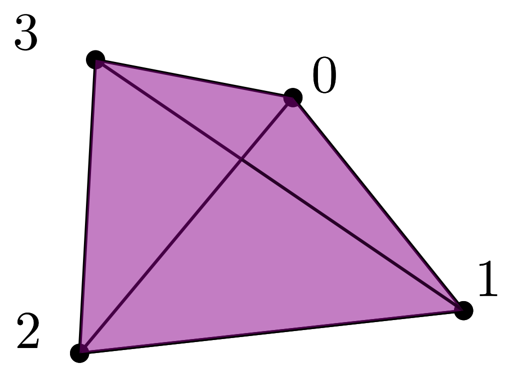
Definition: Let V be a set (called the set of *vertices*). A **simplicial complex** over V is a set K of subsets of V (called the *simplices*) such that, for every $\sigma \in K$ and every non-empty $\tau \subset \sigma$, we have $\tau \in K$.

The **dimension** of a simplex $\sigma \in K$ is defined as $|\sigma| - 1$.

Example: Let $V = \{0, 1, 2, 3\}$ and

$$K = \{[0], [1], [2], [3], [0, 1], [1, 2], [2, 3], [3, 0], [0, 2], [1, 3], [0, 1, 2], [0, 1, 3], [0, 2, 3], [1, 2, 3]\}$$

It a simplicial complex.



It contains four simplices of dimension 0 ($[0]$, $[1]$, $[2]$ and $[3]$), six simplices of dimension 1 ($[0, 1]$, $[1, 2]$, $[2, 3]$, $[3, 0]$,$[0, 2]$ and $[1, 3]$) and four simplices of dimension 2 ($[0, 1, 2]$, $[0, 1, 3]$, $[0, 2, 3]$ and $[1, 2, 3]$).

Euler characteristic

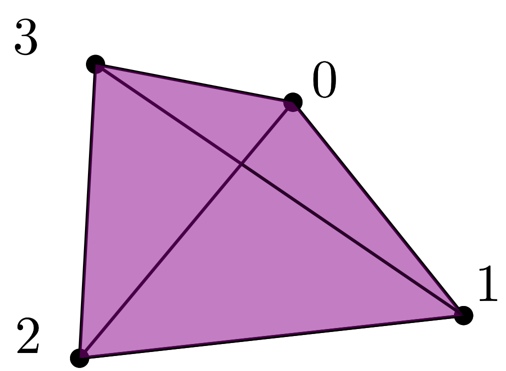
Definition: Let V be a set (called the set of *vertices*). A **simplicial complex** over V is a set K of subsets of V (called the *simplices*) such that, for every $\sigma \in K$ and every non-empty $\tau \subset \sigma$, we have $\tau \in K$.

The **dimension** of a simplex $\sigma \in K$ is defined as $|\sigma| - 1$.

Example: Let $V = \{0, 1, 2, 3\}$ and

$$K = \{[0], [1], [2], [3], [0, 1], [1, 2], [2, 3], [3, 0], [0, 2], [1, 3], [0, 1, 2], [0, 1, 3], [0, 2, 3], [1, 2, 3]\}$$

It a simplicial complex.



(this is a sphere)

It contains four simplices of dimension 0 ($[0]$, $[1]$, $[2]$ and $[3]$), six simplices of dimension 1 ($[0, 1]$, $[1, 2]$, $[2, 3]$, $[3, 0]$,$[0, 2]$ and $[1, 3]$) and four simplices of dimension 2 ($[0, 1, 2]$, $[0, 1, 3]$, $[0, 2, 3]$ and $[1, 2, 3]$).

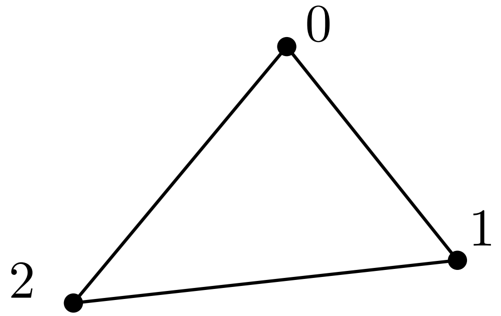
Euler characteristic

Definition: Let K be a simplicial complex of dimension n. Its *Euler characteristic* is the integer

$$\chi(K) = \sum_{0 \leq i \leq n} (-1)^i \cdot (\text{number of simplices of dimension } i).$$

Example: The simplicial complex $K = \{[0], [1], [2], [0,1], [1,2], [2,0]\}$ has Euler characteristic

$$\chi(K) = 3 - 3 = 0$$



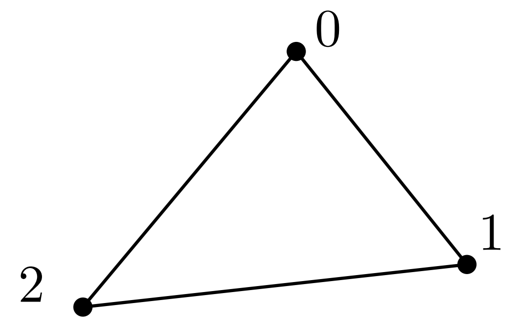
Euler characteristic

Definition: Let K be a simplicial complex of dimension n. Its *Euler characteristic* is the integer

$$\chi(K) = \sum_{0 \leq i \leq n} (-1)^i \cdot (\text{number of simplices of dimension } i).$$

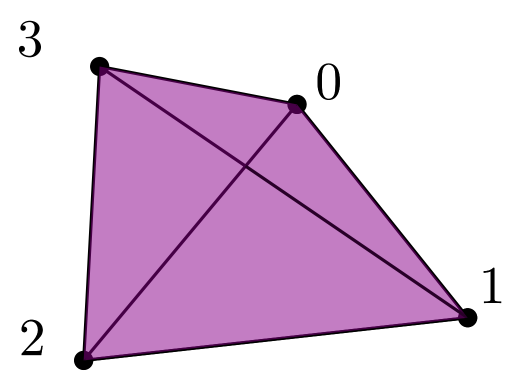
Example: The simplicial complex $K = \{[0], [1], [2], [0,1], [1,2], [2,0]\}$ has Euler characteristic

$$\chi(K) = 3 - 3 = 0$$



Example: The simplicial complex
$K = \{[0], [1], [2], [3], [0,1], [1,2], [2,3], [3,0], [0,2], [1,3], [0,1,2], [0,1,3], [0,2,3], [1,2,3]\}$
has Euler characteristic

$$\chi(K) = 4 - 6 + 4 = 2$$



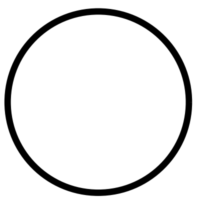
Invariance property - in theory

Proposition: If X and Y are two homotopy equivalent topological spaces, then $\chi(X) = \chi(Y)$.

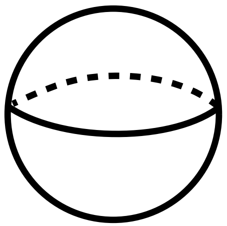
Therefore, the Euler characteristic is an **invariant** of homotopy equivalence classes.

We can use this information to prove that two spaces are not homotopy equivalent.

Example: The circle has Euler characteristic 0, and the sphere Euler characteristic 2. Therefore, they are not homotopy equivalent.



$\chi(\mathbb{S}_1) = 0$



$\chi(\mathbb{S}_2) = 2$

Invariance property - in theory

Proposition: If X and Y are two homotopy equivalent topological spaces, then $\chi(X) = \chi(Y)$.

Therefore, the Euler characteristic is an **invariant** of homotopy equivalence classes.

We can use this information to prove that two spaces are not homotopy equivalent.

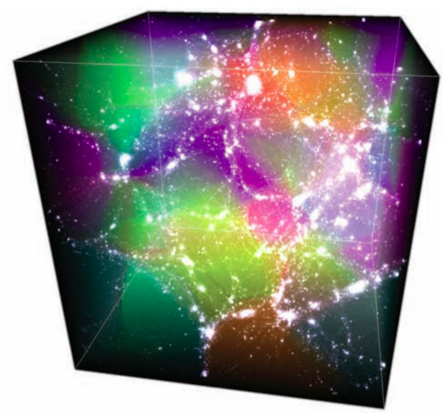
Example: Classification of surfaces.
The homeomorphism classes of *connected and compact surfaces* are classified by their Euler characteristic.

χ $\quad 2 \quad 0 \quad -2 \quad -4 \quad 2 - 2 \times \text{genus}$

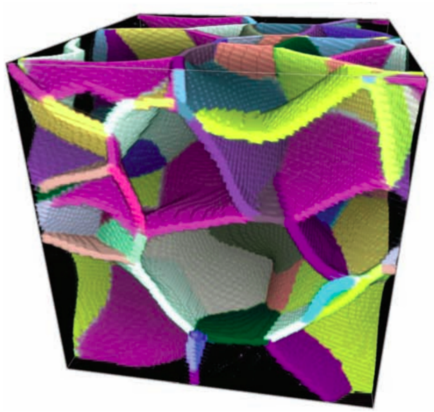
Invariance property - in applications

The Euler characteristic contains information about the homeomorphism class (and homotopy class) of the space.

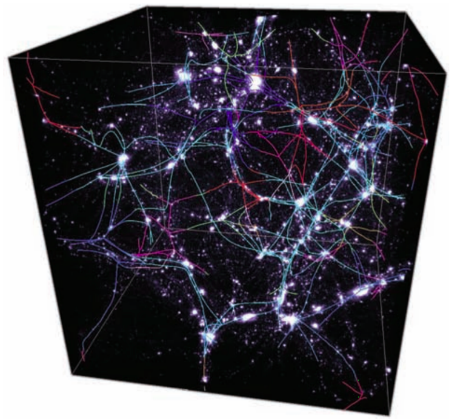
[T. Sousbie, The persistent cosmic web and its filamentary structure, 2011]



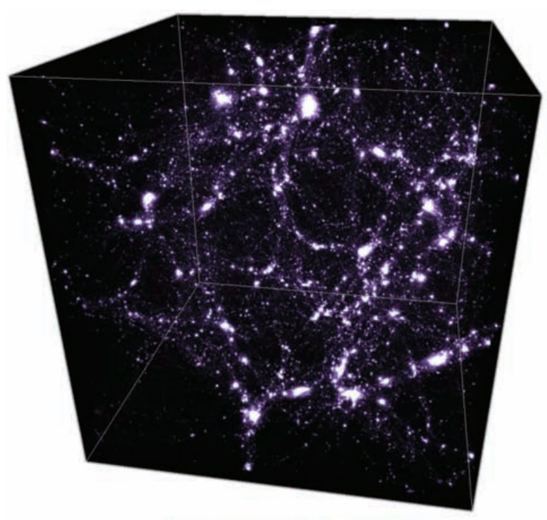
seen as an object of dimension 3



of dimension 2

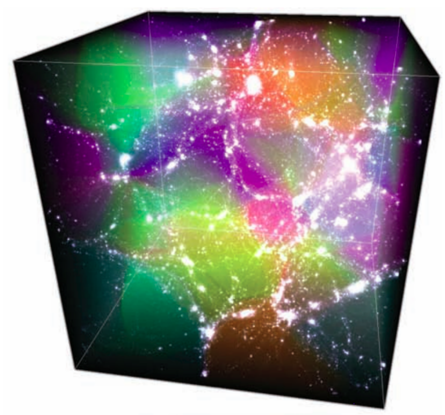


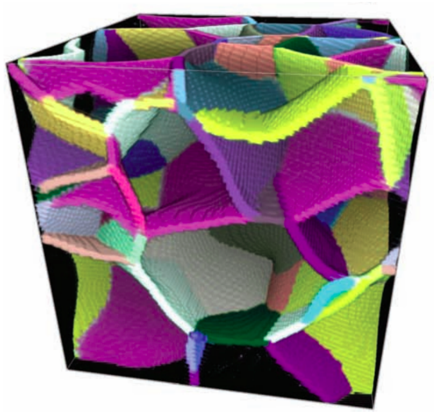
of dimension 1

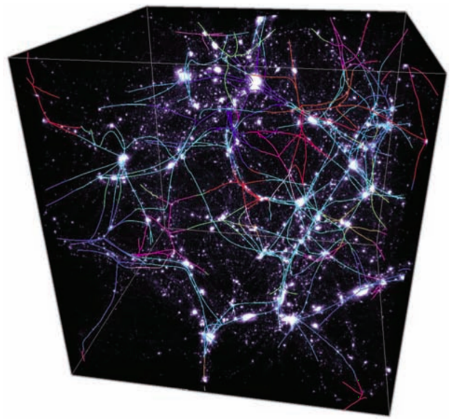


The Euler characteristic contains information about the homeomorphism class (and homotopy class) of the space.

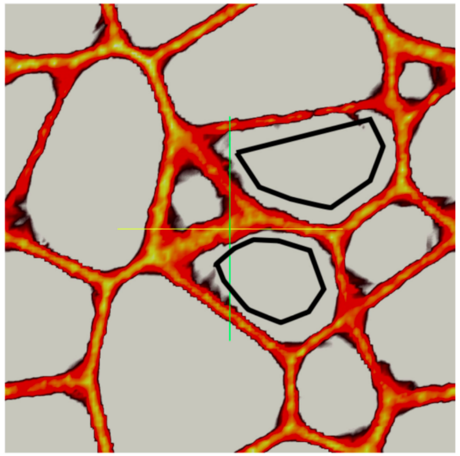
[T. Sousbie, The persistent cosmic web and its filamentary structure, 2011]







[P. Pranav, H. Edelsbrunner, R. de Weygaert, G. Vegter, M. Kerber, B. Jones and M. Wintraecken, The topology of the cosmic web in terms of persistent Betti numbers, 2016]



The Euler characteristic 'counts' the number of holes

I - Comparing topological spaces

1 - Homeomorphic equivalence

2 - Homotopy equivalence

II - Topological invariants

1 - Number of connected components

2 - Euler characteristic

3 - Betti numbers

(Next week - Persistent homology)

Betti numbers

For any topological space X, one defines a sequence of integers

$$\beta_0(X), \quad \beta_1(X), \quad \beta_2(X), \quad \beta_3(X), \quad \ldots$$

called the **Betti numbers**.

Construction of Betti numbers:

Betti numbers

For any topological space X, one defines a sequence of integers

$$\beta_0(X), \quad \beta_1(X), \quad \beta_2(X), \quad \beta_3(X), \quad \ldots$$

called the **Betti numbers**.

Construction of Betti numbers: rendez-vous next week! (based on homology theory)

Betti numbers

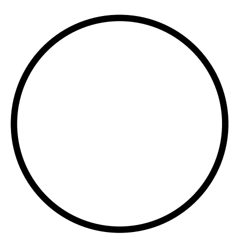
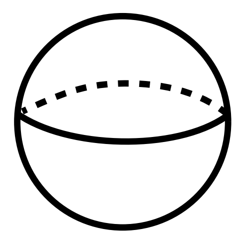
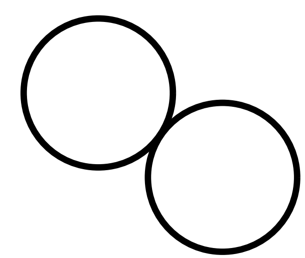
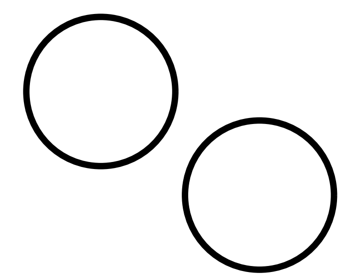
For any topological space X, one defines a sequence of integers

$$\beta_0(X), \quad \beta_1(X), \quad \beta_2(X), \quad \beta_3(X), \quad \ldots$$

called the **Betti numbers**.

Construction of Betti numbers: rendez-vous next week! (based on homology theory)

Example: Let us give some examples instead.

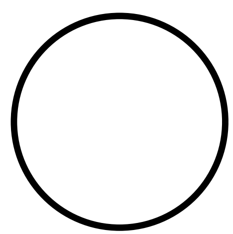
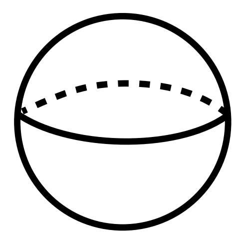
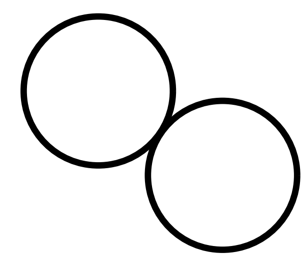
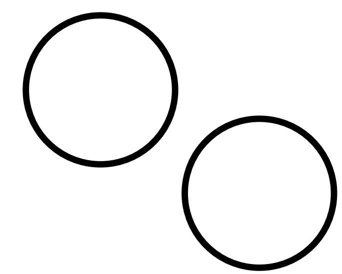
X					
$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

Betti numbers

Interpretation: We have:

- $\beta_0(X)$ is the number of connected components of X
- $\beta_1(X)$ is the number of 'holes' in X
- $\beta_2(X)$ is the number of 'voids' in X
- $\dots$

Example: Let us give some examples instead.

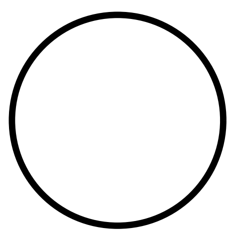
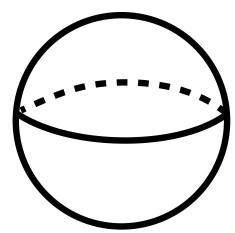
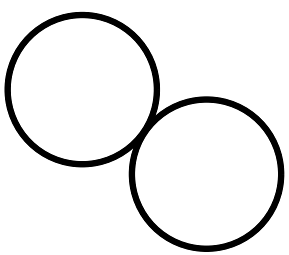
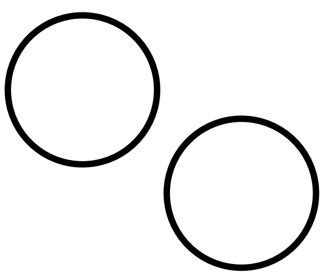
X					
$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

Betti numbers

Interpretation: We have:

- $\beta_0(X)$ is the number of connected components of X
- $\beta_1(X)$ is the number of 'holes' in X
- $\beta_2(X)$ is the number of 'voids' in X
- $\dots$

Example: Let us give some examples instead.

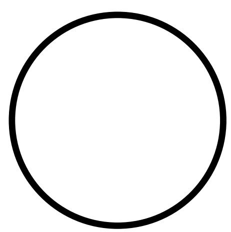
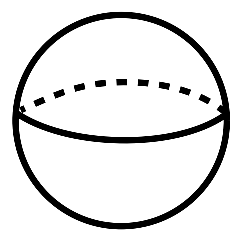
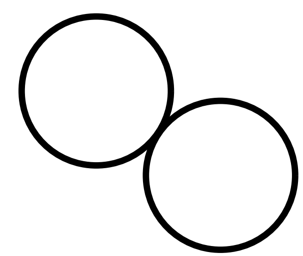
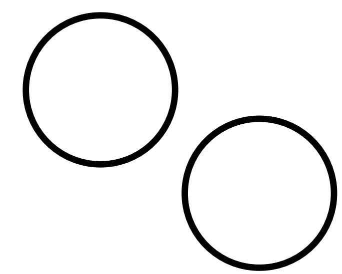
X					
$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

Betti numbers

Interpretation: We have:

- $\beta_0(X)$ is the number of connected components of X
- $\beta_1(X)$ is the number of 'holes' in X
- $\beta_2(X)$ is the number of 'voids' in X
- $\dots$

Example: Let us give some examples instead.

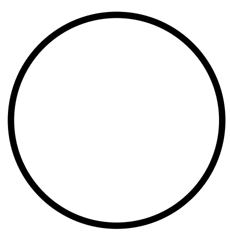
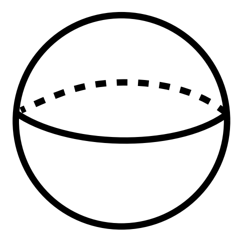
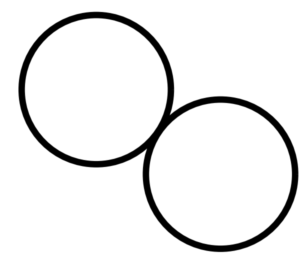
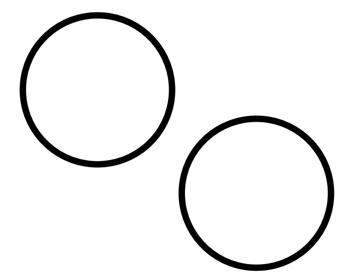
X					
$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

Betti numbers

Interpretation: We have:

- $\beta_0(X)$ is the number of connected components of X
- $\beta_1(X)$ is the number of 'holes' in X
- $\boxed{\beta_2(X) \text{ is the number of 'voids' in } X}$
- ...

Example: Let us give some examples instead.

X					
$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

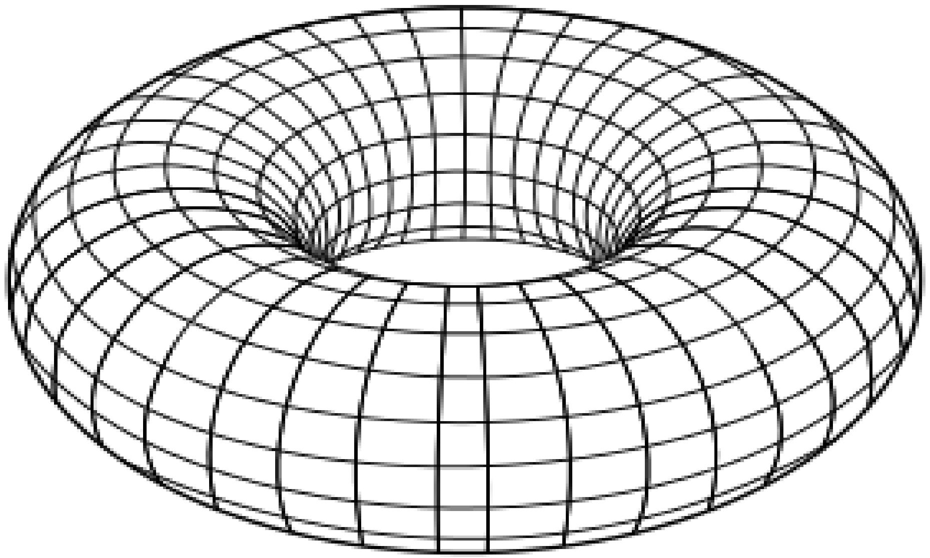
Betti numbers

Interpretation: We have:

- $\beta_0(X)$ is the number of connected components of X
- $\beta_1(X)$ is the number of 'holes' in X
- $\beta_2(X)$ is the number of 'voids' in X
- $\ldots$

Example: Betti numbers of the torus:

$$\beta_0(X) = 1, \quad \beta_1(X) = 2, \quad \beta_2(X) = 1, \quad \beta_3(X) = 0, \quad \ldots$$



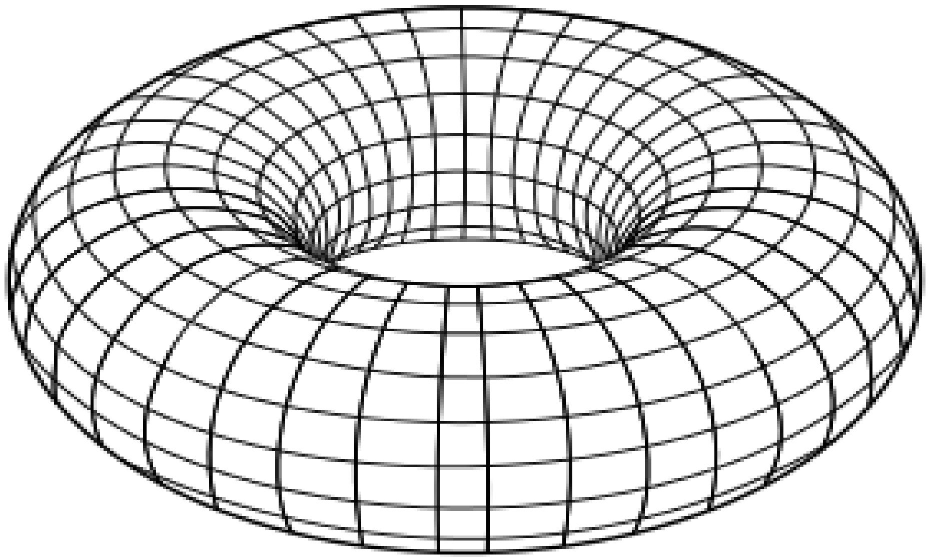
Betti numbers

Interpretation: We have:

- $\beta_0(X)$ is the number of connected components of X
- $\beta_1(X)$ is the number of 'holes' in X
- $\beta_2(X)$ is the number of 'voids' in X
- $\dots$

Example: Betti numbers of the torus:

$$\boxed{\beta_0(X) = 1,} \quad \beta_1(X) = 2, \quad \beta_2(X) = 1, \quad \beta_3(X) = 0, \quad \dots$$



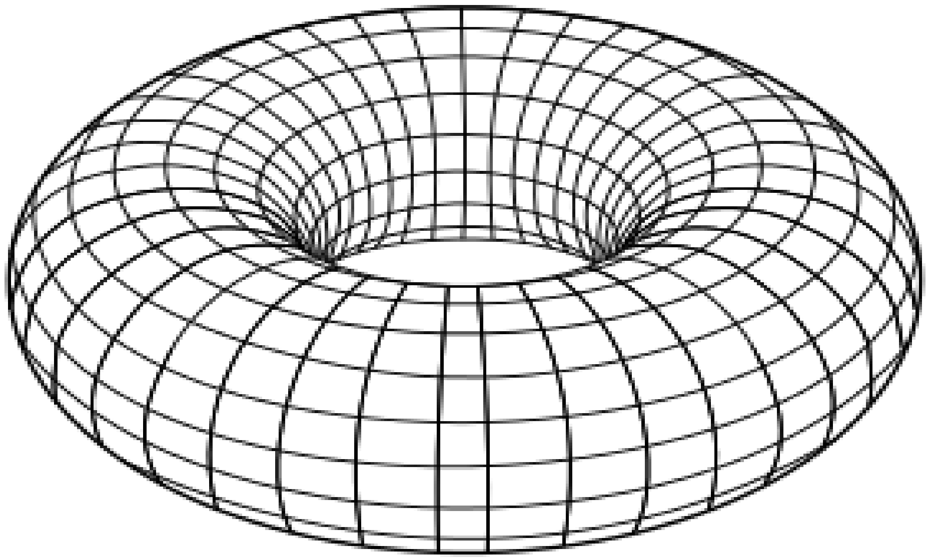
Betti numbers

Interpretation: We have:

- $\beta_0(X)$ is the number of connected components of X
- $\beta_1(X)$ is the number of 'holes' in X
- $\beta_2(X)$ is the number of 'voids' in X
- $\dots$

Example: Betti numbers of the torus:

$$\beta_0(X) = 1, \quad \beta_1(X) = 2, \quad \boxed{\beta_2(X) = 1,} \quad \beta_3(X) = 0, \quad \dots$$



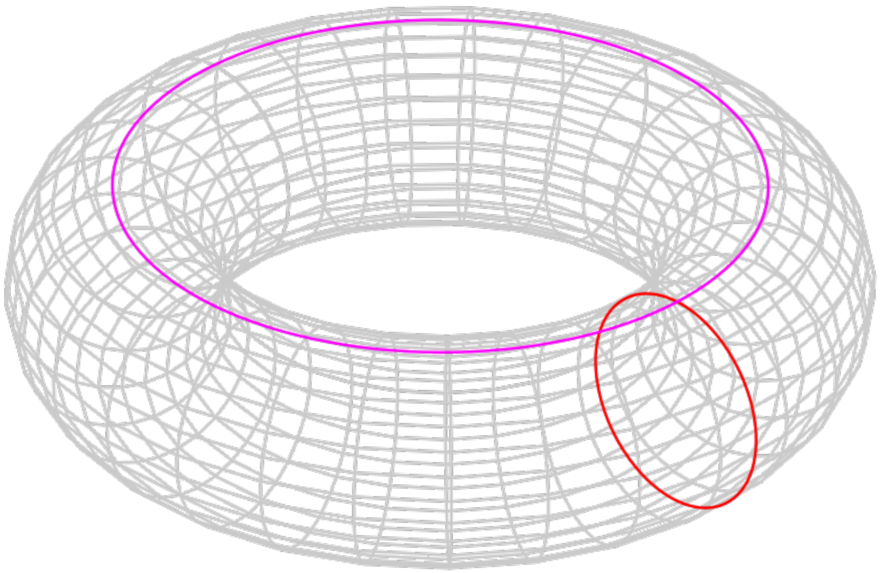
Betti numbers

Interpretation: We have:

- $\beta_0(X)$ is the number of connected components of X
- $\beta_1(X)$ is the number of 'holes' in X
- $\beta_2(X)$ is the number of 'voids' in X
- $\dots$

Example: Betti numbers of the torus:

$$\beta_0(X) = 1, \quad \boxed{\beta_1(X) = 2,} \quad \beta_2(X) = 1, \quad \beta_3(X) = 0, \quad \dots$$



Proposition: If two spaces X and Y are homotopy equivalent, then they have the same Betti numbers.

As a consequence, two spaces with different Betti numbers cannot be homotopy equivalent.

Example: The n-dimensional sphere $\mathbb{S}_n \subset \mathbb{R}^{n+1}$ has Betti numbers

$$\begin{aligned} \beta_i(X) &= 1 \quad \text{if} \;\; i = 0 \text{ or } n, \\ \beta_i(X) &= 0 \quad \text{else.} \end{aligned}$$

Hence, if $n \neq m$, then $\mathbb{S}_n$ and $\mathbb{S}_m$ are not homotopy equivalent.

Invariance property - in theory

Proposition: If two spaces X and Y are homotopy equivalent, then they have the same Betti numbers.

As a consequence, two spaces with different Betti numbers cannot be homotopy equivalent.

Example: Brouwer's invariance of domain.

Let us show that $\mathbb{R}^n$ and $\mathbb{R}^m$, with $n \neq m$, are not homeomorphic.

Let $h \colon \mathbb{R}^n \to \mathbb{R}^m$ be a homeomorphism.
Choose any $x \in \mathbb{R}^n$and consider the restricted map

$$h \colon \mathbb{R}^n \setminus \{x\} \longrightarrow \mathbb{R}^m \setminus \{h(x)\}$$

It is still a homemorphism.

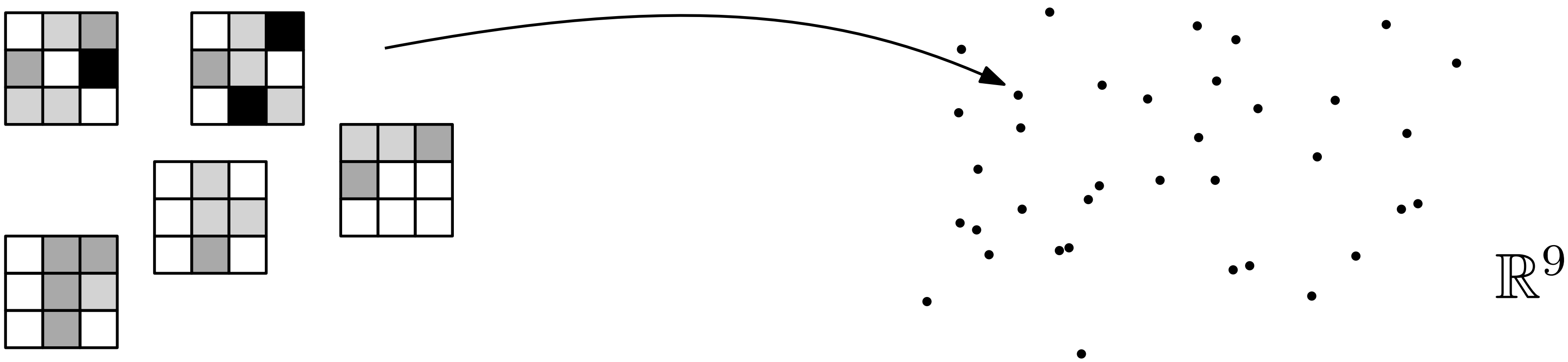
But $\mathbb{R}^n \setminus \{x\}$ is homotopic to the sphere $\mathbb{S}_{n-1}$, and $\mathbb{R}^m \setminus \{x\}$ is homotopic to the sphere $\mathbb{S}_{m-1}$

We have seen before that $\mathbb{S}_{n-1}$ and $\mathbb{S}_{m-1}$ are homotopic if and only if $m = n$. This is a contradiction.

The Betti numbers contain information about the space we study.

[G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, On the Local Behavior of Spaces of Natural Images, 2008.]

From a large collection of natural images, the authors extract 3×3 patches.
Since it consists of 9 pixels, each of these patches can be seen as a 9-dimensional vector, and the whole set as a point cloud in $\mathbb{R}^9$.



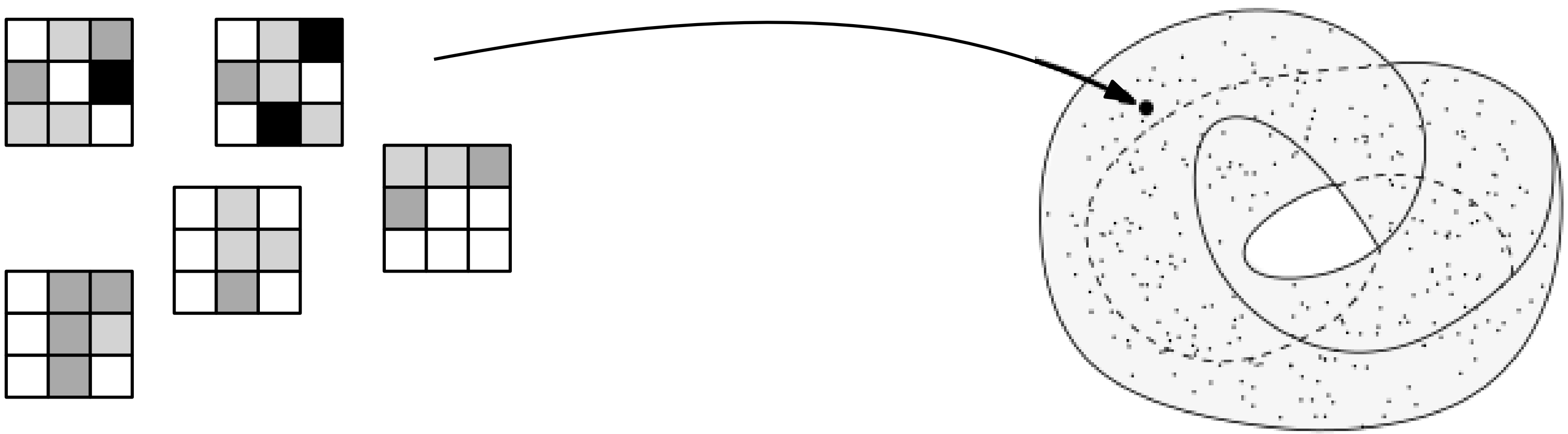
They observe that the point cloud lies close to a shape whose Betti numbers (over $\mathbb{Z}/2\mathbb{Z}$) are

$$\beta_0(X) = 1, \quad \beta_1(X) = 2, \quad \beta_2(X) = 1, \quad \beta_3(X) = 0$$

The Betti numbers contain information about the space we study.

[G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, On the Local Behavior of Spaces of Natural Images, 2008.]

From a large collection of natural images, the authors extract 3×3 patches. Since it consists of 9 pixels, each of these patches can be seen as a 9-dimensional vector, and the whole set as a point cloud in $\mathbb{R}^9$.



$\mathbb{R}^9$

They observe that the point cloud lies close to a shape whose Betti numbers (over $\mathbb{Z}/2\mathbb{Z}$) are

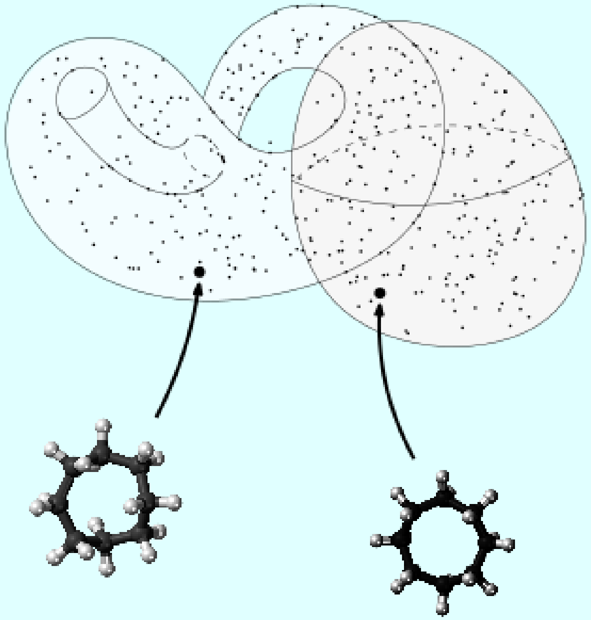
$$\beta_0(X) = 1, \quad \beta_1(X) = 2, \quad \beta_2(X) = 1, \quad \beta_3(X) = 0$$

These are the Betti numbers of a Klein bottle!

(and the authors actually show that the dataset concentrates near a Klein bottle embedded in $\mathbb{R}^9$.)

Conclusion

We can hope to find interesting topology in datasets.



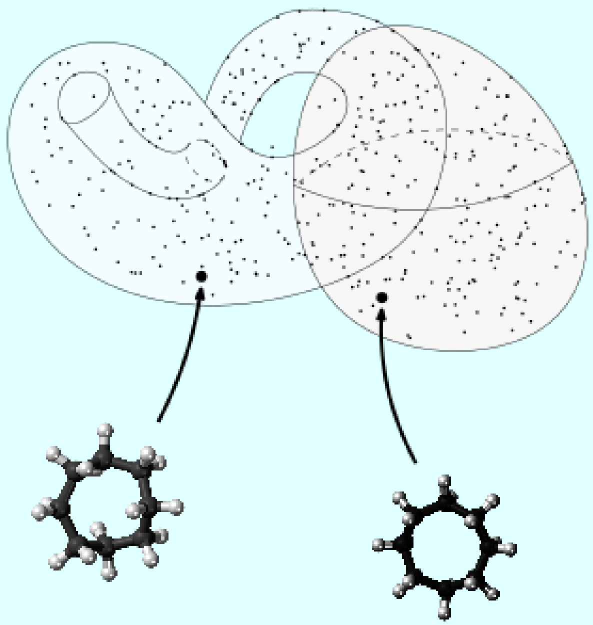
Invariants of homotopy classes allow to describe and understand them.

$$\text{Number of connected components} = 1$$
$$\chi = 2$$
$$\beta_0(X) = 1, \quad \beta_1(X) = 1, \quad \beta_2(X) = 2$$

Next week: how to compute these invariants in practice?

Conclusion

We can hope to find interesting topology in datasets.



Invariants of homotopy classes allow to describe and understand them.

Number of connected components $= 1$
$\chi = 2$
$\beta_0(X) = 1, \quad \beta_1(X) = 1, \quad \beta_2(X) = 2$

Next week: how to compute these invariants in practice?

Obrigado!