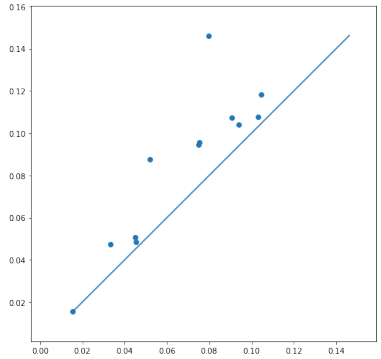
Section 17 Topological Data Analysis - An introduction

1. Overview of topology and data
2. Homology and Betti numbers
3. Persistence homology
4. TDA Mapper
5. Applications

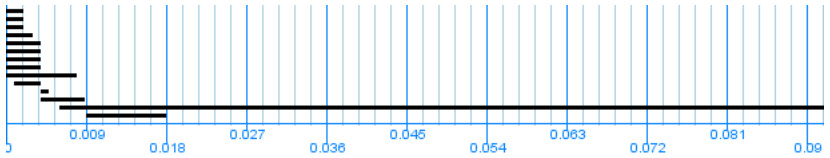
➢ Overview:

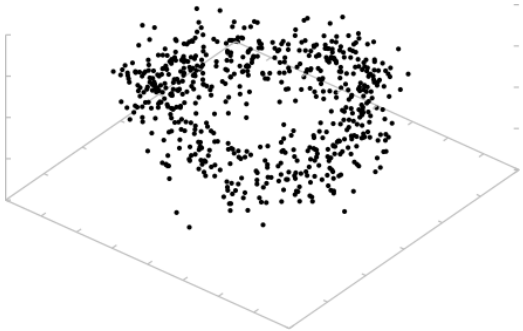
Topological Data Analysis (TDA) an approach to the analysis of datasets using techniques from topology.

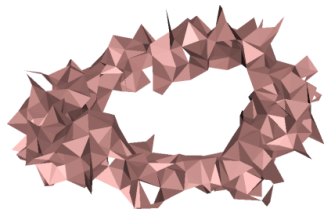
TDA is a generalization of clustering and manifold learning.



1. Fit topological spaces to data.
2. Compute topological invariants of spaces.
3. Apply ML to the invariants.
4. Inference the properties of the data.



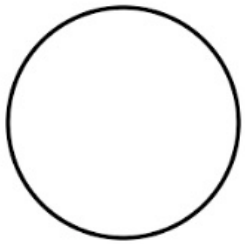


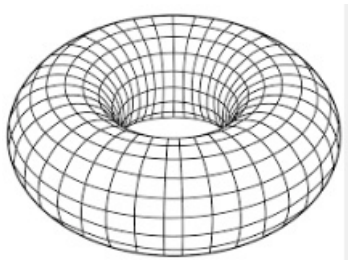


➢ Topological Spaces:

- **Mathematical definition:** A topological space means a set X with a family of subsets, so-called open sets, satisfying the property that the total set X and the empty set ∅ are open, the intersection of any two open sets is open, and an arbitrary union of open sets is open.

- Examples of topological spaces: geometric objects like lines, planes, 3-dimensional spaces, Euclidean Spaces $\mathbb{R}^n$, spheres S^n , torus, Mobiüs Band, Klein Bottle, manifolds, metric spaces etc.



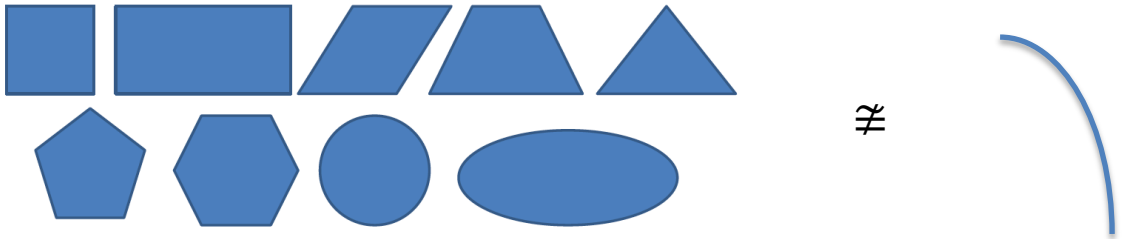


- You can learn topological spaces in Topology I, after you learned real analysis. Then you can learn algebraic topology courses using the methods from group theory, ring and fields, and homological algebra.

Ideal Goal: Classification of all topological spaces up to some **equivalent relations.**

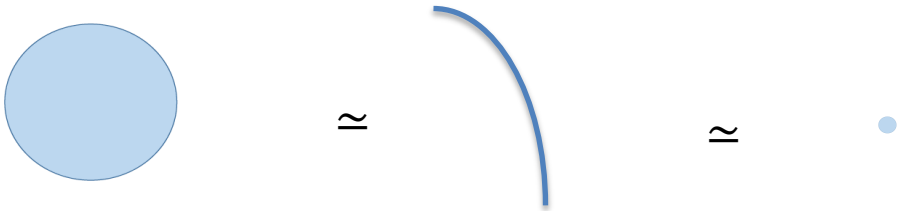
1. Homeomorphism. Continuous deformations on spaces.

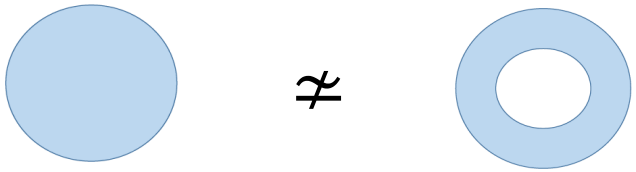
$X \cong Y$



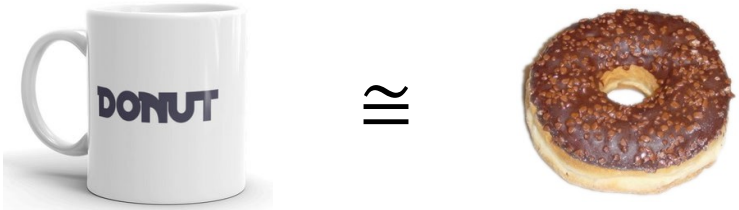
2. Homotopy equivalence. Continuous deformations on maps.

$X \simeq Y$

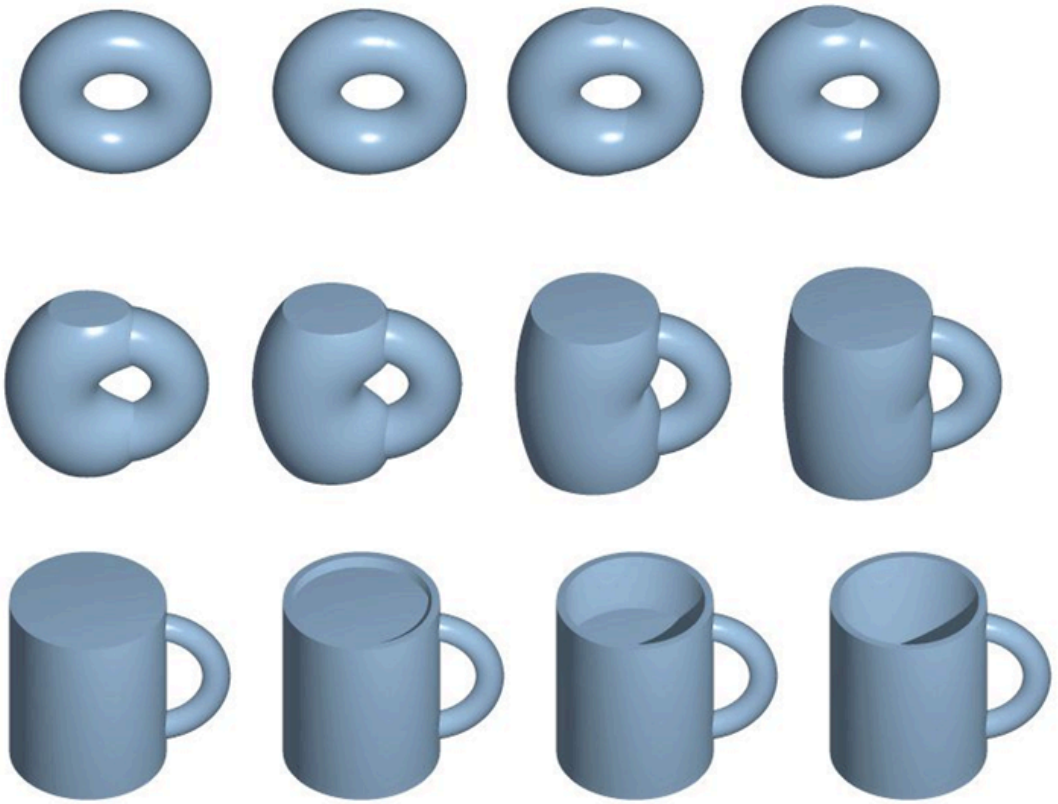




Joke: Topologists cannot tell their donut from their coffee cups.



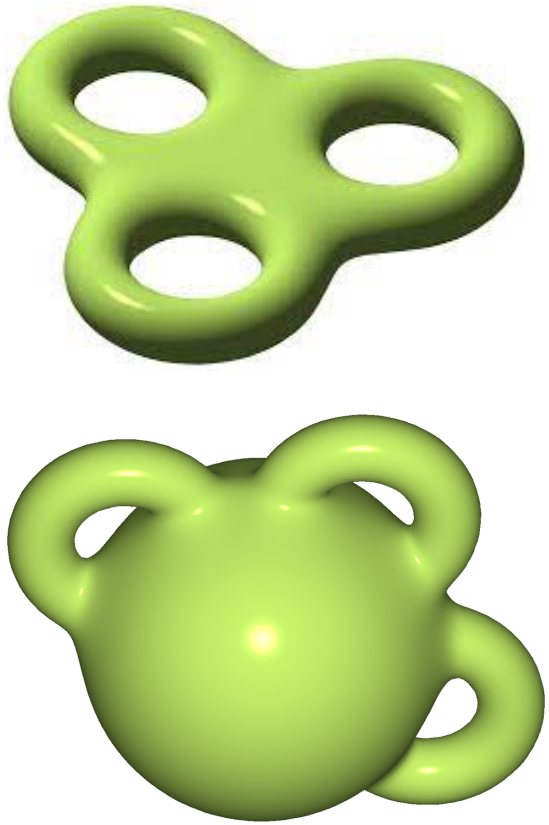
"Proof”:



A story:

- Let M be a smooth compact n-manifold.
- Let S^n be the n-sphere.

If M $\simeq$ S^n, then $M \cong S^n$.



n = 3, Perelman 2003 (2006 Fields medal), Thurston (1982 Fields), Hamilton, ... Poincare
n = 4, Freedman (1982) (1986 Fields medal).
n $\geq$5, Smale (1961) (Fields1966, Wolf). (n = 5, Zeeman (1961), n = 6, Stallings (1962)....)

➢ The Euler Characteristic

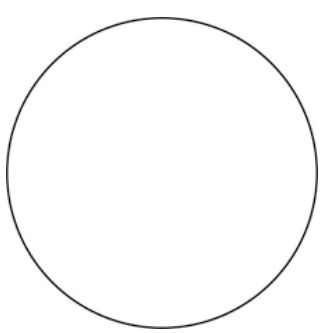
Theorem: The Euler Characteristic defined by

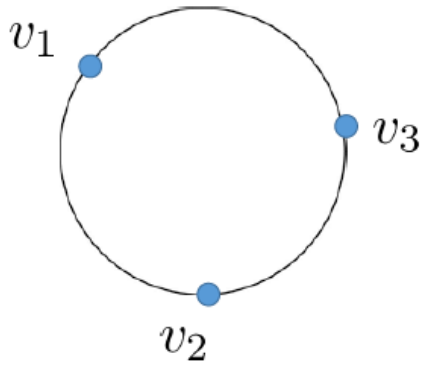
$$\chi := (\#\text{vertices}) - (\#\text{edges}) + (\#\text{faces})$$

is independent of the triangulation, and invariant under topological deformations.

Example: Circle

$$S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$$





$$\#\text{ vertices } = 3$$

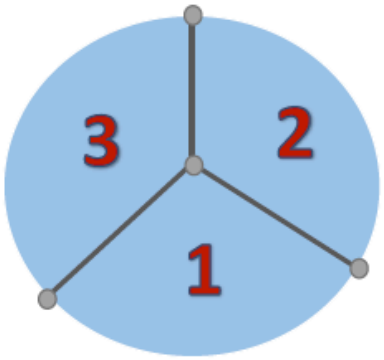
$$\#\text{ edges } = 3$$

$$\#\text{ faces } = 0$$

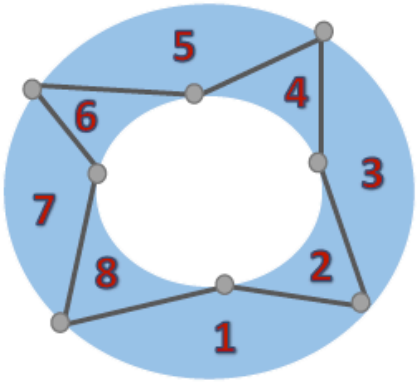
$$\chi(S^1) = 3 - 3 + 0$$

$$= 0$$

Euler Characteristic



Triangulations



D

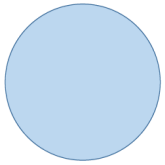
(#vertices) – (# edges) + (#faces)

A

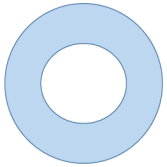
4 – 6 + 3 = 1

8 – 16 + 8 = 0

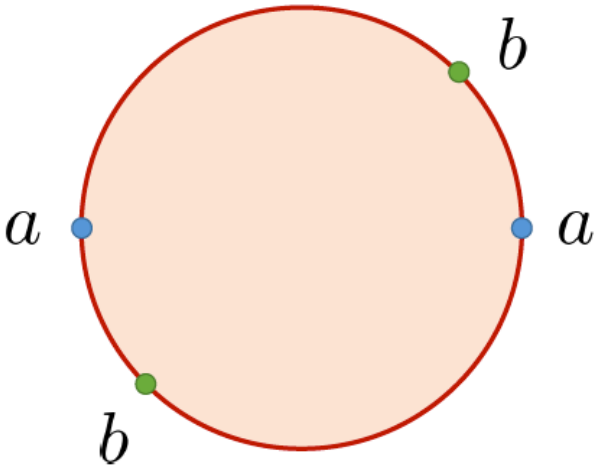
So,

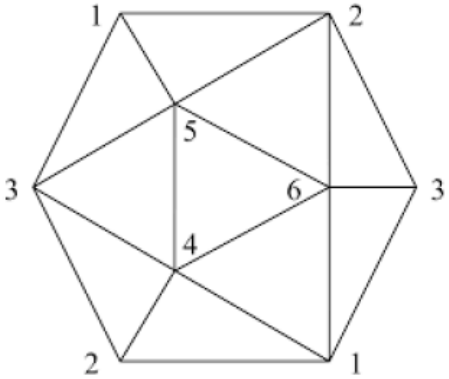


$\not\simeq$



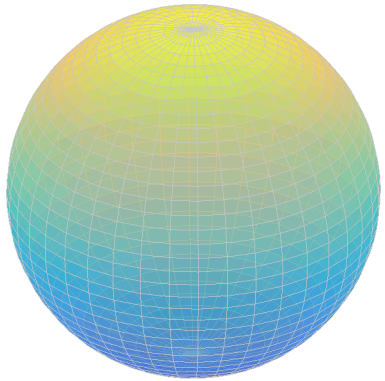
The projective plane $\mathbb{RP}^2 = \{V \leq \mathbb{R}^3 \,:\, \dim_{\mathbb{R}}(V) = 1\}$





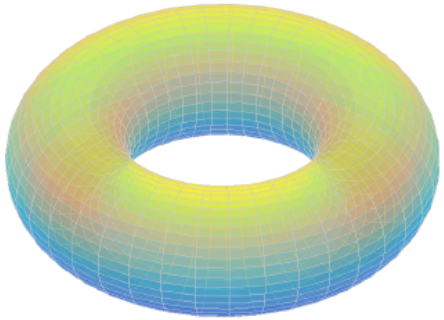
$$\chi(\mathbb{RP}^2) = 1$$

The sphere S^2

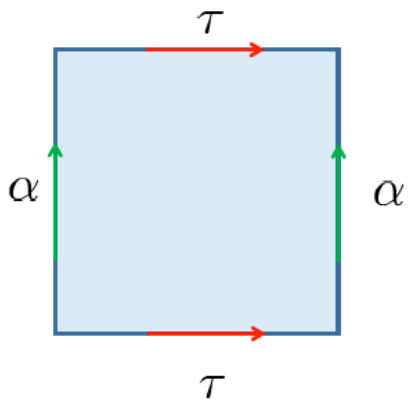


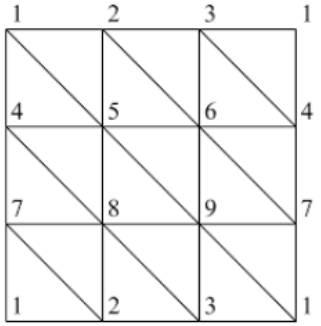
$$\chi(S^2) = 2$$

The torus



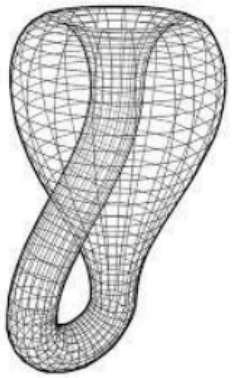
$S^1 \times S^1$



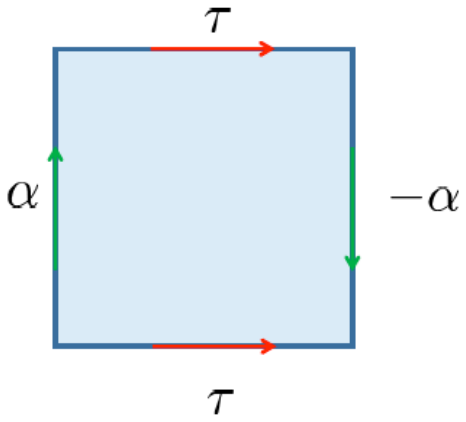


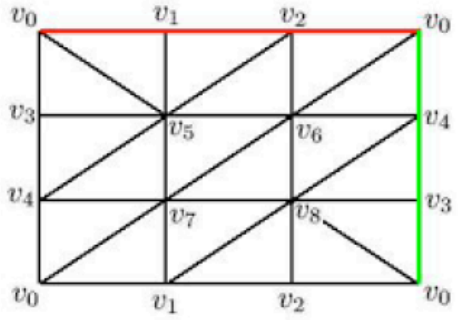
$\chi(S^1 \times S^1) = 0$

The Klein bottle



K





$\chi(K) = 0$

Algebraic Topology

A basic goal in algebraic topology:

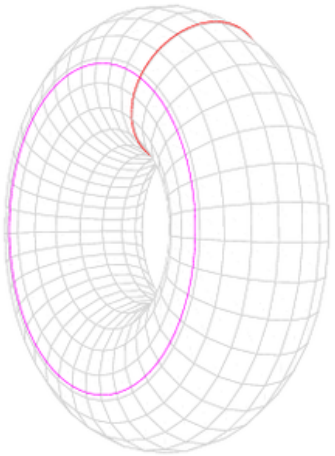
"Find algebraic invariants that classify topological spaces up to homeomorphism, or up to homotopy equivalence."

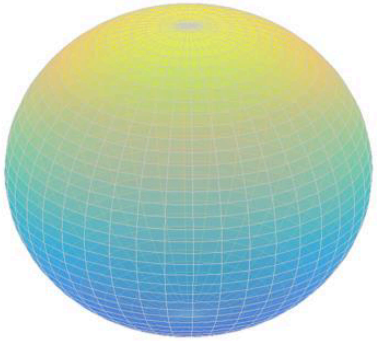
Let X be a connected CW-complex.

- Fundamental group $\pi_1(X)$.
- Homology groups $H_*(X;\mathbb{Q}) \Longrightarrow$ Betti numbers $b_i(X)$.
- Cohomology ring $H^*(X;\mathbb{Q})$.
- Higher rational homotopy groups $\pi_*(X) \otimes \mathbb{Q}$.
- Homotopy Lie algebra $\pi_*(\Omega X) \otimes \mathbb{Q}$.
- .....

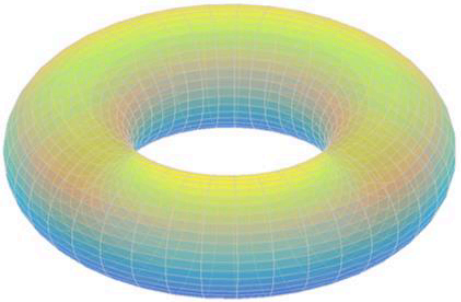
Betti Numbers:

$$\beta_n(X) \sim \ \# \text{ number of } n\text{-dim holes}$$





$$\begin{aligned} \beta_0(S^2) &= 1 \\ \beta_1(S^2) &= 0 \\ \beta_2(S^2) &= 1 \end{aligned}$$



$$\begin{aligned} \beta_0(T) &= 1 \\ \beta_1(T) &= 2 \\ \beta_2(T) &= 1 \end{aligned}$$

Theorem: Euler Characteristic

$$\chi(X) = \beta_0(X) - \beta_1(X) + \beta_2(X) - \cdots$$

- **Coordinate invariance:** topological features/invariants do not rely on any coordinate system.) no need to have data with coordinate or to embed data in spaces with coordinates.
- **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.

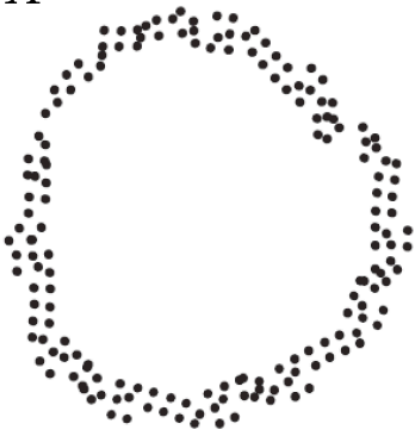
➢ Persistent homology

- A general mathematical framework to encode the evolution of the topology (homology) of families of nested spaces (filtered complex, sublevel sets,...).
- Multiscale topological information.
- Barcodes/persistence diagrams can be efficiently computed.
- Stability properties
- Formalized in its present form by H. Edelsbrunner (2002) et al and G. Carlsson et al (2005) - wide development during the last two decades:

    - 2005: stability of persistence for continuous functions (D. Cohen-Steiner et al).
    - 2009 - 2012: algebraic stability of persistence modules (F.C. et al).
    - 2014: the GUDHI software plateform (J.-D. Boissonnat et al). Also several other softs since 2005: Dionysus, (J)Plex, PHAT,...
    - Last few years: statistical aspects of persistence and machine learning.

➢ The Topology of Point Cloud Data

X



$\beta_0(X) = \#(X)$

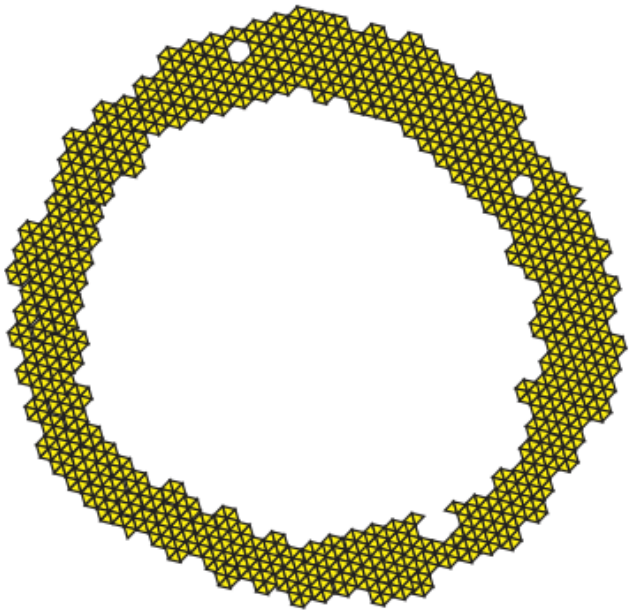
$\beta_1(X) = 0$

- Is it a circle or dots?

  A **point cloud**

- How to find robust topology at different scales?

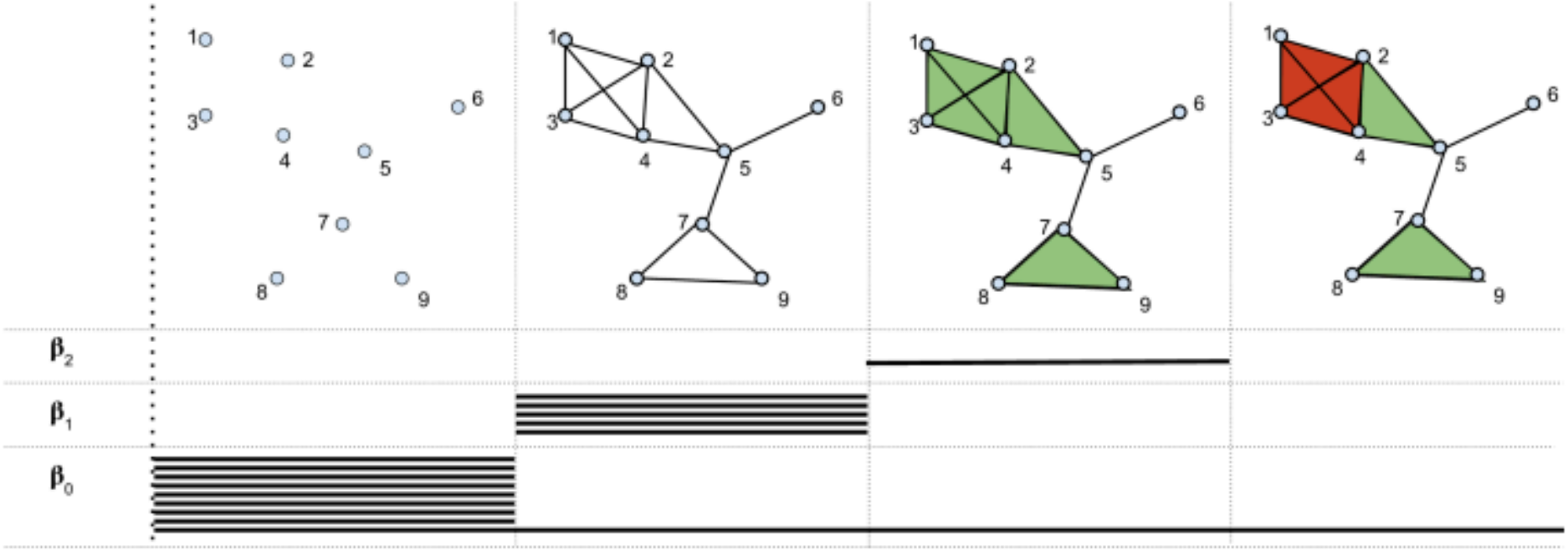
  Associate to a **point cloud** a sequence of simplicial complexes.

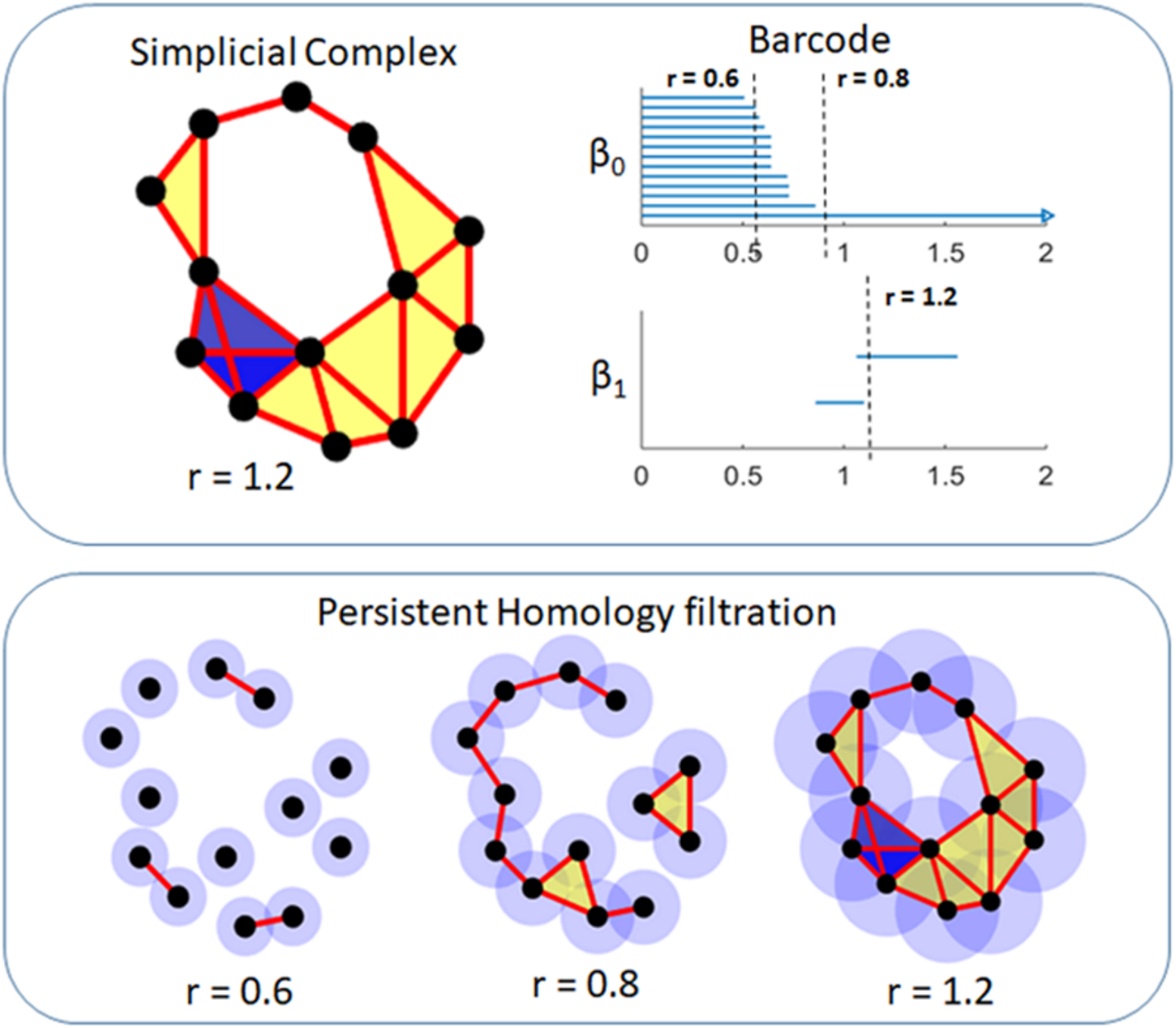


$\beta_0 = 1$
$\beta_1 = 3$

Local connections are noisy, depending on observer's scale!

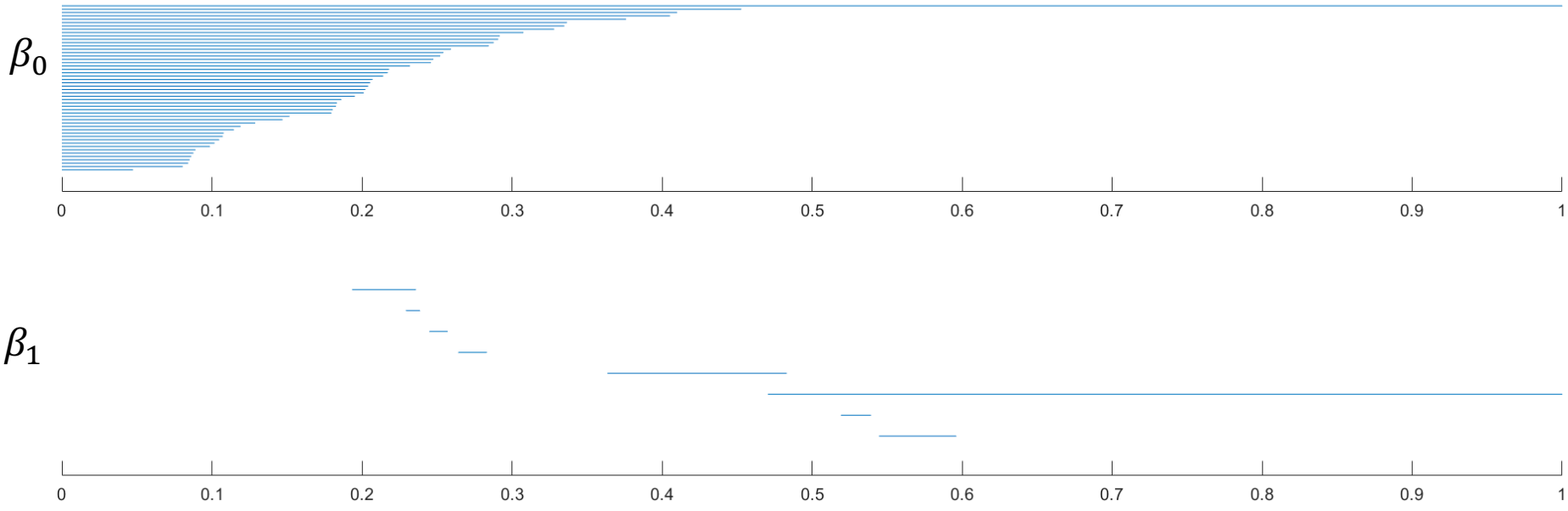
Betti numbers, Persistence barcodes



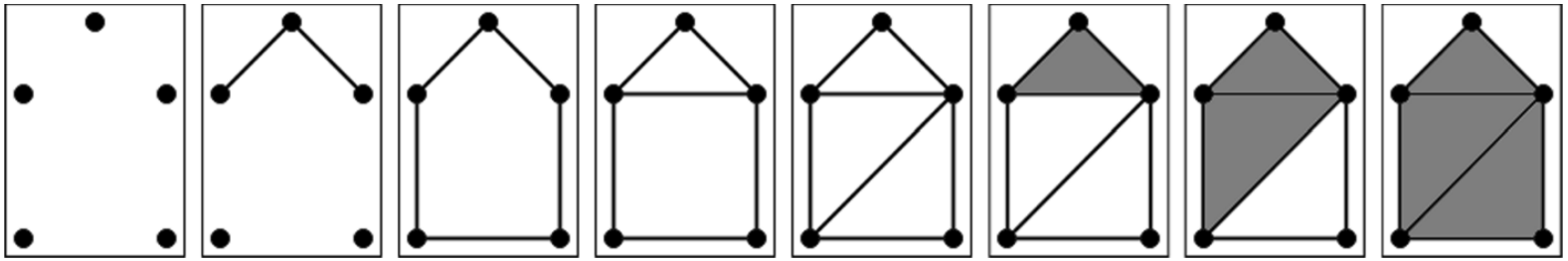

Simplicial Complex
r = 1.2
Barcode
r = 0.6
r = 0.8
β0
0
0.5
1
1.5
2
r = 1.2
β1
0
0.5
1
1.5
2
Persistent Homology filtration
r = 0.6
r = 0.8
r = 1.2

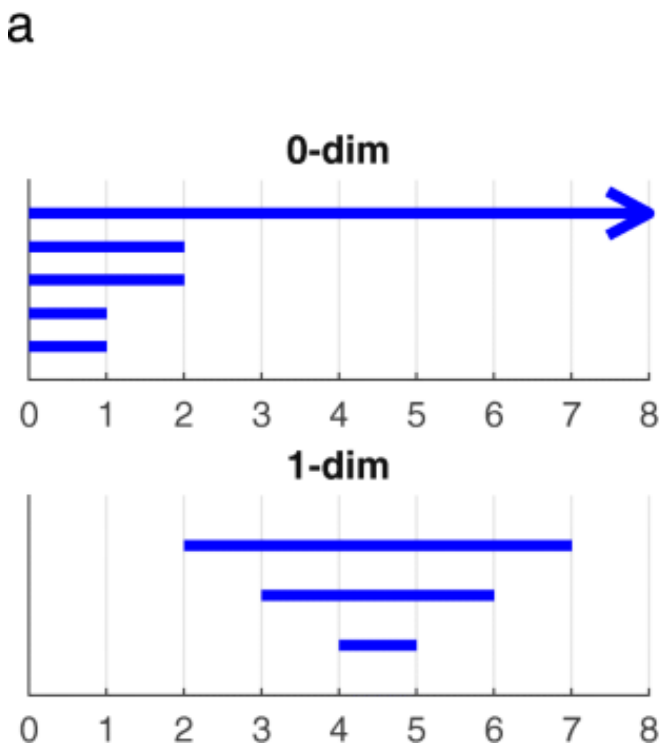
Filtration

Persistence barcodes

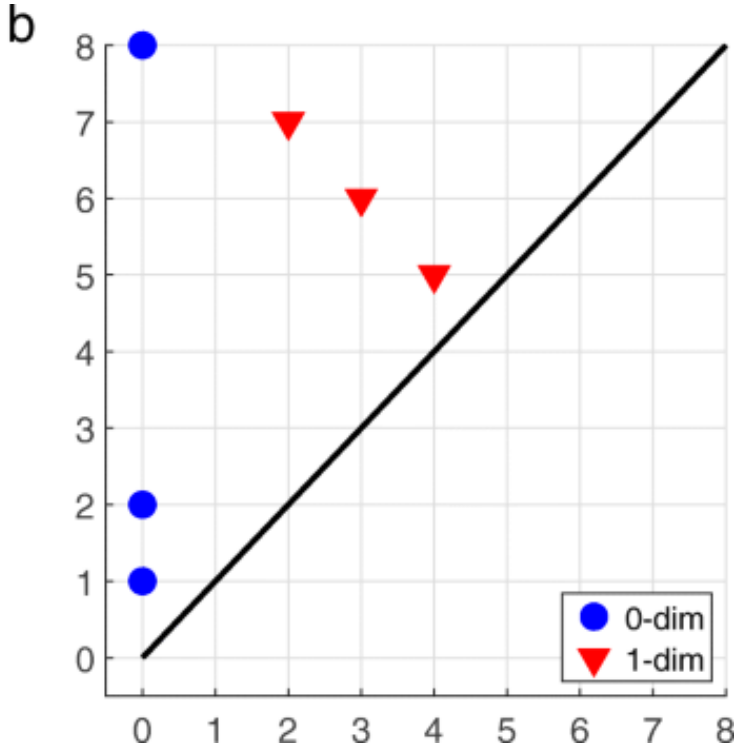


Persistence **barcodes** and Persistence **diagrams**



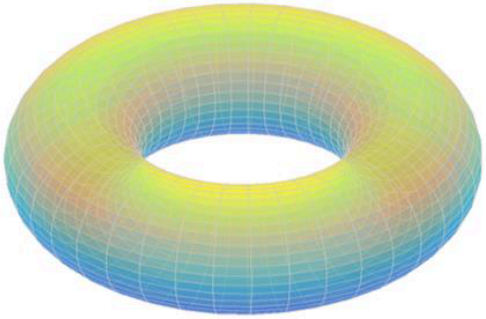


Persistence barcodes

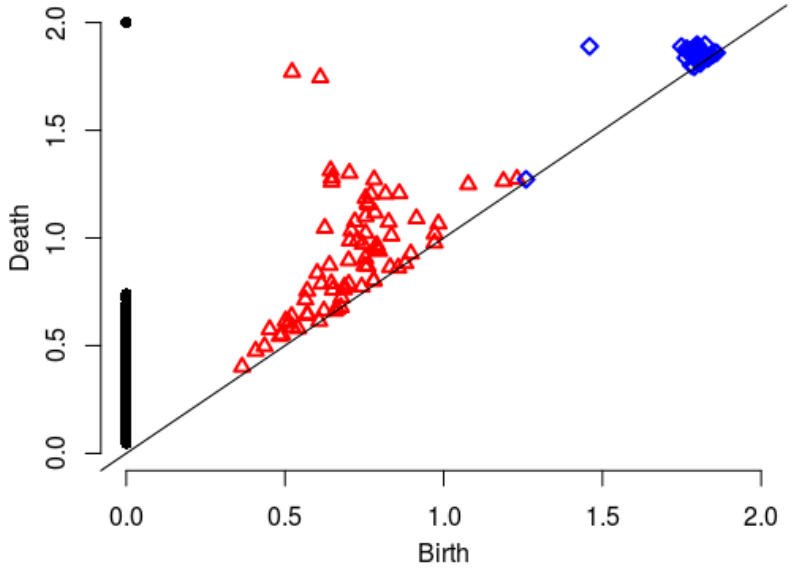


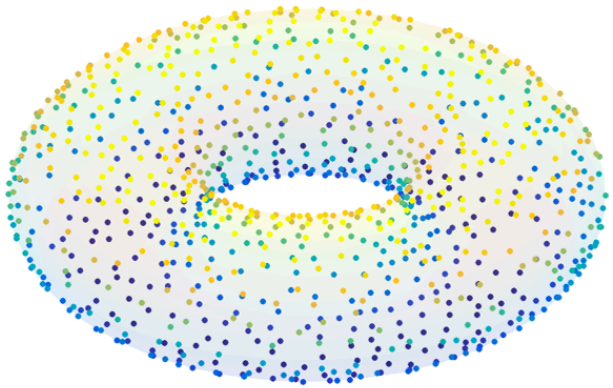
Persistence diagrams

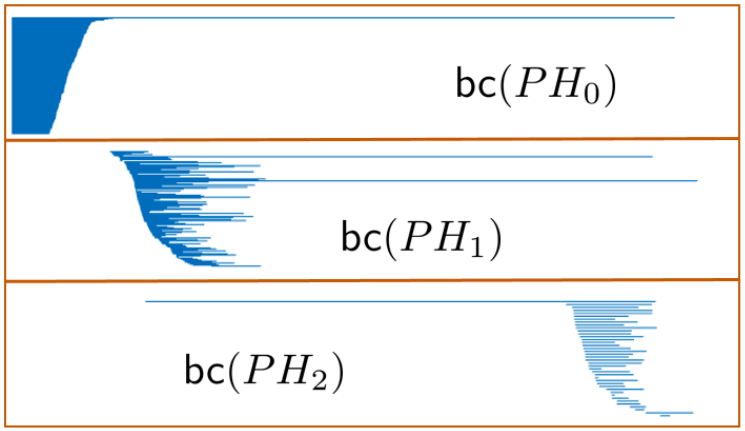
Betti Numbers



$$\beta_0(T) = 1$$
$$\beta_1(T) = 2$$
$$\beta_2(T) = 1$$

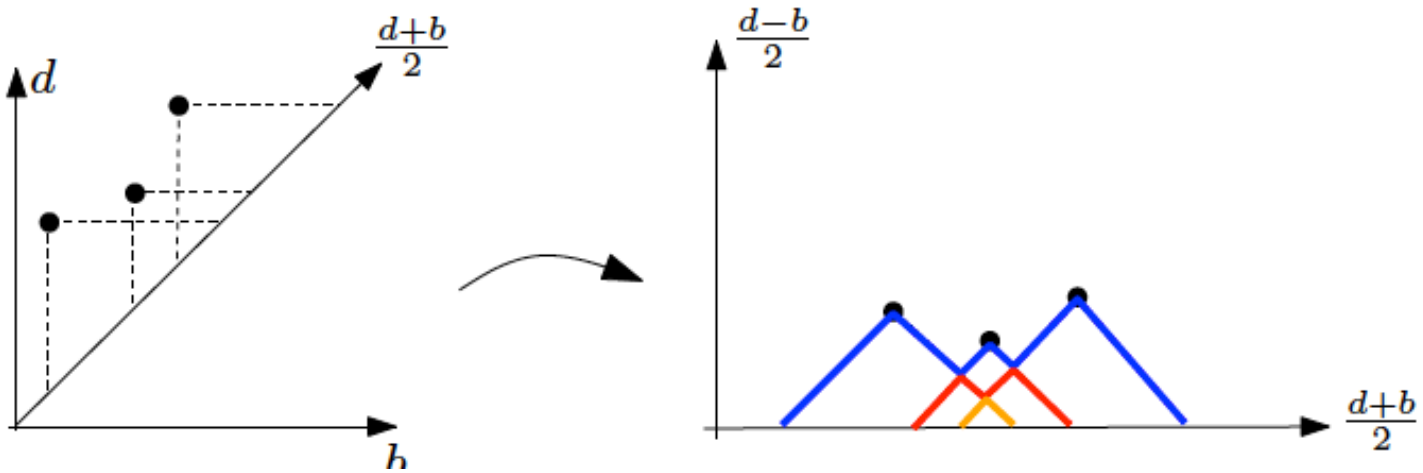






Barcodes

Persistence landscapes

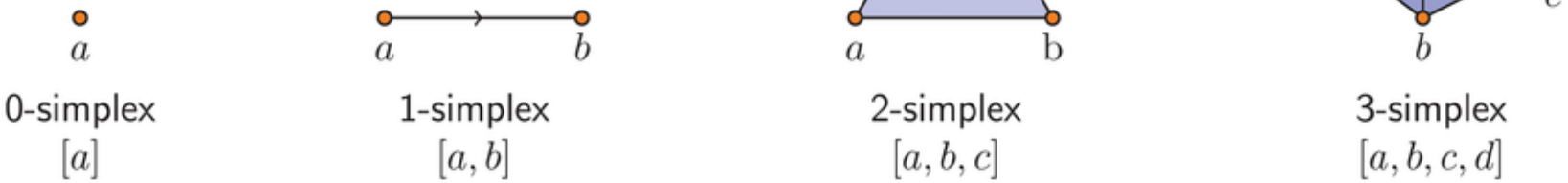


Persistence landscapes

[Bubenik 2012]

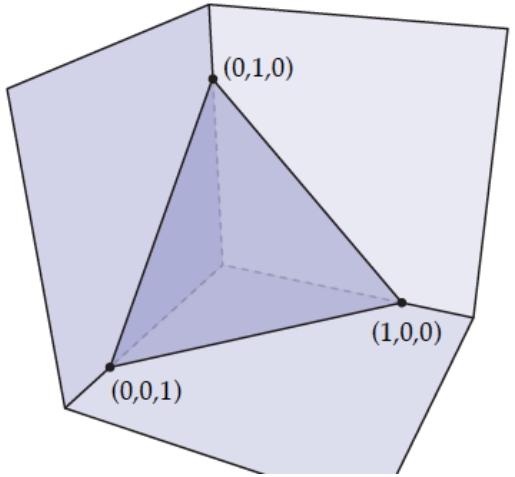
➢ Simplicial complex

Definition: A **k-simplex** is a convex hull of k+1 affinely independent points, which are called vertices.



Standard n-simplex:

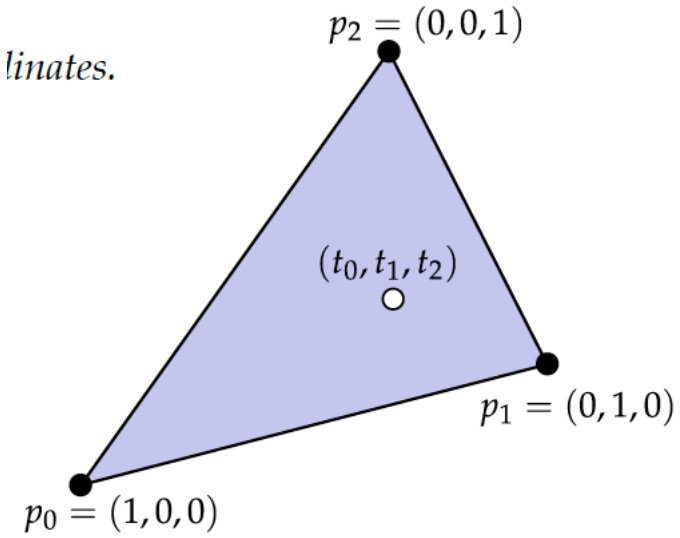
$$\sigma := \left\{ (x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \,\middle|\, \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0\ \forall i \right\}$$



Barycentric coordinates:

$$\sigma = \left\{ \sum_{i=0}^{k} t_i p_i \,\middle|\, \sum_{i=0}^{k} t_i = 1,\ t_i \geq 0\ \forall i \right\}$$

linates.

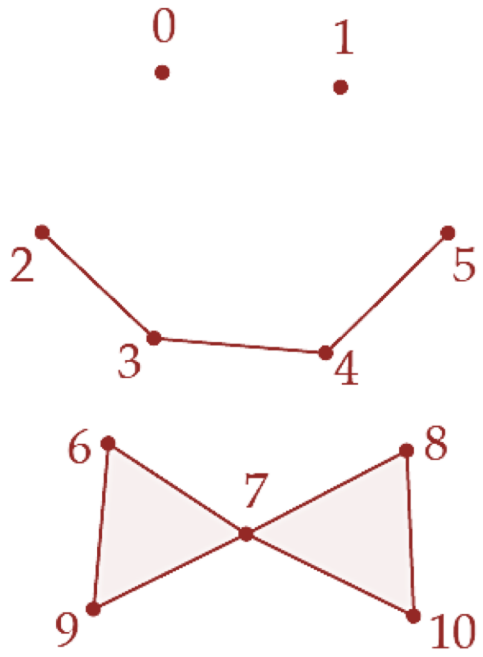


Definition: **A simplicial complex K** is a collection of simplices such that

1. The intersection of any two simplices is a simplex, and
2. Every face of every simplex in the complex is also in the complex.

- **Faces**: the simplices of K.
- **j-skeleton**: the subcomplex made of the simplices of dimension at most j.
- **Dimension of K**: the maximum of the dimensions of the faces. K is homogenous of dimension n if any of its faces is a face of a n-dimensional simplex.

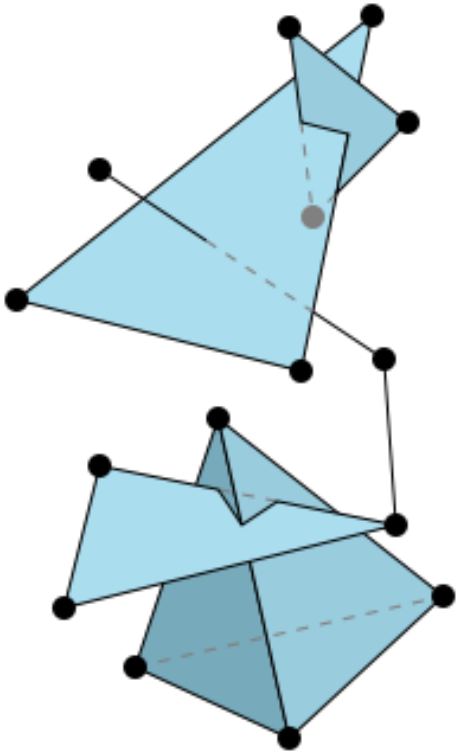
Example: What are all the simplices in this simplicial complex?



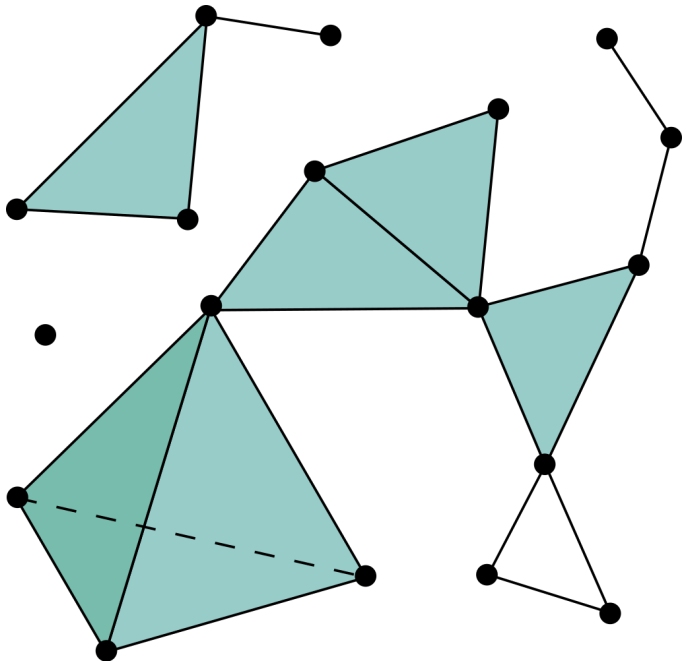
Answer:

{ø}
{0} {1} {2} {3} {4} {5} {6} {7} {8} {9} {10}
{2,3} {3,4} {4,5} {6,7} {7,9} {9,6} {7,8} {8,10} {10,7}
{6,7,9} {7,10,8}

NOT simplicial complex



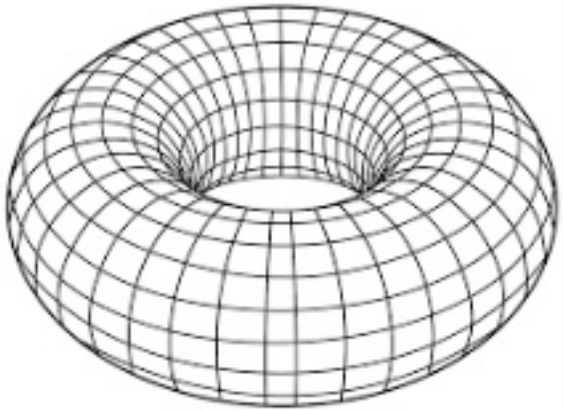
Simplicial complex



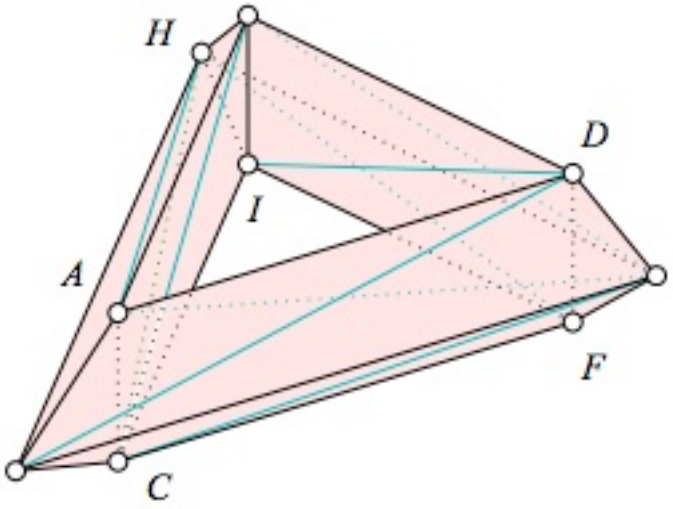
Simplicial complexes can be seen at the same time as geometric/topological spaces (good for top./geom. inference) and as combinatorial objects (abstract simplicial complexes, good for computations).

Topological spaces

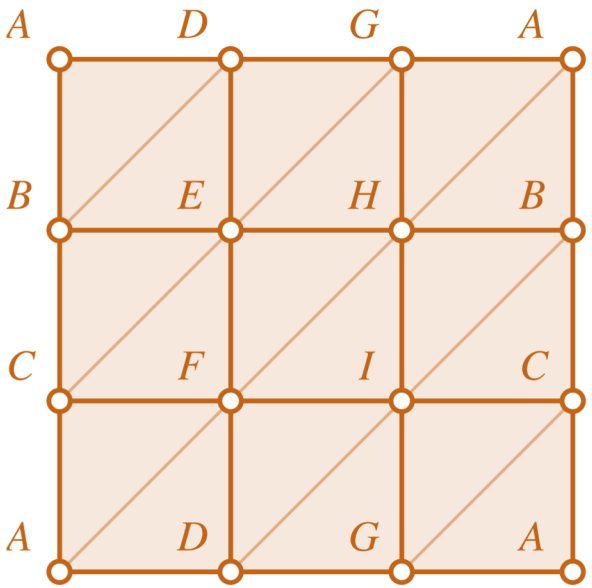
Simplicial complex



≅



Triangulation:



➢ Homology groups

Chain complexes over a field $\mathbb{F}$ or a ring R

Let K be a d-dimensional simplicial complex.

For each $0 \leq k \leq d$, Let $\{\sigma_1, \dots, \sigma_p\}$ be the set of all k simplices of K.

The k-chain:

$$c = \sum_{i=1}^{p} \varepsilon_i \sigma_i$$

Sum and **scalar product** of k-chains:

$$c + c' = \sum_{i=1}^{p} (\varepsilon_i + \varepsilon'_i)\sigma_i \quad \text{and} \quad \lambda.c = \sum_{i=1}^{p} (\lambda \varepsilon'_i)\sigma_i$$

Group (vector space) of k-chains is $C_k(K)$ is the vector spaces of all k-chains.

The **chain complex** $\boldsymbol{C_*(K)}$ is the sequence of chain groups C_p connected by boundary homomorphisms,

$$\ldots \xrightarrow{\partial_{p+2}} \mathrm{C}_{p+1} \xrightarrow{\partial_{p+1}} \mathrm{C}_p \xrightarrow{\partial_p} \mathrm{C}_{p-1} \xrightarrow{\partial_{p-1}} \ldots$$

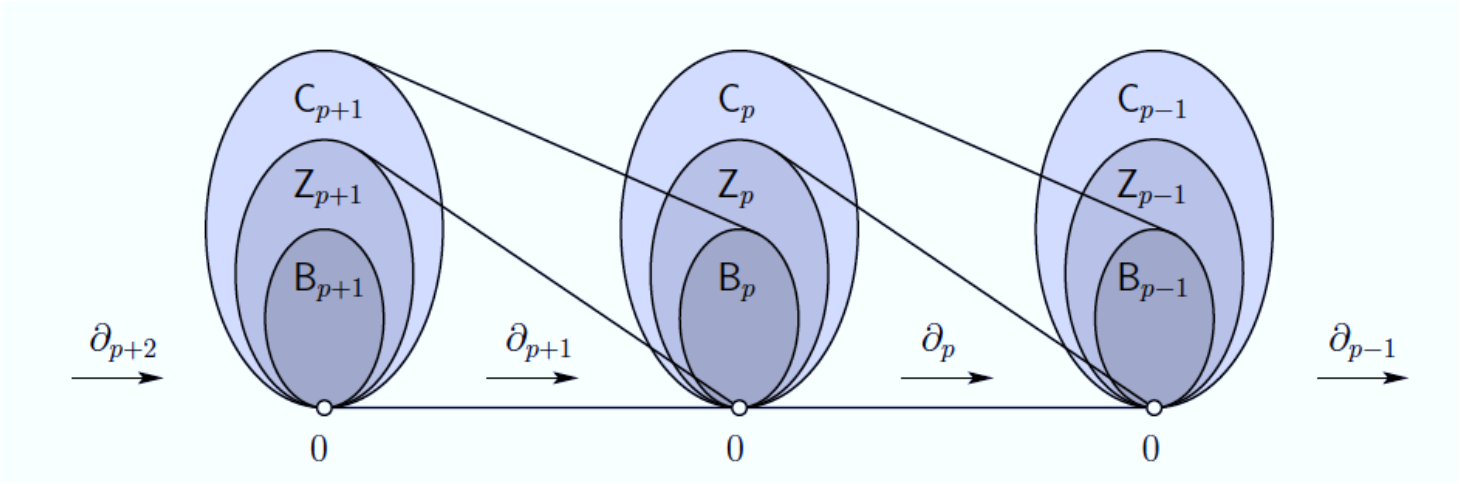
Here the boundary map ∂_p of a p-simplex σ is the sum of its (p-1)-dimensional faces:

$$\partial_p \sigma = \sum_{j=0}^{p} [u_0, \ldots, \hat{u}_j, \ldots, u_p]$$

The group of **cycles**: $Z_p = \ker \partial_p$

The group of **boundary** is $B_p = \mathrm{im}\, \partial_{p+1}$

Lemma: $\partial_p \, \partial_{p+1} = 0$



By Lemma, $B_p \subseteq Z_p$

The ***p*-th homology group** of K is the quotient

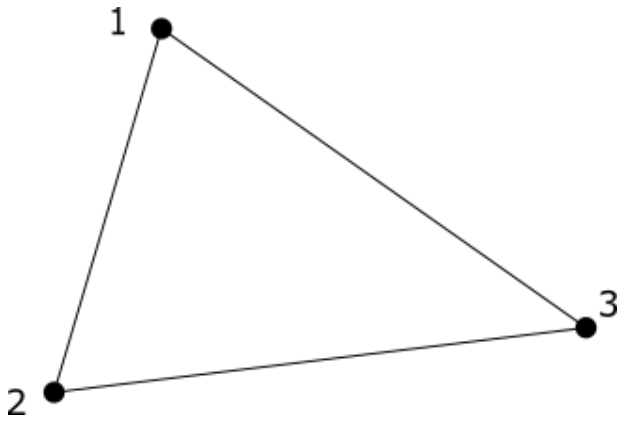
$$H_p = Z_p / B_p$$

The p-th **Betti number** of K is the rank of this group(vector space)

$$\beta_p = \operatorname{rank} H_p$$

Example:

$$\cdots \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$



$C_0 = \mathbb{Z}\langle\{1\},\{2\},\{3\}\rangle$
$C_1 = \mathbb{Z}\langle\{1,2\},\{1,3\},\{2,3\}\rangle$
$C_2 = 0$

$H_0 \equiv \mathbb{Z}$
$H_1 = Ker\partial_1/Im\partial_2 \equiv \mathbb{Z}$

$Ker\partial_0 = \mathbb{Z}\langle\{1,2\} + \{2,3\} - \{1,3\}\rangle$
$Im\partial_2 = 0$

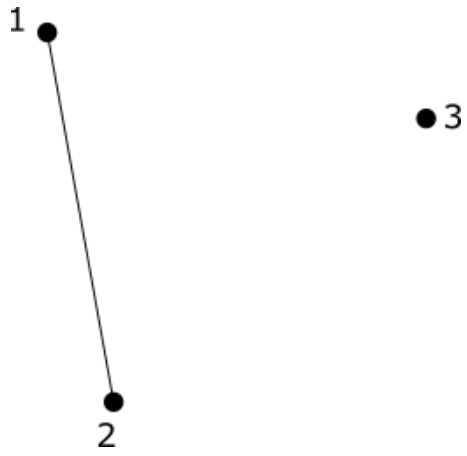
We can see that (1,2)+(2,3)−(1,3) is an example of a k-cycle that is not a kk-boundary.

It is a k-cycle because

$$\partial 1(\{1,2\} + \{2,3\} - \{1,3\}) = \{2\} - \{1\} + \{3\} - \{2\} - (\{3\} - \{1\}) = 0$$

But it is not a k-boundary because there are no 2-simplexes in S.

Example:



$$\cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

$C_0 = \mathbb{Z}\langle\{1\},\{2\},\{3\}\rangle$ $\quad Ker\partial_0 = C_0$ $\quad H_0 = Ker\partial_0/Im\partial_1$

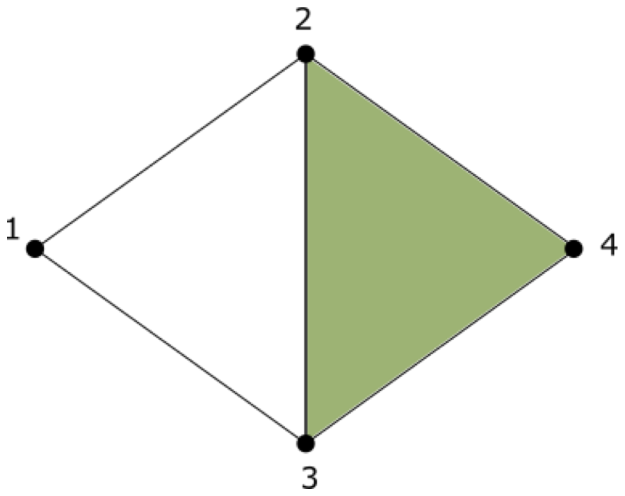
$C_1 = \mathbb{Z}\langle\{1,2\}\rangle$ $\quad Ker\partial_1 = \langle\{2\} - \{1\}\rangle$

$\{1\} = \{2\} - [\{2\} - \{1\}] = \{2\} - \partial_1(\{1,2\})$

So $\{1\} \equiv \{2\}$ in H_0

$H_0 = \mathbb{Z}\langle\{1\}\rangle \times \mathbb{Z}\langle\{2\}\rangle \equiv \mathbb{Z} \times \mathbb{Z}$

Example:

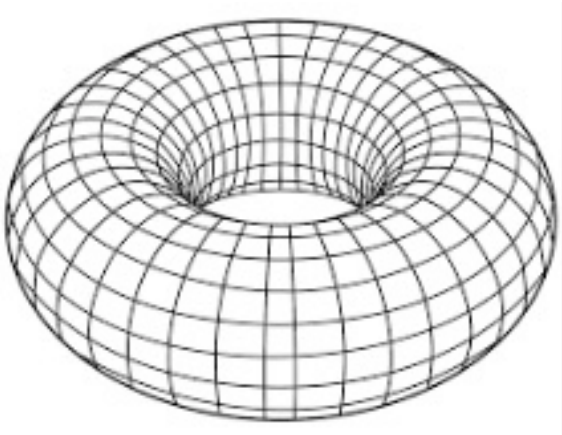


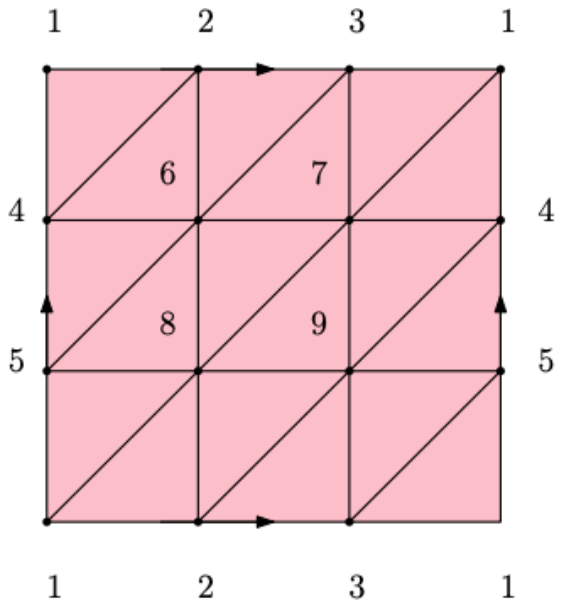
$$\cdots \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

$$C_0 = \mathbb{Z}\langle\{1\},\{2\},\{3\},\{4\}\rangle$$
$$C_1 = \mathbb{Z}\langle\{1,2\},\{1,3\},\{2,3\},\{2,3\},\{3,4\}\rangle$$
$$C_2 = \mathbb{Z}\langle\{2,3,4\}\rangle$$

$$Ker\partial_1 = \mathbb{Z}\Big\langle [\{1,2\}+\{2,3\}-\{1,3\}],\ \ [\{2,3\}+\{3,4\}-\{2,4\}] \Big\rangle$$
$$Im\partial_2 = \mathbb{Z}\langle\{3,4\}-\{2,4\}+\{2,3\}\rangle$$
$$H_1 = Ker\partial_1/Im\partial_2 \equiv \mathbb{Z}\langle\{1,2\}+\{2,3\}-\{1,3\}\rangle$$

Example: Torus





Triangulation:
$[1,2,4], [2,3,6], [3,1,7], [4,2,6],$
$[6,3,7], [7,1,4], [4,6,5], [6,7,8],$
$[7,4,9], [5,6,8], [8,7,9], [9,4,5],$
$[5,8,1], [8,9,2], [9,5,3], [1,8,2],$
$[2,9,3], [3,5,1]$
+ all faces

$$\cdots \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

$$(\partial_2)(T) = \begin{array}{c} \\ [1,2] \\ [2,3] \\ [1,3] \\ [1,4] \\ [1,5] \\ [2,4] \\ [2,6] \\ [2,8] \\ [2,9] \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} [1,2,4] & [2,3,6] & [3,1,7] & [4,2,6] & [6,3,7] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \\ 0 & 1 & 0 & 0 & 0 & \\ 0 & 0 & 1 & 0 & 0 & \\ 1 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & \\ 1 & 0 & 0 & 1 & 0 & \\ 0 & 1 & 0 & 1 & 0 & \\ 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & \\ & & & & & \end{pmatrix} \end{array}$$

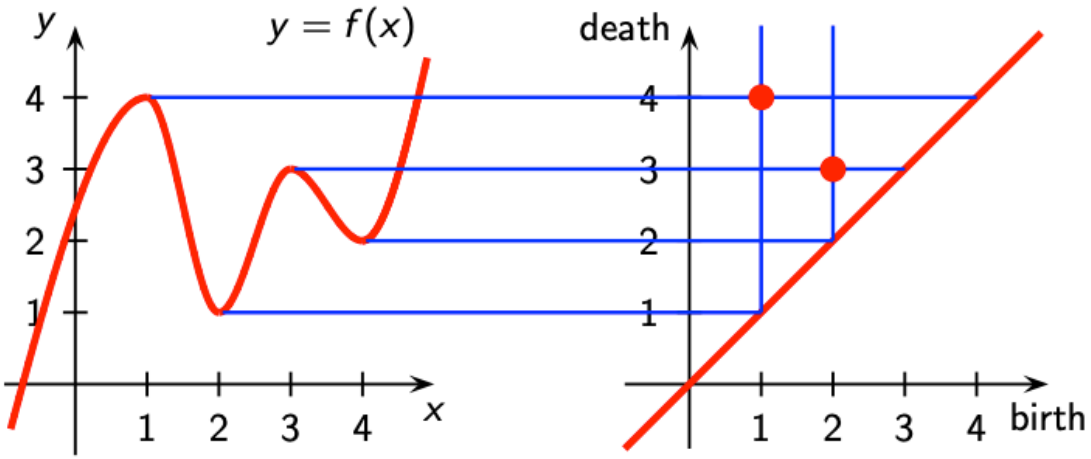
27 × 18 matrix

➢ Persistent homology of functions

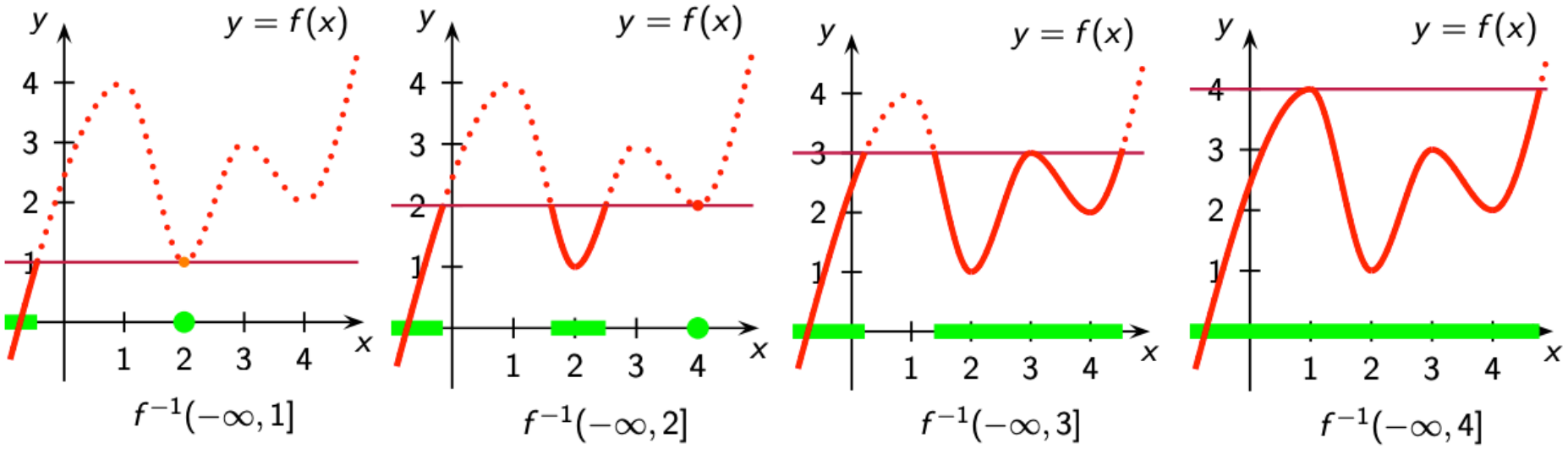
Persistent homology describes the homological features which persist as a single parameter changes.

We take this parameter to be a threshold on the values of a function.

A function $f\colon \mathbb{R} \to \mathbb{R}$ (left) and its 0-th persistence diagram (right).



Local minima create a connected component in the corresponding sublevel set, while local maxima merge connected components. The pairing of birth and death is shown in the persistence diagram.



Persistent homology for functions on manifolds

Let M be a manifold and let $f : M \to \mathbb{R}$.

This function gives an increasing filtration of M by **sublevel sets**

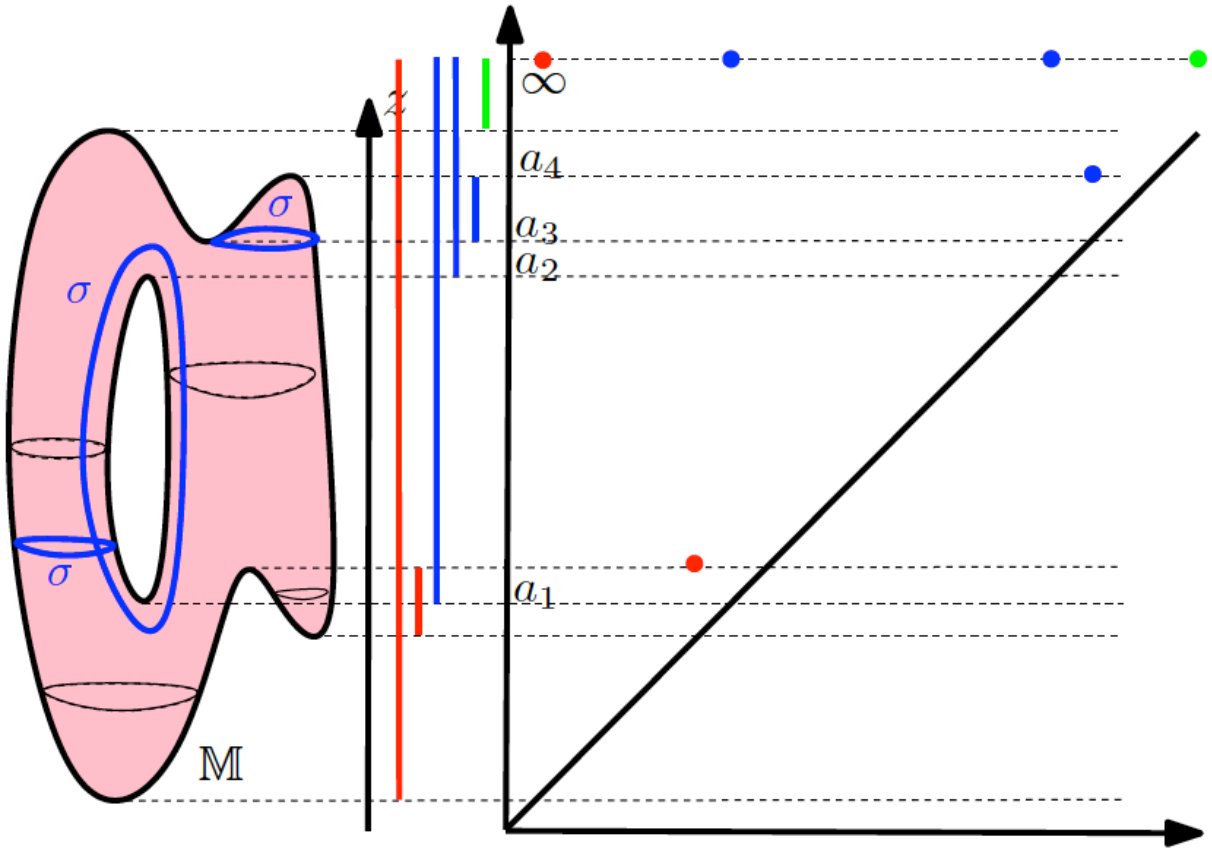
$$M_{f \leq r} = \{x \in M \mid f(x) \leq r\}$$

For $s \leq t$, the inclusion $i_s^t : \mathcal{M}_{f \leq s} \to \mathcal{M}_{f \leq t}$ induces

$$H_*(i_a^b) : H_k(\mathcal{M}_{f \leq s}) \to H_k(\mathcal{M}_{f \leq t}),$$

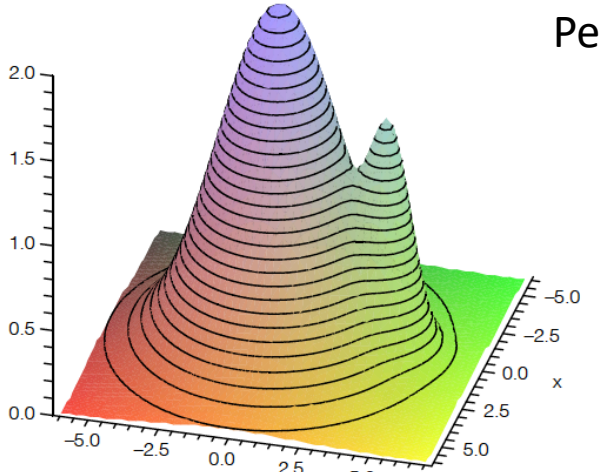
whose image is the persistent homology from s to t of f.

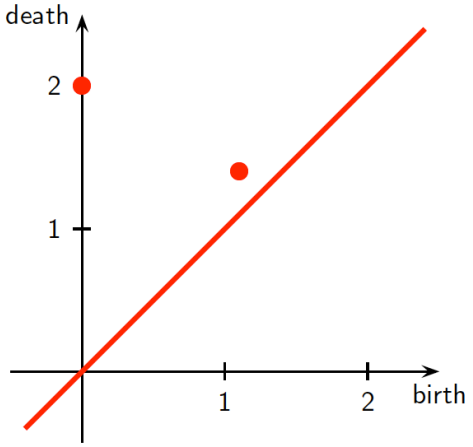
By Morse theory, the sublevel sets $M_{f \leq t}$ only changes (up to homotopy) if t is a critical point.

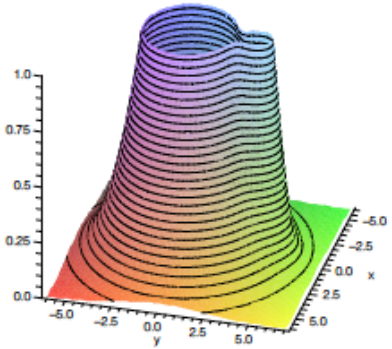


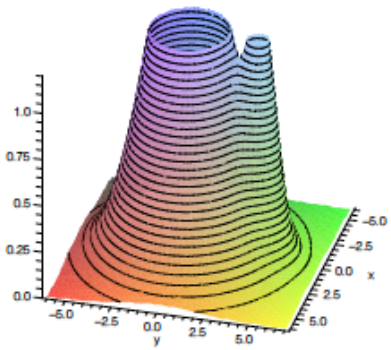
Tracking and encoding the evolution of the 0-dimensional homology, 1-dimensional homology and 2-dimensional homology of the sublevel sets.

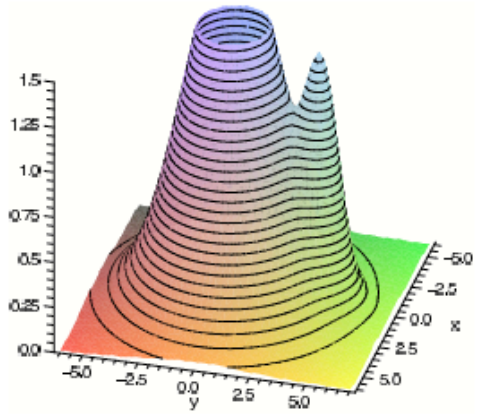
Persistence Diagram

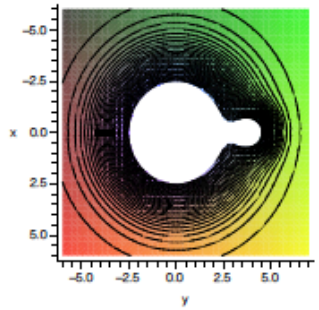


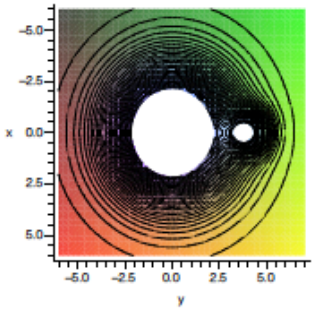


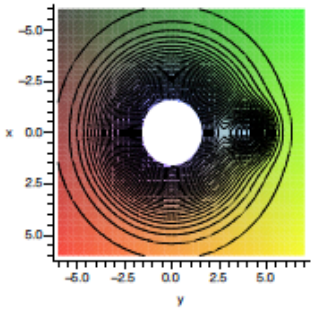




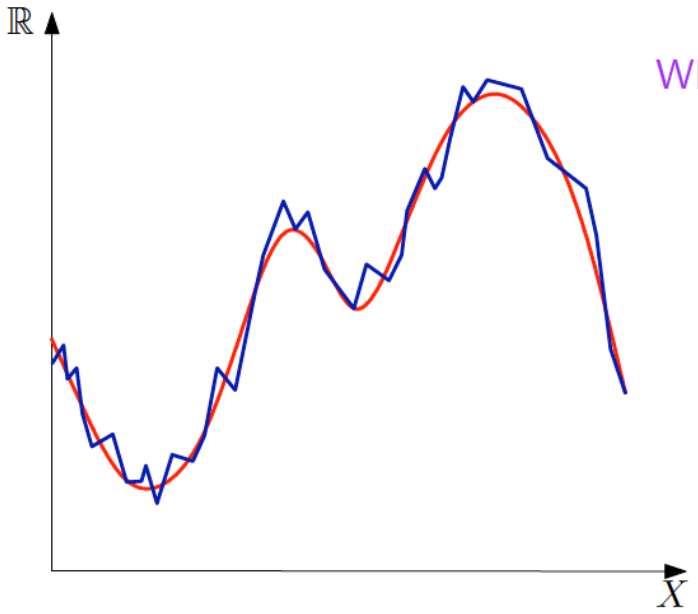




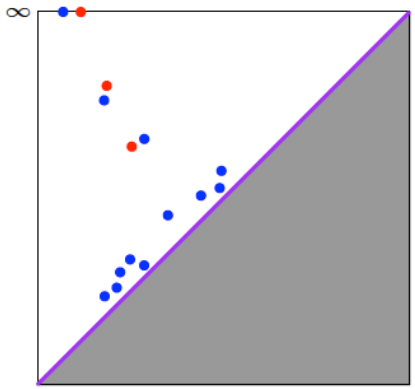




➢ Stability properties



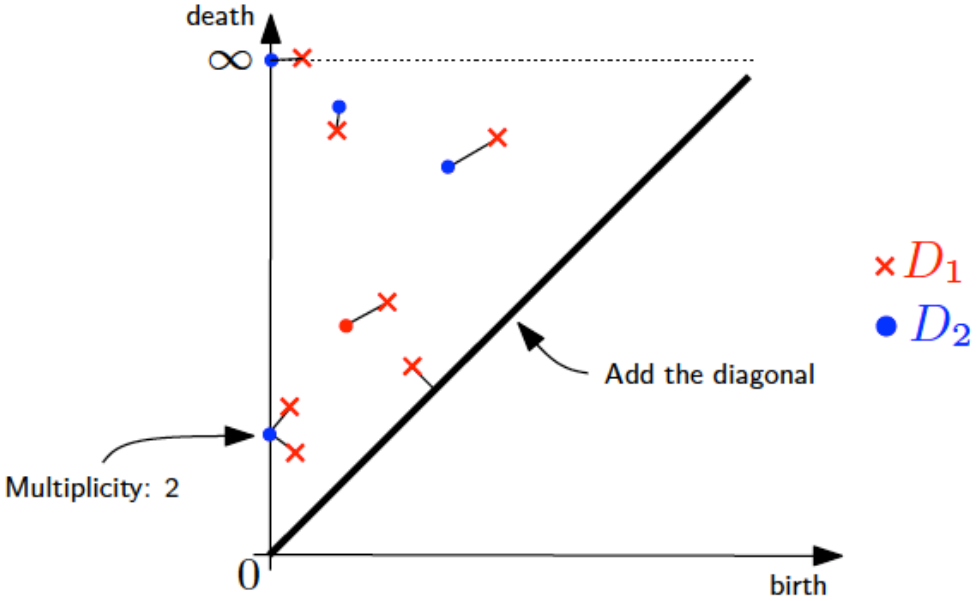
What if f is slightly perturbed?



Theorem (Stability):
For any *tame* functions $f, g : \mathbb{X} \to \mathbb{R}$, $\mathrm{d}_{\mathrm{B}}^{\infty}(\mathrm{D}_f, \mathrm{D}_g) \leq \|f - g\|_{\infty}$.

[Cohen-Steiner, Edelsbrunner, Harer 05], [C., Cohen-Steiner, Glisse, Guibas, Oudot - SoCG09], [C., de Silva, Glisse, Oudot 12]

➢ Distance between persistence diagrams



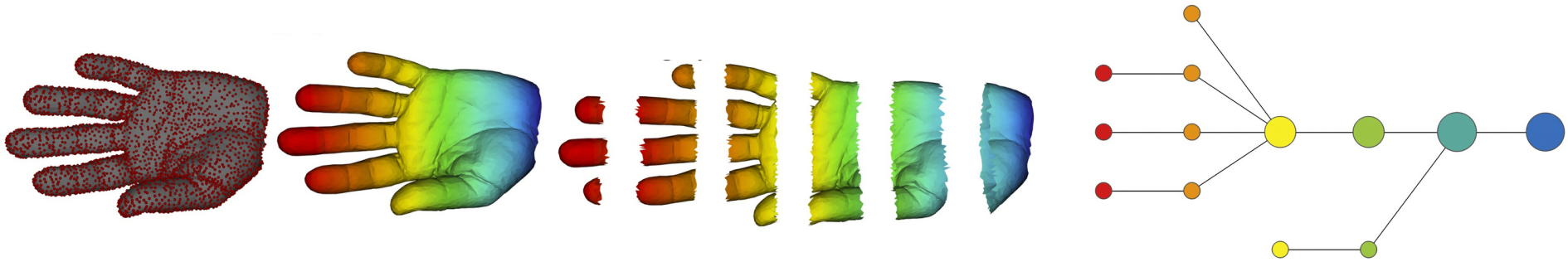
The bottleneck distance between two diagrams D_1 and D_2 is

$$d_B(D_1, D_2) = \inf_{\gamma \in \Gamma} \sup_{p \in D_1} \|p - \gamma(p)\|_\infty$$

where Γ is the set of all the bijections between D_1 and D_2 and $\|p - q\|_\infty = \max(|x_p - x_q|, |y_p - y_q|)$.

➢ Data Visualization with TDA Mapper

Data → coloring → Overlapping bins → Graph



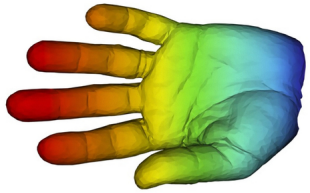
Extracting insights from the shape of complex data using topology
P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson & G. Carlsson
http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

https://www.ayasdi.com/

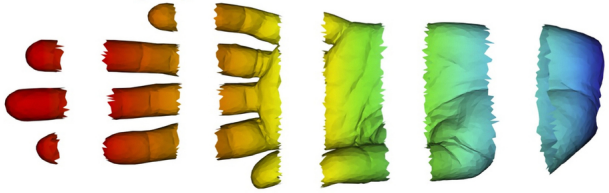
https://research.math.osu.edu/tgda/mapperPBG.pdf

A Original Point Cloud

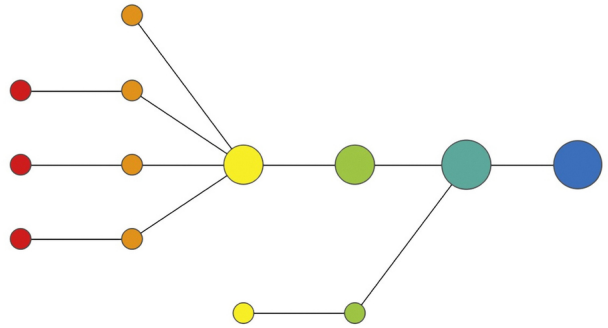
B Coloring by filter value



C Binning by filter value



D Clustering and network construction



A) Data Set

Example: Point cloud data representing a hand.

B) Function f : Data Set $\rightarrow \mathbb{R}$

Example: x-coordinate

$f : (x, y, z) \rightarrow x$

C) Put data into overlapping bins.

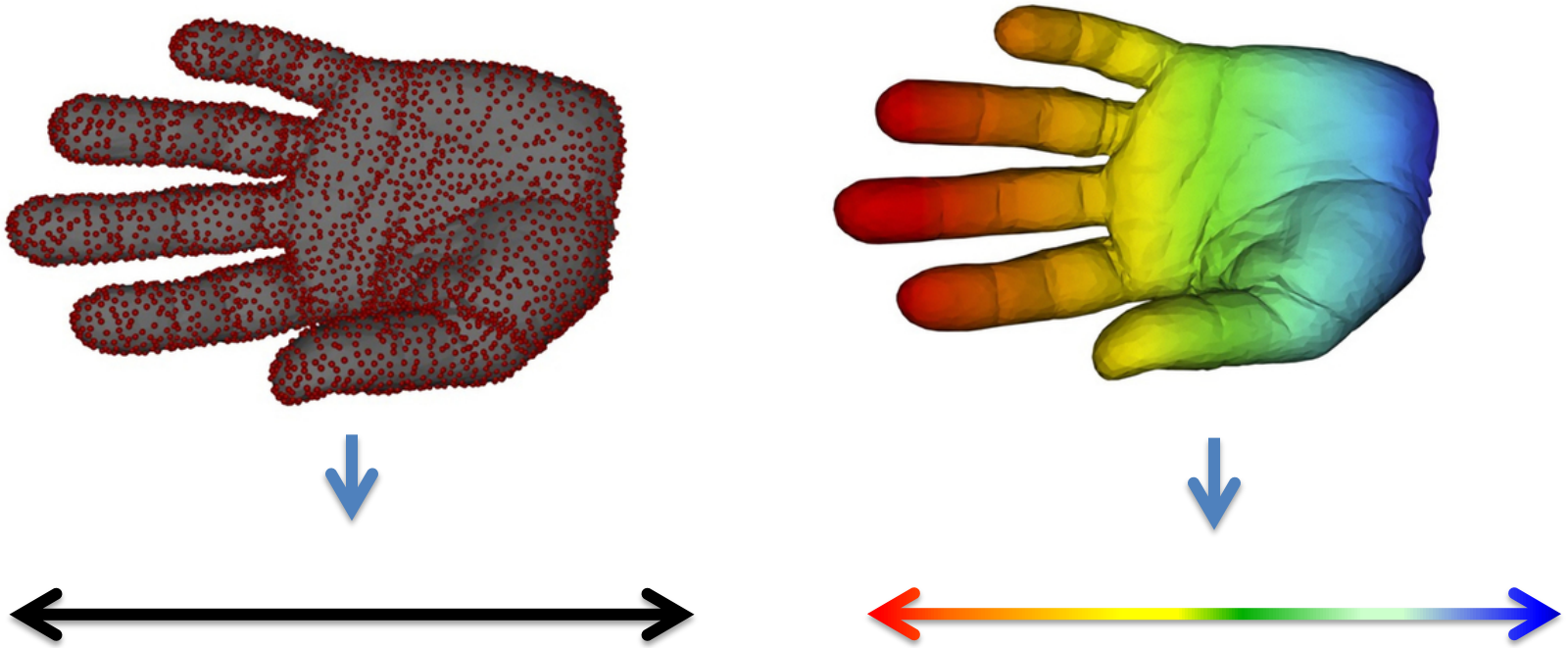
Example: $f^{-1}(a_i, b_i)$

D) Cluster each bin & create network.

Vertex = a cluster of a bin.

Edge = nonempty intersection between clusters

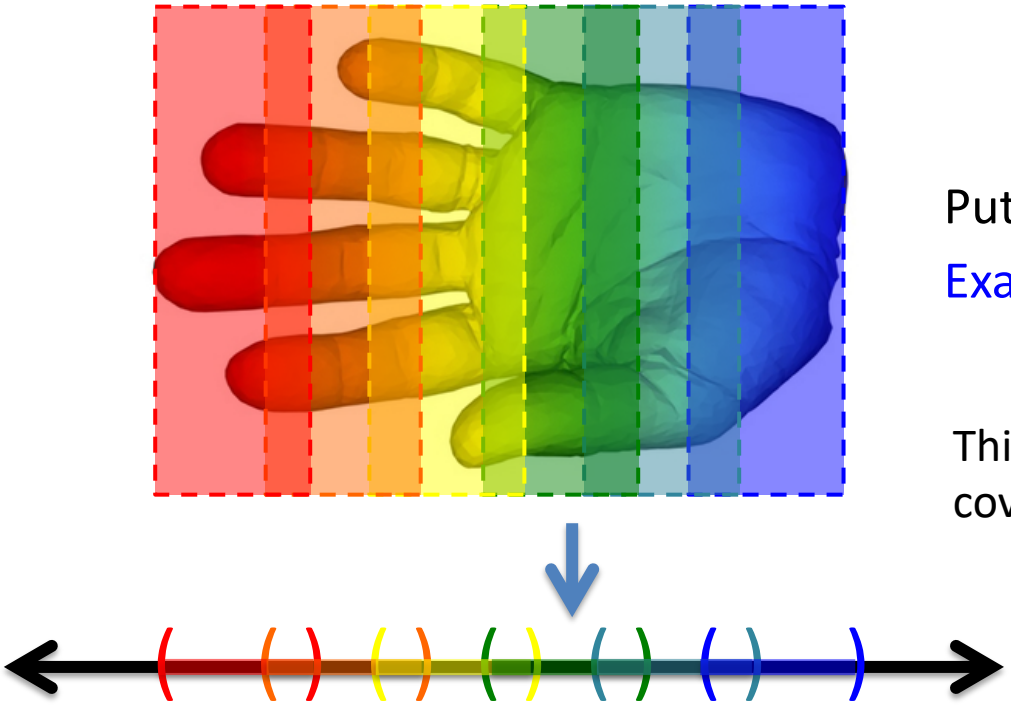
B. Coloring by filter value



Function f : Data Set → $\mathbb{R}$

Ex 1: x-coordinate

f : (x, y, z) → x



Put data into overlapping bins.

Example: $f^{-1}(a_i, b_i)$

This comes from the finite covering in topology.

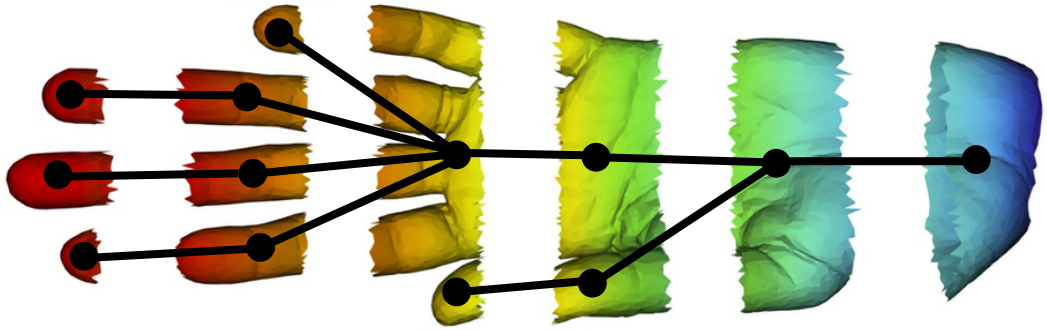
Function f : Data Set → $\mathbb{R}$

Ex 1: x-coordinate

f : (x, y, z) → x

Function f can also be the probability density function for a Gaussian distribution.

C Binning by filter value

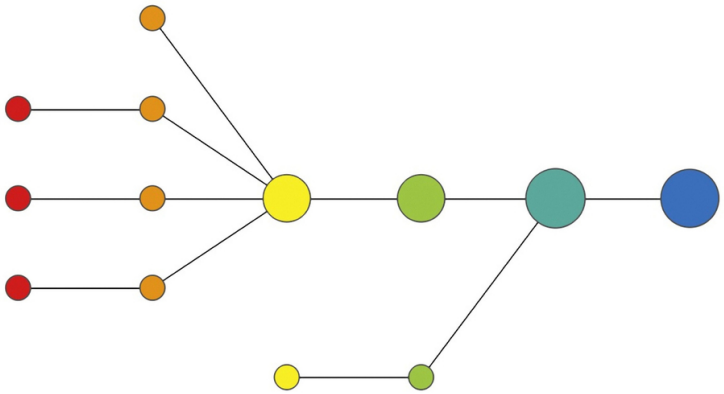


D) Cluster each bin

& create network.

Vertex = a cluster of a bin.

Edge = nonempty intersection between clusters

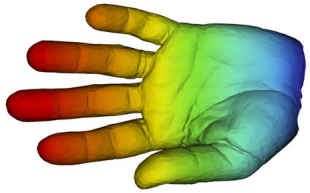


The **color** of a node indicates the value of the function f (red being high and blue being low)

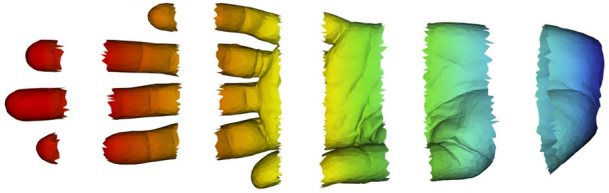
The **size** of a node indicates the number of points in the set represented by the node.

A Original Point Cloud

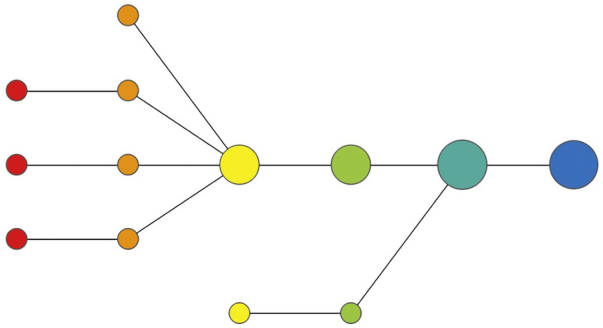
B Coloring by filter value



C Binning by filter value



D Clustering and network construction



A) Data Set

Example: Point cloud data representing a hand.

B) Function f : Data Set → $\mathbb{R}$

Example: x-coordinate

$f : (x, y, z) \rightarrow x$

C) Put data into overlapping bins.

Example: $f^{-1}(a_i, b_i)$

D) Cluster each bin & create network.

Vertex = a cluster of a bin.

Edge = nonempty intersection between clusters

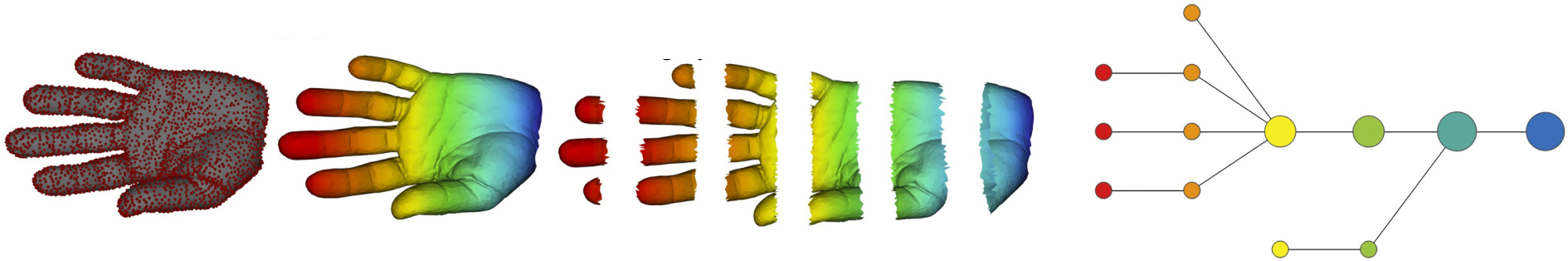
In the next few examples, please note

1. The different types of data to which we can apply TDA mapper.

2. Several choices need to be made when applying TDA mapper. For example:

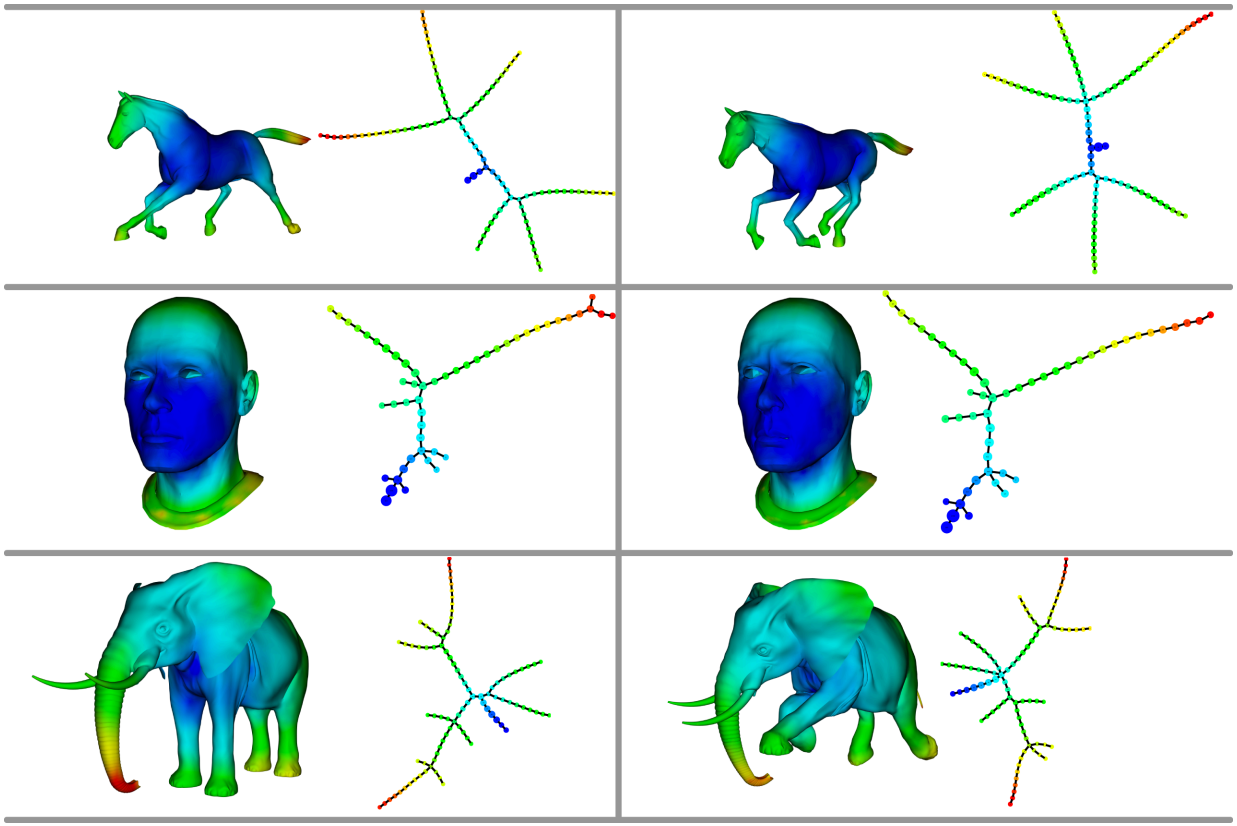
   - How is the data modeled including how is the distance between data points calculated?

   - How are the data put into overlapping bins?



Data → coloring → Overlapping bins → Graph

Mapper on 3D Shape Database



Each row of this image shows two poses of the same shape along with the Mapper result. For each Mapper computation, they used 15 intervals in the range of the filter with a 50% overlap. Euclidean distance is used between points.

Gurjeet Singh , Facundo Mémoli and Gunnar Carlsson, Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. Eurographics Symposium on Point-Based Graphics (2007)

Three key ideas of topology that make extracting of patterns via shape possible.

1.) coordinate free.

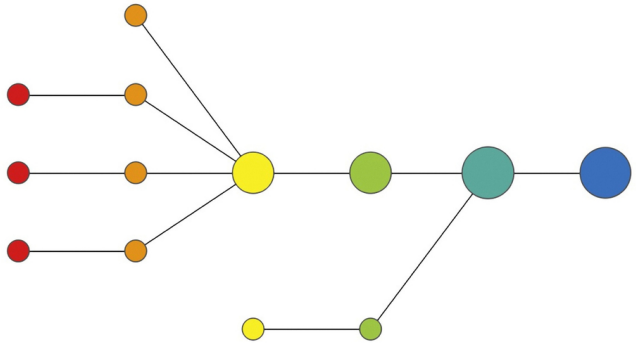
- No dependence on the coordinate system chosen.
- Can compare data derived from different platforms

2.) invariant under "small" deformations.

- less sensitive to noise

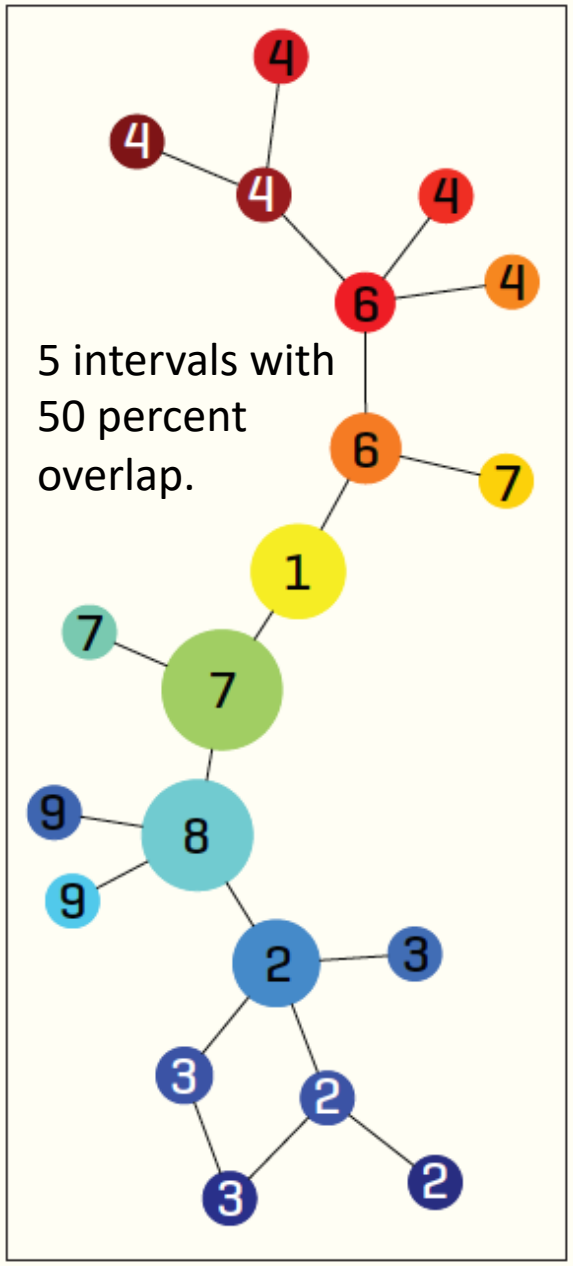
3.) compressed representations of shapes.

- Input: dataset with thousands of points
- Output: network with 13 vertices and 12 edges.



http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Handwritten digits example: https://dl.acm.org/doi/pdf/10.1145/2627814



1,797 data points

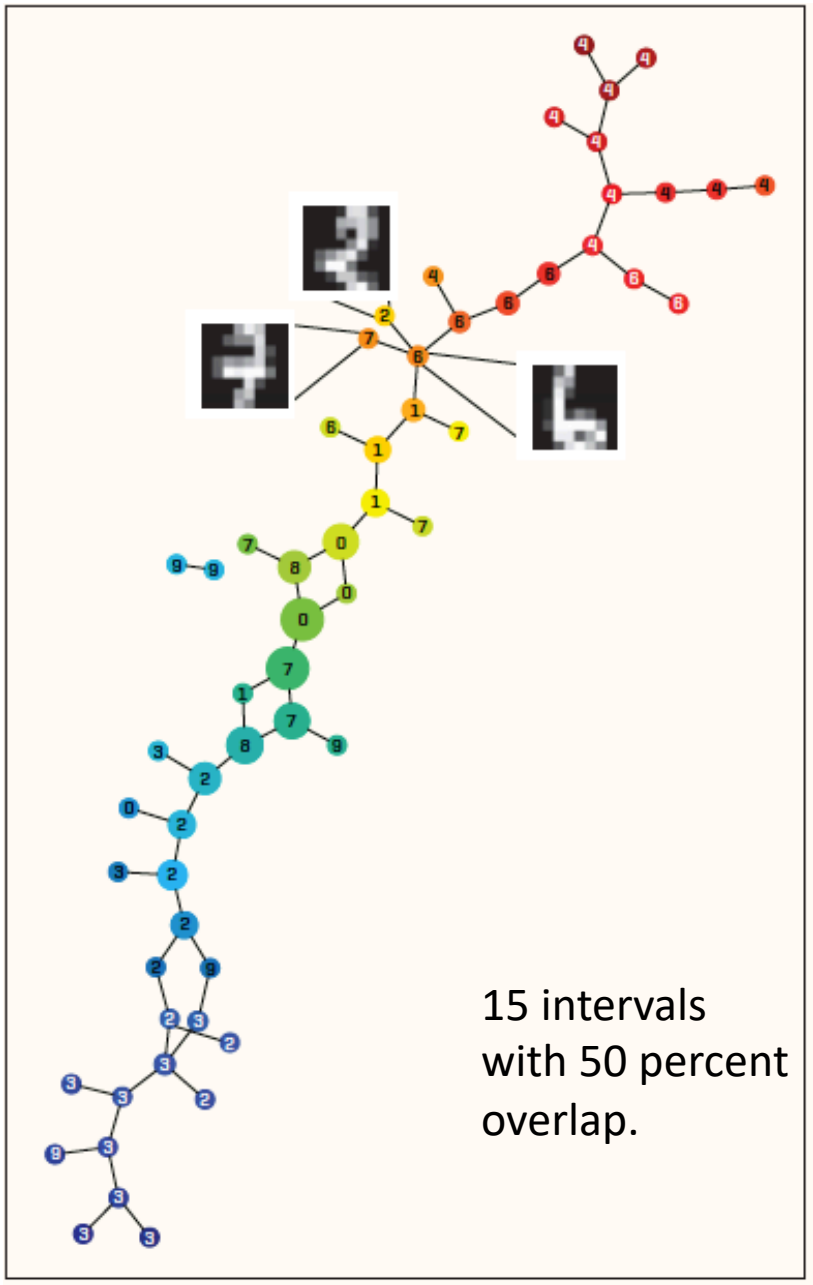
data point:
8x8 matrix

Distance metric:
Euclidean

Filter function:
principal SVD values

Node colors:
filter values,
red = high and
blue = low

Nodes labels:
most frequently occurring digit in the associated clusters



Applications from paper

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Application: Basketball

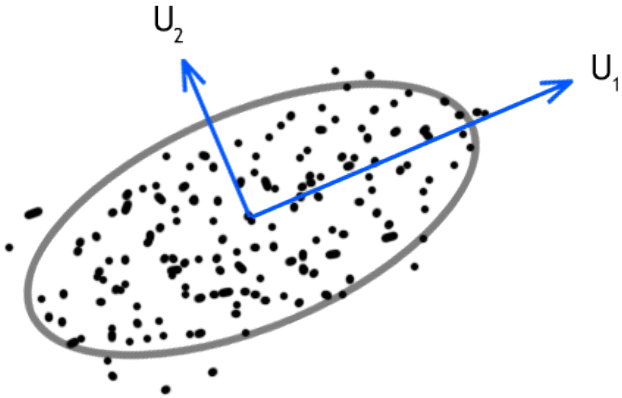
Data: rates (per minute played) of rebounds, assists, turnovers, steals, blocked shots, personal fouls, and points scored for 452 players.

➔ Input: 452 points in $\mathbb{R}^7$

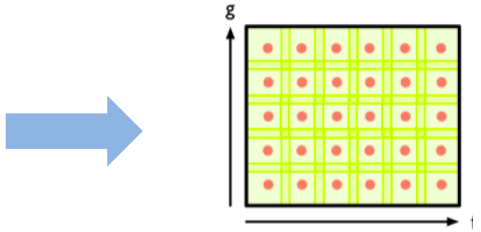
For each player, we have a vector
(rebounds/min, assists /min, turnovers /min, steals /min, blocked shots /min, personal fouls /min, points scored /min)= (r, a, t, s, b, f, p) in $\mathbb{R}^7$

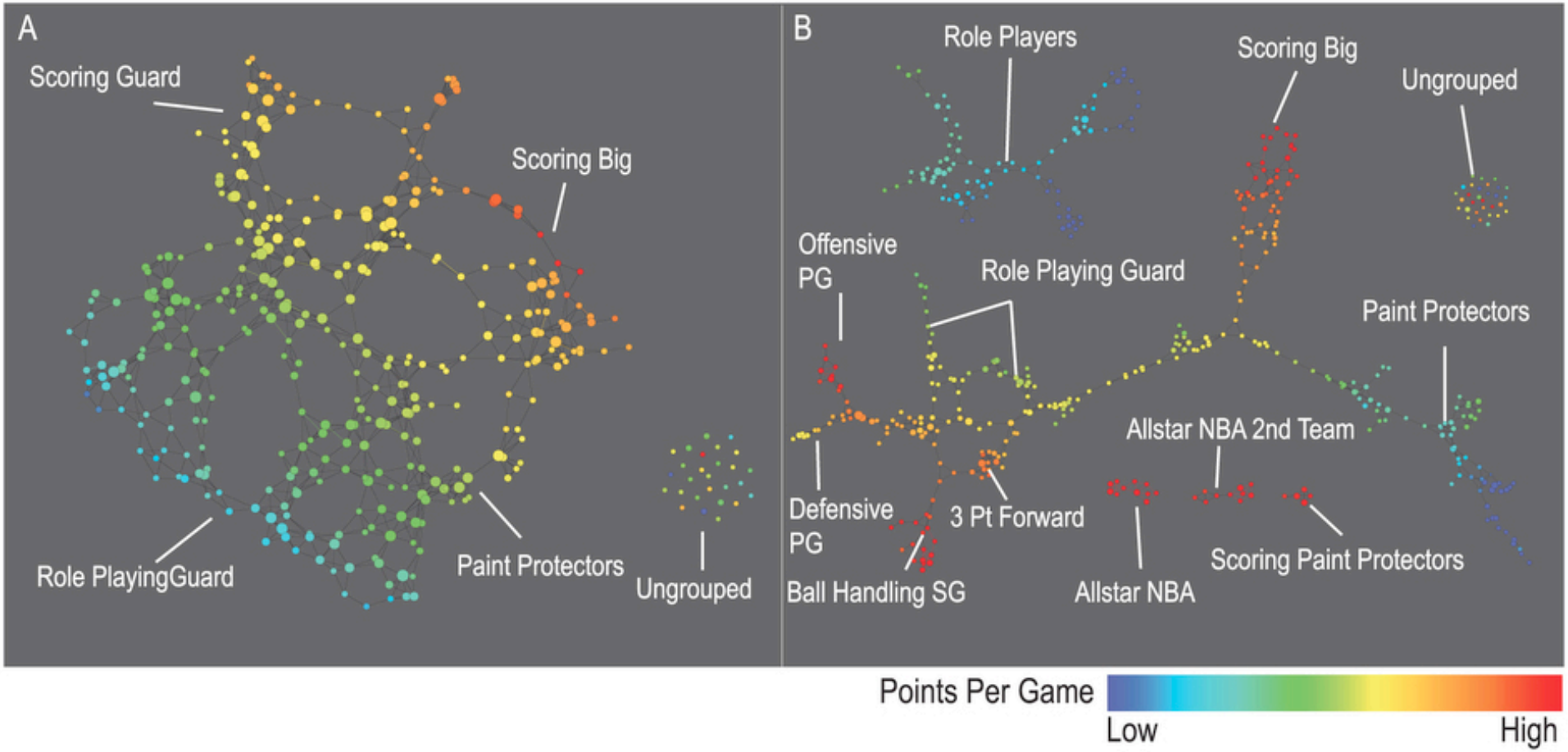
Distance: variance normalized Euclidean distance.
Clustering: Single linkage.

Filters: principle and secondary SVD values.



Data





A) Low resolution map at 20 intervals for each filter B) High resolution map at 30 intervals for each filter. The overlap is such at that each interval overlaps with half of the adjacent intervals, the graphs are colored by points per game, and a variance normalized Euclidean distance metric is applied. Metric: Variance Normalized Euclidean; Lens: Principal SVD Value (Resolution 20, Gain 2.0x, Equalized) and Secondary SVD Value (Resolution 20, Gain 2.0x, Equalized). Color: red: high values, blue: low values.

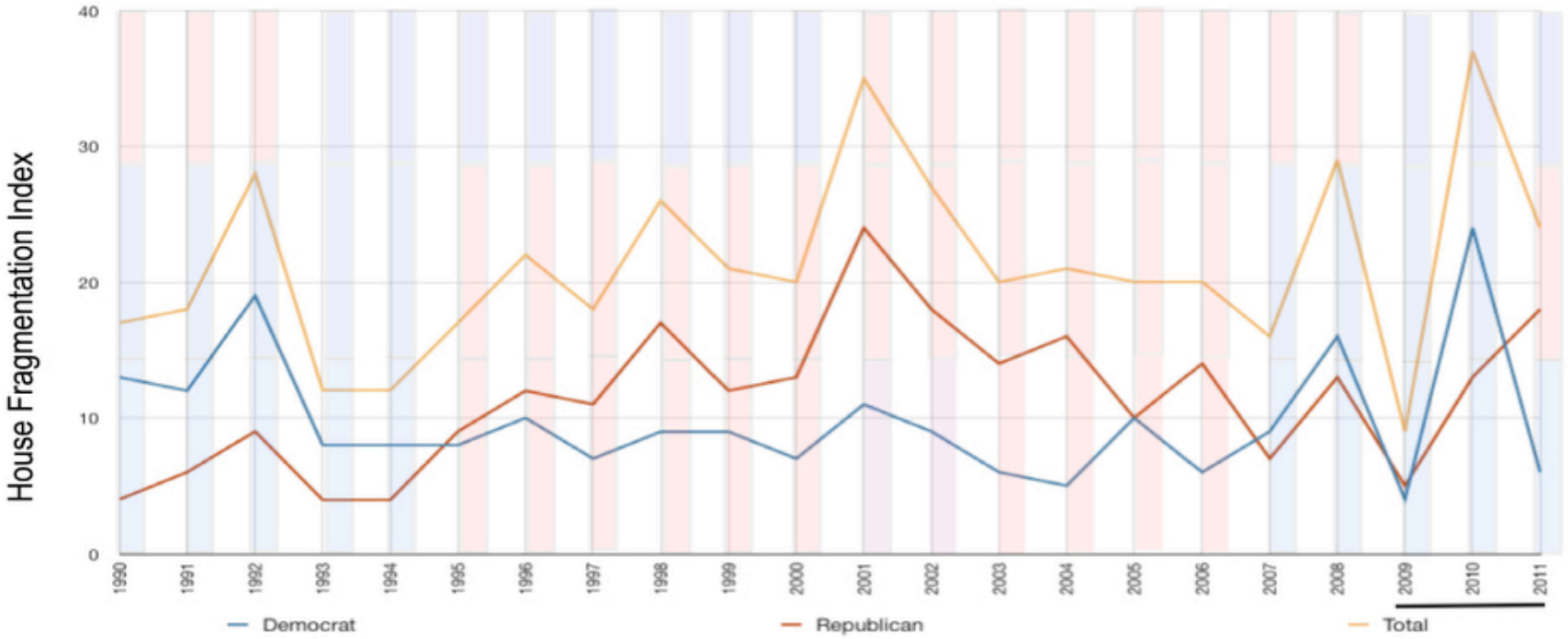
Application 2: US House of Representatives Voting records

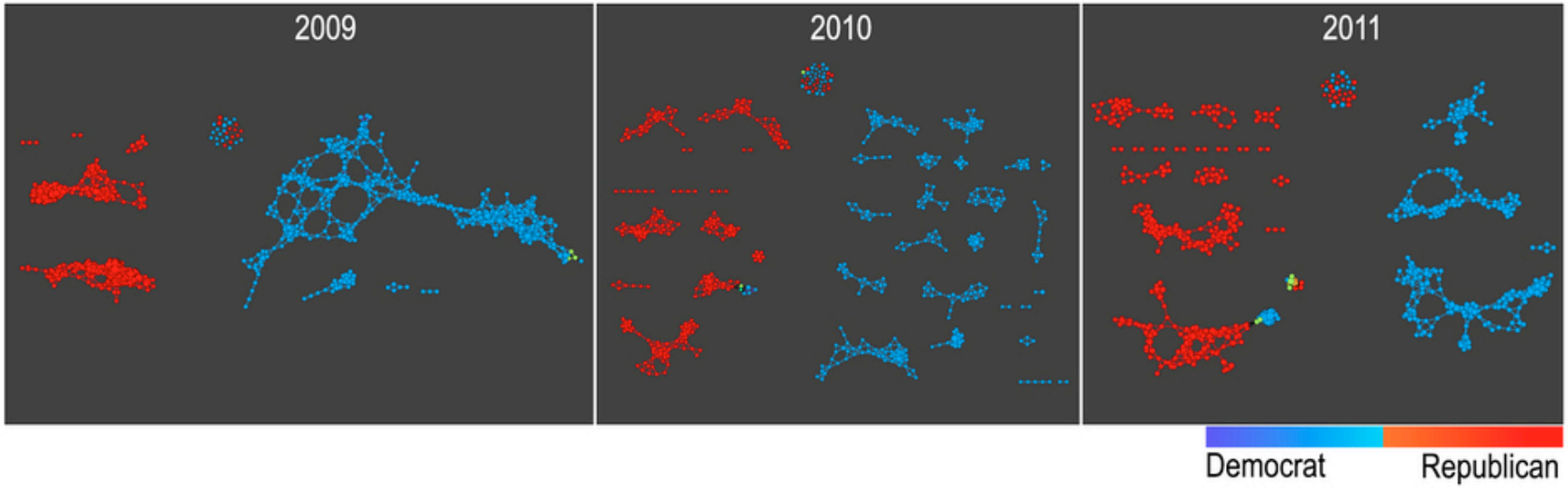
Data: (aye, abstain, nay, ….)= (+1 , 0 , -1 , …)

Distance: Pearson correlation

Filters: principal and secondary metric SVD

Clustering: Single linkage.





X-axis: 1990–2011. Y-axis: Fragmentation index. Color bars denote, from top to bottom, party of the President, party for the House, party for the Senate (red: republican; blue: democrat; purple: split). The bottom 3 panels are the actual topological networks for the members. Networks are constructed from voting behavior of the member of the house, with an “aye” vote coded as a 1, “abstain” as zero, and “nay” as a -1. Each node contains sets of members. Each panel labeled with the year contains networks constructed from all the members for all the votes of that year. Note high fragmentation in 2010 in both middle panel and in the Fragmentation Index plot (black bar). The distance metric and filters used in the analysis were Pearson correlation and principal and secondary metric SVD. Metric: Correlation; Lens: Principal SVD Value (Resolution 120, Gain 4.5x, Equalized) and Secondary SVD Value (Resolution 120, Gain 4.5x, Equalized). Color: Red: Republican; Blue: Democrats.

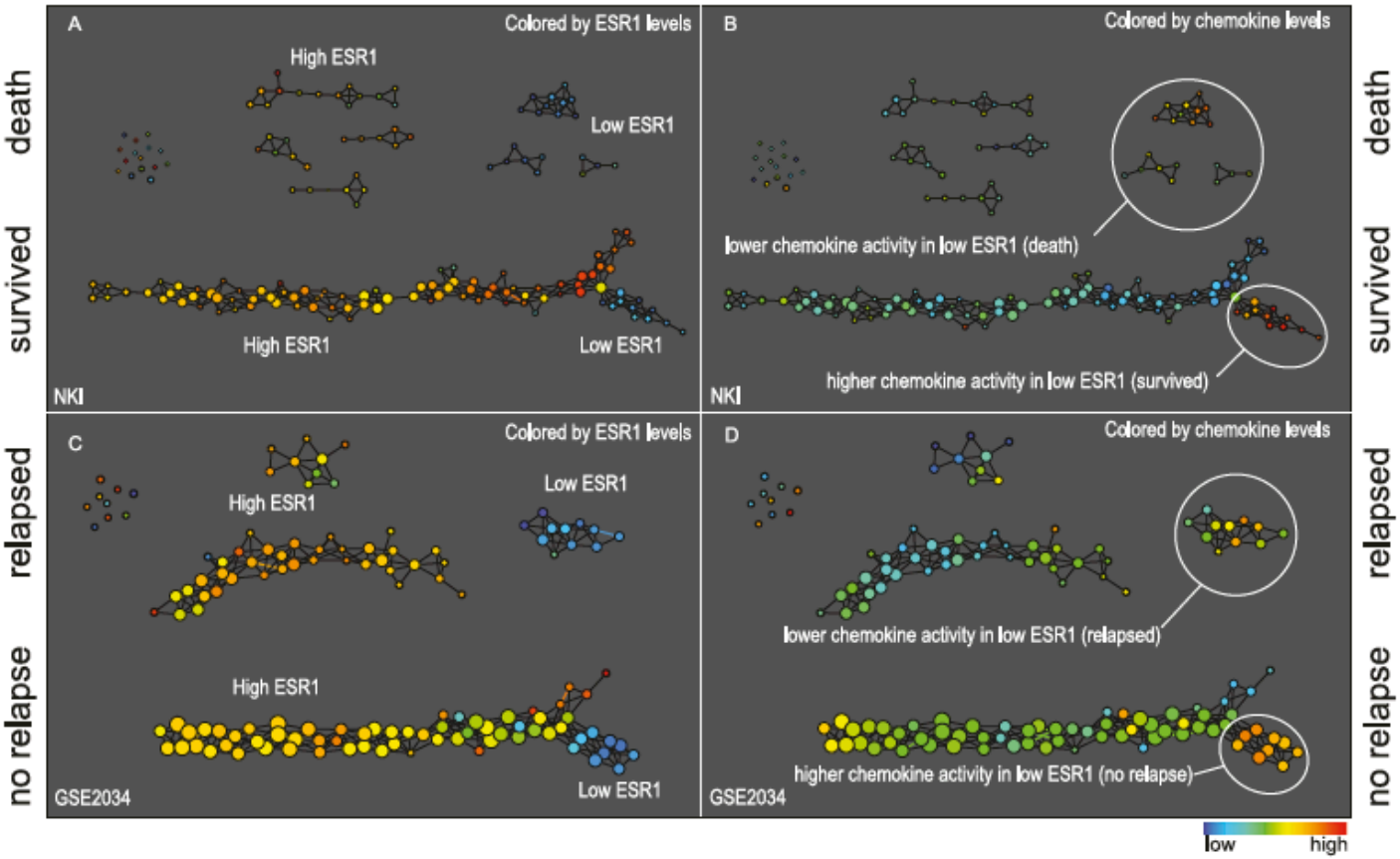
Application : breast cancer gene expression

Data: microarray gene expression data from 2 data sets, NKI and GSE2034

Distance: correlation distance

Filters: (1) L-infinity centrality: f(x) = max{d(x, p) : p in data set} captures the structure of the points far removed from the center or norm.

(2) NKI: survival vs. death. GSE2034: no relapse vs. relapse



Compare to

Gene expression profiling predicts clinical outcome of breast cancer

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH

Nature. 2002 Jan 31; 415(6871):530-6.

Breast cancer data sets:

1.) NKI (2002):
gene expression levels of 24,000 from 272 tumors. Includes node-negative and node-positive patients, who had or had not received adjuvant systemic therapy. Also includes survival information.

2.) GSE203414 (2005)
expression of 22,000 transcripts from total RNA of frozen tumour samples from 286 lymph-node-negative patients who had not received adjuvant systemic treatment. Also includes time to relapse information.

http://bioinformatics.nki.nl/data.php

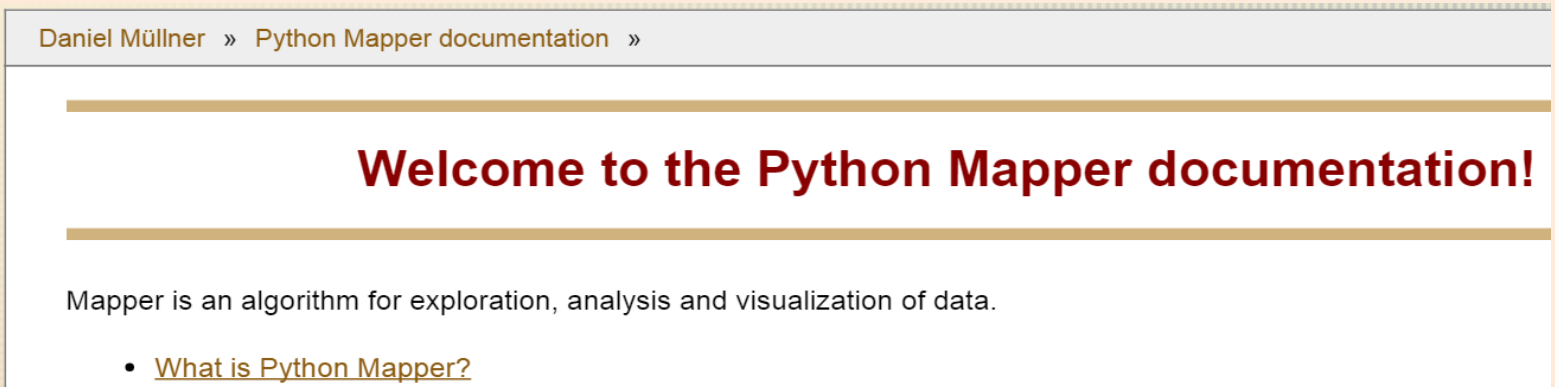
https://cran.r-project.org/web/packages/TDAmapper/

TDAmapper: Analyze High-Dimensional Data Using Discrete Morse Theory

Topological Data Analysis using Mapper (discrete Morse theory). Generate a 1-dimensional simpli within within each level set and generate one node (vertex) for each cluster. 3. For each pair of clus a filter function with codomain R, while the the function mapper2D uses a filter function with codo

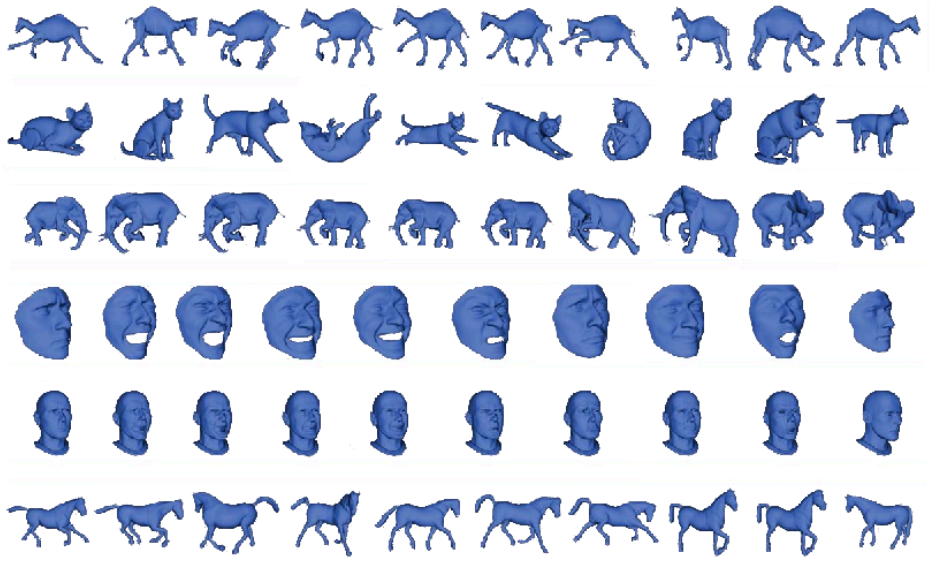
Version:	1.0
Depends:	R (≥ 3.1.2)
Suggests:	fastcluster, igraph
Published:	2015-05-31
Author:	Paul Pearson [aut, cre, trl], Daniel Muellner [aut, ctb], Gurjeet Singh [aut, ctb]
Maintainer:	Paul Pearson <pearsonp at hope.edu>

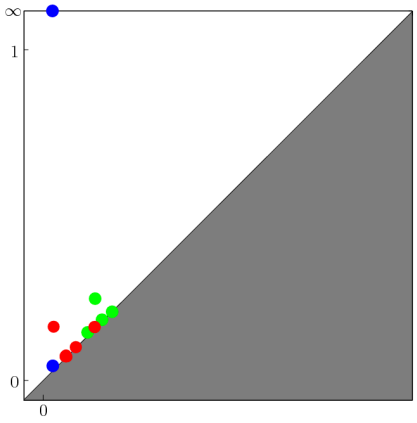
http://danifold.net/mapper/

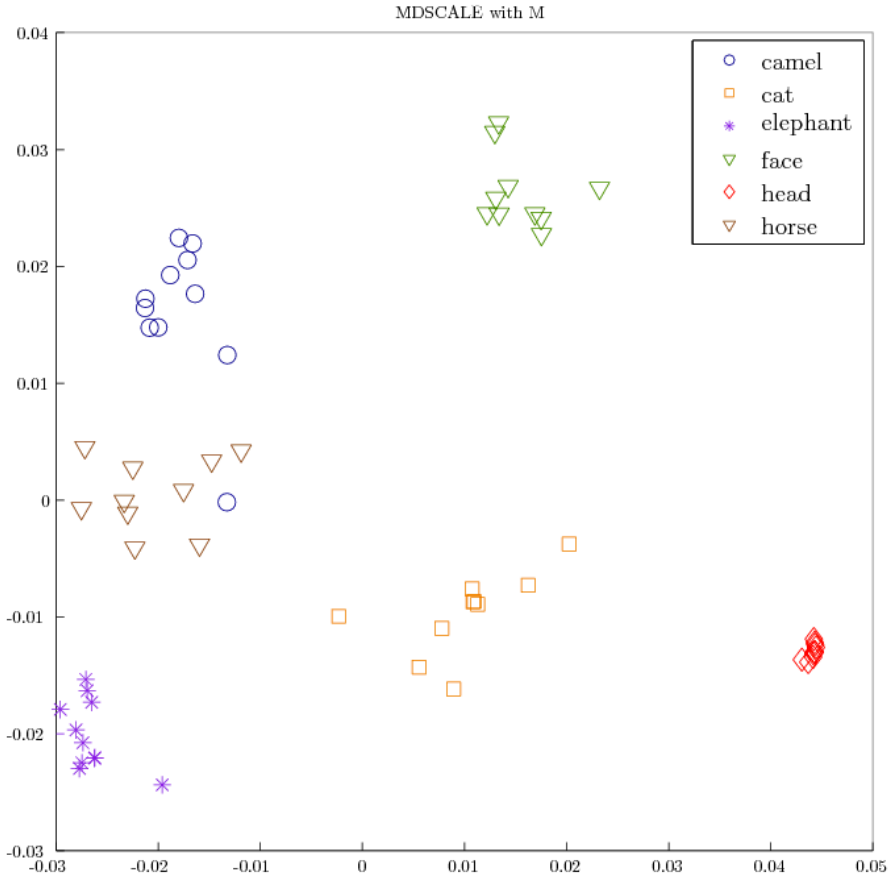


Applications to clustering, segmentations, sensor networks,...

Persistence diagrams are defined and stable for a large class of continuous functions defined over (pre-)compact metric spaces.

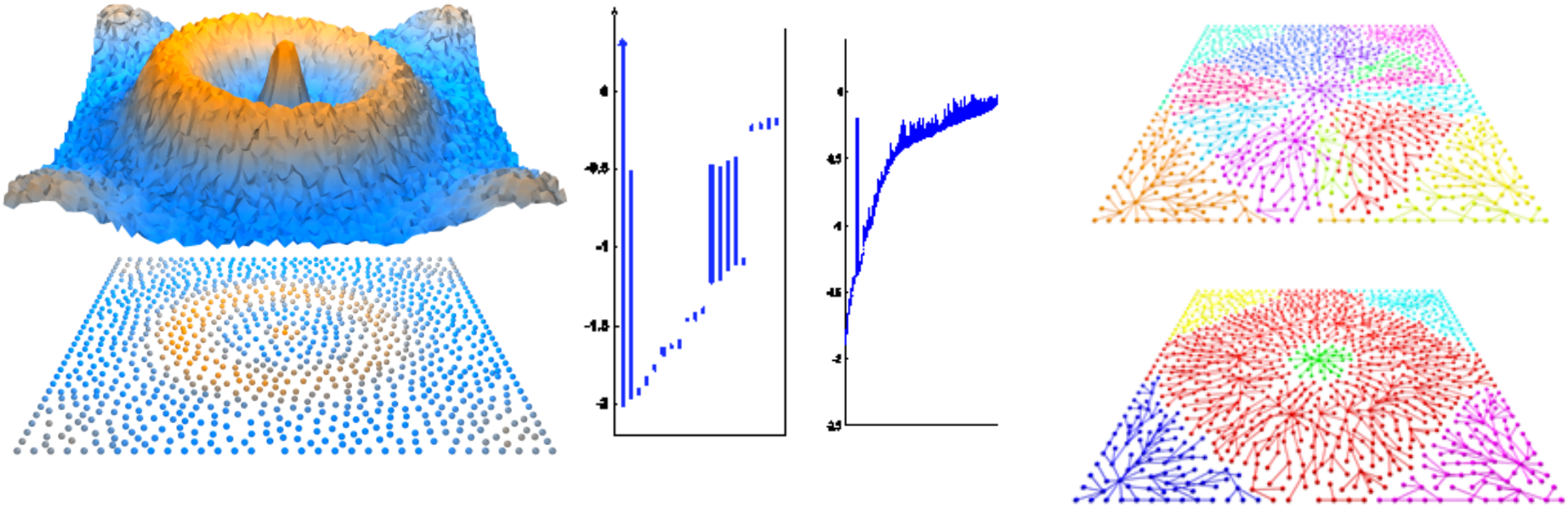






Ref: F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Memoli, S. Oudot, Gromov-Hausdor Stable Signatures for Shapes using Persistence, Computer GraphicsForum (proc. SGP 2009), pp. 1393-1403, 2009.

Persistence diagrams can be reliably estimated from data (functions known through a point cloud data set approximating a topological space).



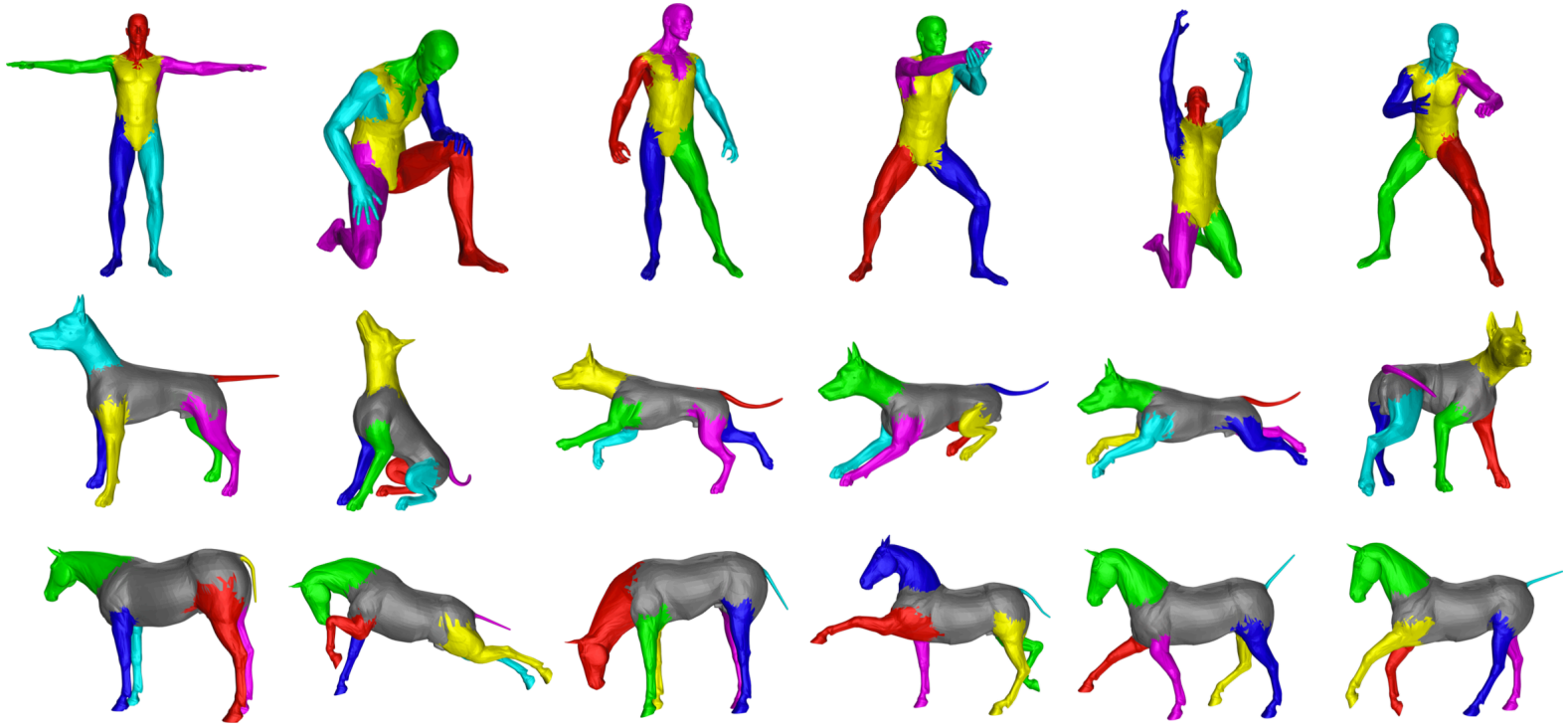
Previous approach can be generalized, leading to robust algorithms to compute the topological persistence of functions defined over point clouds sampled around unknown shapes

Ref:

F. Chazal, L. Guibas, S. Oudot, P. Skraba, Analysis of Scalar Fields over Point Cloud Data, proc. ACM Symposium on Discrete Algorithms 2009.

F. Chazal, S. Oudot, Toward Persistence-Based Reconstruction in Euclidean Spaces, proc. ACM Symposium on Computational Geometry 2008.

Applications to non rigid shapes segmentation



• P. Skraba, M. Ovsjanikov, F. Chazal, L. Guibas, Persistence-Based Segmentation of Deformable Shapes, Proc. Workshop on Nonrigid Shape Analysis and Deformable Image Alignment (NORDIA), Proc. CVPR 2010

Topological Data Analysis and Deep Learning

PLLay: Efficient Topological Layer based on Persistence Landscapes,
by Kwangho Kim, Jisu Kim, Manzil Zaheer, Joon Sik Kim, Frederic Chazal, Larry Wasserman.

https://github.com/jisuk1/pllay/

References:

https://scikit-tda.org/

Papers and Books

Persistence Theory: From Quiver Representations to Data Analysis, **Book** by Steve Oudot.

Topology and data, by Gunnar Carlsson
Topological pattern recognition for point cloud data, by Gunnar Carlsson
Persistent Homology and Applied Homotopy Theory, by Gunnar Carlsson

Courses

http://homepage.divms.uiowa.edu/~idarcy/COURSES/TDA/SPRING18/3900.html

http://graphics.stanford.edu/courses/cs233-21-spring/

https://geometrica.saclay.inria.fr/team/Fred.Chazal/

https://people.clas.ufl.edu/peterbubenik/intro-to-tda/

https://www.joperea.com/teaching/spring2020

https://yao-lab.github.io/2019_csic5011/