Topological Data Analysis: The New Frontier of Data Science

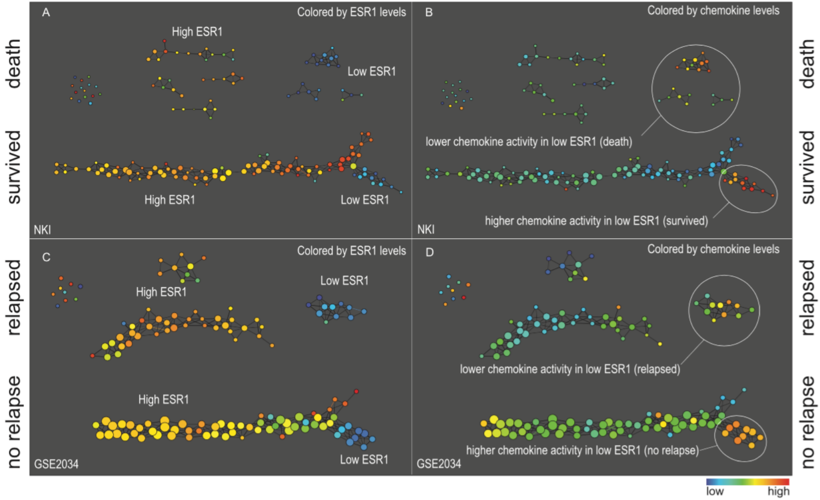
ATEE Plan

Kunlin

Topological Data Analysis

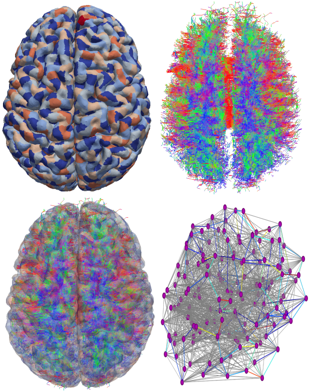
- A marriage: math and computer science
- Topological data analysis is cool: many data applications!
- Many great and fun people are players in this field: mathematicians, computer scientists, statisticians...
- Interdisciplinary: CS, math (algebraic topology, differential topology, i.e. Homology, Morse theory), statics (machine learning, manifold learning), electrical engineering (sensor networks), physics (universe)
- It is young (15+ years), and a lot of open problems, that is, challenges and opportunities! (Imaging the field of computational geometry at its infancy...)
- The researchers are only in their 2nd generation (approximately): room to grow!
- Topological data analysis and visualization is inseparable

Market/Gene Segmentation



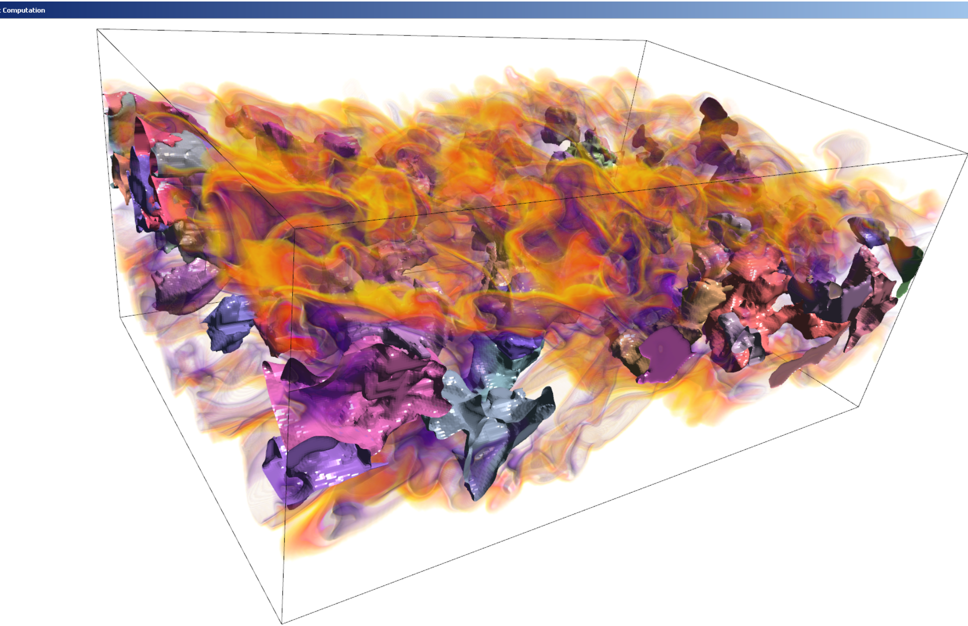
[P. Y. Lum et. al 2013]

Brain Networks

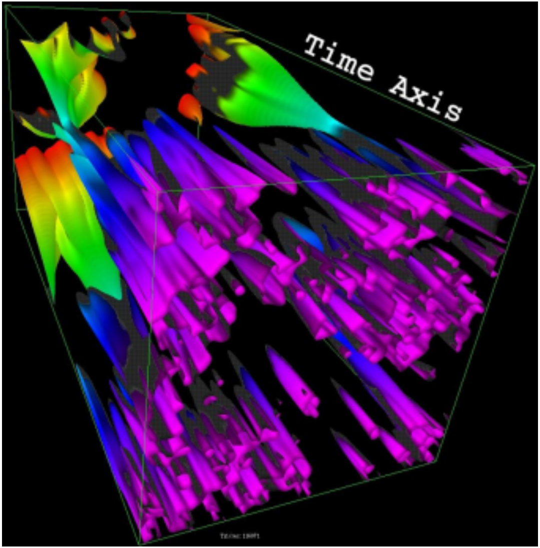


[Wong et. al 2016]

Combustion simulation

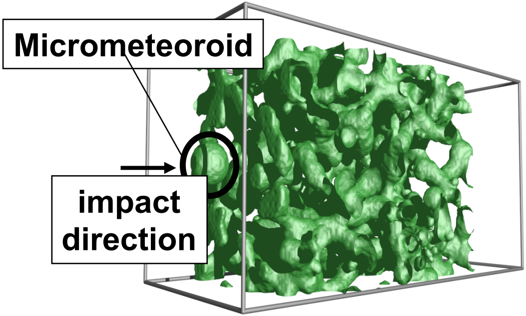


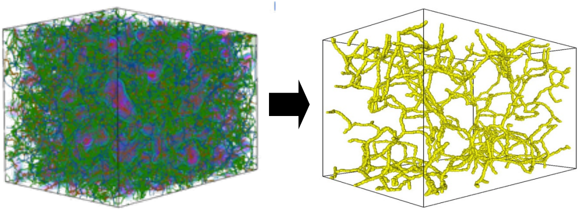
Tracking 2D Combustion



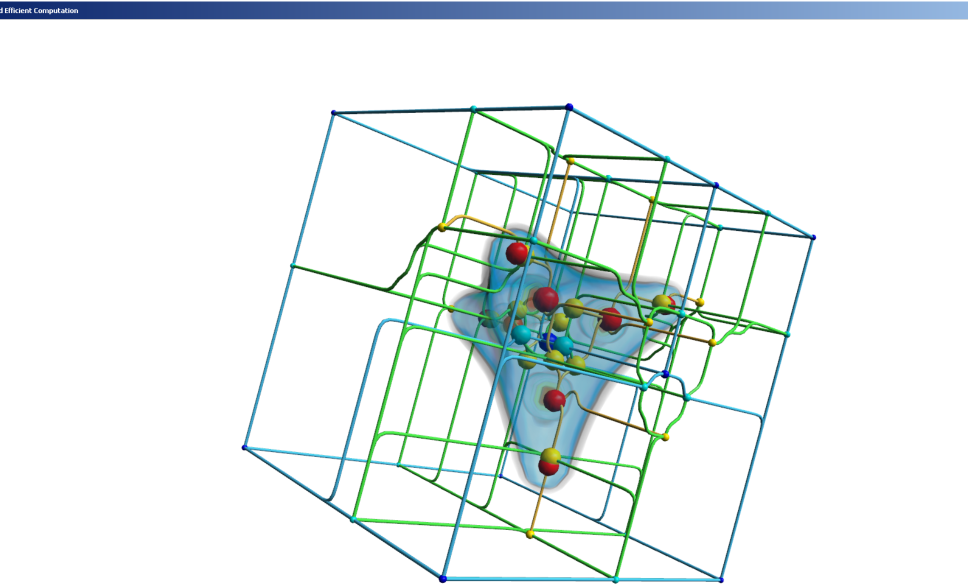
Material science

Quantitative Analysis of the Impact of a Micrometeoroid in a Porous Medium; reconstruction the structure of porous medium



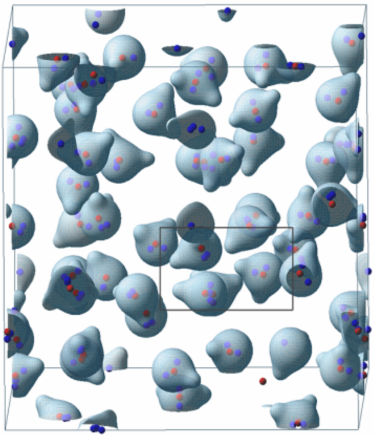


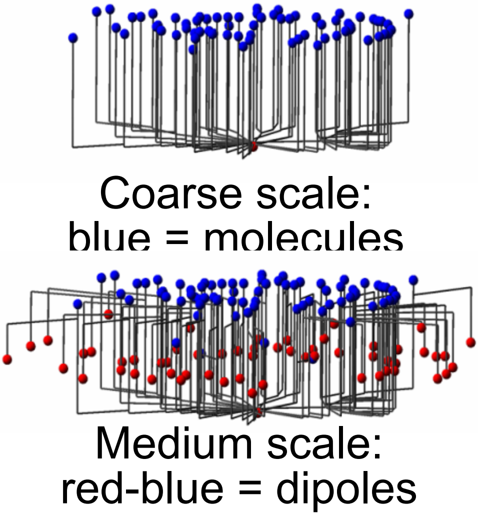
Chemical compound: $C4H4$



Molecular dynamics

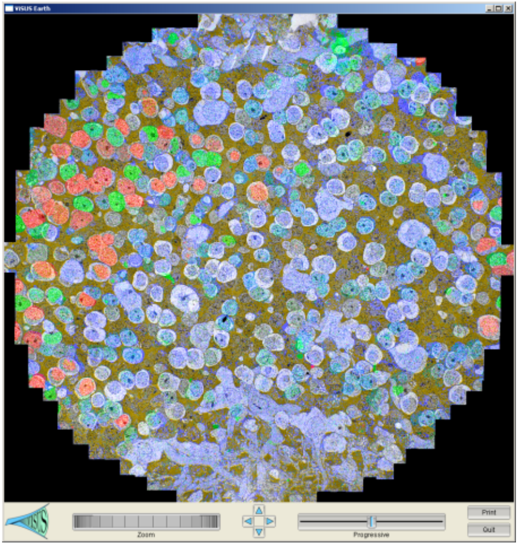
Molecular dynamics simulation (left) with abstract graph representation of its features at two scales (right)





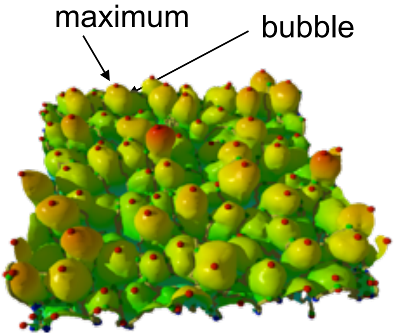
Retinal connectome

A connectome is a comprehensive map of neural connections in the brain [wiki]



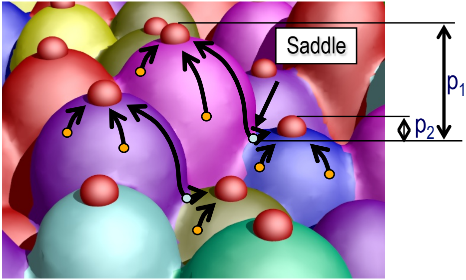
Case study: feature definition

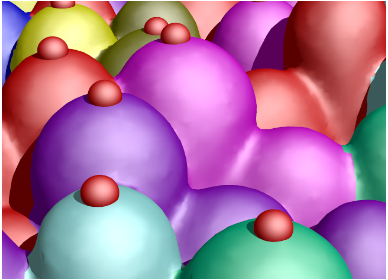
Analyze high-resolution Rayleigh Taylor instability simulations



Case study: persistence simplification

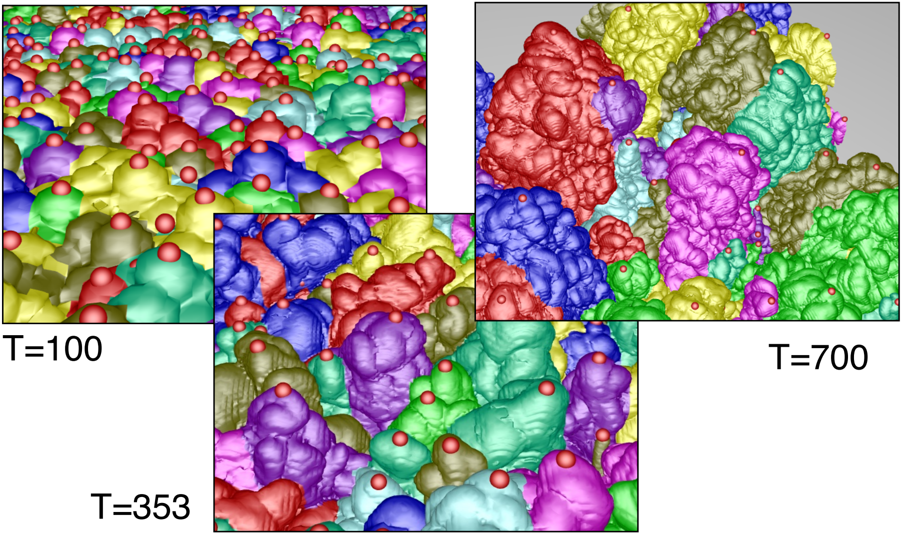
Analyze high-resolution Rayleigh Taylor instability simulations





Case study: robust segmentation

The segmentation method is robust from early mixing to late turbulence



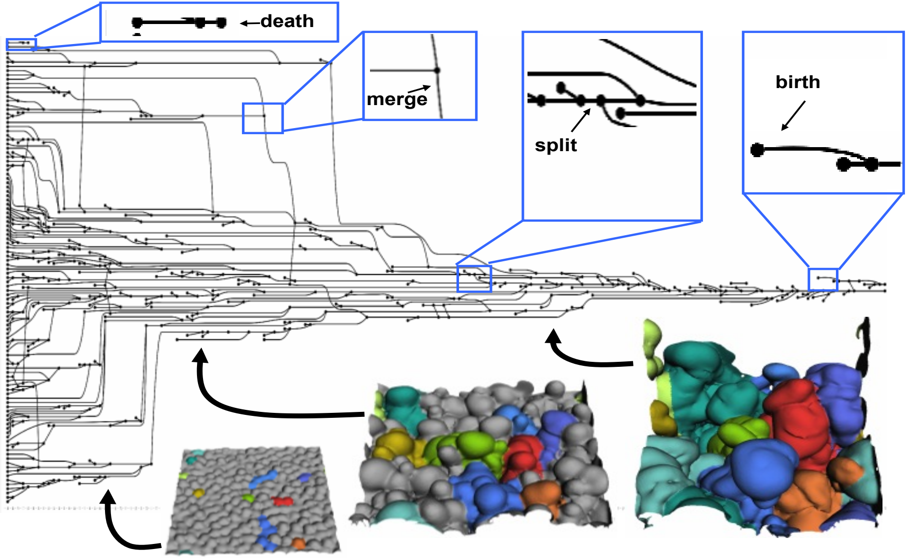
T=100

T=353

T=700

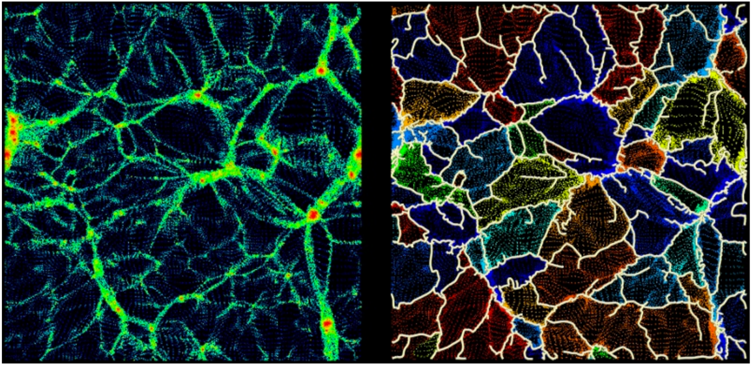
Case study: event characterization

We characterize events that occur in the mixing process



Study the universe!

TDA+ASTRONEMY
POTENTIALS



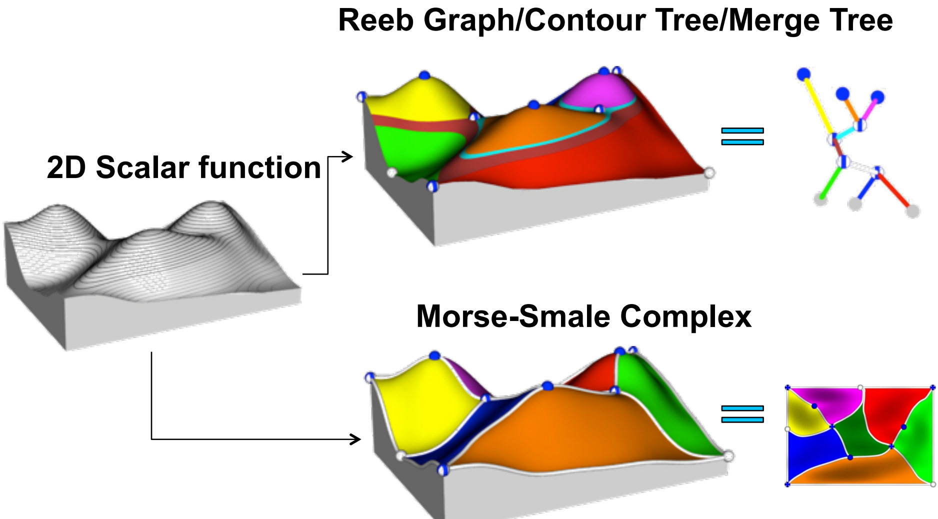
FILAMENTS STRUCTURE T. SOUSBIE, *DISPERSE*

A really old joke...

Who thinks the coffee mug and a donut is the same?

Key development in topological data analysis (TDA)

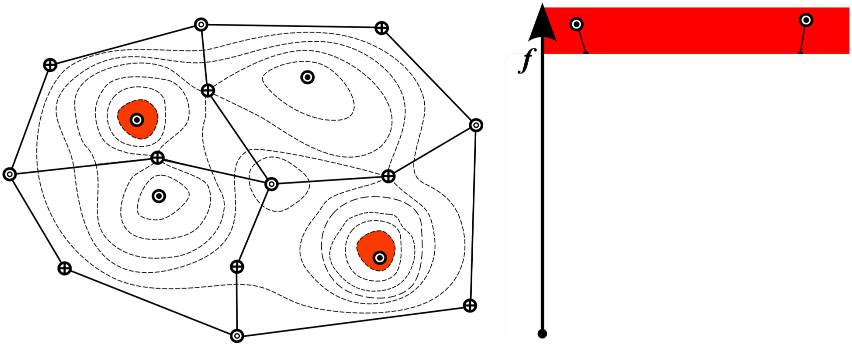
1. Abstraction of the data: topological structures and their combinatorial representations
2. Seperate features from noise: persistent homology



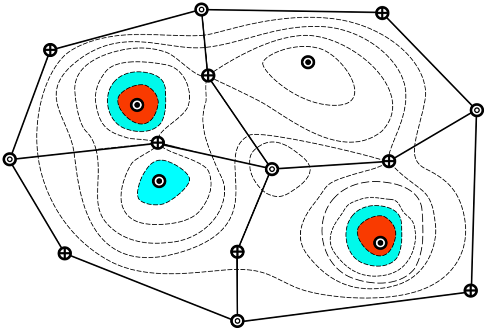
Contour tree

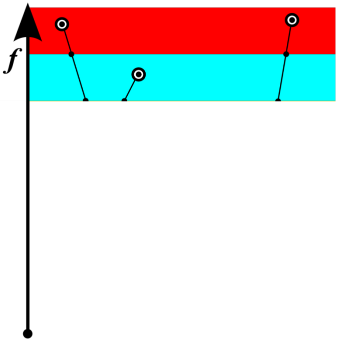
[M. Kreveld, R. Oostrumy, C. Bajaj, V. Pascucci, D. Schikore. Contour Trees and Small Seed Sets for Isosurface Traversal. 1997]
[H. Carr,J. Snoeyink, U. Axen, Computing Contour Trees in All Dimensions, 2001]

Contour tree

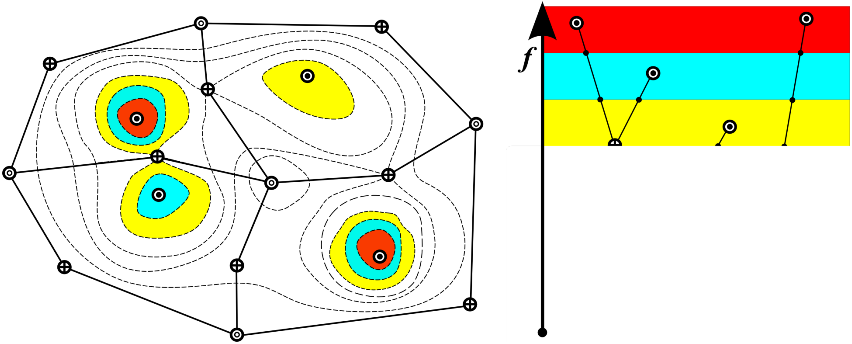


Contour tree

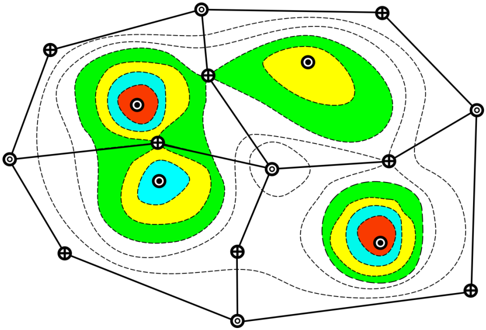


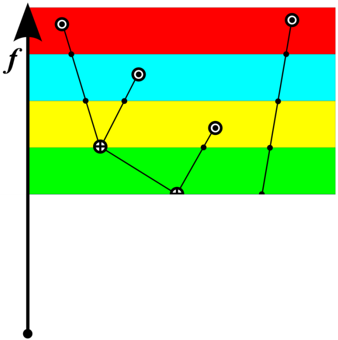


Contour tree

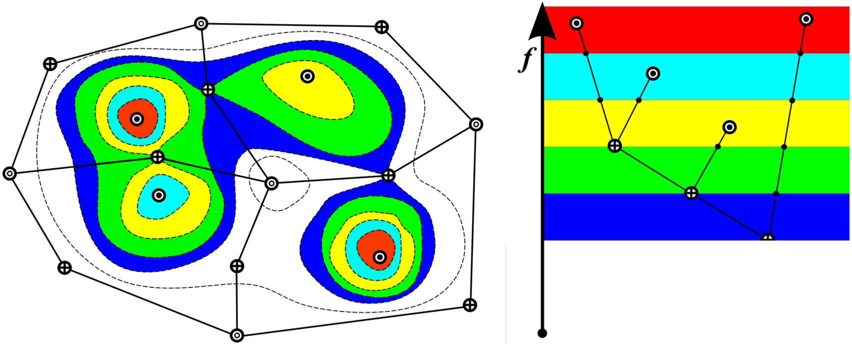


Contour tree

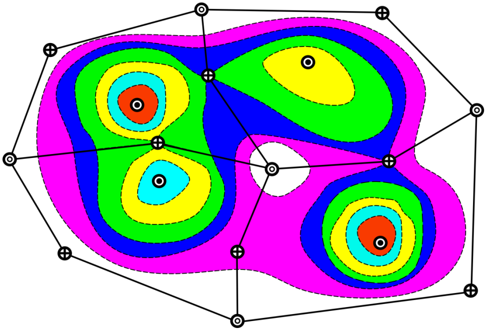


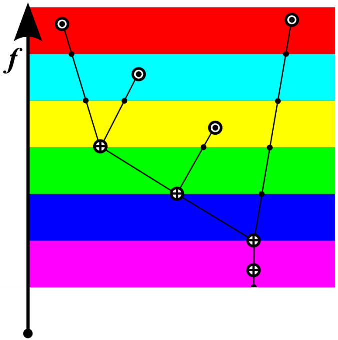


Contour tree

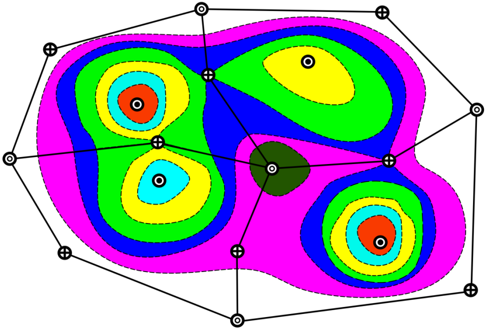


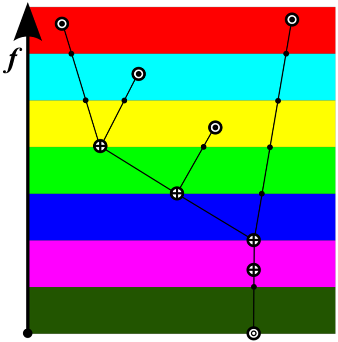
Contour tree





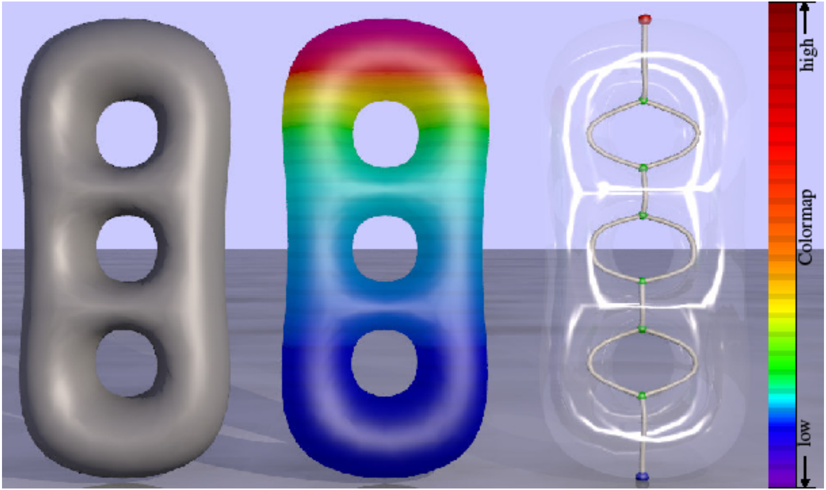
Contour tree





Reeb graph

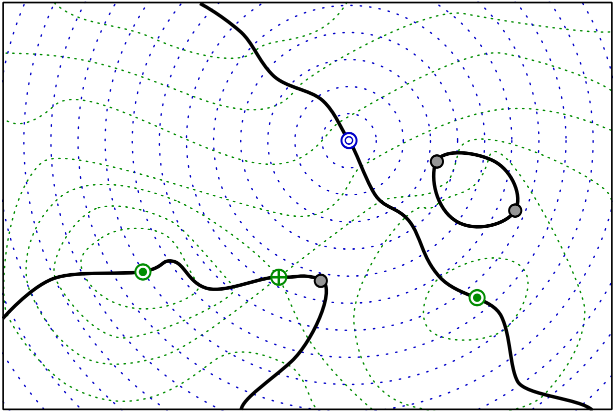
Graph obtained by continuos contraction of all the contours in a scalar field, where each contour is collapsed to a distinct point.



[K. Cole-McLaughlin, H. Edelsbrunner, J. Harer, V. Natarajan and V. Pascucci. Loops in Reeb Graphs of 2-Manifolds. 2004]

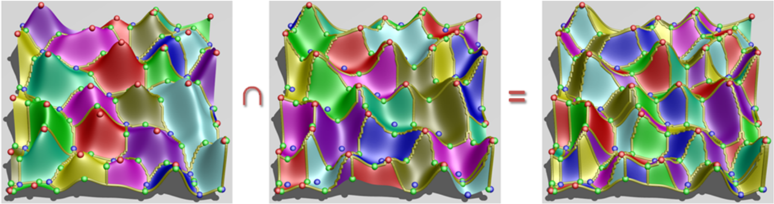
Jacobi Set

[H. Edelsbrunner and J. Harer. Jacobi sets of multiple Morse functions. 2002]



Morse-Smale complex

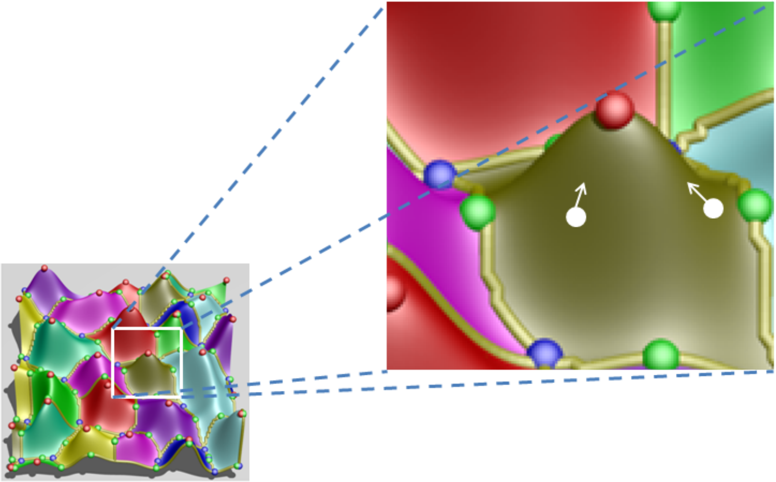
[H. Edelsbrunner, J. Harer and A. Zomorodian. Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds. 2003]
[H. Edelsbrunner, J. Harer, V. Natarajan and V. Pascucci. Morse-Smale complexes for piecewise linear 3-manifolds. 2003]
A partition of the data into monotonic regions



Ascending Manifolds Descending Manifolds Morse-Smale Complex

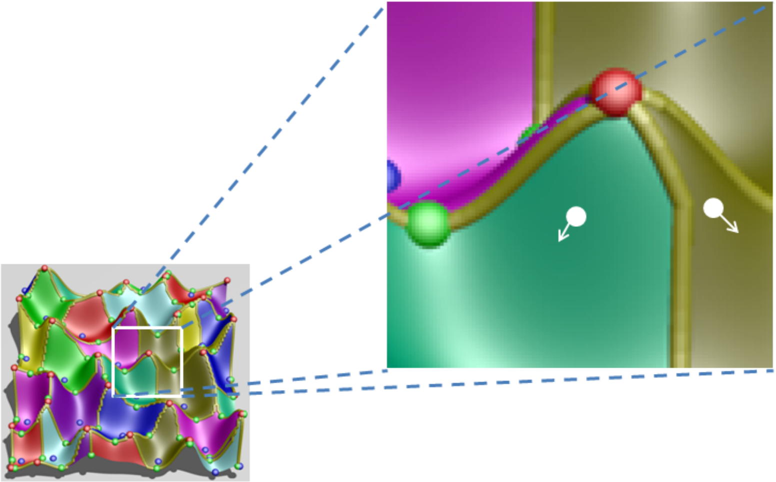
Ascending Manifolds

Compute steepest ascent gradient from each point in dataset



Descending Manifolds

Compute steepest descent gradient from each point in dataset



Morse-Smale complex

[P.-T Bremer, H. Edelsbrunner, B. Hamann and V. Pascucci. A Multi-resolution Data Structure for Two-dimensional Morse-Smale Functions. 2003]

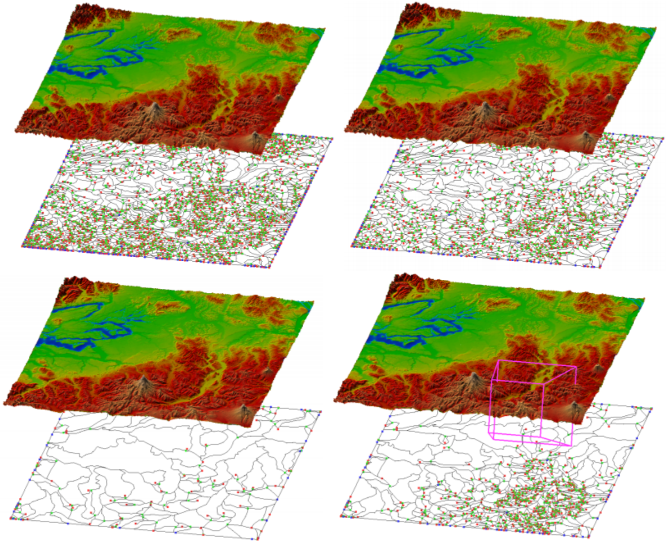
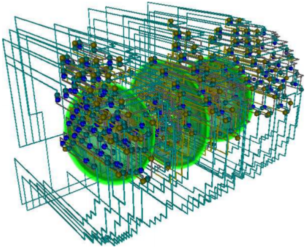


Figure 11: (Upper-left) Puget Sound data after topological noise removal. (Upper-right) Data at persistence of 1.2% of the maximum height. (Lower-left) Data at persistence 20% of the maximum height. (Lower-right) View-dependent re□nement (purple: view frustum).

Morse-Smale complex

[A. Gyulassy, V. Natarajan, V. Pascucci, P.-T. Bremer, B. Hamann. Topology-based Simplication for Feature Extraction from 3D Scalar Fields, 2005]



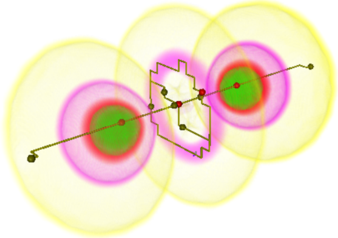
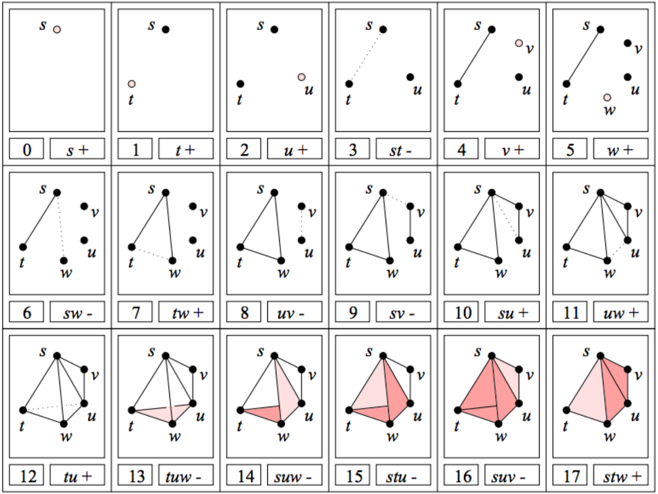


Figure: Topology simplication applied on electron density data for a hydrogen atom: the input has a large number of critical points, several of which are identied as being insignicant and removed by repeated application of two atomic operations. Features are identied by the surviving critical points and enhanced in a volume rendered image by an automatically designed transfer function

Persistent homology

[H. Edelsbrunner, D. Letscher and A. Zomorodian. Topological persistence and simplification. 2002] [A. Zomorodian, G. Carlsson. Computing Persistent Homology. 2004] Persistence diagram v.s. barcodes and persistence modules.



Persistent homology

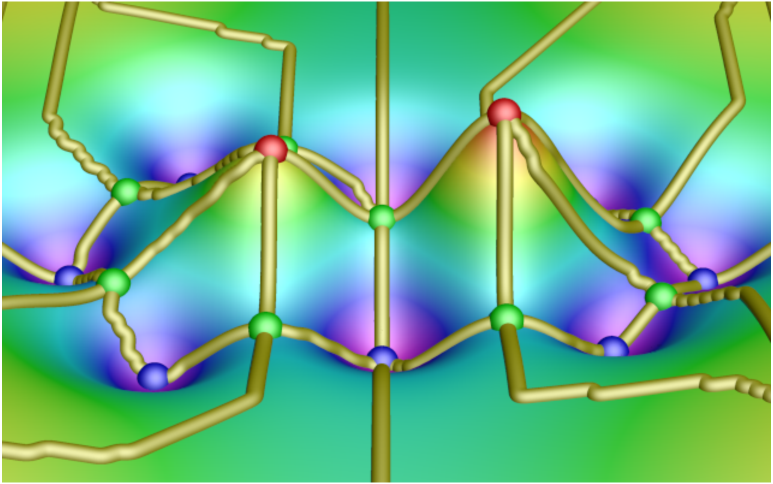
When data is corrupted by noise, how can we tell features from noise?

"The eye, or the brain, performs the marvelous task of taking the sense data of individual points and assembling them into a coherent image of a continuumit infers the continuous from the discrete."

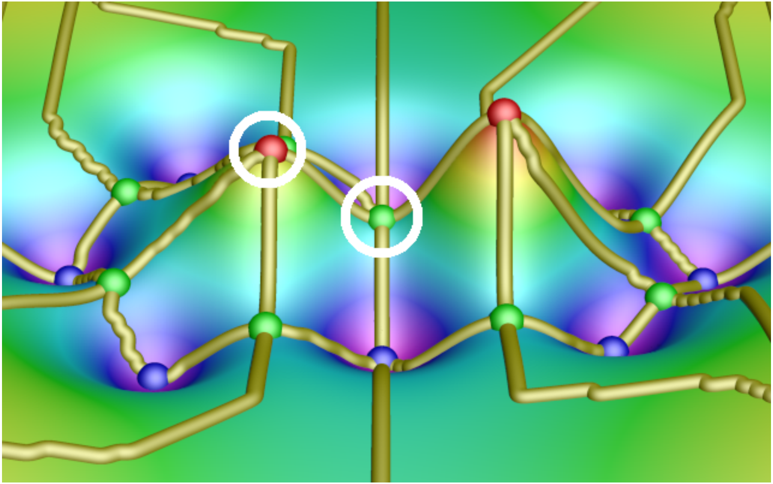
Figure: The Seine at La Grande Jatte by Georges Seurat

[S. Weinberger. What is persistent homology? 2011]

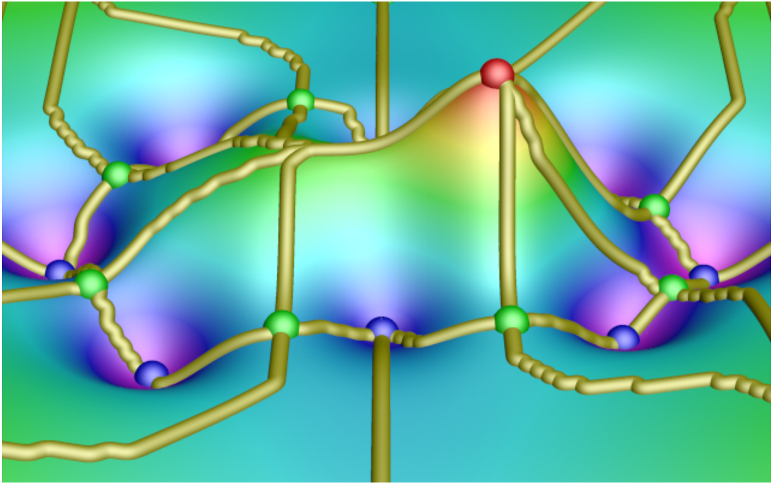
Simplifying topological features



Simplifying topological features



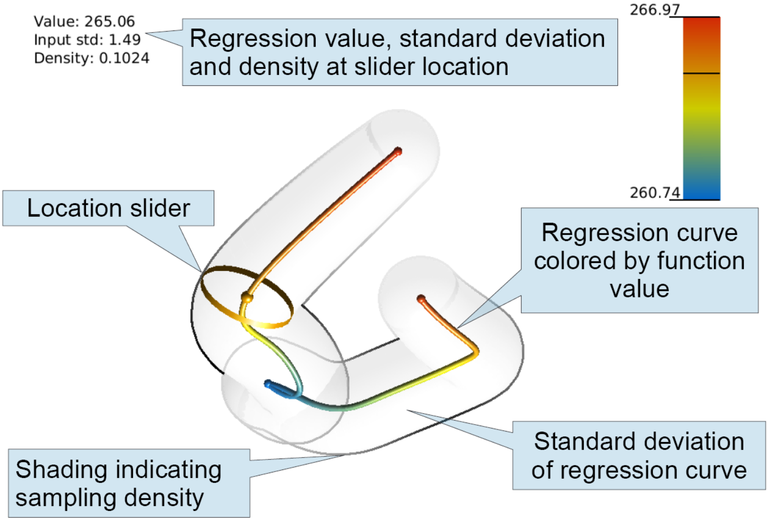
Simplifying topological features



What about hight dimensional data? More data analysis than visualization...

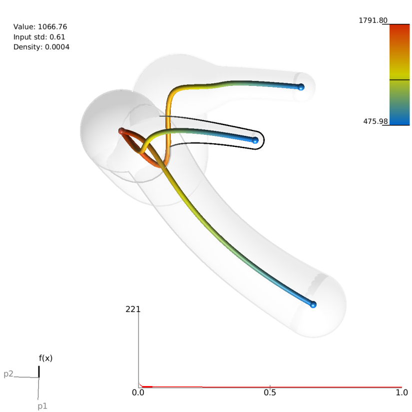
High dimensional scalar function

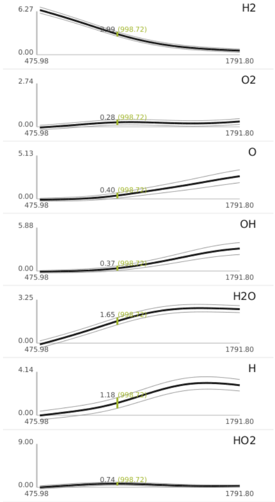
[S. Gerber, P.-T. Bremer, V. Pascucci, R. Whitaker. Visual Exploration of High Dimensional Scalar Functions. 2010]



High dimensional scalar function

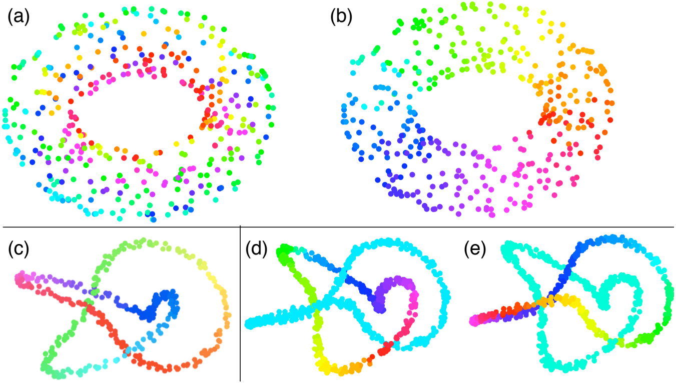
[S. Gerber, P.-T. Bremer, V. Pascucci, R. Whitaker. Visual Exploration of High Dimensional Scalar Functions. 2010]





10 dimensional data set describing the heat release wrt. to various chemical species in a combustion simulation

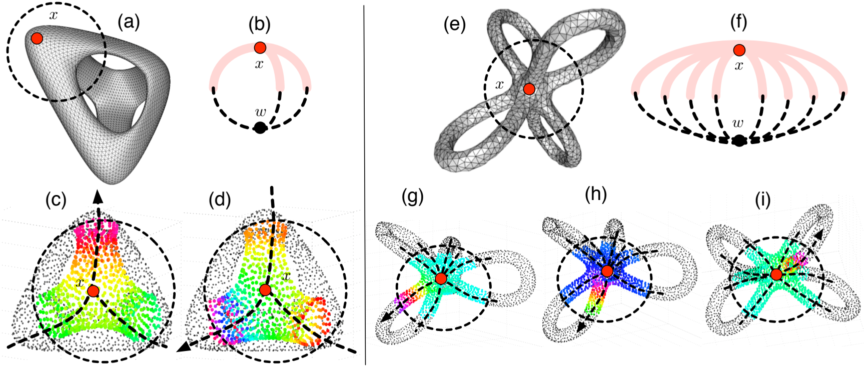
Parametrizing data (for circular features) in high-dimensions.



[Silva, Morozov, Vejdemo-Johansson 2009]

Detect branching features in high-dim data

Parametrizing data in high-dimensions.

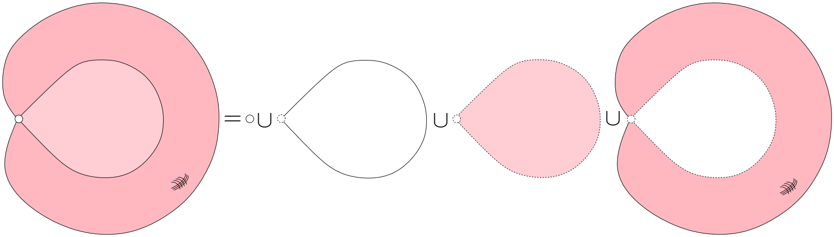


[W, Summa, Pascucci, Vejdemo-Johansson 2011]

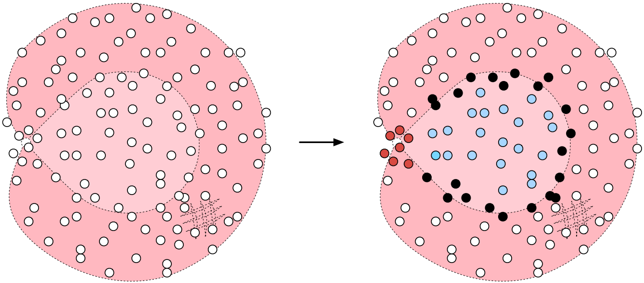
Stratification learning in high dimensions

The coarsest stratification of a pinched torus

1. Decompose into manifold pieces (strata). 2. Pieces fit "nicely".



Stratification learning in high dimensions



What are some of the cool open problems?

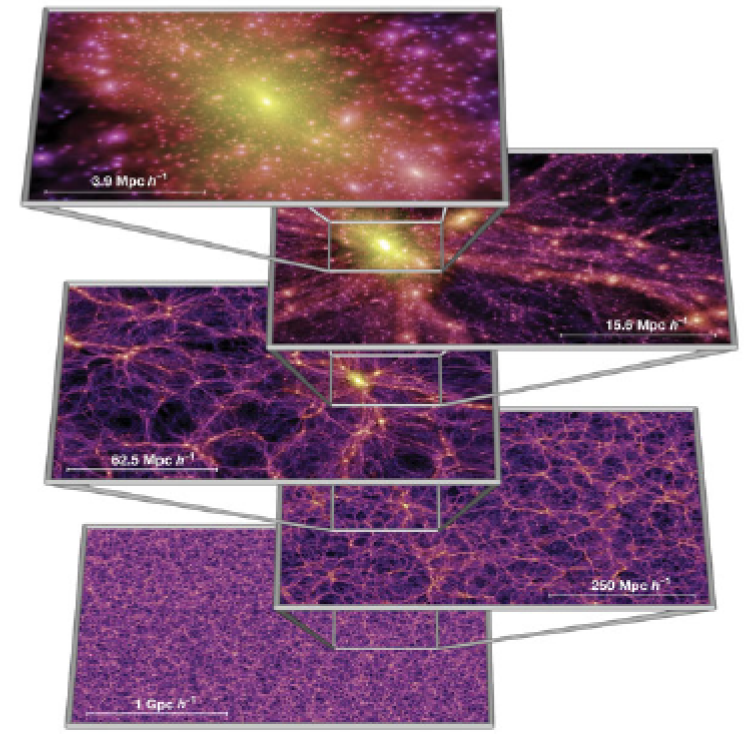
For data analysis...

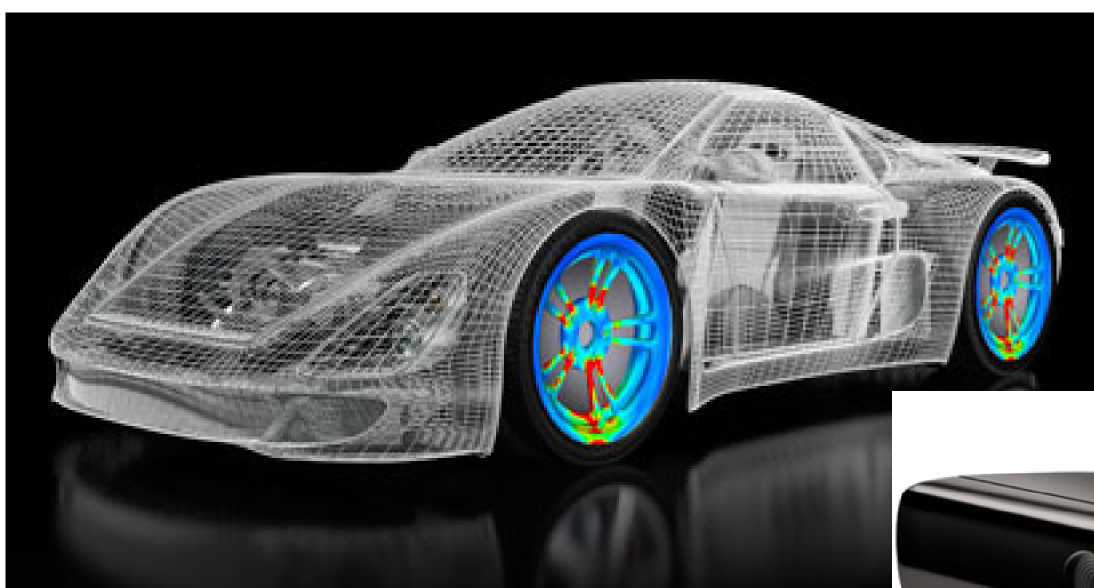
- Robustness of topological structures
- Scalability, approximation
- High-dimensional data
- Integration with statistics and machine learning
- Usability

Context: The data deluge

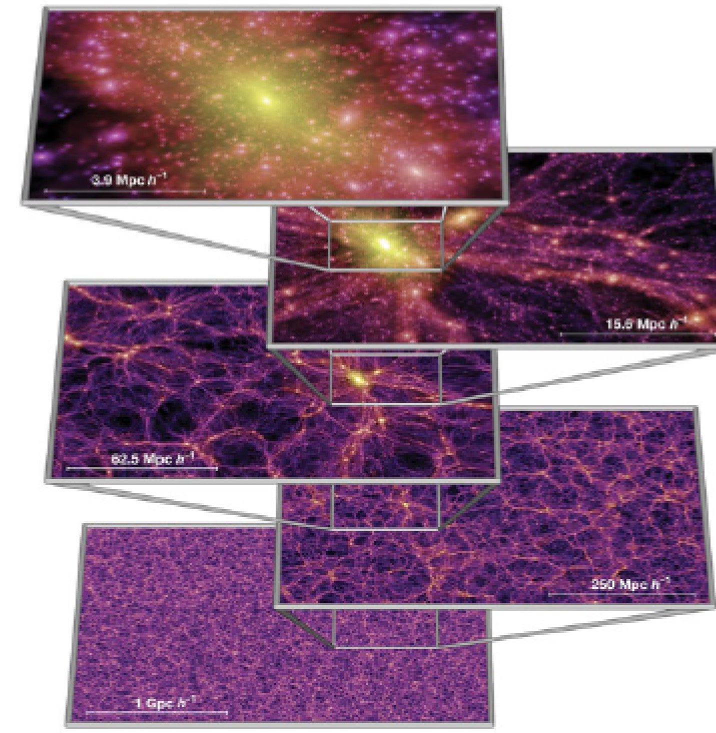
Data are generated at an unprecedented rate by:

- academia
- industry
- general public



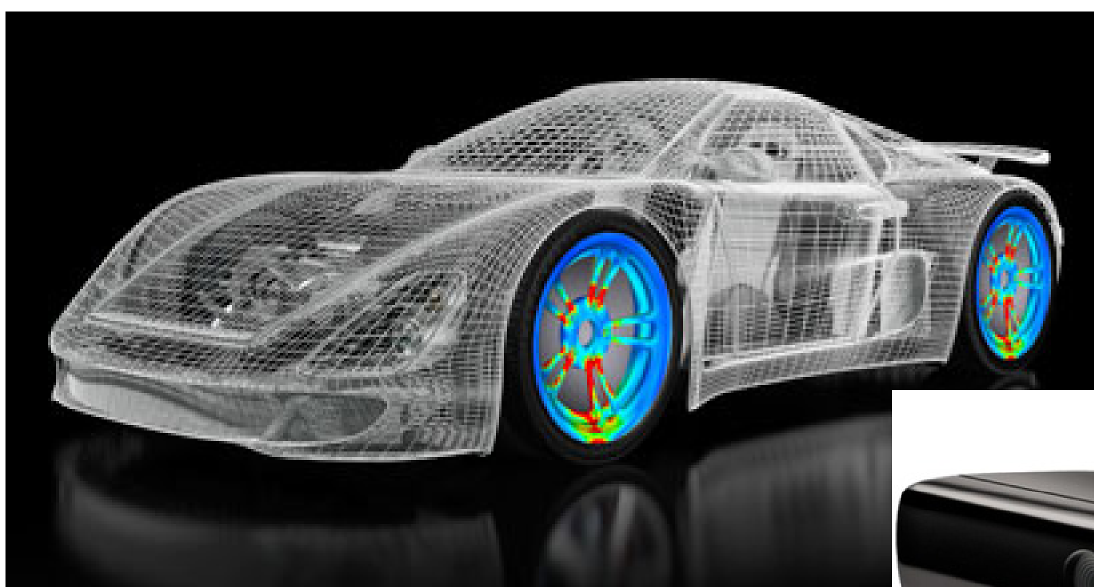


Context: The data deluge



Data are generated at an unprecedented rate by:

- academia
- industry
- general public

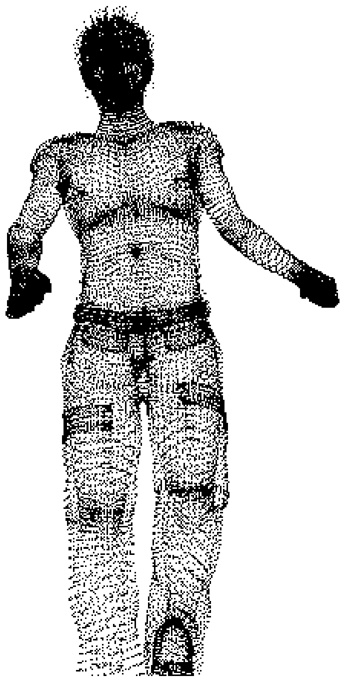


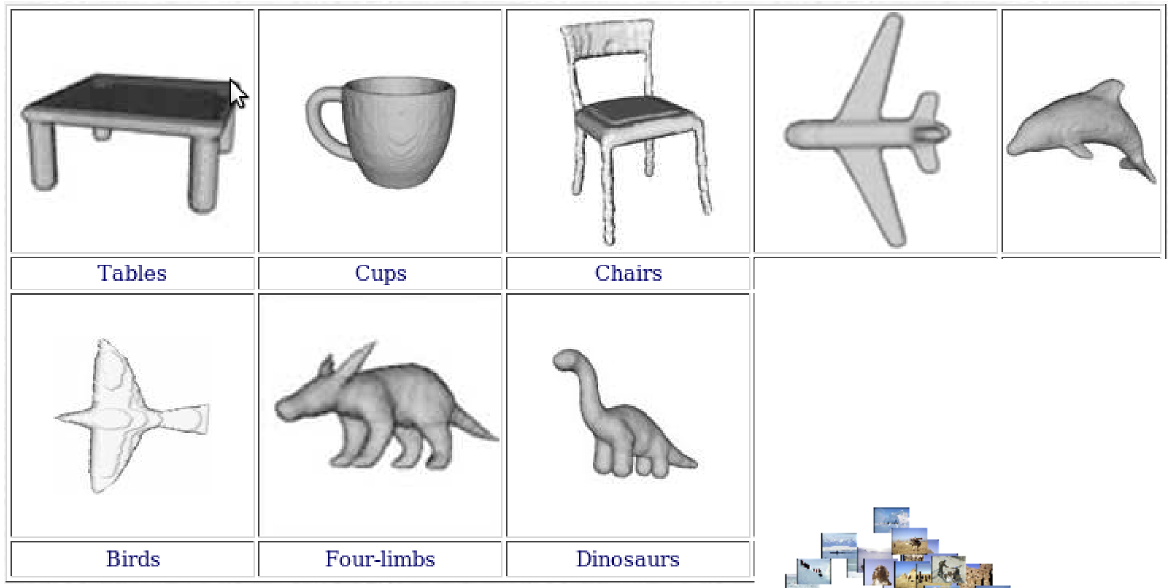
Need for new scalable methods to analyze and classify these data automatically

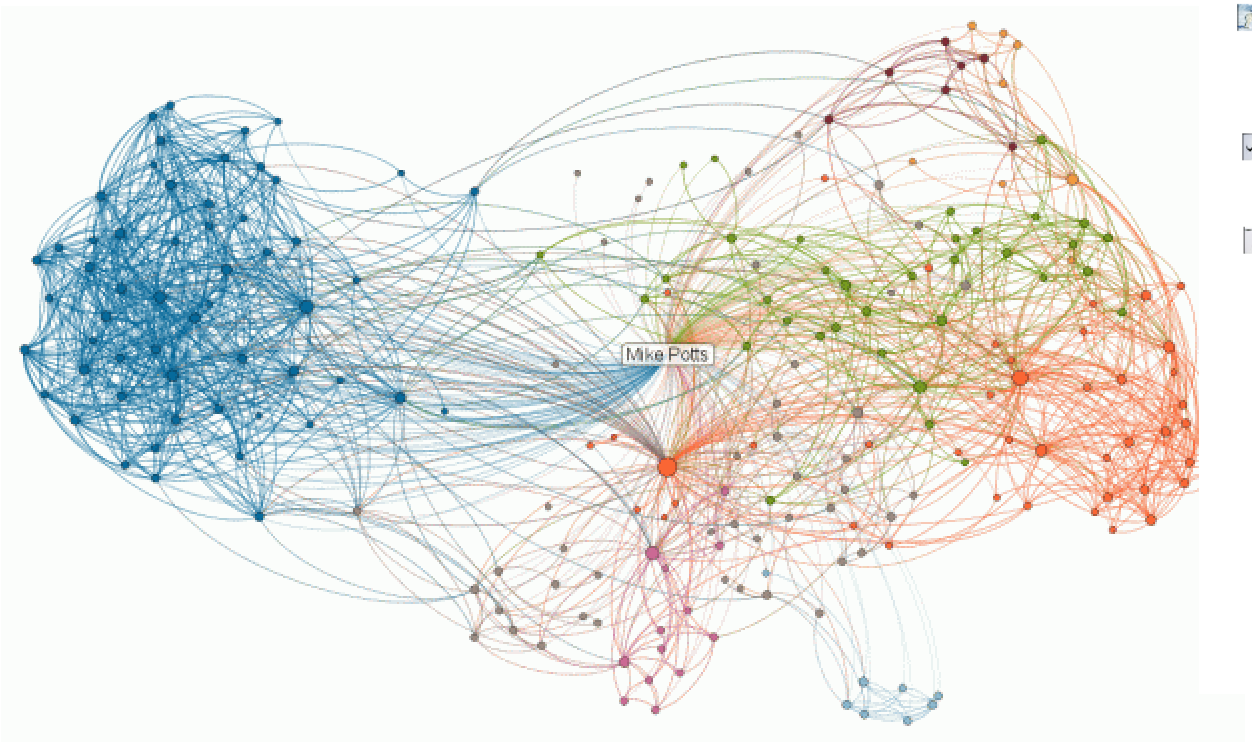
Exploratory analysis of geometric data

Input: set of data points with metric or (dis-)similarity measure

data point $\equiv$ 3d point, image patch, image or 3d shape in collection, Facebook user, etc.



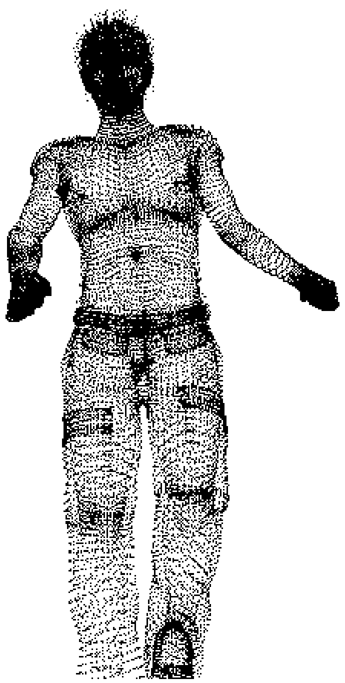


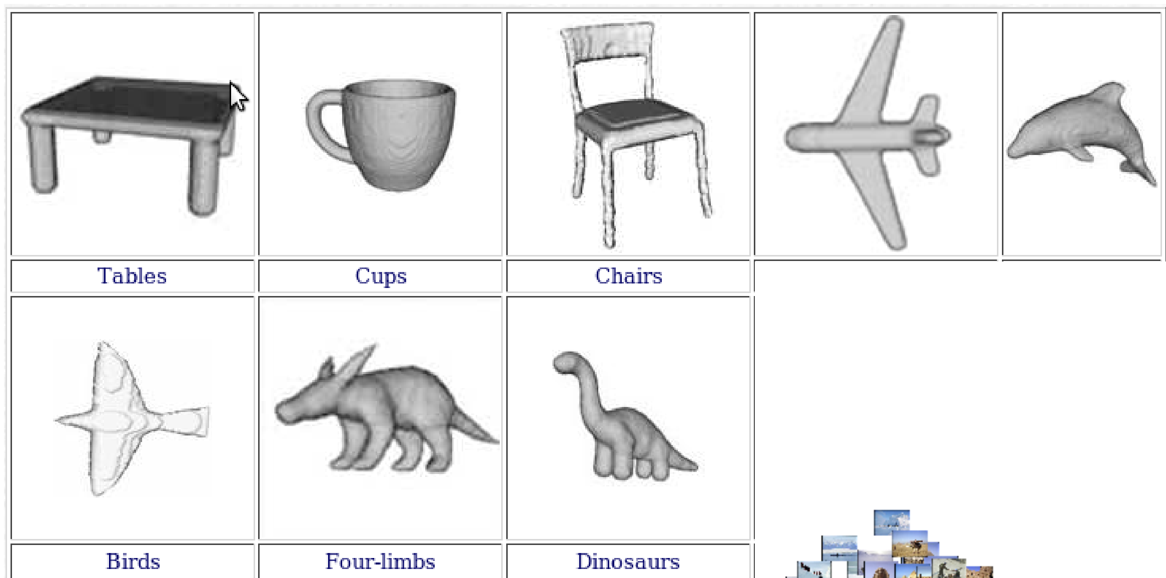


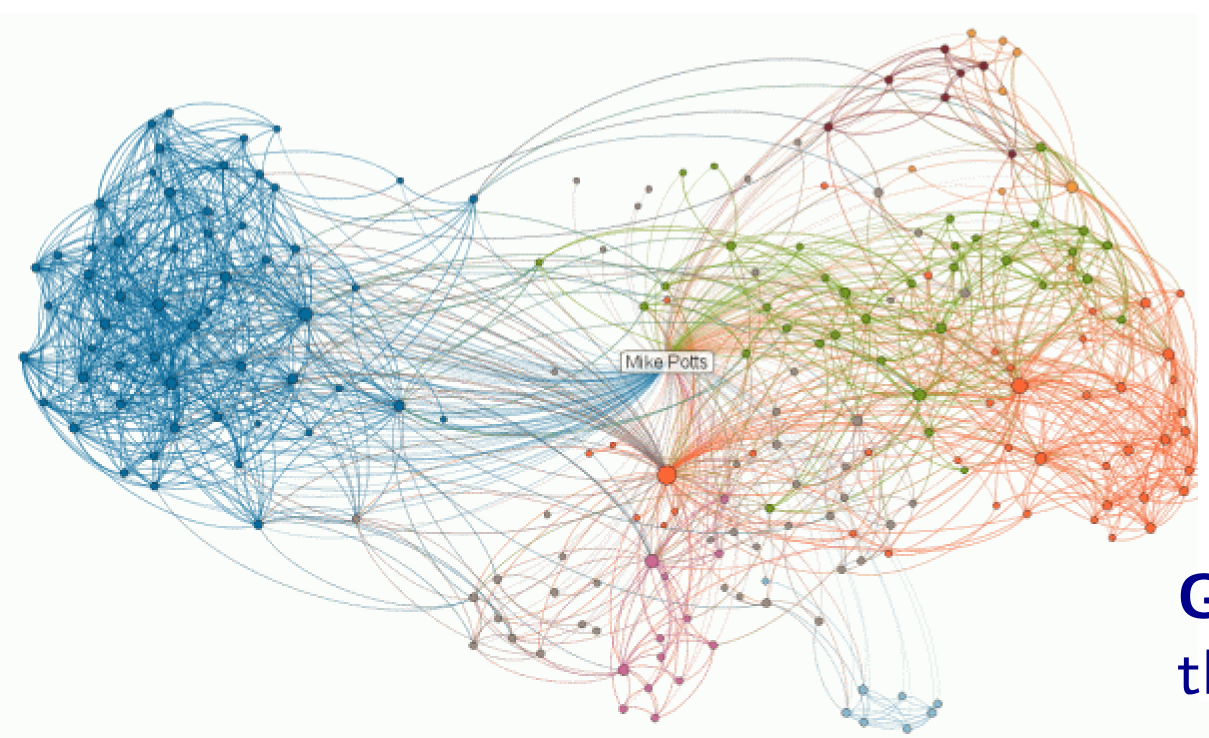
Exploratory analysis of geometric data

Input: set of data points with metric or (dis-)similarity measure

data point $\equiv$ 3d point, image patch, image or 3d shape in collection, Facebook user, etc.

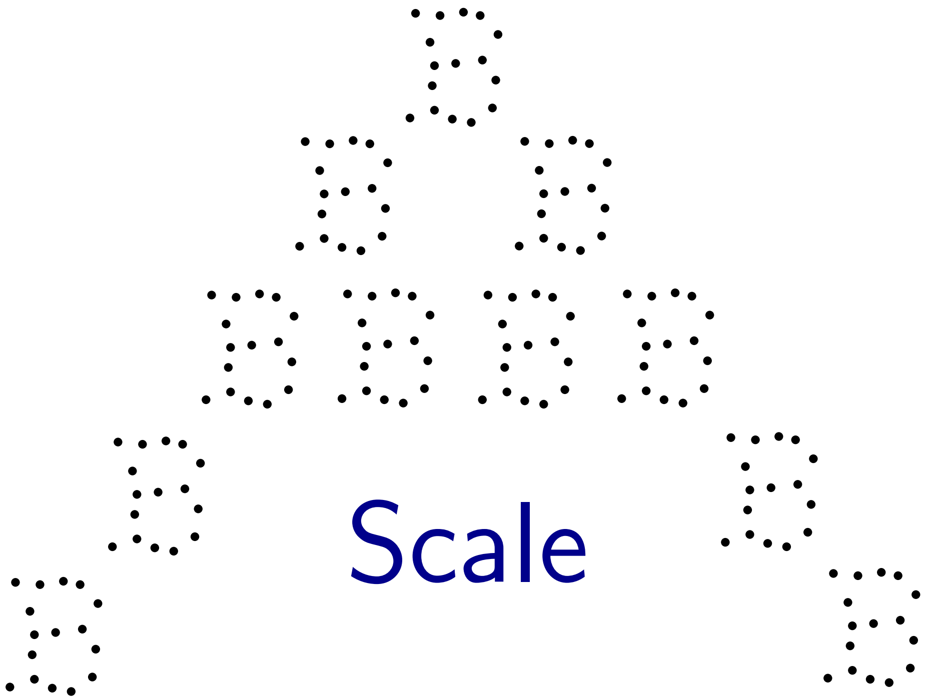




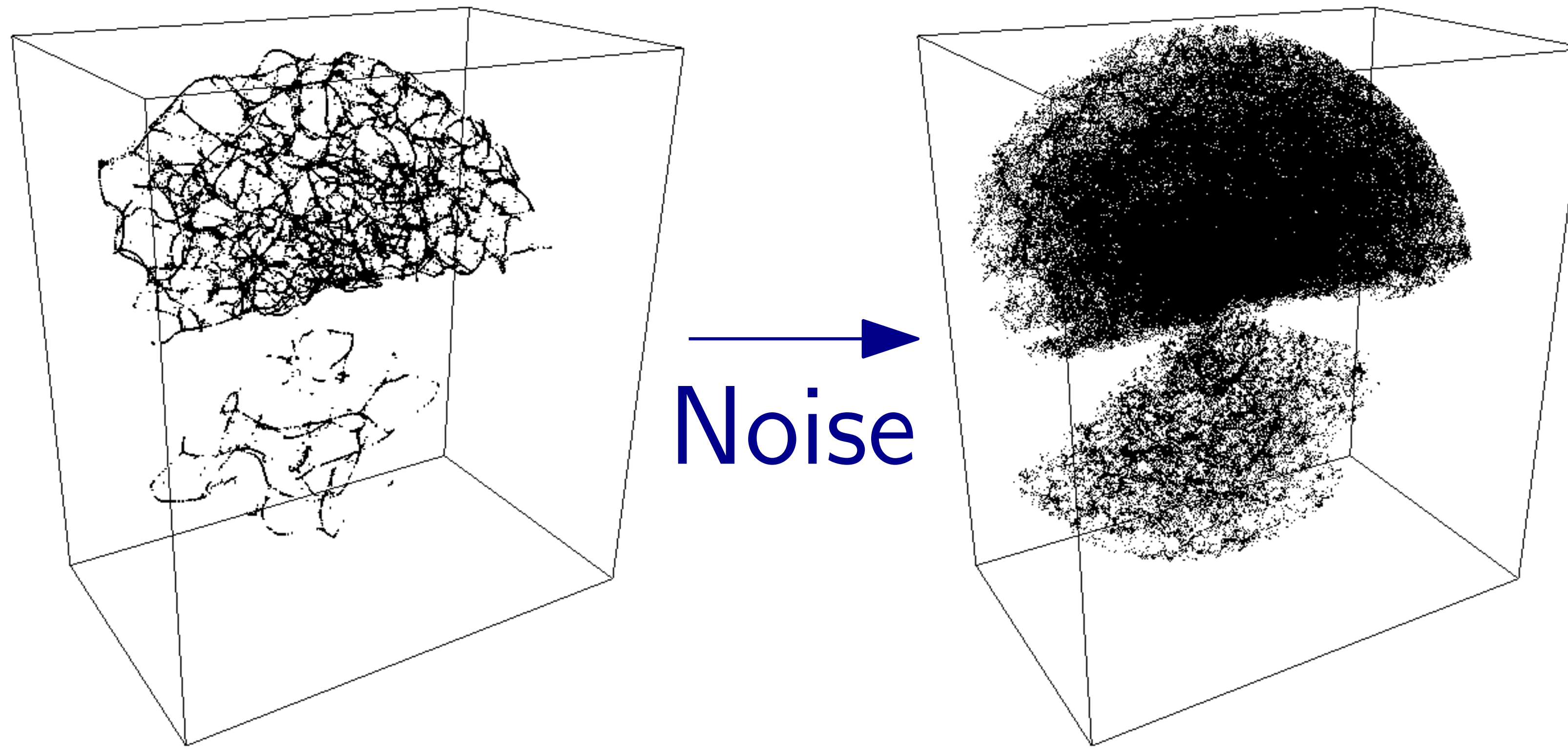


Goal: describe the underlying structure of the data, for interpretation or summary

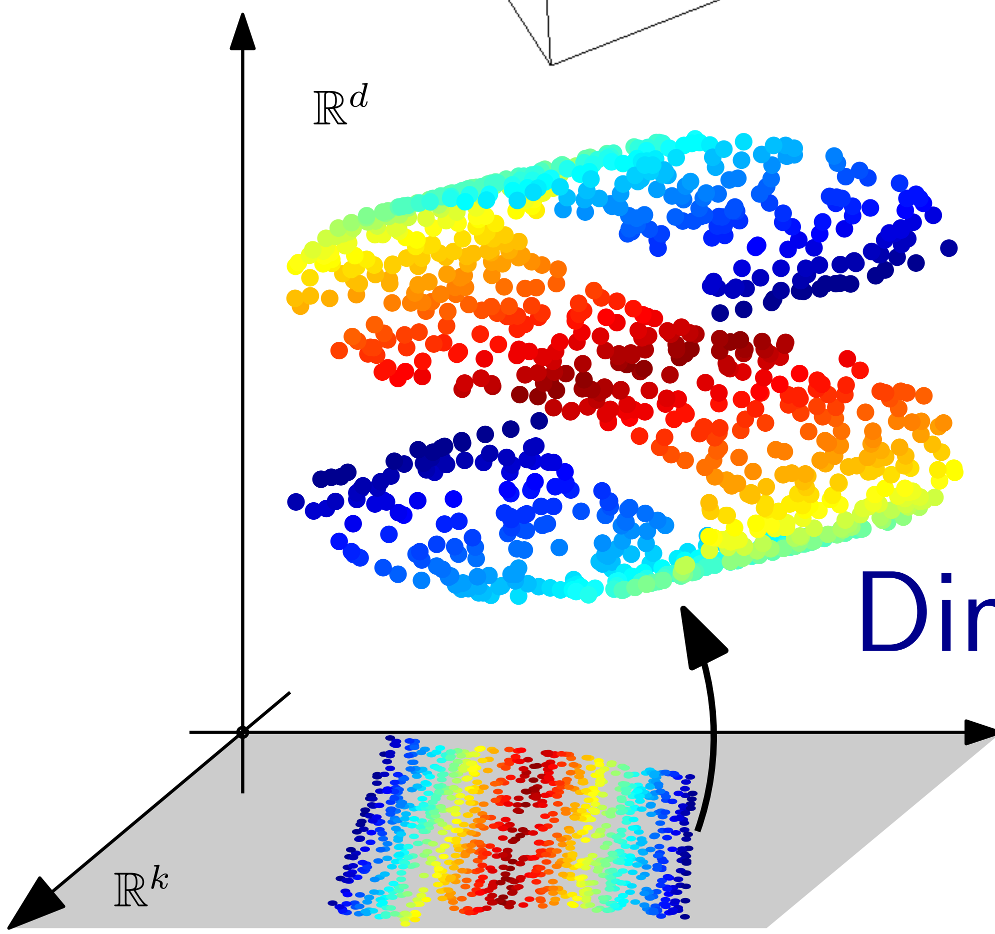
Challenges



Scale

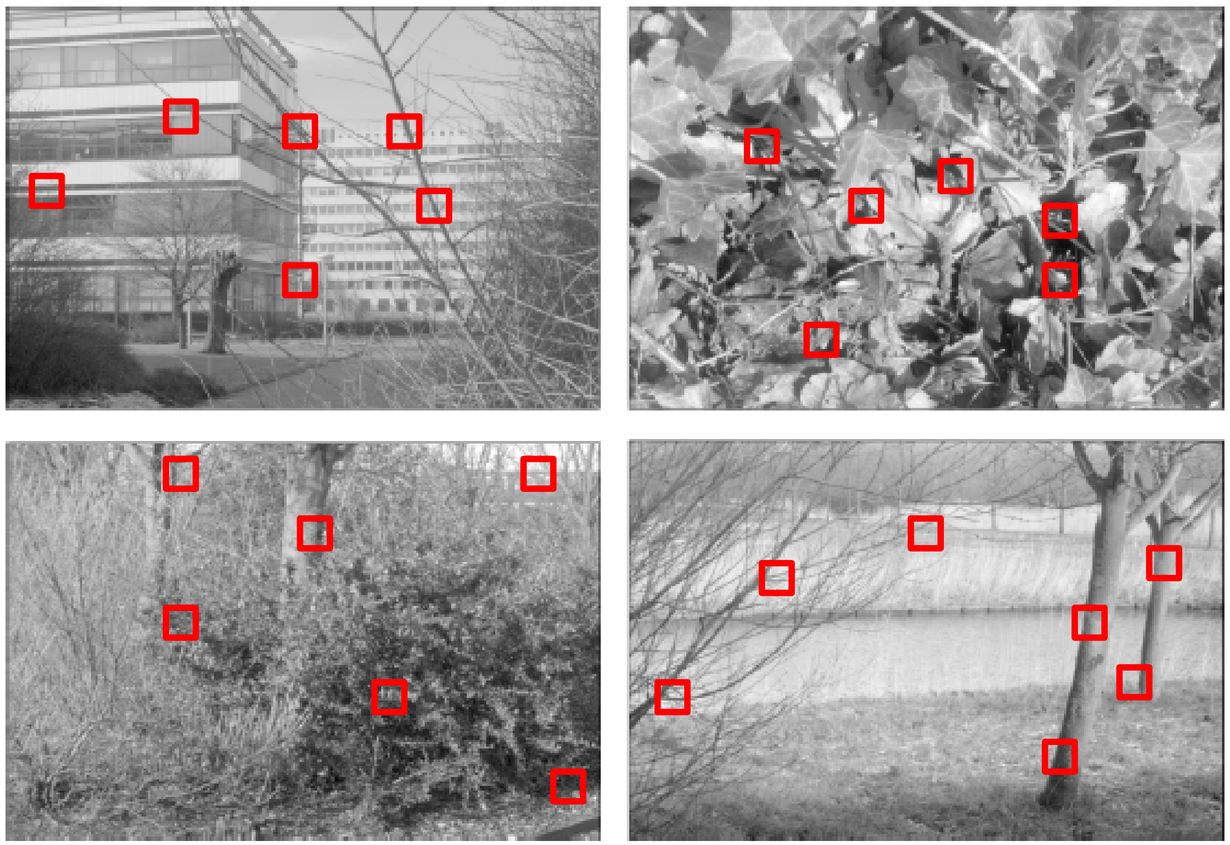


Noise



Dimensionality

Challenges

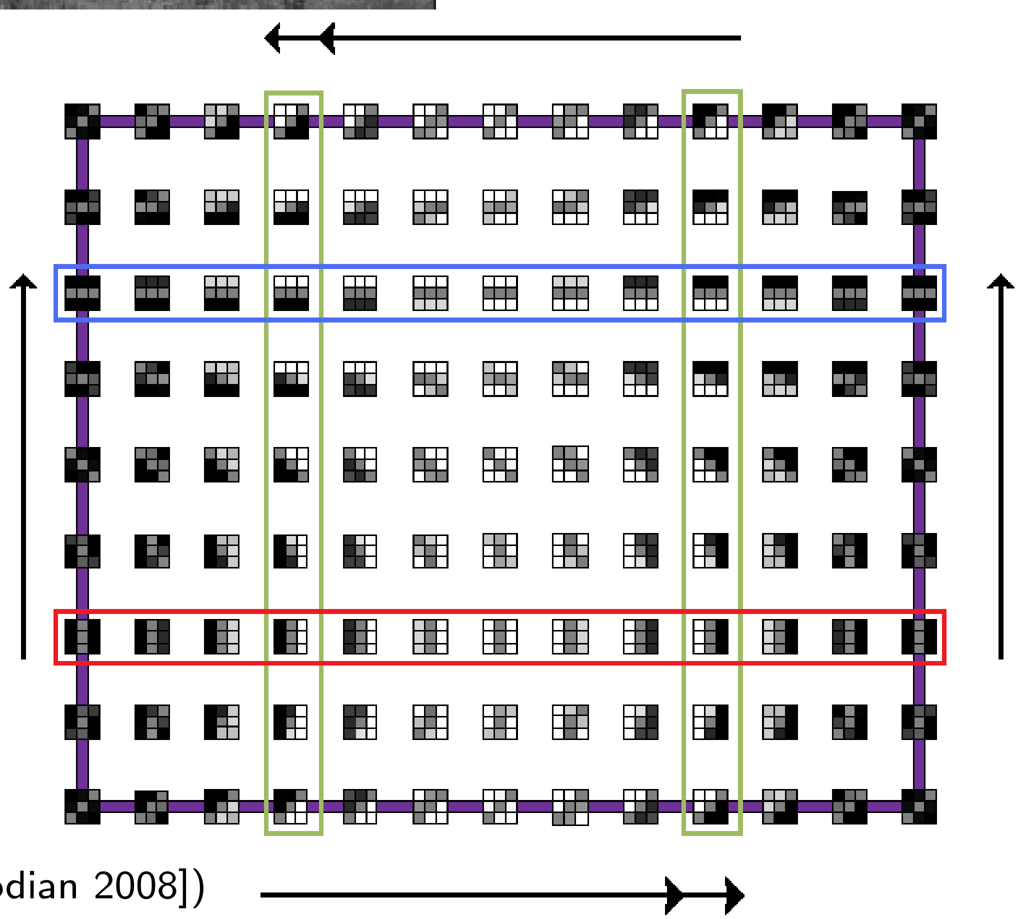


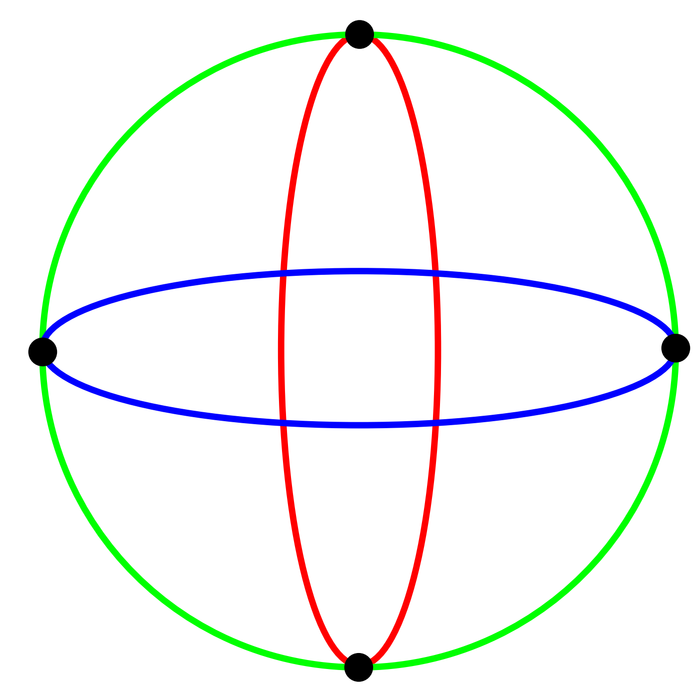
4 million data points in $\mathbb{R}^9$

(source: [Lee, Pederson, Mumford 2003])

Motivation: study cognitive representation of space of images

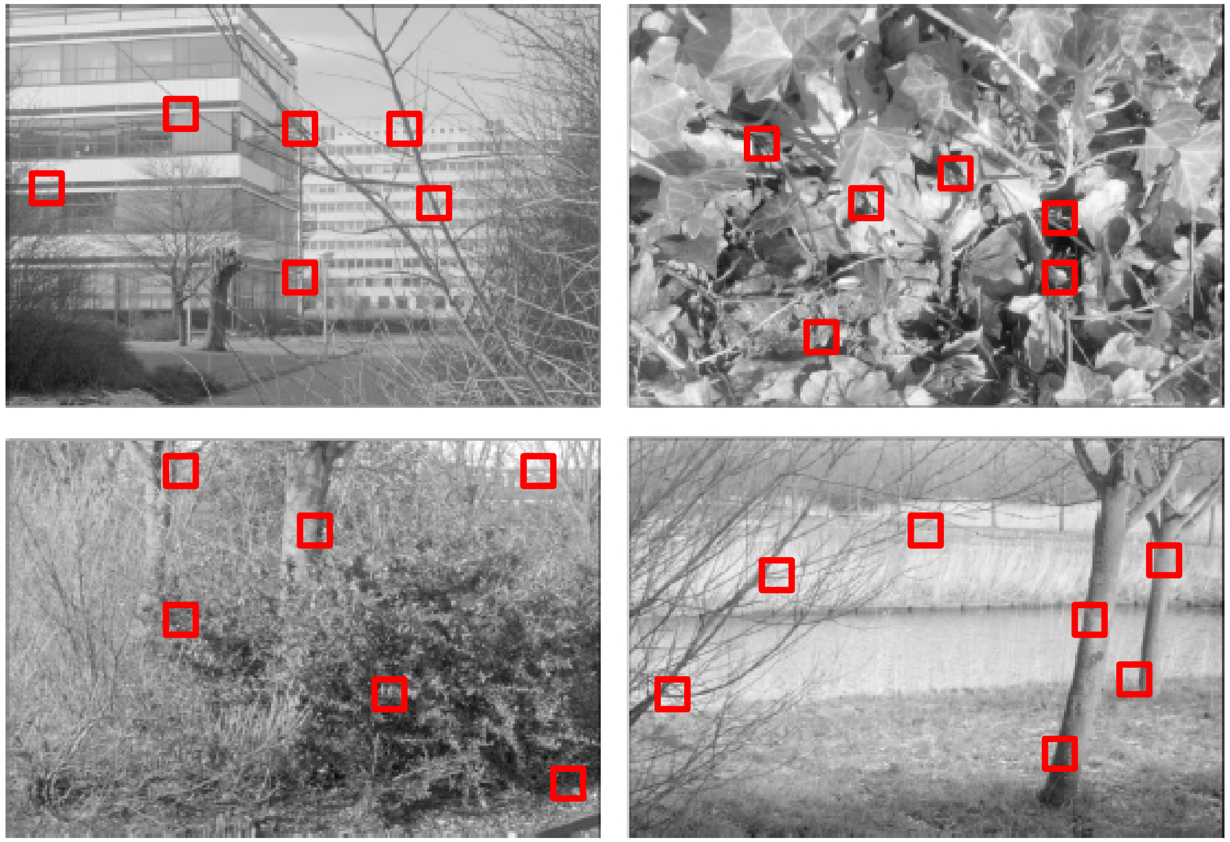
Topology





(source: [Carlsson, Ishkhanov, de Silva, Zomorodian 2008])

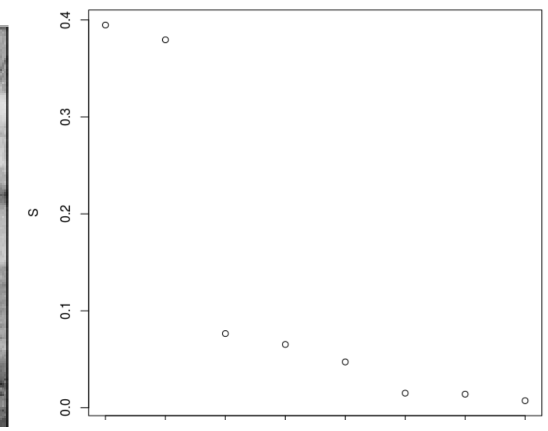
Challenges

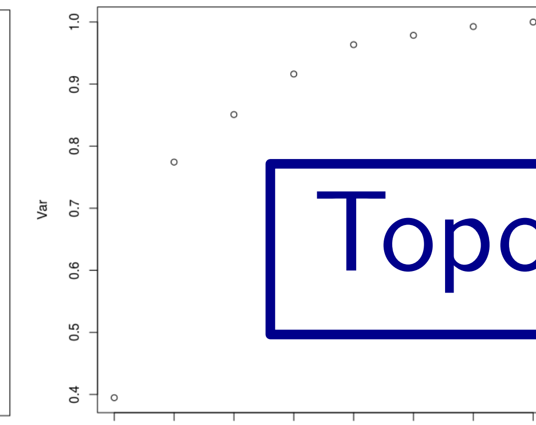


4 million data points in $\mathbb{R}^9$

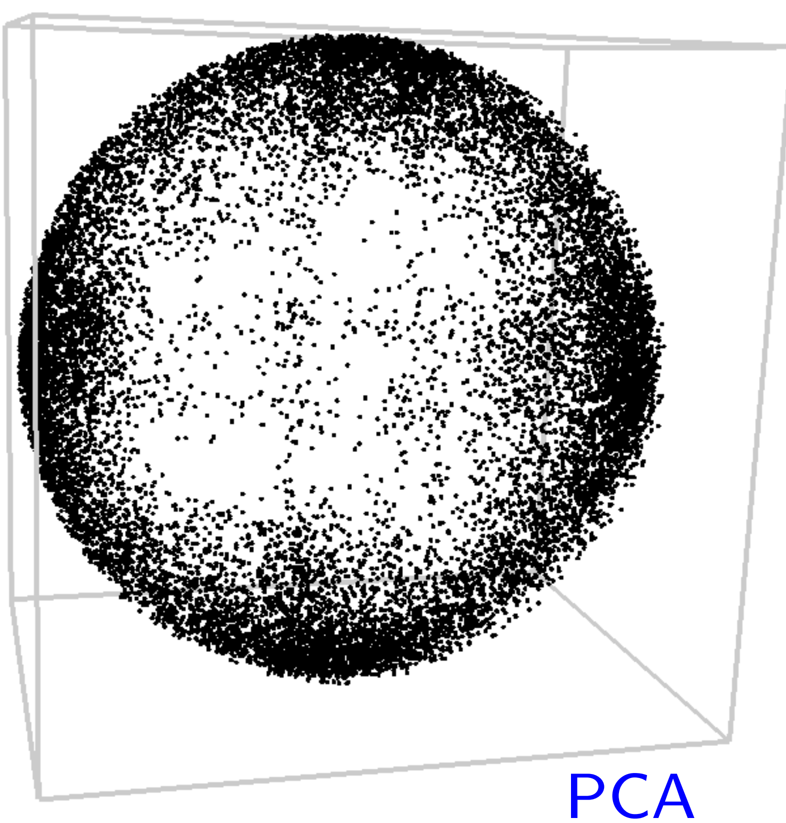
(source: [Lee, Pederson, Mumford 2003])

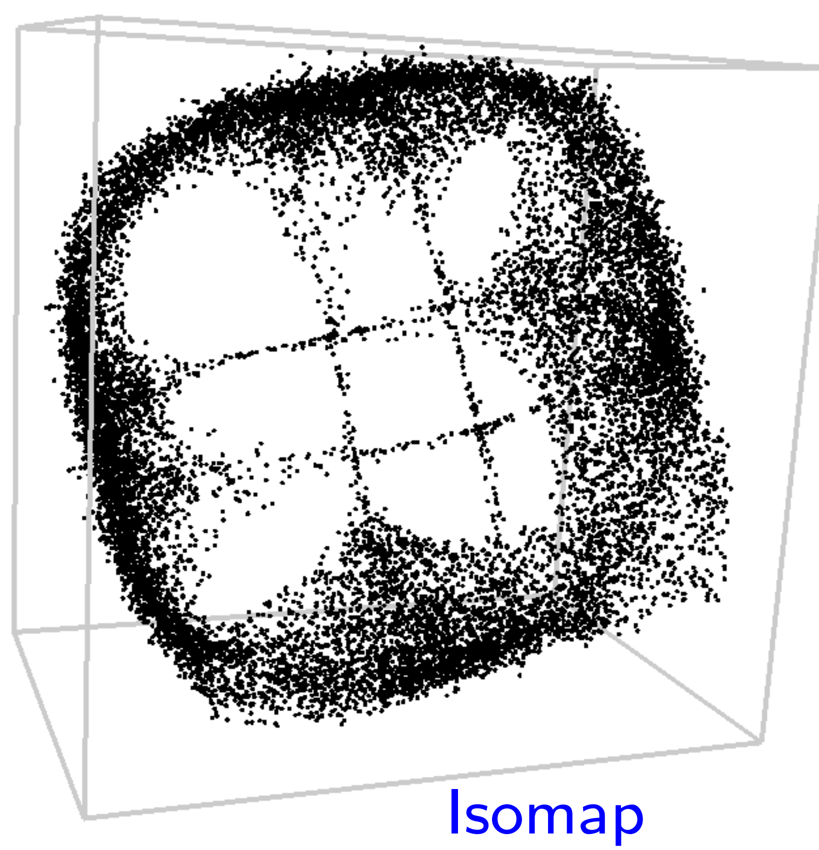
Motivation: study cognitive representation of space of images

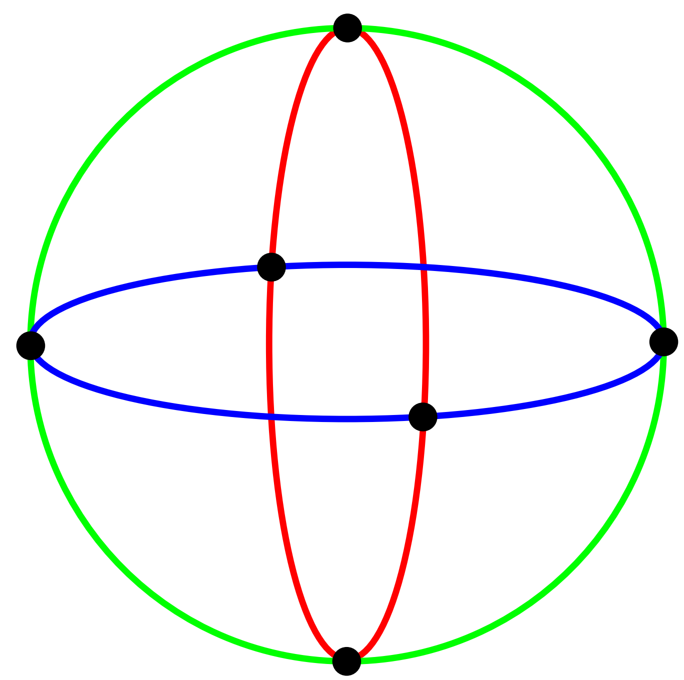




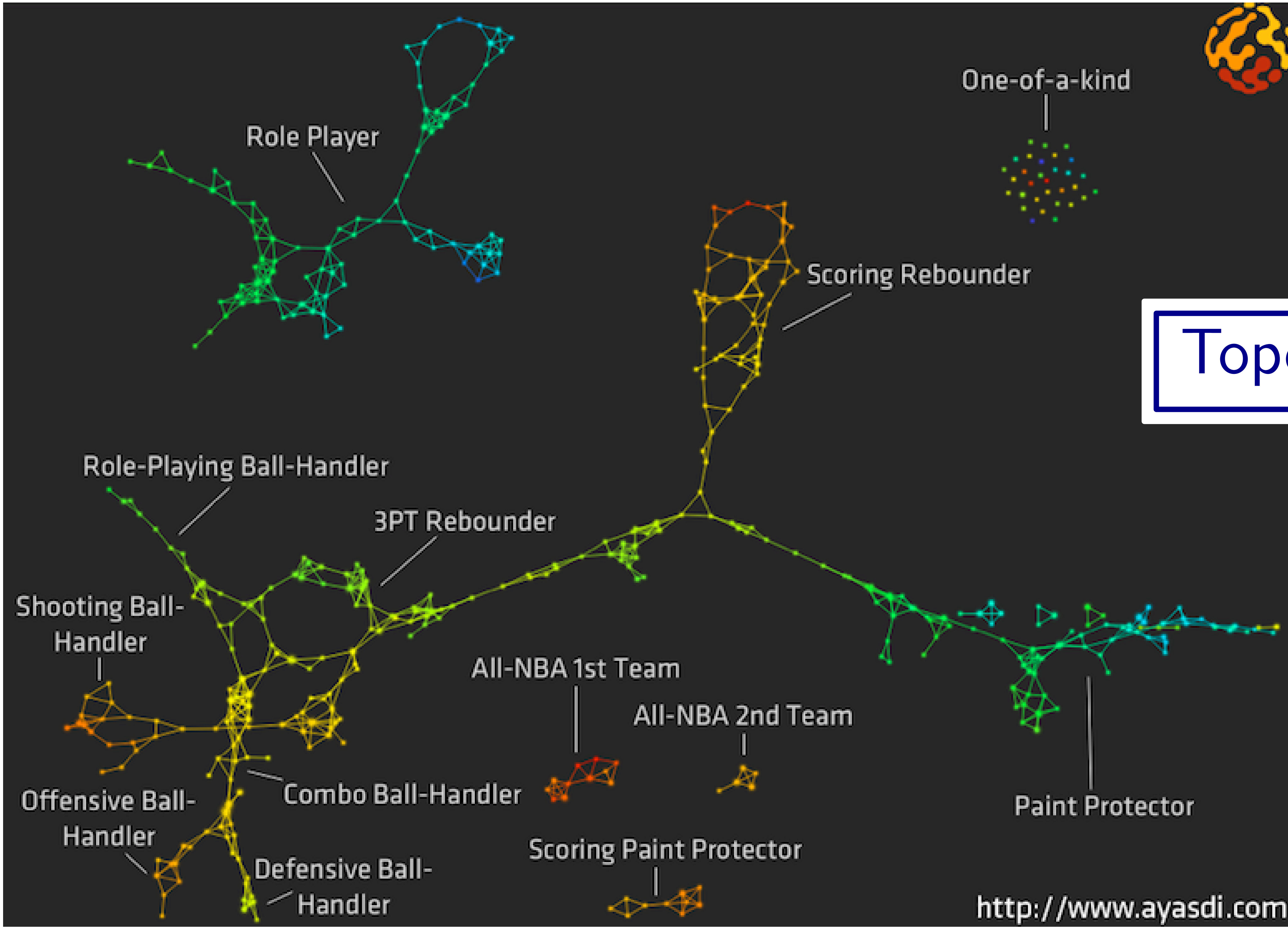
Topology







Challenges



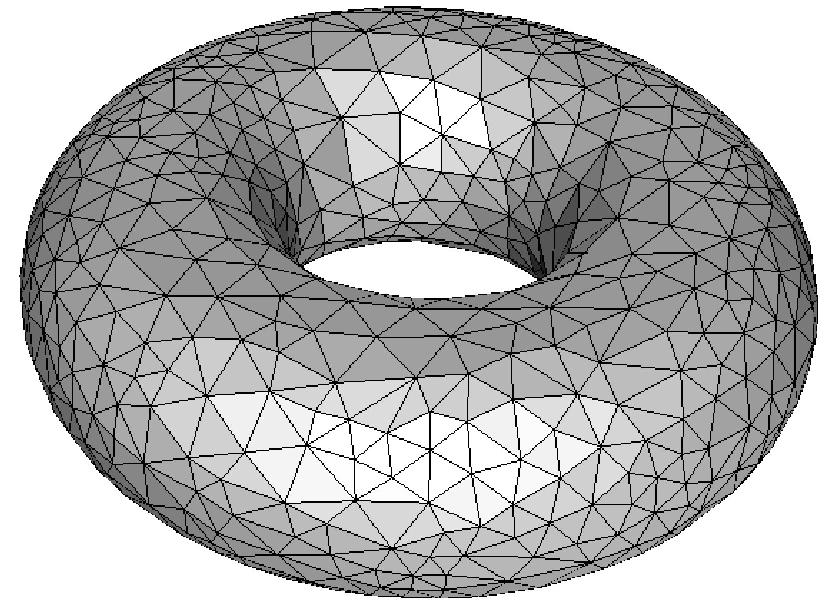
Topology

(source: http://www.sloansportsconference.com/wp-content/uploads/2012/03/Alagappan-Muthu-EOSMarch2012PPT.pdf)

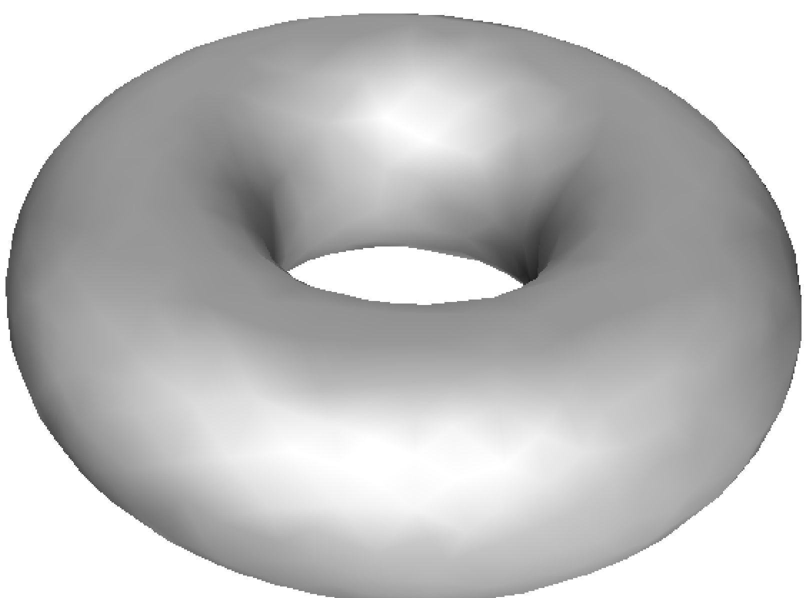
Topological Data Analysis (TDA)

topological invariants for classification

$$\beta_0 = \beta_2 = 1$$
$$\beta_1 = 2$$



triangulation

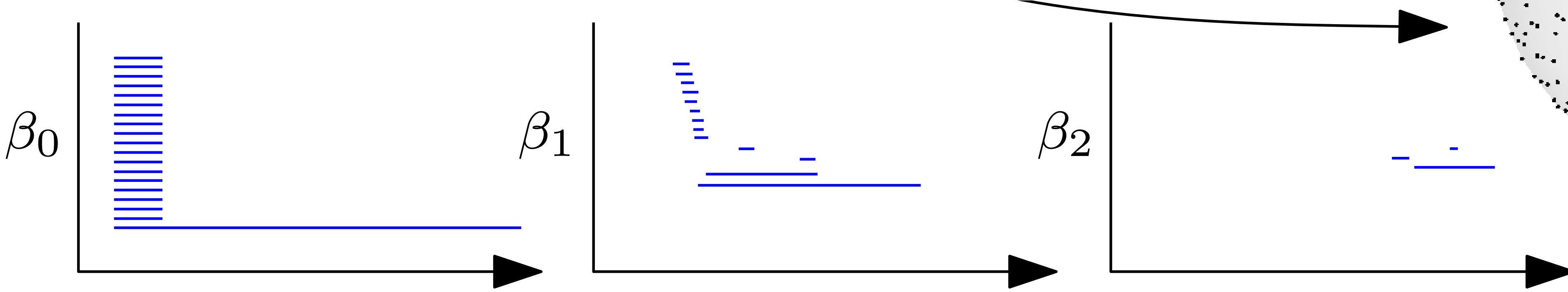


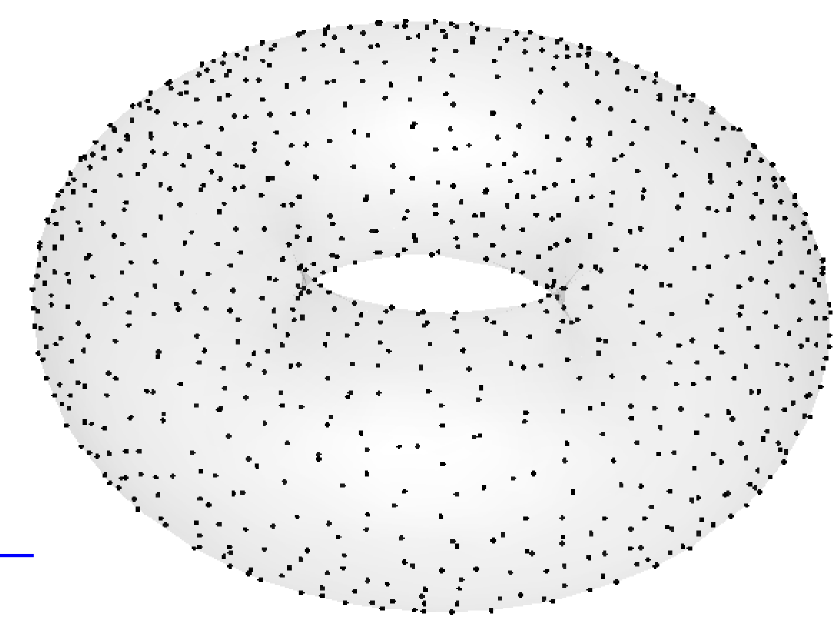
compact set

Algebraic topology

Applied algebraic topology

topological descriptors for inference and comparison



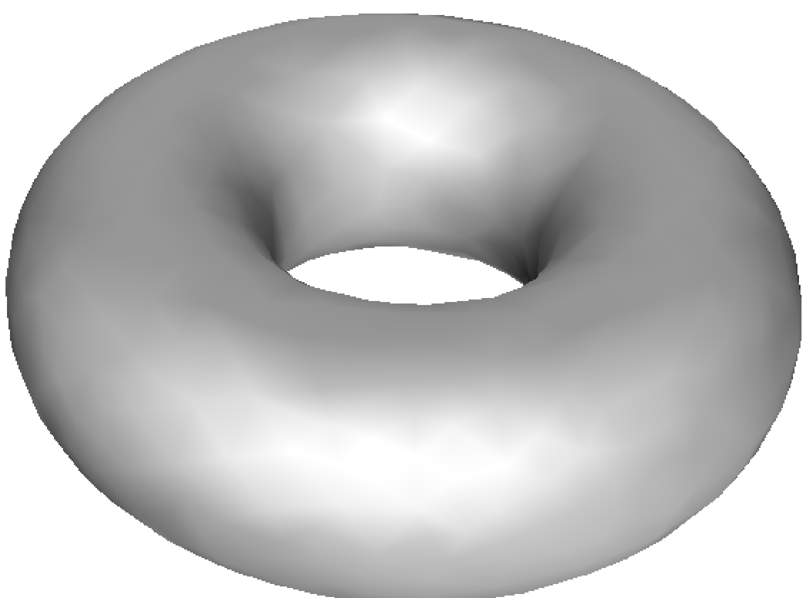


point cloud

Topological Data Analysis (TDA)

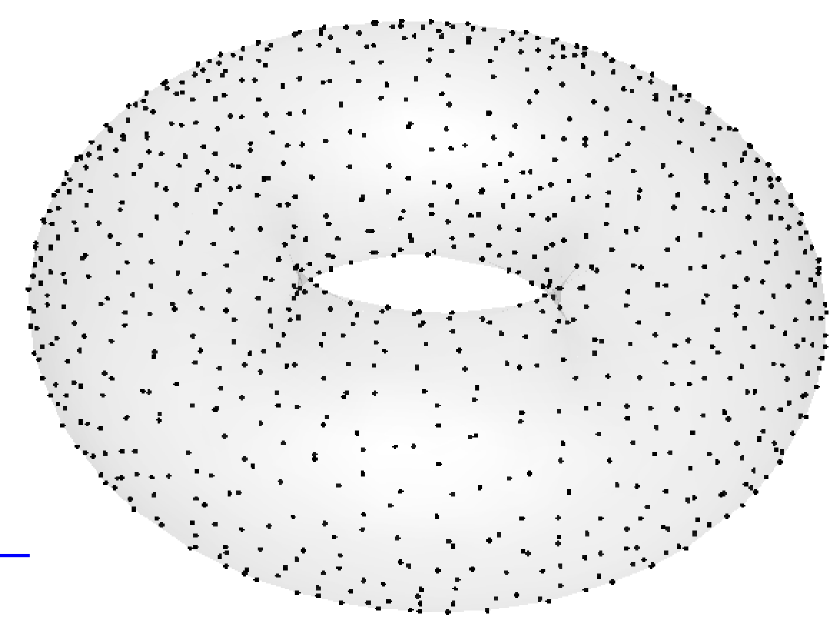
Properties of topological descriptors:

- invariant under coordinate changes
- stable with respect to perturbations
- informative
- applicable generally



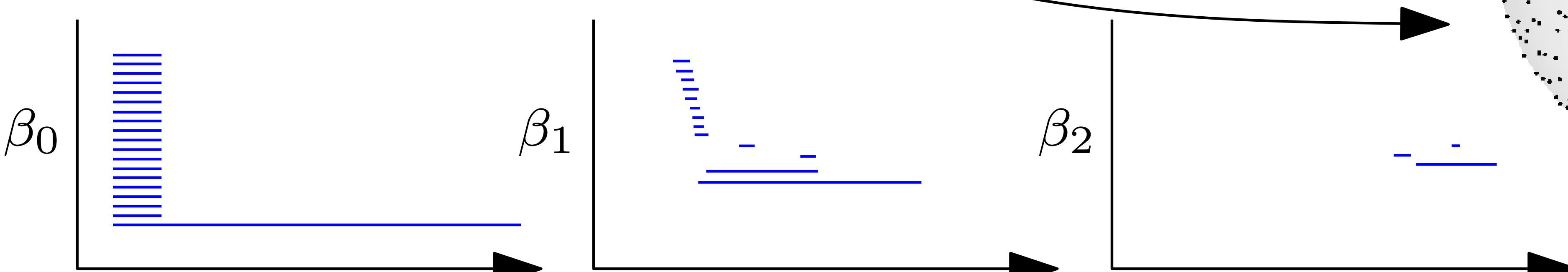
compact set

Applied algebraic topology

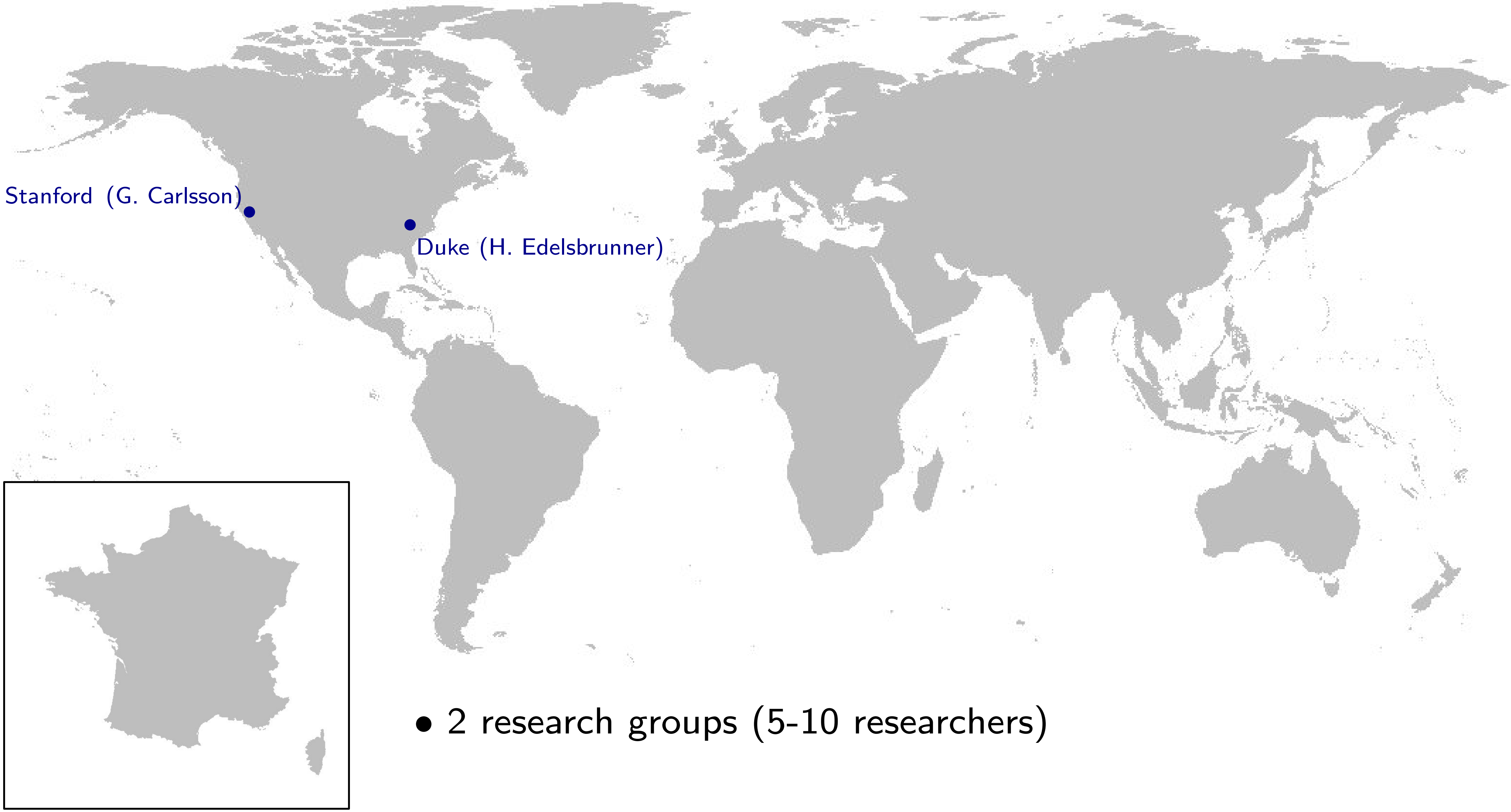


point cloud

topological descriptors for inference and comparison

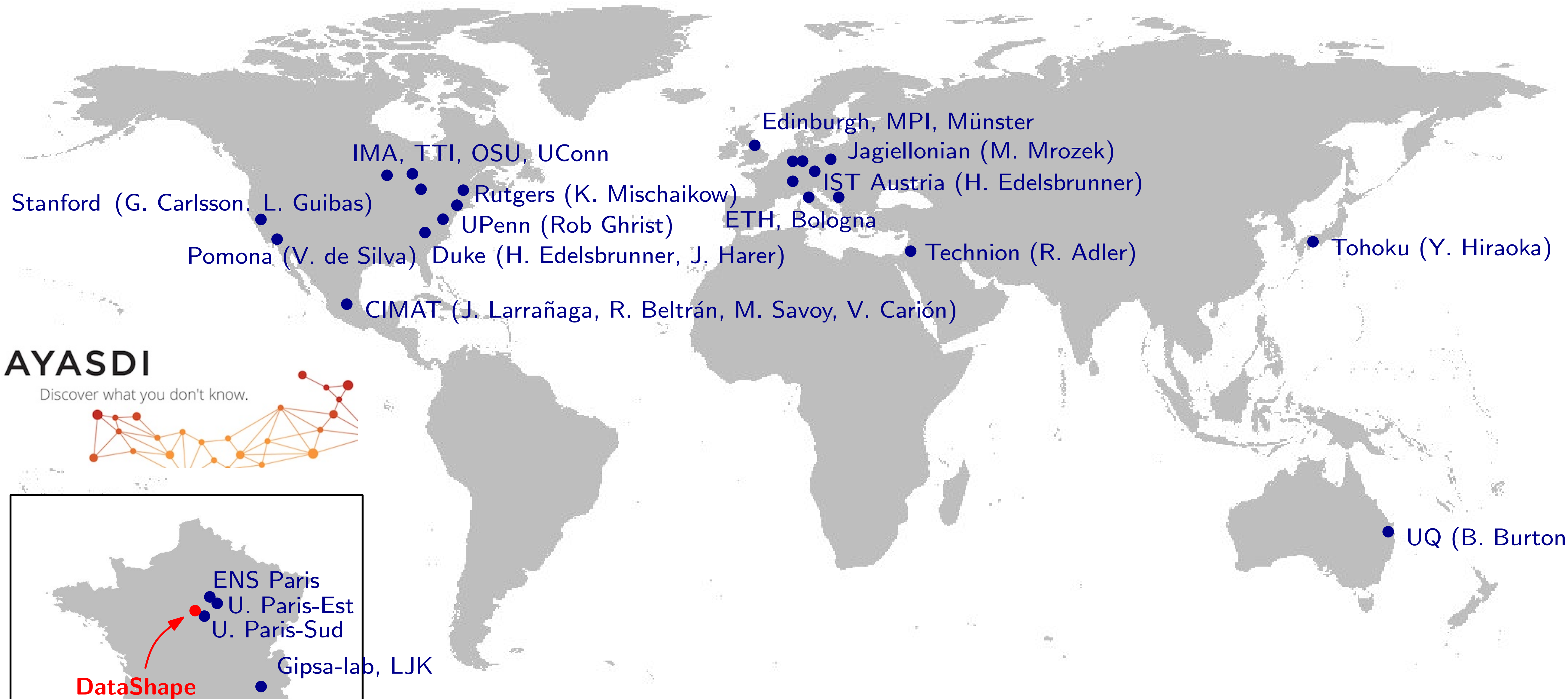


The TDA community (as of 2003)



- 2 research groups (5-10 researchers)

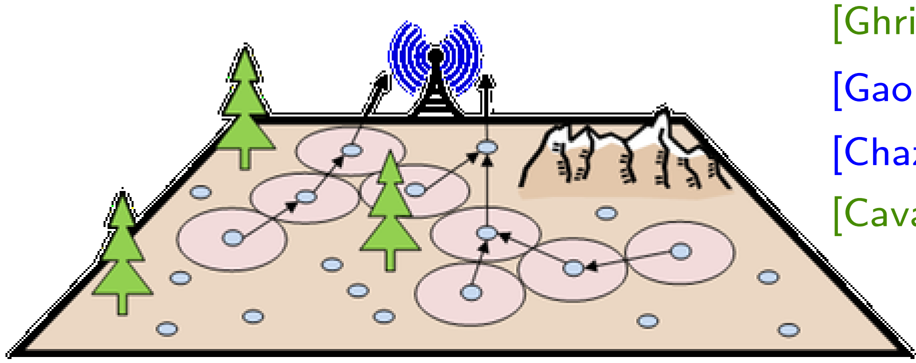
The TDA community (as of 2018)



- 50-100 researchers working on theoretical foundations
- 200-300 researchers at the interface with applications
- very successful applications and company (Ayasdi)
- In France: leading role of Geometrica / DataShape since 2007

A few applications of TDA

Topology of wireless sensor fields

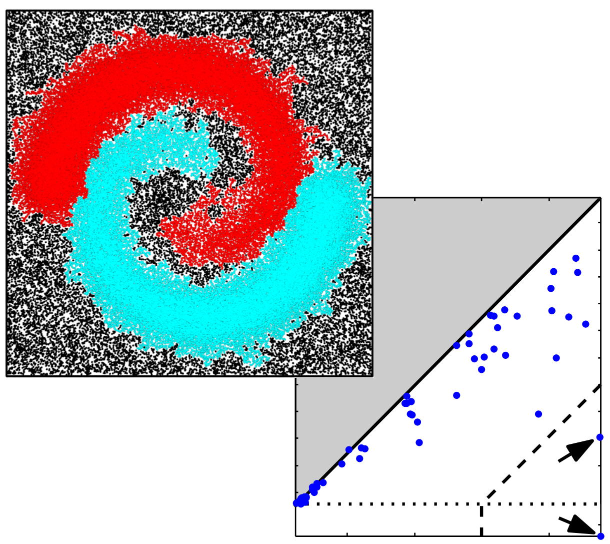


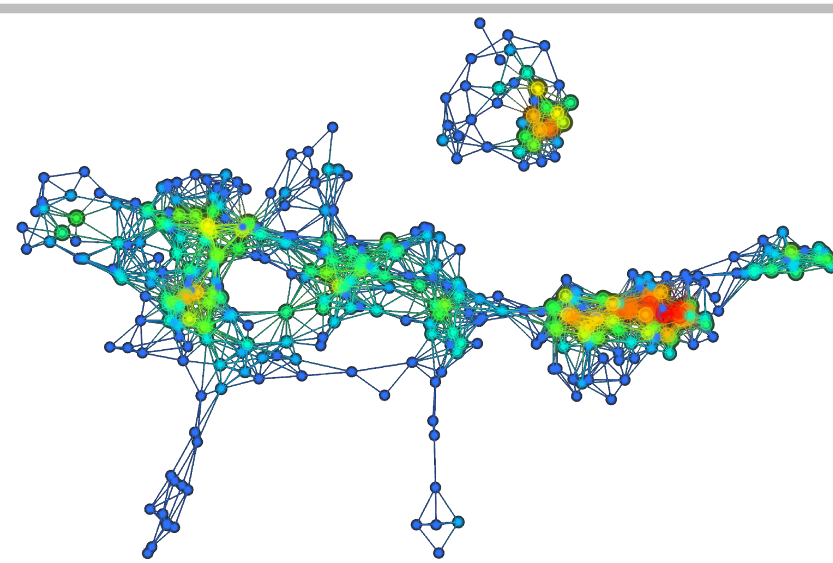
[Ghrist, de Silva 2005]

[Gao, Guibas, O., Wang 2010]

[Chazal, Guibas, O. Skraba 2011]

[Cavanna, Gardner, Sheehy 2015]





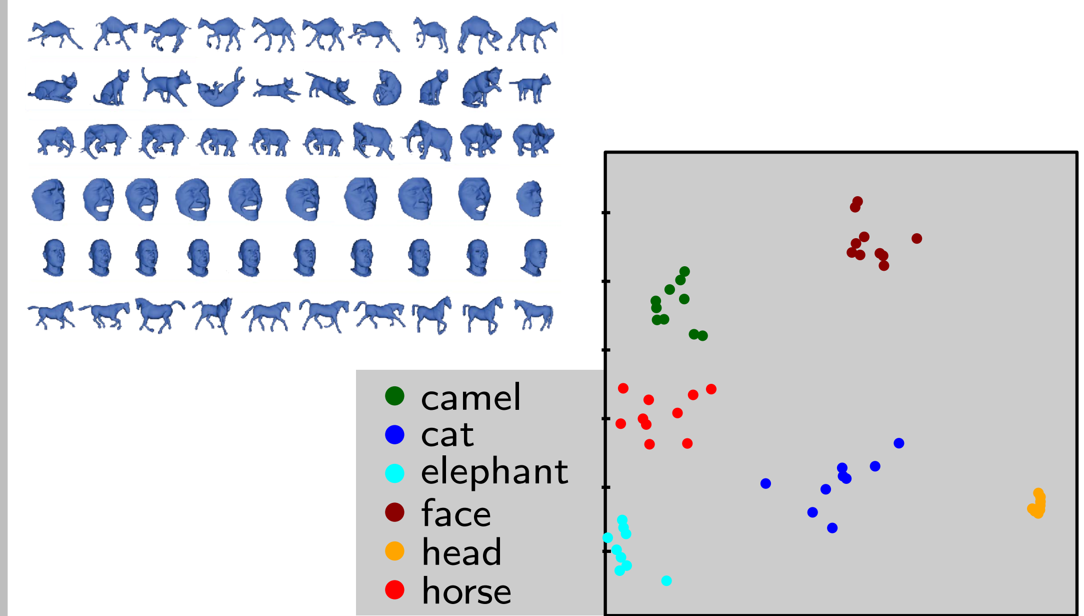
Clustering enhanced with topological invariants

[Chazal, Guibas, O., Skraba 2013]

[Bonis, O. 2015]

[Singh, Mémoli, Carlsson 2007]

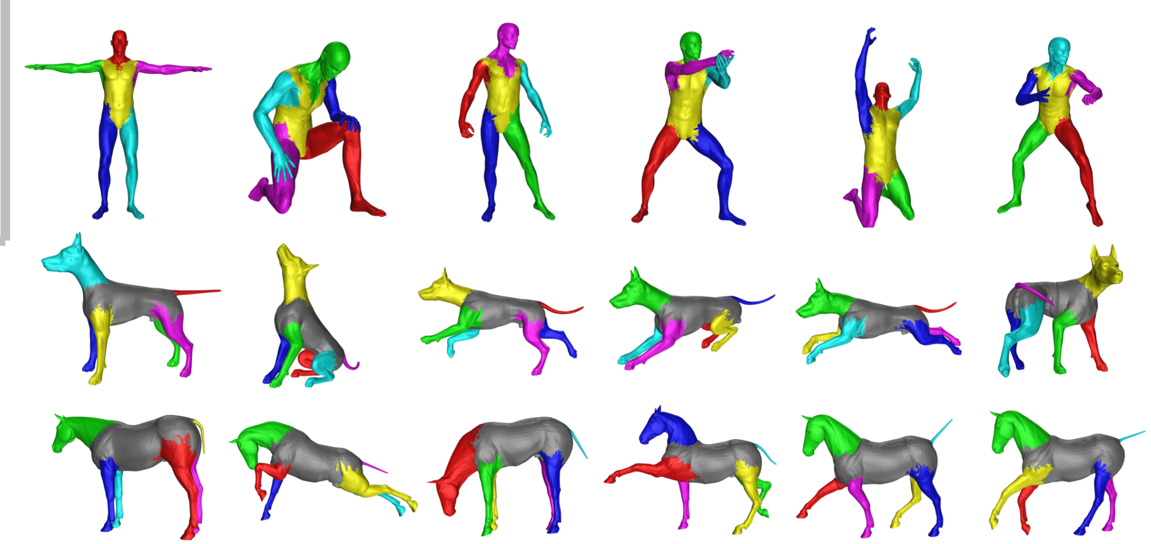
[Carrière, O. 2016] [Munch, Wang 2016]



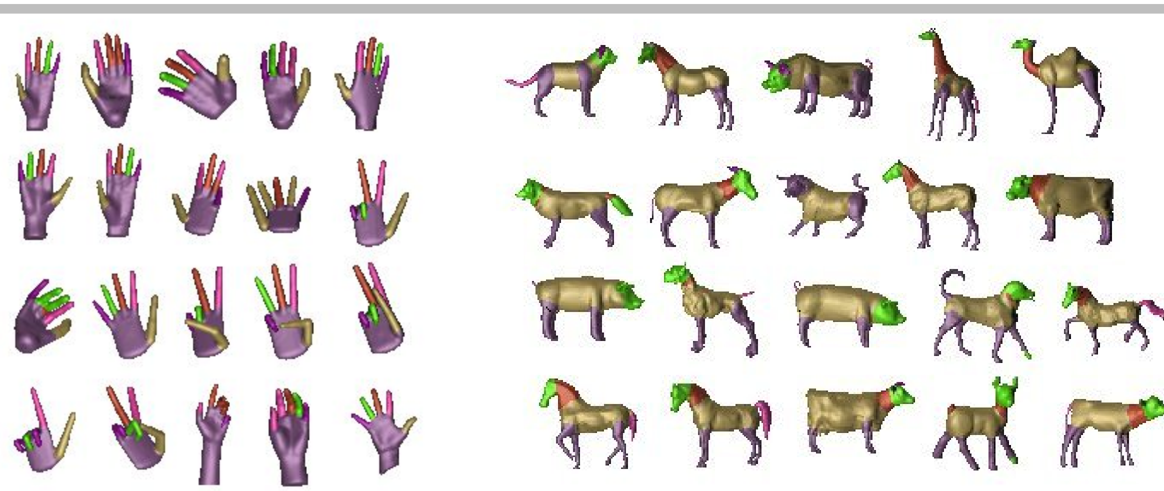
[Chazal, Cohen-Steiner, Guibas, Mémoli, O. 2009]

[Chazal, de Silva, O. 2013]

image/shape classification, segmentation, matching



[Skraba, Ovsjanikov, Chazal, Guibas 2010]



[Carrière, O., Ovsjanikov 2015]

Thank you!
ATEE