

Topological Data Analysis

A Primer on Topology

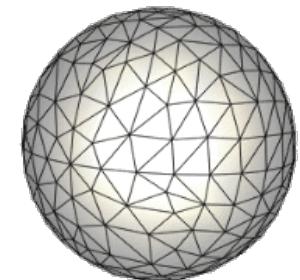
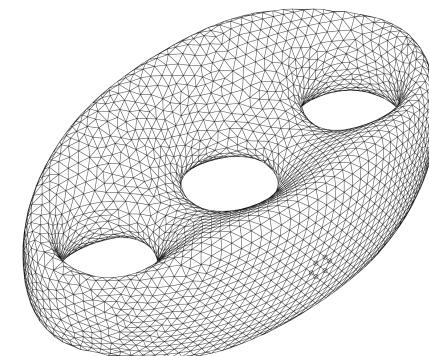
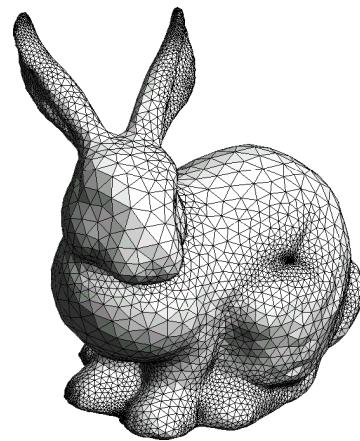
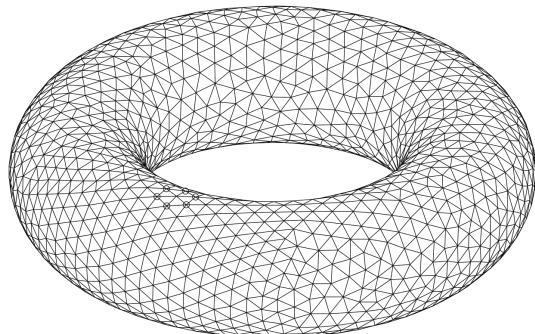
Ulderico Fugacci

CNR - IMATI



Topological Data Analysis

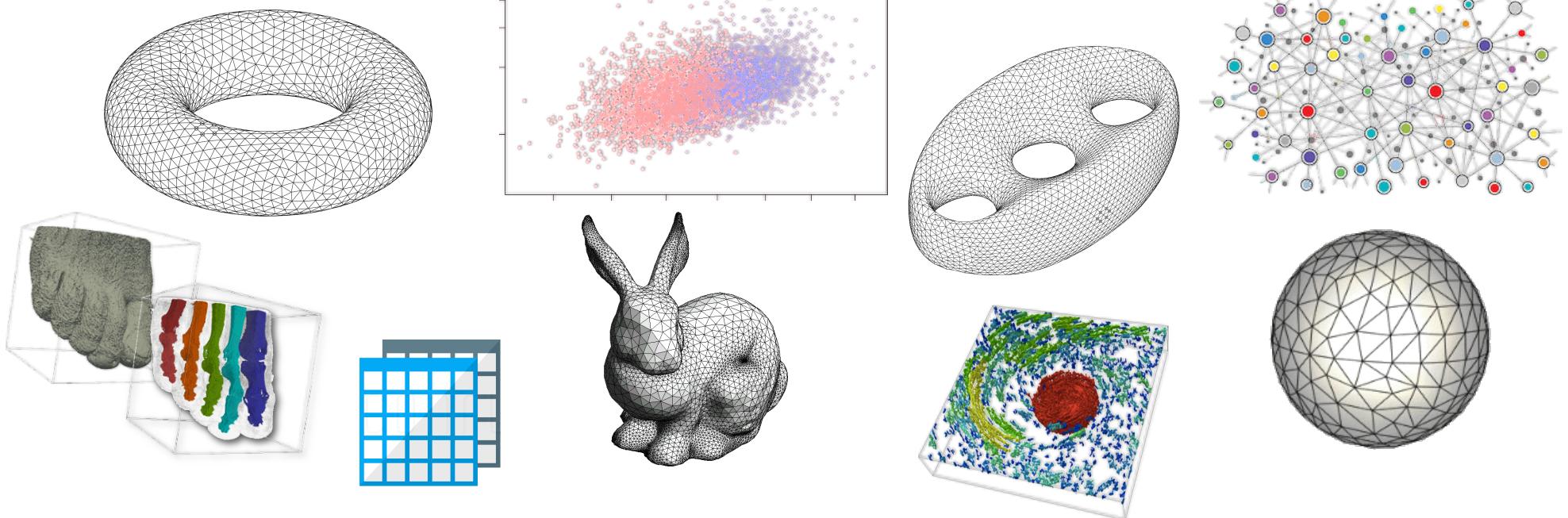
Topology describes, characterizes, and discriminates *shapes* by studying their properties that are preserved under *continuous deformations*, such as *stretching* and *bending*, but *not tearing* or *gluing*



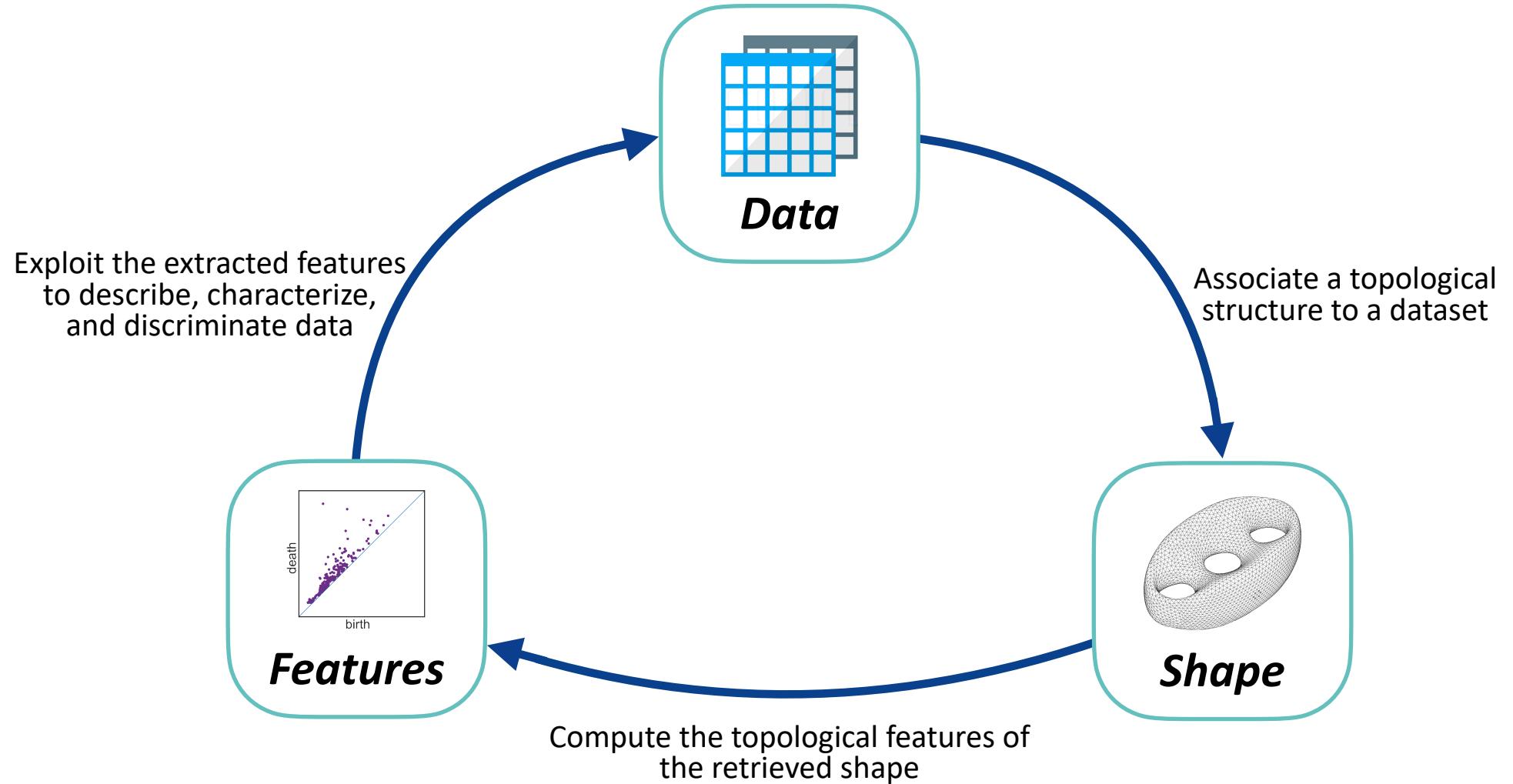
Topological Data Analysis

Assumption in TDA: *Any data* can be endowed with a *shape*.

So, any data can be studied in terms of its *topological features*

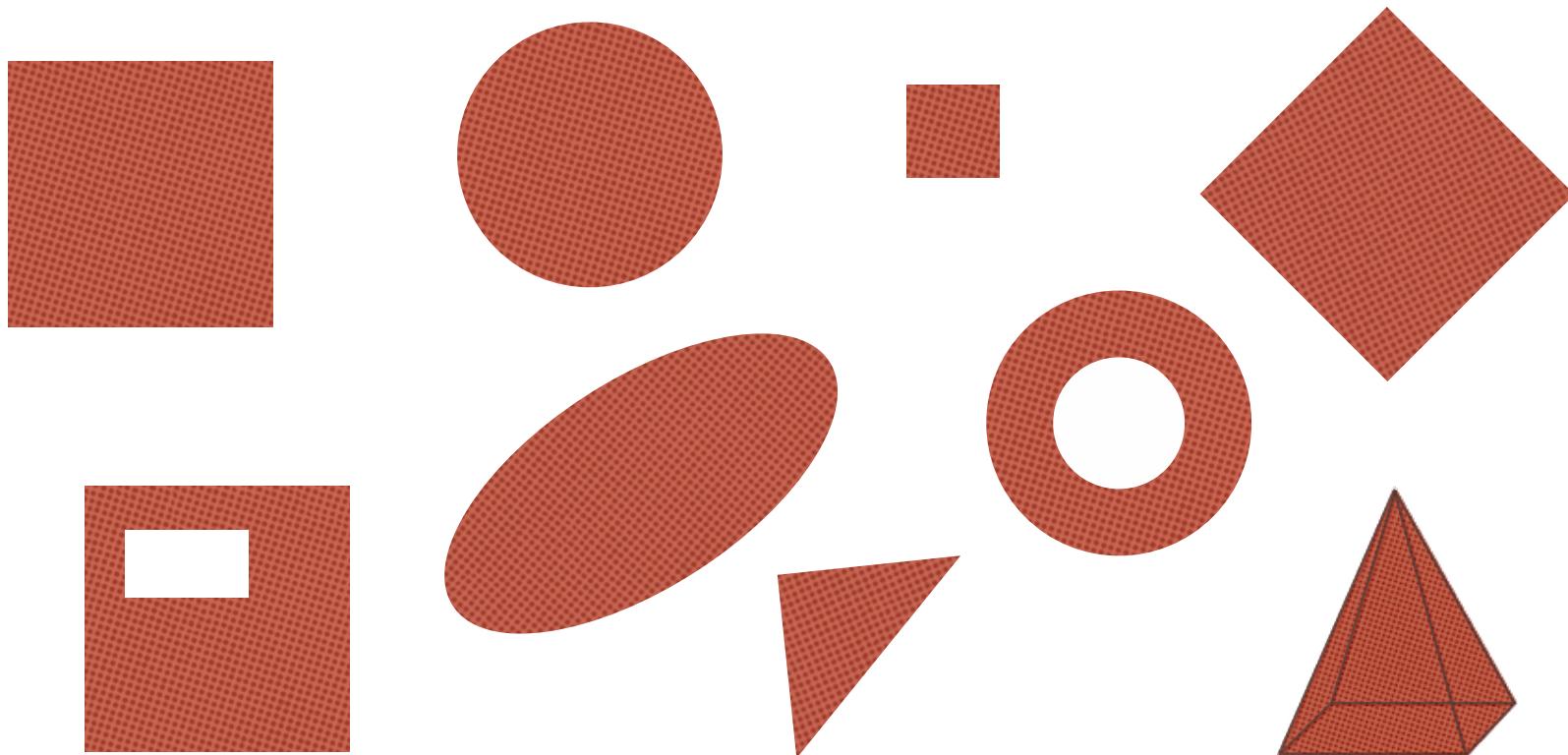


Topological Data Analysis



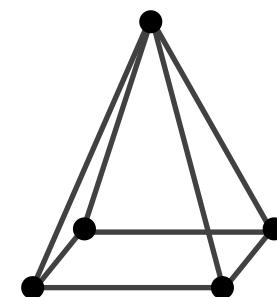
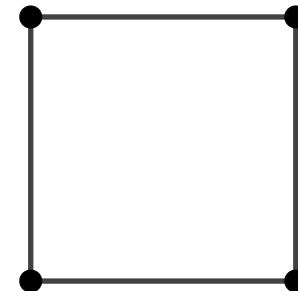
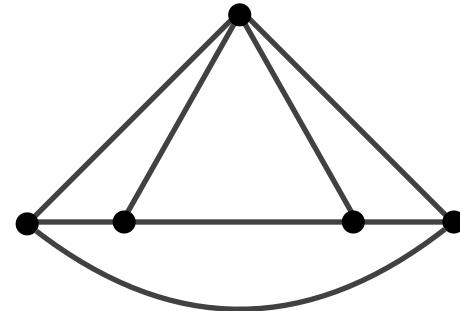
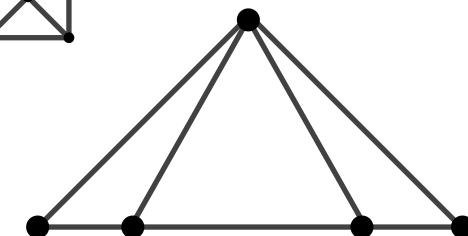
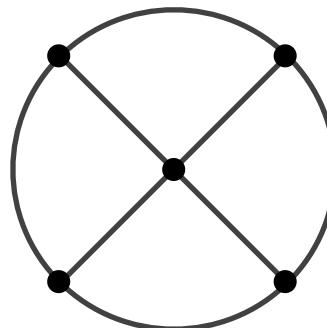
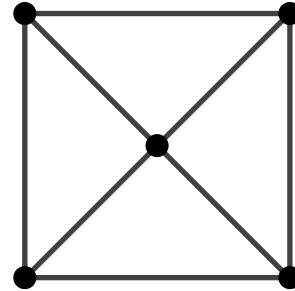
Geometry or Topology?

Which of these domains look similar?



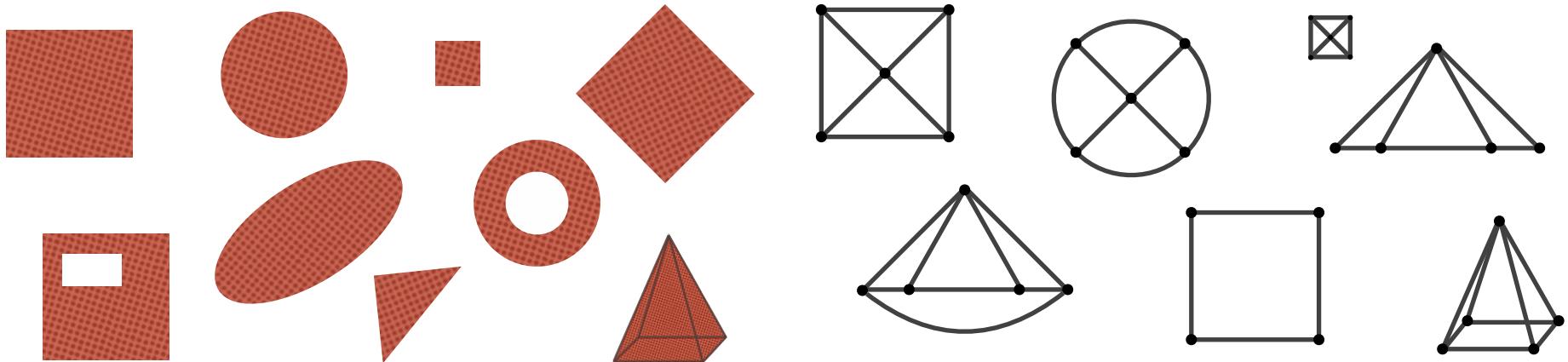
Geometry or Topology?

And what about these ones?



Geometry or Topology?

The answer depends on the *point of view* we adopt

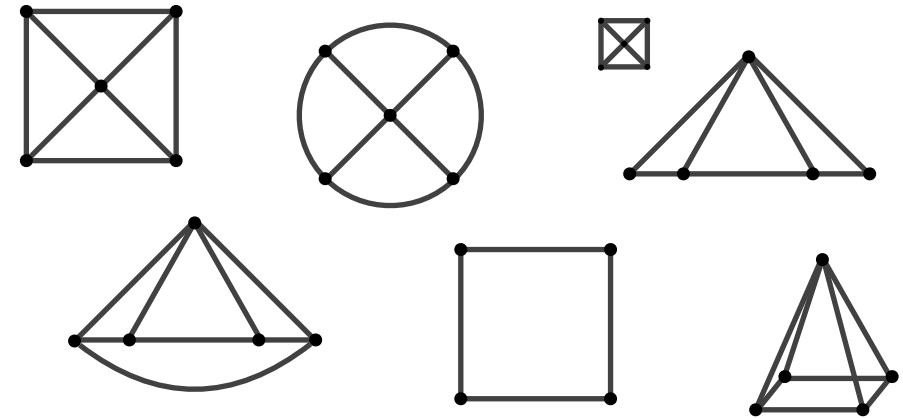
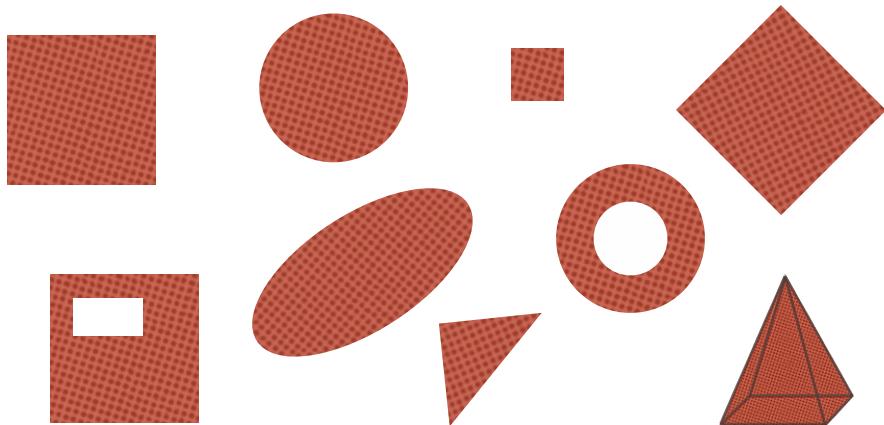


Geometry cares about those properties which **change**
when an object is continuously **deformed**

E.g. length, area, volume, angles, curvature, ...

Geometry or Topology?

The answer depends on the *point of view* we adopt



Topology

~~Geometry~~ cares about those properties which *change*
when an object is continuously *deformed*

E.g. connectivity, orientation, manifoldness, ...

do not

Why Topology?

In life or social sciences, **distances (metric)** are constructed using a notion of **similarity (proximity)**: e.g. distance between faces, gene expression profiles, Jukes-Cantor distance between sequences

We have that:

- ◆ Construction of a distance has ***no theoretical backing***
- ◆ Small distances still represent similarity, but ***long distance comparisons hardly make sense***
- ◆ Distance measurements are ***typically noisy***
- ◆ Physical devices, e.g. human eyes, may ***ignore differences in proximity***

Topology is the crudest way to capture invariants under distortions of distances
(even if, at the presence of noise, one needs topology varied with scales)

Topological Spaces

Definition:

A **topological space** (X, T) is a non-empty set X endowed with a family T , called **topology**, of subsets of X satisfying the following properties:

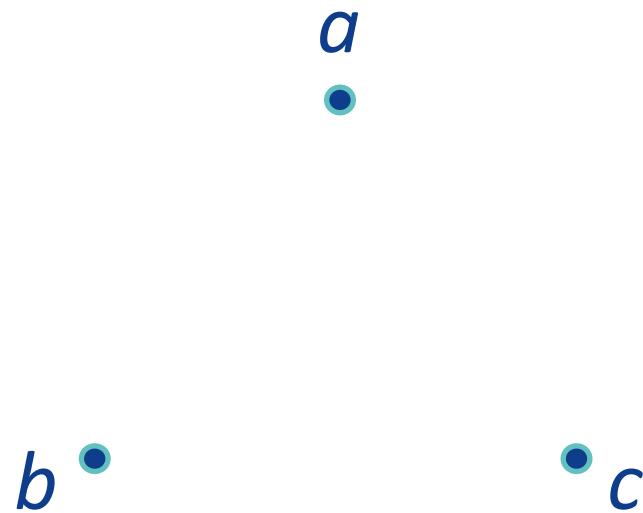
- ◆ X and the **empty set** \emptyset belong to T
- ◆ **Union of any collection** of elements of T is in T
- ◆ **Intersection of any finite collection** of elements of T is in T

A set U in T is called **open set**. A set F such that $X \setminus F$ is in T is called **closed set**.
Dually to the above definition, a topological space can be characterized by defining its closed sets

Topological Spaces

Exercise:

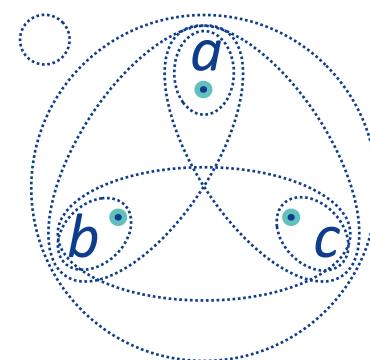
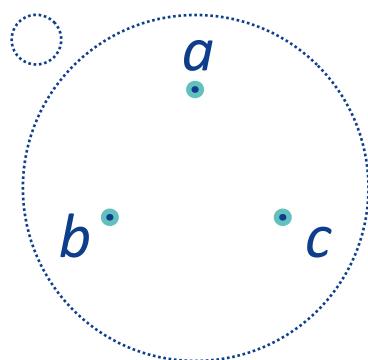
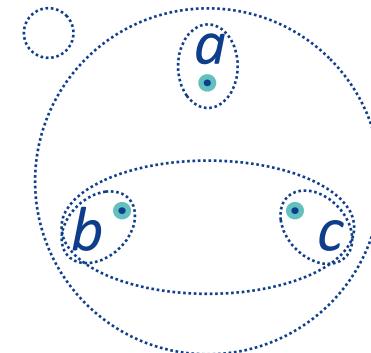
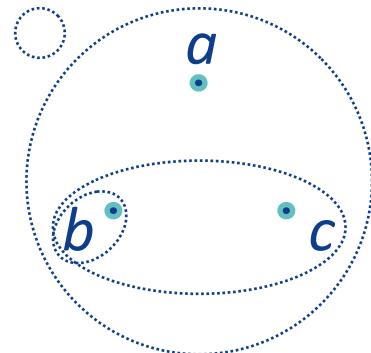
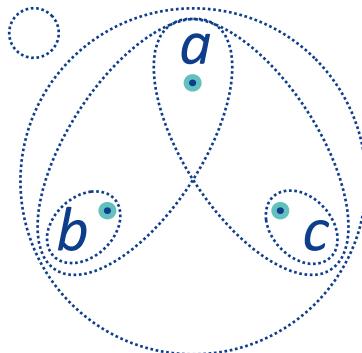
Given the set $X := \{a, b, c\}$, define a topology T for X



Topological Spaces

Exercise:

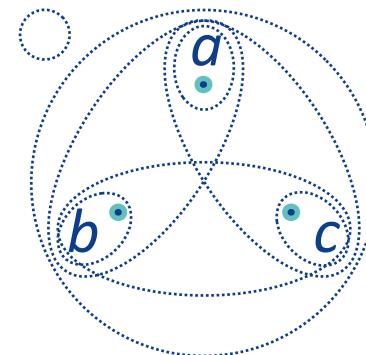
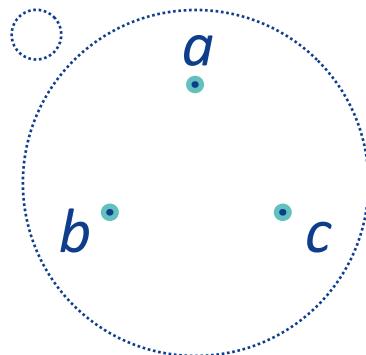
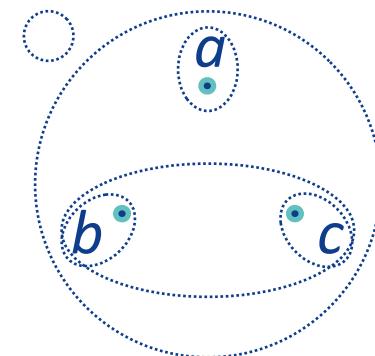
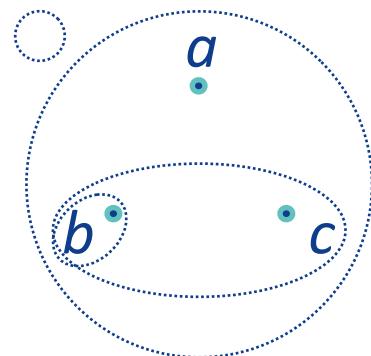
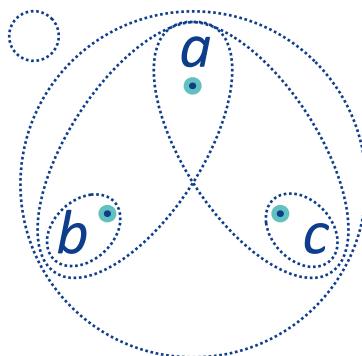
Which of the following families are topologies for X ?



Topological Spaces

Exercise:

Which of the following families are topologies for X ?



Trivial topology:
i.e. $T := \{\emptyset, X\}$

Discrete topology:
i.e. $T := P(X)$

Topological Spaces

Definition:

Let T be a topology of a non-empty set X . A **basis** of T is a family of open sets $\mathcal{B} \subseteq T$ such that ***each open sets of T is union of elements of \mathcal{B}***

Proposition:

Let X be a non-empty set and \mathcal{B} be a family of subsets of X such that:

- ◆ $\bigcup_{B \in \mathcal{B}} B = X$
- ◆ For any $A, B \in \mathcal{B}$, $A \cap B$ is union of elements of \mathcal{B}

Then, ***there exists a (unique) topology*** T of X of which \mathcal{B} is a basis

Metric Spaces as Topological Spaces

Definition:

A **metric space** (X, d) is a non-empty set X on which is defined a function $d: X \times X \rightarrow \mathbb{R}$, called **distance**, such that, for any $x, y, z \in X$:

- ◆ $d(x, y) \geq 0$
- ◆ $d(x, y) = 0$ if and only if $x = y$ *(identity of indiscernibles)*
- ◆ $d(x, y) = d(y, x)$ *(symmetry)*
- ◆ $d(x, z) \leq d(x, y) + d(y, z)$ *(subadditivity or triangle inequality)*

Proposition:

Each **metric space** (X, d) is a **topological space** (X, T) with respect to the topology T having as basis $\mathcal{B} := \{B(x, r) \mid x \in X, r > 0\}$, where

$B(x, r)$ is the **open ball of radius r centered in x** defined as $B(x, r) := \{y \in X \mid d(x, y) < r\}$

Metric Spaces as Topological Spaces

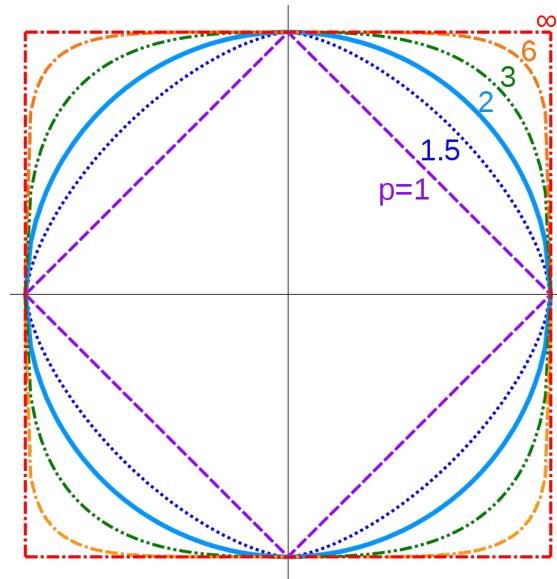
Example:

The *n-dimensional Euclidean space* \mathbb{E}^n is the topological space induced by the metric space (\mathbb{R}^n, d) where d is defined as

$$d(x, y) := \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

For any $p \geq 1$, the *Minkowski distance* d_p induces the same topology on \mathbb{R}^n

$$d_p(x, y) := \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

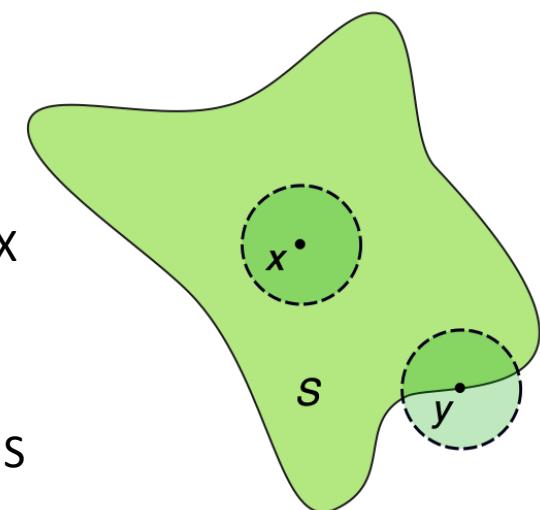


Topological Spaces

Some Basic Notions:

Given a topological space (X, T) , an element x of X , and a subset S of X :

- ◆ A **neighborhood of x** is a subset V of X that includes an open set U containing x (i.e. $x \in U \subseteq V$)
- ◆ The **interior $i(S)$** of S is the union of all subset of S that are open of X
 - ❖ $i(S)$ consists of the elements x of X for which there exists an open neighborhood V of x completely contained in S
- ◆ The **closure $c(S)$** of S is the intersection of all closed sets containing S
 - ❖ $c(S)$ consists of the elements x of X for which every open neighborhood V of x contains a element of S
- ◆ The **boundary $\partial(S)$** of S is the set of elements in the closure of S not belonging to the interior of S (i.e. $\partial(S) = c(S) \setminus i(S)$)
 - ❖ $\partial(S)$ consists of the elements x of X for which every open neighborhood V of x intersects both S and $X \setminus S$



Topological Spaces

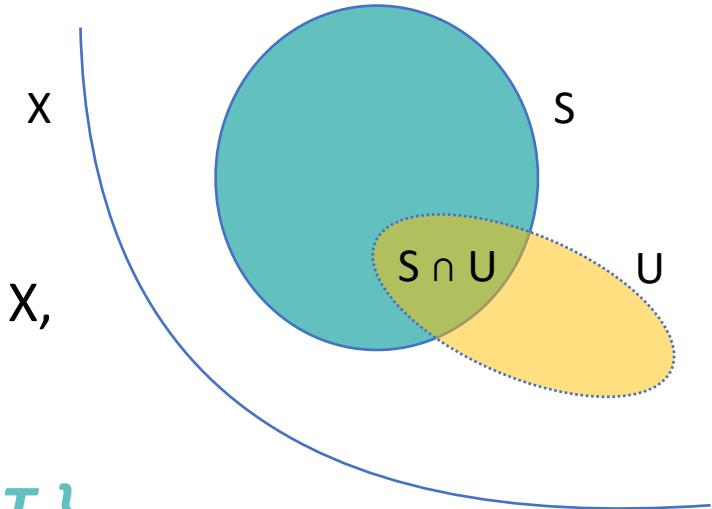
Definition:

Given a topological space (X, T) and a subset S of X ,
the **subspace topology T_S** on S is defined as

$$T_S := \{ S \cap U \mid U \in T \}$$

i.e. a subset of S is an open set of T_S if and only if it is the intersection of S with an open set of X

S equipped with the subspace topology T_S is called a **subspace** of (X, T)

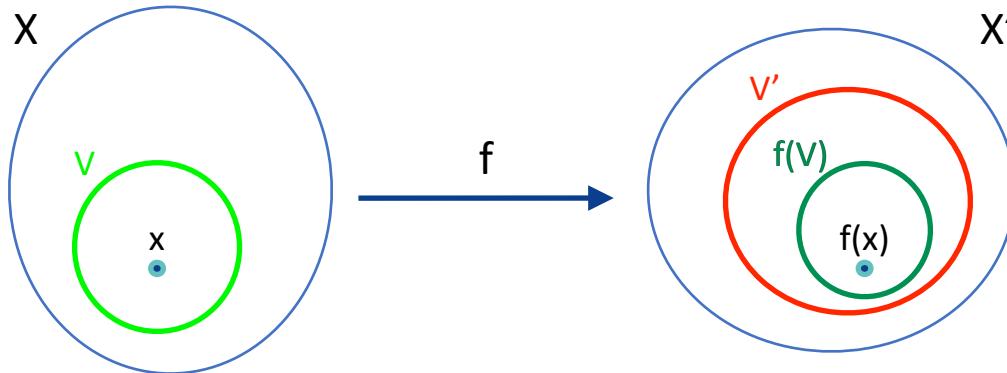


Continuous Functions

Definition:

Given two topological spaces (X, T) and (X', T') , a *function $f: X \rightarrow X'$* is called

- ◆ **Continuous in $x \in X$** if, for each neighborhood V' of $f(x)$ in X' , there exists a neighborhood V of x in X such that $f(V) \subseteq V'$
- ◆ **Continuous** if it is continuous in each element $x \in X$ or, equivalently, if, *for each open set U' of X' , $f^{-1}(U')$ is an open set of X*



Continuous Functions

Exercise:

Let X be a non-empty set X and let T, T' be the discrete and the trivial topologies on X , respectively. Which of the following functions is continuous?

- ◆ the identity map $\text{id}: (X, T) \rightarrow (X, T')$
- ◆ the identity map $\text{id}': (X, T') \rightarrow (X, T)$

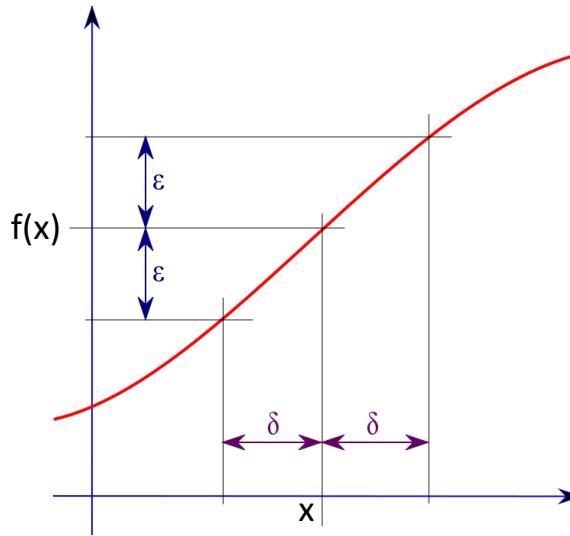
Continuous Functions

Proposition:

Given two metric spaces (X, d) and (X', d') , a function $f: X \rightarrow X'$ is continuous in $x \in X$

if and only if

$\forall \varepsilon > 0 \exists \delta > 0$ such that, for any $y \in X$ with $d(x, y) < \delta$, $d'(f(x), f(y)) < \varepsilon$

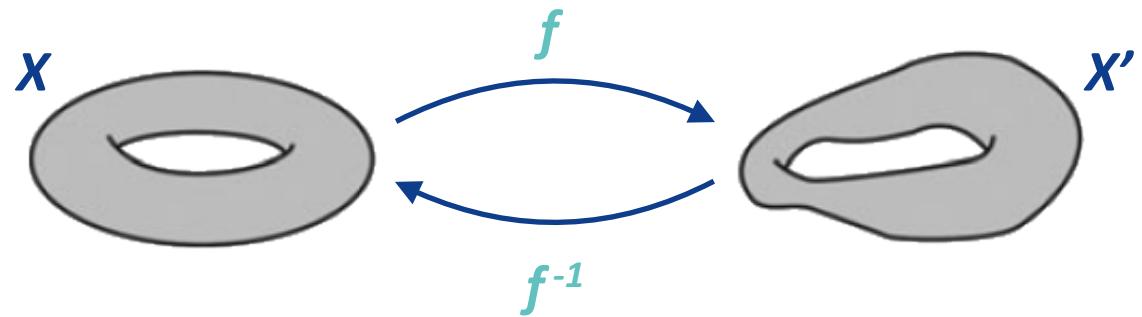


Homeomorphisms

Definition:

Given two topological spaces (X, T) and (X', T') ,
a function $f: X \rightarrow X'$ is called **homeomorphism** if:

- ◆ f is a **bijection**
- ◆ f is **continuous**
- ◆ f^{-1} is **continuous**

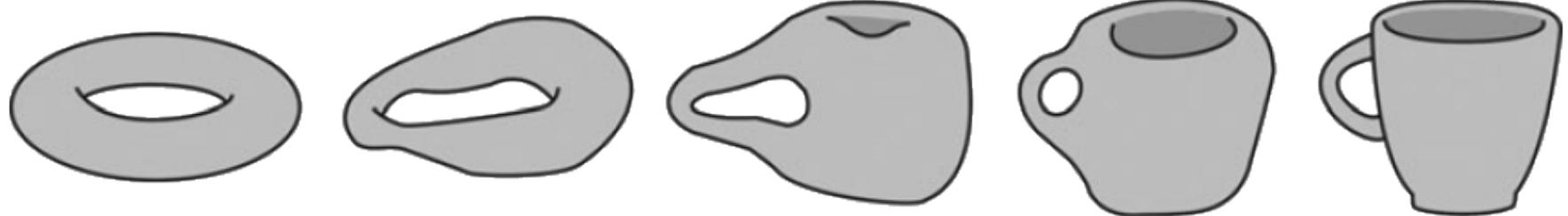


Two topological spaces (X, T) and (X', T') are **homeomorphic** and denoted $X \cong X'$ if there exists a homeomorphism $f: X \rightarrow X'$

Homeomorphisms induce an **equivalence relation** of topological spaces partitioning them into equivalence classes

Homeomorphisms

Intuitively:



The notion of homeomorphism captures the idea of continuous deformation



\cong



Homeomorphisms

Intuitively:

One can:

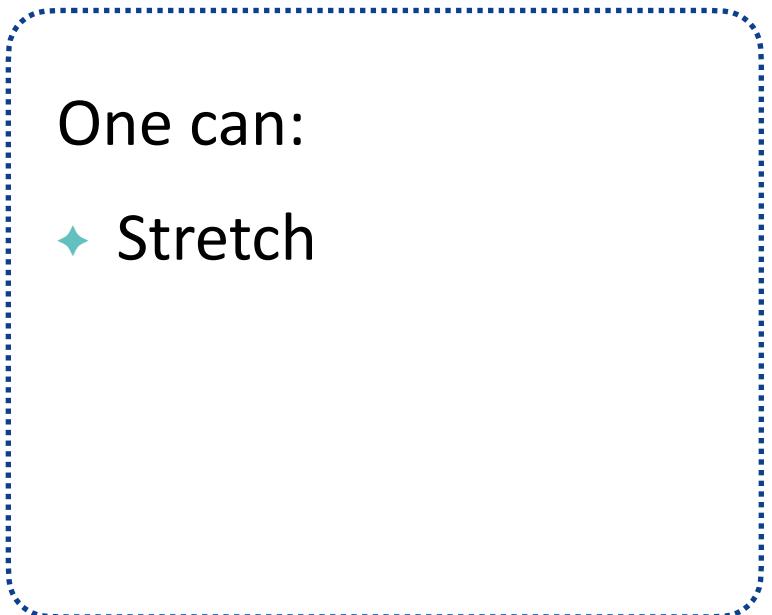


Homeomorphisms

Intuitively:

One can:

- ◆ Stretch

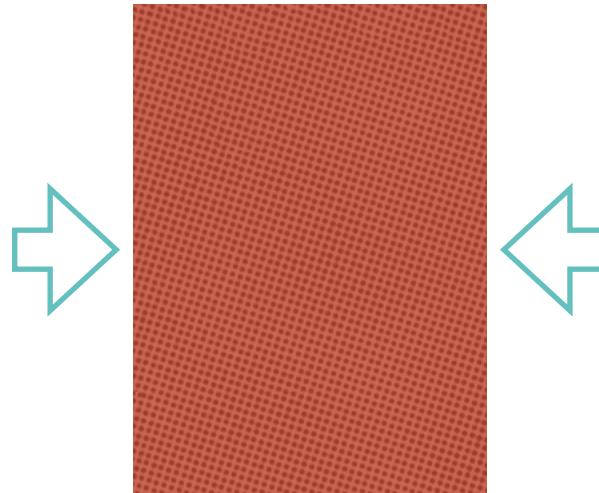


Homeomorphisms

Intuitively:

One can:

- ◆ Stretch
- ◆ Compress



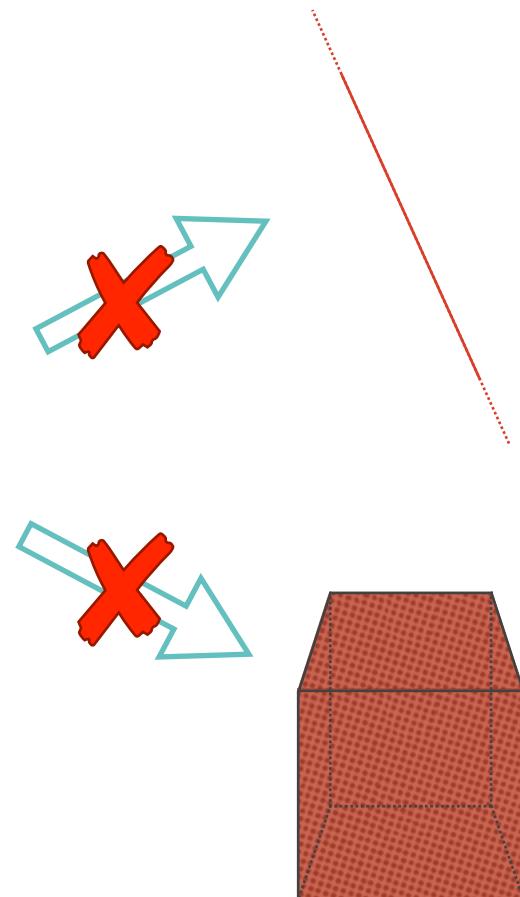
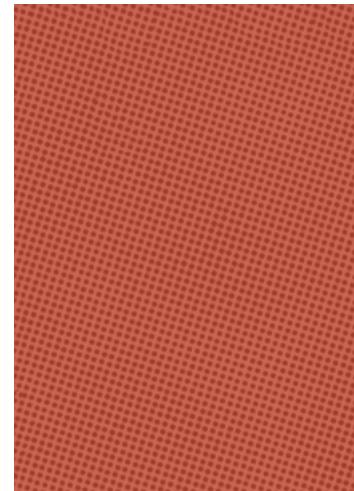
Homeomorphisms

Intuitively:

One can:

- ◆ Stretch
- ◆ Compress

But not too much!



Homeomorphisms

Intuitively:

Moreover:

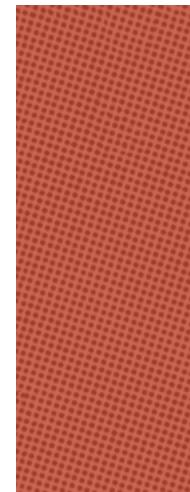
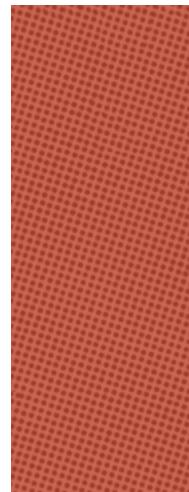
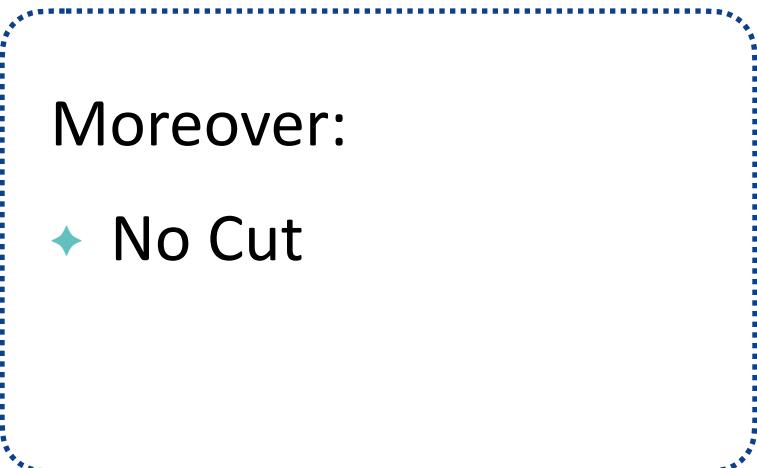


Homeomorphisms

Intuitively:

Moreover:

- ◆ No Cut



Homeomorphisms

Intuitively:

Moreover:

- ◆ No Cut
- ◆ No Glue

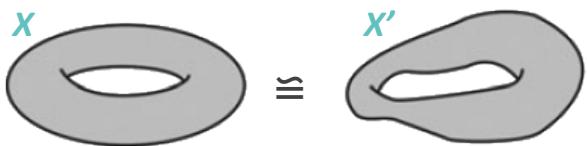


Topological Invariants

Definition:

I is a **topological invariant** if, given two topological spaces (X, T) and (X', T') ,

X is homeomorphic to X'



X and X' have the same
topological invariant

$$I(X) = I(X')$$

Some classical topological invariants:

- ◆ *Connectedness*
- ◆ *Compactness*
- ◆ *Manifoldness*

- ◆ *Orientability*
- ◆ *Euler characteristic*
- ◆ *Homology*
- ◆ *Homotopy*

Topological Invariants

Question:

Is there a “perfect” topological invariant I such that

$X \cong X'$ if and only if $I(X) = I(X')$?

Topological Invariants

Question:

Is there a “perfect” topological invariant I such that

$X \cong X'$ if and only if $I(X) = I(X')$?

Let us **simplify the question** and let focus on:

- ◆ Considering a specific topological invariant I (e.g. the **homology**)
- ◆ Completely characterizing just the **spheres** $S^n := \{x \in \mathbb{R}^n : |x| = 1\}$

The above question turns into the following:

If X and S^n have the same homology, then $X \cong S^n$?

Topological Invariants

Question:

Is there a “perfect” topological invariant I such that

$X \cong X'$ if and only if $I(X) = I(X')$?

Let us **simplify the question** and let focus on:

- ◆ Considering a specific topological invariant I (e.g. the **homology**)
- ◆ Completely characterizing just the **spheres** $S^n := \{x \in \mathbb{R}^n : |x| = 1\}$

The above question turns into the following:

If X and S^n have the same homology, then $X \cong S^n$?

NO

Topological Invariants

But:

*Replacing homology with **homotopy**, the answer is positive!*

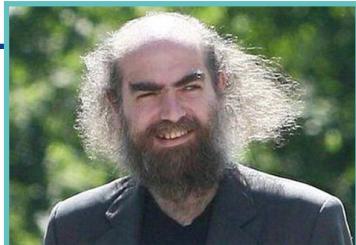
Topological Invariants

But:

*Replacing homology with **homotopy**, the answer is positive!*

Poincaré Conjecture (3rd Millennium Prize Problem):

*If X is a closed n -manifold **homotopy equivalent** to S^n , then $X \cong S^n$*



Proven by Grigori Perelman in 2003

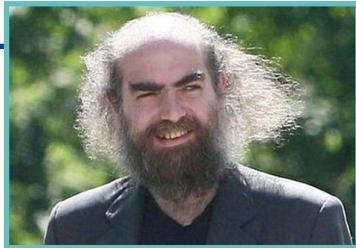
Topological Invariants

But:

*Replacing homology with **homotopy**, the answer is positive!*

Poincaré Conjecture (3rd Millennium Prize Problem):

*If X is a closed n -manifold **homotopy equivalent** to S^n , then $X \cong S^n$*



Proven by Grigori Perelman in 2003

So:

Why we will mainly focus on homology rather than homotopy?

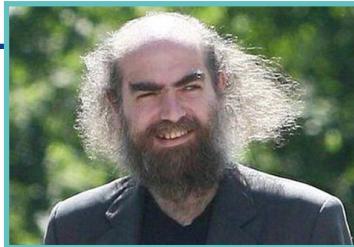
Topological Invariants

But:

*Replacing homology with **homotopy**, the answer is positive!*

Poincaré Conjecture (3rd Millennium Prize Problem):

*If X is a closed n -manifold **homotopy equivalent** to S^n , then $X \cong S^n$*



Proven by Grigori Perelman in 2003

So:

Why we will mainly focus on homology rather than homotopy?

*Because, in practice, computing homotopy groups is **nearly impossible**!*

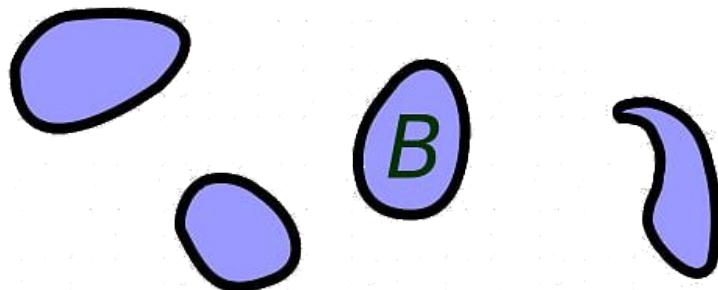
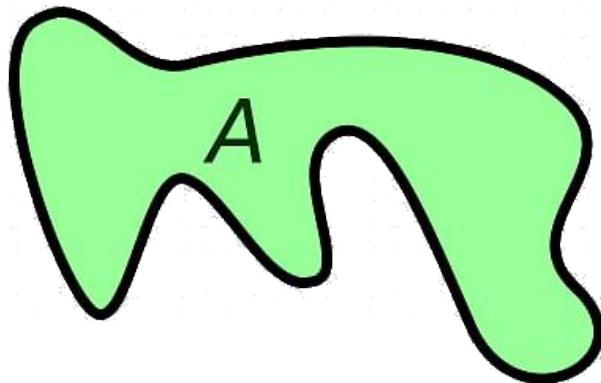
Connected Spaces

Definition:

A topological space (X, T) is **connected** if, given any two disjoint open sets U and V s.t.

$$X = U \cup V, \text{ then } U = \emptyset \text{ or } V = \emptyset$$

i.e. X cannot be written as the union of two non-empty disjoint open sets of X



A **connected component** of X is a maximal connected subset of X

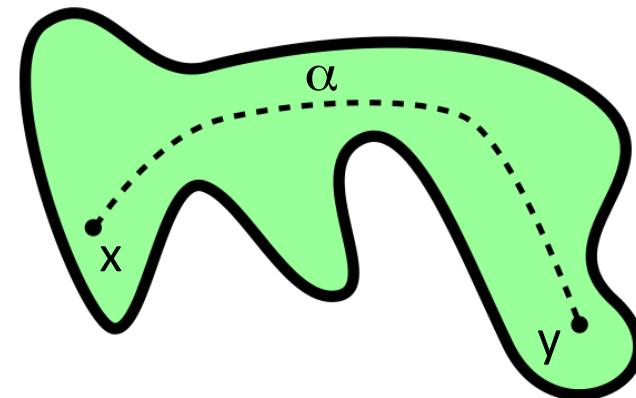
Connected Spaces

Definition:

A topological space (X, T) is **path-connected** if, for every pair $x, y \in X$, there exists a **continuous map** $\alpha: [0,1] \rightarrow X$ such that $\alpha(0) = x$ and $\alpha(1) = y$

The map α is called a **path** from x to y

A **path-connected component** of X is a maximal path-connected subset of X



Proposition:

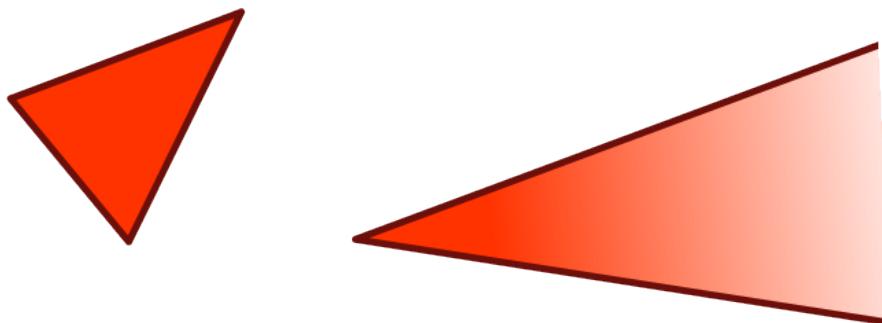
If X is path-connected, then X is connected. The converse is not true

Compact Spaces

Definition:

An **open cover** of a topological space (X, T) is a collection C of open sets U_i of X whose union is the whole space X , i.e. $X \supseteq \bigcup_{i \in I} U_i$. A **subcover** of C is a subset of C that still covers X .

A topological space (X, T) is called **compact** if any of its open covers has a **finite subcover**.



Heine-Borel Theorem:

A subset S of the Euclidean space \mathbb{E}^n is **compact if and only if** S is **closed** and **bounded** (i.e. there exists $r > 0$ such that, for any $x, y \in S$, $d(x, y) < r$)

Manifolds

Definitions:

A topological space (X, T) is called

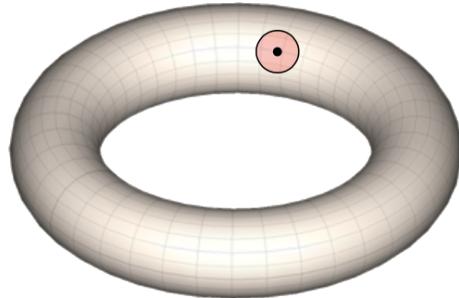
- ◆ **locally homeomorphic** to \mathbb{E}^n if every element $x \in X$ has a neighborhood which is homeomorphic to the n-dimensional Euclidean space \mathbb{E}^n
- ◆ **Hausdorff** if any pair of distinct elements $x, y \in X$ admits disjoint neighborhoods (any metric space and so any subspace of an Euclidean space is Hausdorff)

A **(topological) n-manifold** M is a **Hausdorff** space that is **locally homeomorphic** to the n-dimensional Euclidean space \mathbb{E}^n

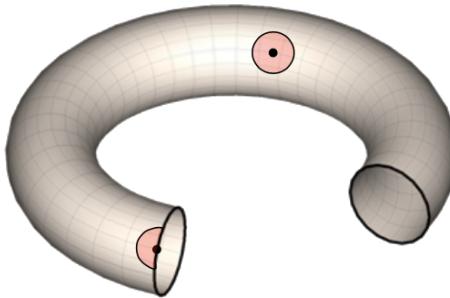
A **(topological) n-manifold with boundary** M is a **Hausdorff** space in which every element has a **neighborhood homeomorphic** to the n-dimensional **Euclidean space** \mathbb{E}^n or to the n-dimensional **Euclidean half-space** $H^n := \{x \in \mathbb{R}^n \mid x_n \geq 0\}$

Manifolds

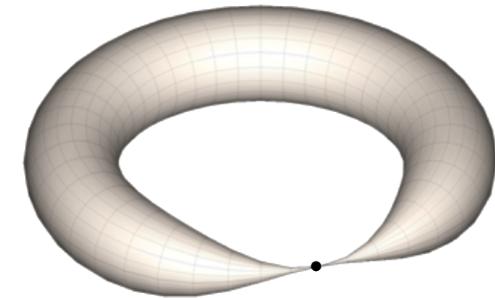
Examples:



manifold

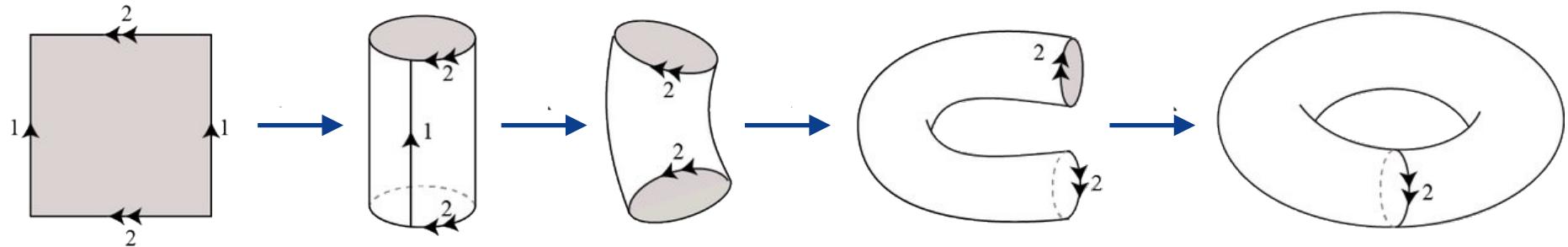


manifold with boundary



non-manifold

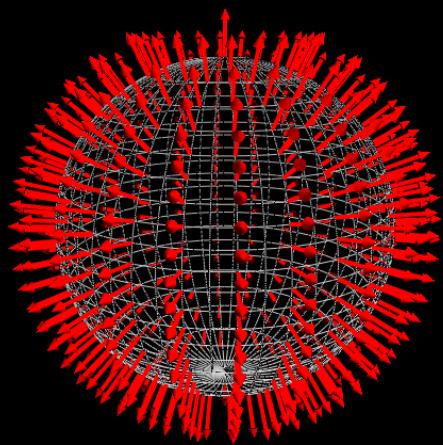
Recall that a **torus** can be built from a **unit square** by the following construction



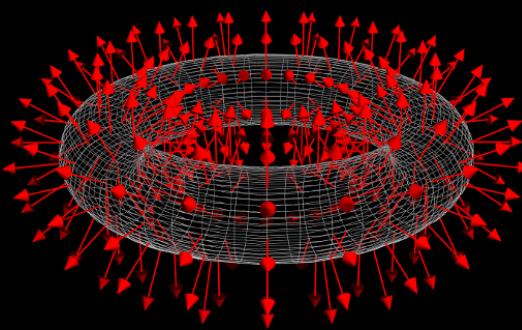
Orientable Surfaces

Definition:

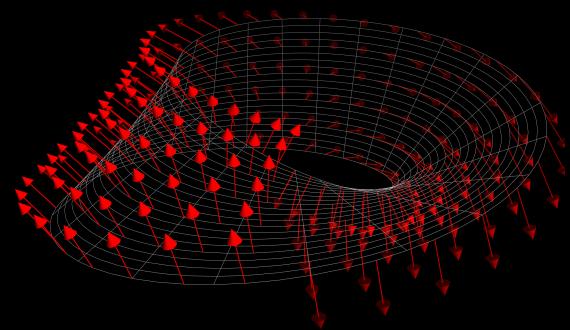
A *surface* (i.e. a topological 2-manifold with or without boundary) S is called *orientable* if it is possible to make a *consistent choice of surface normal vector* at every point of S



orientable



orientable

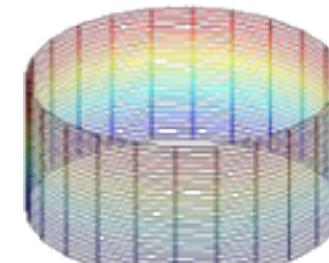
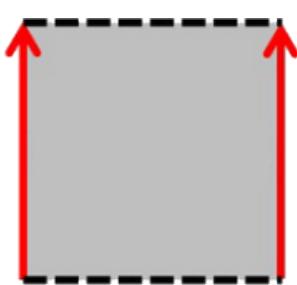


non-orientable

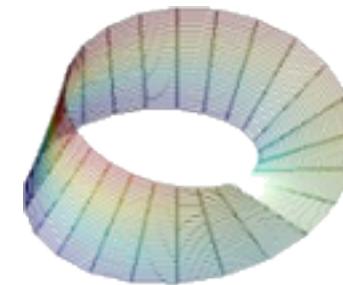
Orientable Surfaces

Remark:

As for the *torus* and the *cylinder*,
the *Möbius strip* can be built from a unit square via edge identification



cylinder



Möbius strip

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:

- ◆ **Intro to (Algebraic) Topology:**
 - ❖ E. Sernesi. *Geometria 2*. Bollati Boringhieri, Torino, 1994.
 - ❖ A. Hatcher. *Algebraic topology*. 2002.