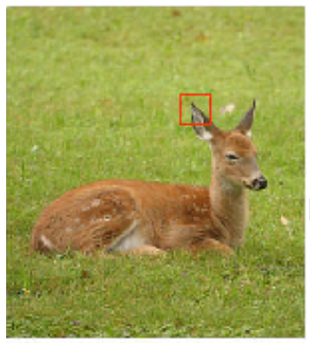
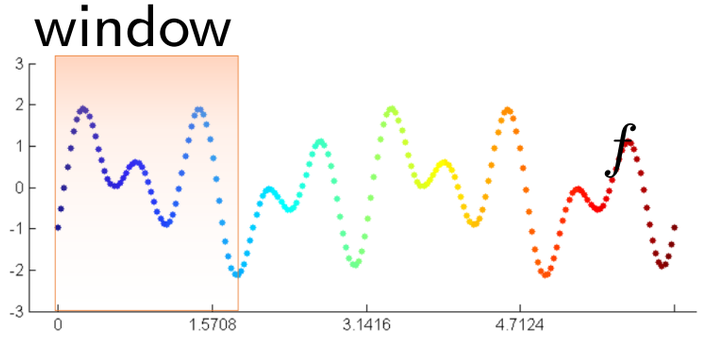
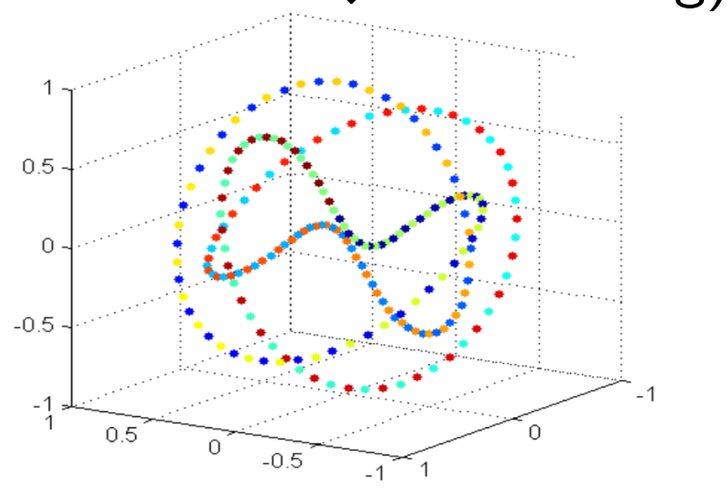
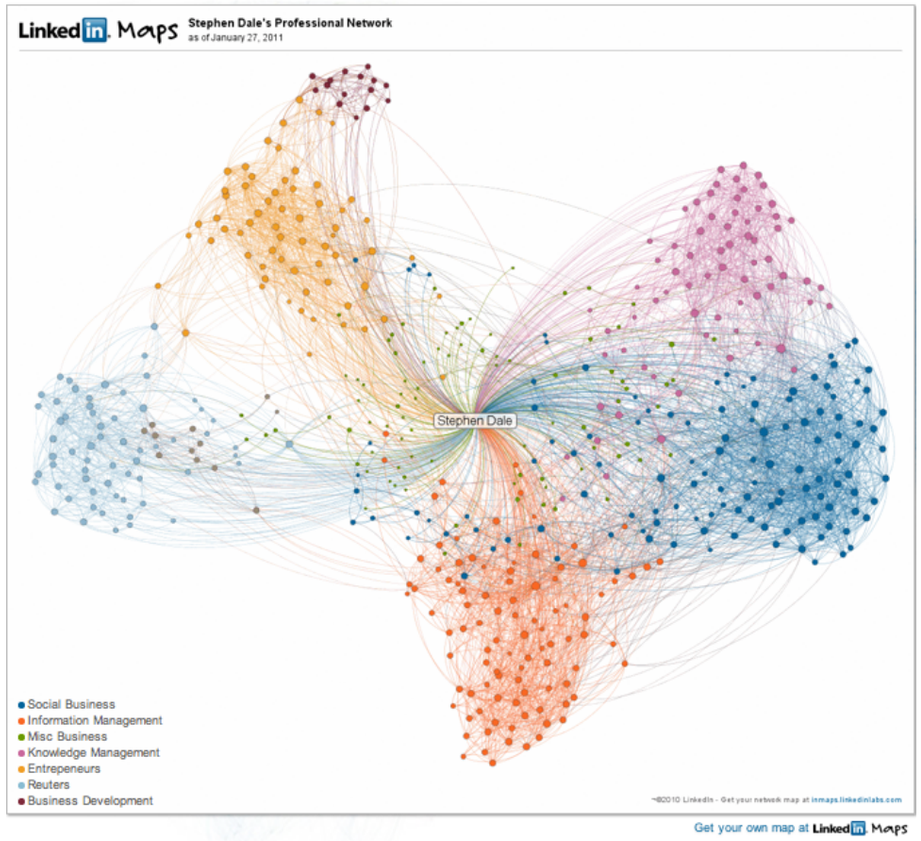
Topology in data

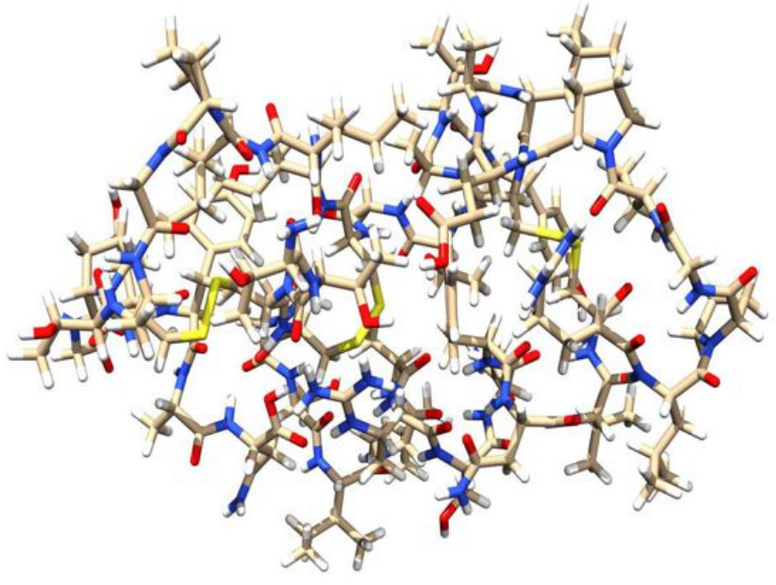


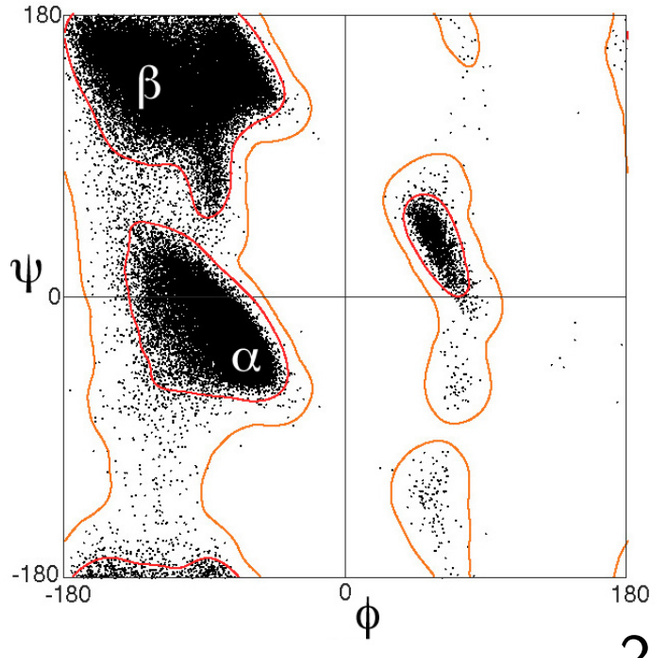


(time-delay embedding)





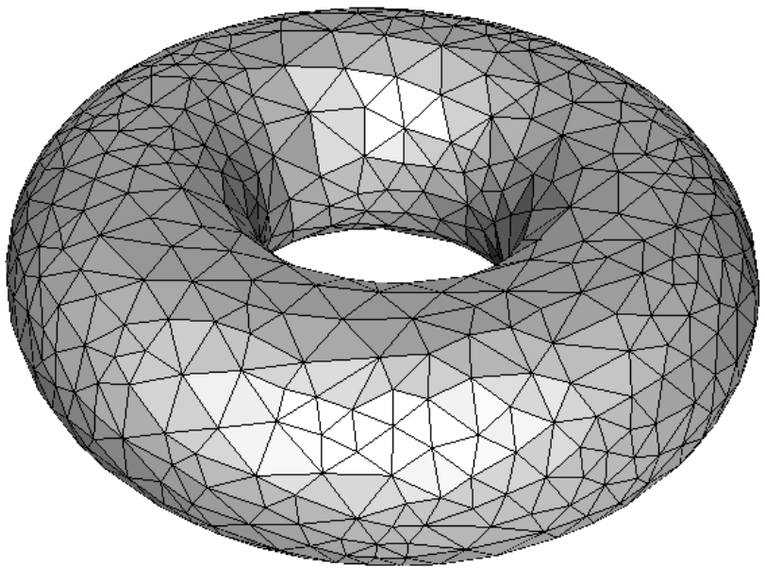




Topological Data Analysis (TDA)

topological invariants for classification

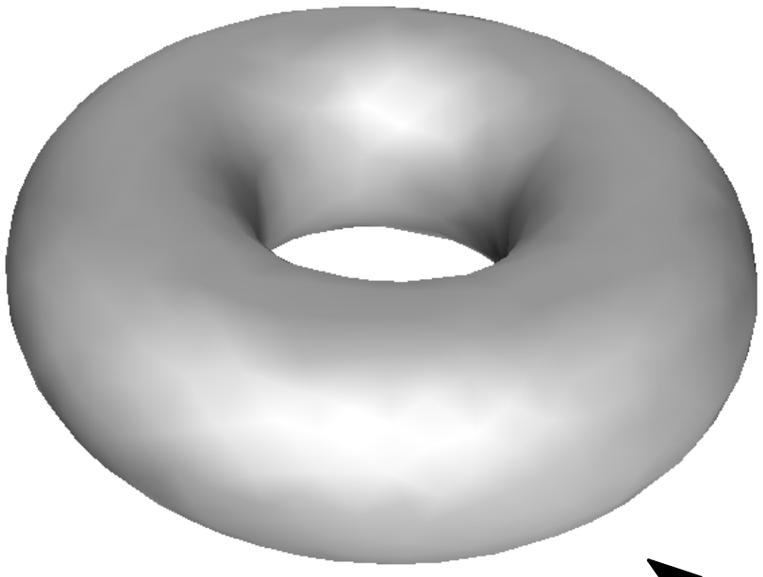
$$\beta_0 = \beta_2 = 1$$
$$\beta_1 = 2$$



triangulation

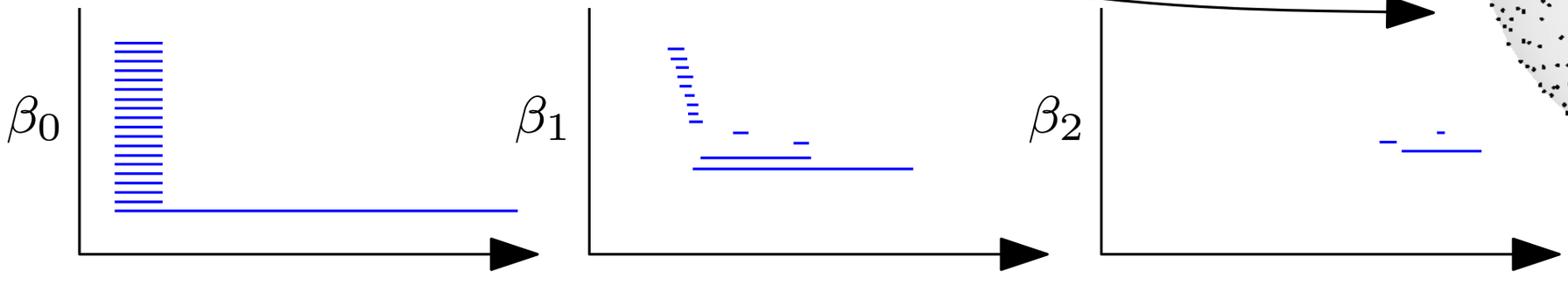
Algebraic topology

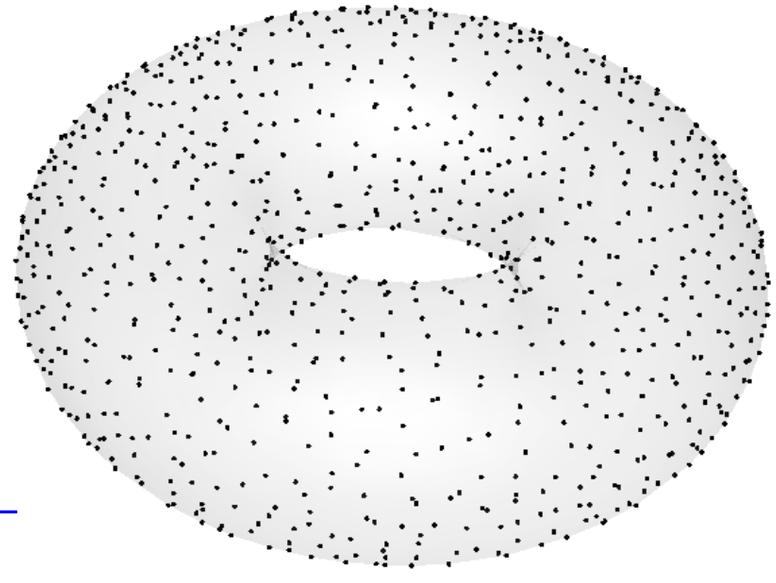
Applied algebraic topology



compact set

topological descriptors for inference and comparison



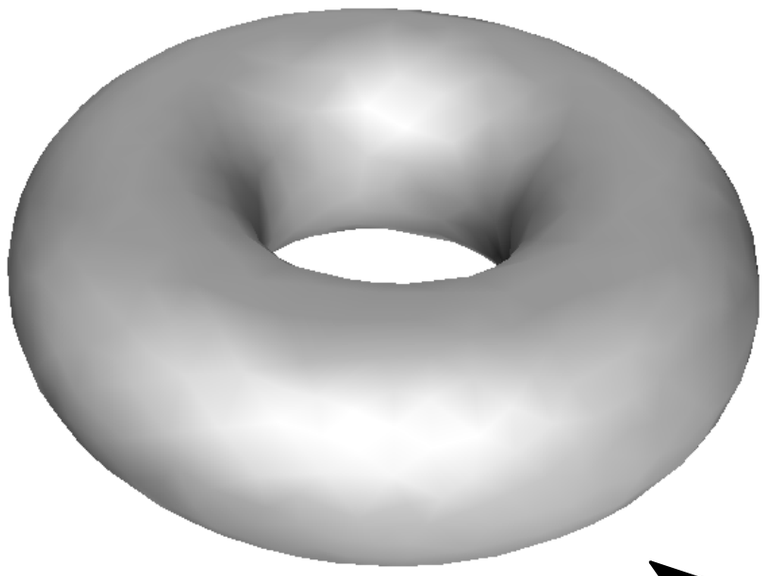


point cloud

Topological Data Analysis (TDA)

Properties of topological descriptors:

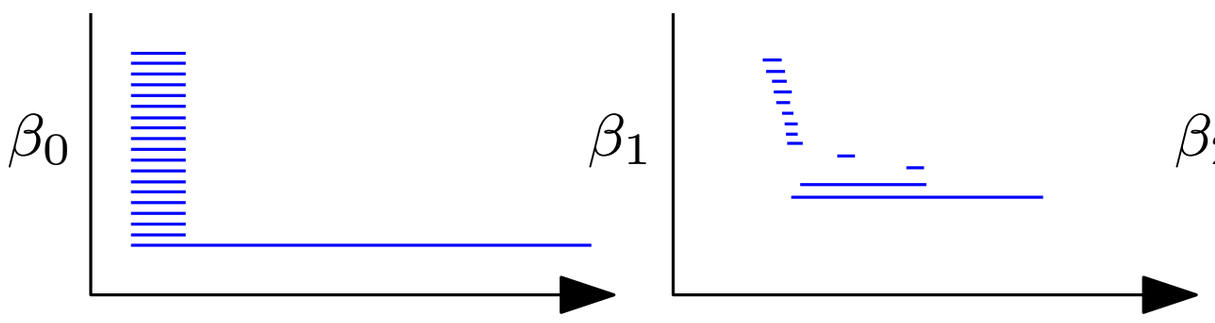
- invariant under coordinate changes
- stable with respect to perturbations
- informative

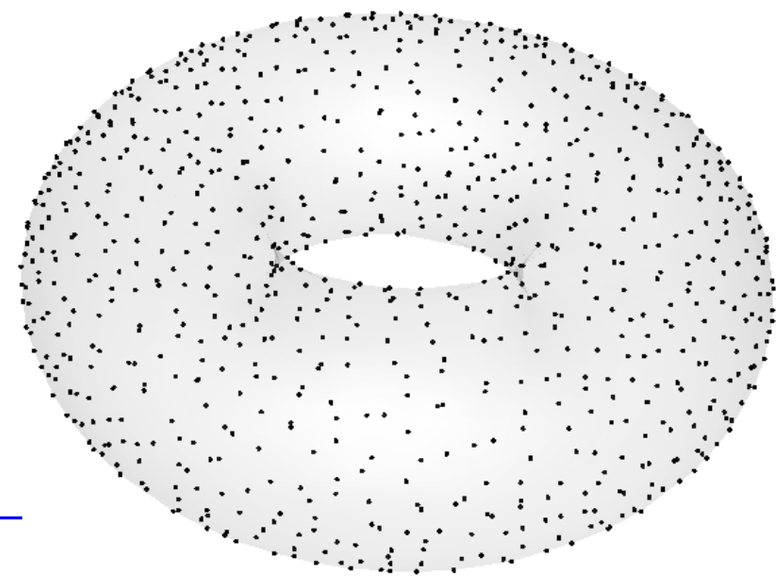


compact set

Applied algebraic topology

topological descriptors for inference and comparison



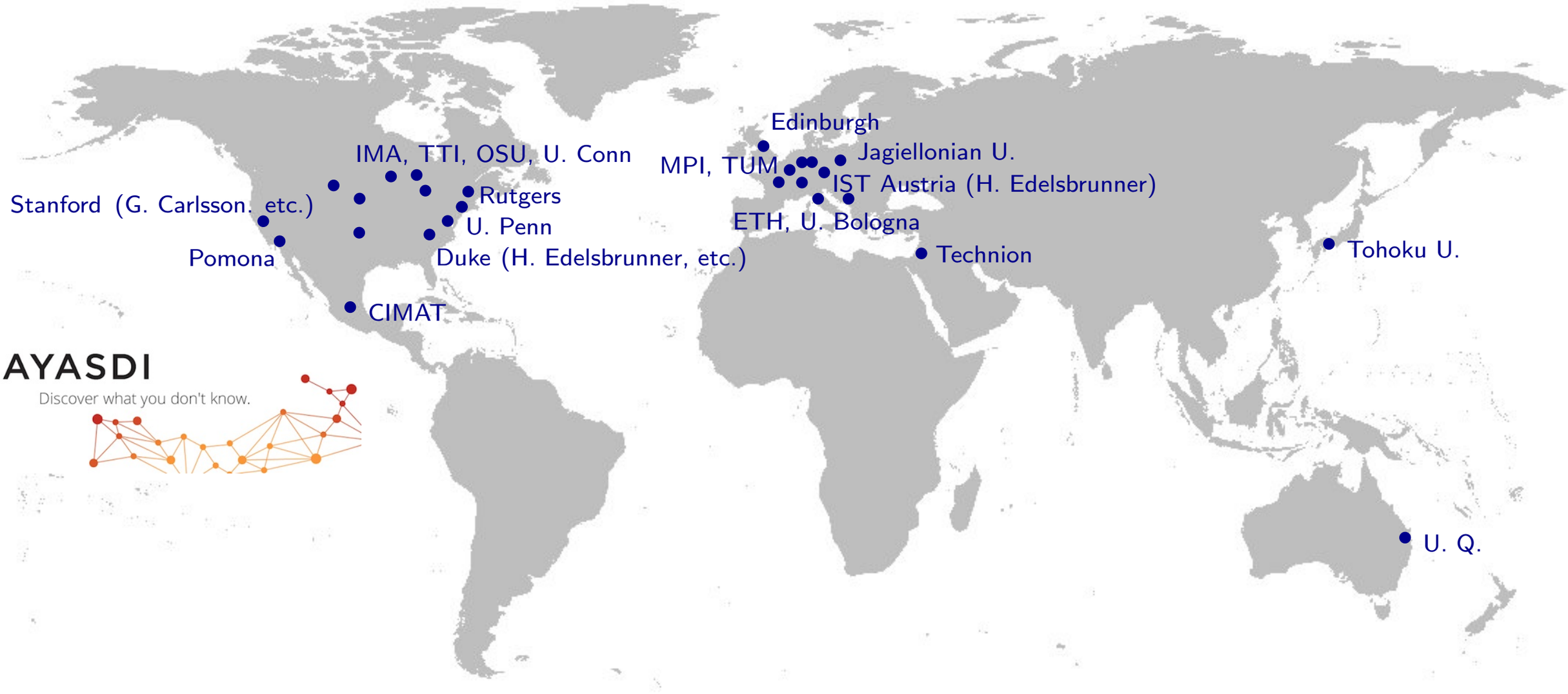


point cloud

The TDA community (as of 2002)

- 2 research groups (5-10 researchers)

The TDA community (as of today)



AYASDI

Discover what you don't know.

- ∼100 researchers working on theoretical foundations
- 200-300 researchers at the interface with applications
- successful applications and company (Ayasdi)

Some applications

- analysis of random, modular and non-modular scale-free networks and networks with exponential connectivity distribution,
- analysis of social and spatial networks like neurons, genes, online messages, air passengers, Twitter, face-to-face contact, etc.,
- coverage and hole detection in wireless sensor fields,
- multiple hypothesis tracking on urban vehicular data,
- analysis of the statistics of high-contrast image patches,
- image segmentation,
- 1d signal denoising,
- 3d shape classification/segmentation/matching,
- clustering of protein conformations,
- measurement of protein compressibility,

Some applications

- analysis of random, modular and non-modular scale-free networks and networks with exponential connectivity distribution,
- analysis of social and spatial networks like neurons, genes, online messages, air passengers, Twitter, face-to-face contact, etc.,
- coverage and hole detection in wireless sensor fields,
- multiple hypothesis tracking on urban vehicular data,
- analysis of the statistics of high-contrast image patches,
- image segmentation,
- 1d signal denoising,
- 3d shape classification/segmentation/matching,
- clustering of protein conformations,
- measurement of protein compressibility,
- identification of breast cancer subtypes,
- analysis of activity patterns in the primary visual cortex,
- identification of hidden networks in the U.S. house of representatives,

Some applications

- analysis of random, modular and non-modular scale-free networks and networks with exponential connectivity distribution,
- analysis of social and spatial networks like neurons, genes, online messages, air passengers, Twitter, face-to-face contact, etc.,
- coverage and hole detection in wireless sensor fields,
- multiple hypothesis tracking on urban vehicular data,
- analysis of the statistics of high-contrast image patches,
- image segmentation,
- 1d signal denoising,
- 3d shape classification/segmentation/matching,
- clustering of protein conformations,
- measurement of protein compressibility,
- identification of breast cancer subtypes,
- analysis of activity patterns in the primary visual cortex,
- identification of hidden networks in the U.S. house of representatives,
- analysis of 2d cortical thickness data,
- time series analysis,

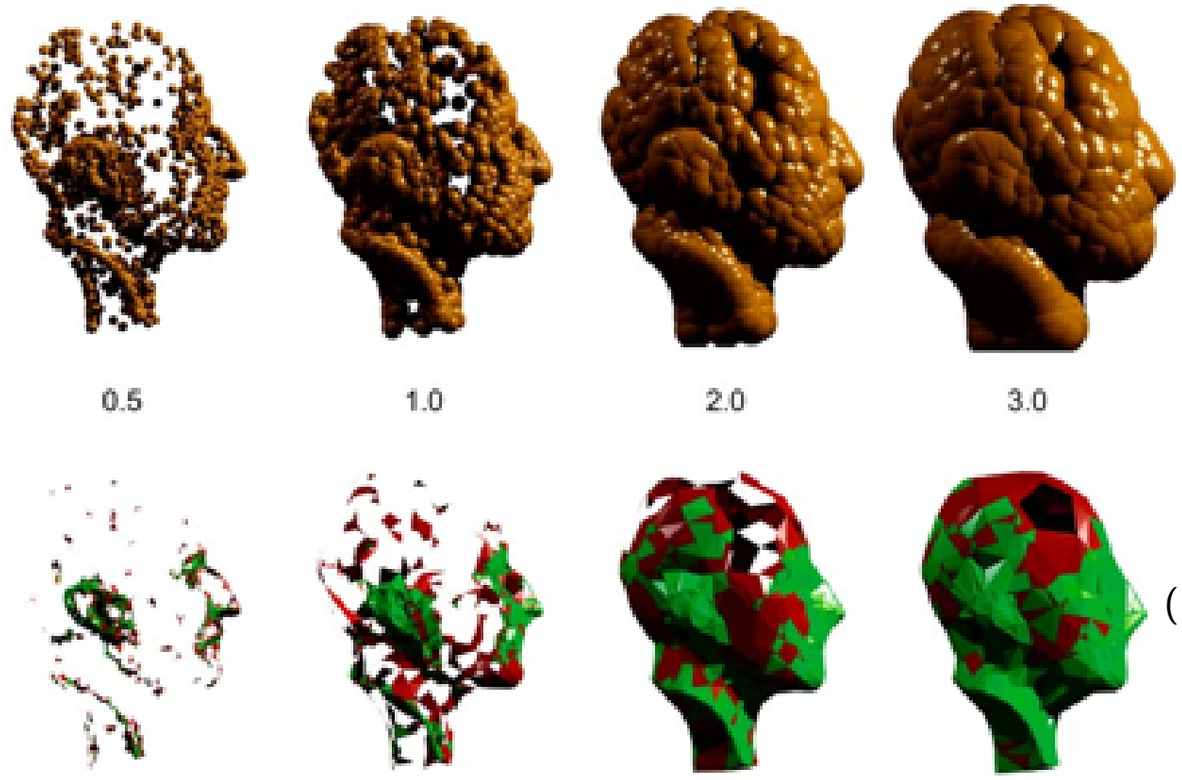
Some applications

- analysis of random, modular and non-modular scale-free networks and networks with exponential connectivity distribution,
- analysis of social and spatial networks like neurons, genes, online messages, air passengers, Twitter, face-to-face contact, etc.,
- coverage and hole detection in wireless sensor fields,
- multiple hypothesis tracking on urban vehicular data,
- analysis of the statistics of high-contrast image patches,
- image segmentation,
- 1d signal denoising,
- 3d shape classification/segmentation/matching,
- clustering of protein conformations,
- measurement of protein compressibility,
- identification of breast cancer subtypes,
- analysis of activity patterns in the primary visual cortex,
- identification of hidden networks in the U.S. house of representatives,
- analysis of 2d cortical thickness data,
- time series analysis,
- refinement of the classification of NBA players,
- discrimination of electroencephalogram signals recorded before and during epileptic seizures,
- statistical analysis of orthodontic data,
- measurement of structural changes during lipid vesicle fusion,
- characterization of the frequency and scale of lateral gene transfer in pathogenic bacteria,
- pattern detection in gene expression data,
- study of the cosmic web and its filamentary structure,

The TDA pipeline in a nutshell

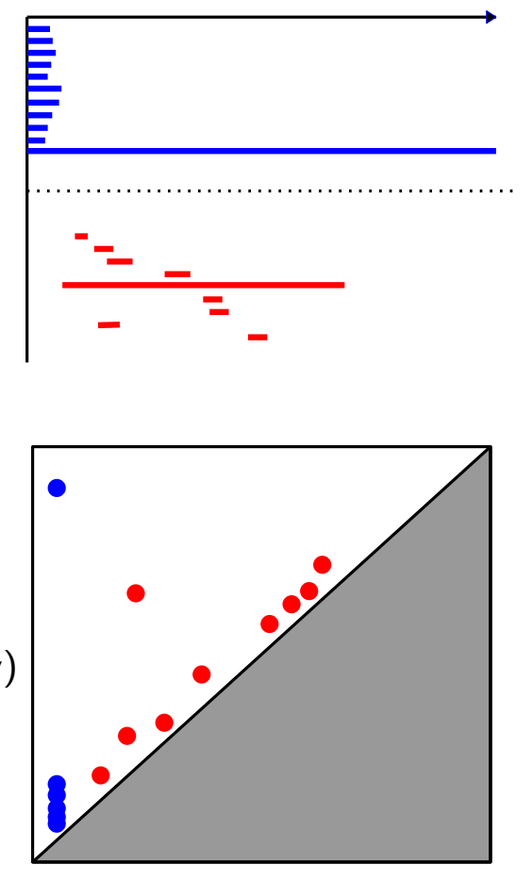
metric space

(geometry)



filtration

(homology)

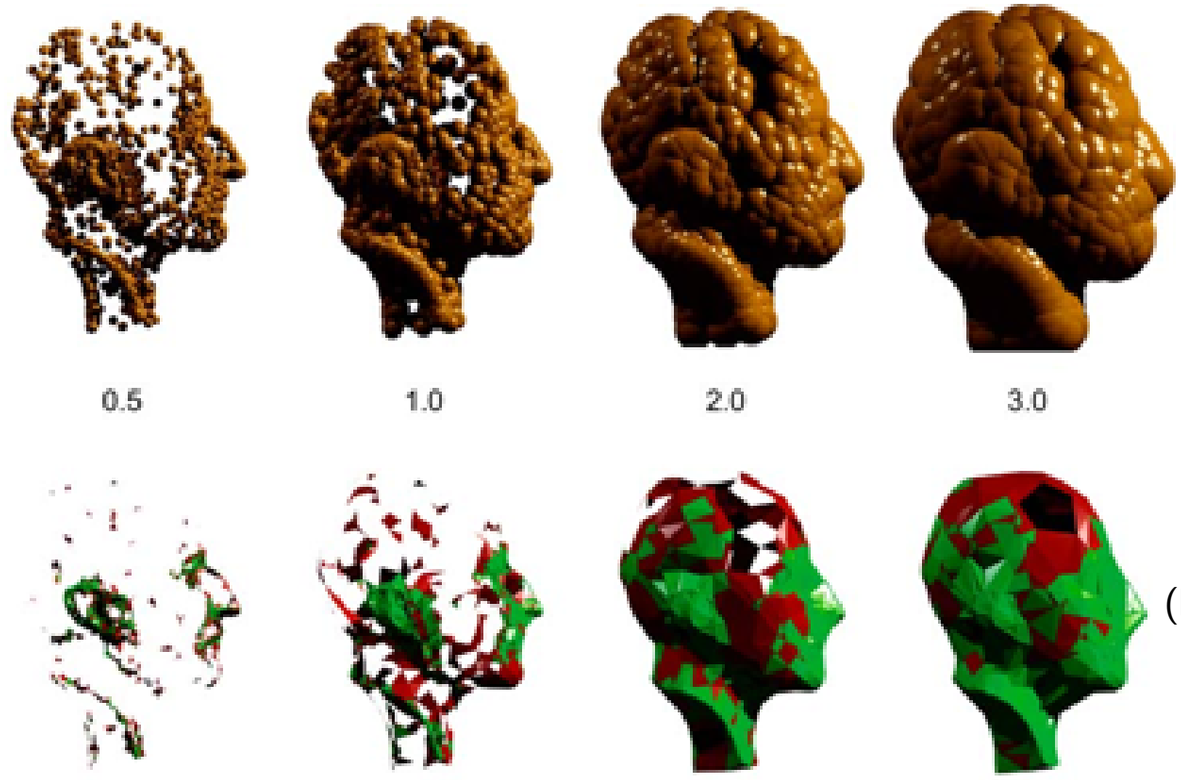


pers. barcode/diag.

The TDA pipeline in a nutshell

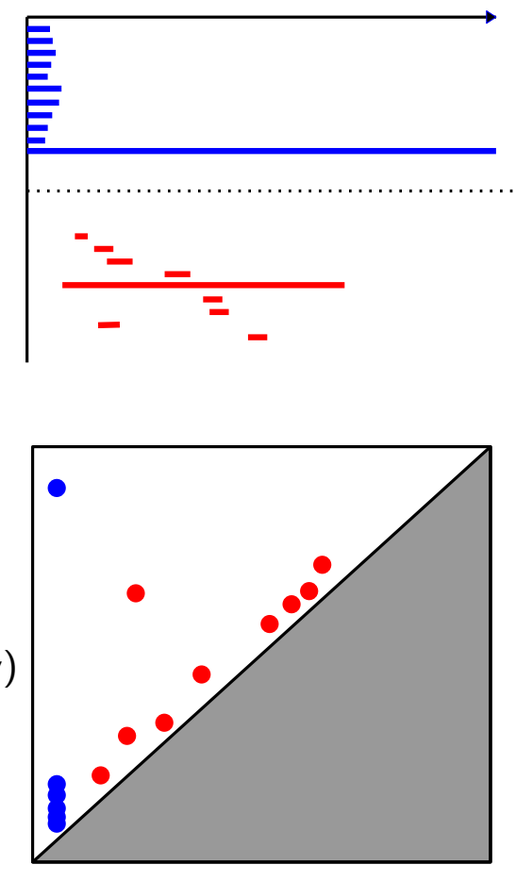
metric space

(geometry)



filtration

(homology)



pers. barcode/diag.

4 pillars to the theory (topological persistence):

- decomposition theorems ($\exists$ barcodes)
- algorithms (computation of barcodes)
- stability theorems (barcodes as stable descriptors)
- statistics / learning (barcodes as observations)