An Introduction to Applied Topology

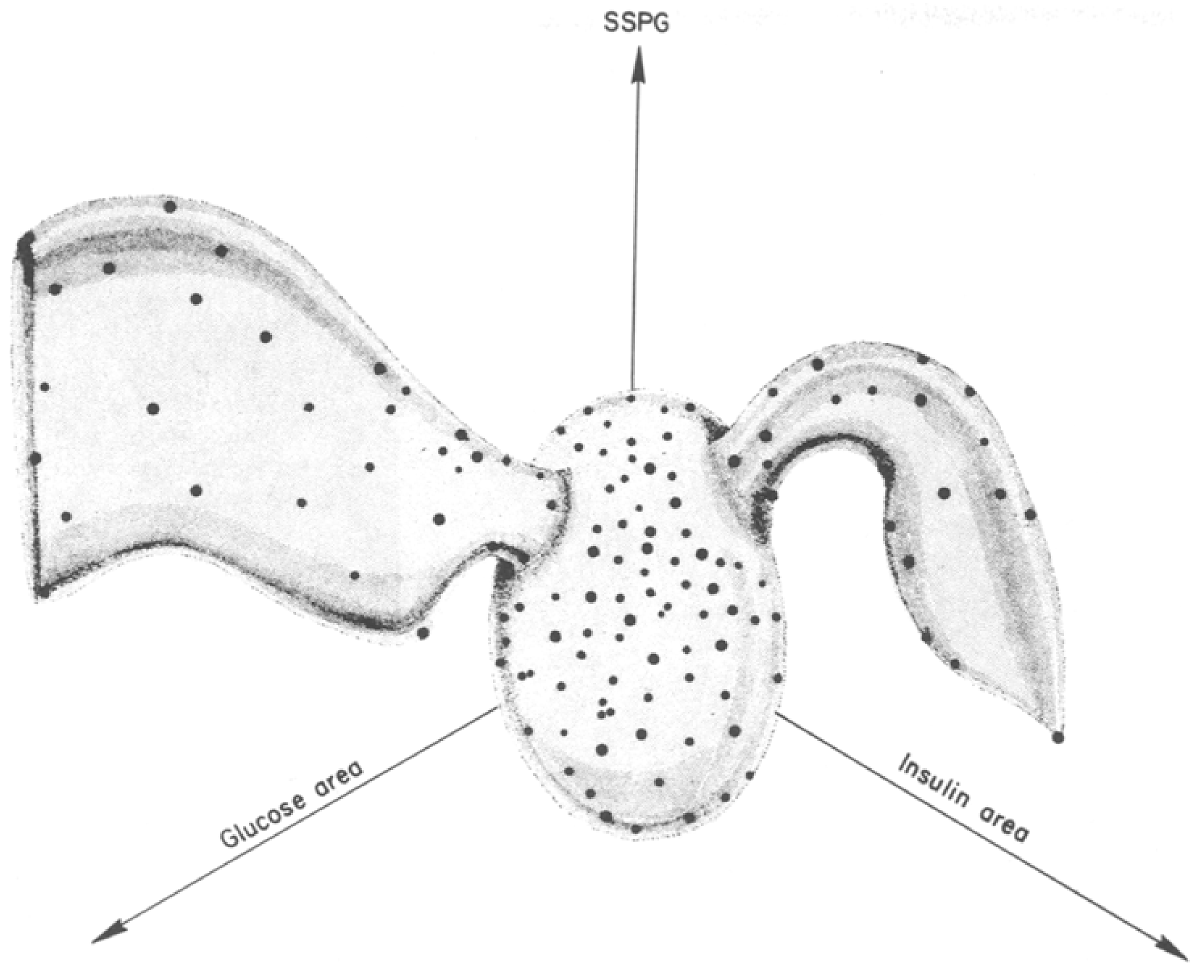
Henry Adams

University of Florida

Datasets have shapes

Example: Diabetes study

145 points in 5-dimensional space

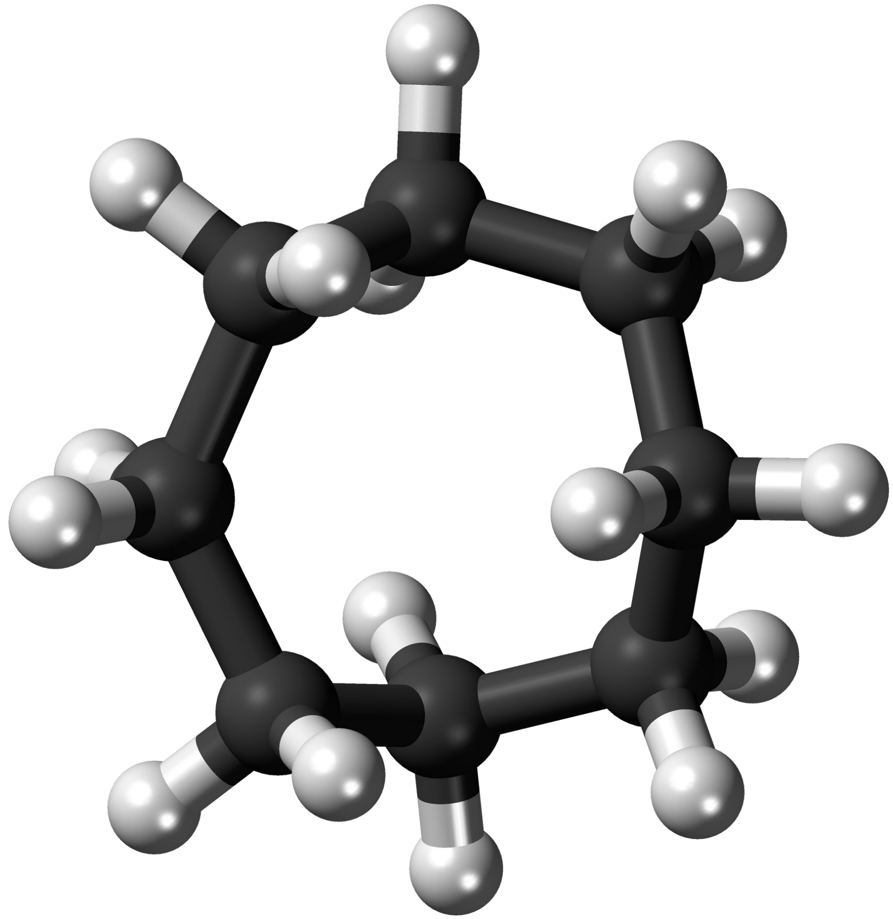


An attempt to define the nature of chemical diabetes using a multidimensional analysis by G. M. Reaven and R. G. Miller, 1979

Datasets have shapes

Example: Cyclo-Octane (C_8H_{16}) data

1,000,000+ points in 24-dimensional space

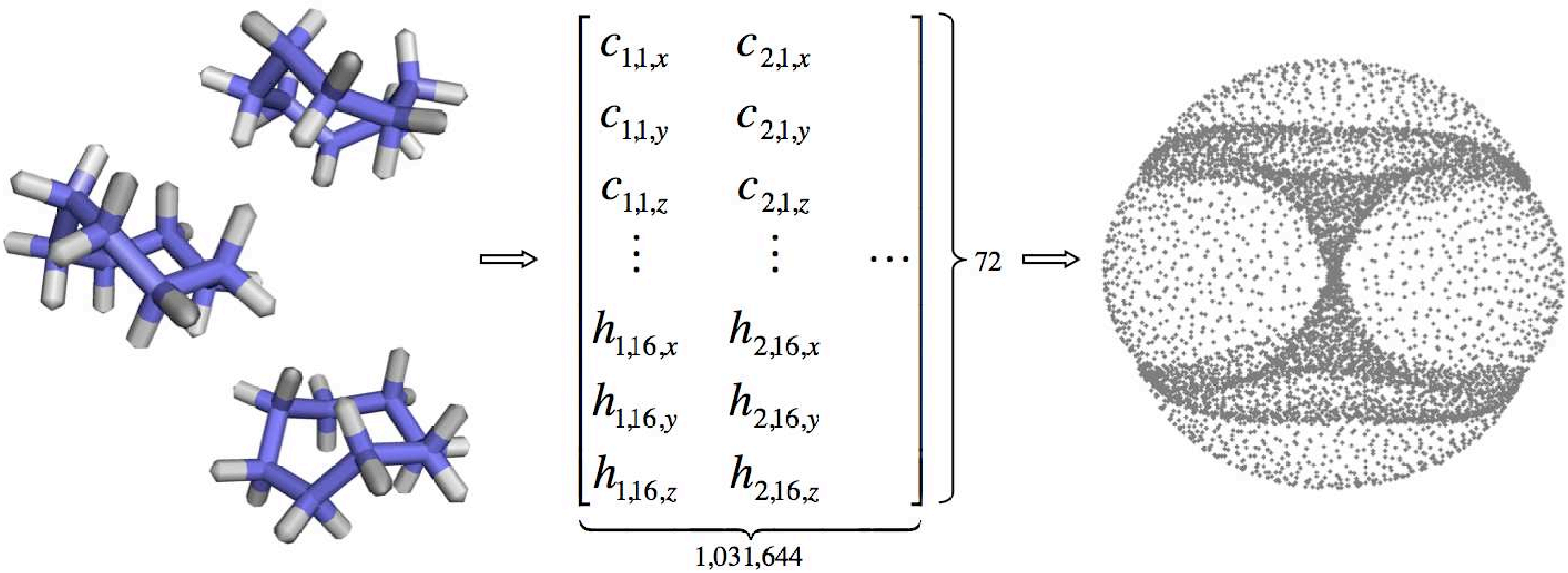


Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Datasets have shapes

Example: Cyclo-Octane (C_8H_{16}) data

1,000,000+ points in 24-dimensional space

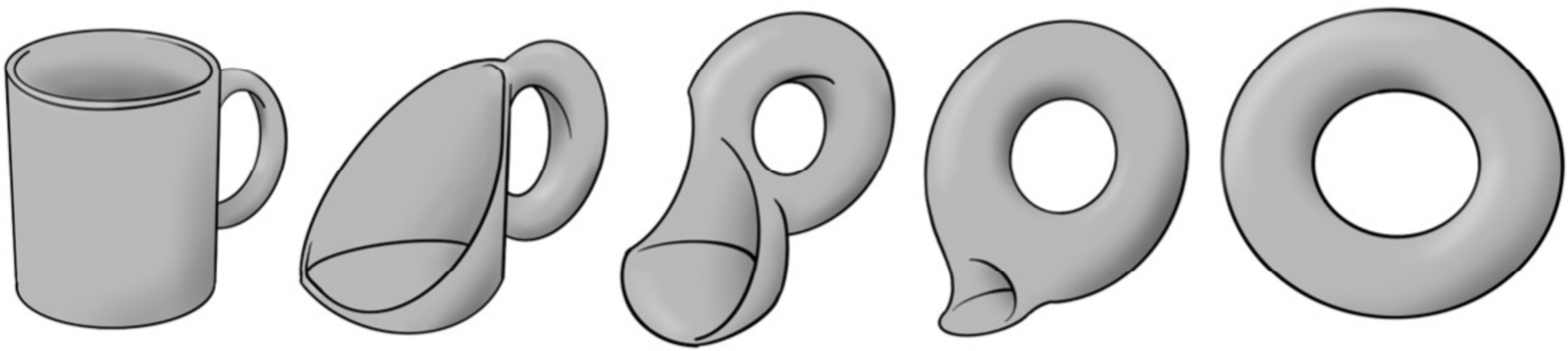


Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Datasets have shapes

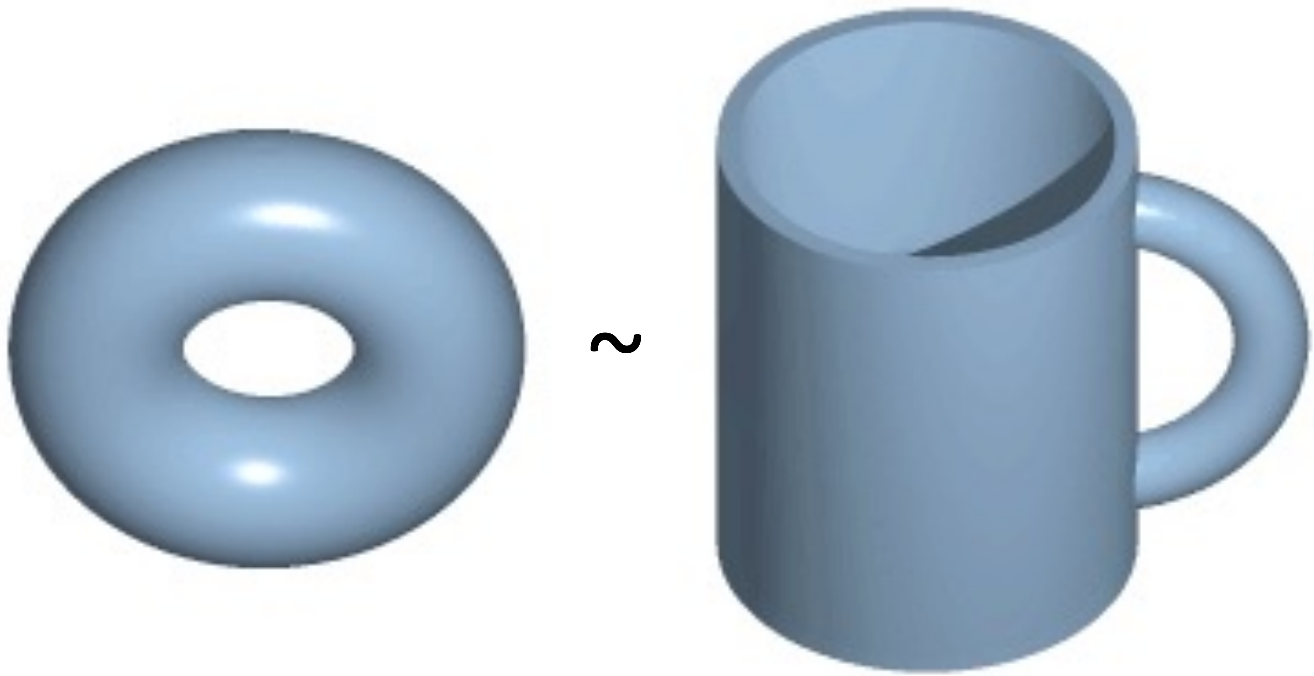
Topology studies shapes

A donut and coffee mug are "homotopy equivalent", and considered to be the same shape. You can bend and stretch (but not tear) one to get the other.

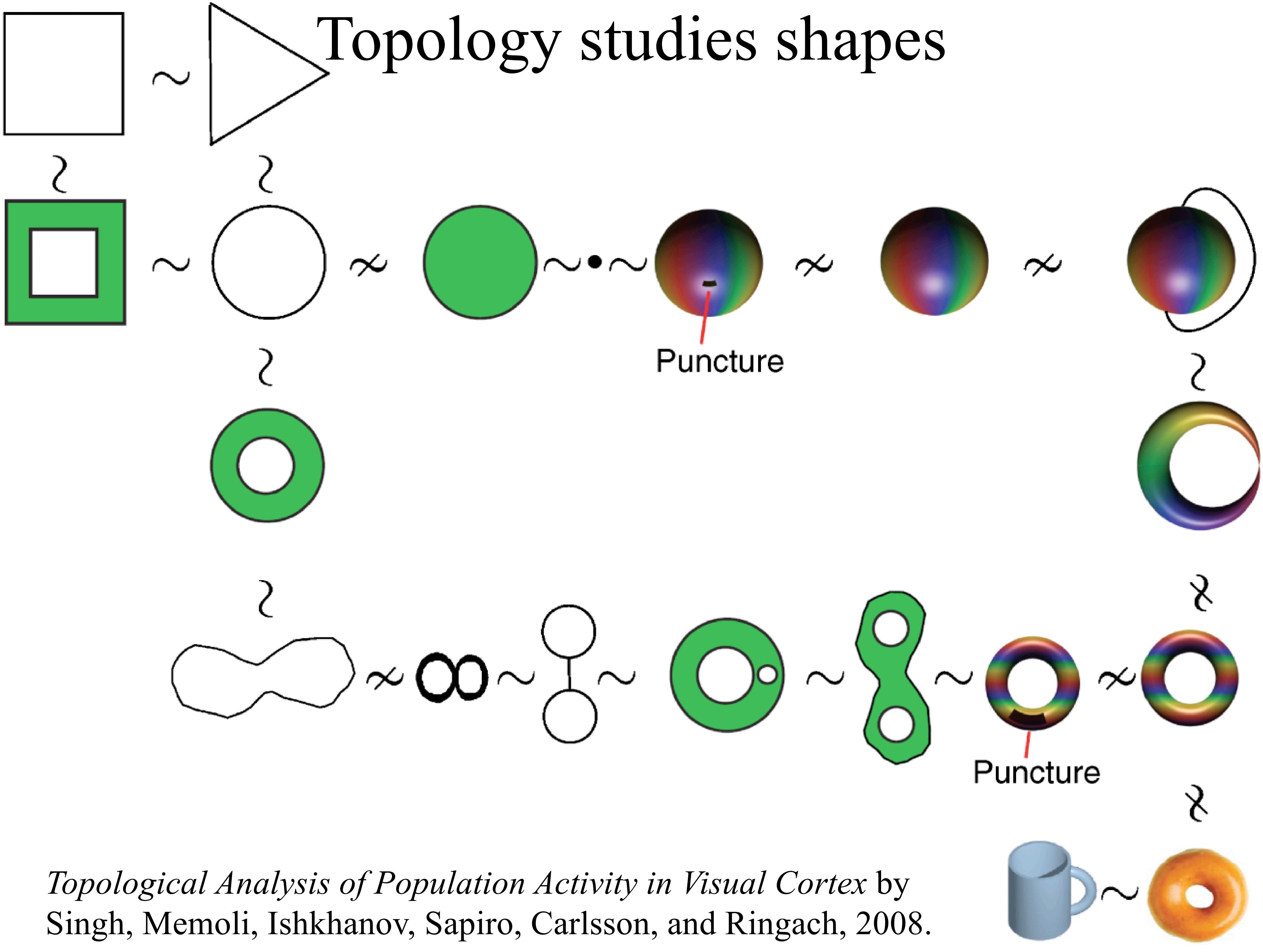


Topology studies shapes

A donut and coffee mug are “homotopy equivalent”, and considered to be the same shape. You can bend and stretch (but not tear) one to get the other.



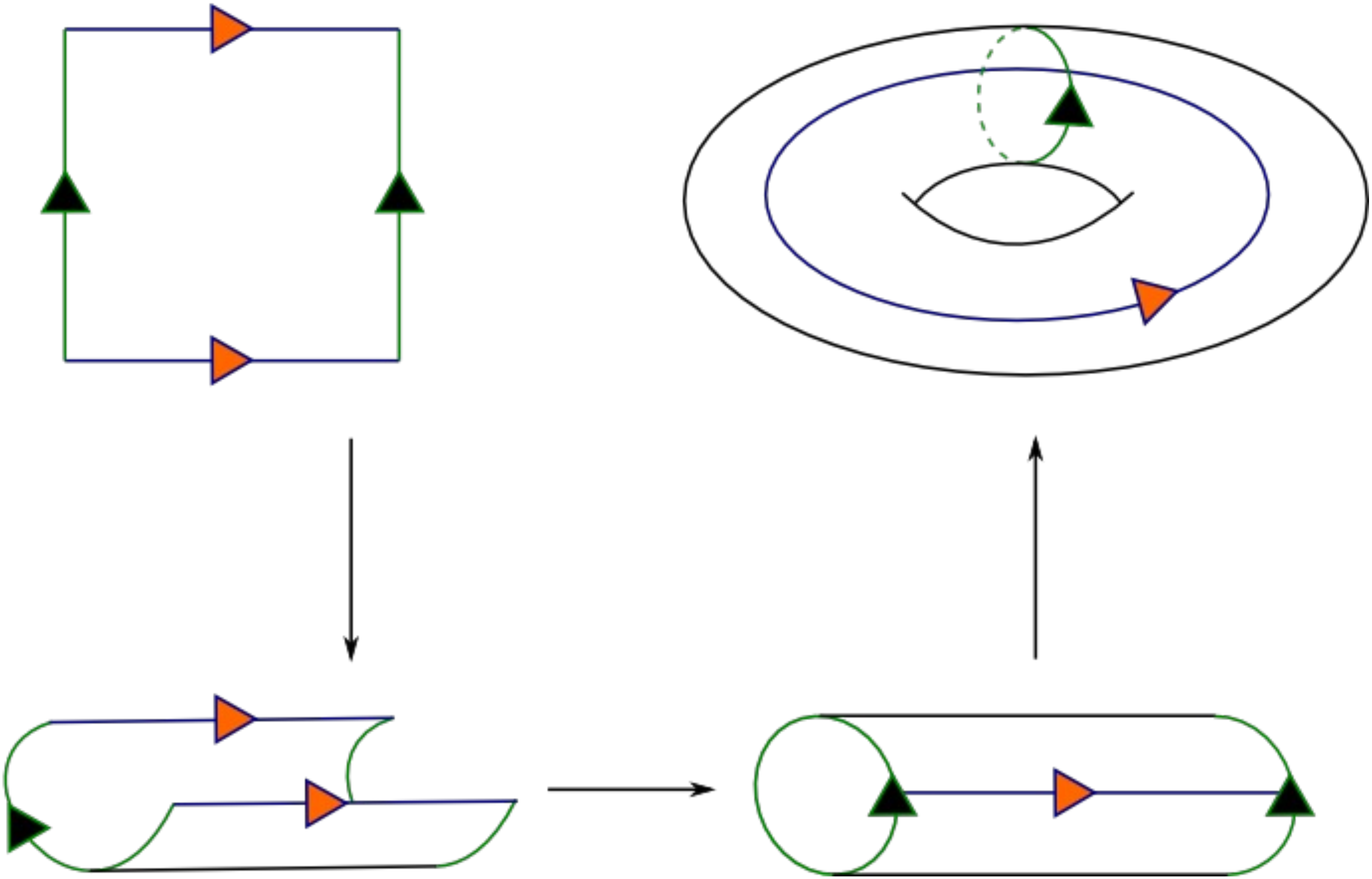
Topology studies shapes



Topological Analysis of Population Activity in Visual Cortex by Singh, Memoli, Ishkhanov, Sapiro, Carlsson, and Ringach, 2008.

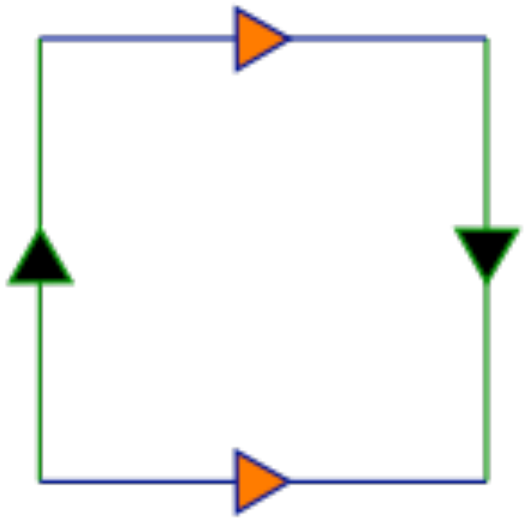
Topology studies shapes

Torus



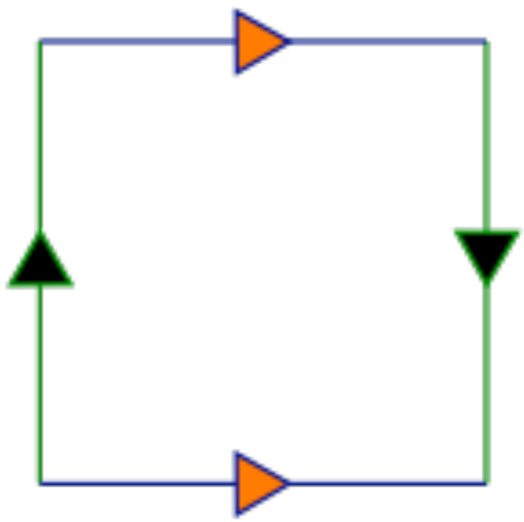
Topology studies shapes

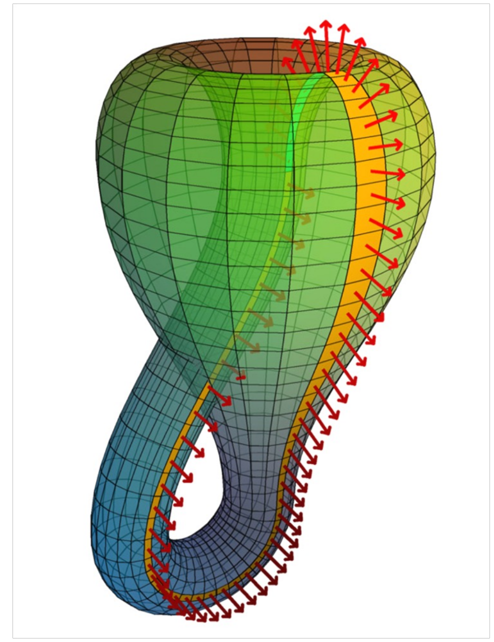
Klein bottle



Topology studies shapes

Klein bottle





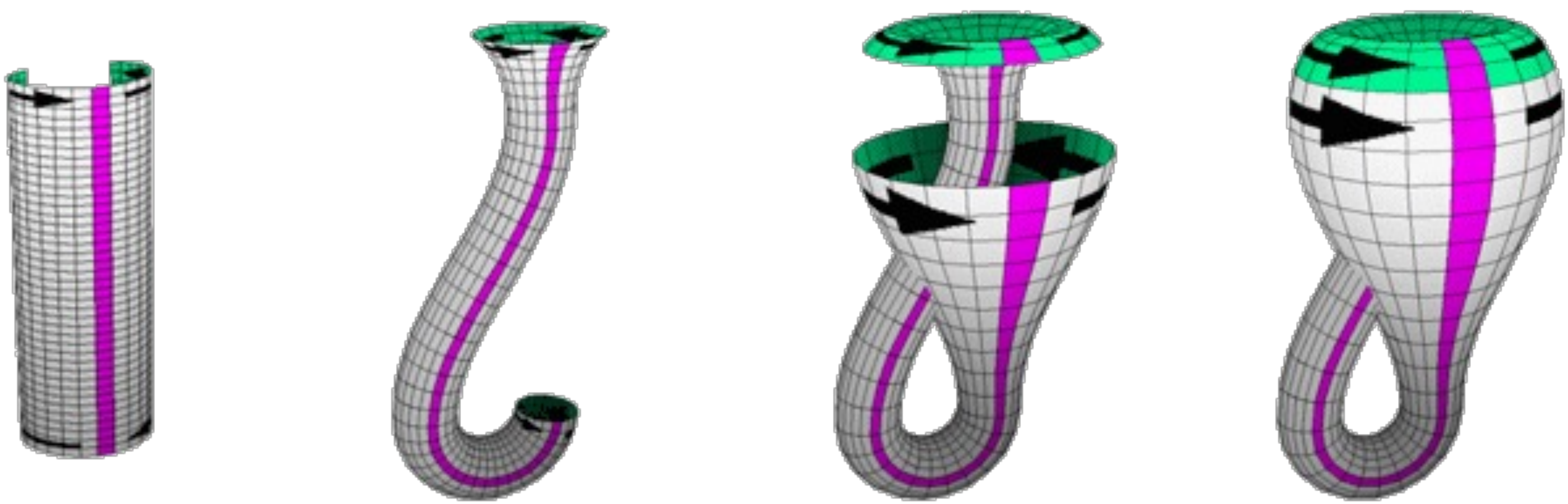
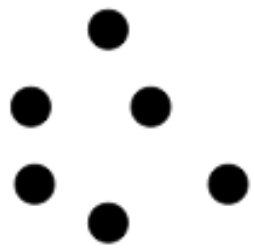


Image credit: https://plus.maths.org/content/imaging-maths-inside-klein-bottle

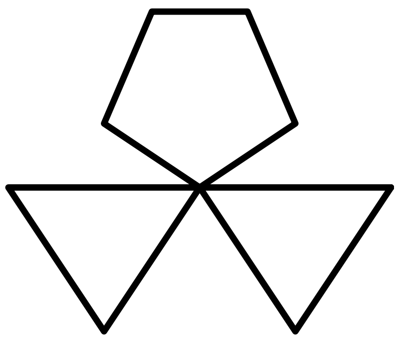
Homology

- i-dimensional homology H_i "counts the number of i-dimensional holes"
- i-dimensional homology H_i actually has the structure of a vector space!



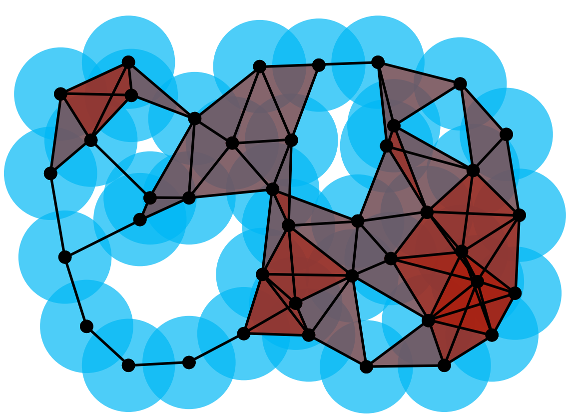
0-dimensional homology H_0: rank 6

1-dimensional homology H_1: rank 0



0-dimensional homology H_0: rank 1

1-dimensional homology H_1: rank 3

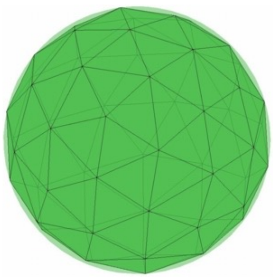


0-dimensional homology H_0: rank 1

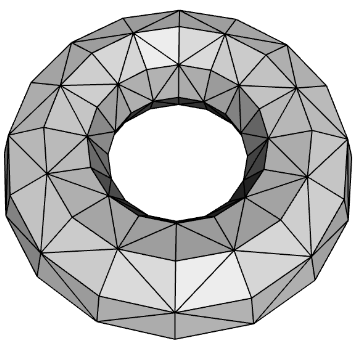
1-dimensional homology H_1: rank 6

Homology

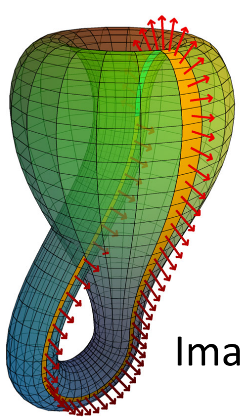
- *i*-dimensional homology "counts the number of *i*-dimensional holes"
- *i*-dimensional homology actually has the structure of a vector space!



0-dimensional homology H_0: rank 1
1-dimensional homology H_1: rank 0
2-dimensional homology H_2: rank 1



0-dimensional homology H_0: rank 1
1-dimensional homology H_1: rank 2
2-dimensional homology H_2: rank 1

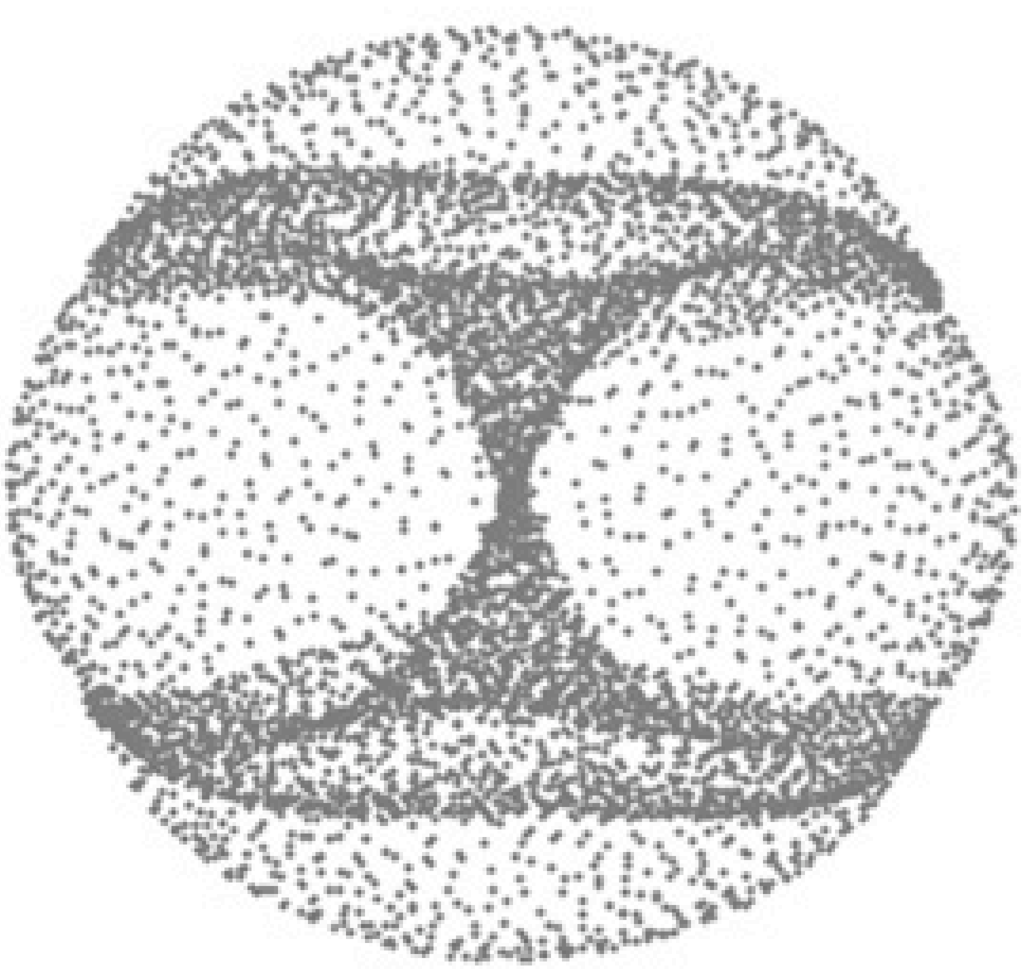


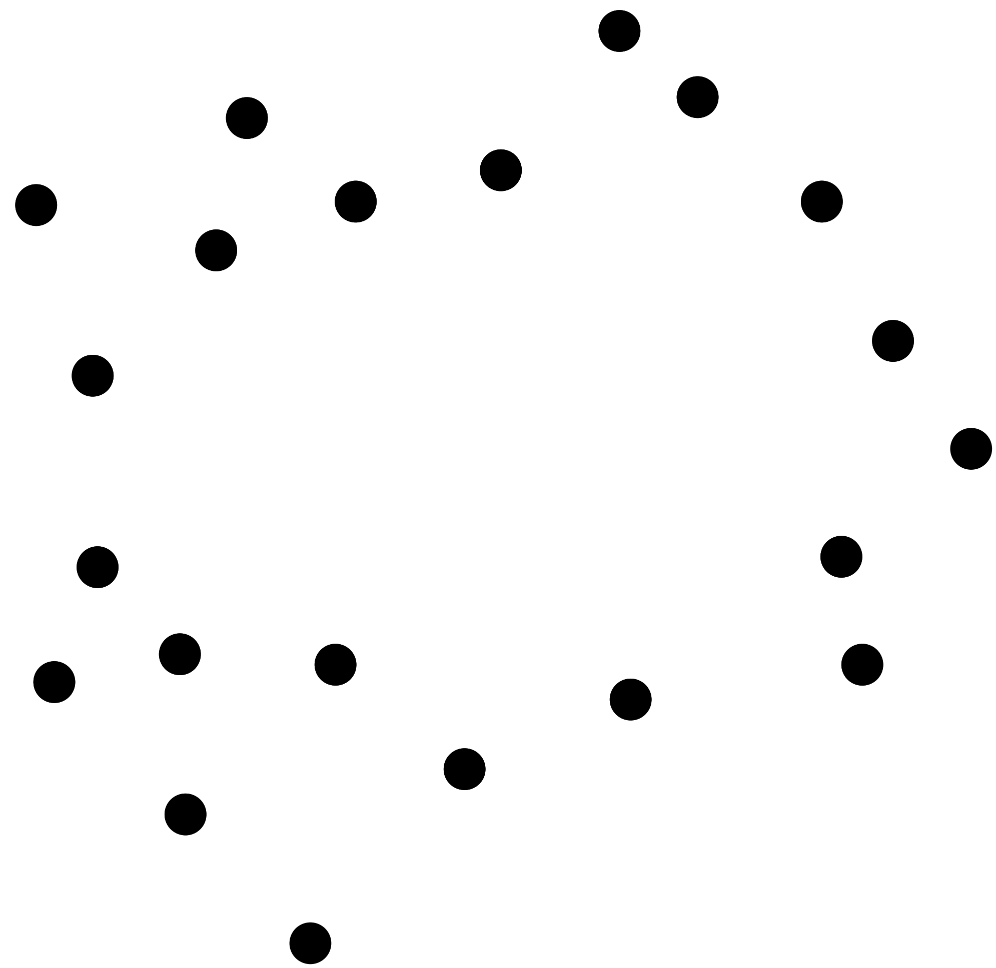
Be careful! (Same as torus over $\mathbb{Z}/2\mathbb{Z}$)

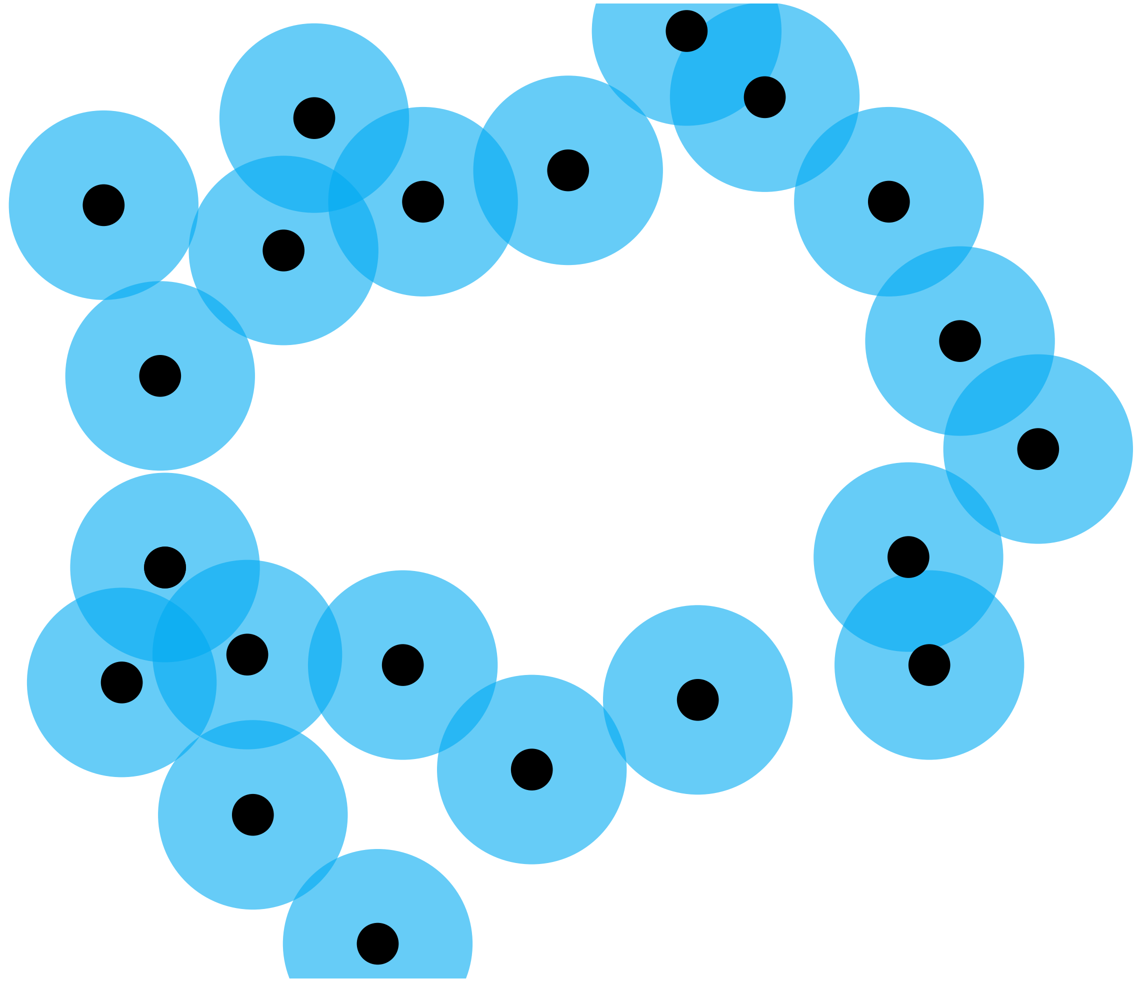
Image credit: https://plus.maths.org/content/imaging-maths-inside-klein-bottle

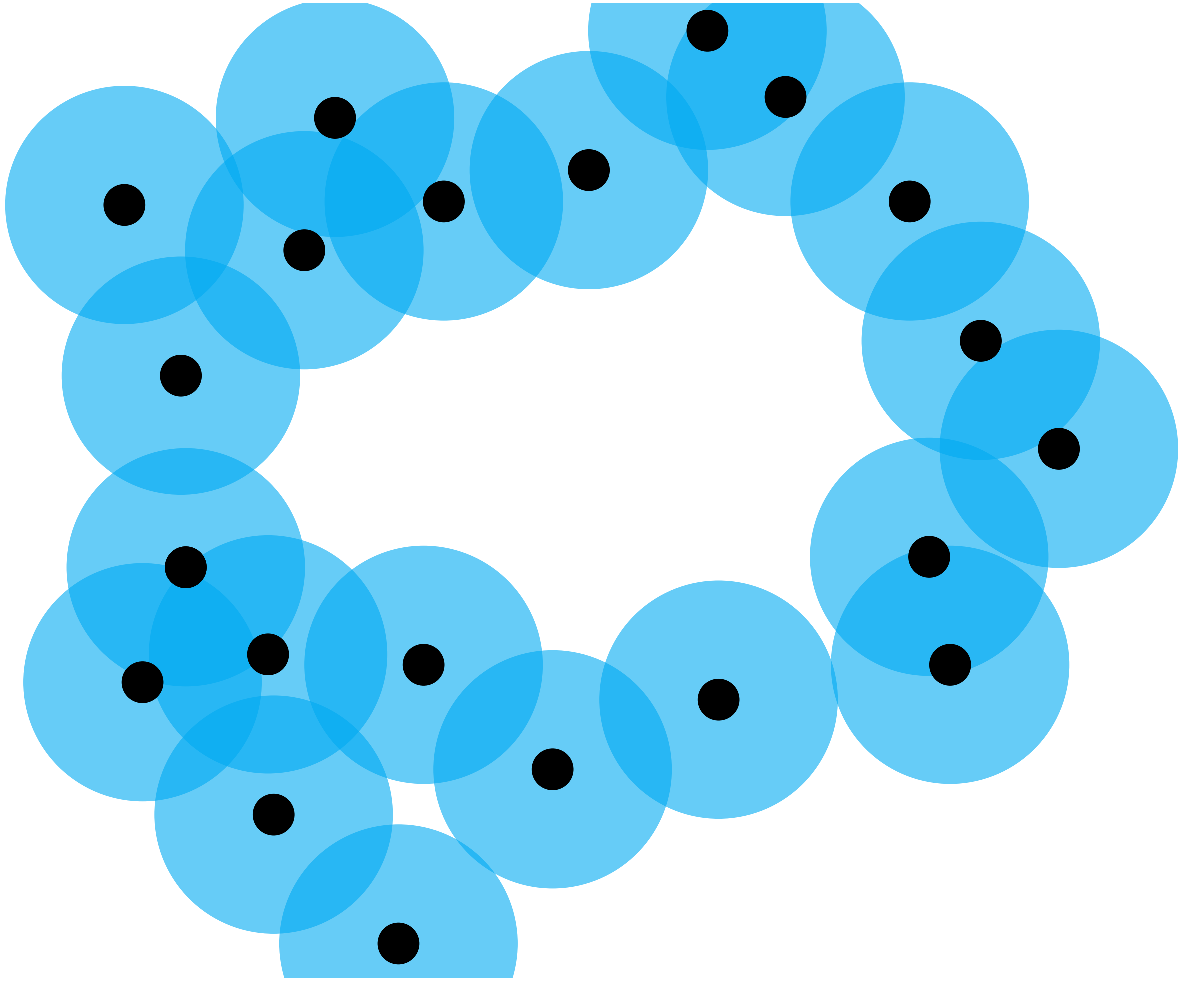
Topology studies shapes

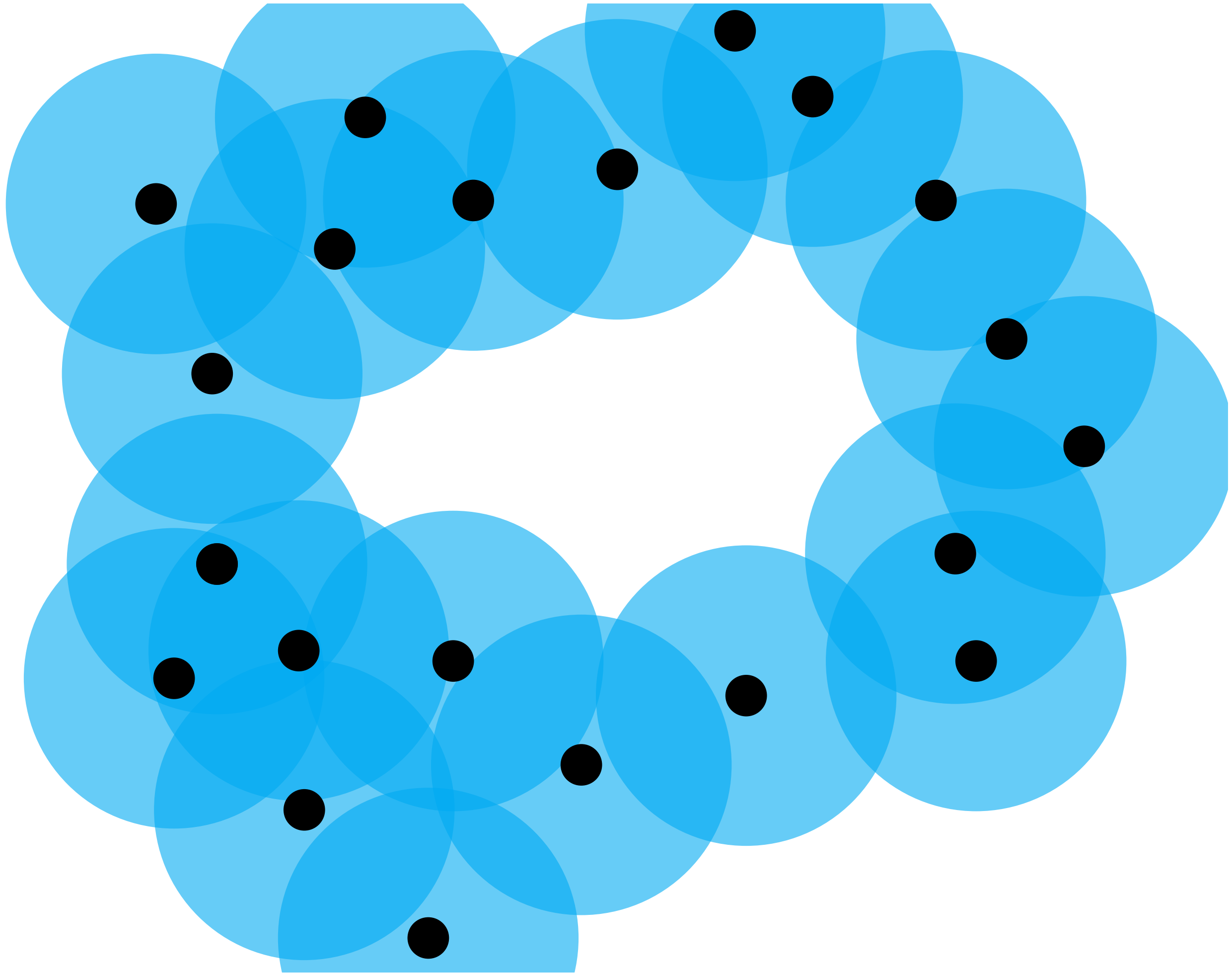
What shape is this?

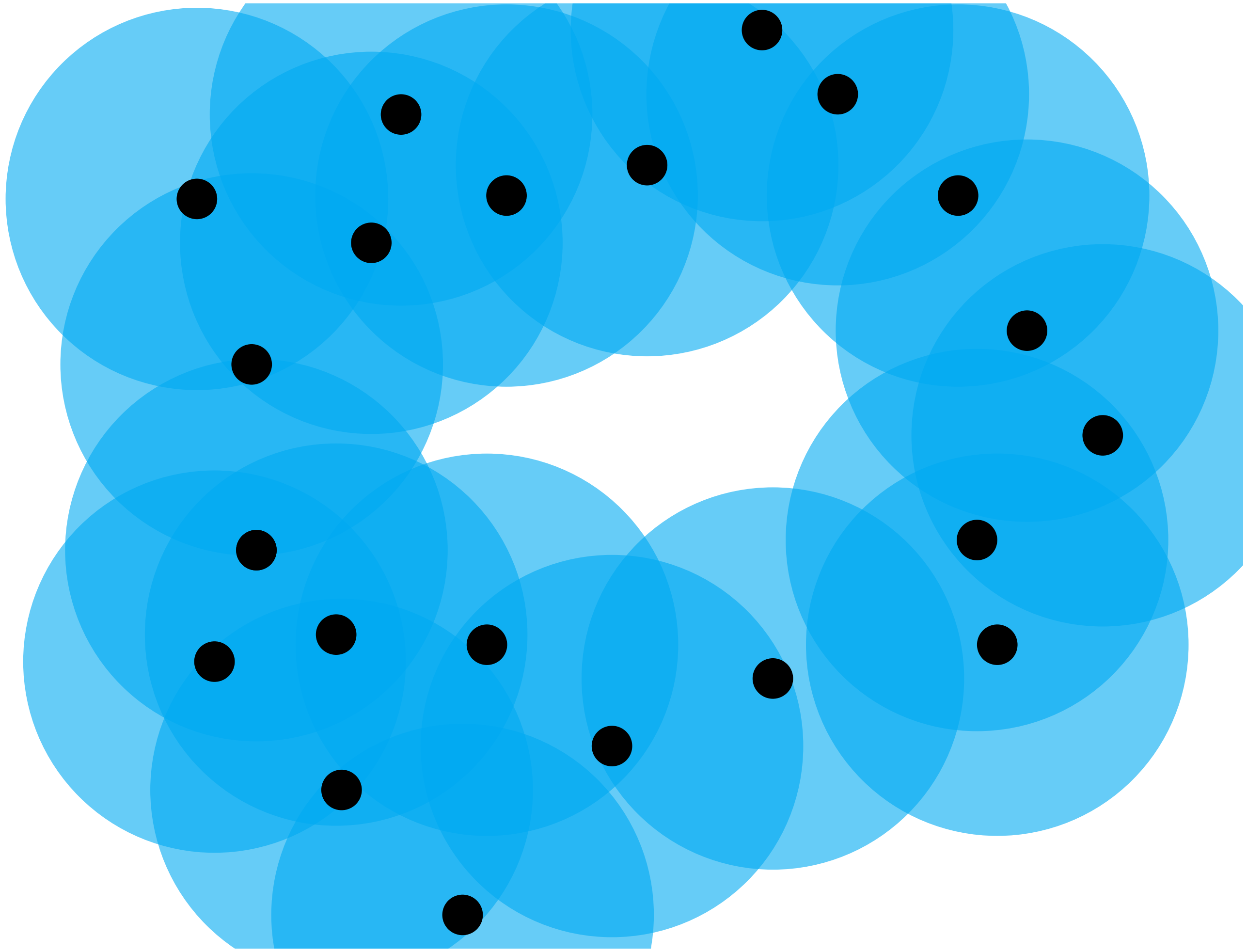


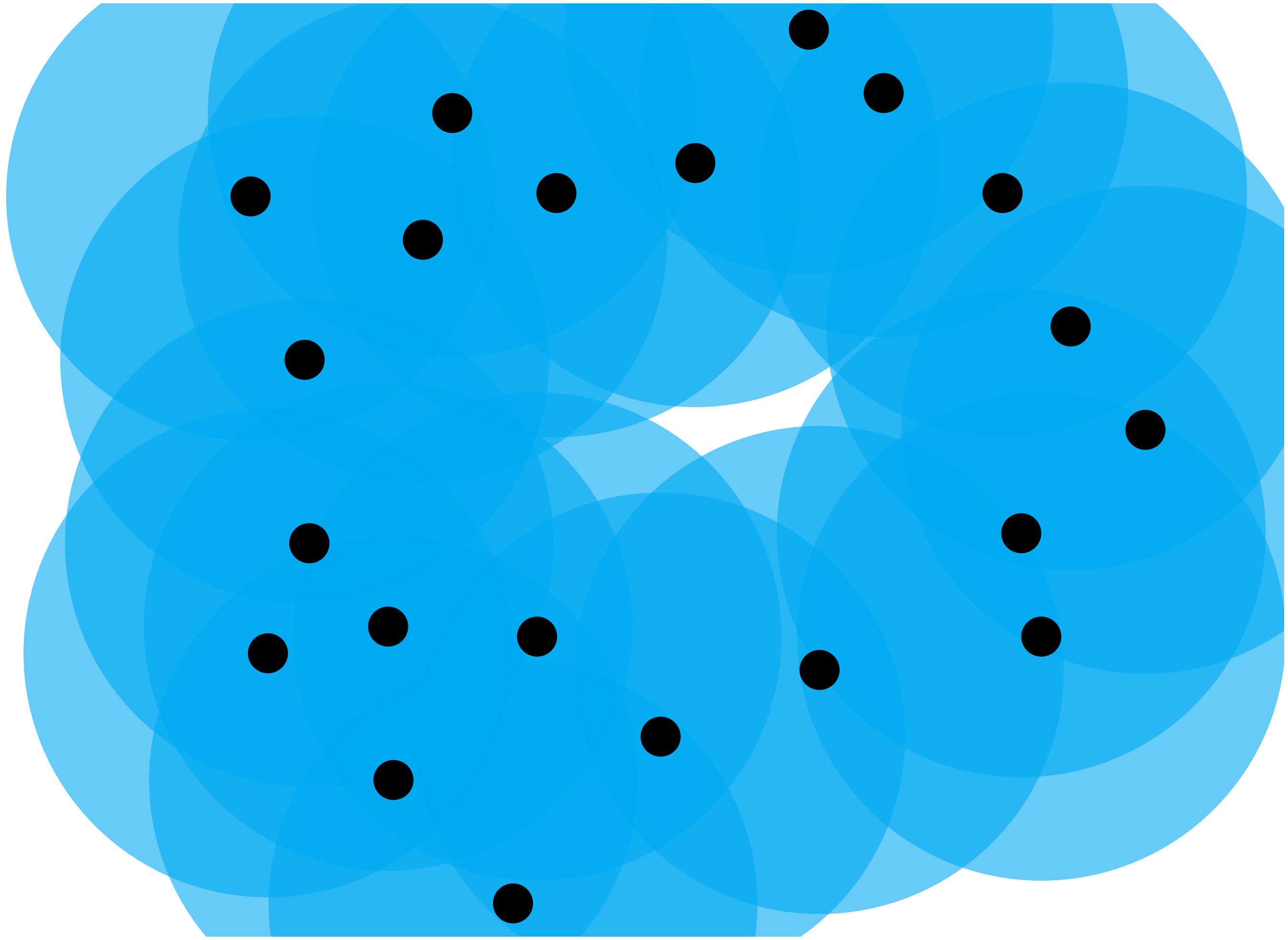


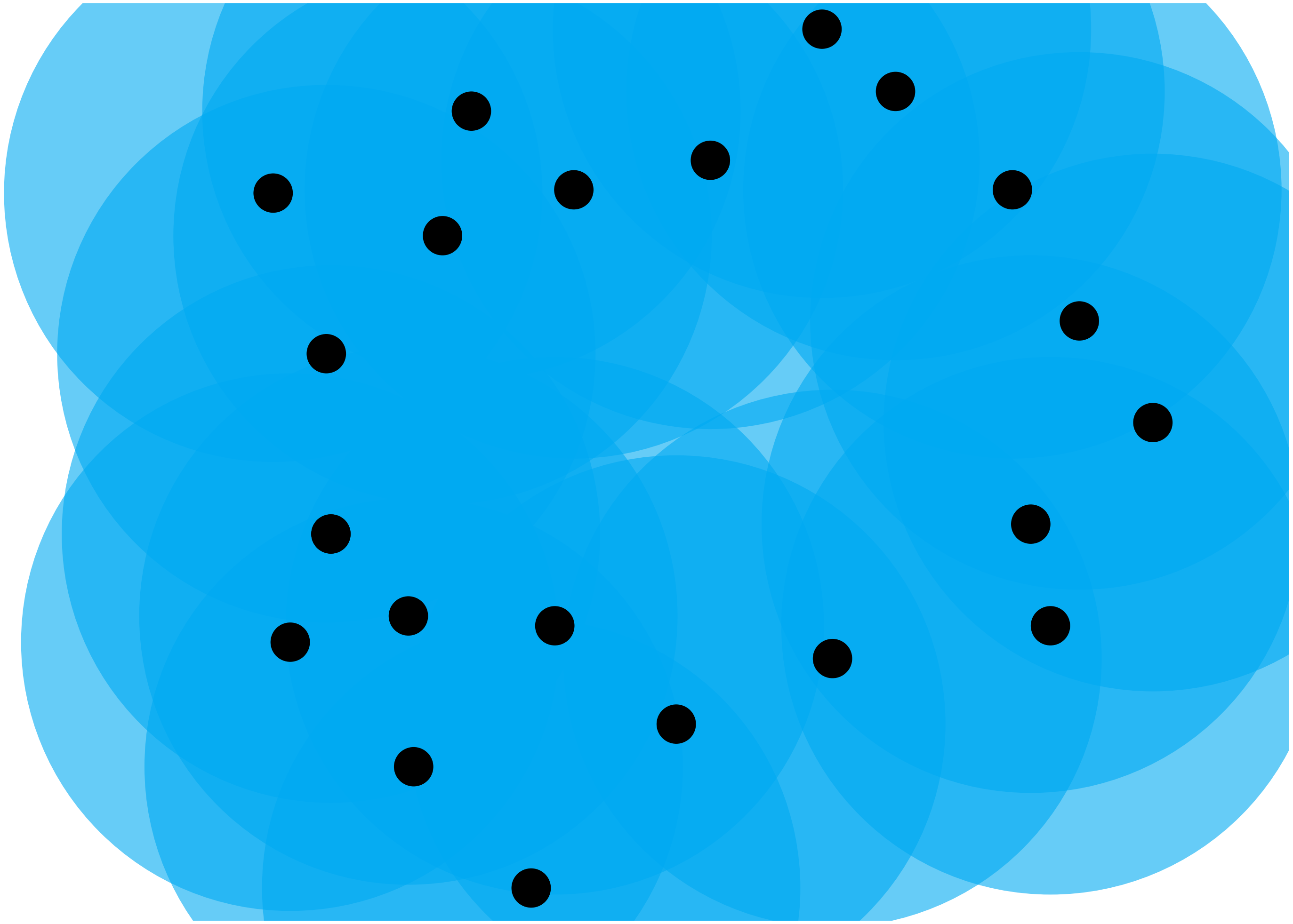


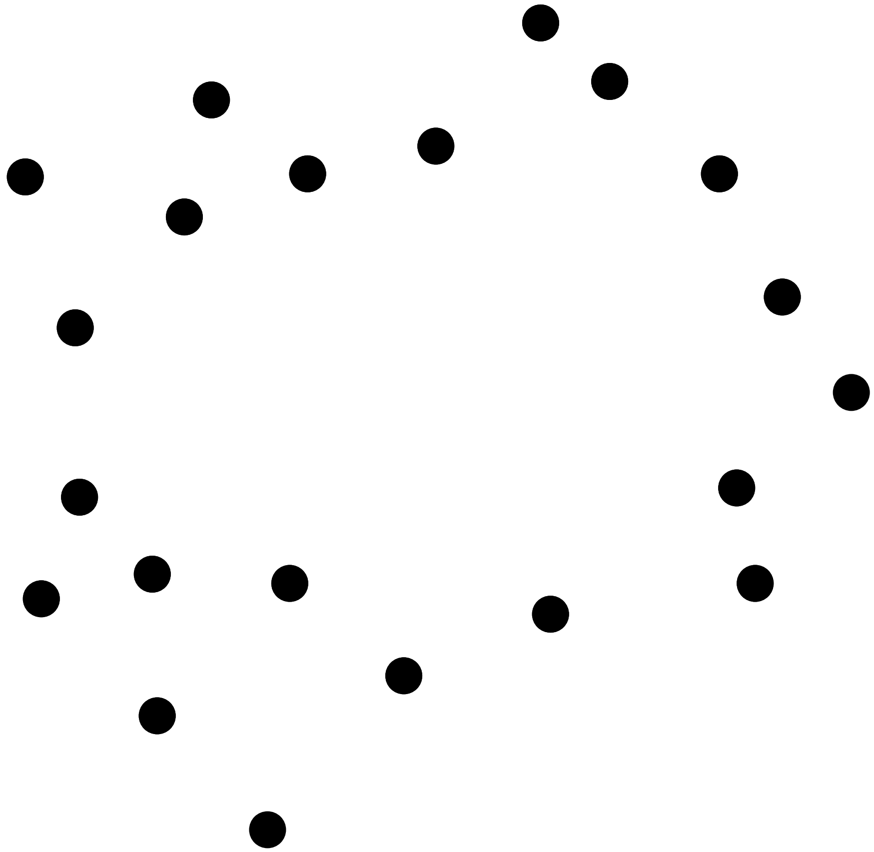








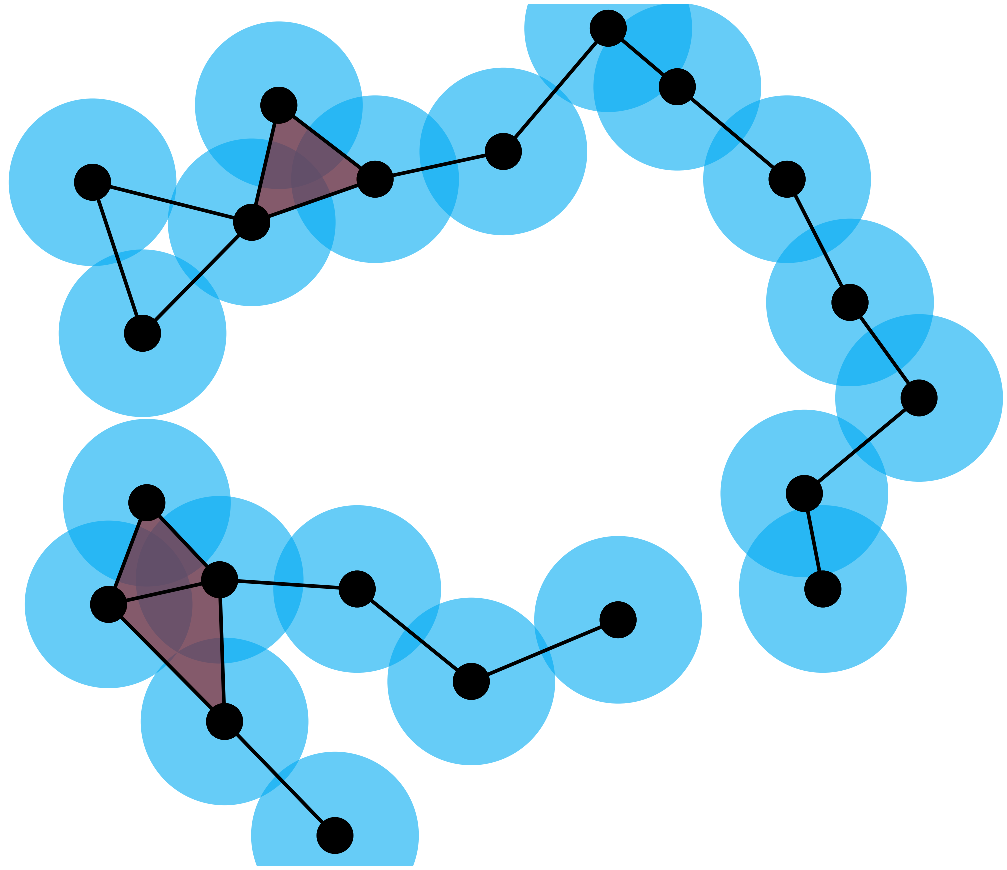




Definition

For a data set $X \subseteq \mathbb{R}^n$ and scale $r \geq 0$, the *Čech simplicial complex* $\check{\text{C}}\text{ech}(X; r)$ has

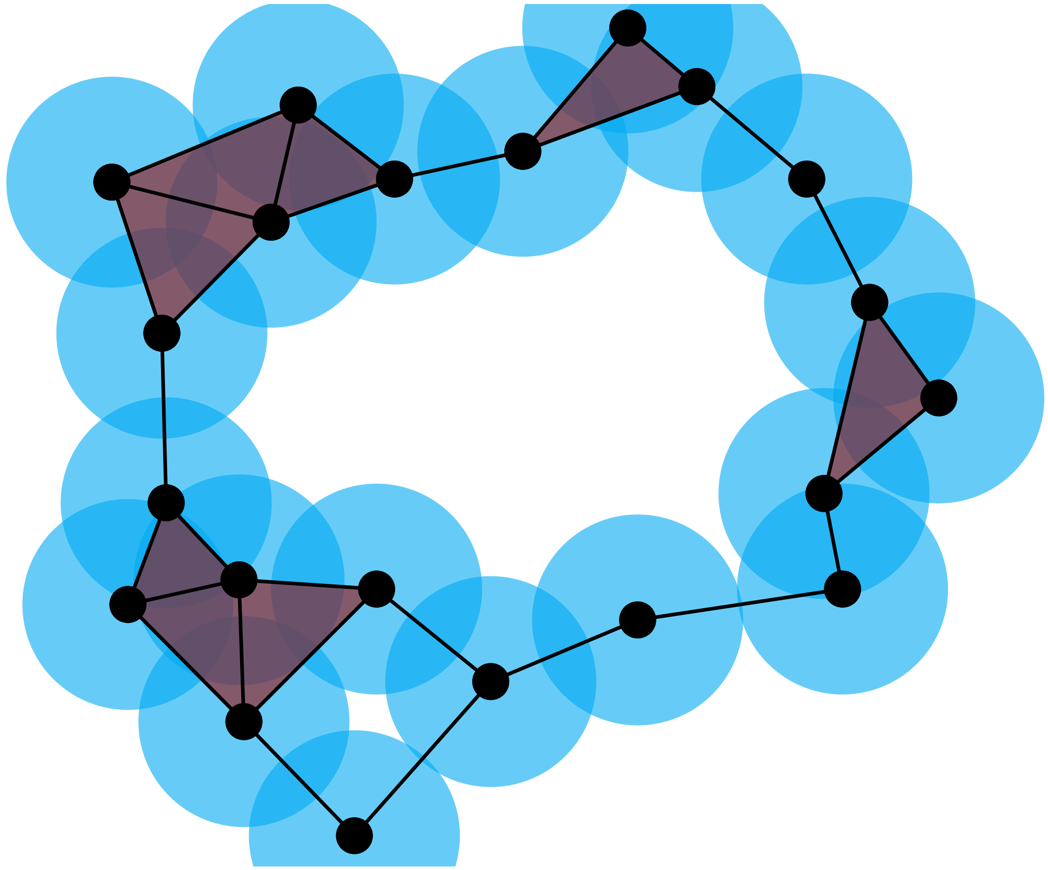
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

For a data set $X \subseteq \mathbb{R}^n$ and scale $r \geq 0$, the *Čech simplicial complex* $\check{\mathrm{C}}\mathrm{ech}(X; r)$ has

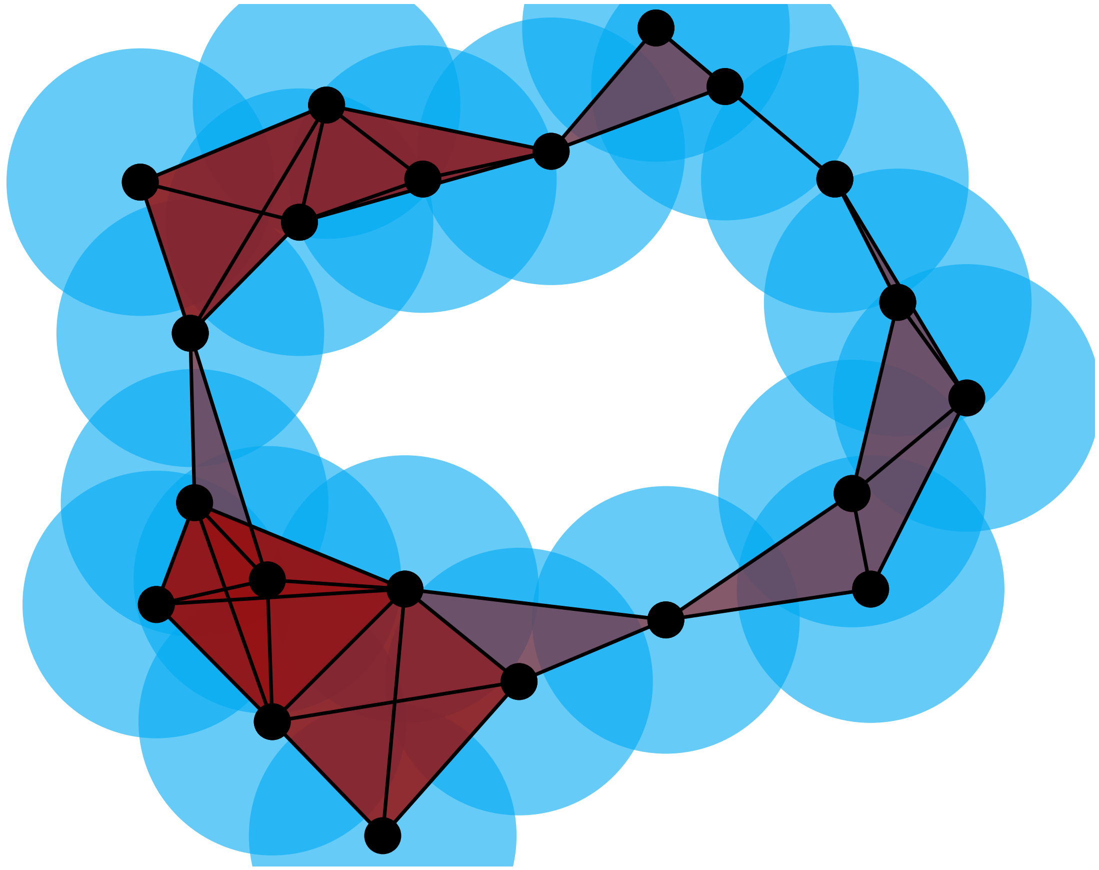
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

For a data set $X \subseteq \mathbb{R}^n$ and scale $r \geq 0$, the *Čech simplicial complex* $\check{\mathrm{C}}\mathrm{ech}(X; r)$ has

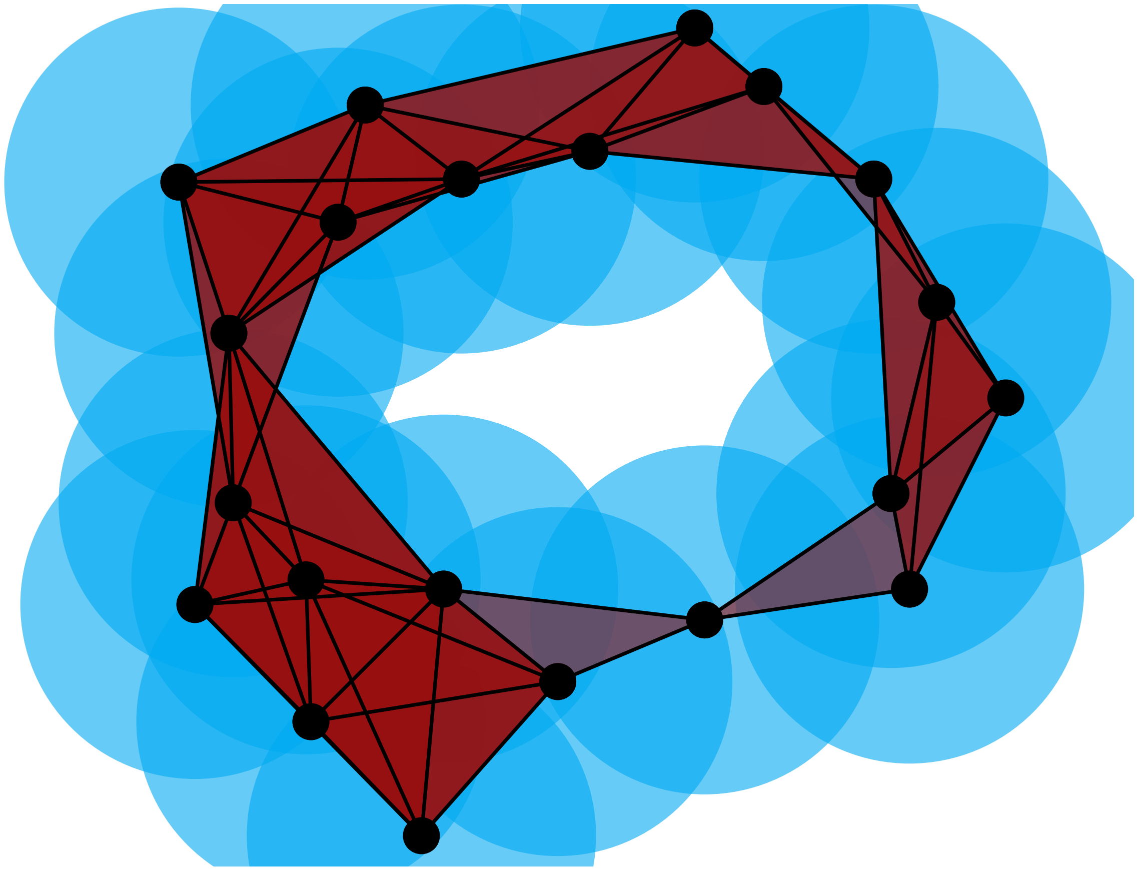
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

For a data set $X \subseteq \mathbb{R}^n$ and scale $r \geq 0$, the *Čech simplicial complex* $\text{Čech}(X; r)$ has

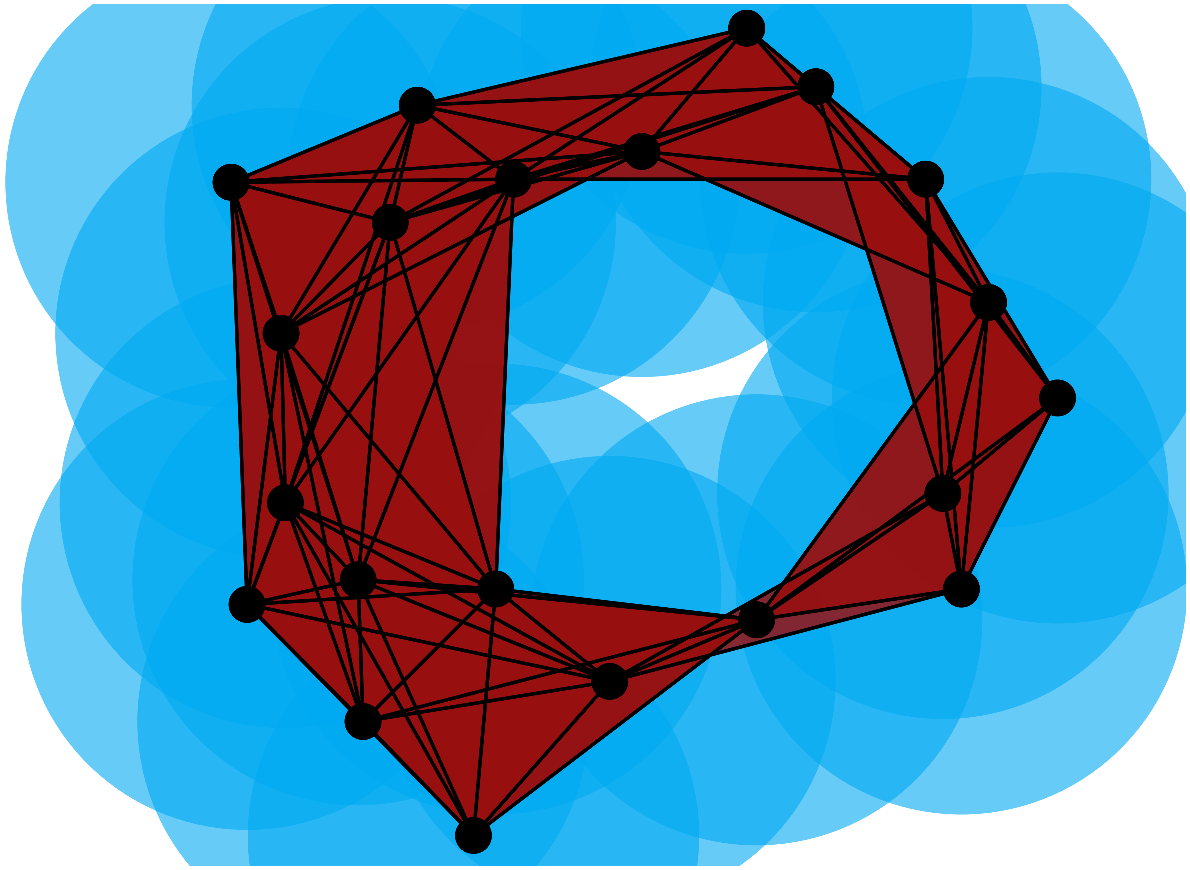
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

For a data set $X \subseteq \mathbb{R}^n$ and scale $r \geq 0$, the *Čech simplicial complex* $\check{\mathrm{C}}\mathrm{ech}(X; r)$ has

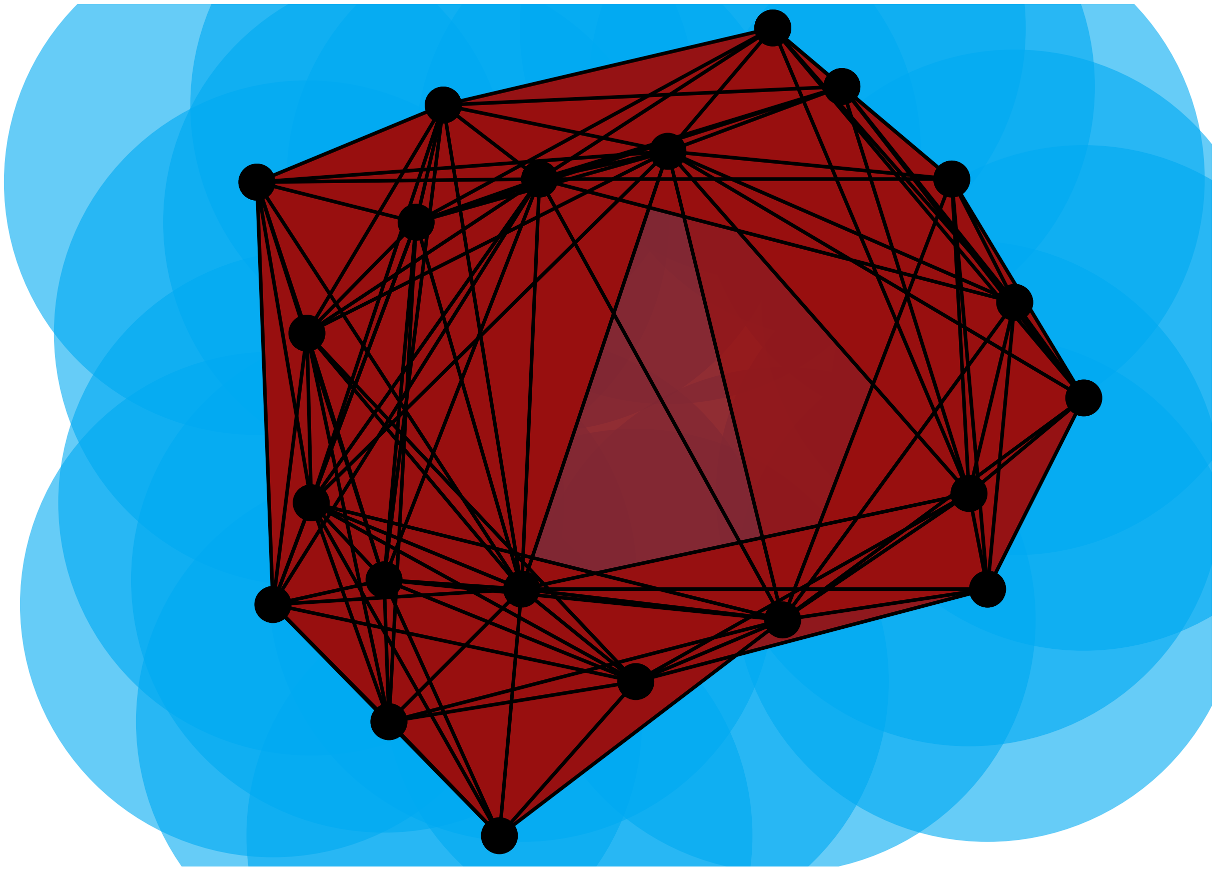
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

For a data set $X \subseteq \mathbb{R}^n$ and scale $r \geq 0$, the *Čech simplicial complex* $\check{\mathrm{C}}\mathrm{ech}(X; r)$ has

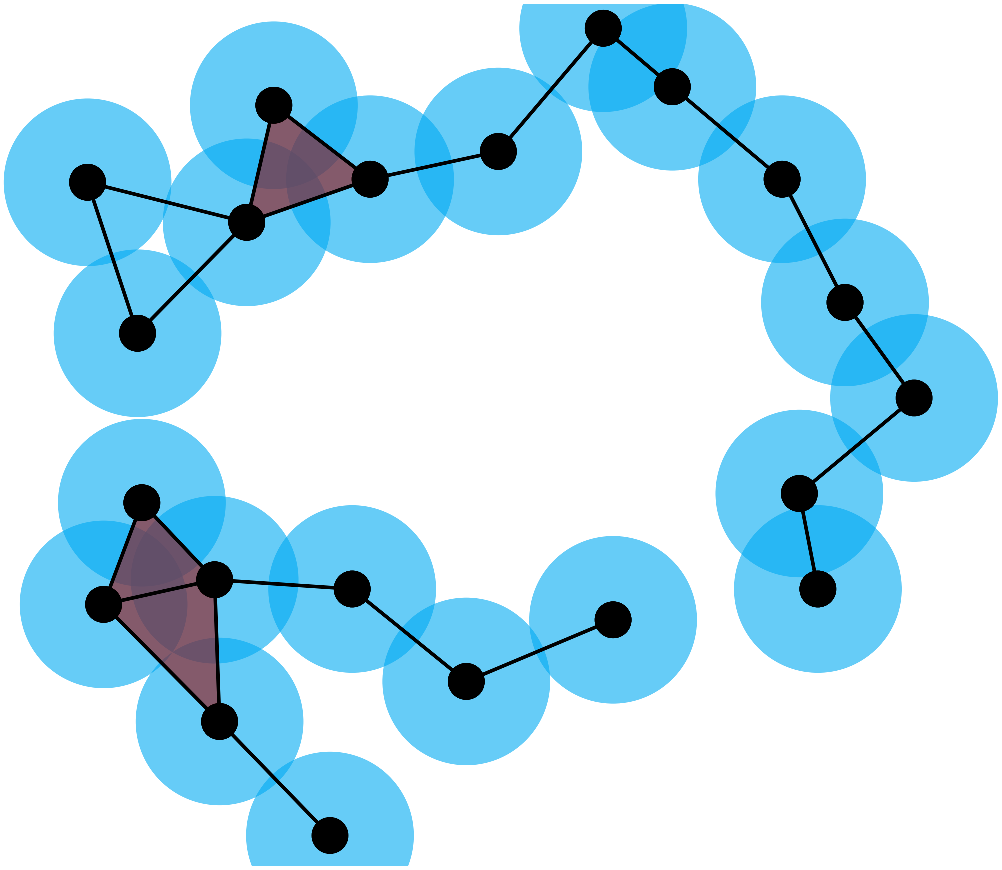
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

For a data set $X \subseteq \mathbb{R}^n$ and scale $r \geq 0$, the *Čech simplicial complex* $\check{\mathrm{C}}\mathrm{ech}(X; r)$ has

- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



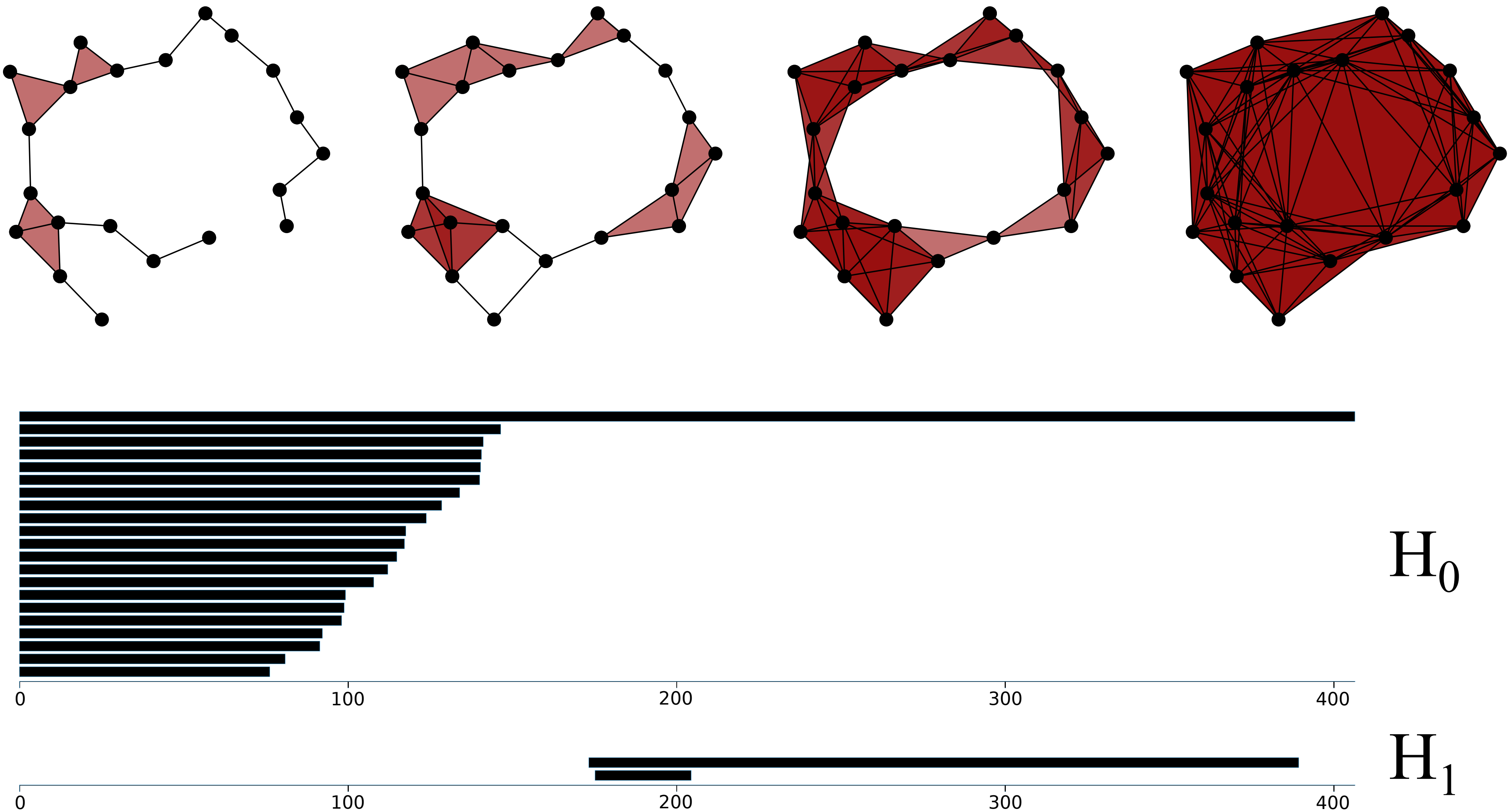
Nerve Lemma. $\check{\mathrm{C}}\mathrm{ech}(X;r) \simeq$ union of balls

Definition

For a data set $X \subseteq \mathbb{R}^n$ and scale $r \geq 0$, the *Čech simplicial complex* $\check{\mathrm{C}}\mathrm{ech}(X;r)$ has

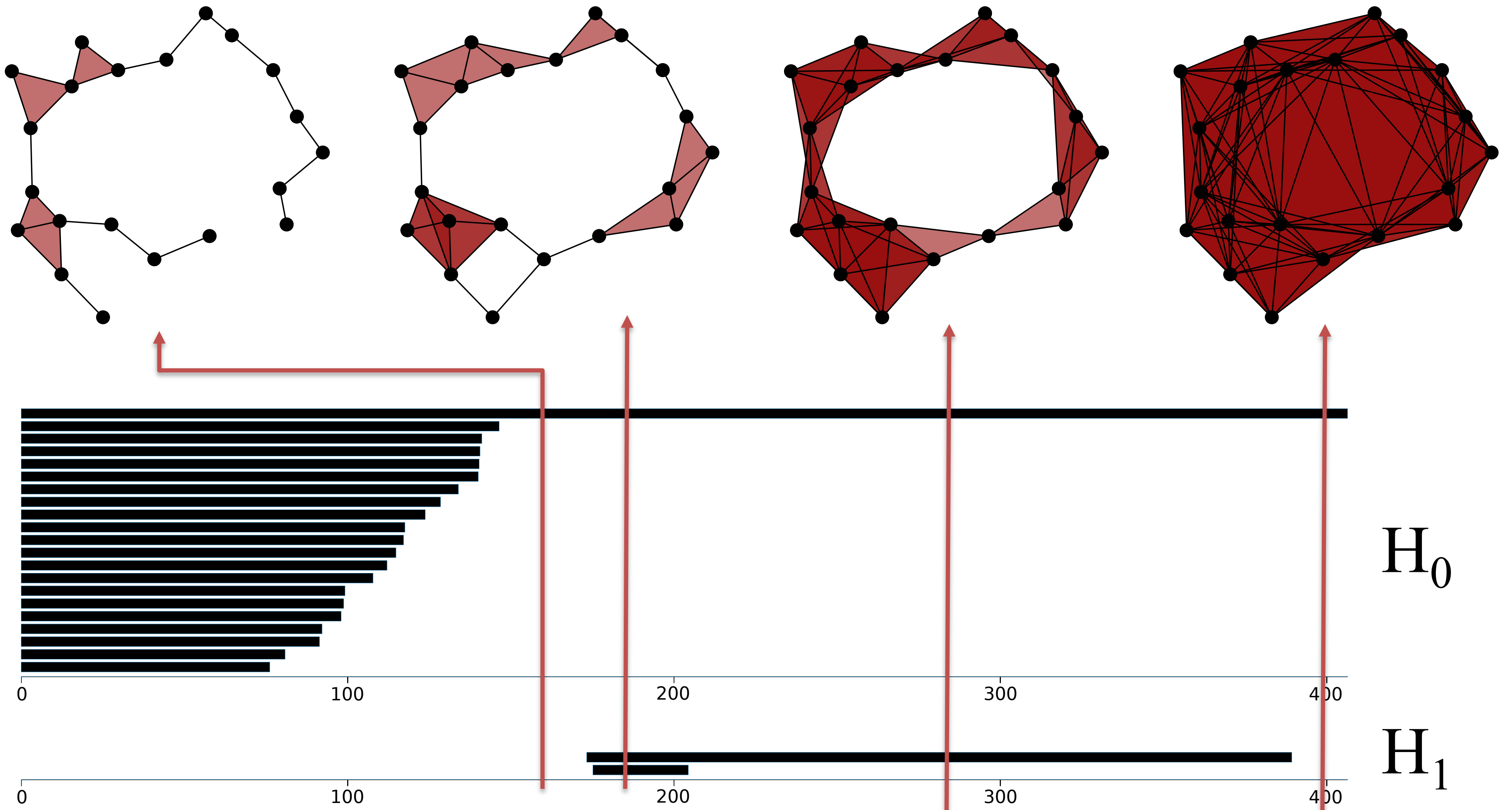
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.

Persistent homology



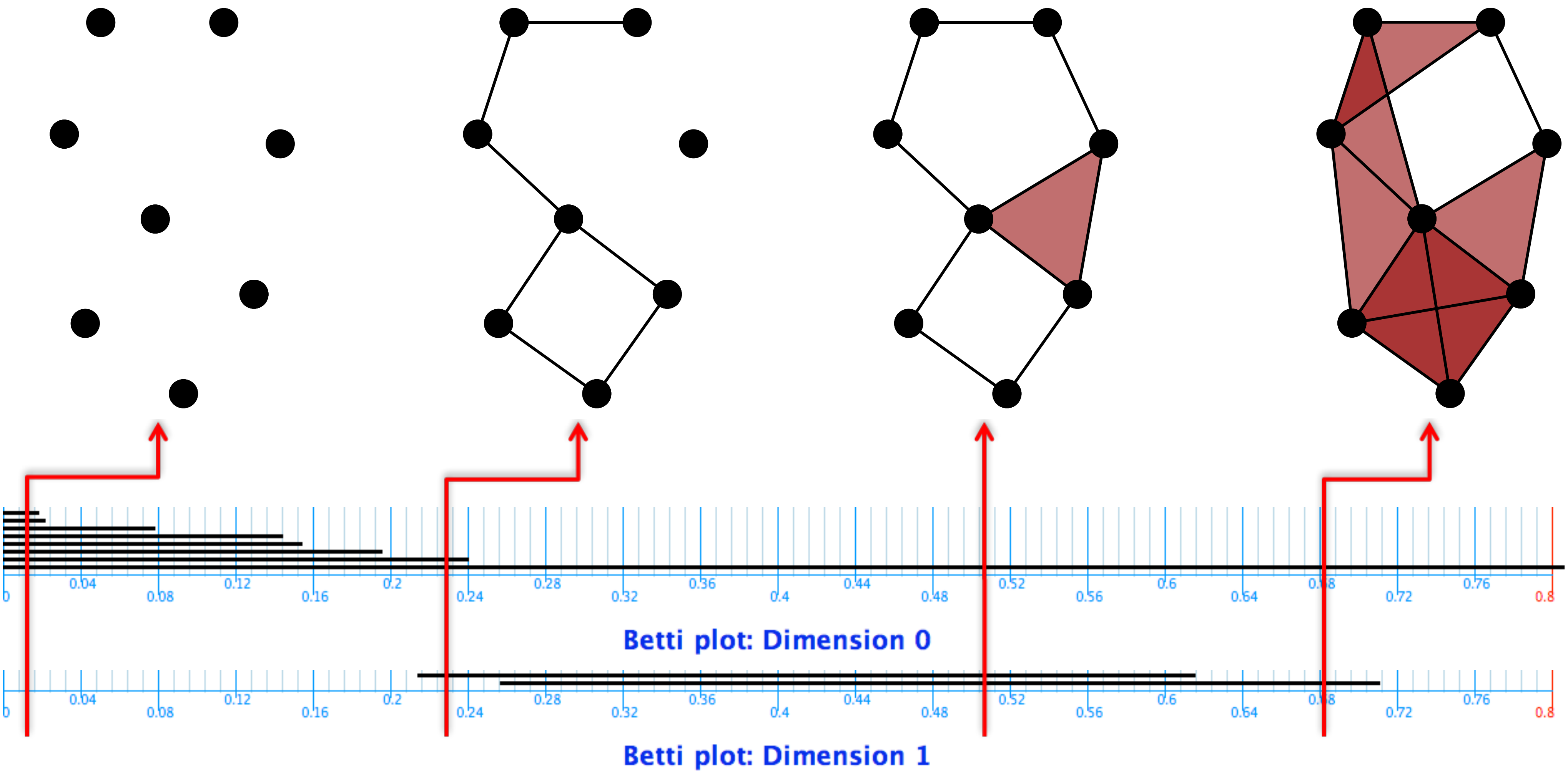
- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology



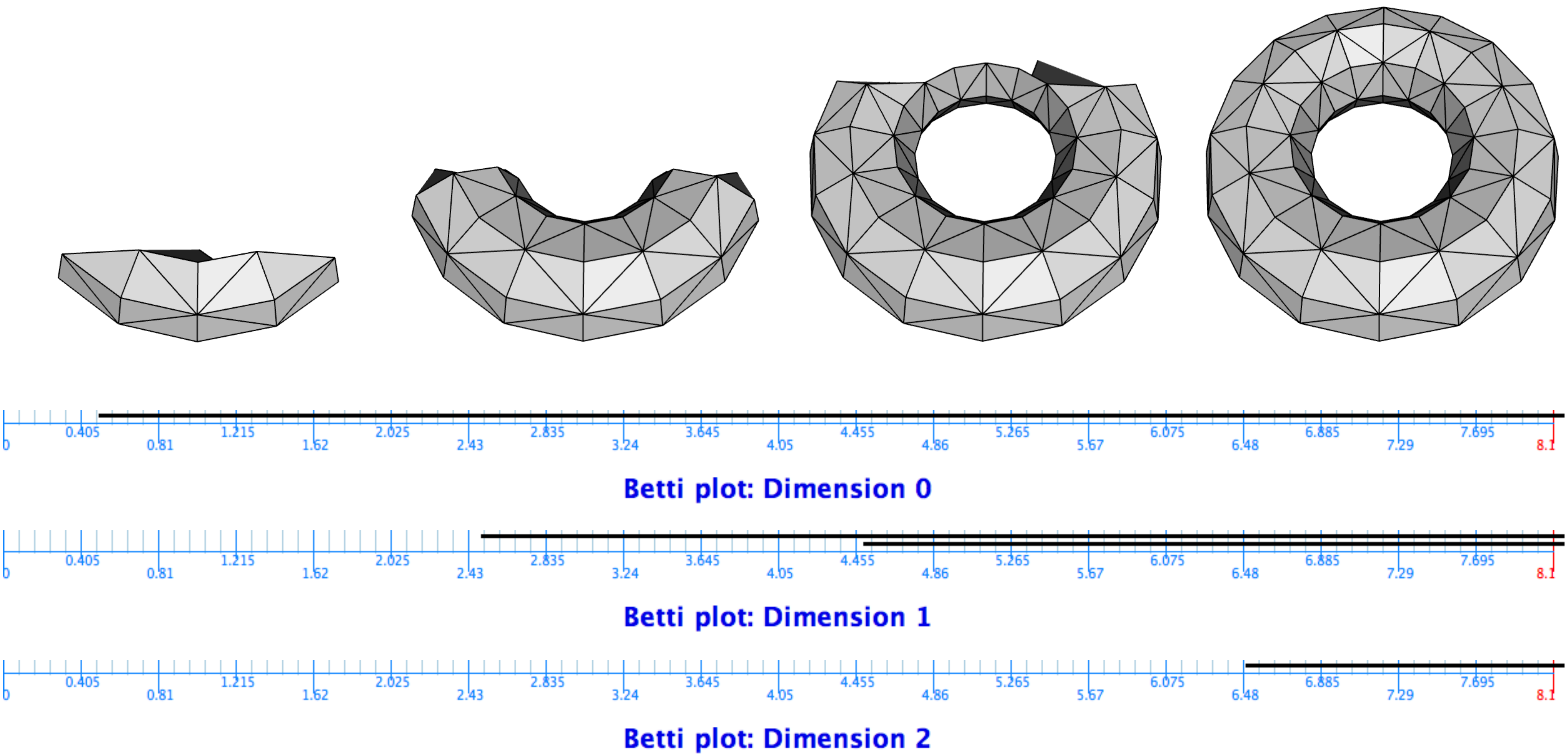
- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology



- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology

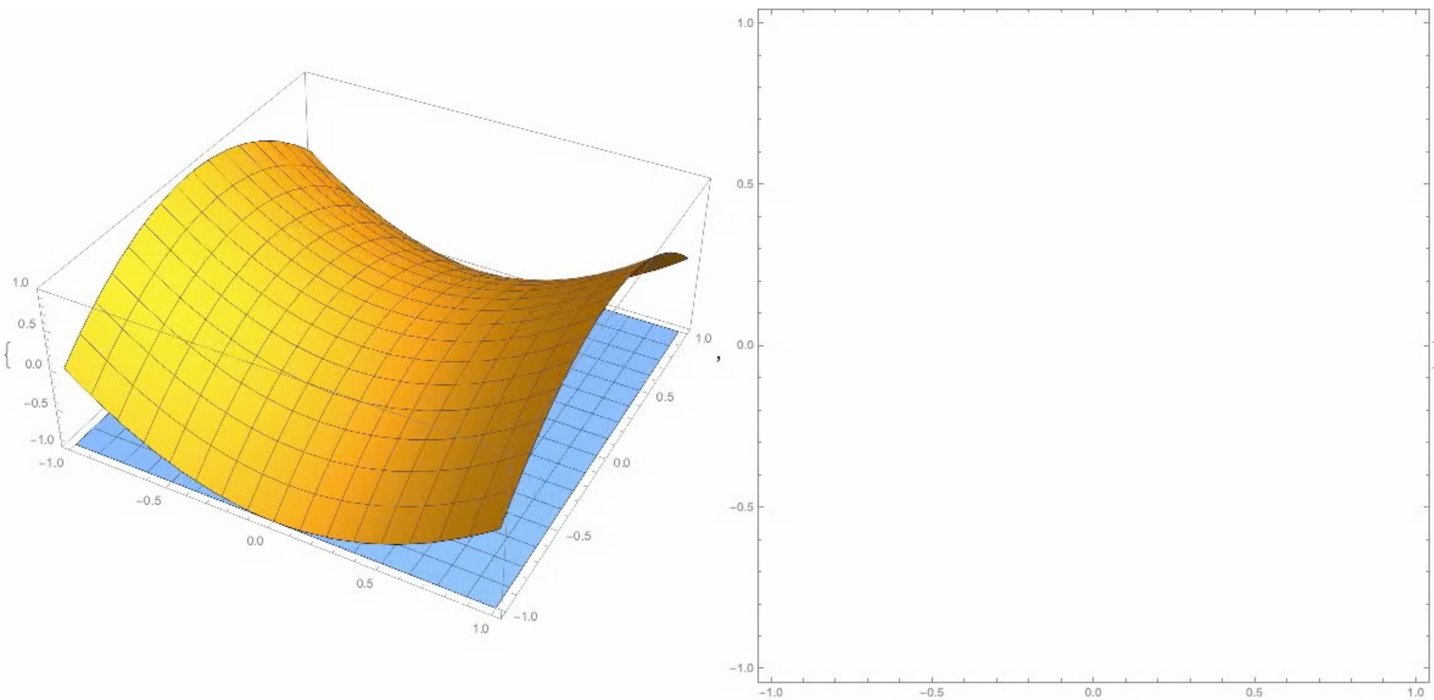


- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Sublevel set persistent homology

Given a space X and a real-valued function $f: X \to \mathbb{R}$, the **sublevel set** for $a \in \mathbb{R}$ consists of all the points in X with $f(x) \leq a$.

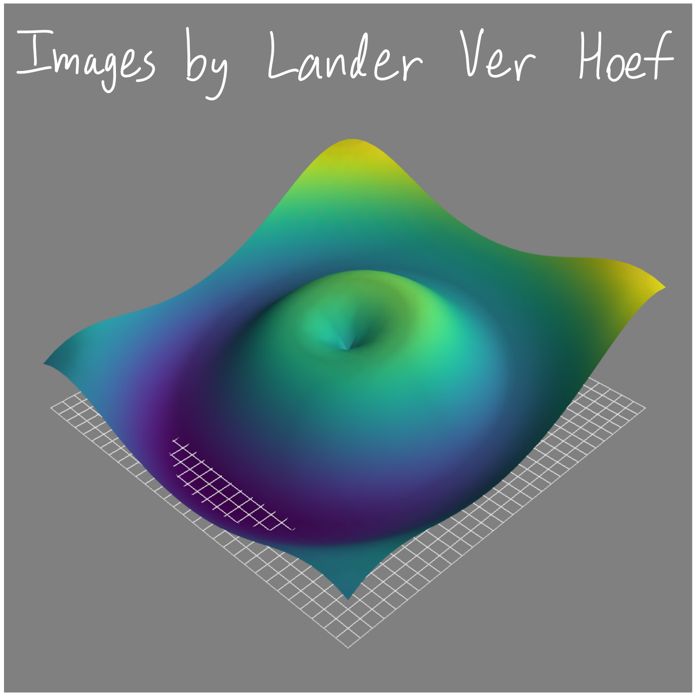
Example: $f(x) = x^2 - y^2$ on $X = [-1,1] \times [-1,1]$.

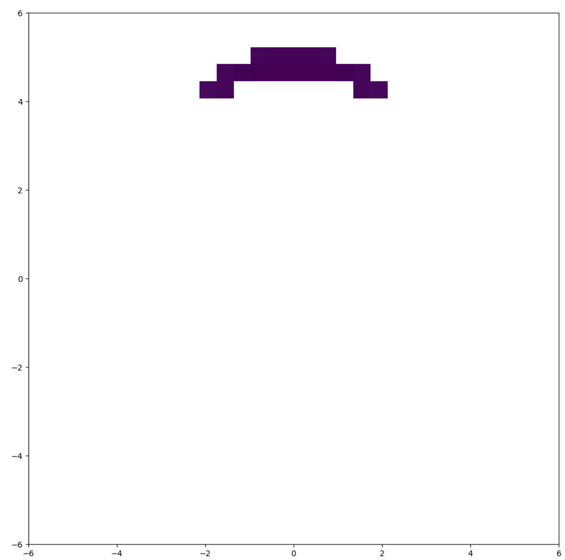


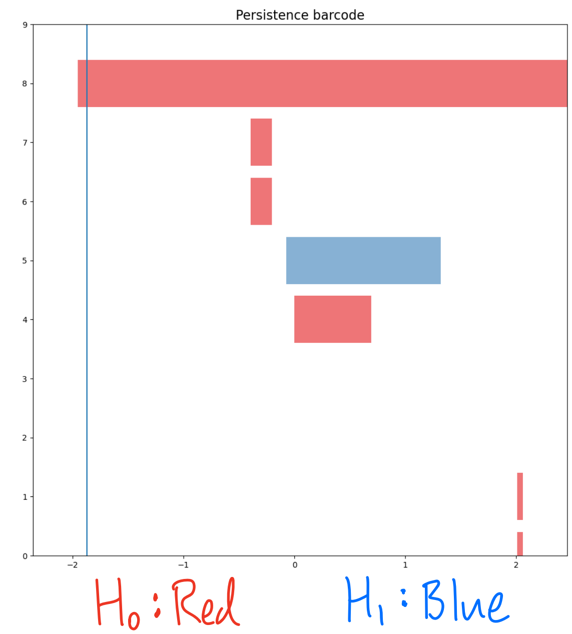
Input: Real-valued function on a space. Output: barcode.

Persistent homology

Sublevelset persistence



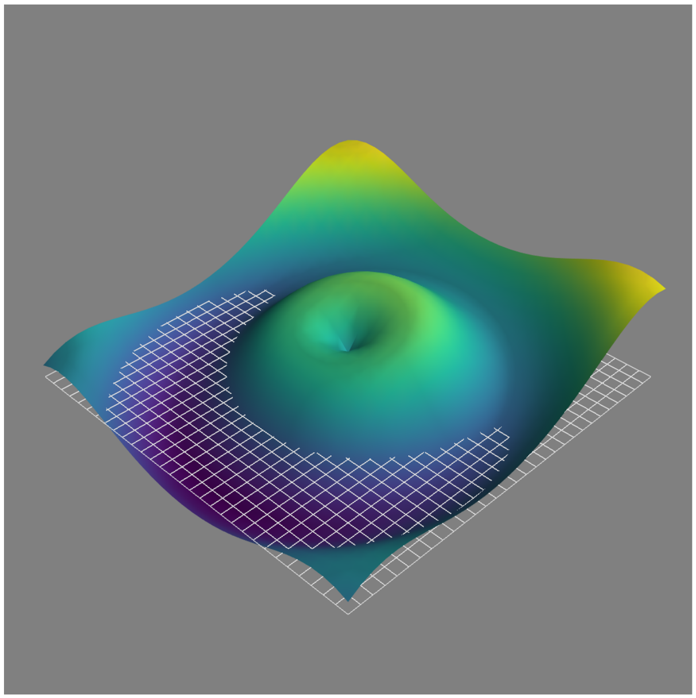


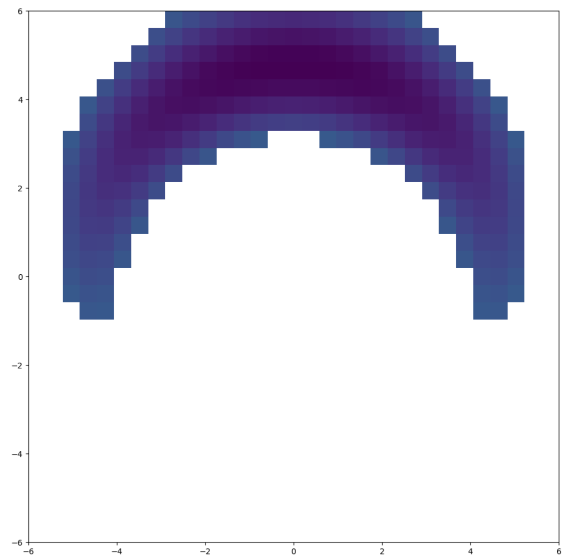


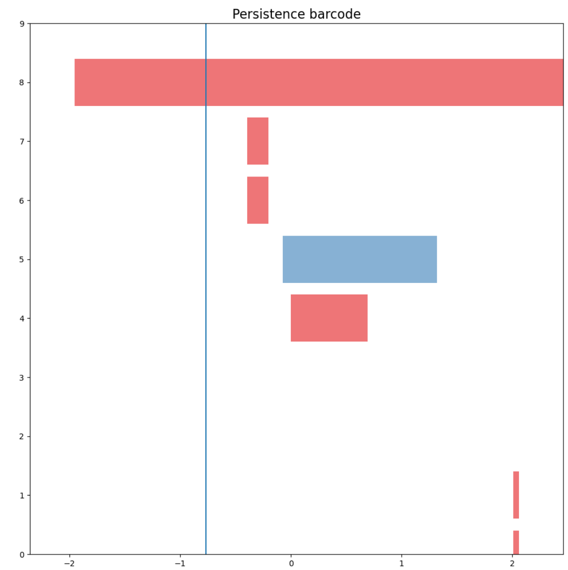
- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology

Sublevelset persistence





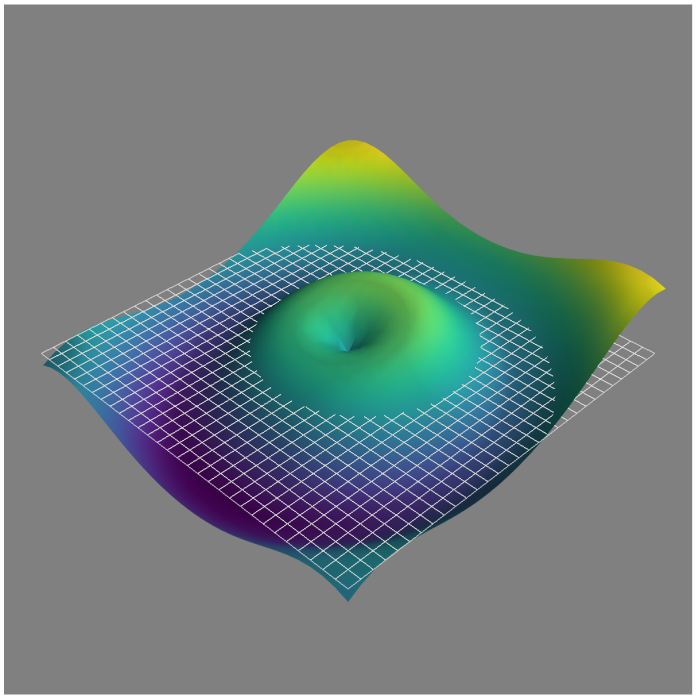


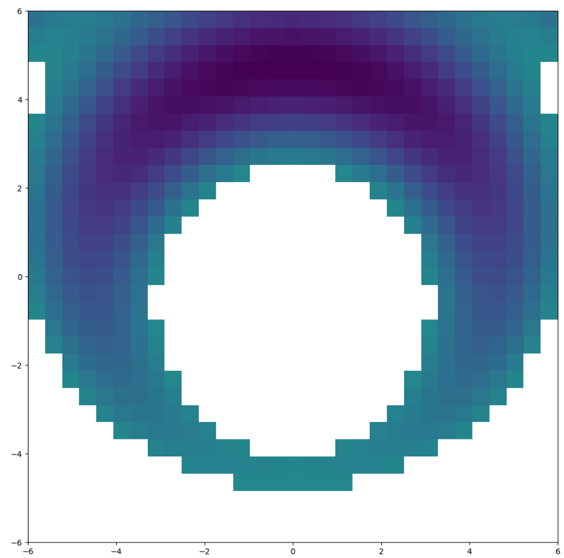
H_0: Red H_1: Blue

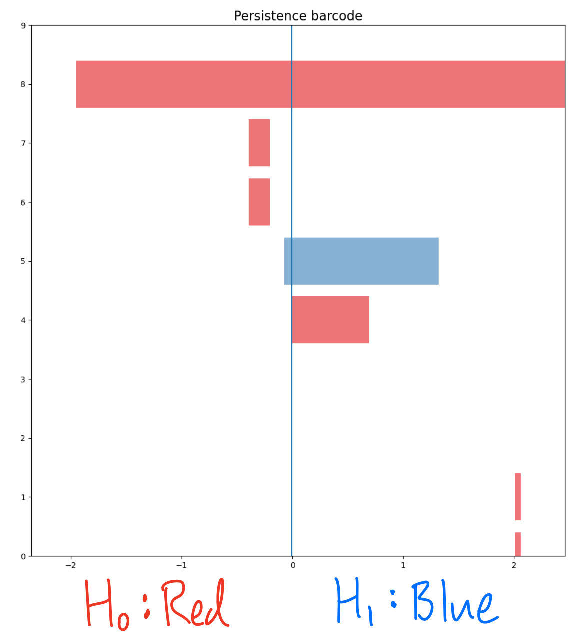
- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology

Sublevelset persistence



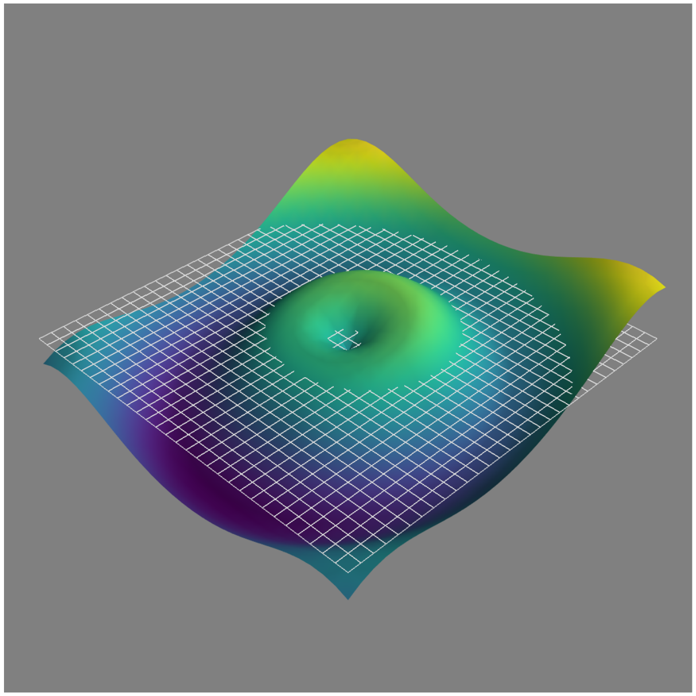


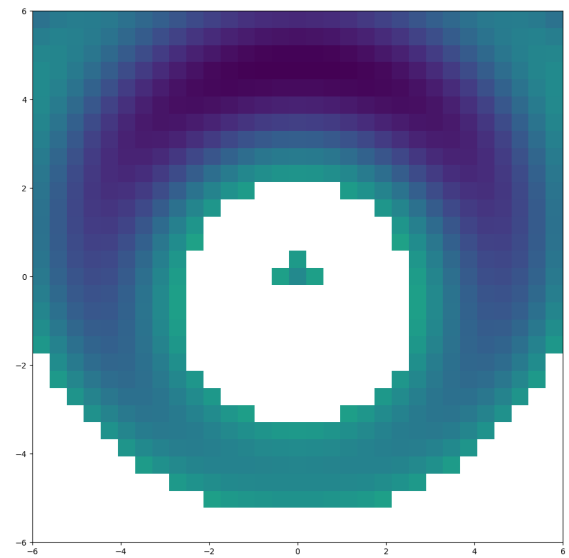


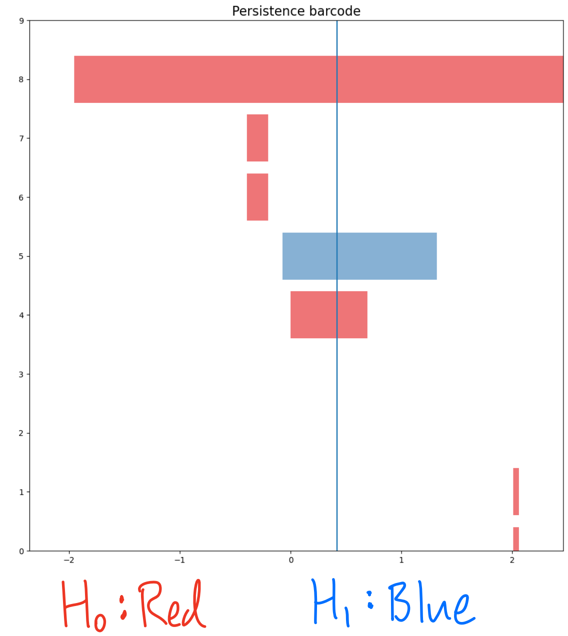
- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology

Sublevelset persistence



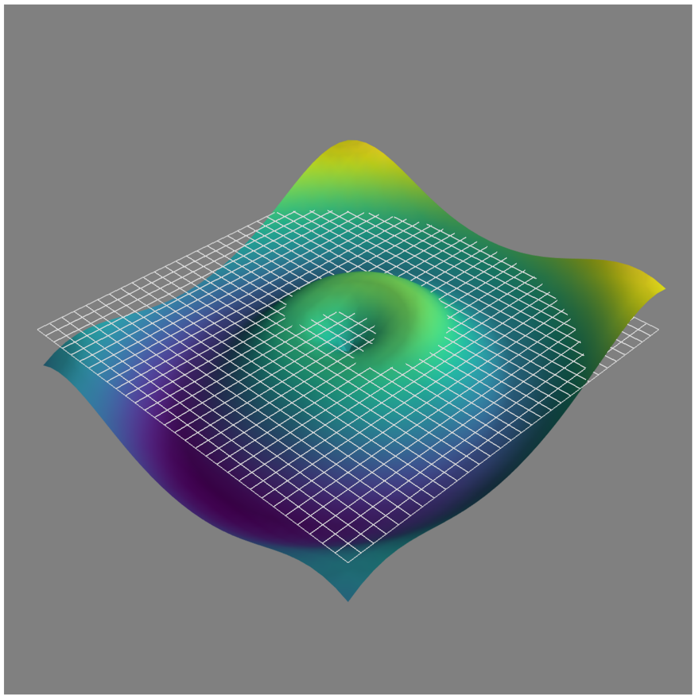


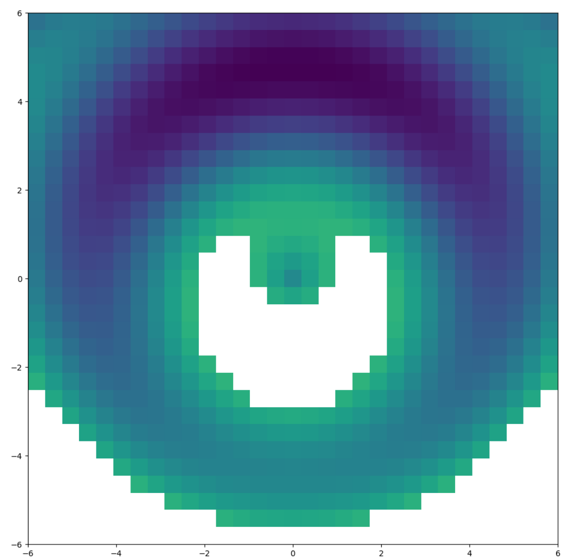


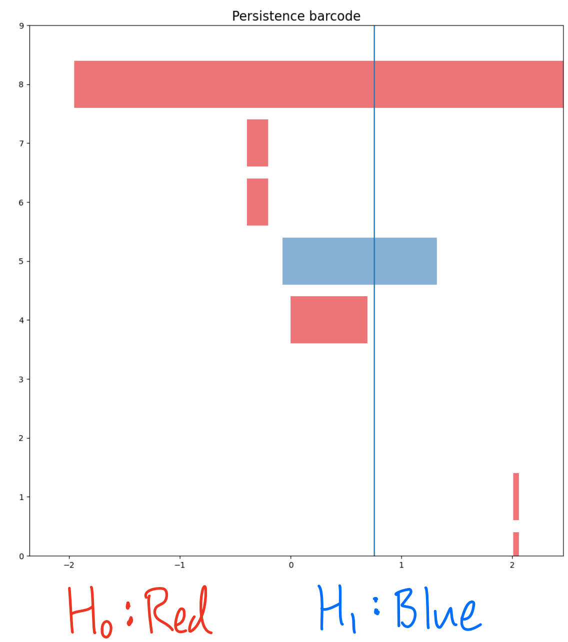
- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology

Sublevelset persistence



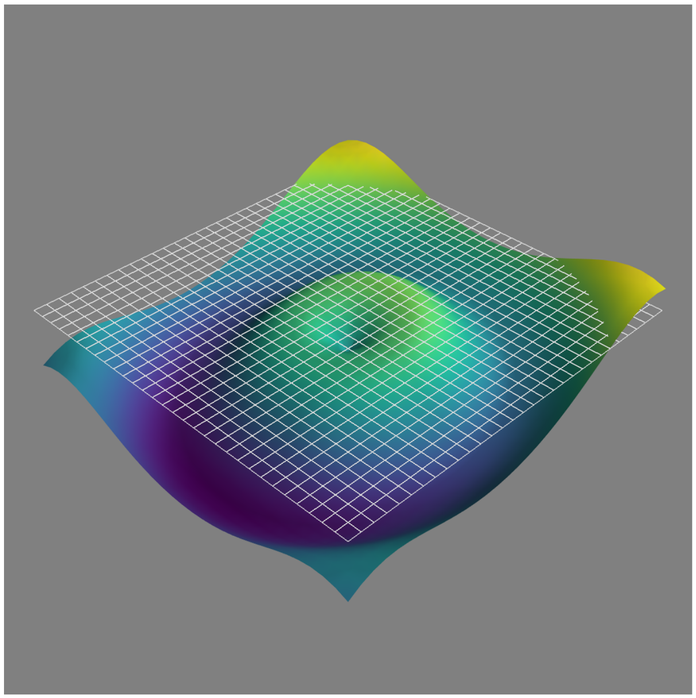


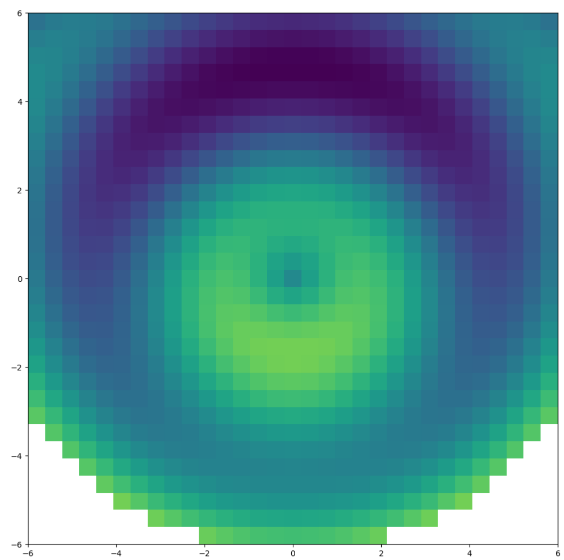


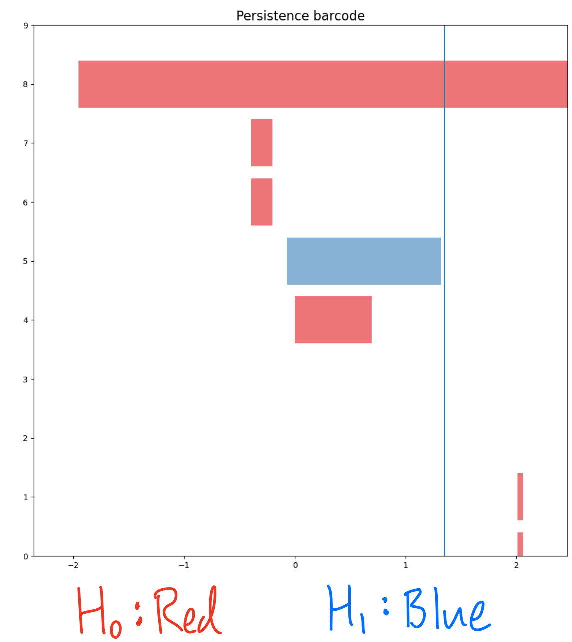
- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology

Sublevelset persistence



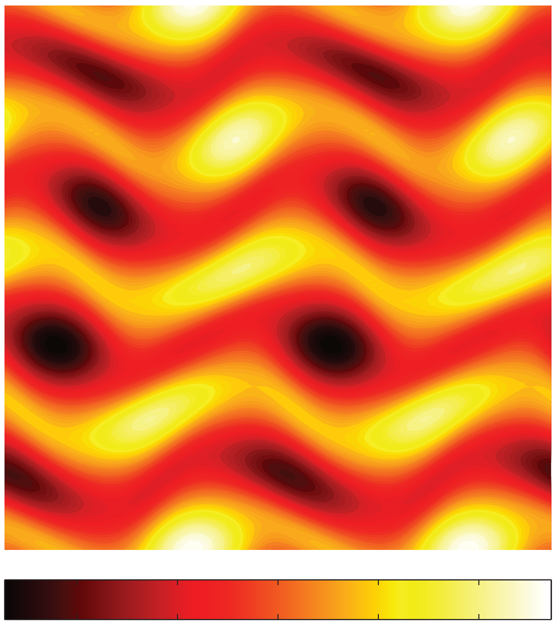




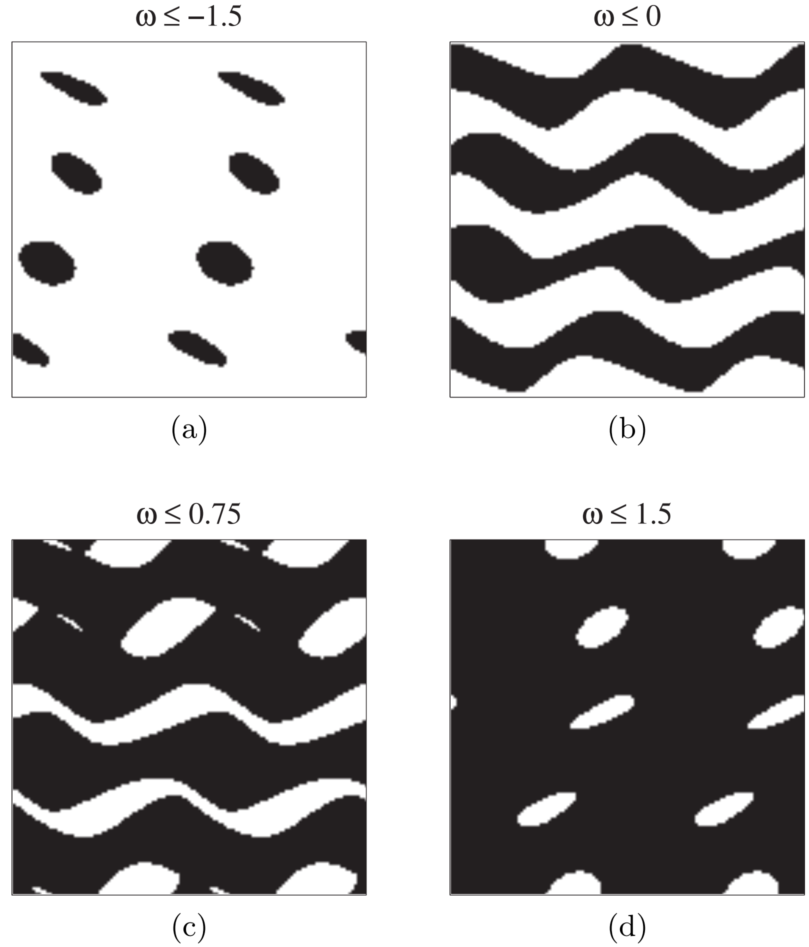
- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology

Sublevelset persistence



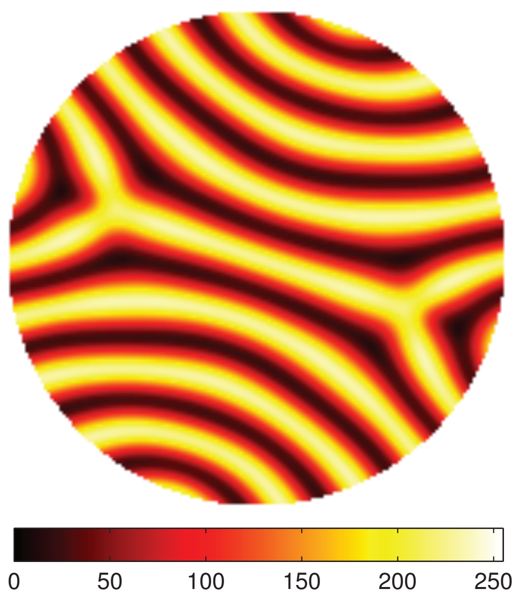
Analysis of Kolmogorov flow and Rayleigh–Bénard convection using persistent homology by Miroslav Kramár, Rachel Levanger, Jeffrey Tithof, Balachandra Suri, Mu Xu, Mark Paul, Michael F Schatz, Konstantin Mischaikow



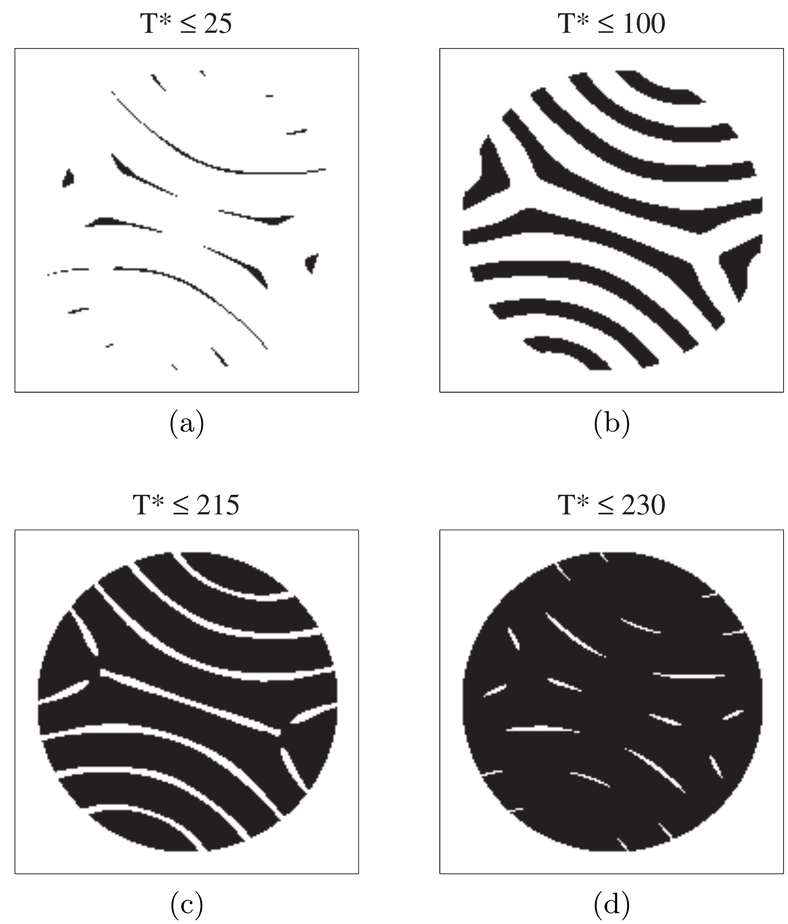
- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology

Sublevelset persistence



Analysis of Kolmogorov flow and Rayleigh–Bénard convection using persistent homology by Miroslav Kramár, Rachel Levanger, Jeffrey Tithof, Balachandra Suri, Mu Xu, Mark Paul, Michael F Schatz, Konstantin Mischaikow

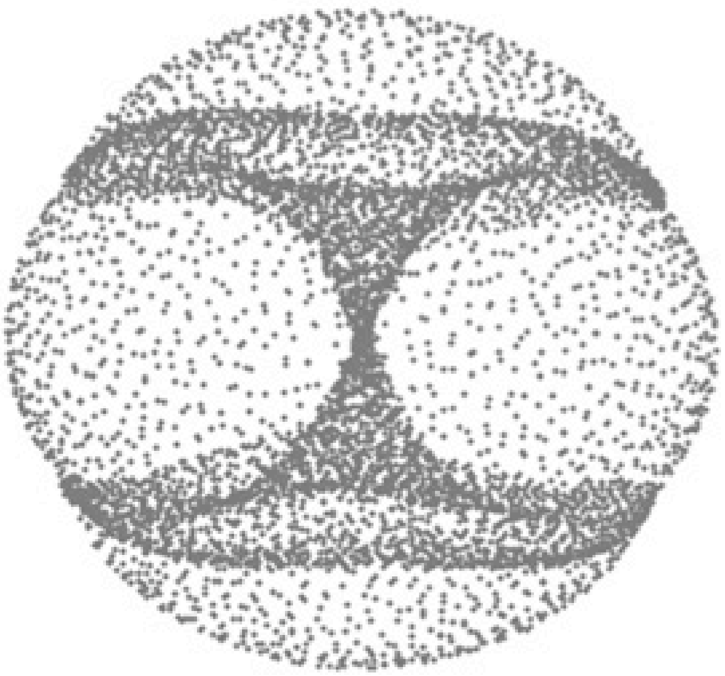


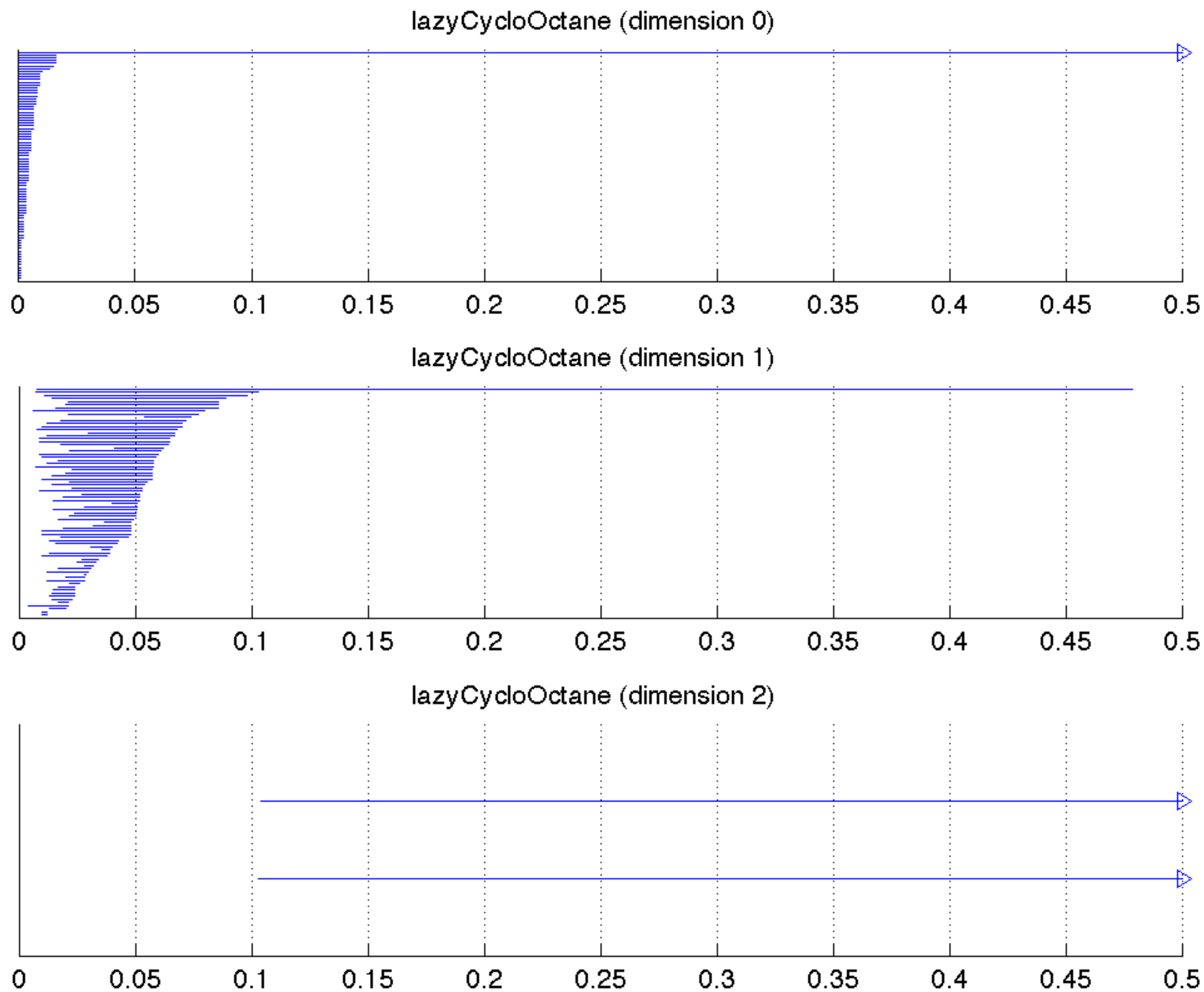
- Input: Increasing spaces. Output: barcode.
- Significant features persist.
- Cubic computation time in the number of simplices.

Persistent homology applied to data

Example: Cyclo-Octane (C_8H_{16}) data

1,000,000+ points in 24-dimensional space



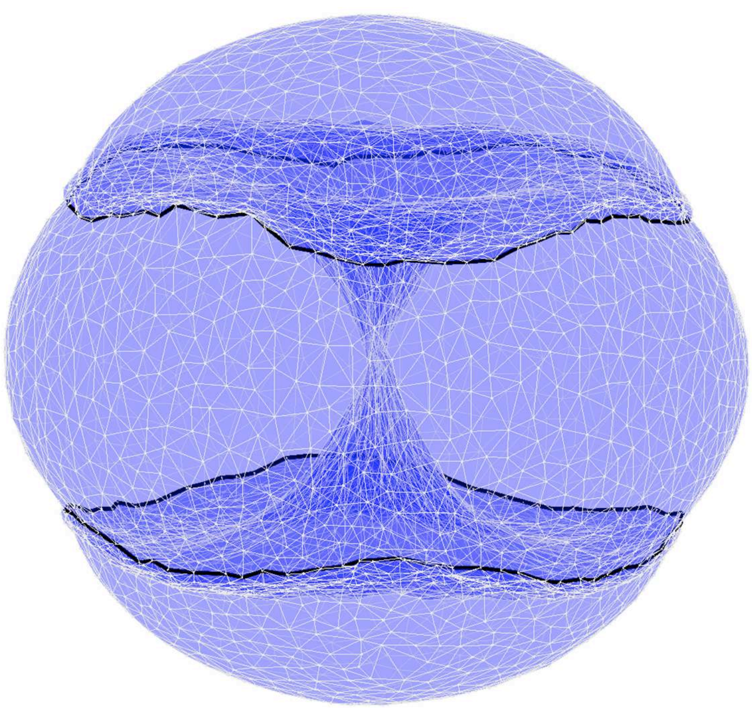


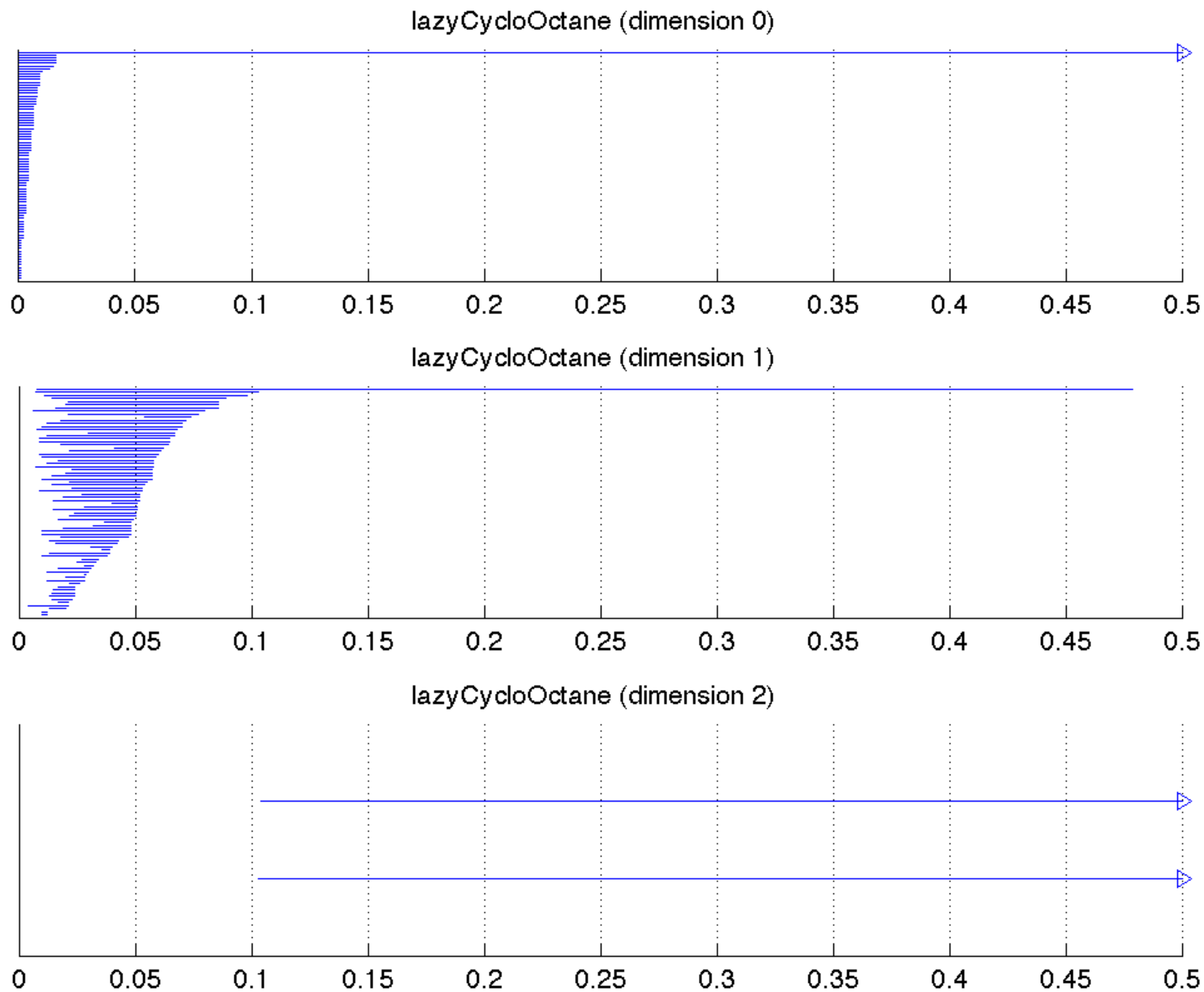
Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

Example: Cyclo-Octane (C_8H_{16}) data

1,000,000+ points in 24-dimensional space

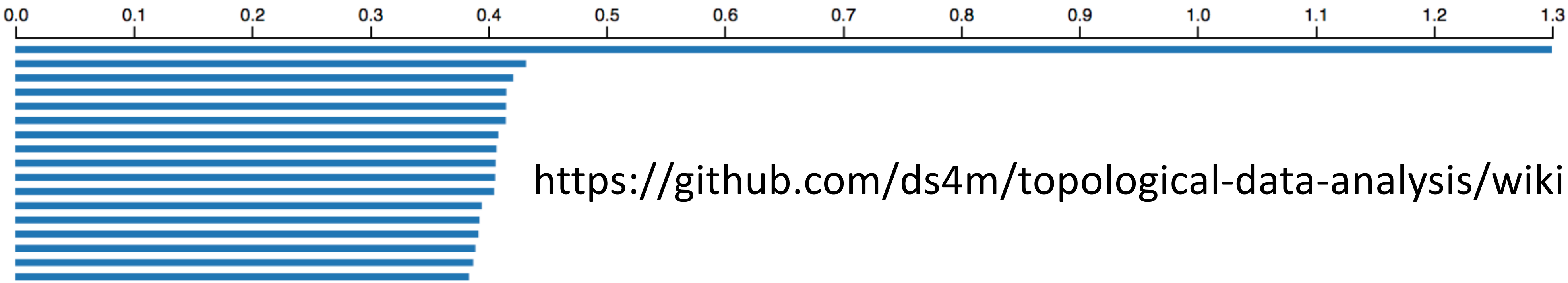




Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

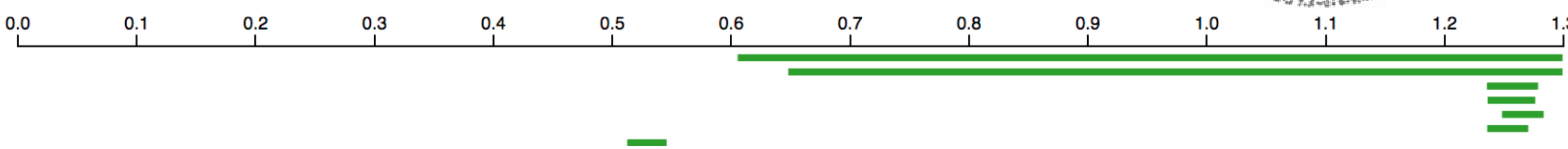
Persistence intervals in dimension 0:



https://github.com/ds4m/topological-data-analysis/wiki

Persistence intervals in dimension 1:

Persistence intervals in dimension 2:

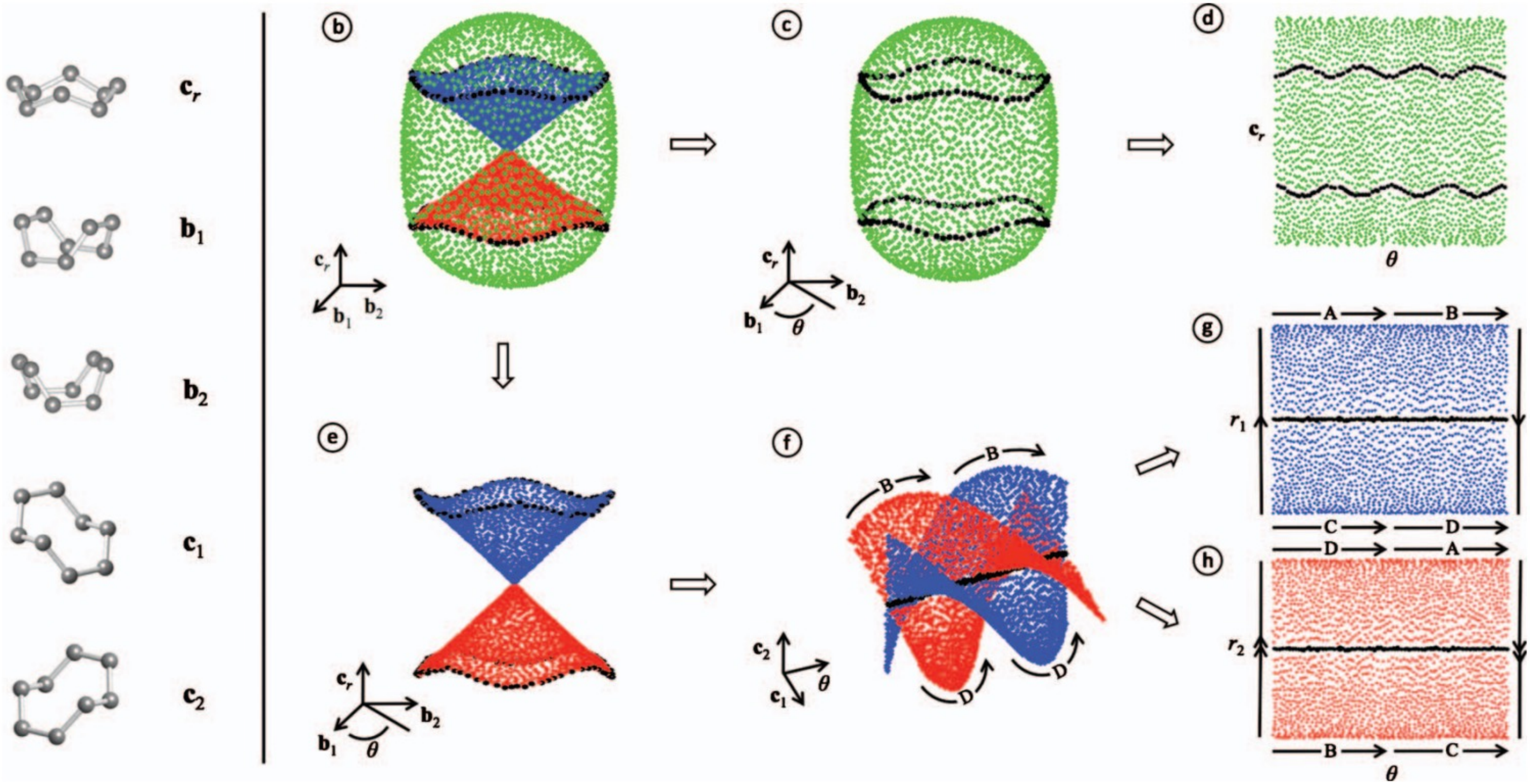


Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

Example: Cyclo-Octane (C_8H_{16}) data

1,000,000+ points in 24-dimensional space

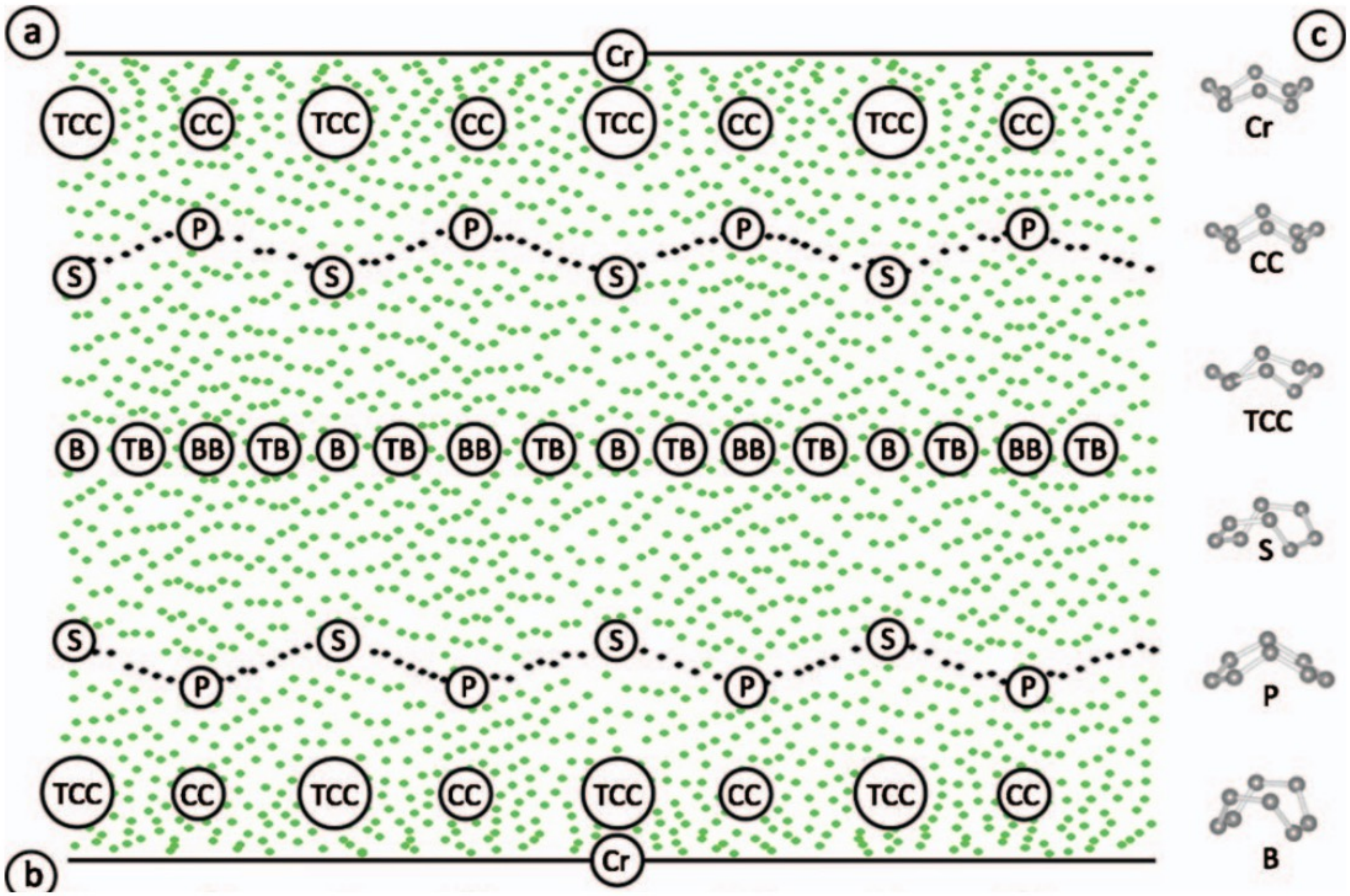


Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

Example: Cyclo-Octane (C_8H_{16}) data

1,000,000+ points in 24-dimensional space

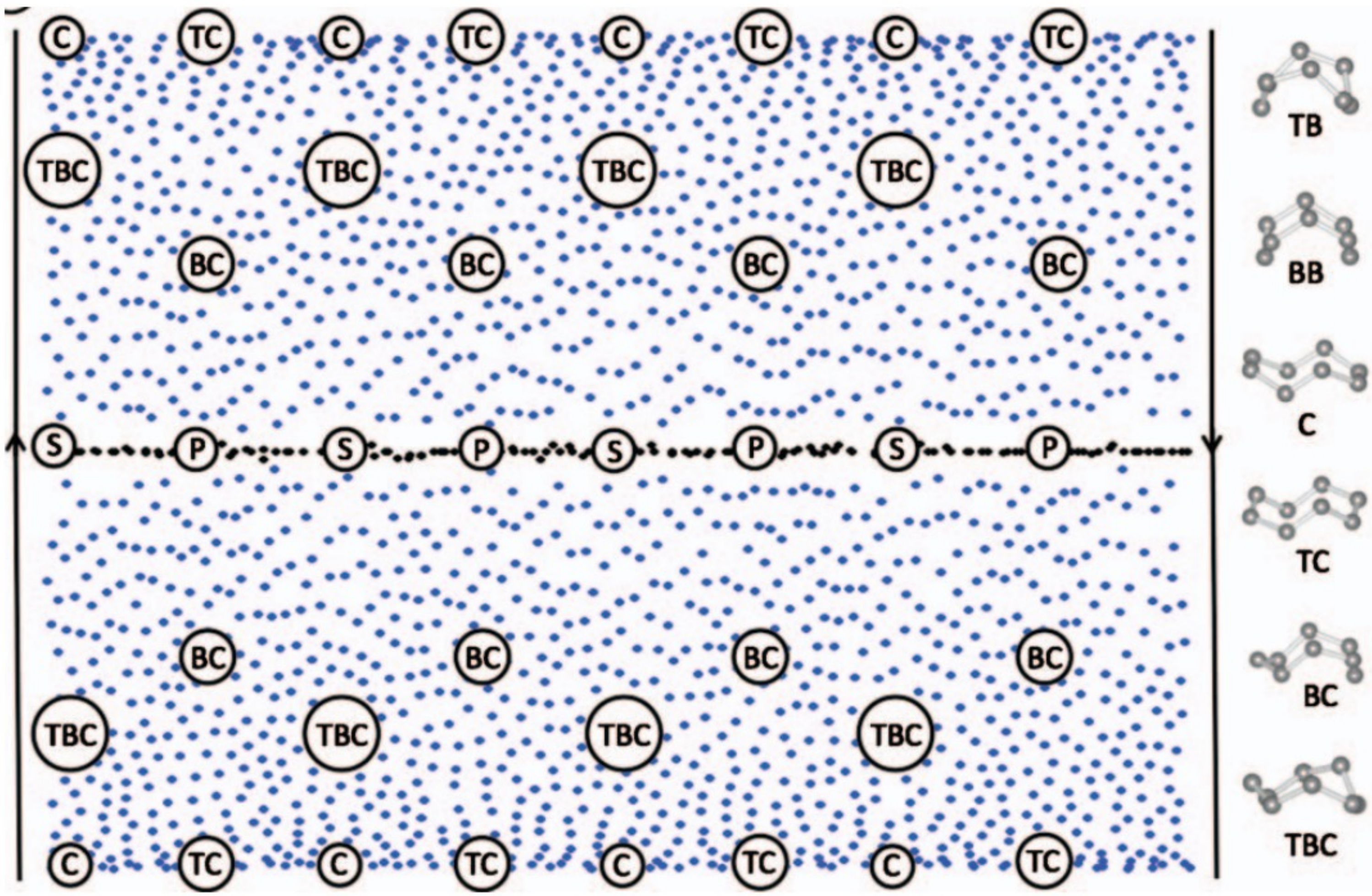


Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

Example: Cyclo-Octane (C_8H_{16}) data

1,000,000+ points in 24-dimensional space

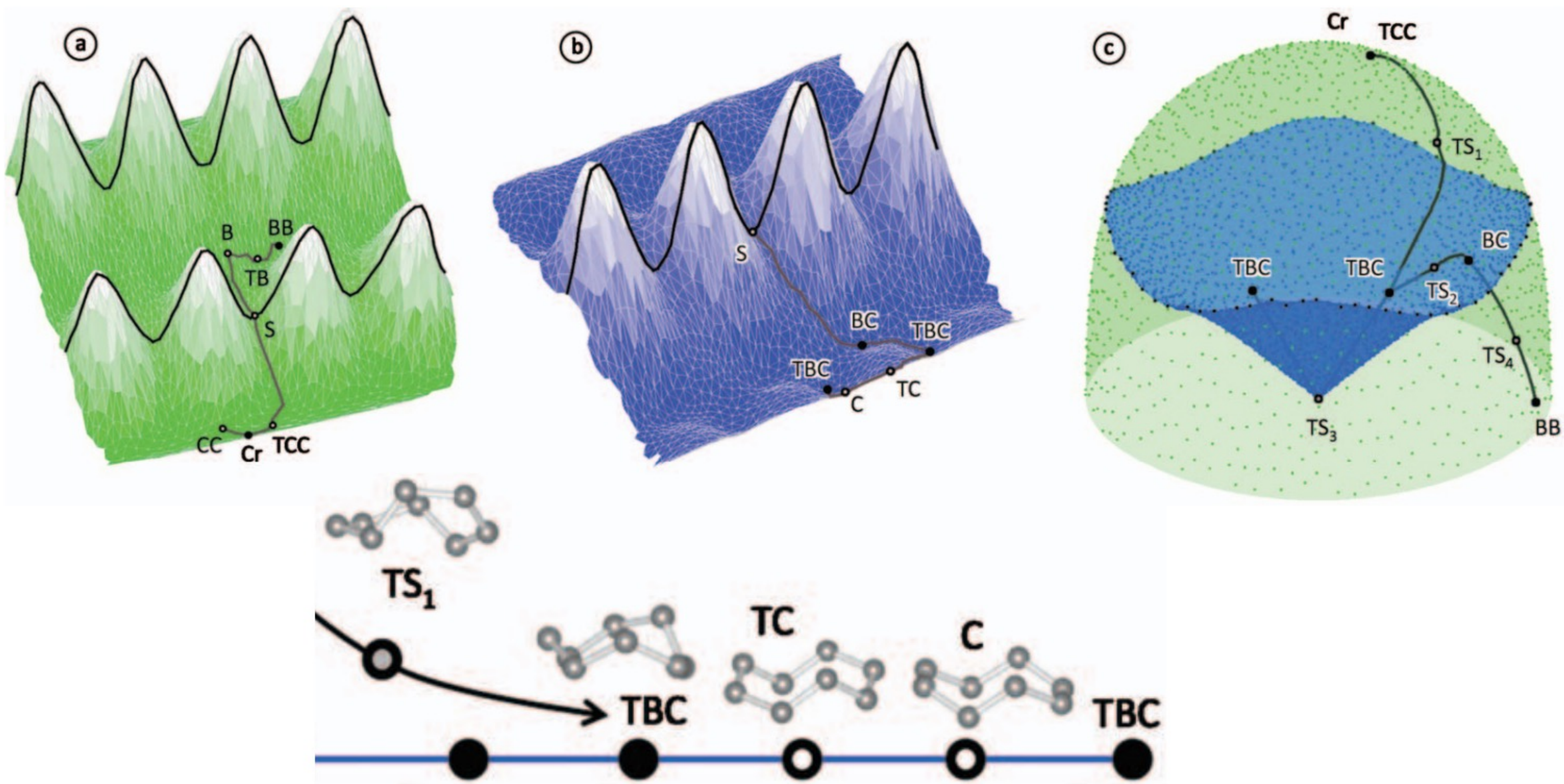


Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

Example: Cyclo-Octane (C_8H_{16}) data

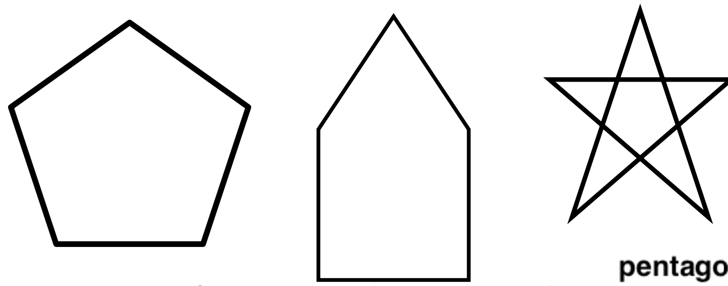
1,000,000+ points in 24-dimensional space



Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

Example: Equilateral pentagons in the plane



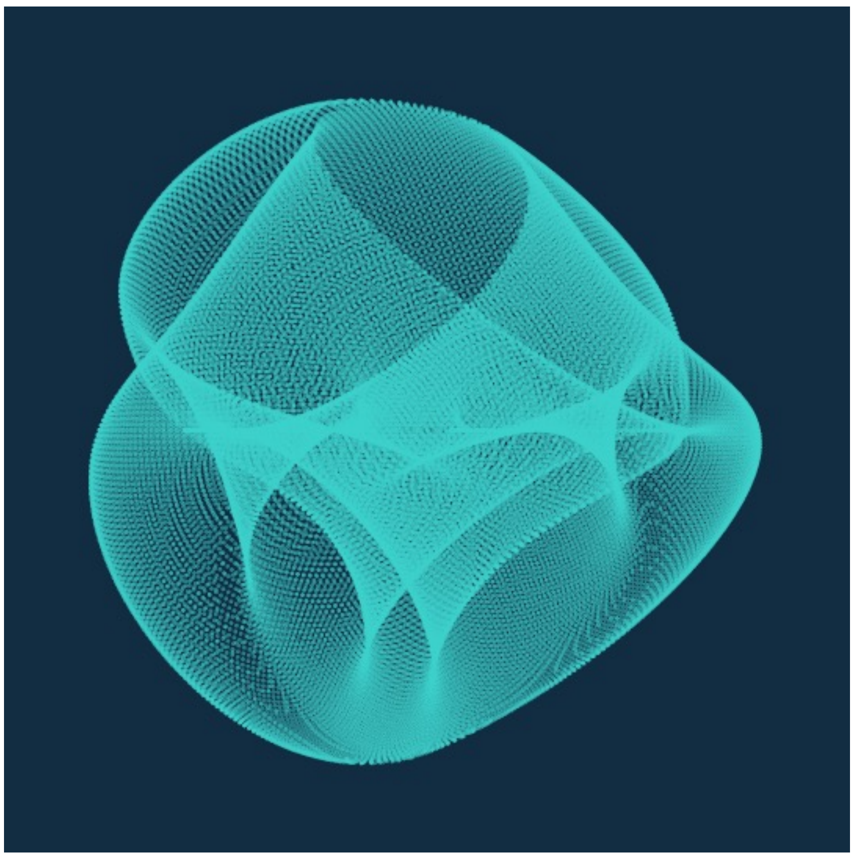
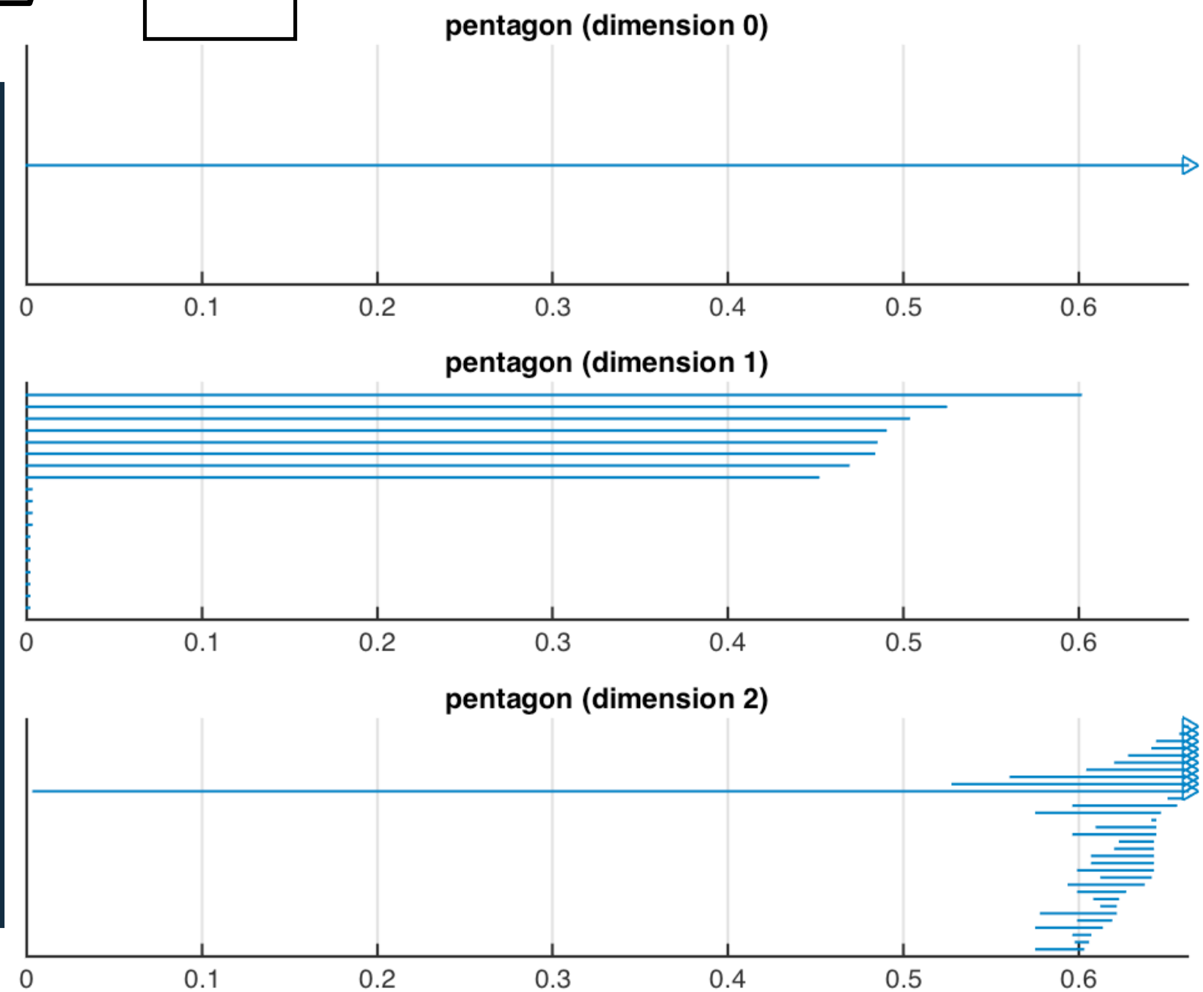
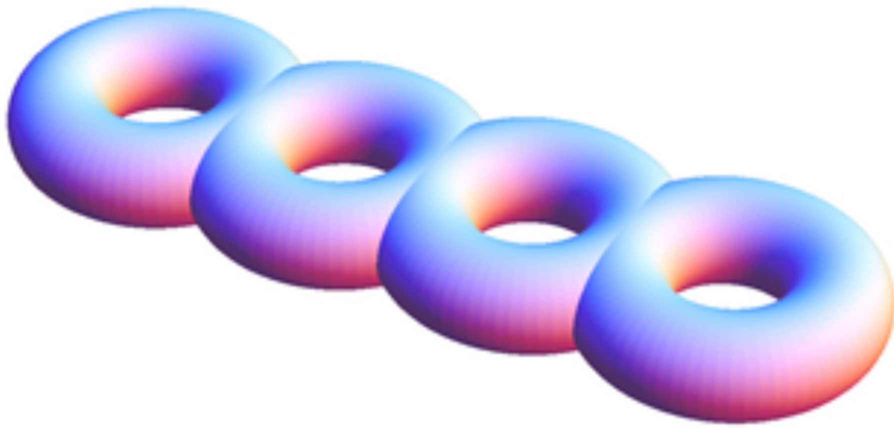


Image credit: Clayton Shonkwiler



Persistent homology applied to data



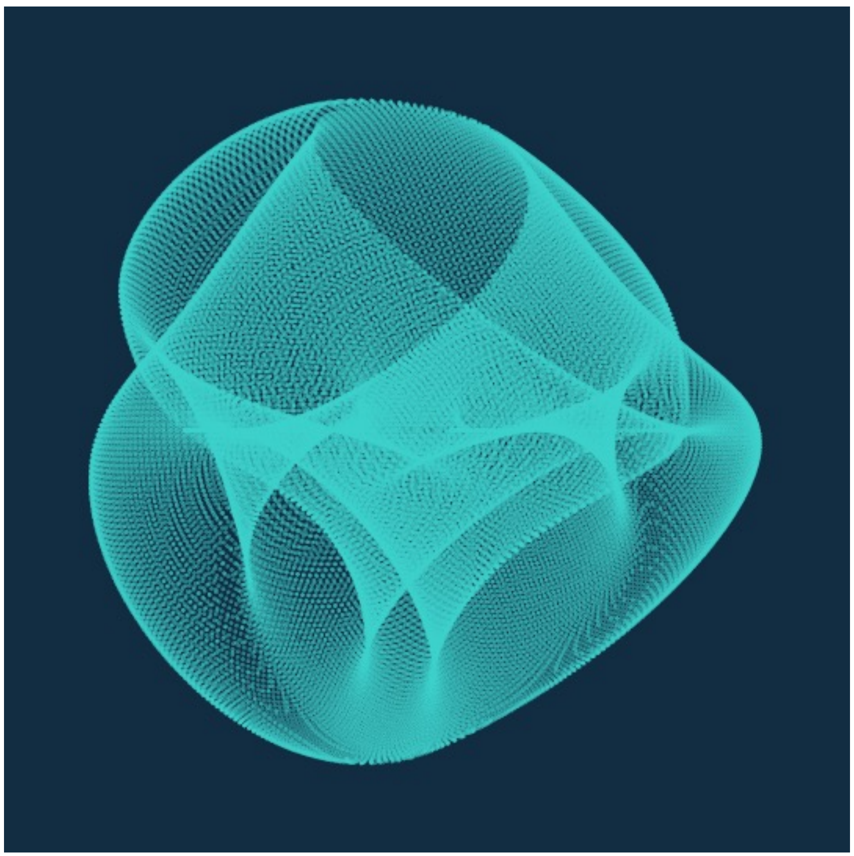
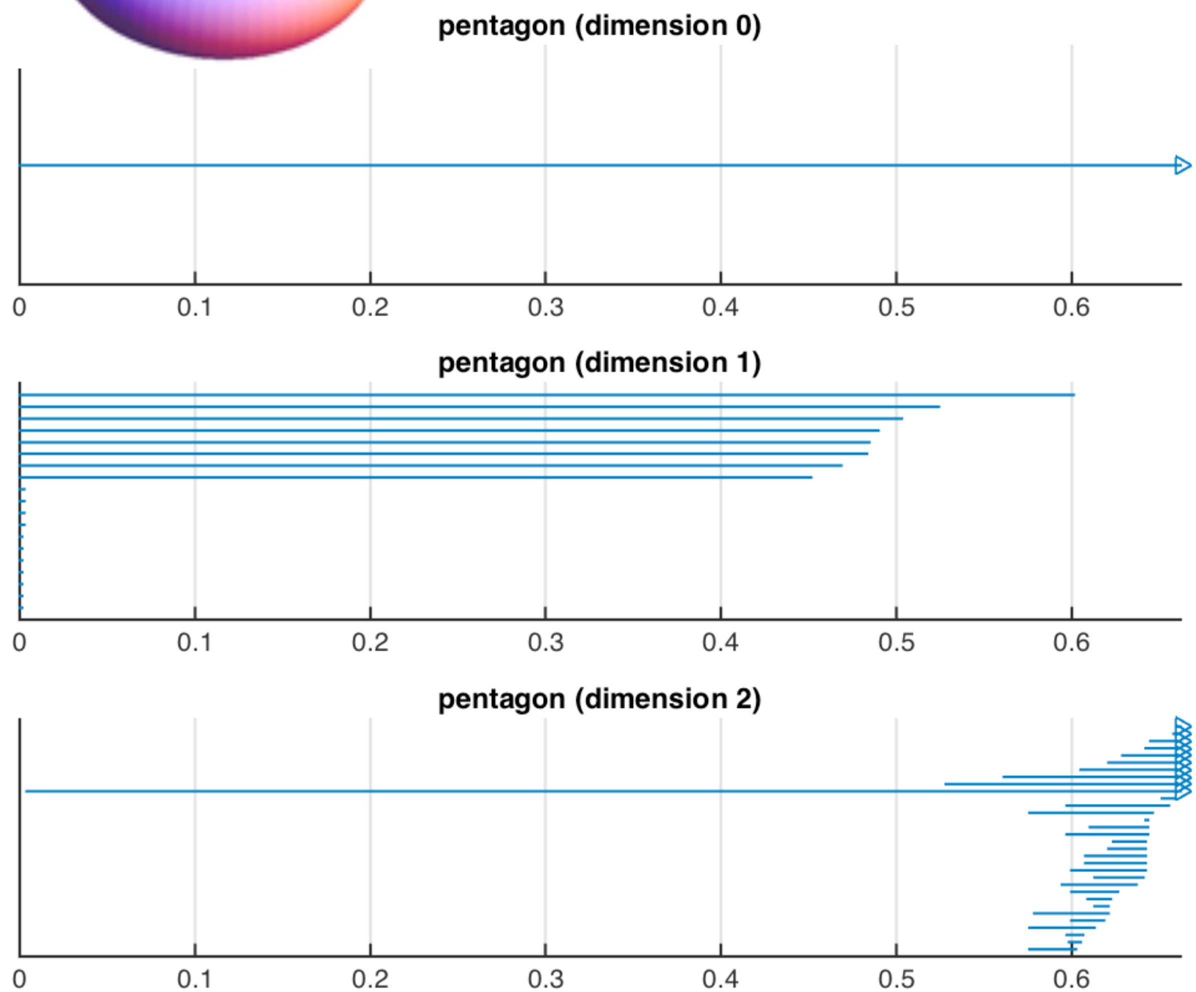


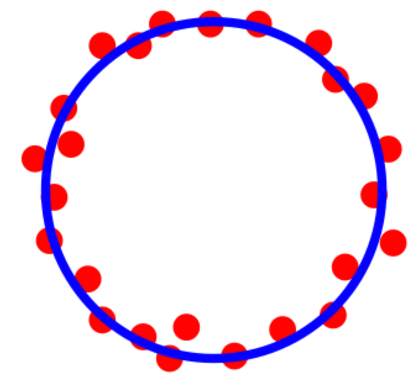
Image credit: Clayton Shonkwiler



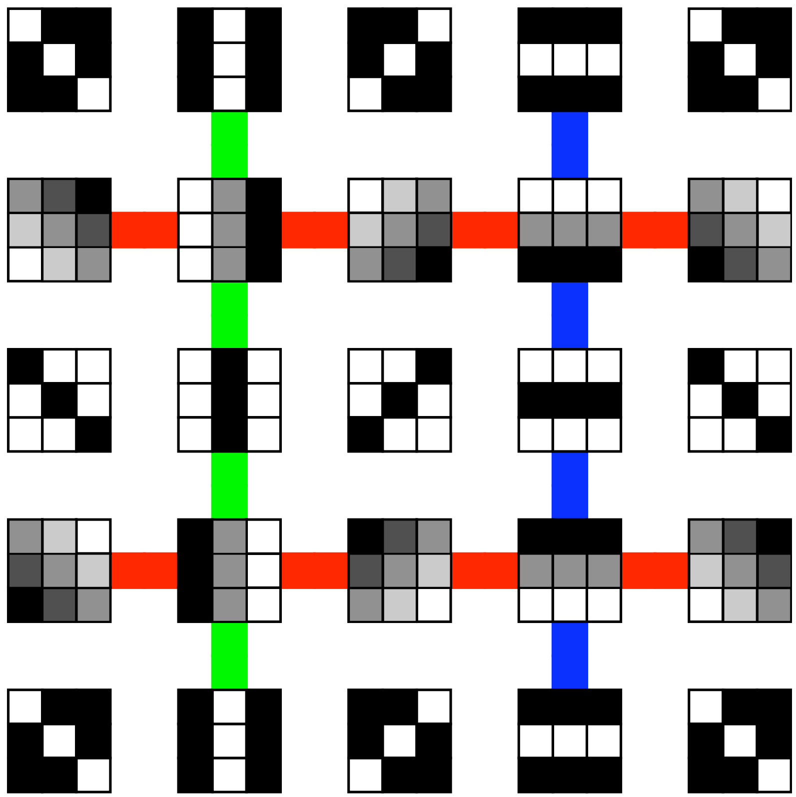
Persistent homology applied to data

- <u>Stability Theorem.</u>
  If X and Y are metric spaces, then

$$d_b(\mathrm{PH}(\check{\mathrm{C}}\mathrm{ech}(X)), \mathrm{PH}(\check{\mathrm{C}}\mathrm{ech}(Y))) \leq 2 d_{\mathrm{GH}}(X, Y)$$



Topology applied to image data



Persistent homology applied to data

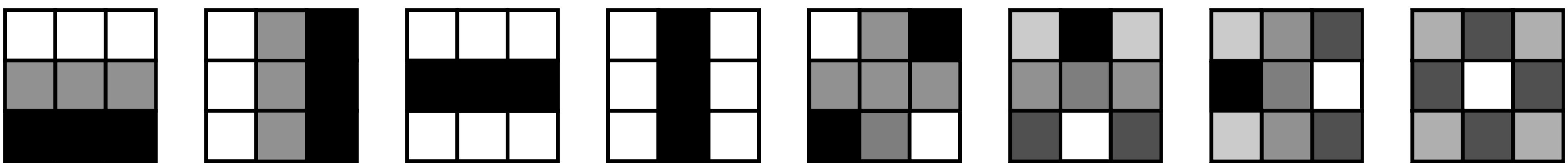
The receptive fields of cells in our primary visual cortex (V1) are related to the statistics natural images.

Independent component filters of natural images compared with simple cells in primary visual cortex by JH van Hateren and A van der Schaaf, 1997

Persistent homology applied to data

3x3 *high-contrast* patches from images

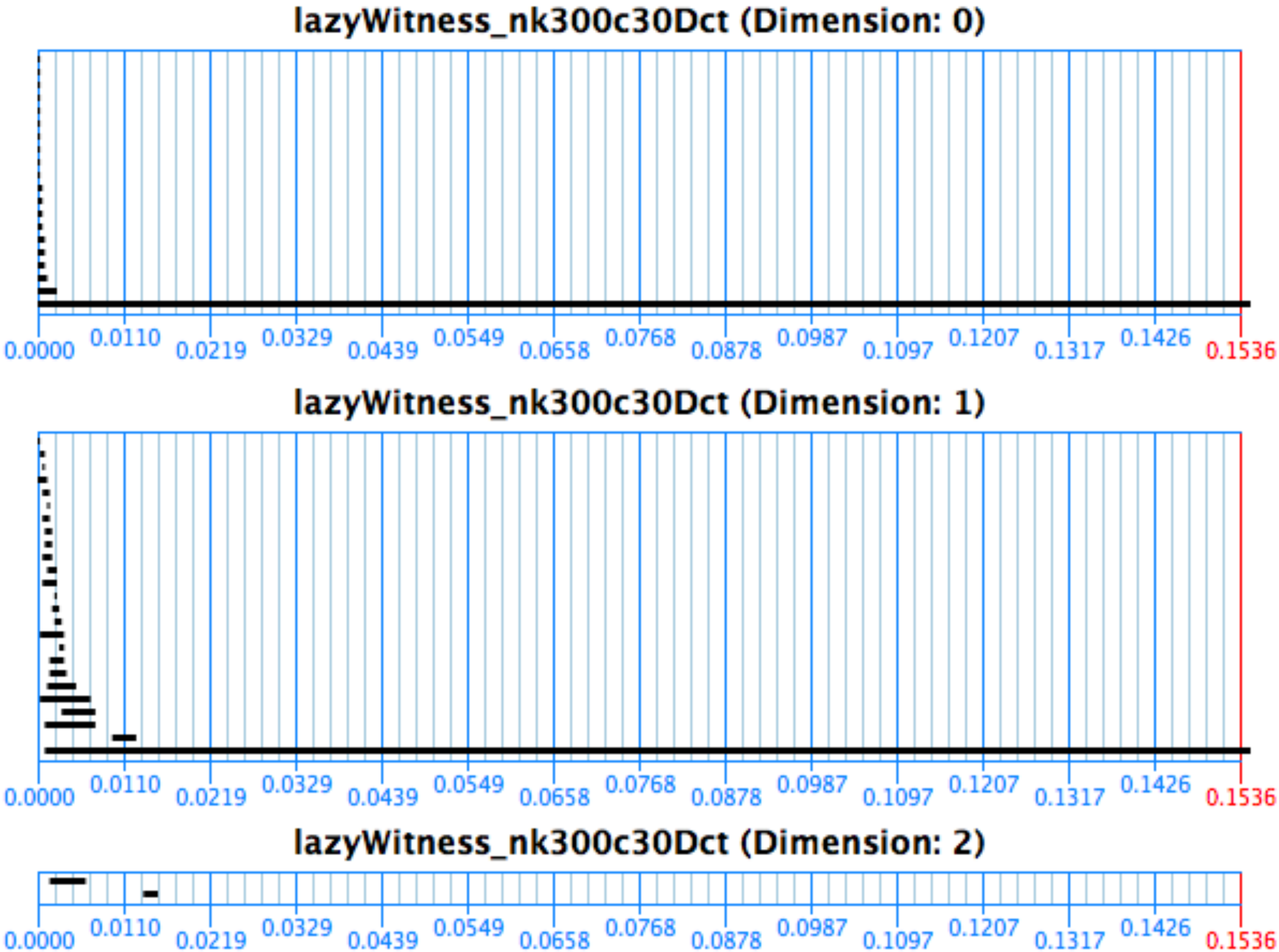
Points in 9-dimensional space, normalized to have average color gray and contrast norm one (on 7-sphere).



On the Local Behavior of Spaces of Natural Images by Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian, 2008.

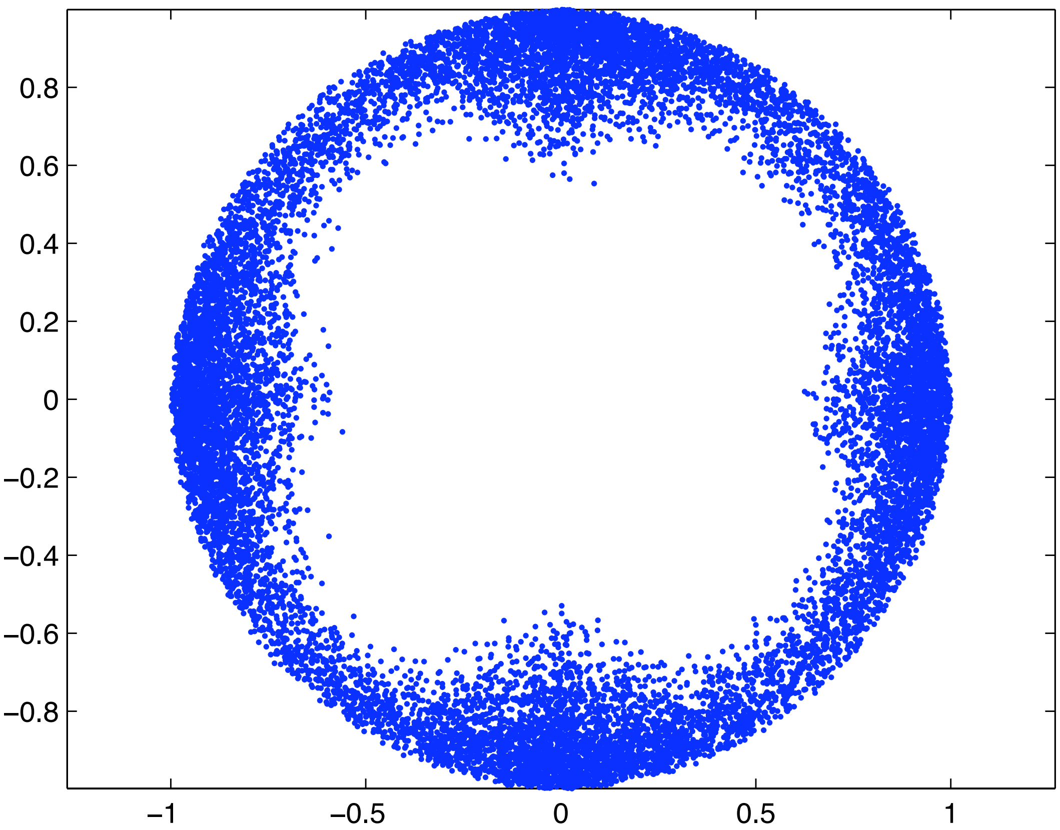
Persistent homology applied to data

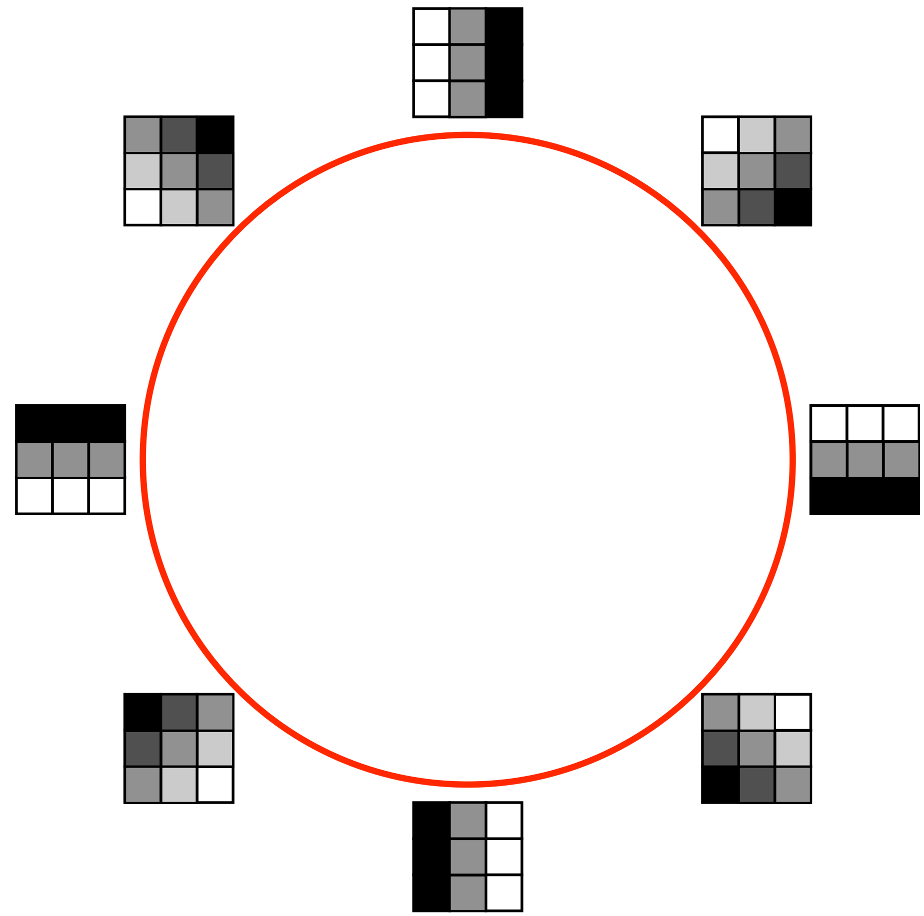
1. Densest patches according to a global estimate



Persistent homology applied to data

1. Densest patches according to a global estimate

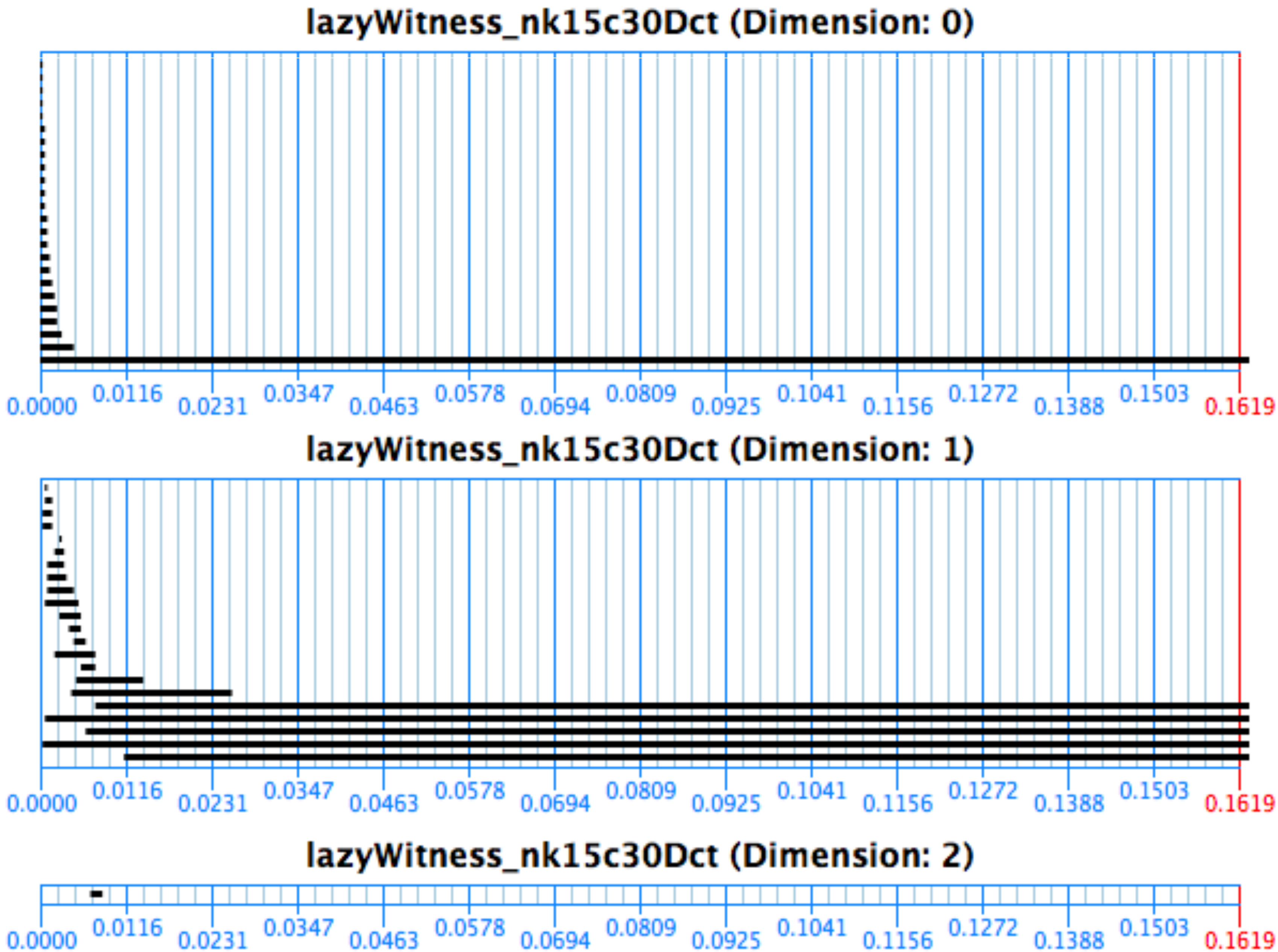




Interpretation: nature prefers linearity

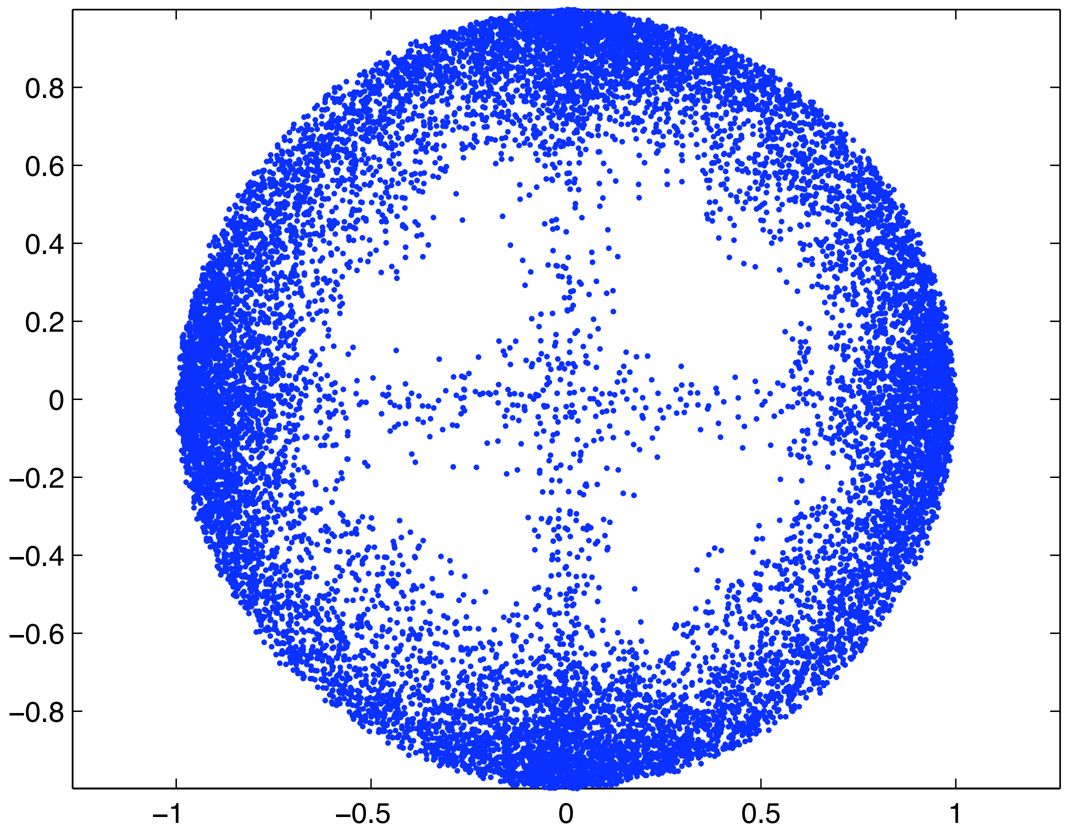
Persistent homology applied to data

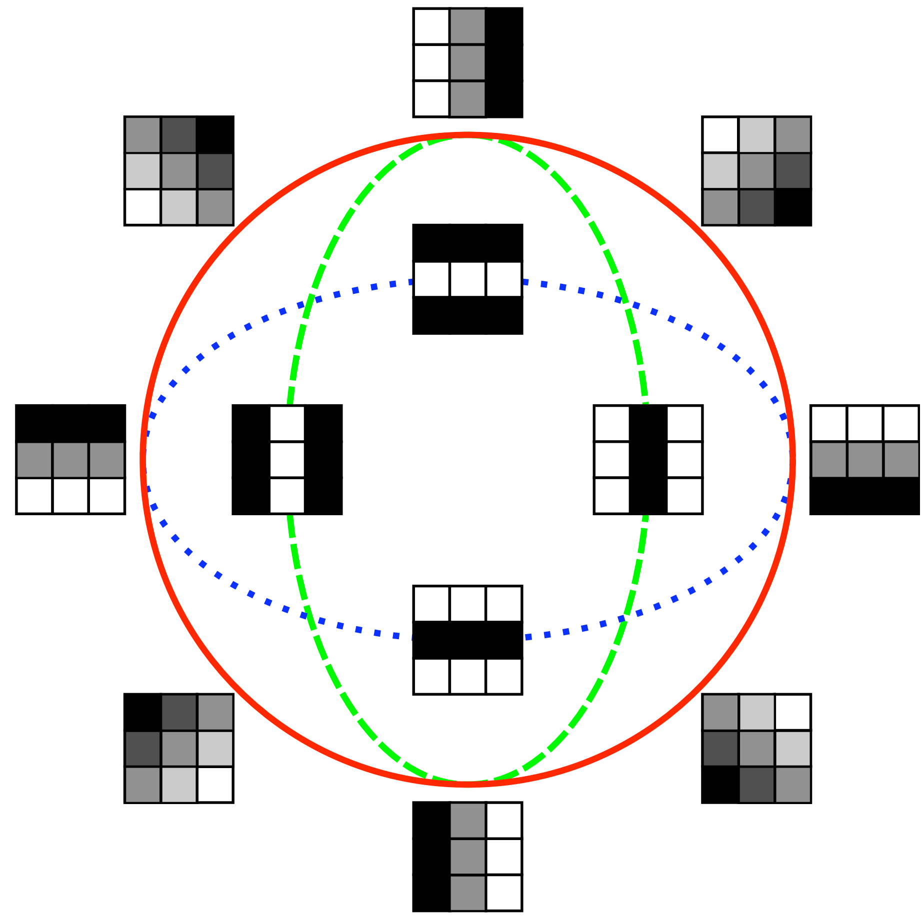
2. Densest patches according to an intermediate estimate



Persistent homology applied to data

2. Densest patches according to an intermediate estimate

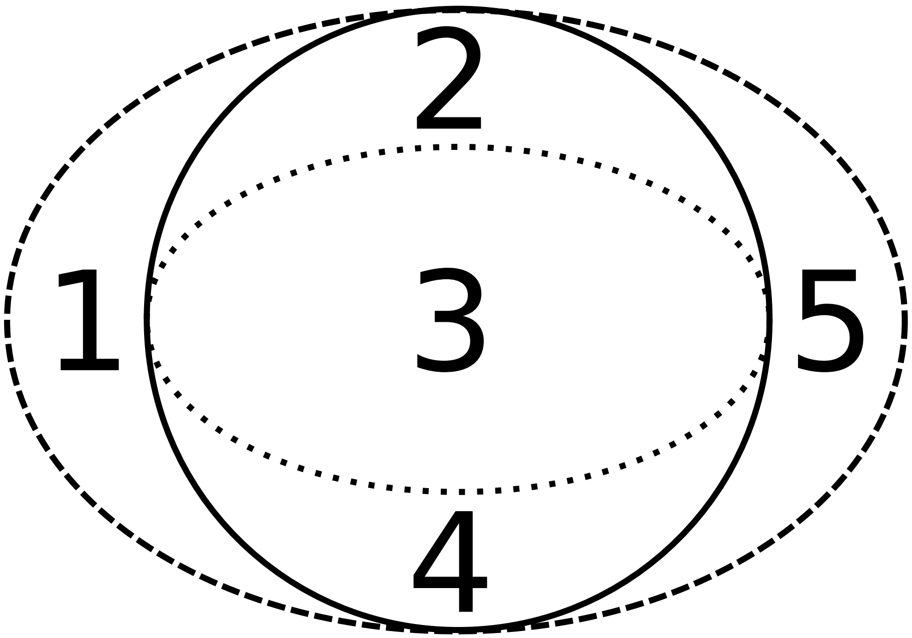


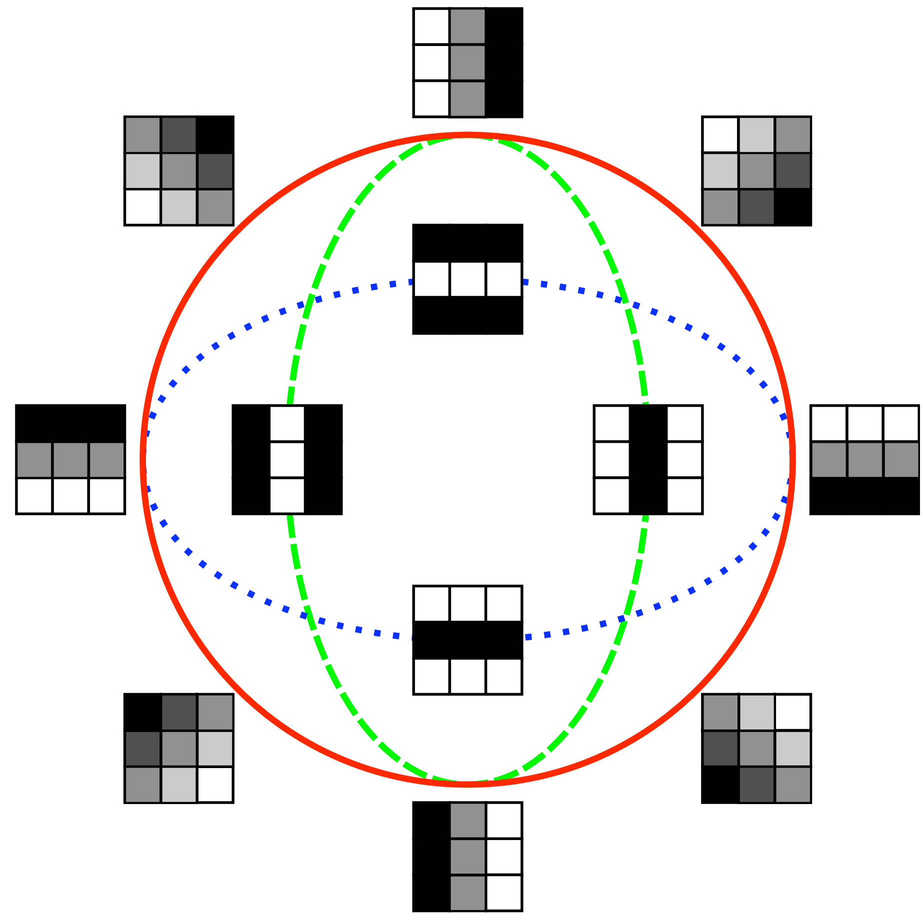


Interpretation: nature prefers horizontal and vertical directions

Persistent homology applied to data

2. Densest patches according to an intermediate estimate

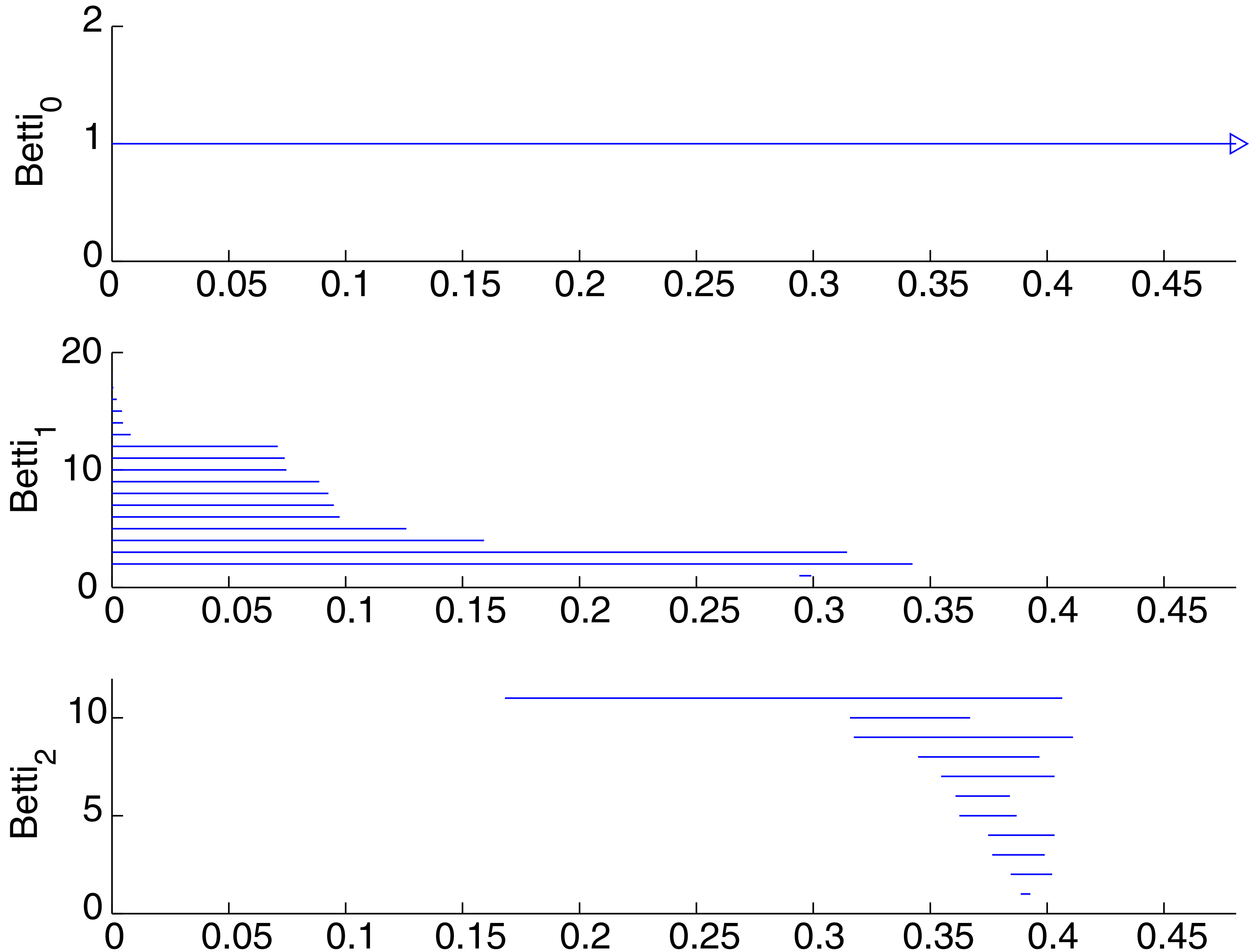




Interpretation: nature prefers horizontal and vertical directions

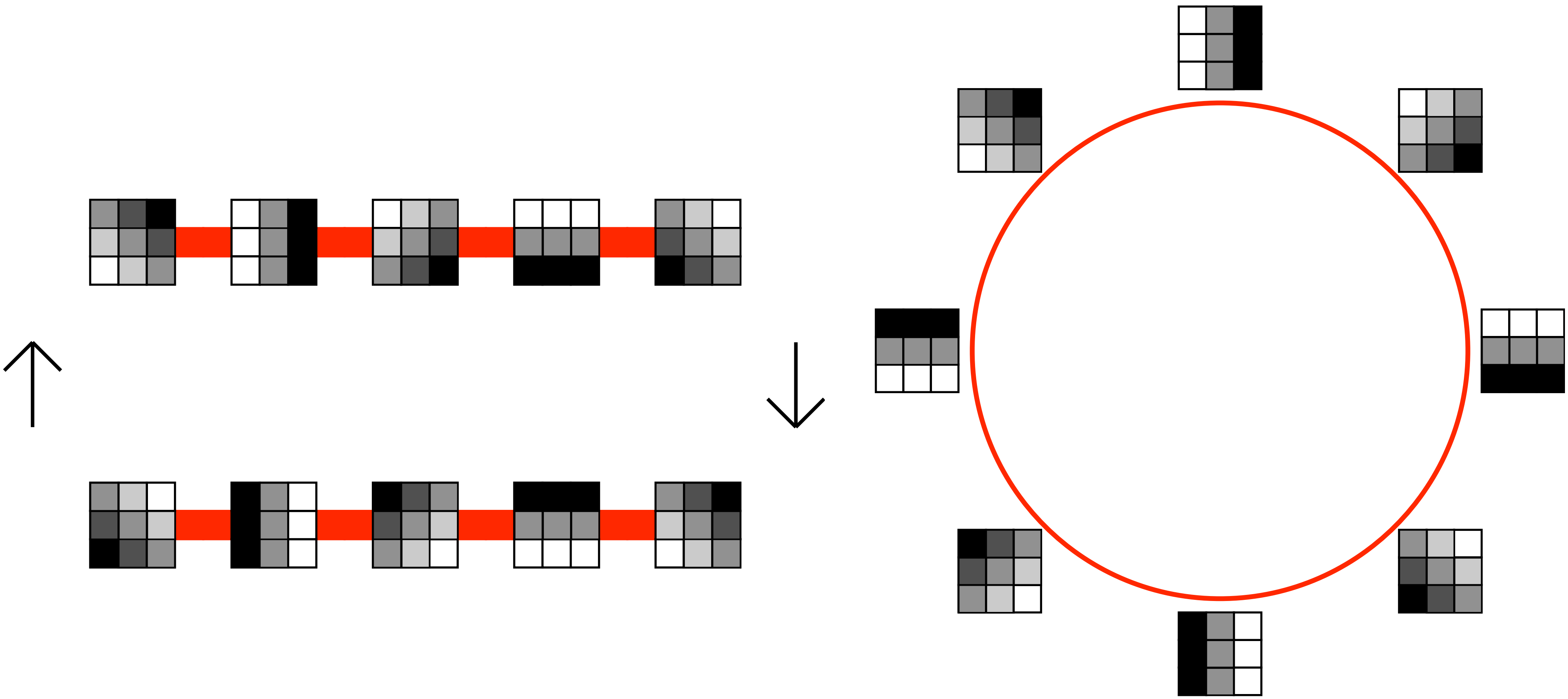
Persistent homology applied to data

3. Densest patches according to a local estimate



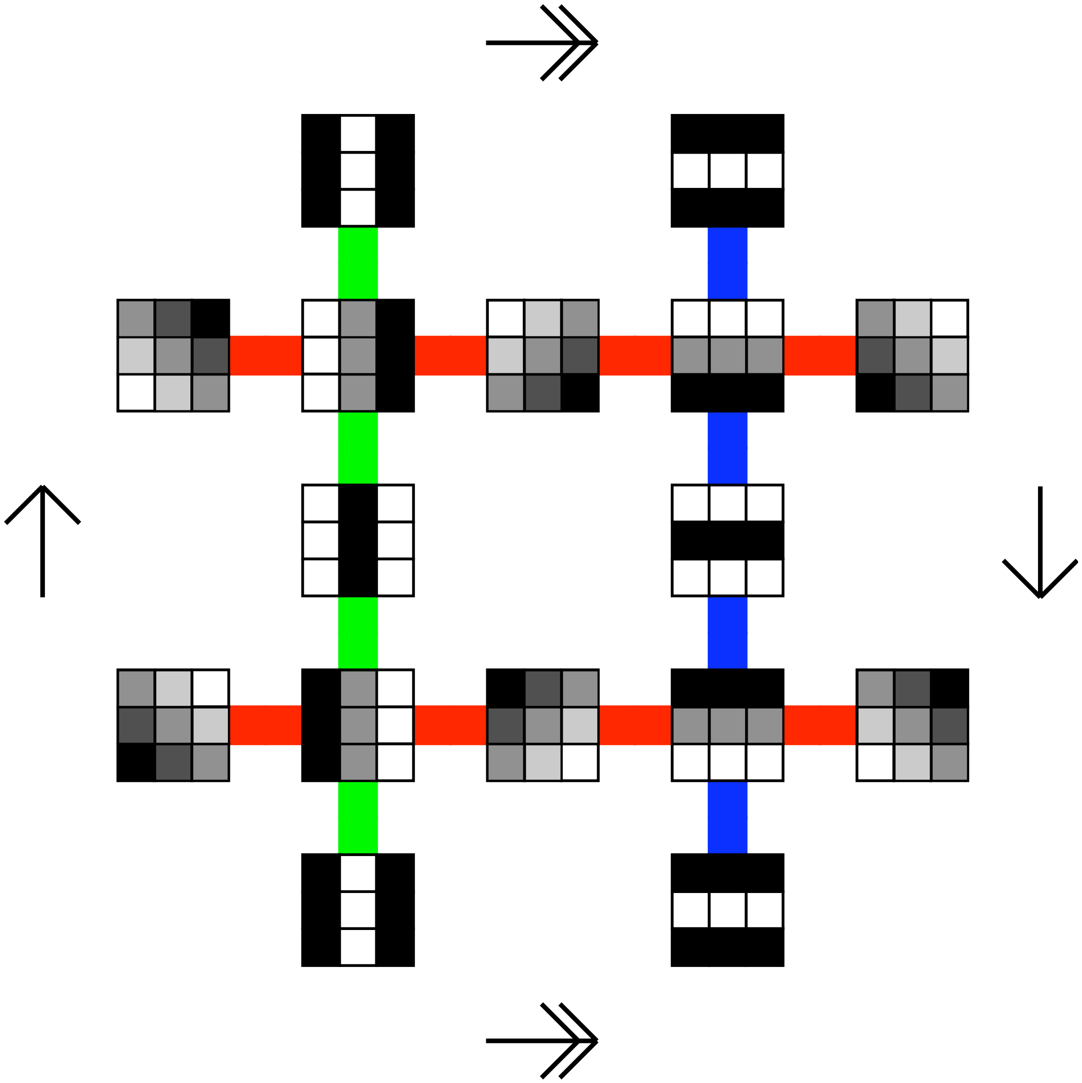
Persistent homology applied to data

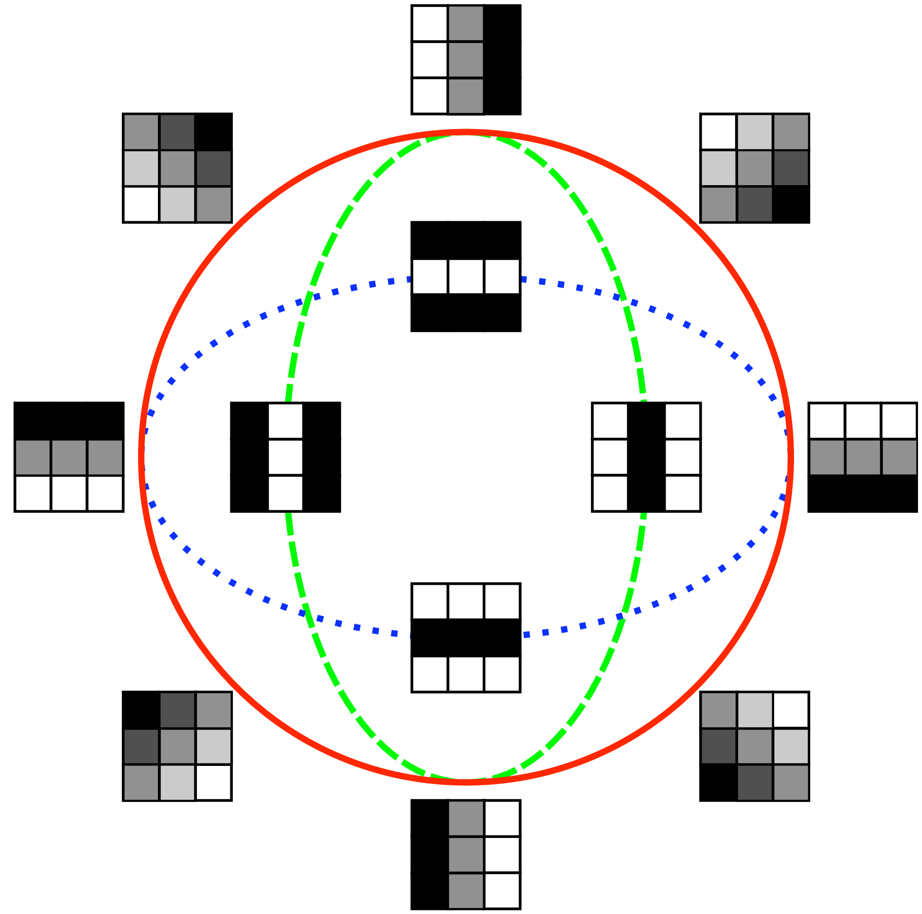
3. Densest patches according to a local estimate



Persistent homology applied to data

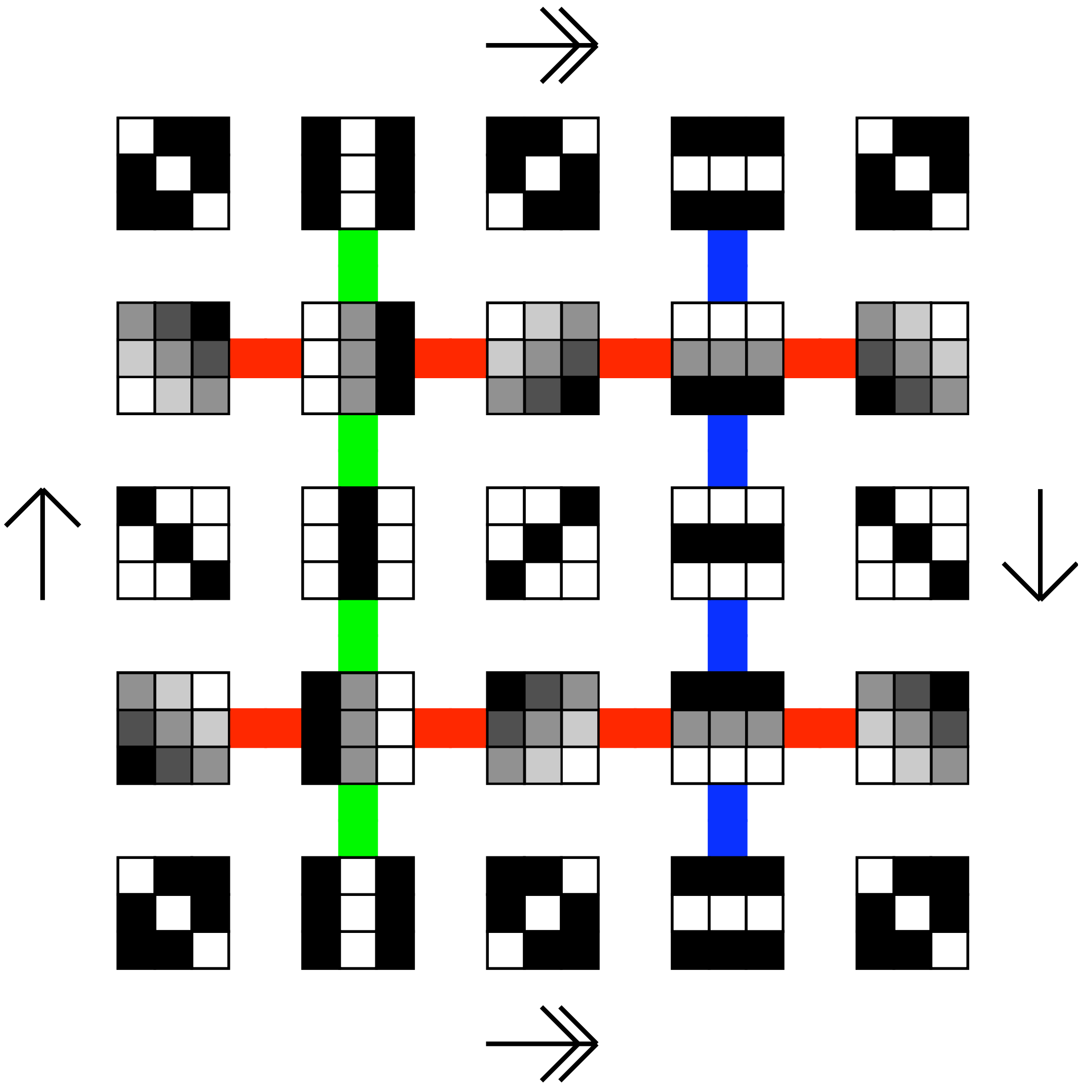
3. Densest patches according to a local estimate





Persistent homology applied to data

3. Densest patches according to a local estimate



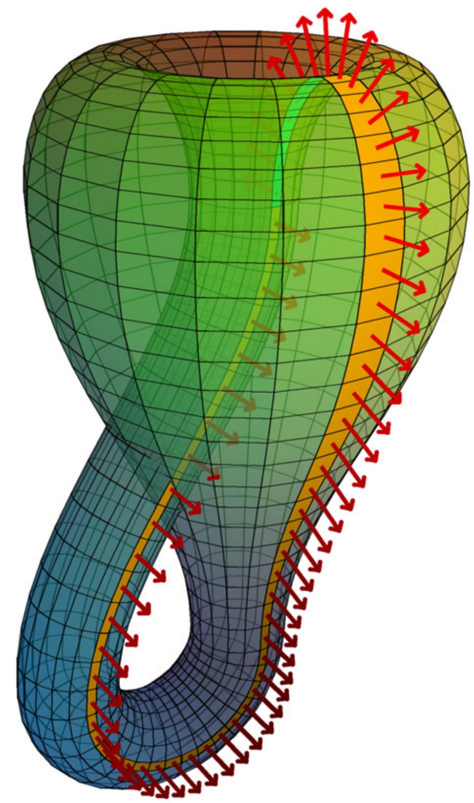
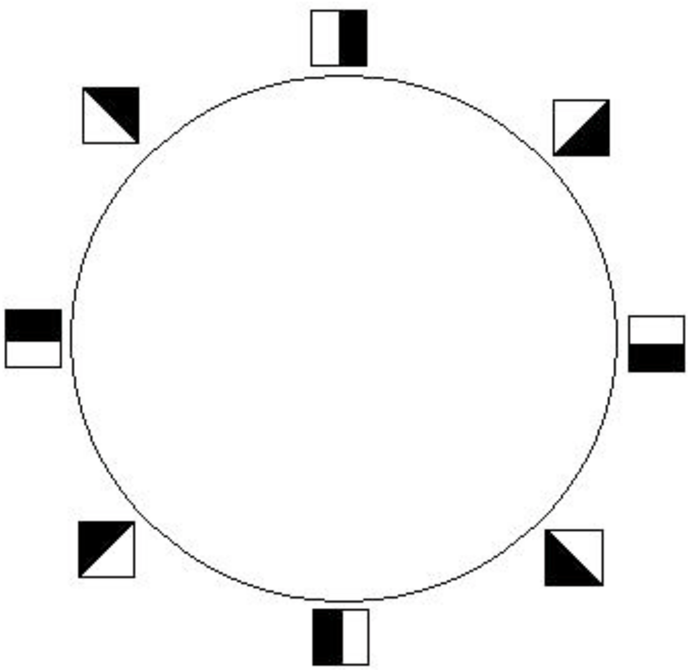


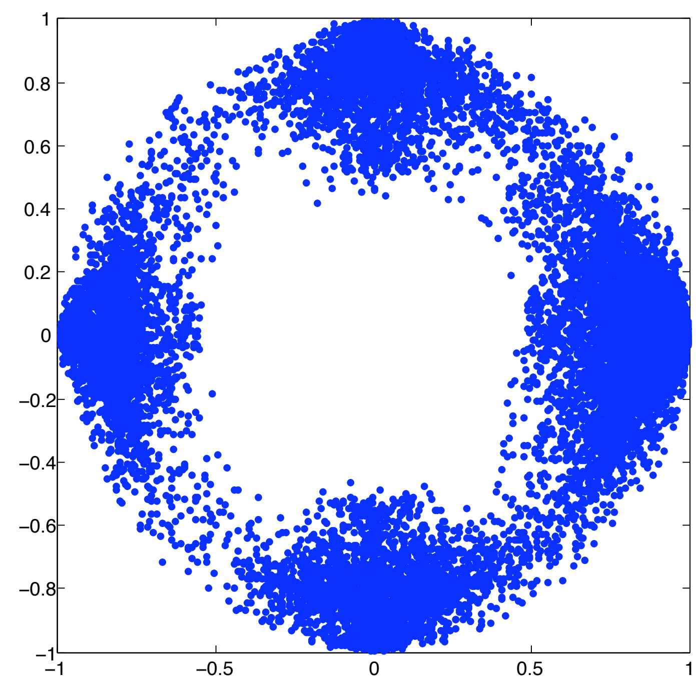
Image credit: https://plus.maths.org/ content/imaging-maths-inside-klein-bottle

Interpretation: nature prefers linear and quadratic patches at all angles

Persistent homology applied to data

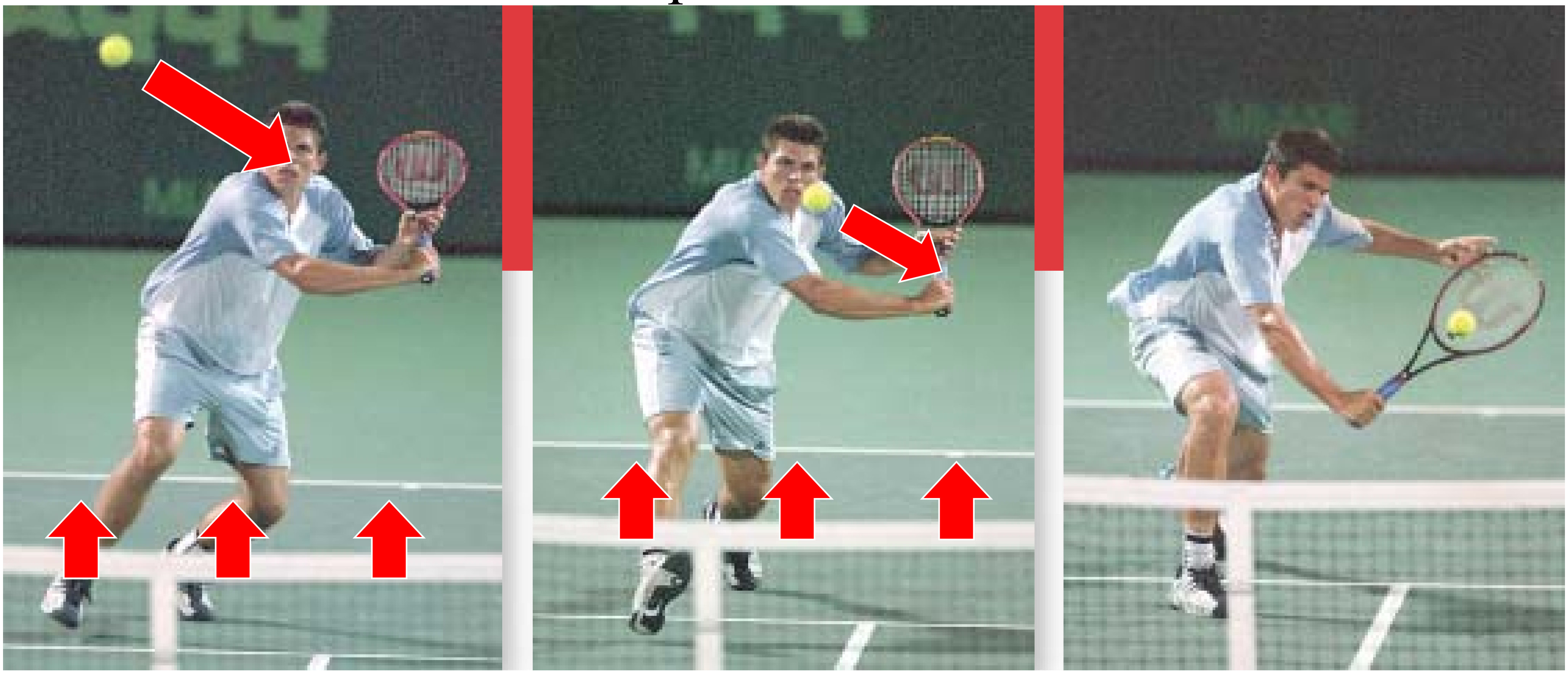
Range Images





Persistent homology applied to data

Optical Flow



Optical flow is a vector field representing the apparent motion (or projected motion) in a video.

On the nonlinear statistics of optical flow by HA, Johnathan Bush, Brittany Carr, Lara Kassab, and Joshua Mirth, 2018

Persistent homology applied to data

Optical Flow

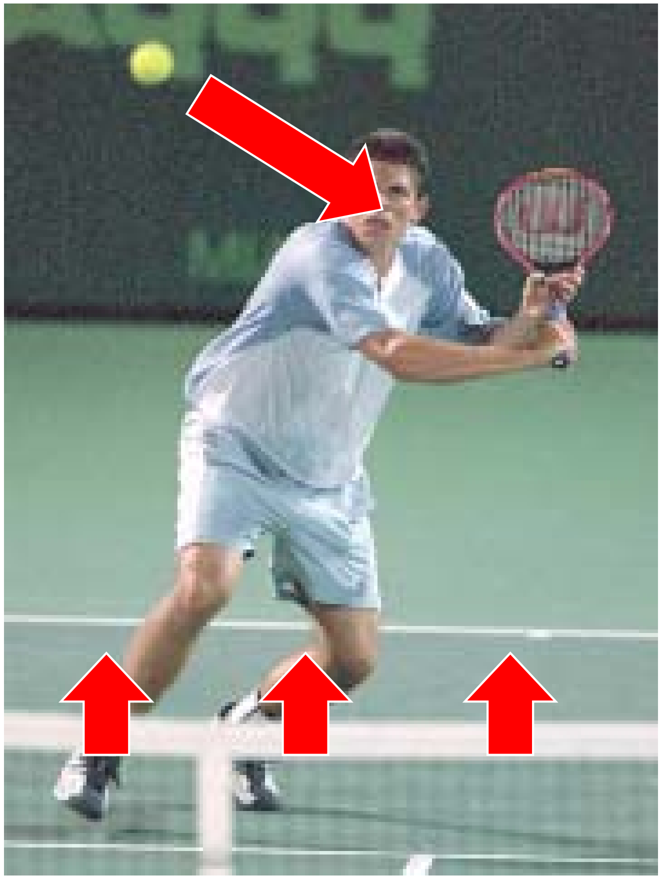
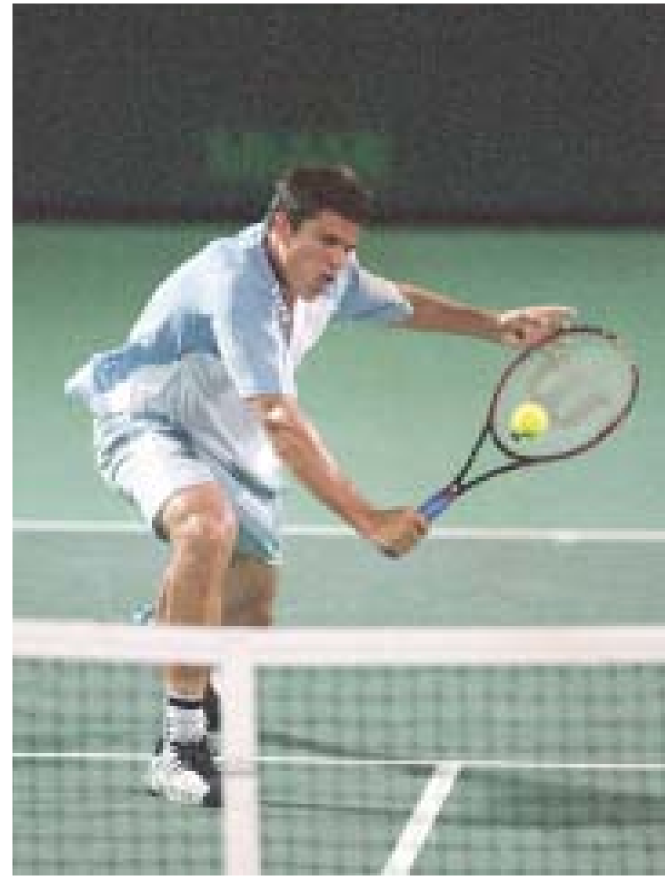


Image: Wikipedia

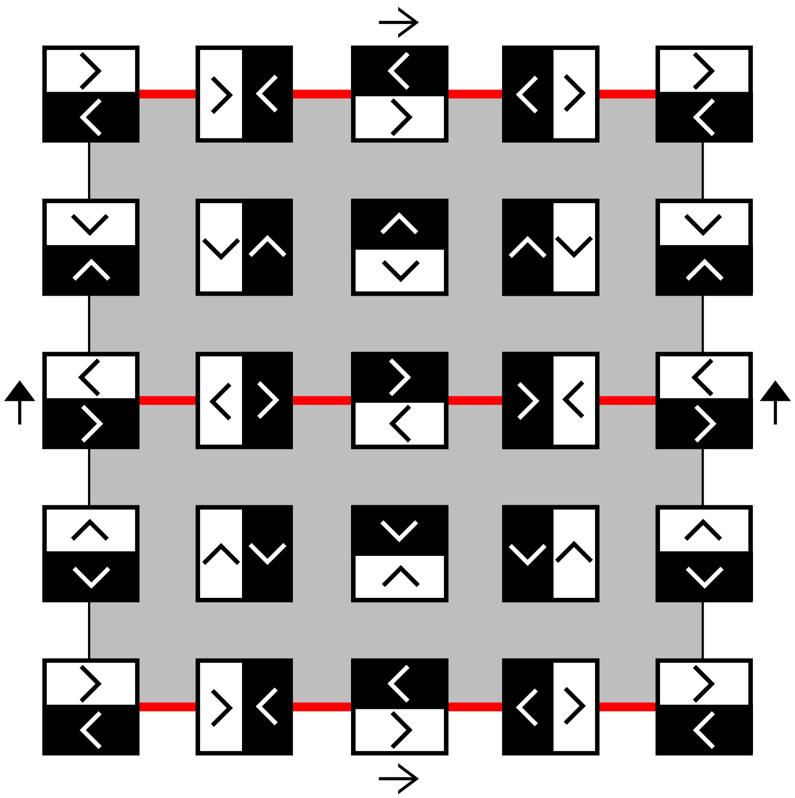


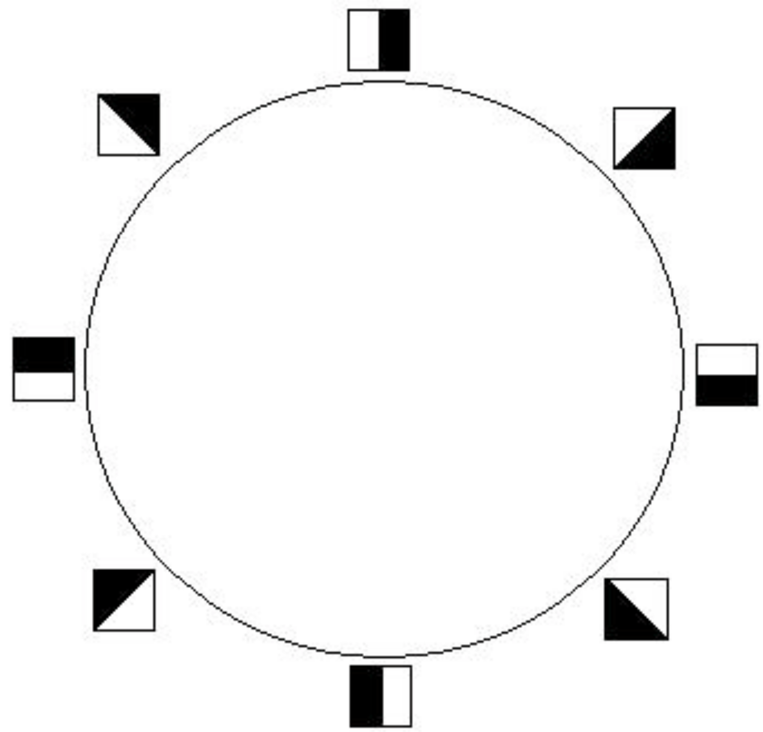
Optical flow is a vector field representing the apparent motion (or projected motion) in a video.

On the nonlinear statistics of optical flow by HA, Johnathan Bush, Brittany Carr, Lara Kassab, and Joshua Mirth, 2018

Persistent homology applied to data

Optical Flow

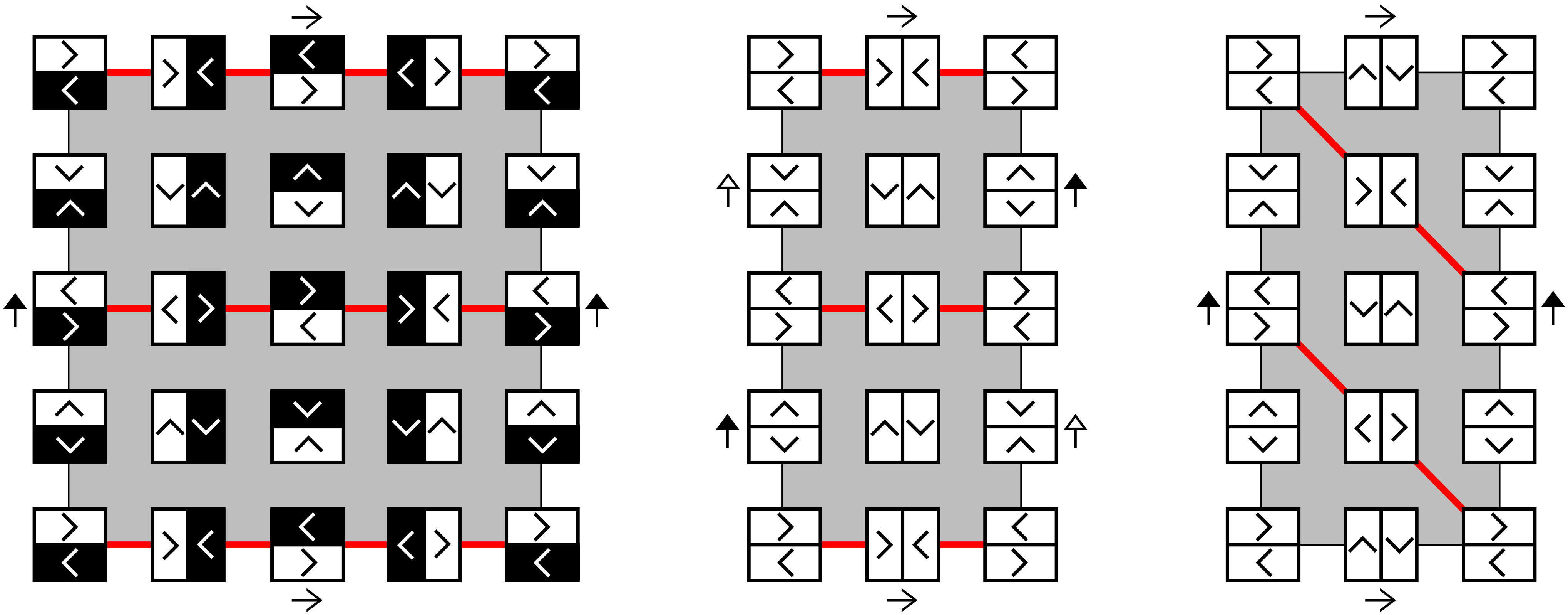




On the nonlinear statistics of optical flow by HA, Johnathan Bush, Brittany Carr, Lara Kassab, and Joshua Mirth, 2018

Persistent homology applied to data

Optical Flow

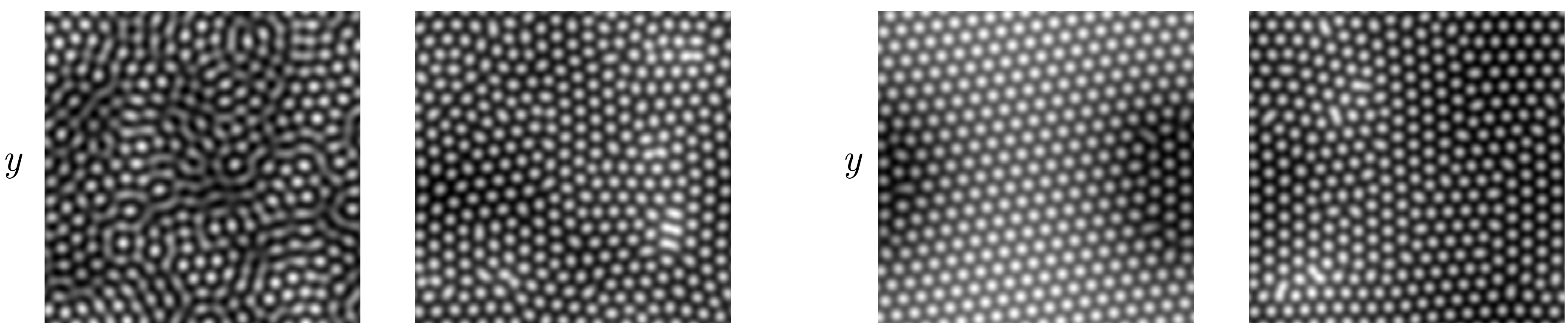


On the nonlinear statistics of optical flow by HA, Johnathan Bush, Brittany Carr, Lara Kassab, and Joshua Mirth, 2018

Why is applied topology popular when few datasets have Klein bottles?

- Many datasets have clusters & flares (as in the diabetes example)
- Motivates interesting questions in many pure disciplines: mathematics, computer science (computational geometry), statistics
- Interest from domain experts in biology, neuroscience, computer vision, dynamical systems, sensor networks, ...
- Materials science, pattern formation
- Machine learning: small features matter
- Agent-based modeling (swarming)

Possible answer: Persistent homology measures both the local geometry and the global topology of a dataset.

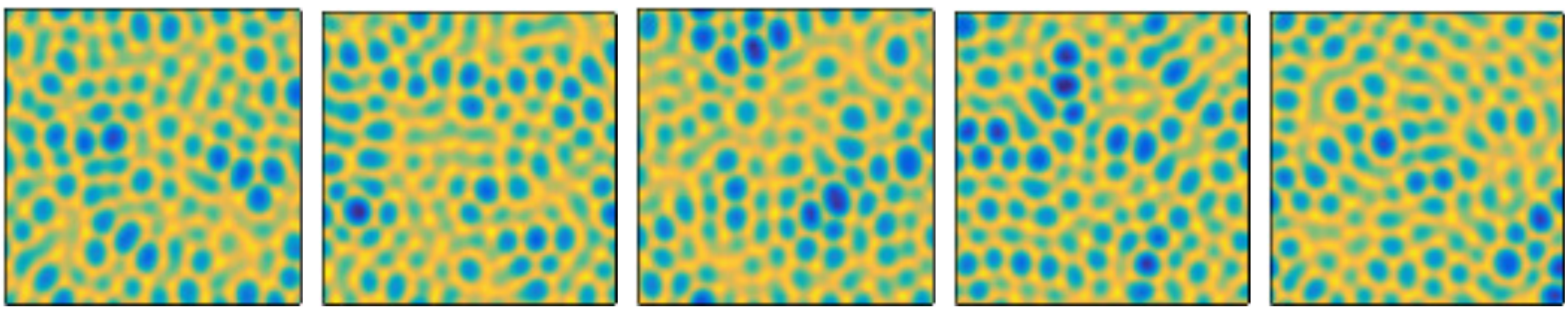


Measures of Order for nearly hexagonal lattices by Francis Motta, Rachel Neville, Patrick Shipman, Daniel Pearson, and Mark Bradley, 2018.

Why is applied topology popular when few datasets have Klein bottles?

- Many datasets have clusters & flares (as in the diabetes example)
- Motivates interesting questions in many pure disciplines: mathematics, computer science (computational geometry), statistics
- Interest from domain experts in biology, neuroscience, computer vision, dynamical systems, sensor networks, ...
- Materials science, pattern formation
- Machine learning: small features matter
- Agent-based modeling (swarming)

Possible answer: Persistent homology measures both the local geometry and the global topology of a dataset.

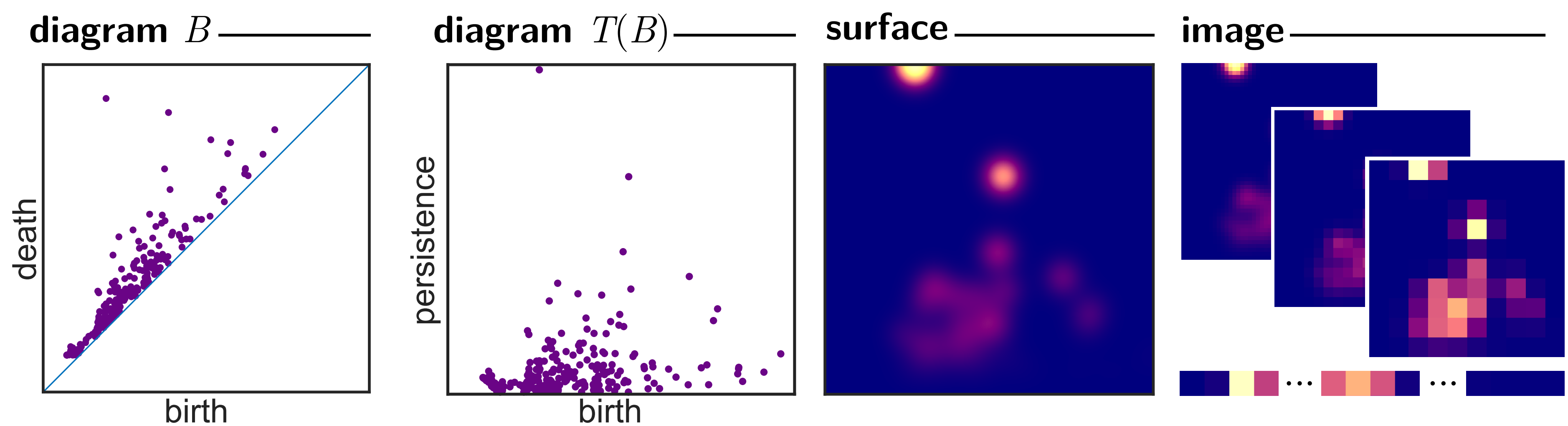


Answer: (from left) $r = 1.75, 2, 1.75, 2, 2.$

Why is applied topology popular when few datasets have Klein bottles?

- Many datasets have clusters & flares (as in the diabetes example)
- Motivates interesting questions in many pure disciplines: mathematics, computer science (computational geometry), statistics
- Interest from domain experts in biology, neuroscience, computer vision, dynamical systems, sensor networks, ...
- Materials science, pattern formation
- Machine learning: small features matter
- Agent-based modeling (swarming)

Possible answer: Persistent homology measures both the local geometry and the global topology of a dataset.



Agent-Based Modeling

Collective phenomenon, self-organization

Dutch Starling murmuration
filmed by Roald van Stijn

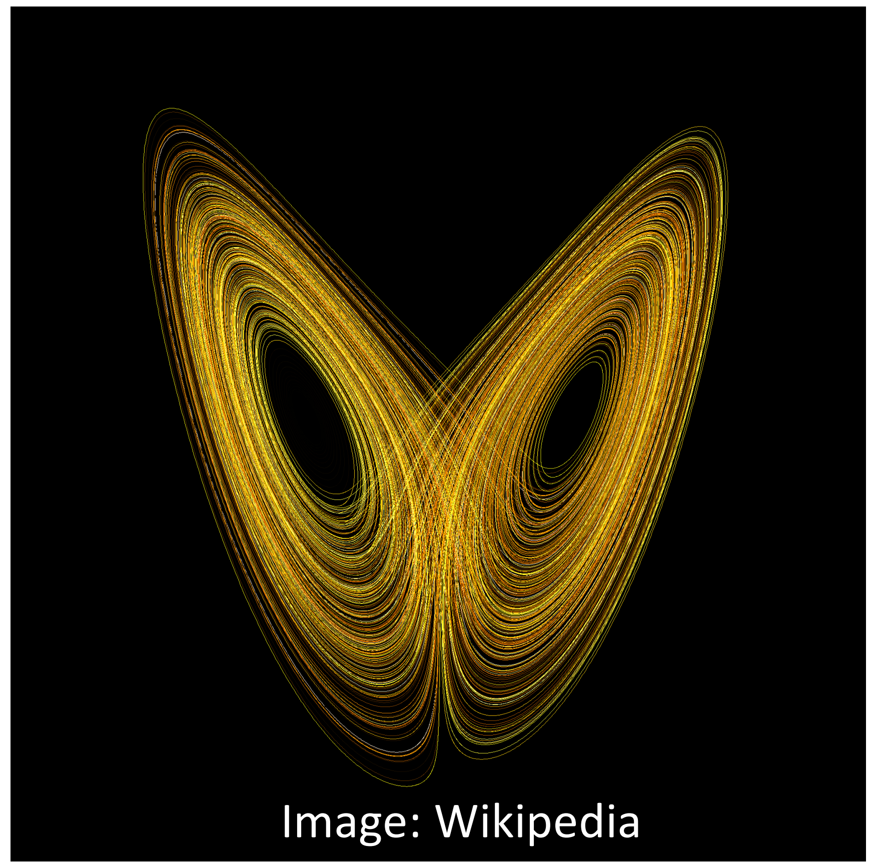
https://www.youtube.com/watch?v=YjDYE5CUb7Q

Takens' Theorem

Roughly speaking: Let M be a d-dimensional compact manifold, let $\phi\colon M \to M$ be a flow, and let $f\colon M \to \mathbb{R}$ be a measurement. Then *generically*,

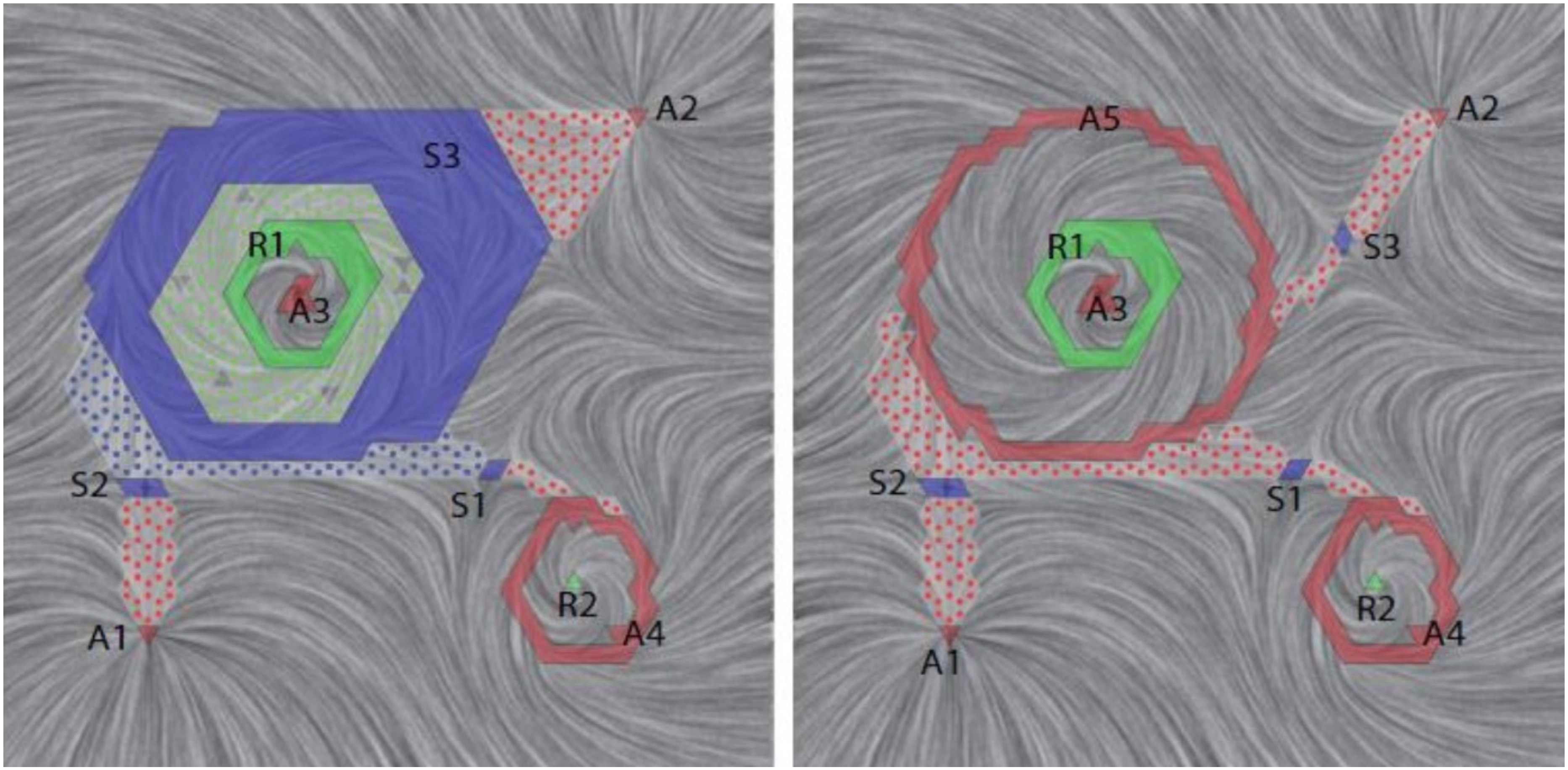
$$m \mapsto \big(f(m), f(\phi(m), f(\phi^2(m), \ldots, f(\phi^{2d}(m))\big)$$

is an embedding $M \hookrightarrow \mathbb{R}^{2d+1}$.



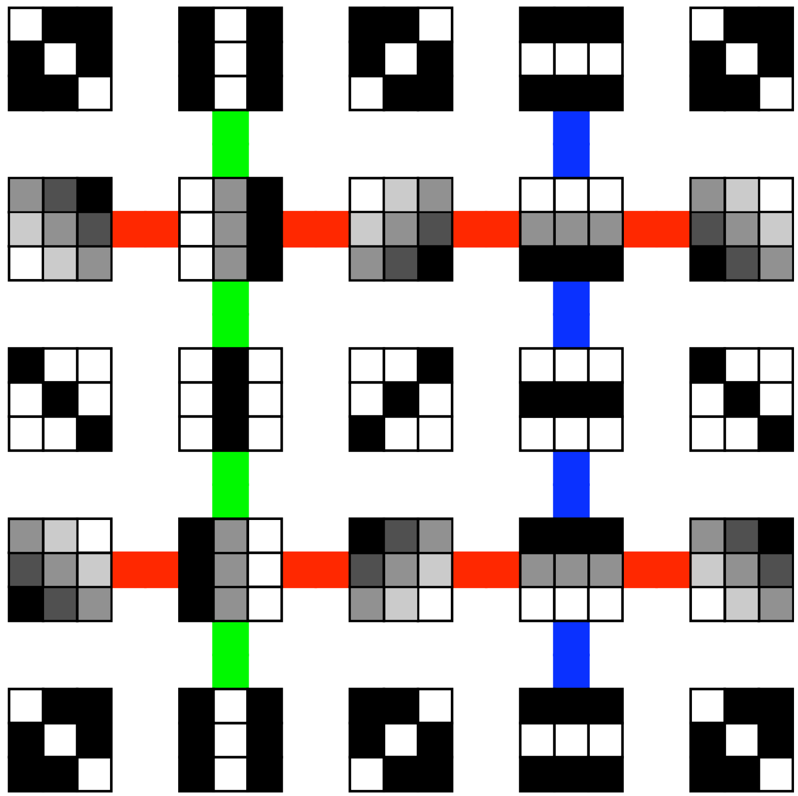
Detecting strange attractors in turbulence by Floris Takens, 1982

Conley index theory



Conclusions

- Datasets have shape, which are reflective of patterns within.
- Persistent homology is a way to measure some of the local geometry and global topology of a dataset.



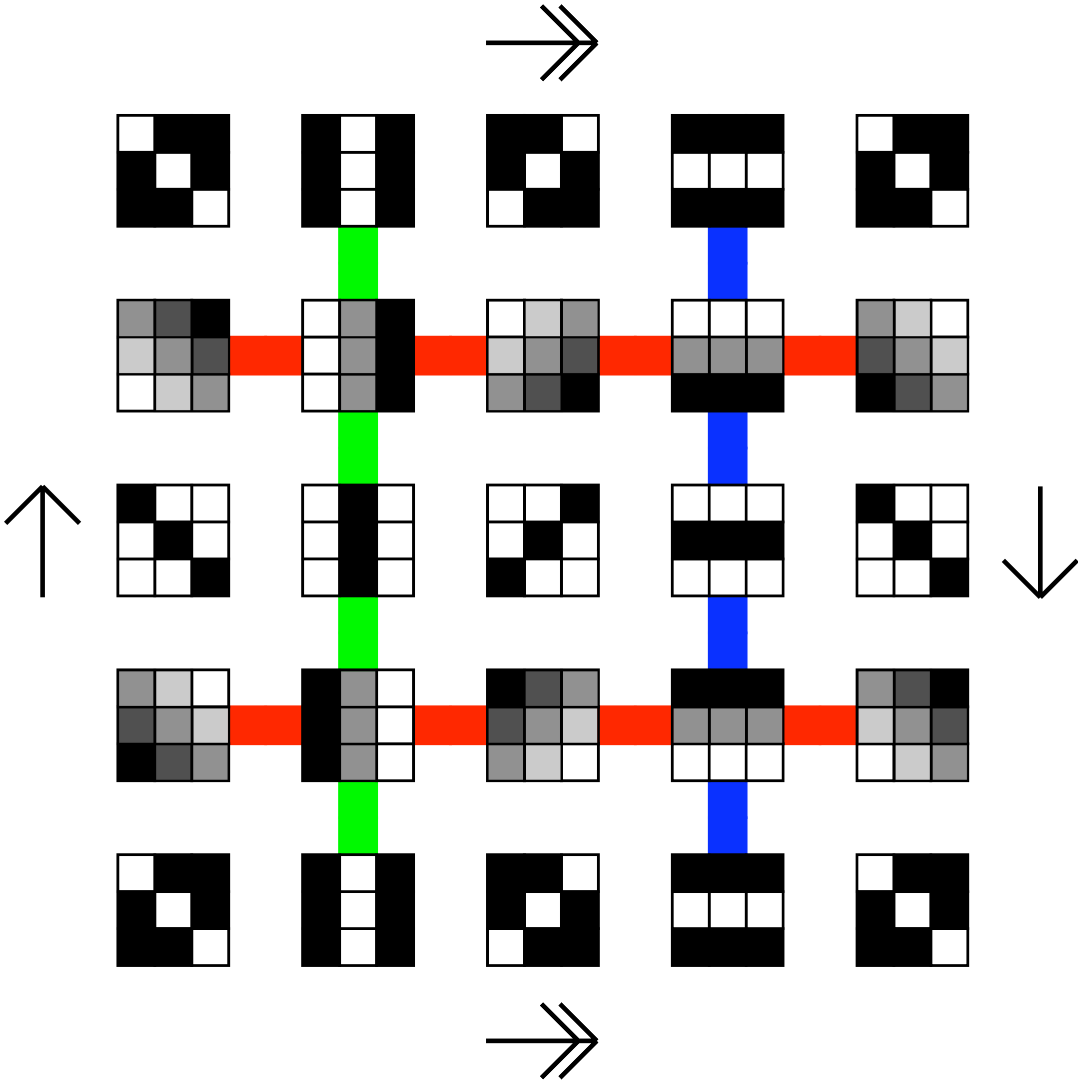
"Topology! The stratosphere of human thought! In the twenty-fourth century it might possibly be of use to someone …"

- Aleksandr Solzhenitsyn, *The First Circle*

Where can I find resources if I am interested in applied topology?

- You may be interested in the Applied Algebraic Topology Research Network. Become a member to receive email invites to the online research seminars. Recorded talks are available at the YouTube Channel. There is also a forum.
- Another source of applied topology news is appliedtopology.org.
- A second online research seminar is GEOTOP-A: Applications of Geometry and Topology.
- Mailing lists with announcements in applied topology include WinCompTop and ALGTOP-L.

https://www.math.colostate.edu/~adams/advising



Thank you!