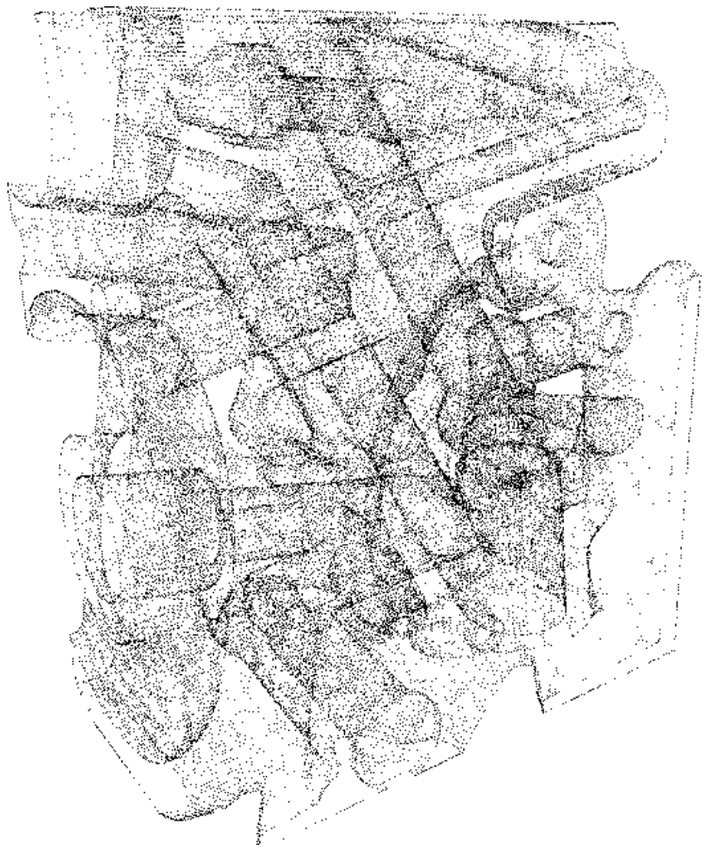
Barcelona, June 2016

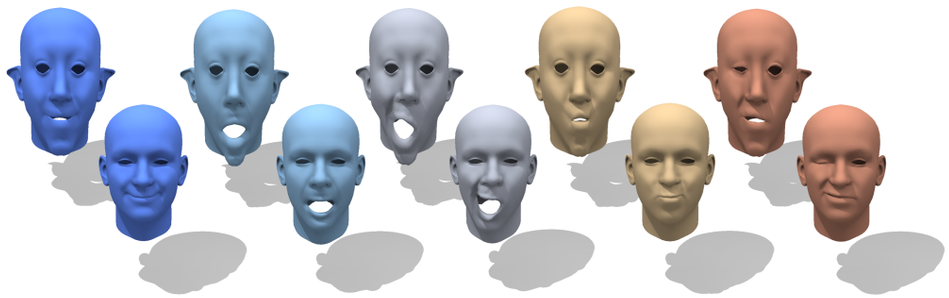
An introduction to Topological Data Analysis

Frédéric Chazal and Bertrand Michel

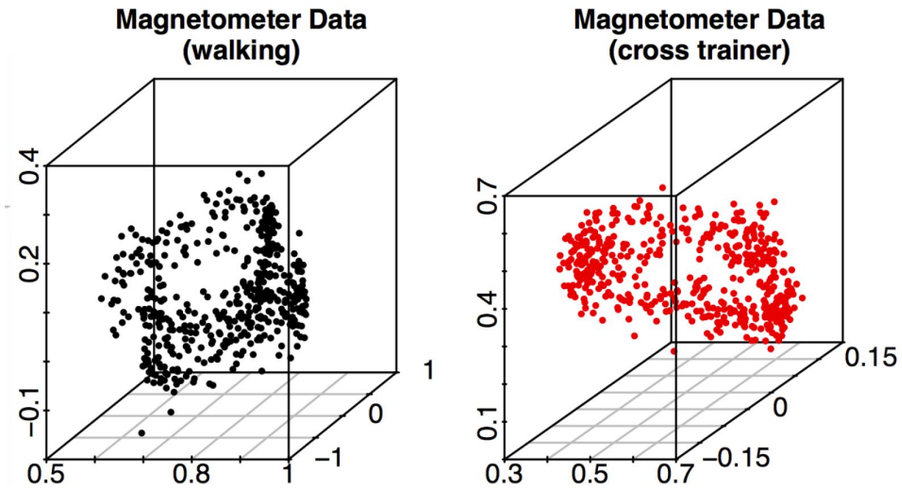
Introduction

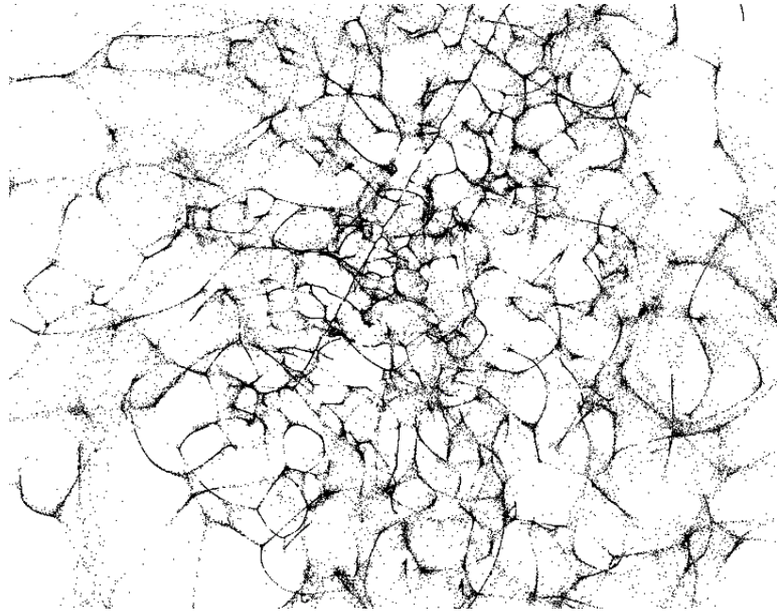


[Scanned 3D object]



[Shape database]

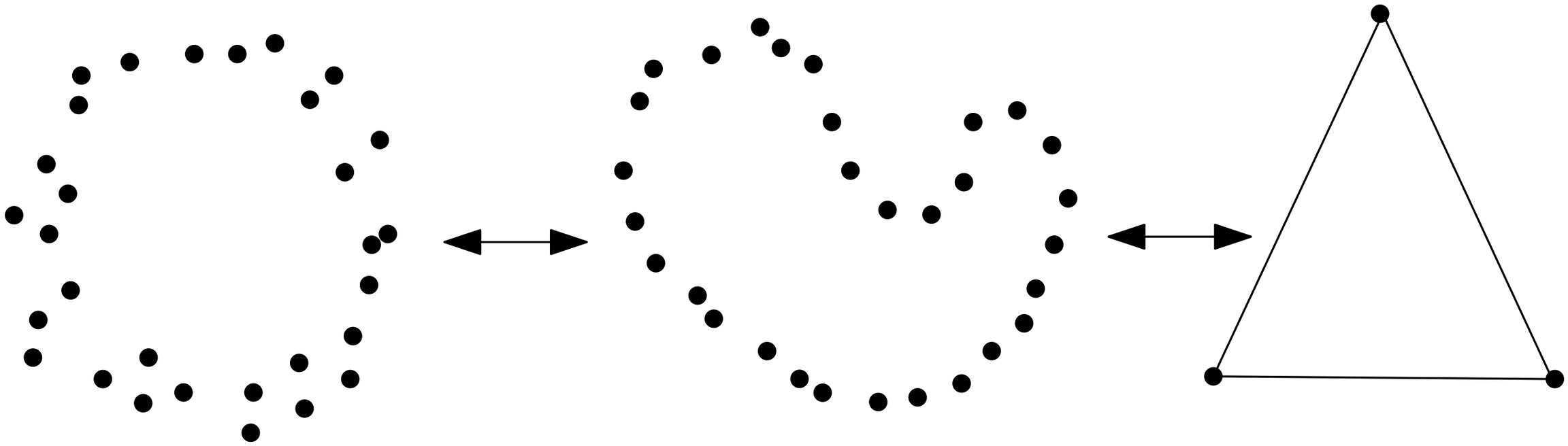




[Galaxies data]

- Data often come as (sampling of) metric spaces or sets/spaces endowed with a similarity measure with, possibly complex, topological/geometric structure.
- Data carrying geometric information are becoming high dimensional.
- Topological Data Analysis (TDA):
  - infer relevant topological and geometric features of these spaces.
  - take advantage of topol./geom. information for further processing of data (classification, recognition, learning, clustering, parametrization...).

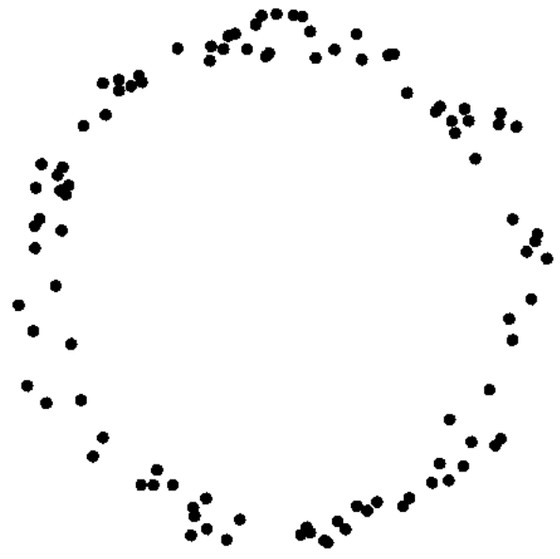
Why is topology interesting for data analysis?

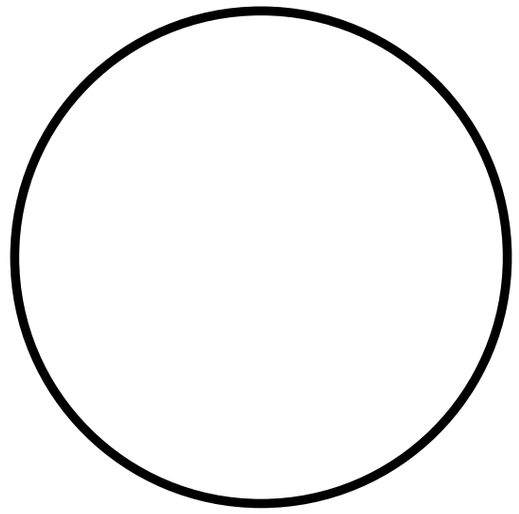


- **Coordinate invariance:** topological features/invariants do not rely on any coordinate system. $\Rightarrow$ no need to have data with coordinate or to embed data in spaces with coordinates... But the metric (distance/similarity between data points) is important.
- **Deformation invariance:** topological features are invariant under homeomorphism.
- **Compressed representation:** Topology offer a set of tools to summarize and represent the data in compact ways while preserving its global topological structure.

Challenges and goals

Problem(s):
- how to visualize the topological structure of data?
- how to compare topological properties (invariants) of close shapes/data sets?





- Challenges and goals:
  $\rightarrow$ no direct access to topological/geometric information: need of intermediate constructions (simplicial complexes);
  $\rightarrow$ distinguish topological "signal" from noise;
  $\rightarrow$ topological information may be multiscale;
  $\rightarrow$ statistical analysis of topological information.