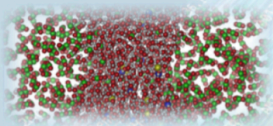
An Introduction to Persistent Homology

Henry Adams, Johnathan Bush, Mark Heim, Joshua Mirth
Colorado State University

DELTA

Descriptors of Energy Landscapes Using Topological Analysis



3N Energy Landscape (Simulation/ Experiment)

Dimensionality Reduction

- PCA
- Non-linear Methods
- Generalized Collective Coordinates

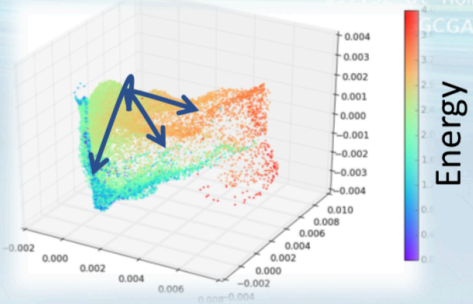
Topology of Reduced Energy Landscapes

- Morse Theory
- Persistent Homology
- Catastrophe Theory
- Singularity Theory

Predictive Machine Learning

- Optimized Synthetic Conditions
- Phase Behavior
- Tuning Catalytic Pathways

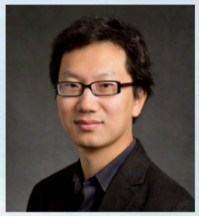
Accelerated Sampling











Washington State University · Colorado State University · Rensselaer · Auburn University · University of Colorado Boulder · University of Arkansas · Illinois

An Introduction to Persistent Homology

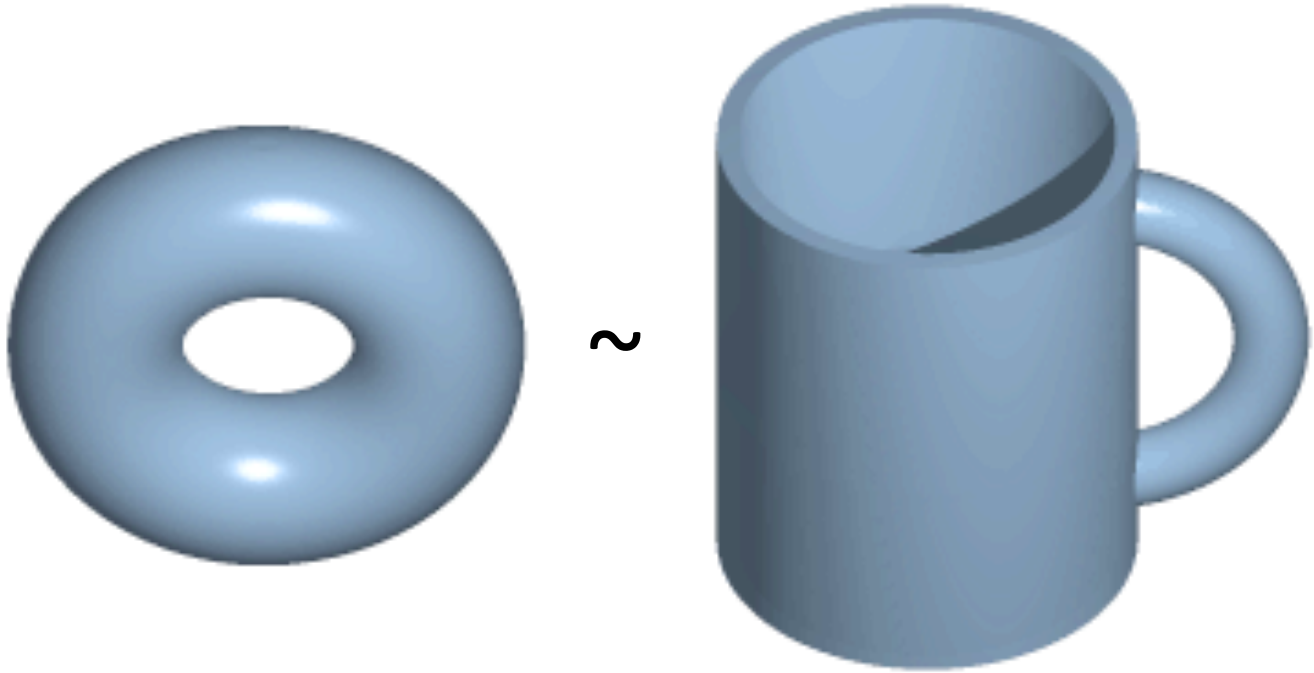
1. Review of topology and homology
2. Introduction to persistent homology
3. Sublevel set persistent homology
4. Persistent homology applied to point cloud data

An Introduction to Persistent Homology

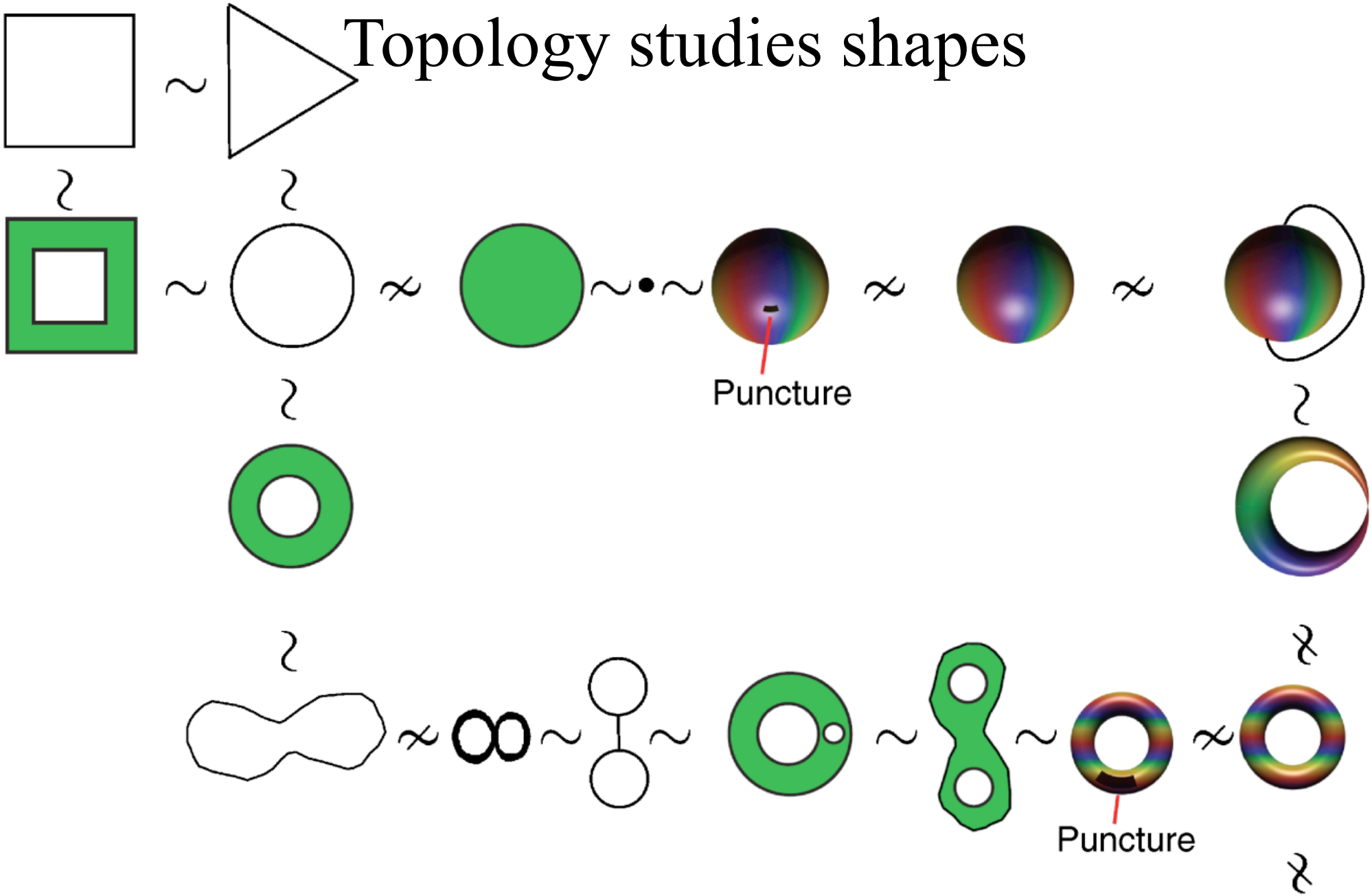
1. Review of topology and homology
2. Introduction to persistent homology
3. Sublevel set persistent homology
4. Persistent homology applied to point cloud data

Topology studies shapes

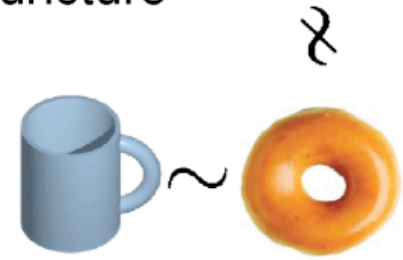
A donut and coffee mug are “homotopy equivalent” and considered to be the same shape. You can bend and stretch (but not tear) one to get the other.



Topology studies shapes

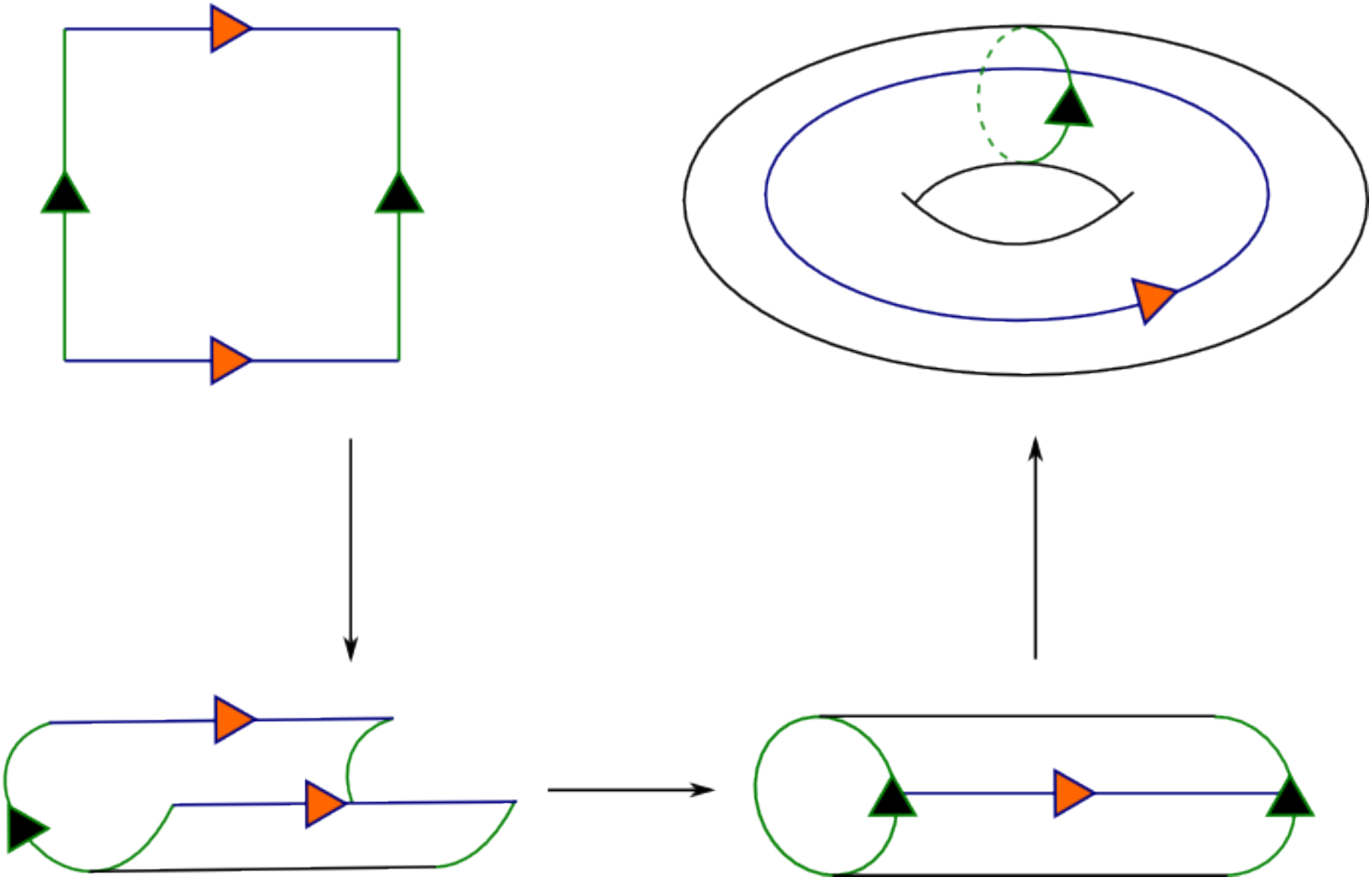


Topological Analysis of Population Activity in Visual Cortex by Singh, Memoli, Ishkhanov, Sapiro, Carlsson, and Ringach, 2008.



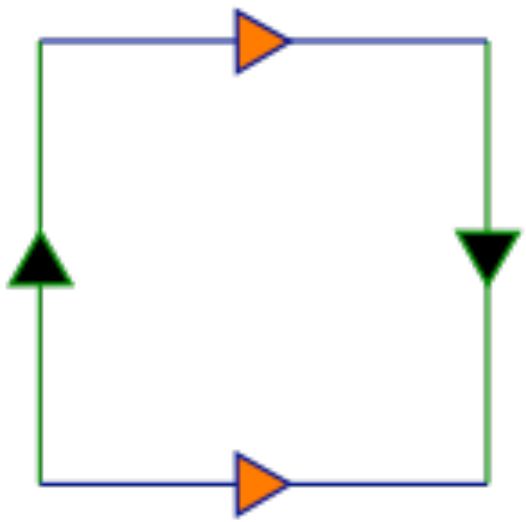
Topology studies shapes

Torus



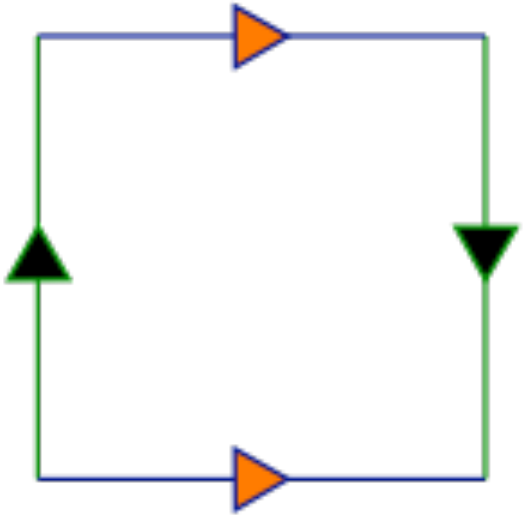
Topology studies shapes

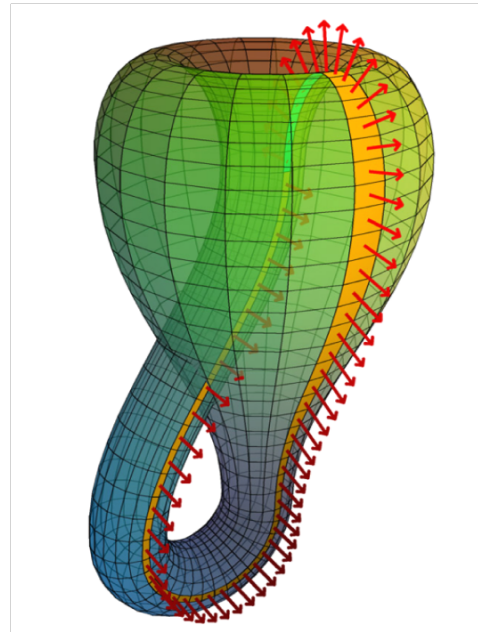
Klein bottle



Topology studies shapes

Klein bottle





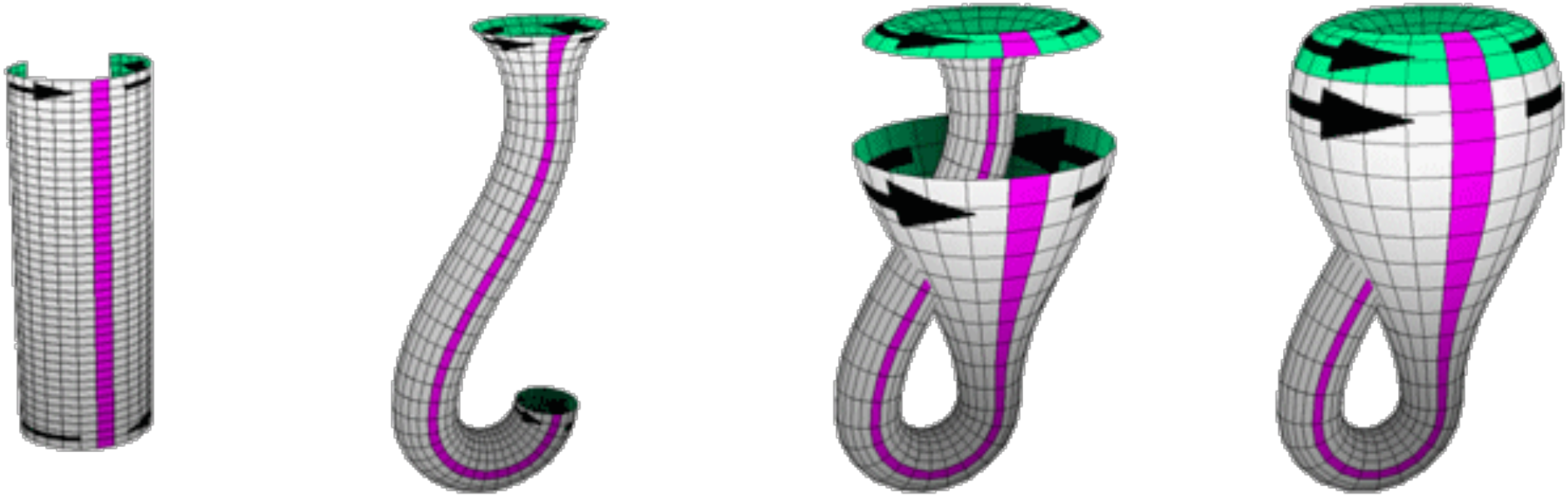
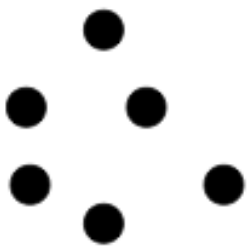


Image credit: https://plus.maths.org/content/imaging-maths-inside-klein-bottle

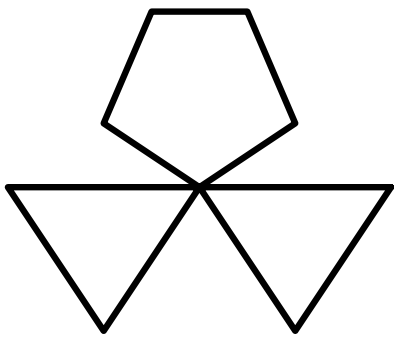
Homology

- 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!



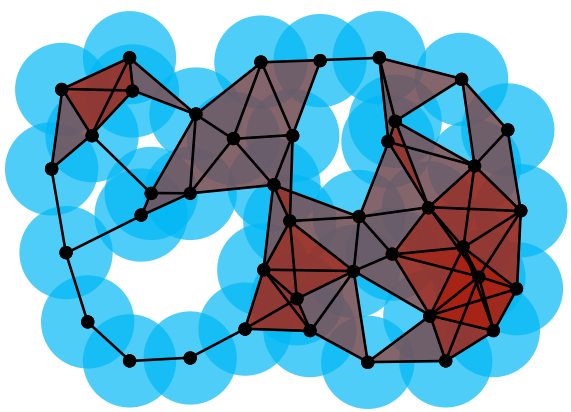
0-dimensional homology H_0: rank 6

1-dimensional homology H_1: rank 0



0-dimensional homology H_0: rank 1

1-dimensional homology H_1: rank 3

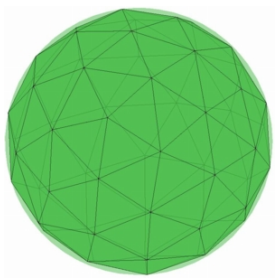


0-dimensional homology H_0: rank 1

1-dimensional homology H_1: rank 6

Homology

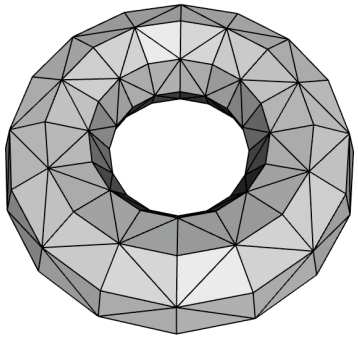
- i-dimensional homology H_i "counts the number of i-dimensional holes"
- i-dimensional homology H_i actually has the structure of a vector space!



0-dimensional homology H_0: rank 1

1-dimensional homology H_1: rank 0

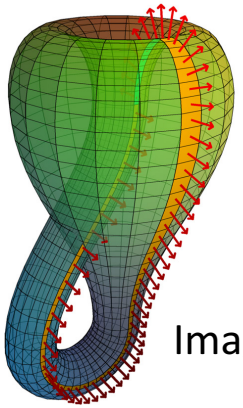
2-dimensional homology H_2: rank 1



0-dimensional homology H_0: rank 1

1-dimensional homology H_1: rank 2

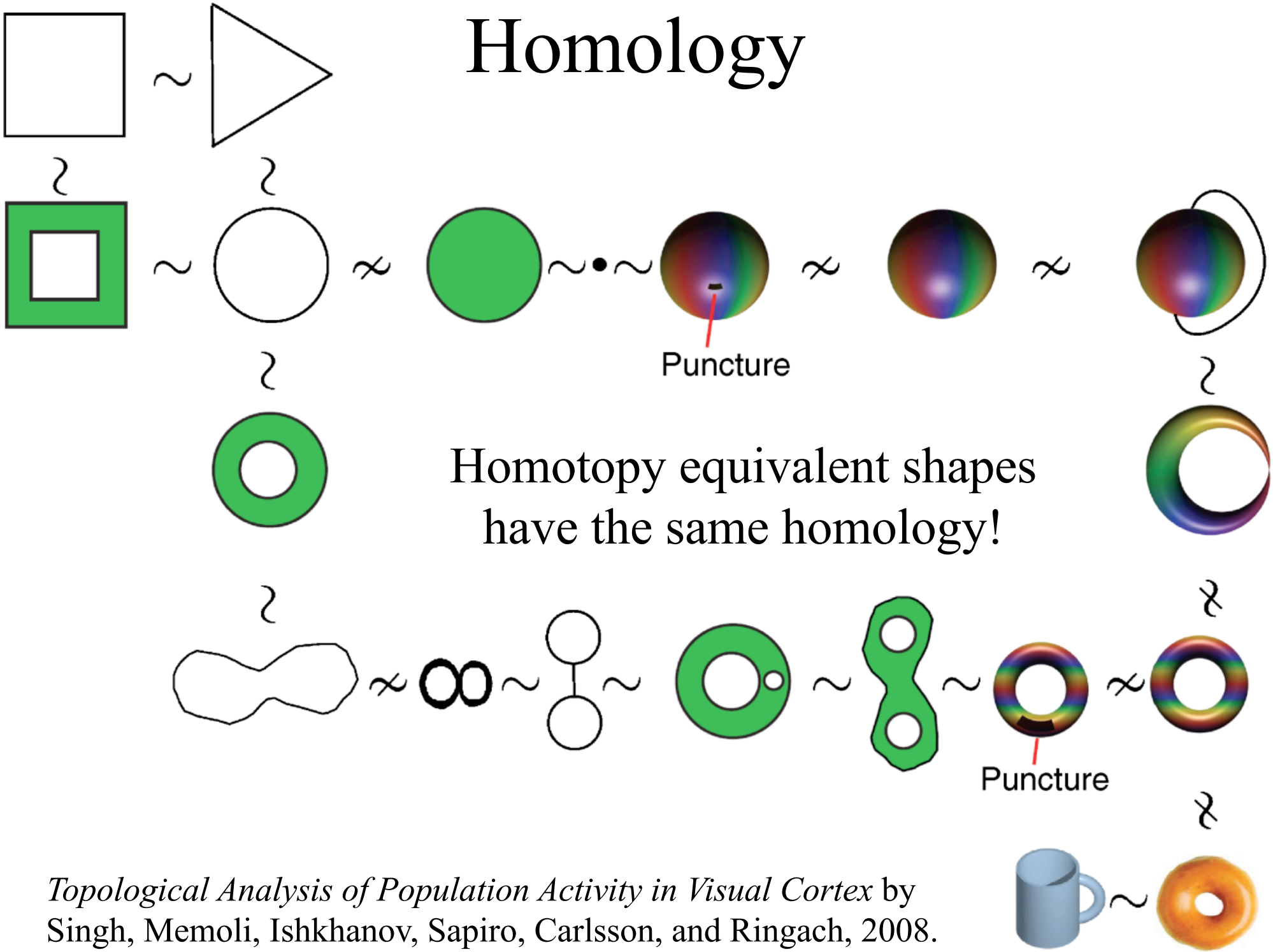
2-dimensional homology H_2: rank 1



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

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

Homology



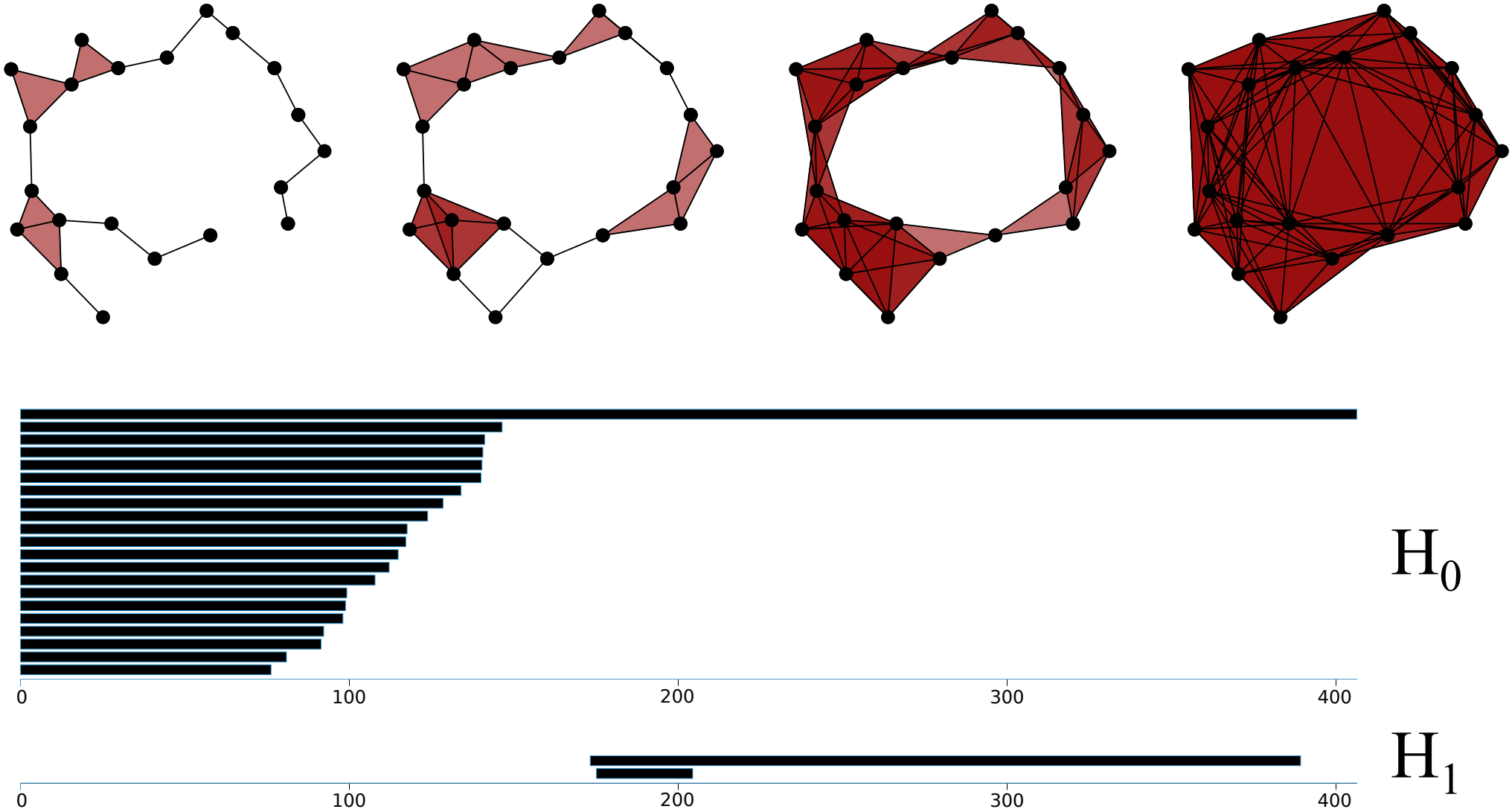
Homotopy equivalent shapes have the same homology!

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

An Introduction to Persistent Homology

1. Review of topology and homology
2. Introduction to persistent homology
3. Sublevel set persistent homology
4. Persistent homology applied to point cloud data

Persistent homology

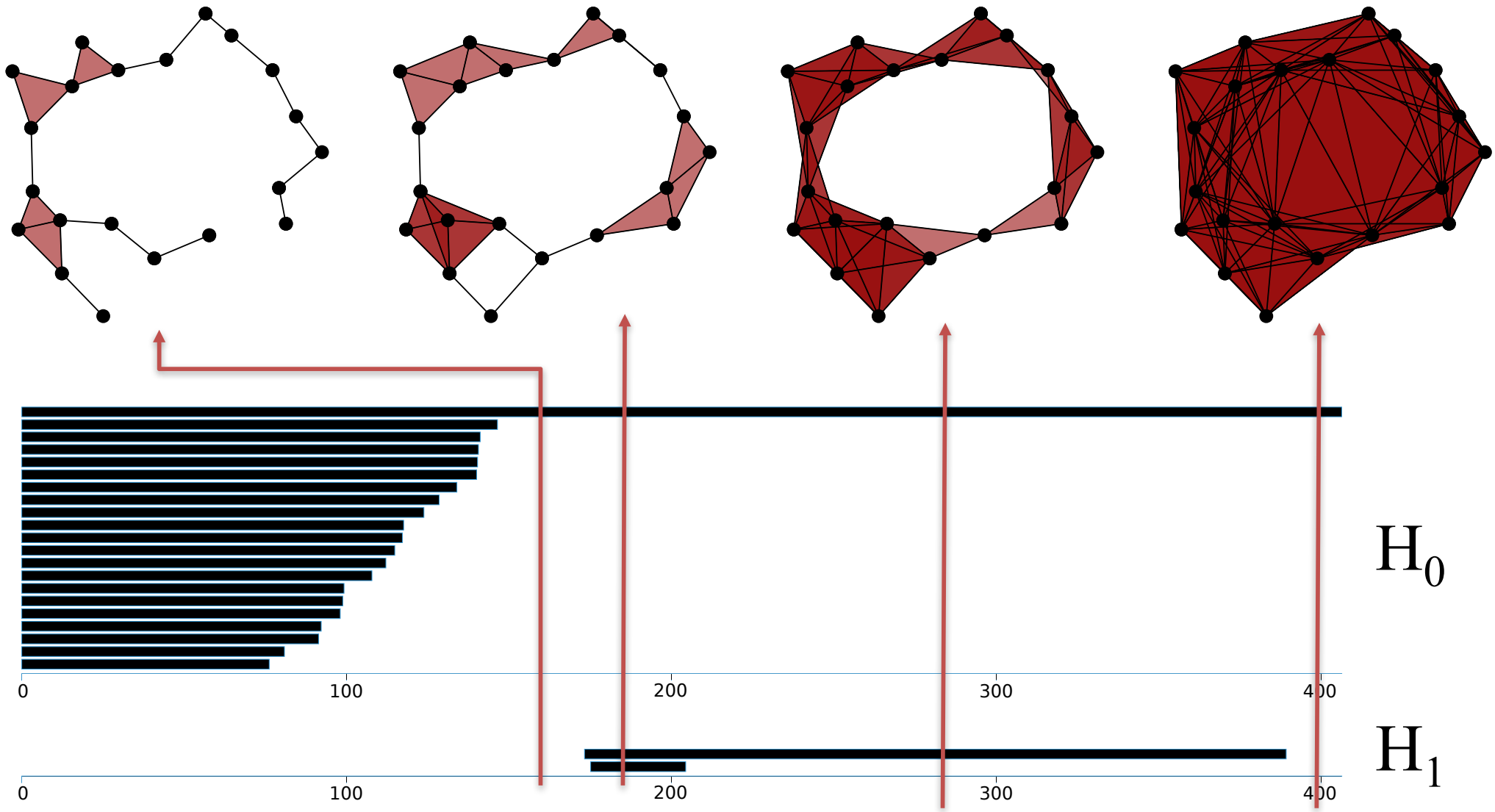


Input: Increasing spaces. Output: barcode.

Significant features persist.

Cubic computation time in the number of simplices.

Persistent homology

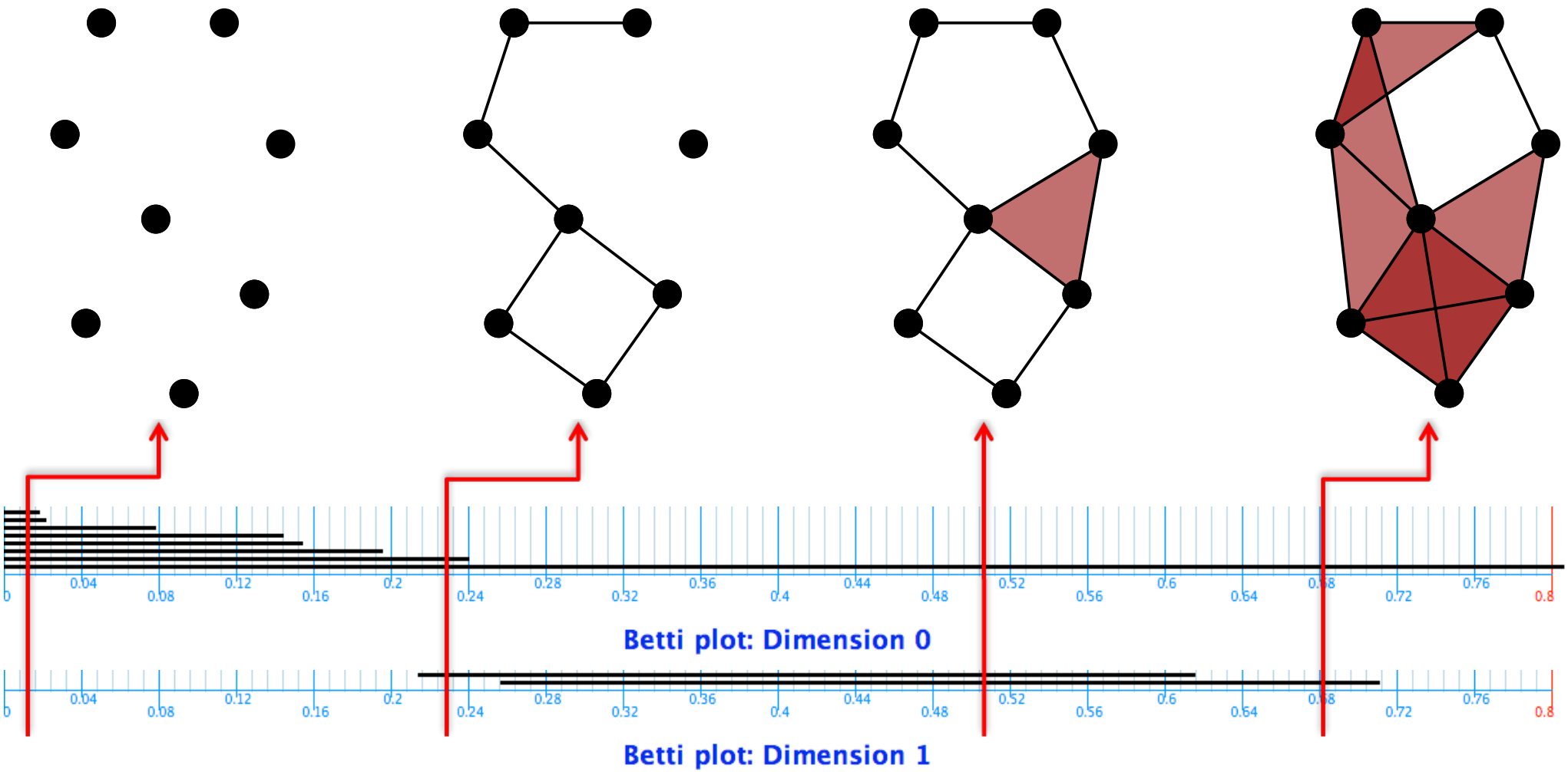


Input: Increasing spaces. Output: barcode.

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.

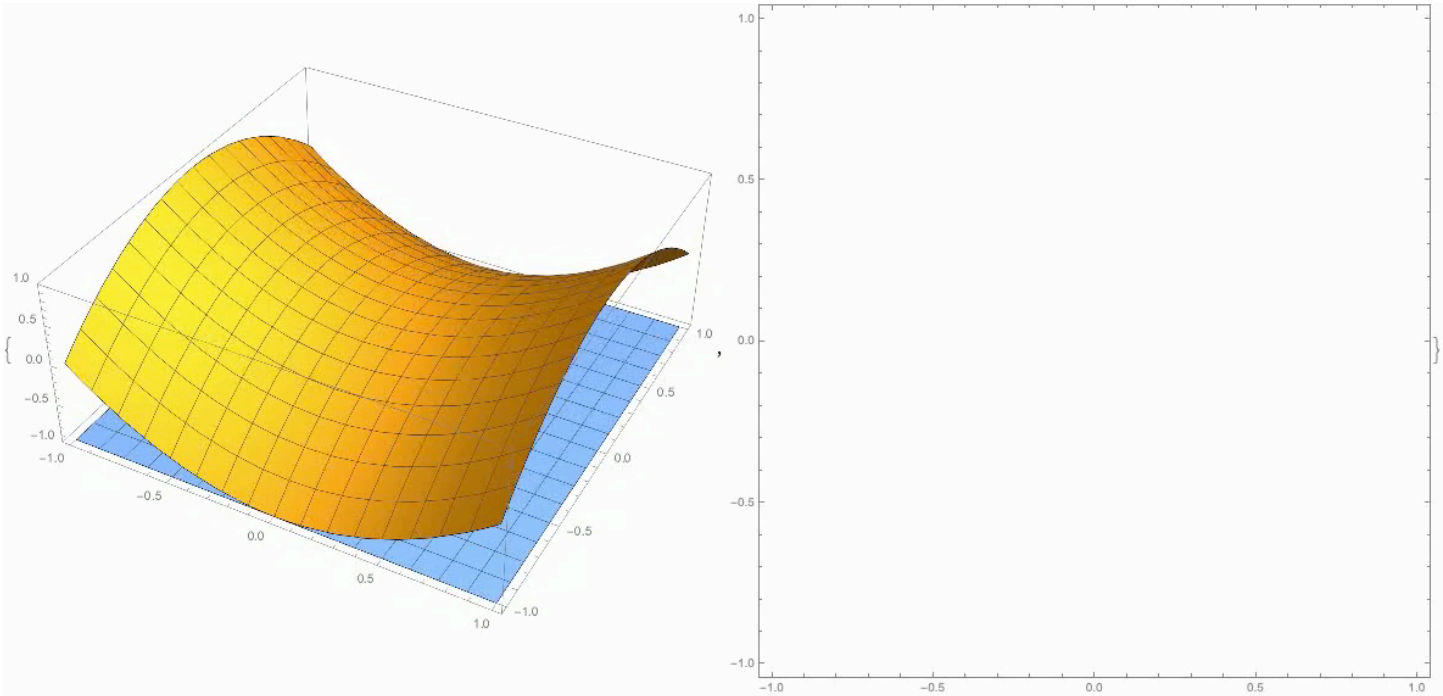
An Introduction to Persistent Homology

1. Review of topology and homology
2. Introduction to persistent homology
3. Sublevel set persistent homology
4. Persistent homology applied to point cloud data

Sublevel set persistent homology

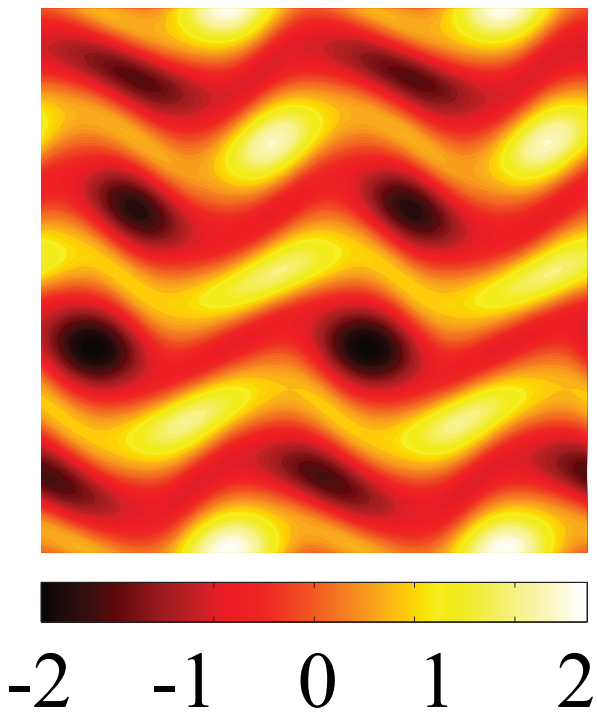
Given a space X and a real-valued function $f: X \to \mathbb{R}$, the **sublevel set** for $a \in \mathbb{R}$ consists of all the points in X with $f(x) \leq a$.

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

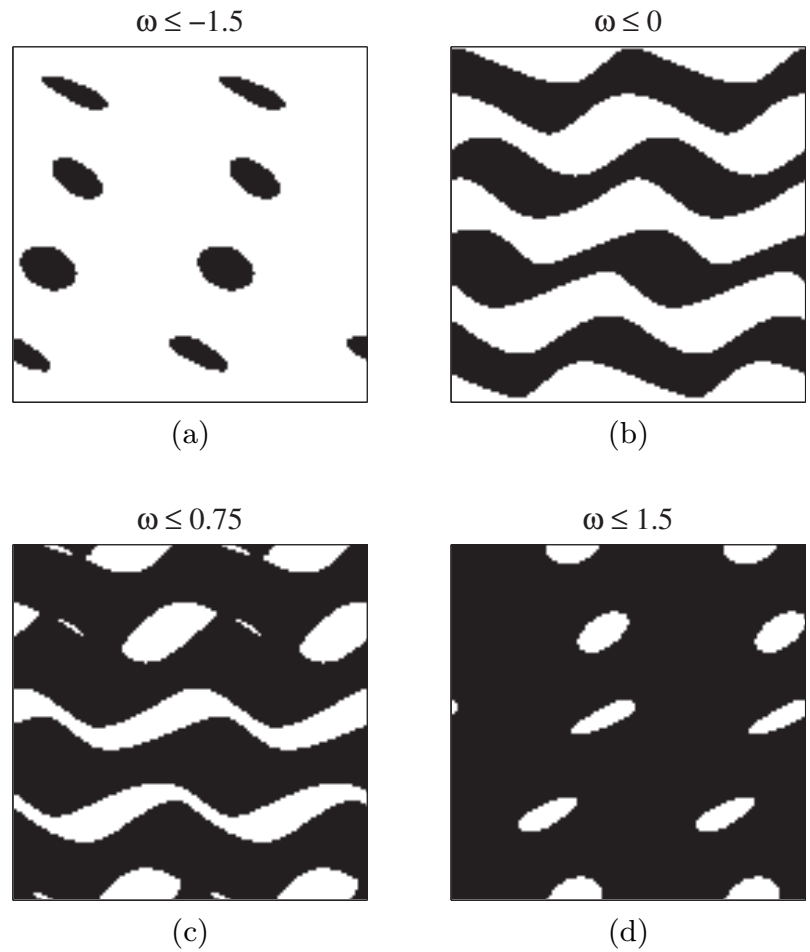


Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

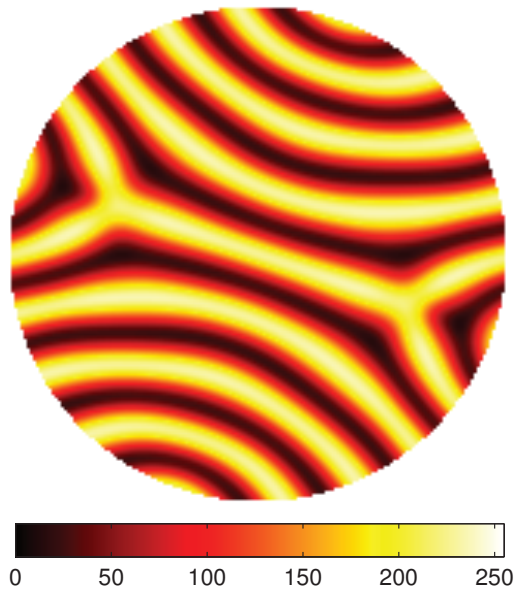


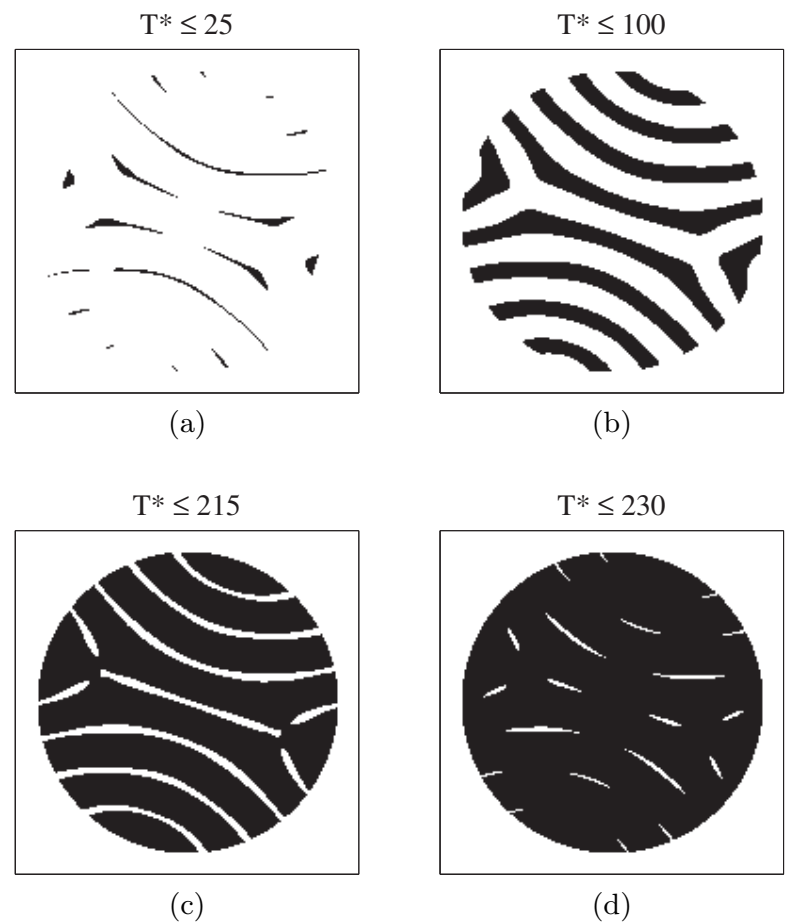
Analysis of Kolmogorov flow and Rayleigh–Bénard convection using persistent homology by Miroslav Kramár, Rachel Levanger, Jeffrey Tithof, Balachandra Suri, Mu Xu, Mark Paul, Michael F Schatz, Konstantin Mischaikow



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

Sublevel set persistent homology

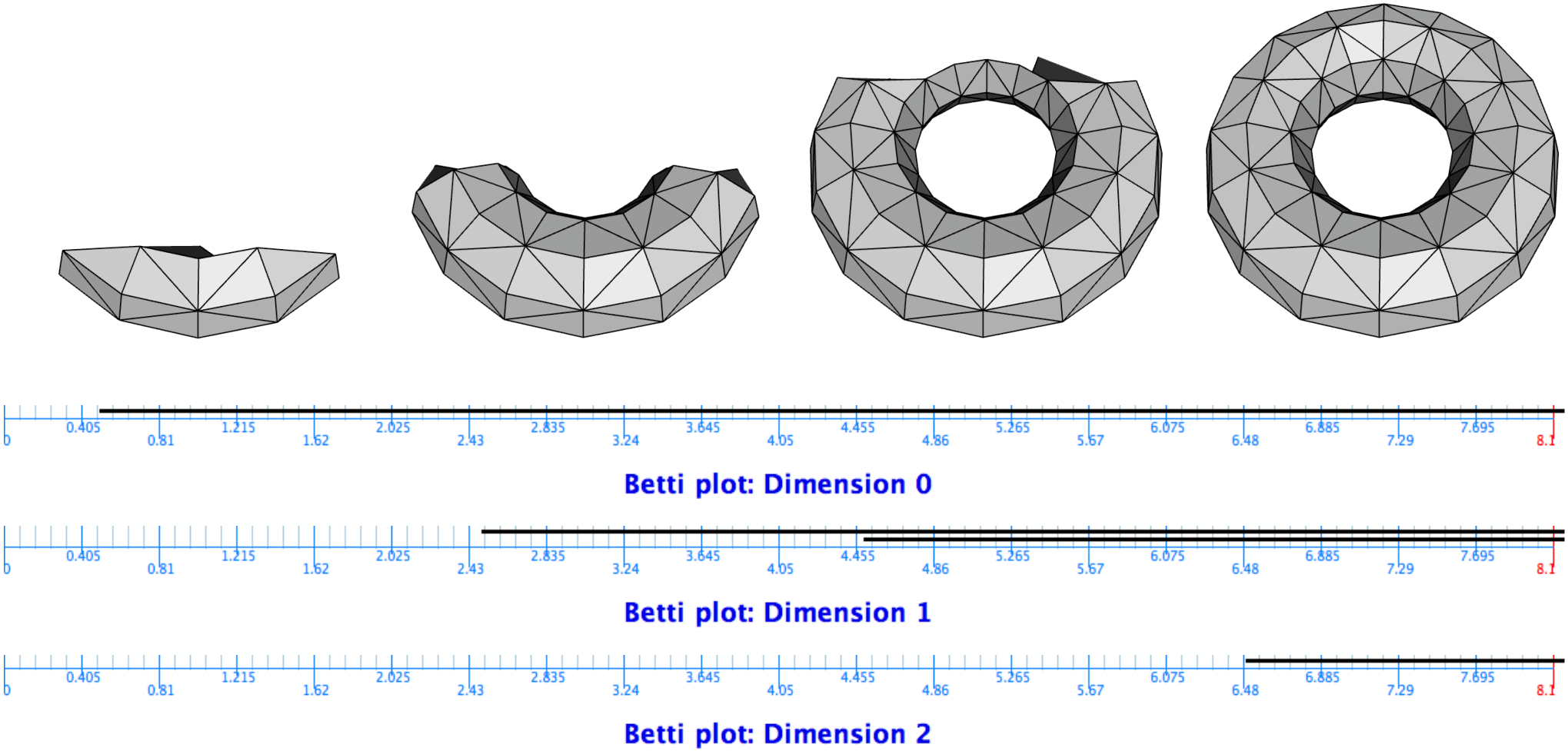




Analysis of Kolmogorov flow and Rayleigh–Bénard convection using persistent homology by Miroslav Kramár, Rachel Levanger, Jeffrey Tithof, Balachandra Suri, Mu Xu, Mark Paul, Michael F Schatz, Konstantin Mischaikow

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

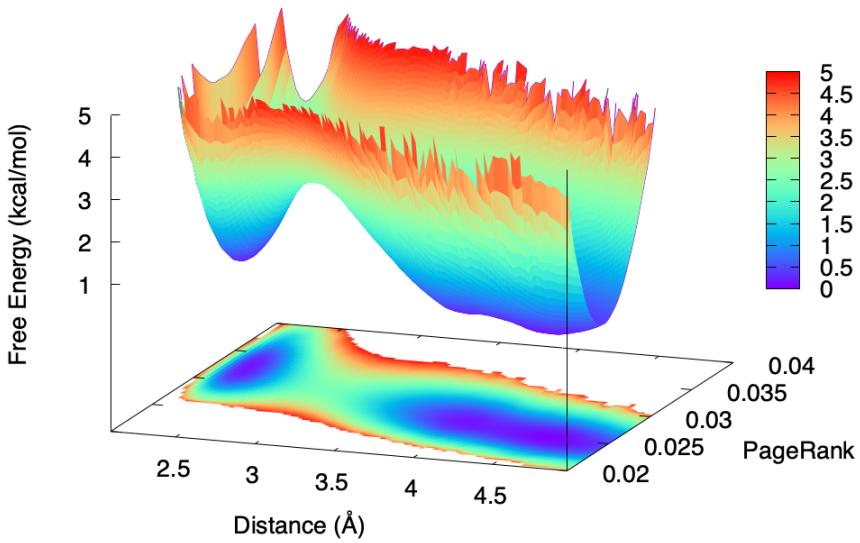
Sublevel set persistent homology

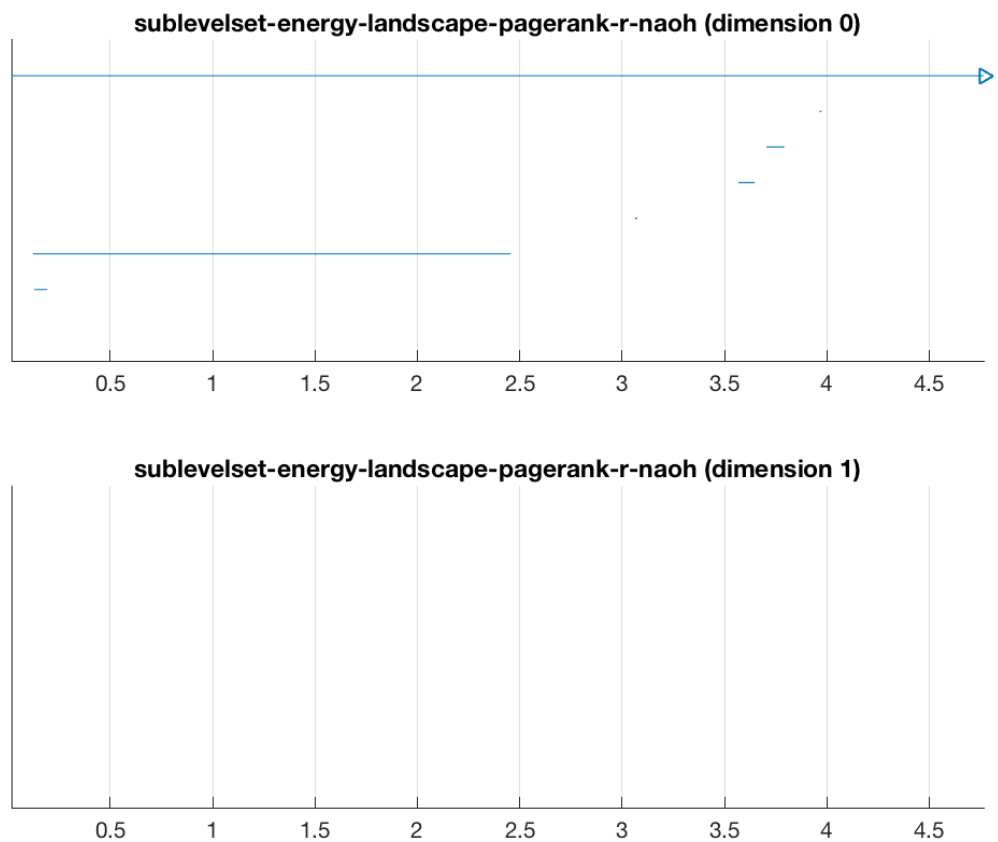


Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

EL (PageRank) of ion-pair formation of Na^+ and OH^-: superlevel sets

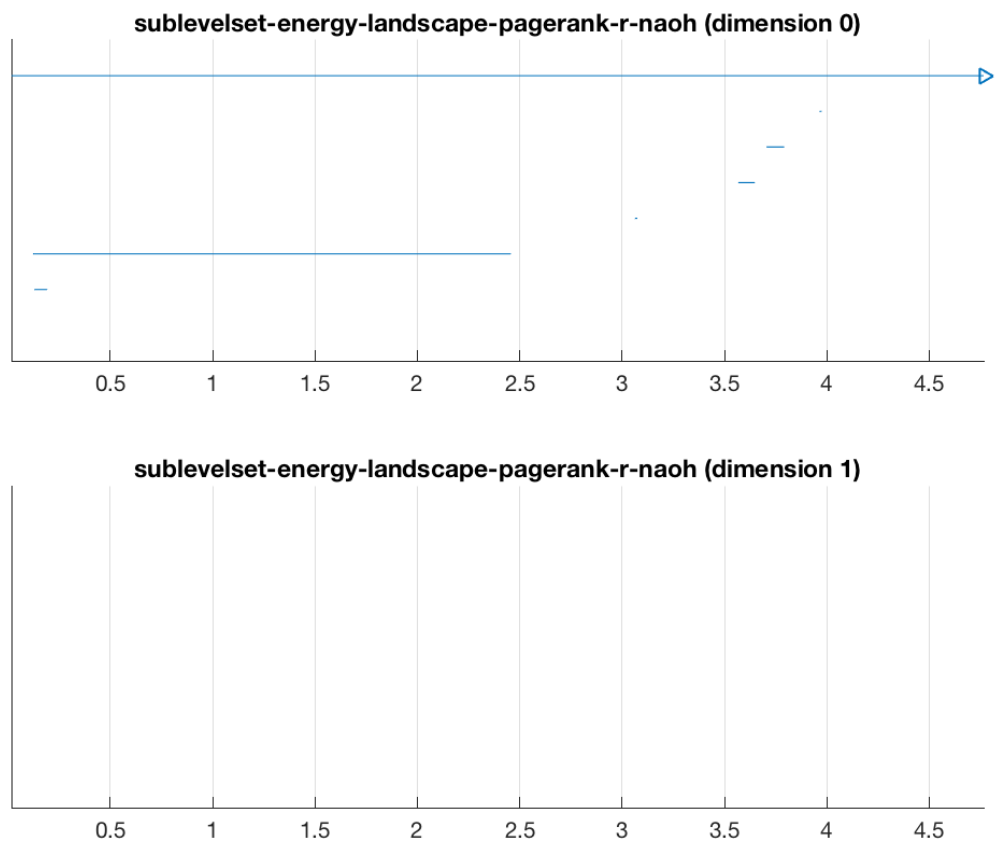




Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

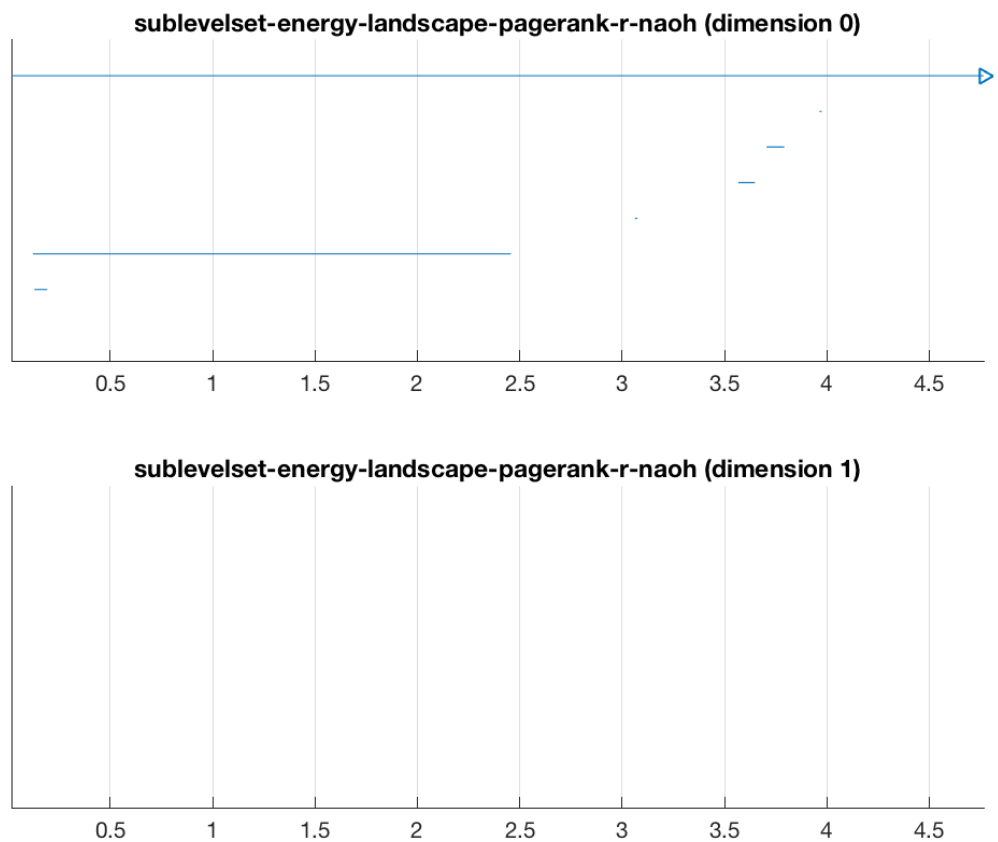
EL (PageRank) of ion-pair formation of Na^+ and OH^-: sublevel sets



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

Sublevel set persistent homology

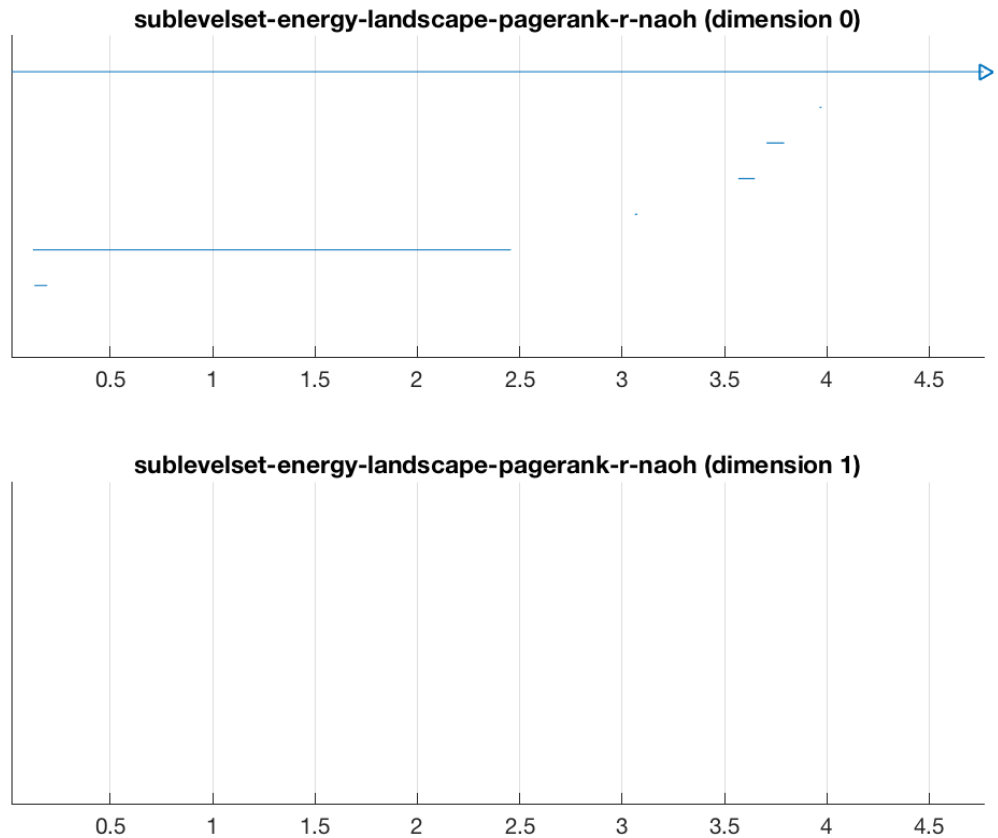
EL (PageRank) of ion-pair formation of Na^+ and OH^-: sublevel sets



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

Sublevel set persistent homology

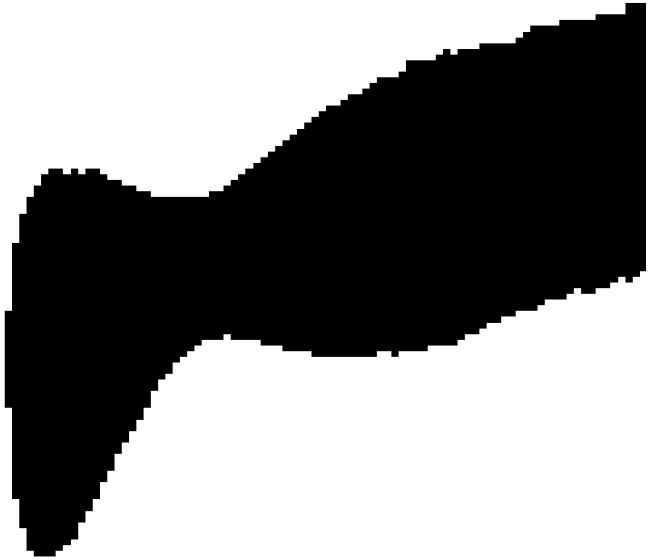
EL (PageRank) of ion-pair formation of Na^+ and OH^-: sublevel sets

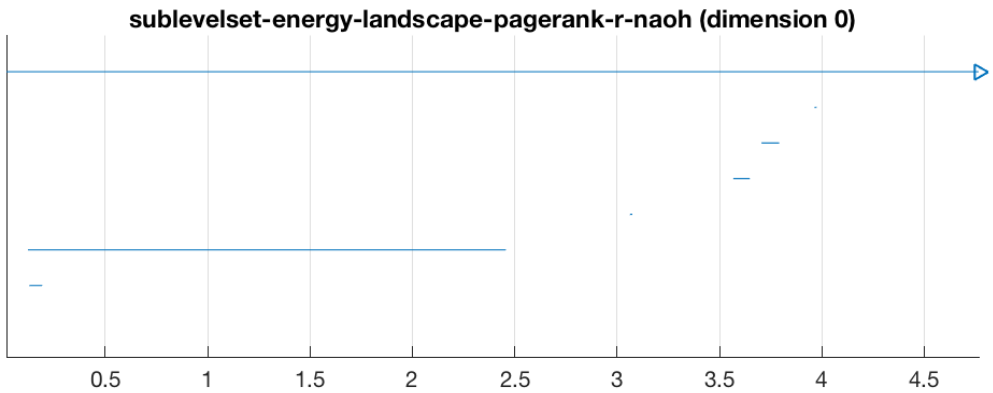


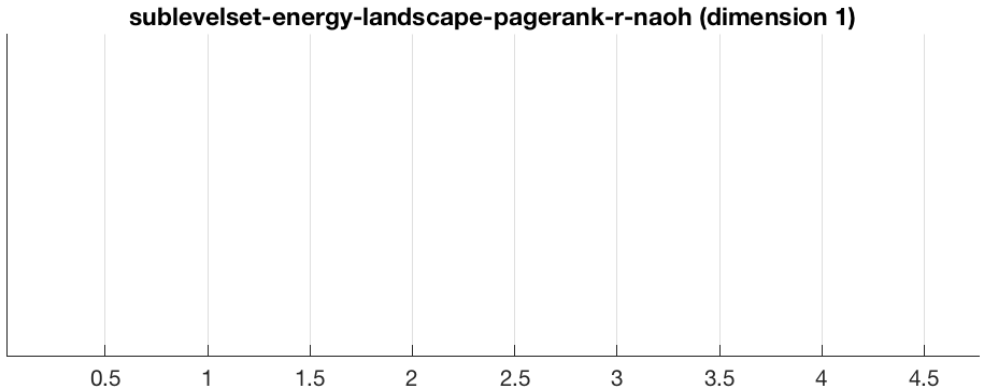
Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

EL (PageRank) of ion-pair formation of Na^+ and OH^-: sublevel sets



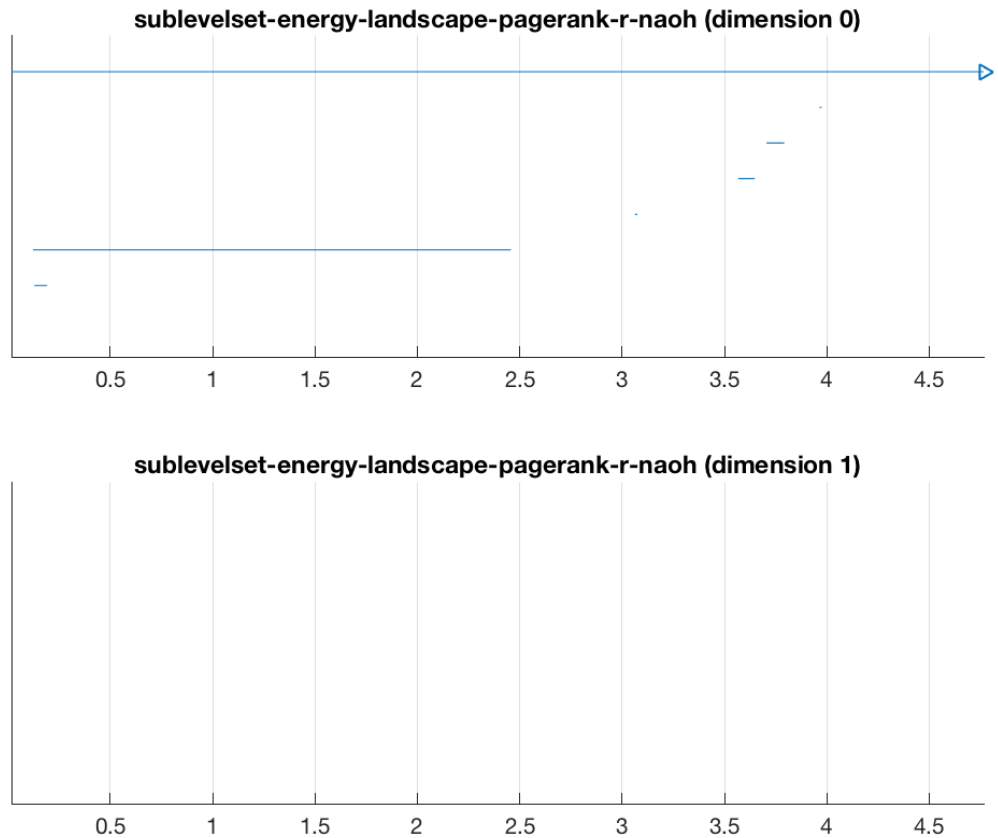




Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

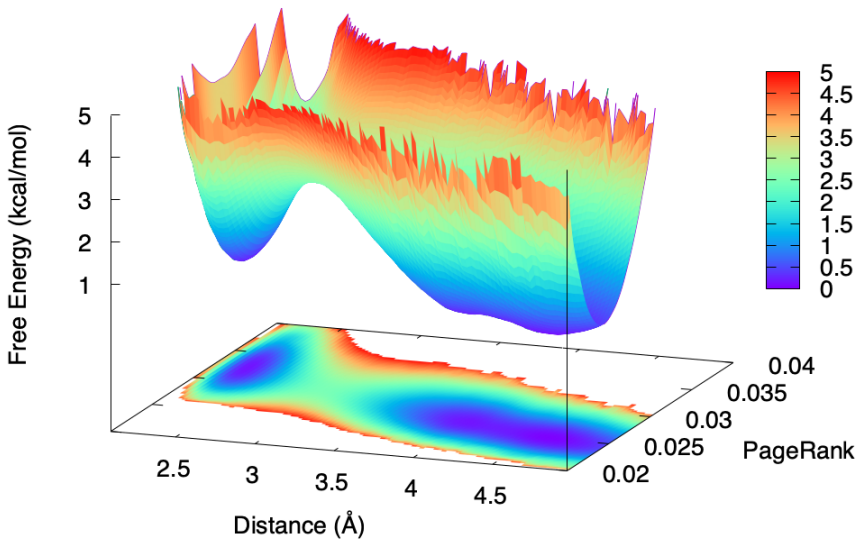
EL (PageRank) of ion-pair formation of Na^+ and OH^-: sublevel sets

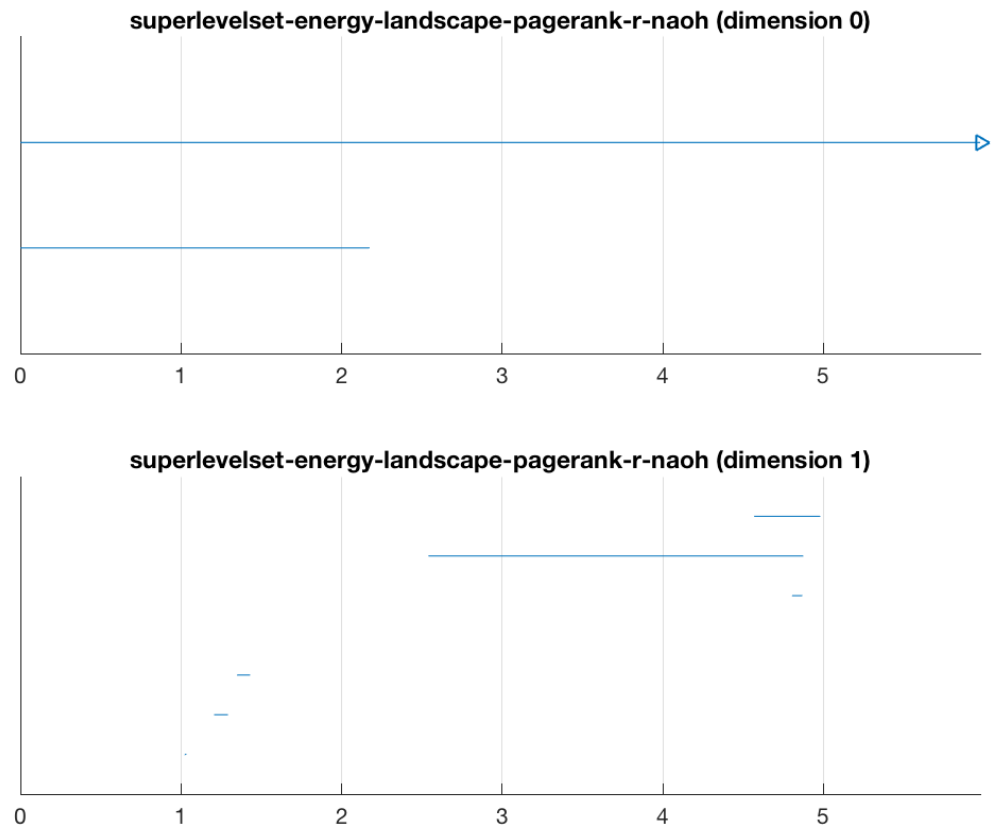


Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

EL (PageRank) of ion-pair formation of Na^+ and OH^-: superlevel sets

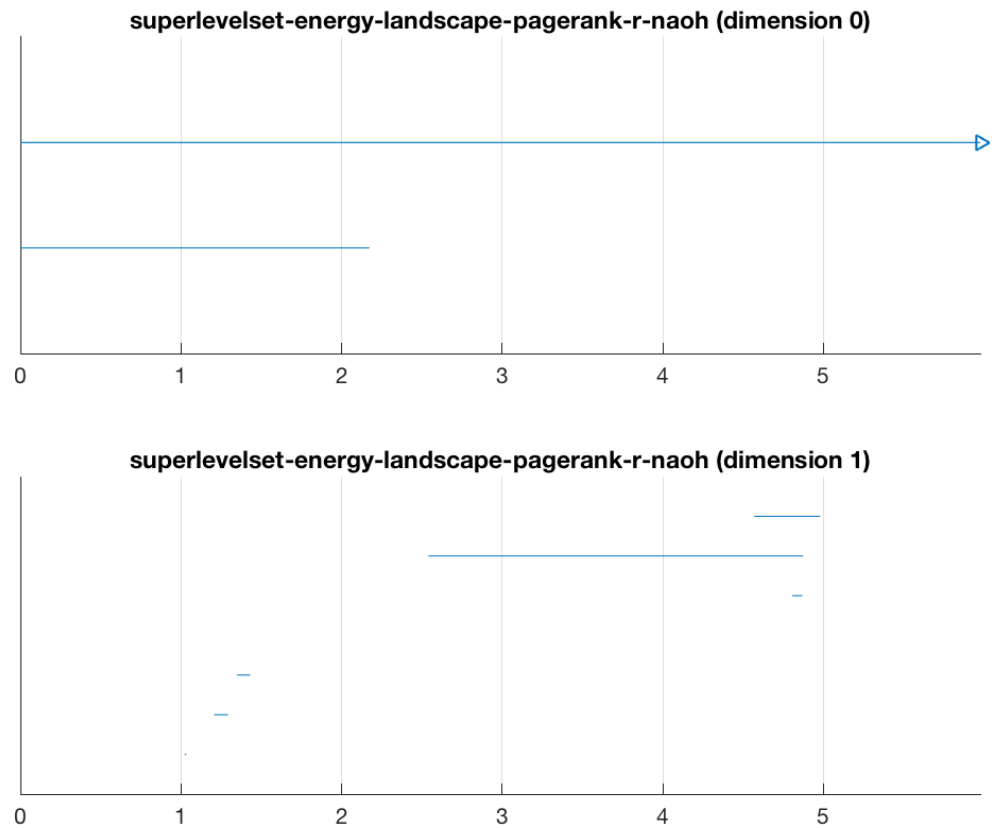




Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

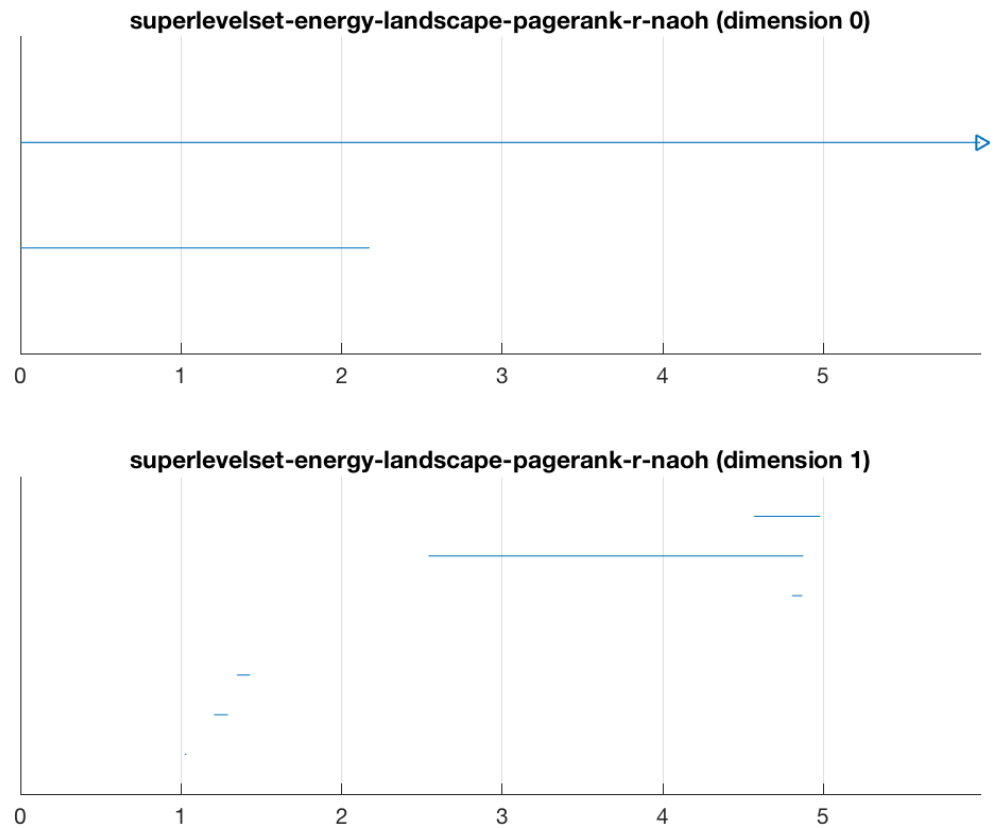
EL (PageRank) of ion-pair formation of Na^+ and OH^-: superlevel sets



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

Sublevel set persistent homology

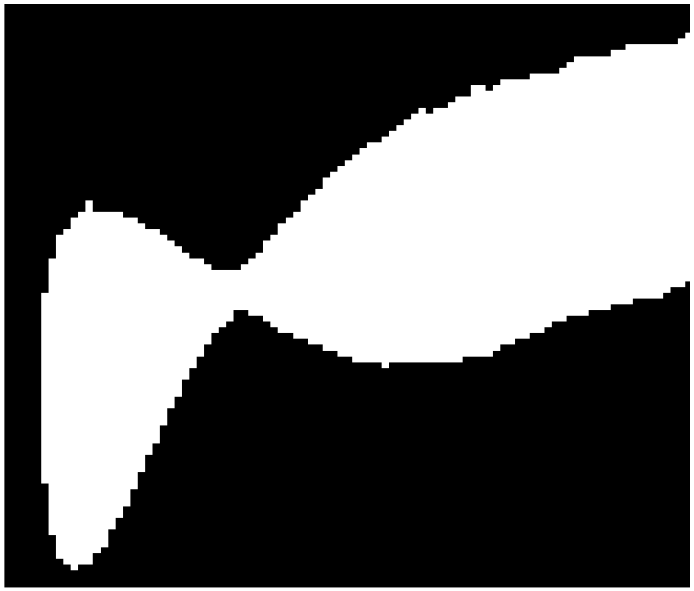
EL (PageRank) of ion-pair formation of Na^+ and OH^-: superlevel sets

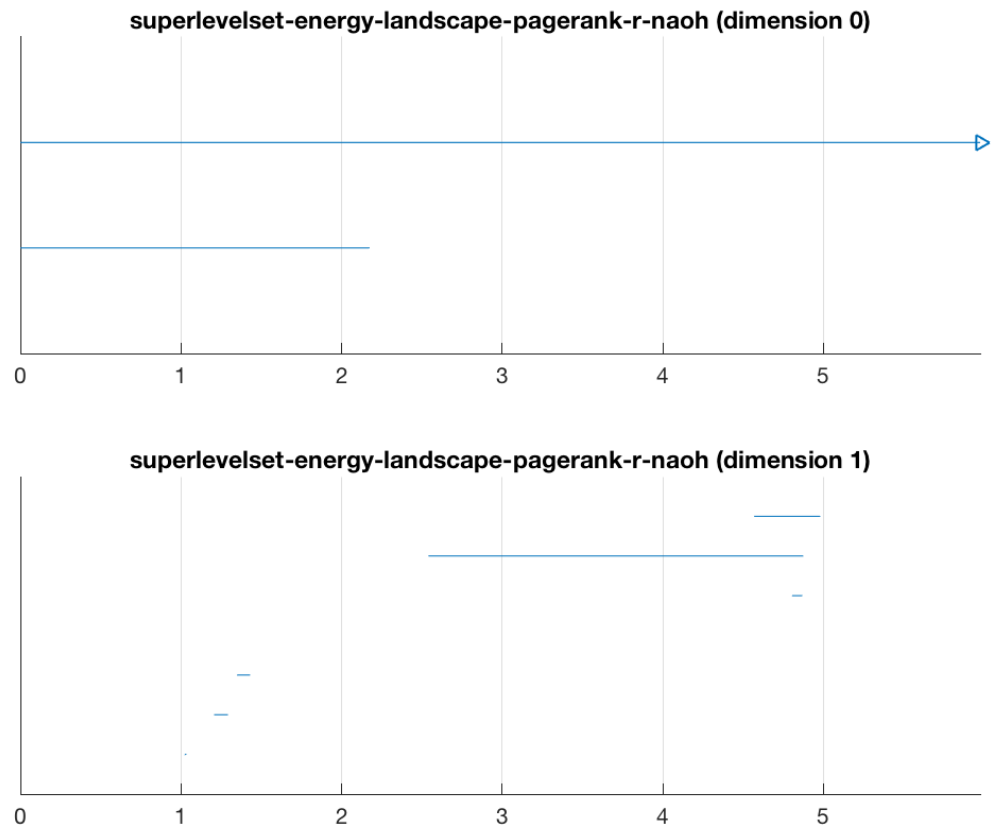


Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

EL (PageRank) of ion-pair formation of Na^+ and OH^-: superlevel sets

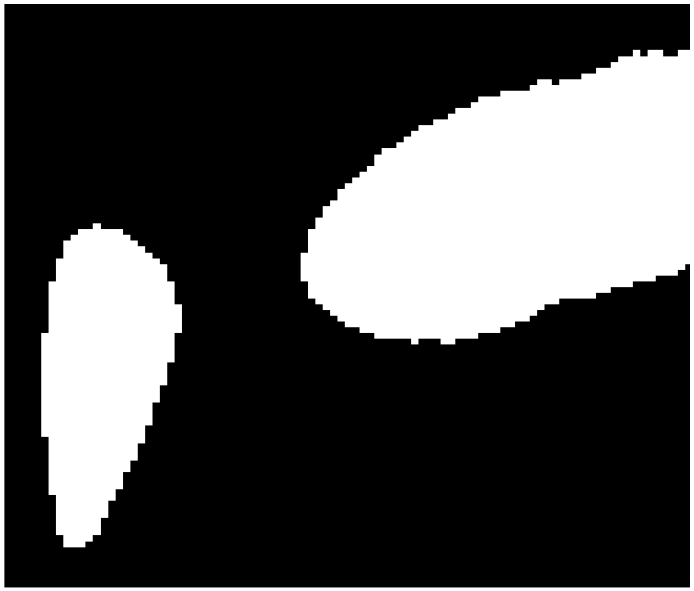


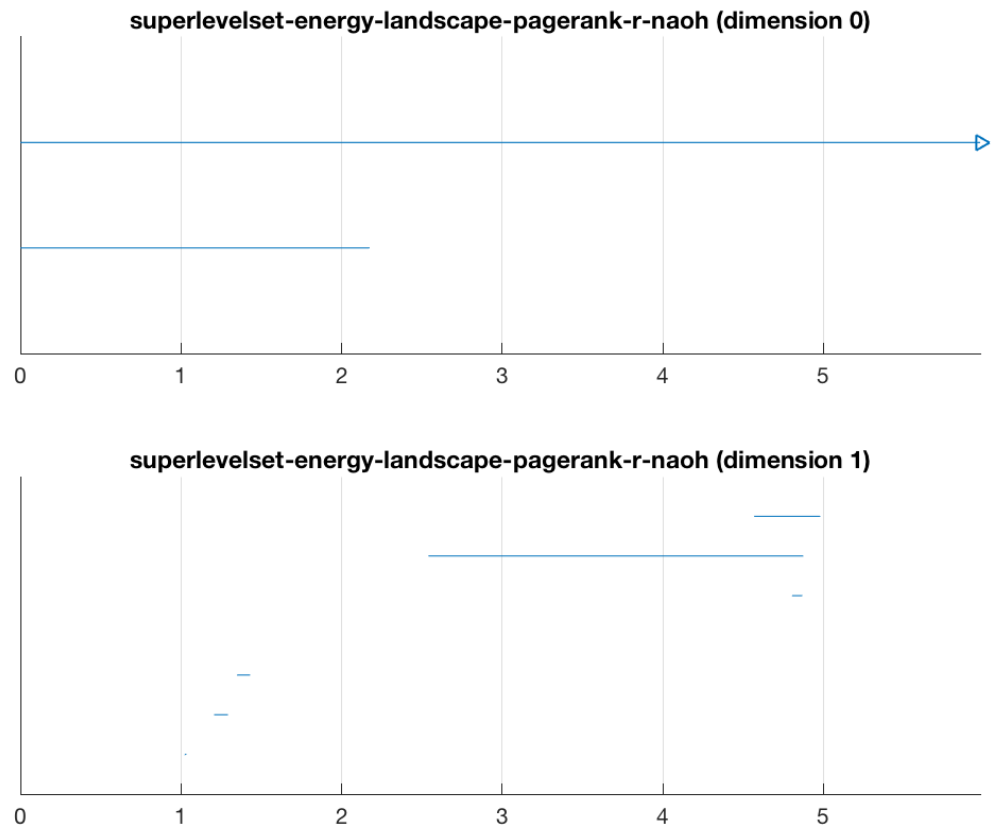


Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

EL (PageRank) of ion-pair formation of Na^+ and OH^-: superlevel sets

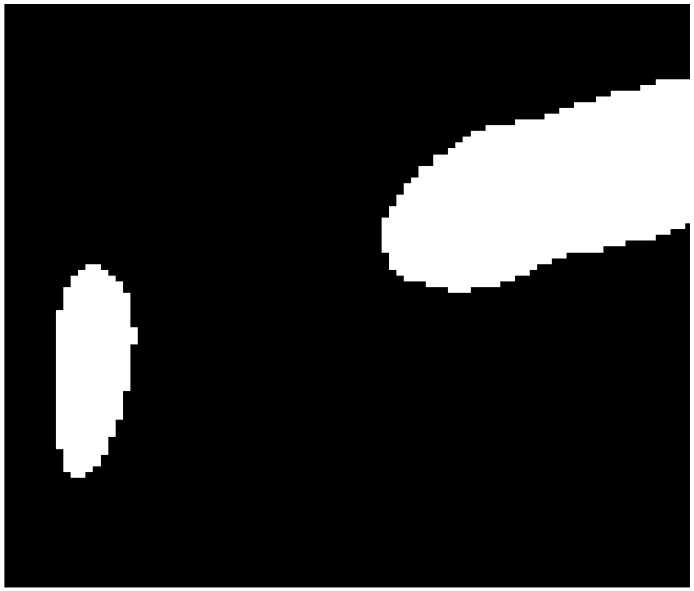


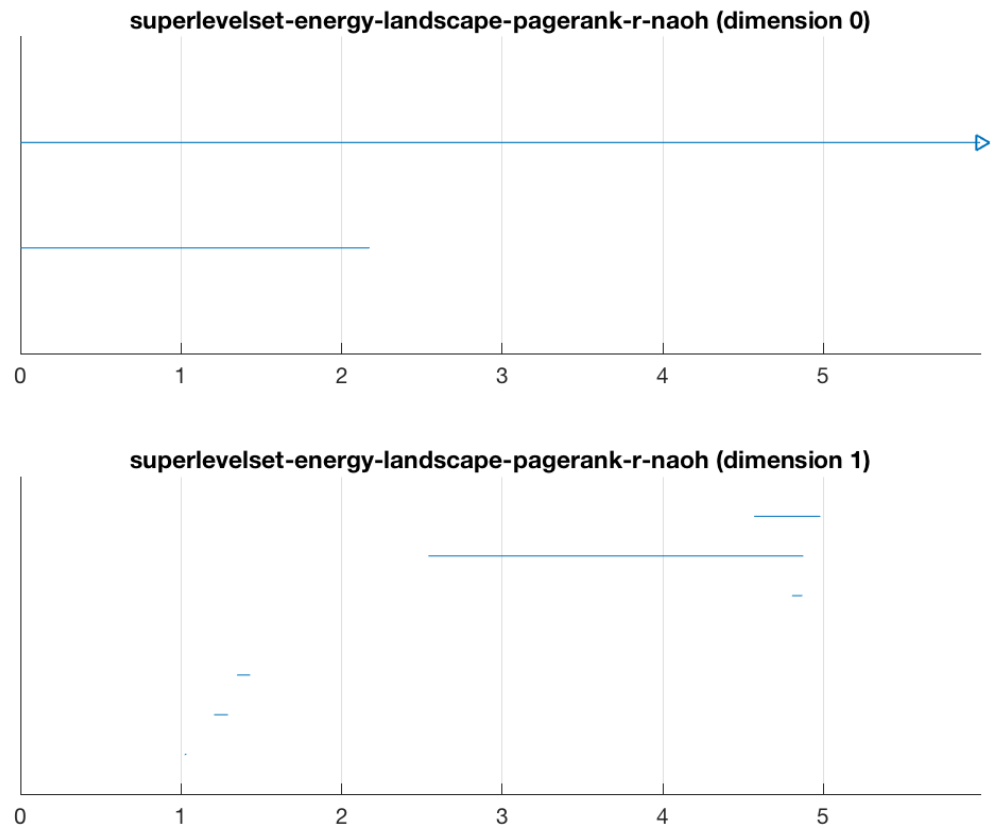


Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

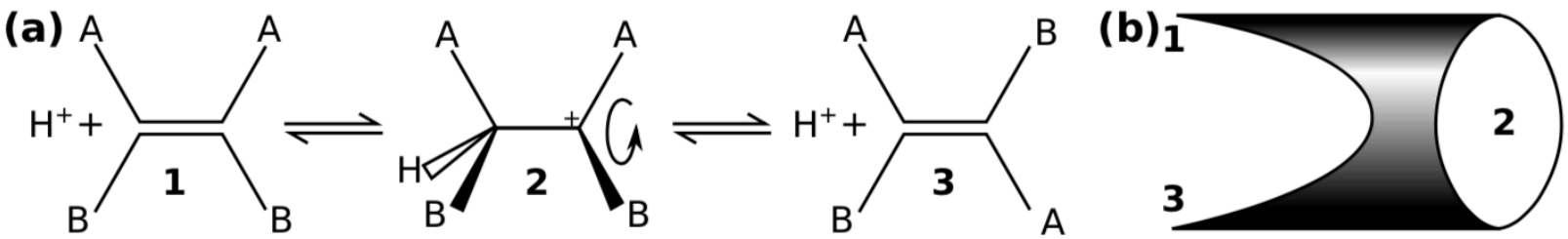
EL (PageRank) of ion-pair formation of Na^+ and OH^-: superlevel sets



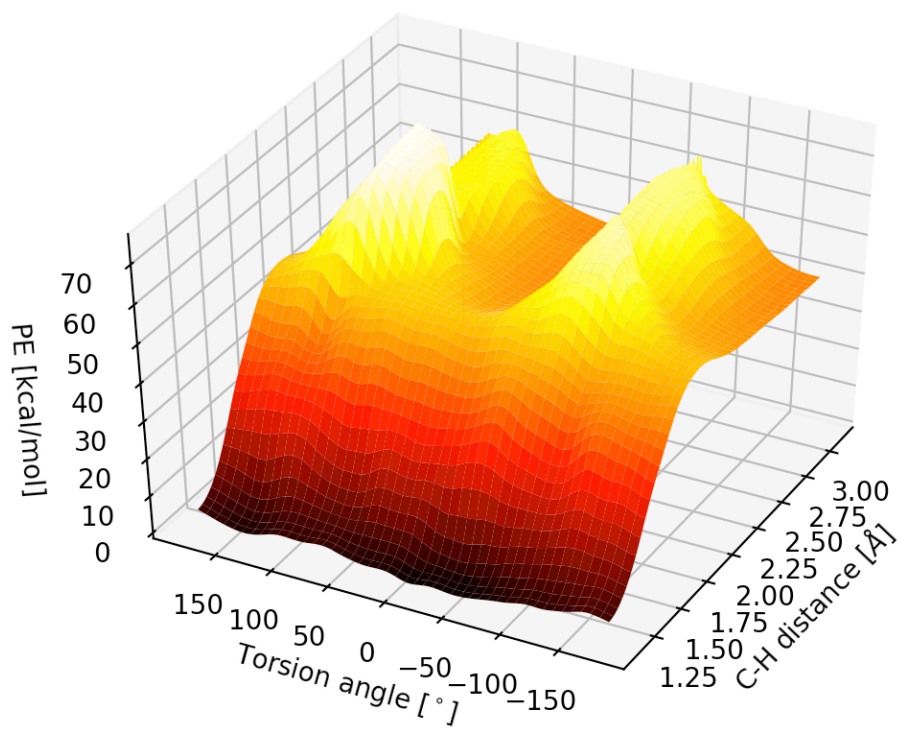


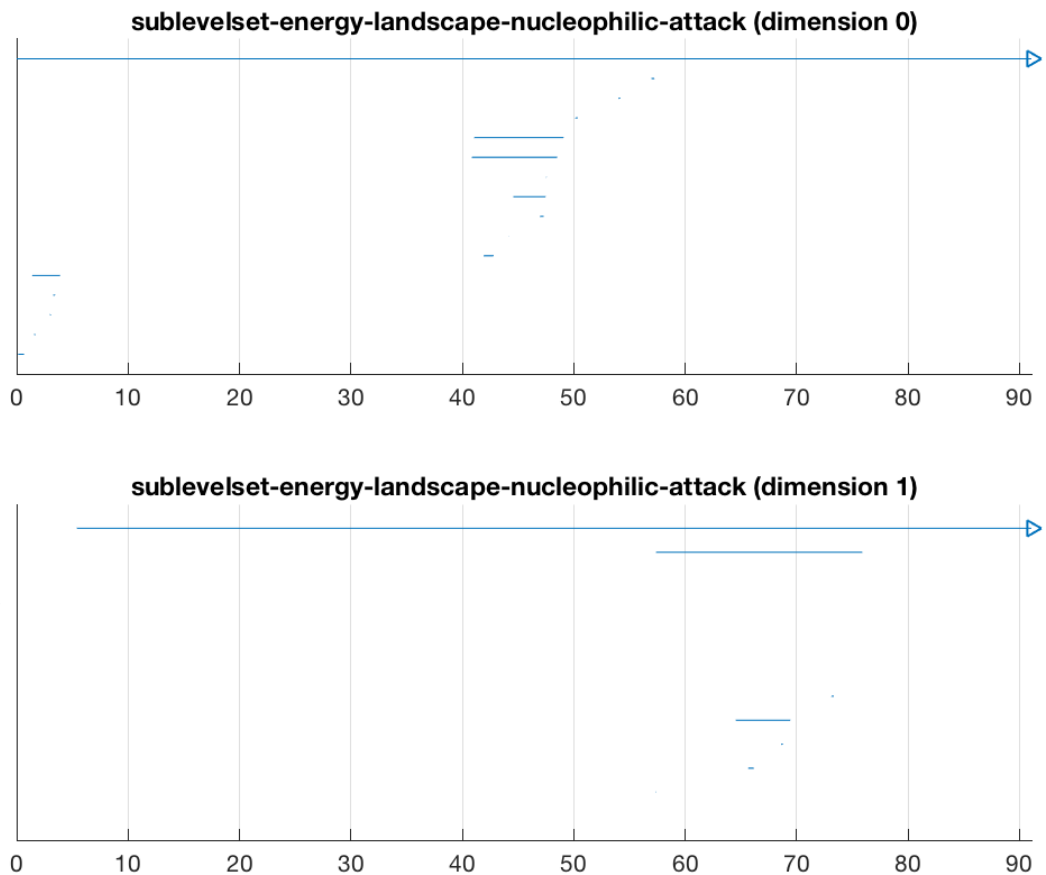
Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology



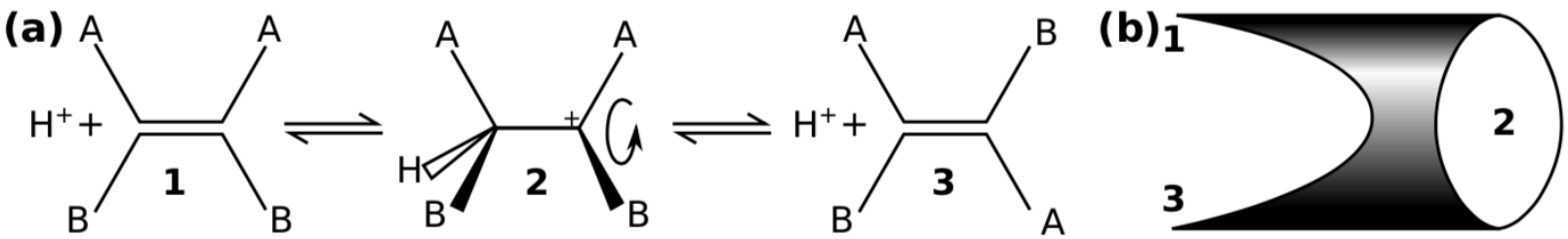
Nucleophilic attack EL: sublevel sets



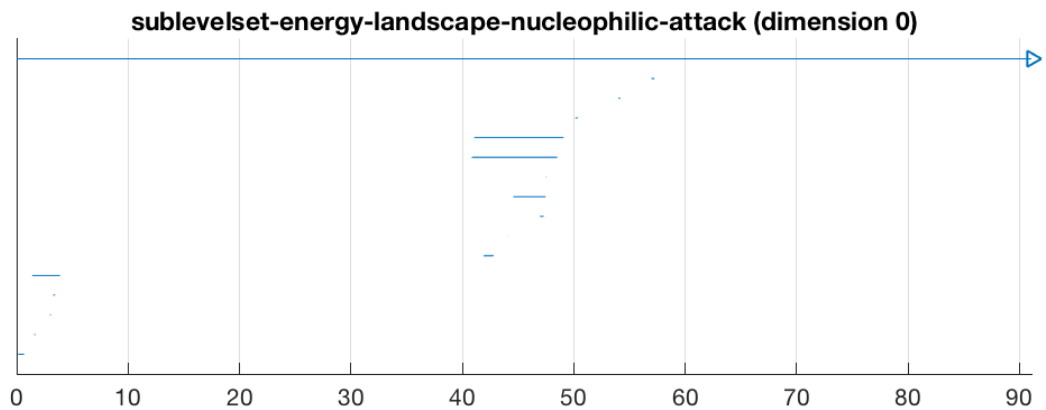


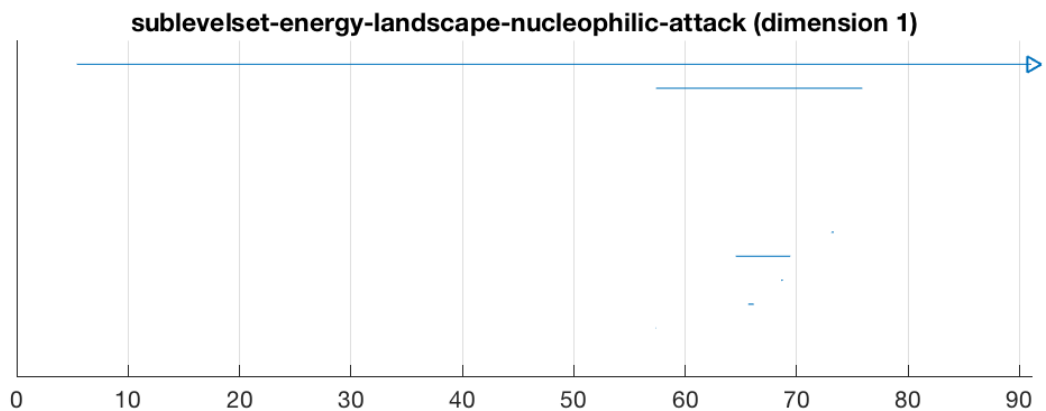
Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology



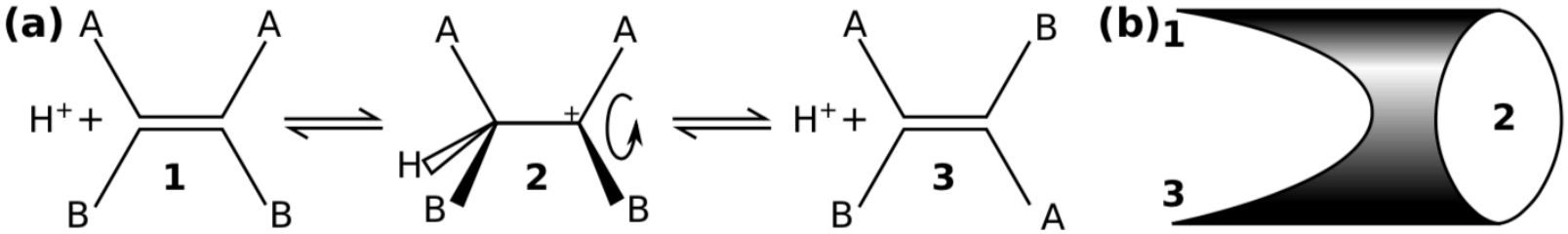
Nucleophilic attack EL: sublevel sets



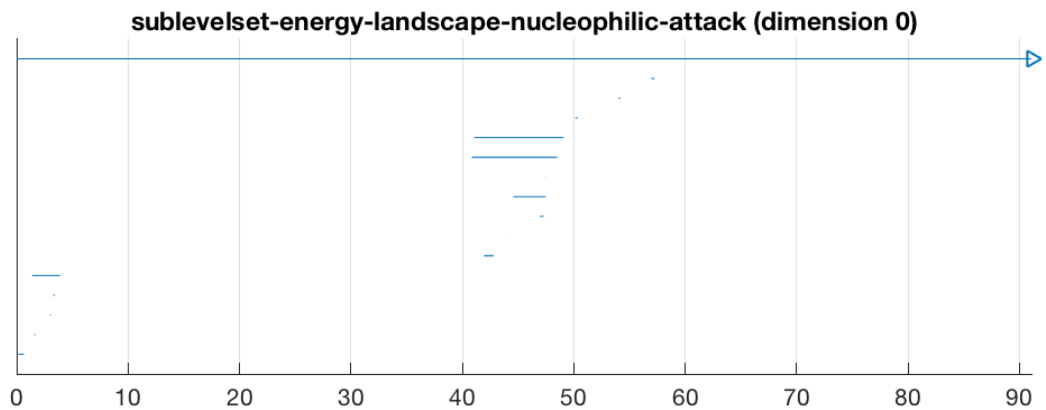


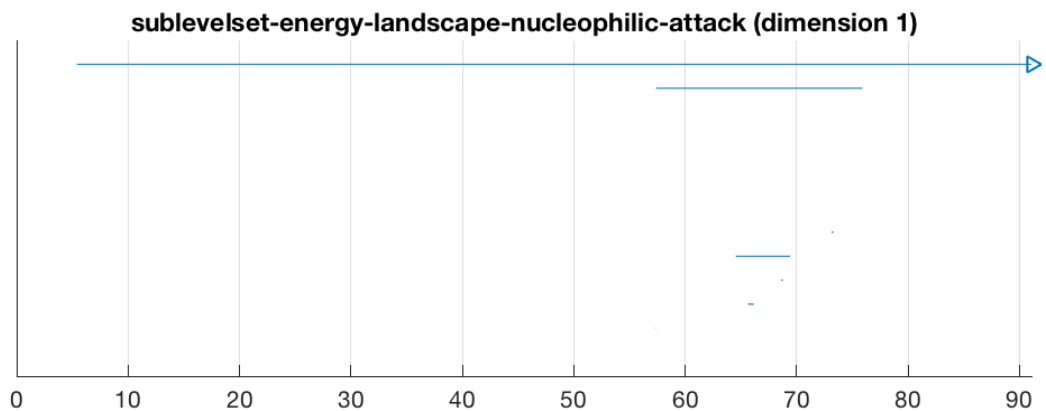
Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology



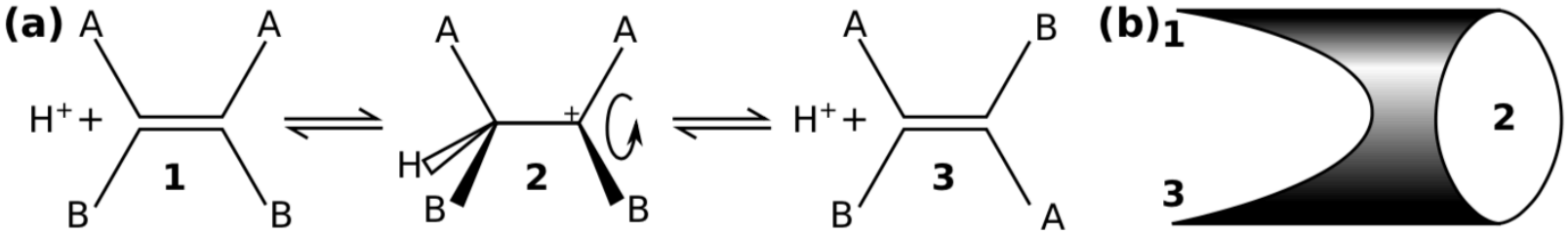
Nucleophilic attack EL: sublevel sets



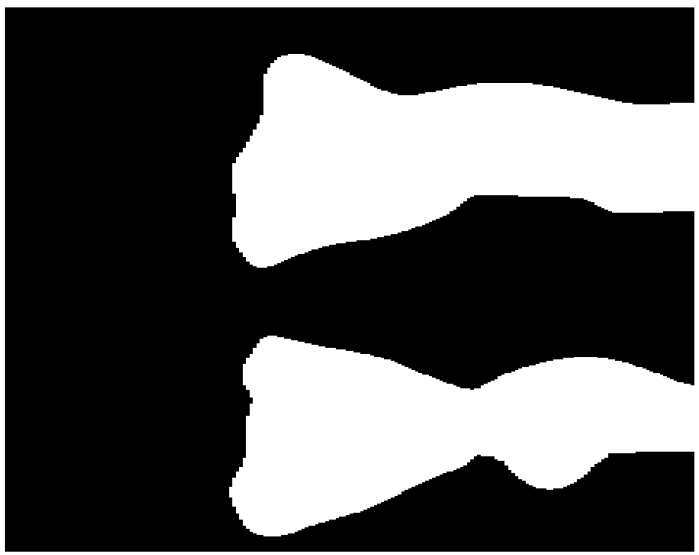


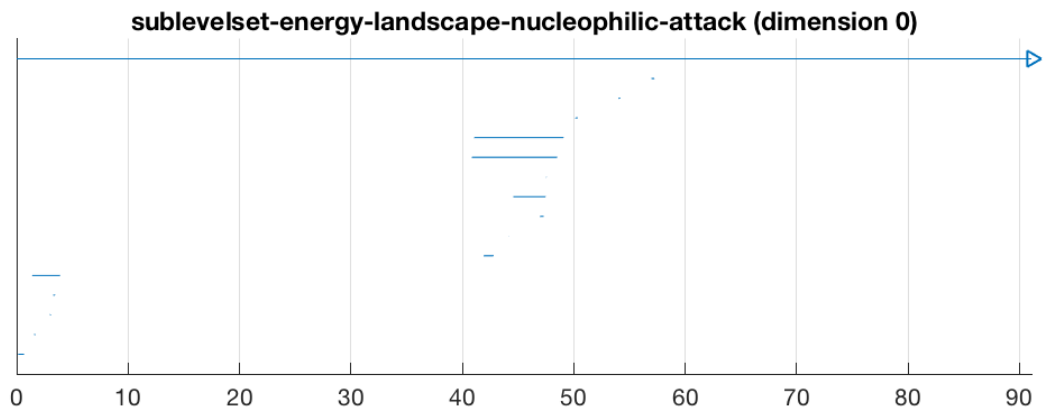
Input: Real-valued function on a space. Output: barcode.

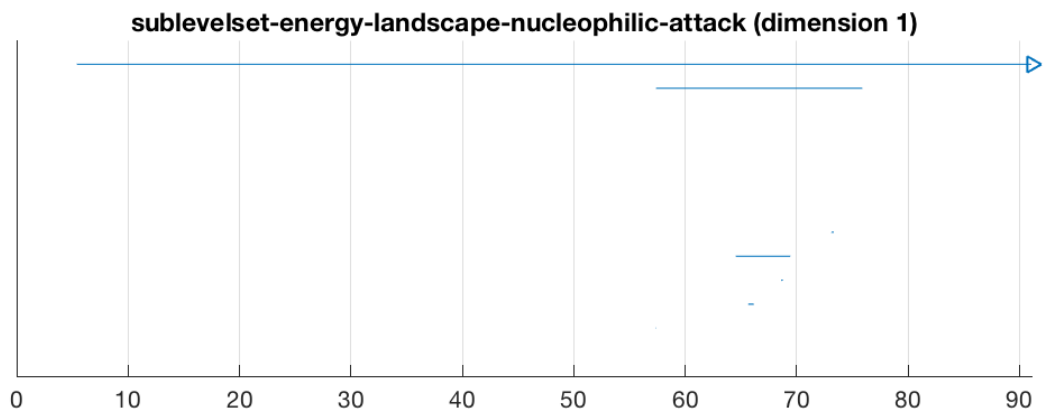
Sublevel set persistent homology



Nucleophilic attack EL: sublevel sets

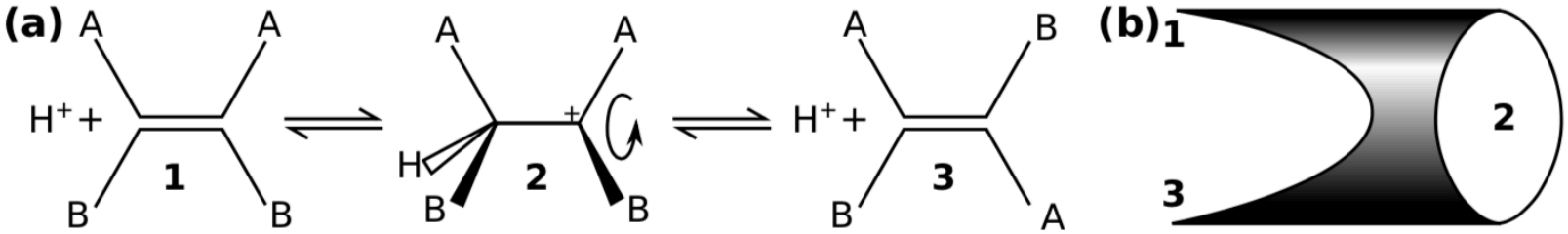




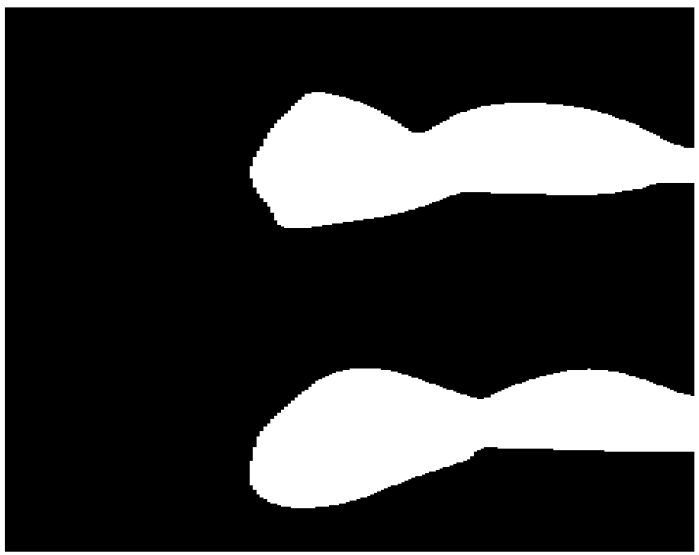


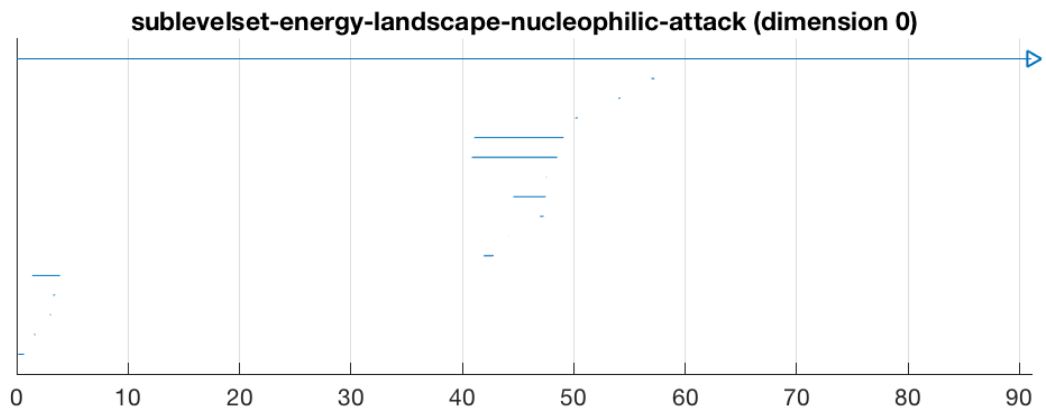
Input: Real-valued function on a space. Output: barcode.

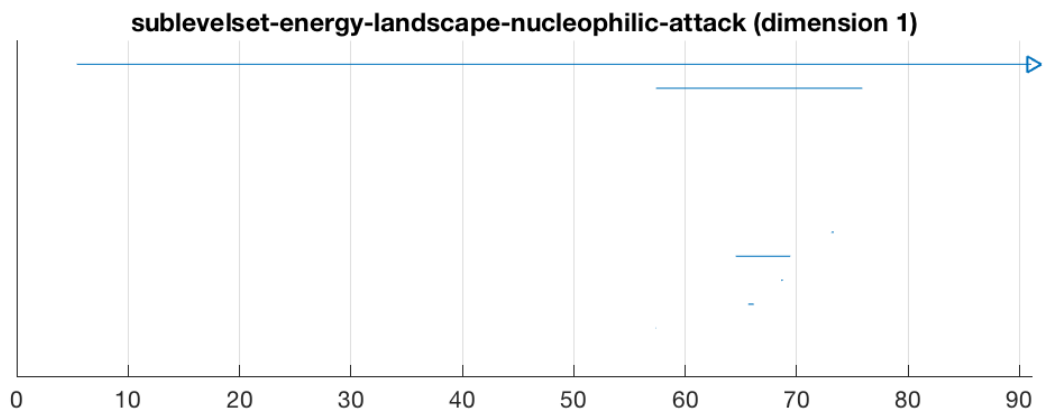
Sublevel set persistent homology



Nucleophilic attack EL: sublevel sets

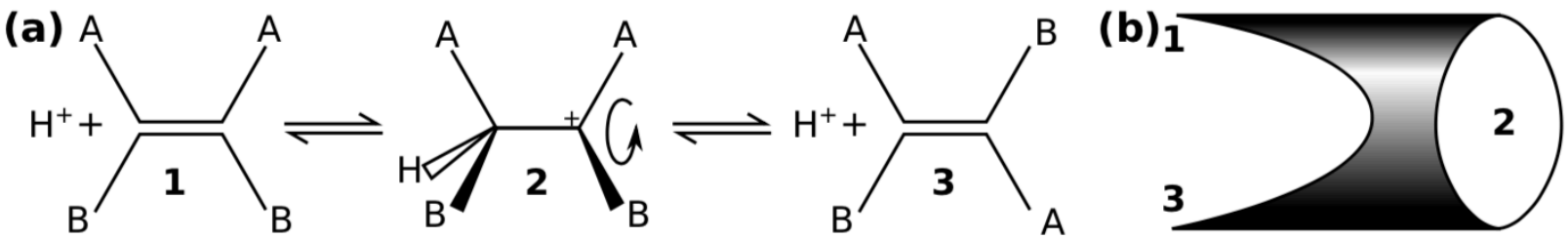




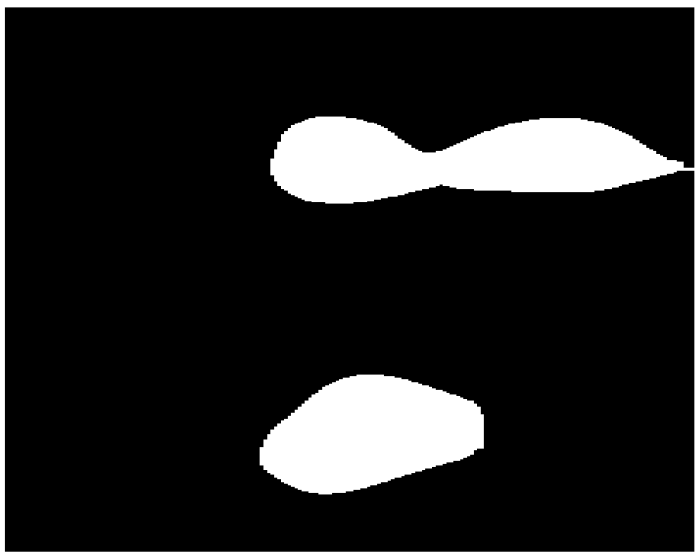


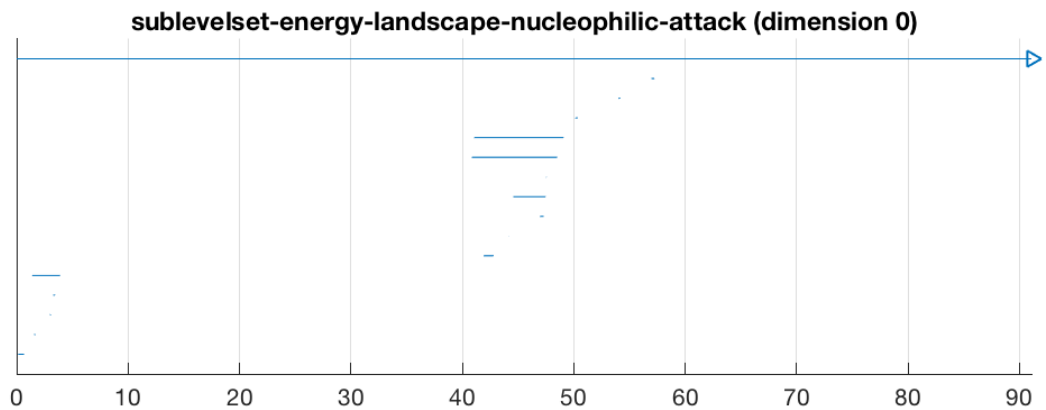
Input: Real-valued function on a space. Output: barcode.

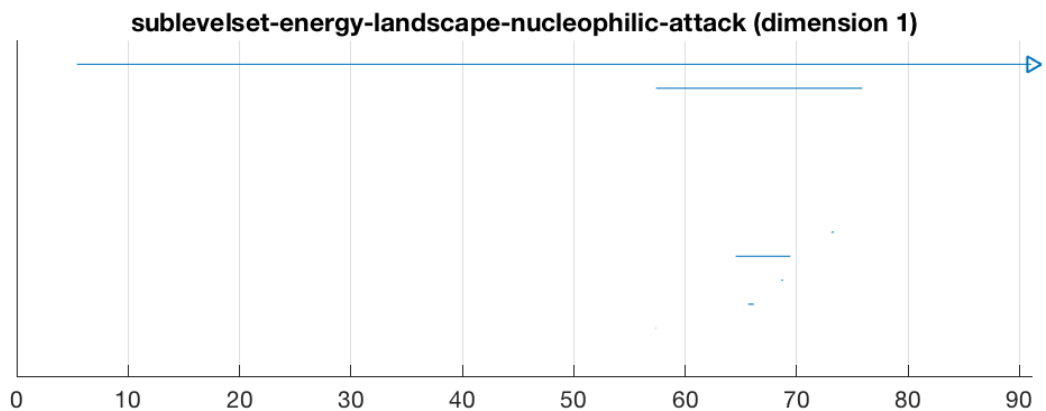
Sublevel set persistent homology



Nucleophilic attack EL: sublevel sets

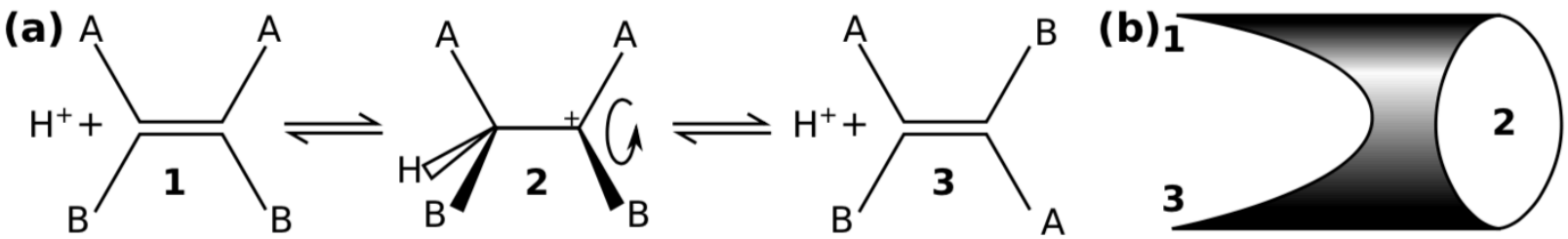




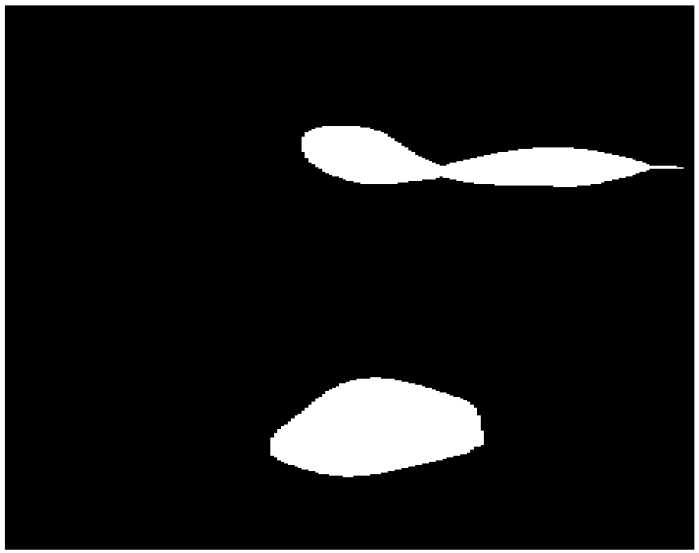


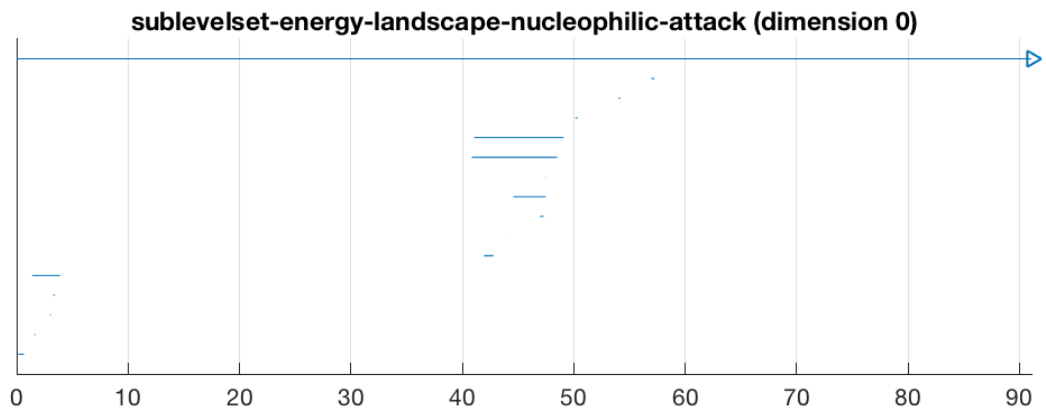
Input: Real-valued function on a space. Output: barcode.

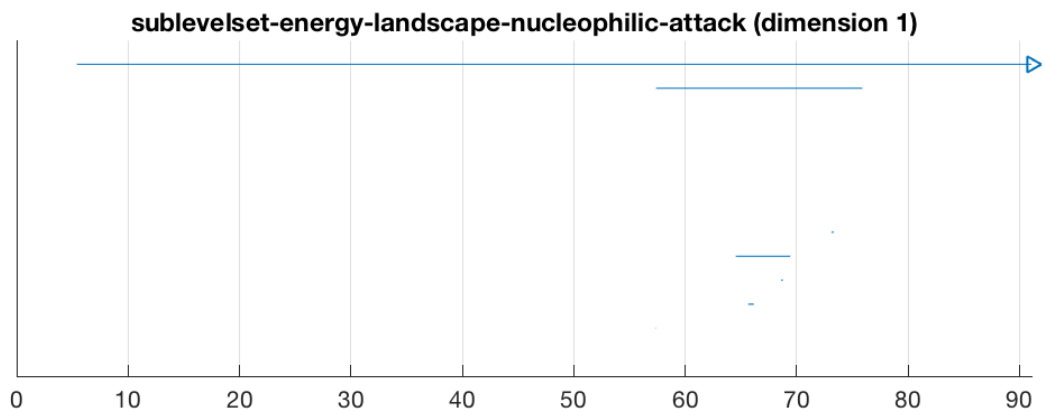
Sublevel set persistent homology



Nucleophilic attack EL: sublevel sets

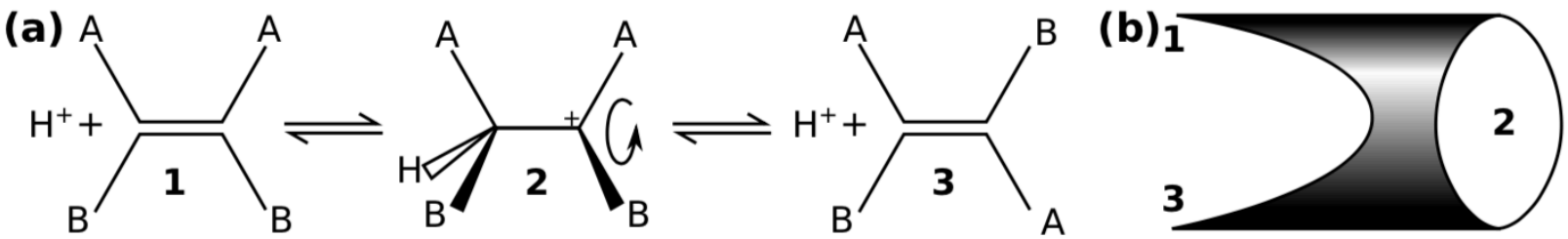




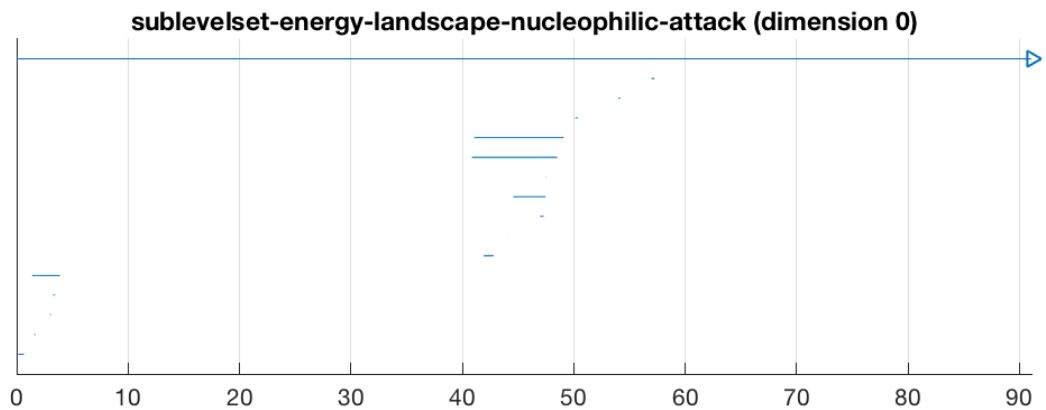


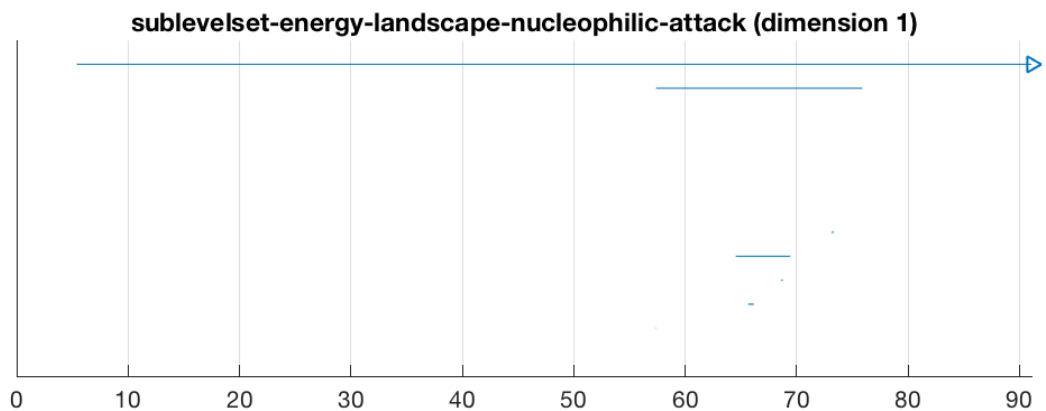
Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology



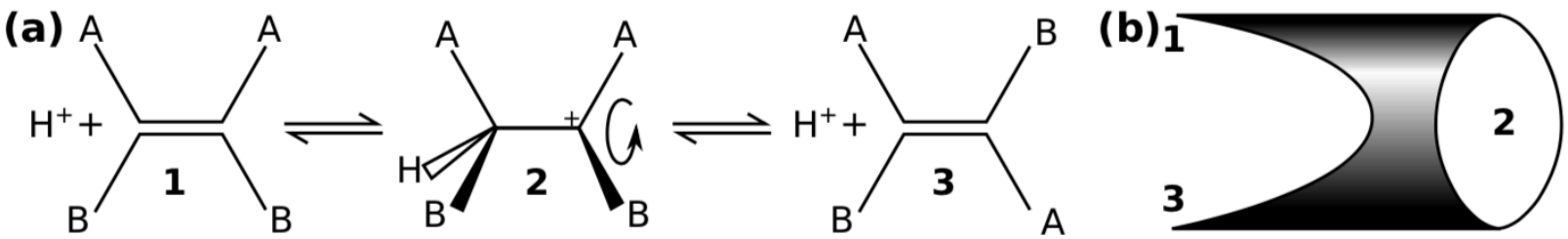
Nucleophilic attack EL: sublevel sets



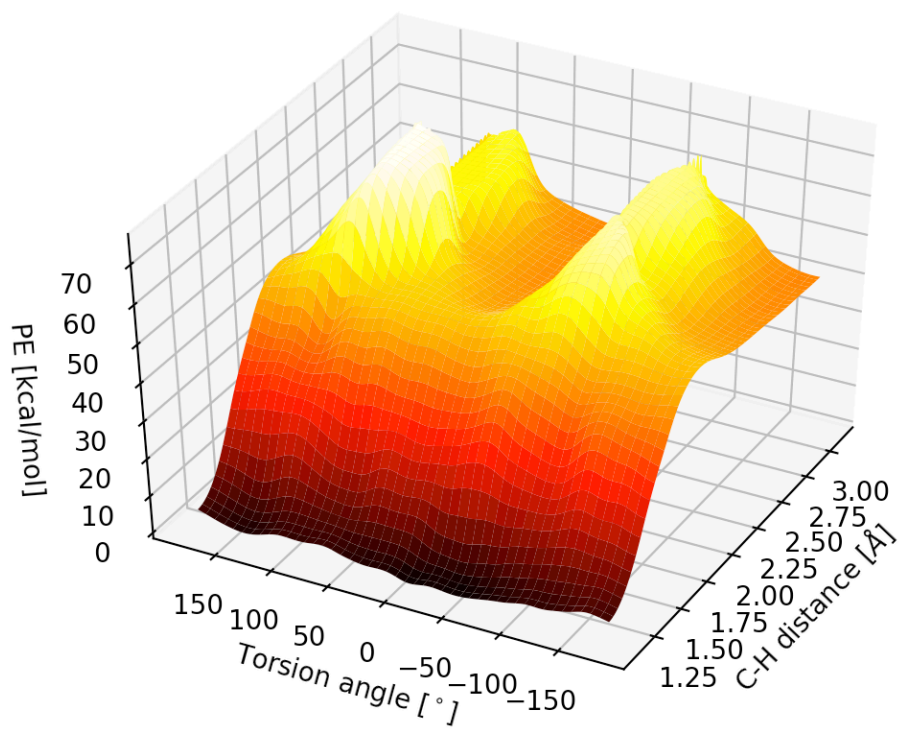


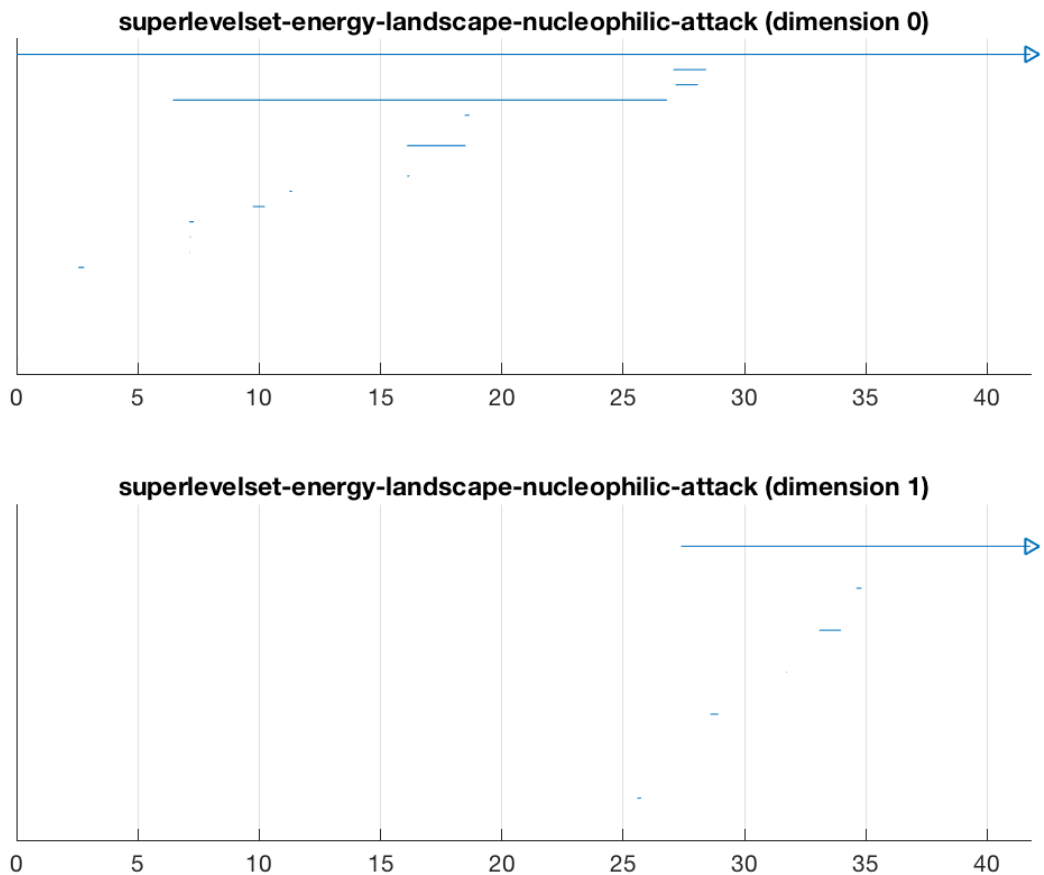
Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology



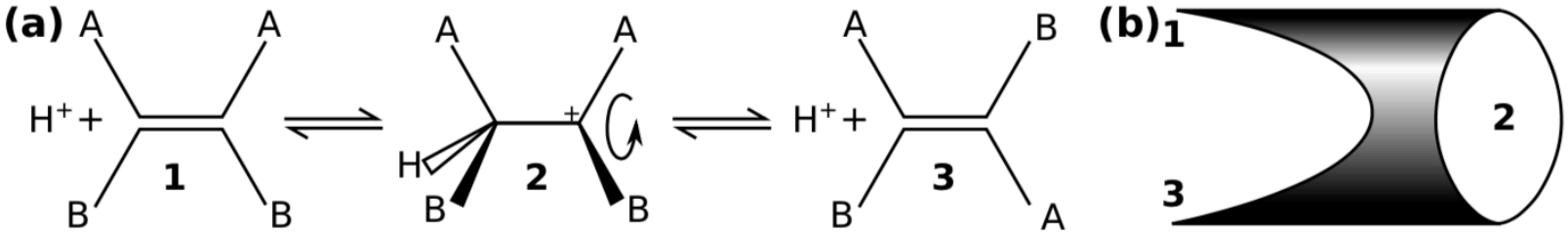
Nucleophilic attack EL: superlevel sets



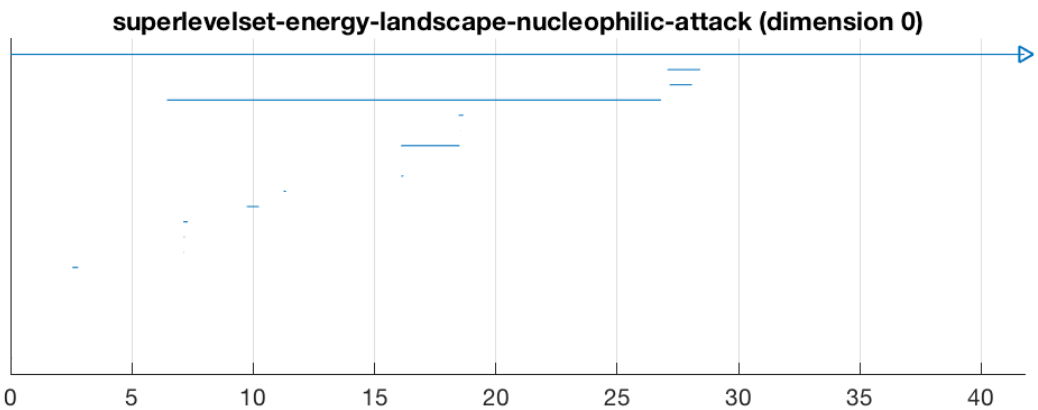


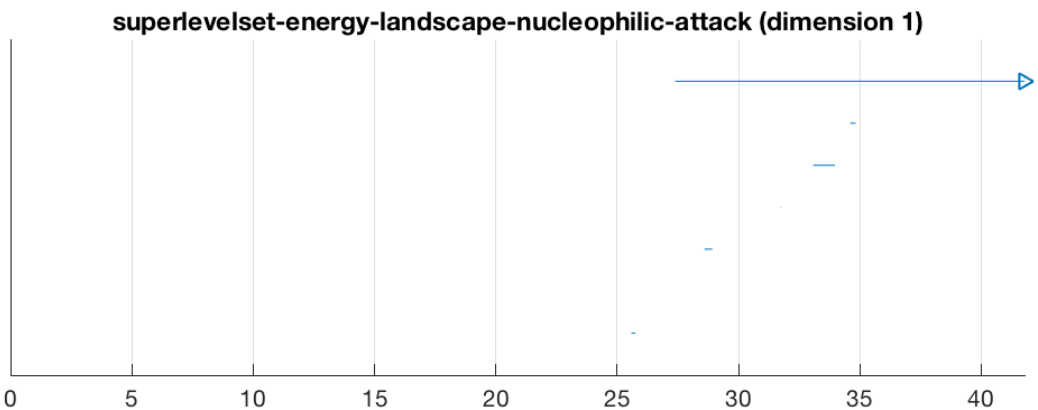
Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology



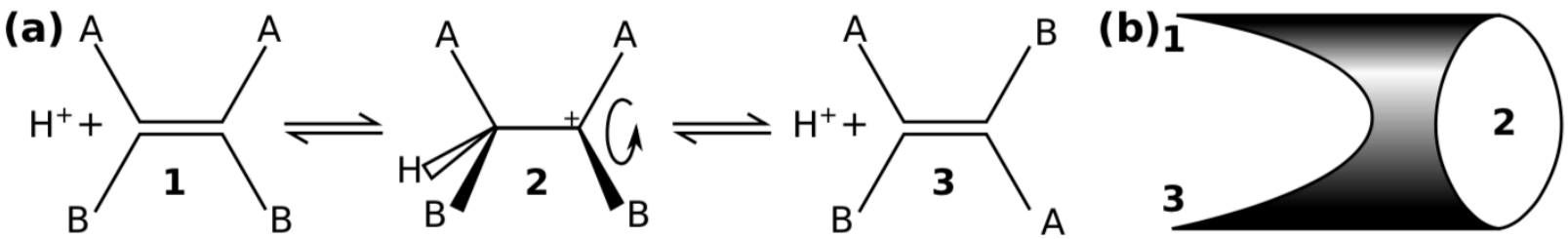
Nucleophilic attack EL: superlevel sets



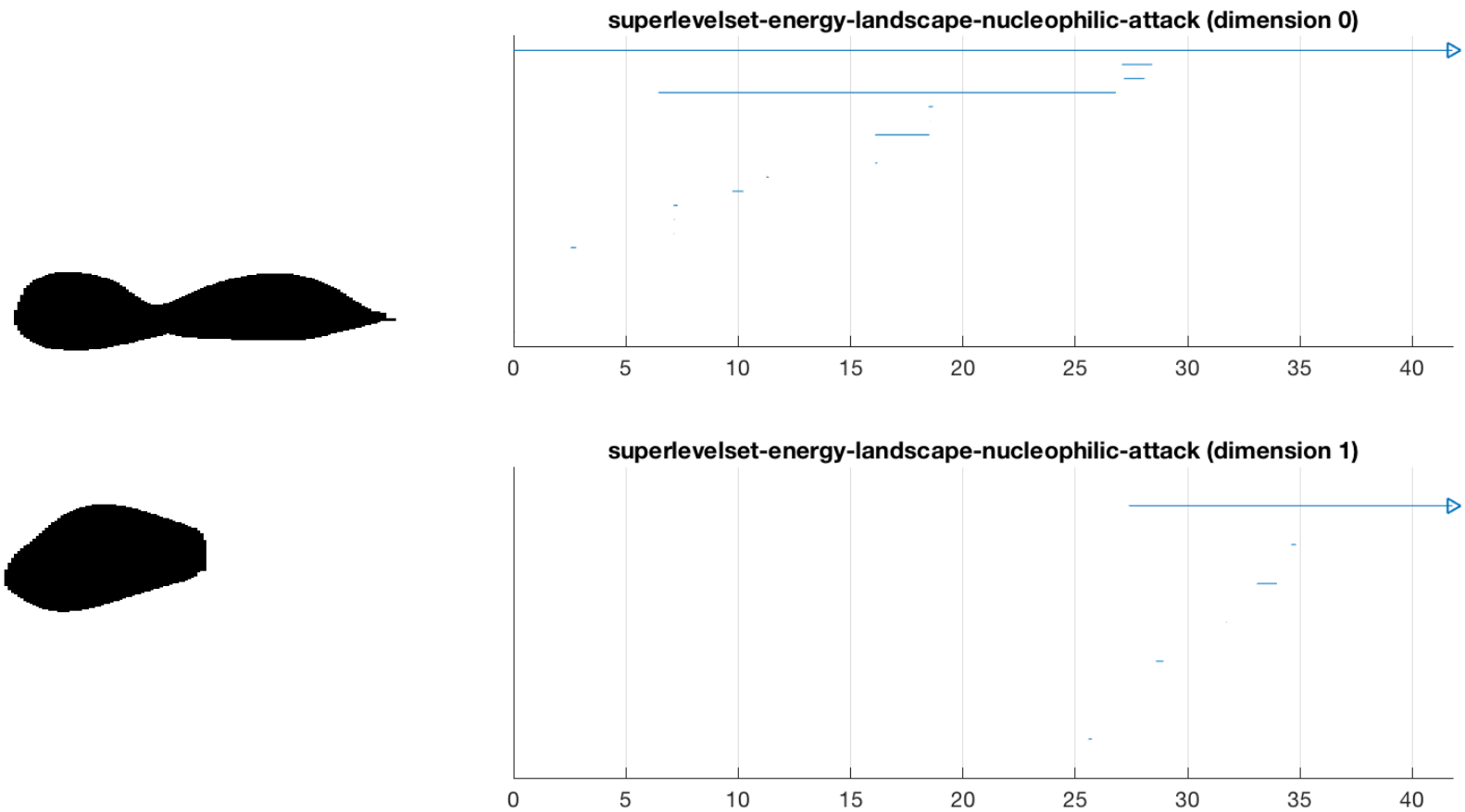


Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

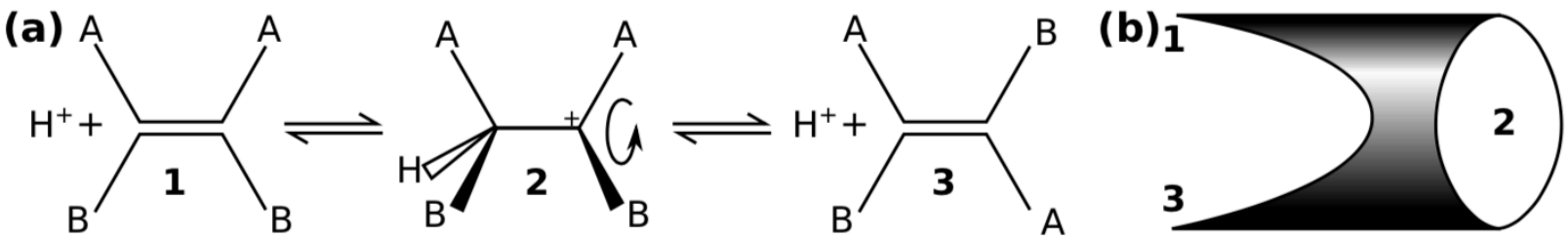


Nucleophilic attack EL: superlevel sets

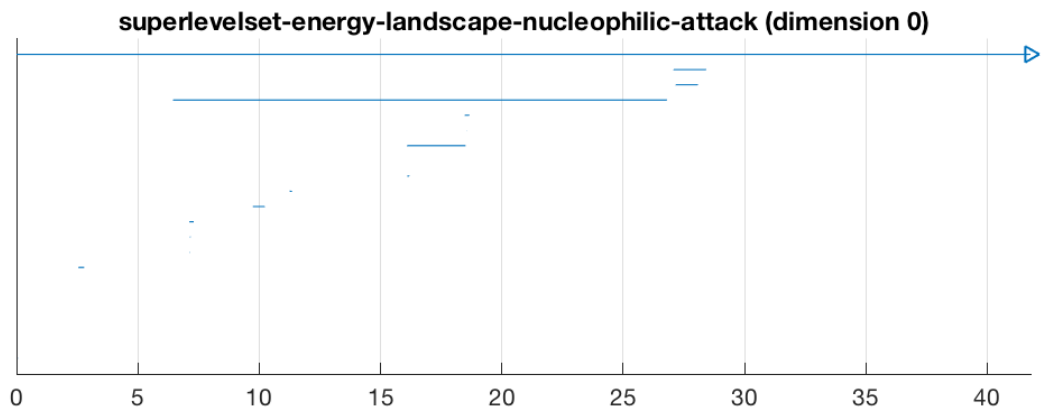


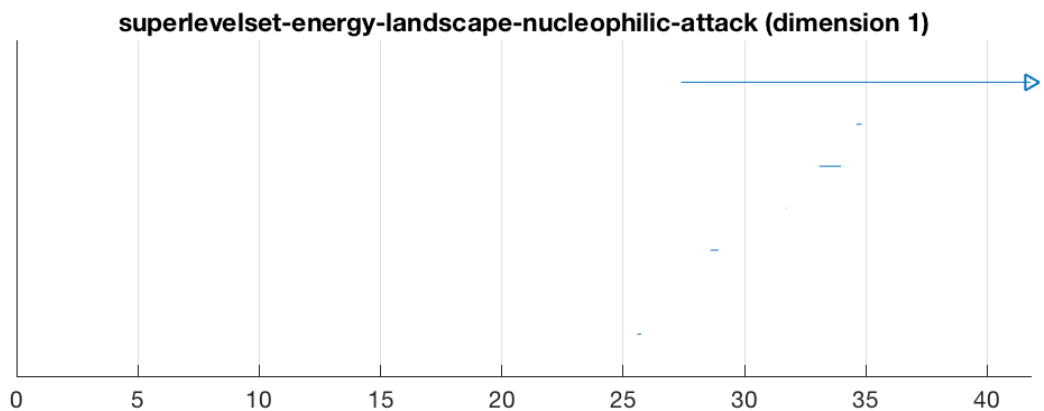
Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology



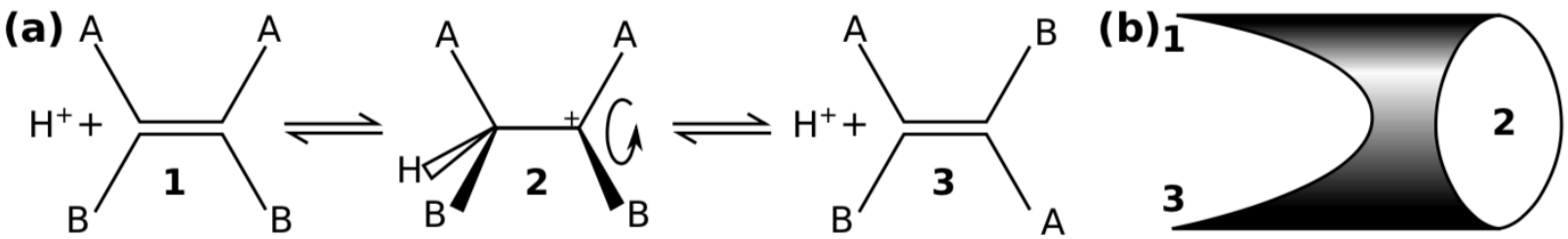
Nucleophilic attack EL: superlevel sets



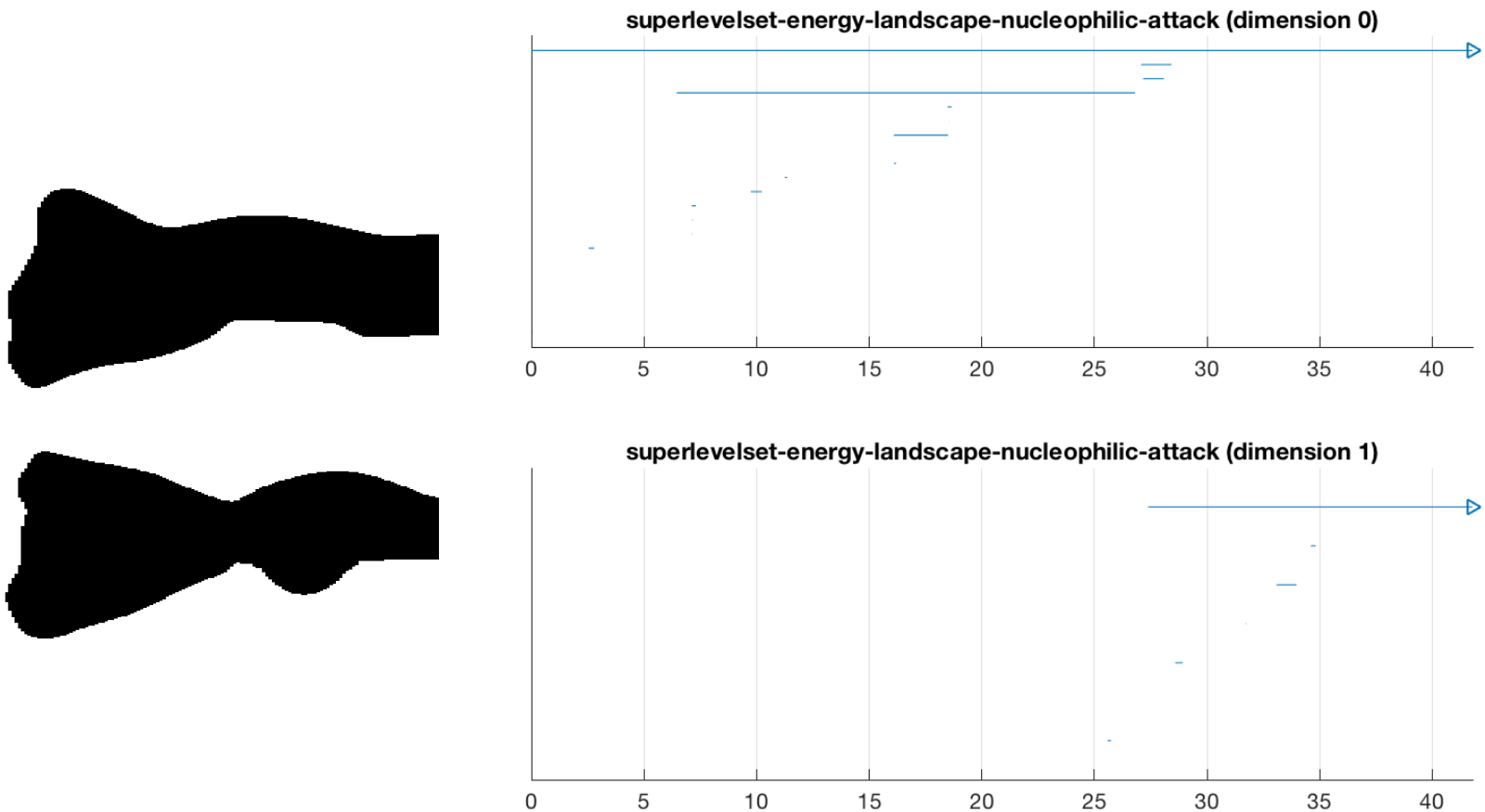


Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

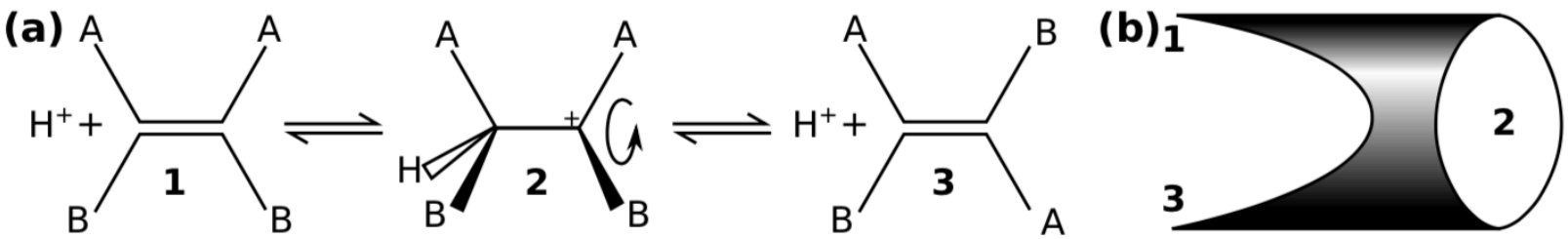


Nucleophilic attack EL: superlevel sets

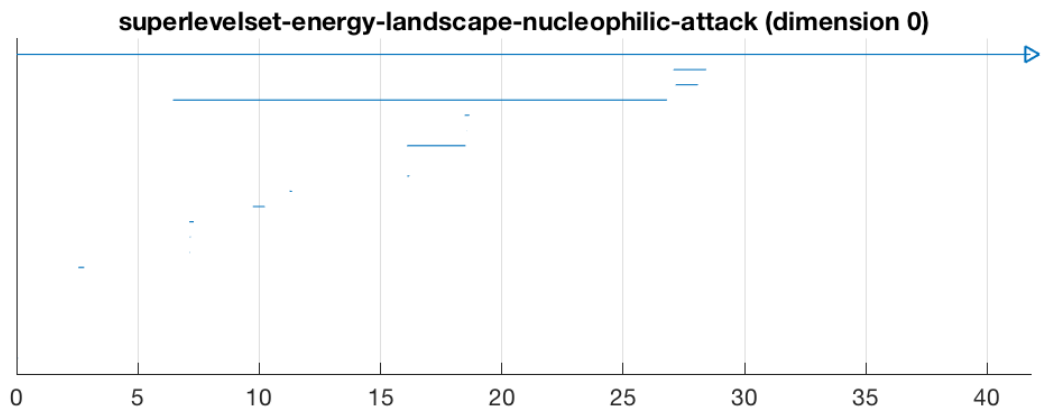


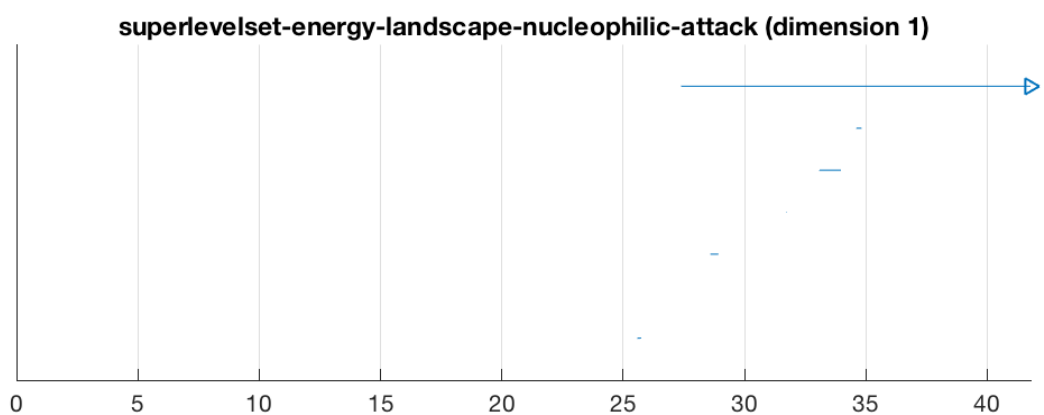
Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology



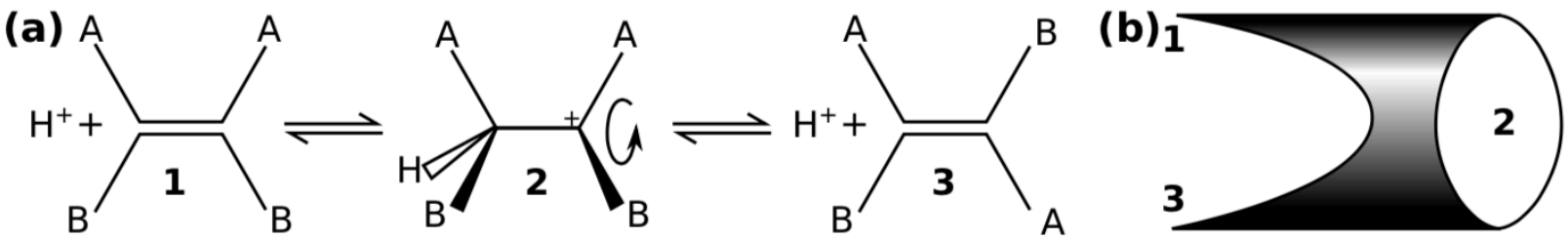
Nucleophilic attack EL: superlevel sets



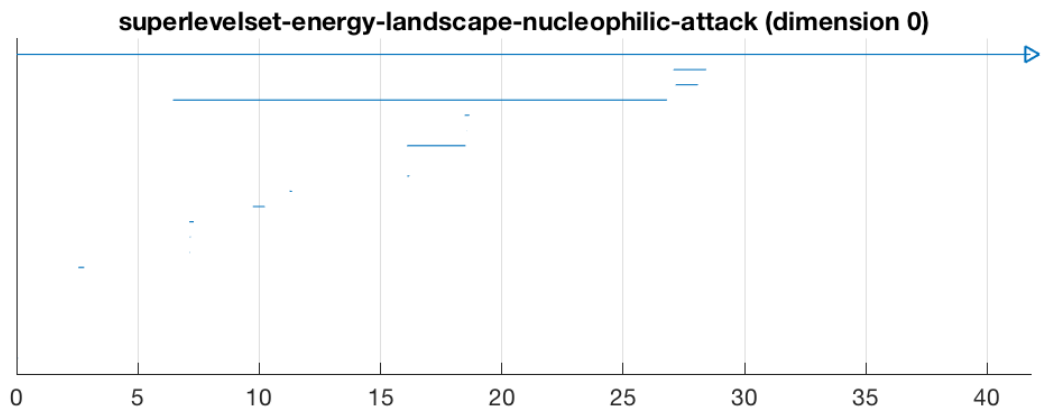


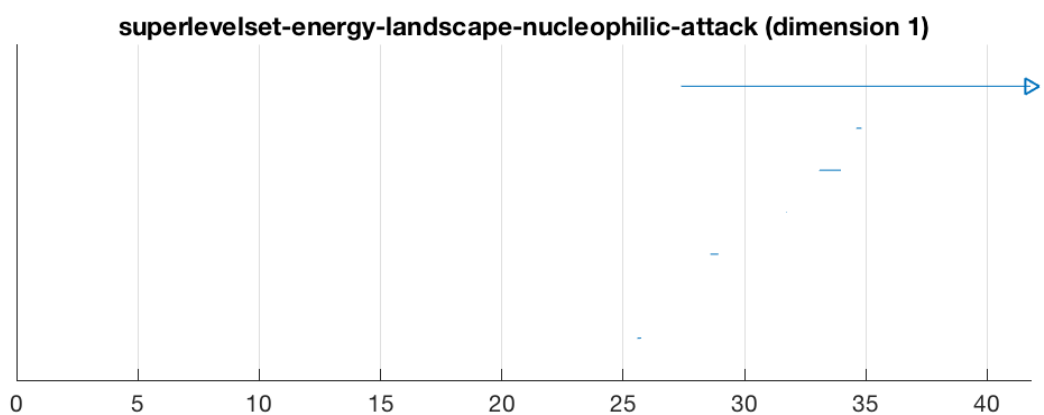
Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology



Nucleophilic attack EL: superlevel sets

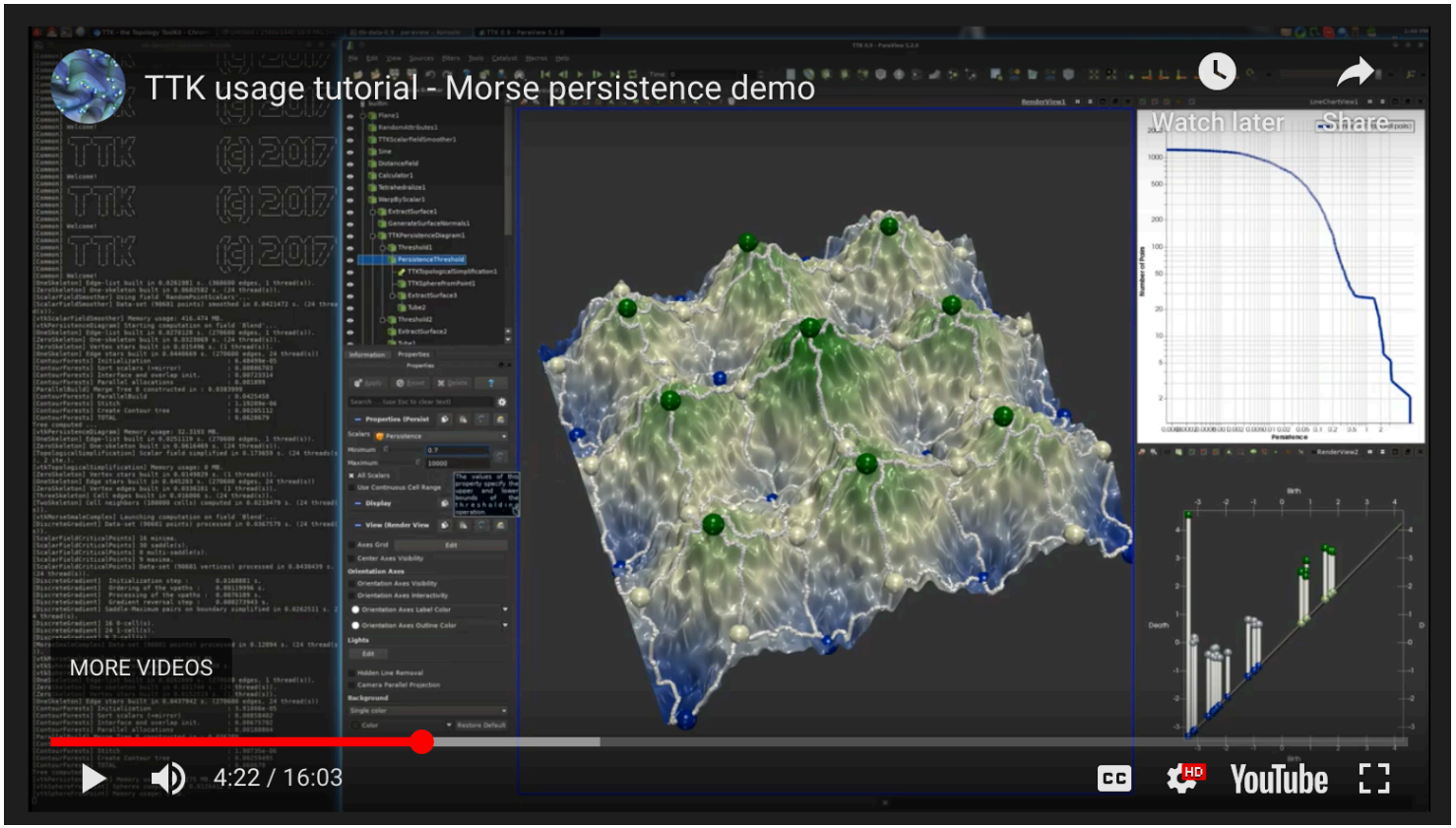




Input: Real-valued function on a space. Output: barcode.

Sublevel set persistent homology

Question: How does sublevel set persistent homology relate to Morse-Smale complexes (and TTK)?

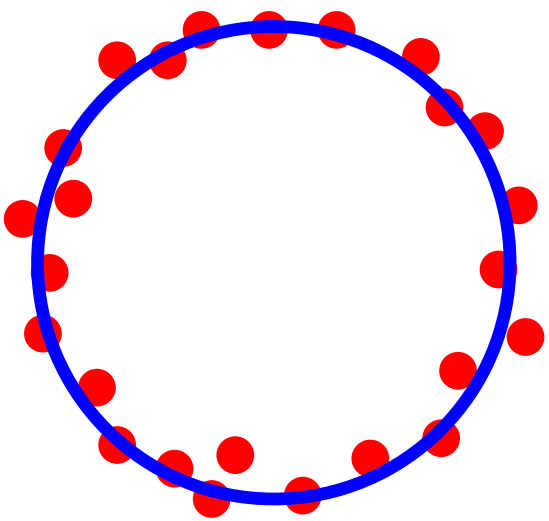


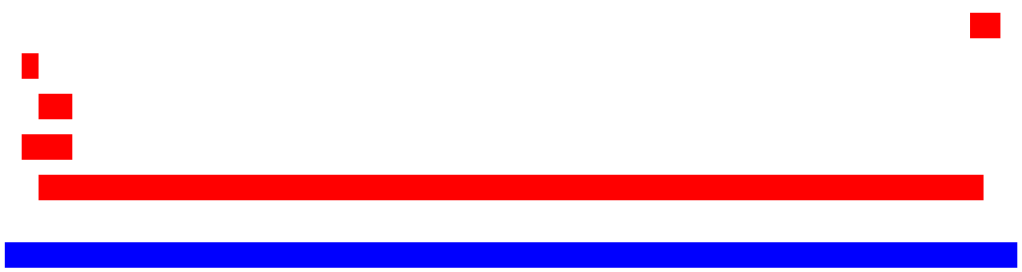
Answer: Sublevel set persistent homology encodes most of the topology, but little of the geometry, in the Morse-Smale complex.

Sublevel set persistent homology

- <u>Stability Theorem.</u> If f and g are real-valued functions on a space, X, and U_f and U_g are the barcodes of the associated sublevel set filtrations, then

$$d_B\left(U_f, U_g\right) \leq ||f - g||_\infty = \max_{x \in X} |f(x) - g(x)|$$



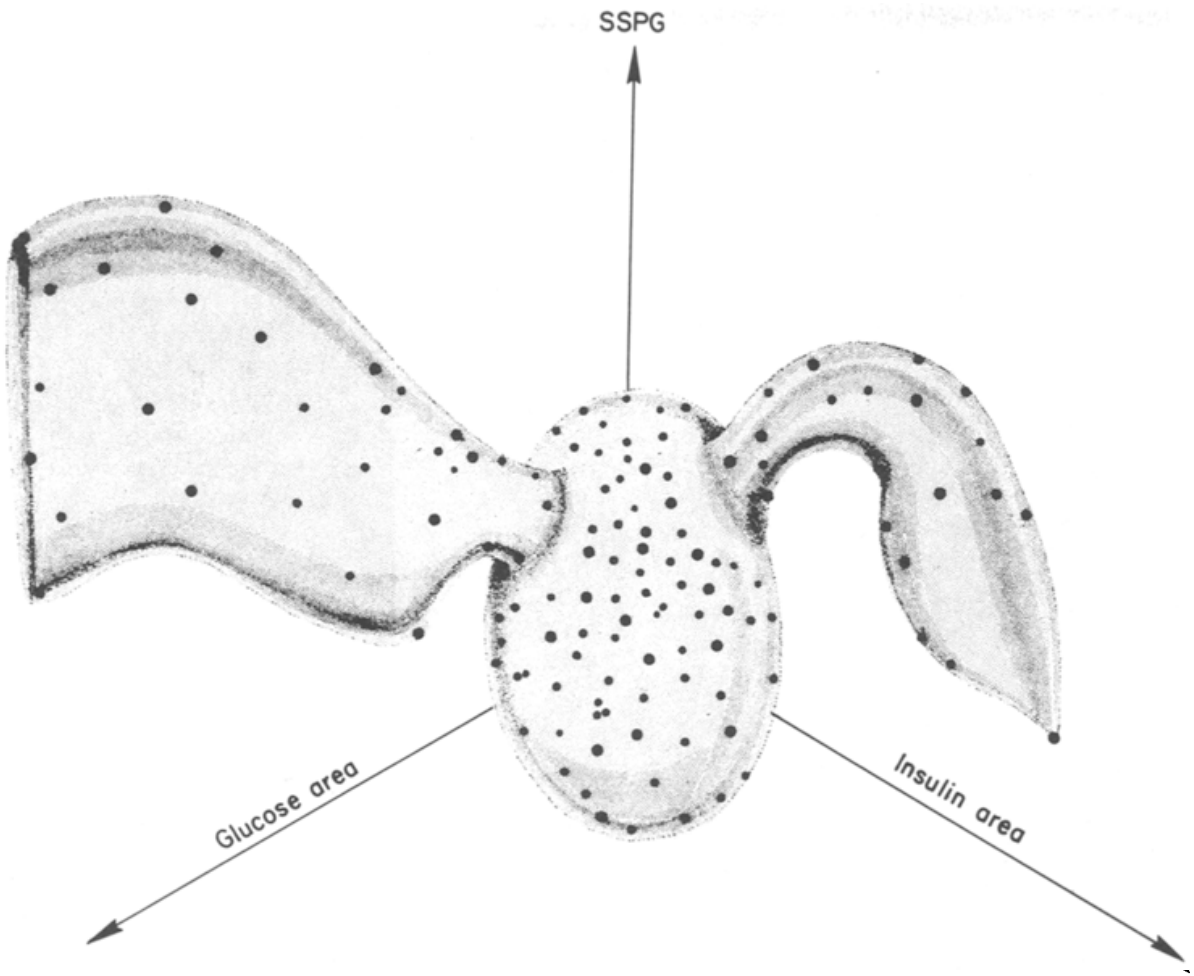


An Introduction to Persistent Homology

1. Review of topology and homology
2. Introduction to persistent homology
3. Sublevel set persistent homology
4. Persistent homology applied to point cloud data

Datasets have shapes

Example: Diabetes study
145 points in 5-dimensional space

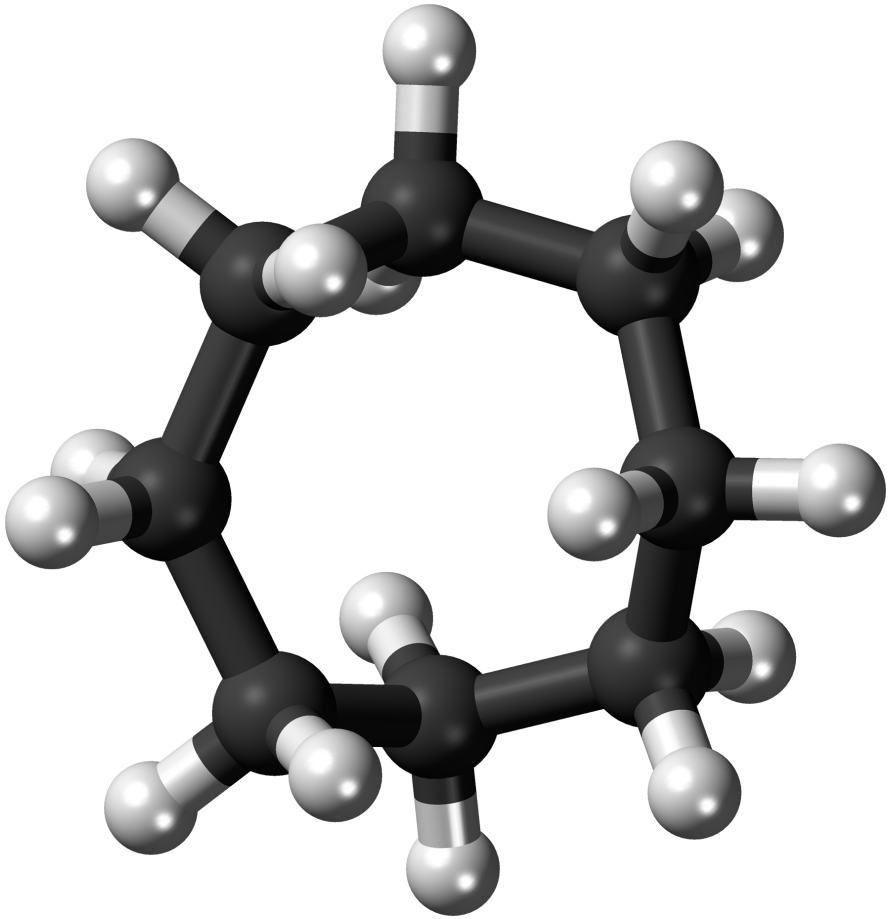


An attempt to define the nature of chemical diabetes using a multidimensional analysis by G. M. Reaven and R. G. Miller, 1979

Datasets have shapes

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

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

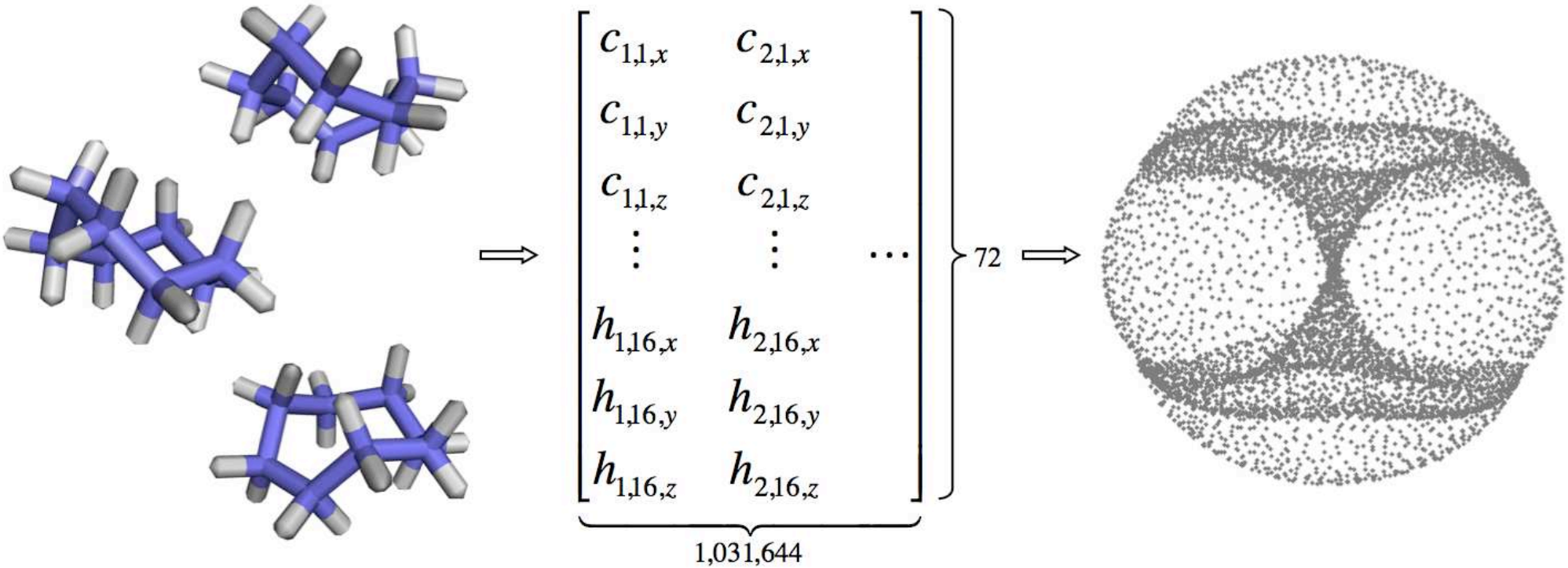


Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Datasets have shapes

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

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

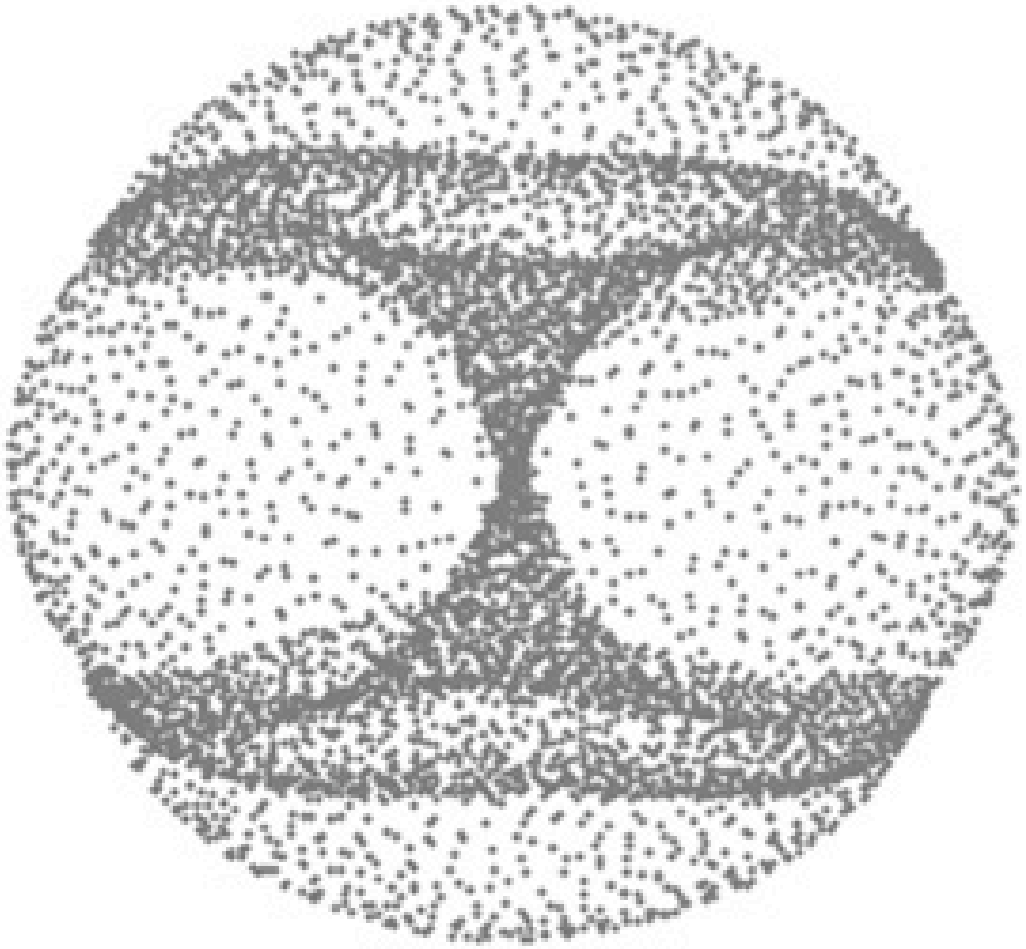


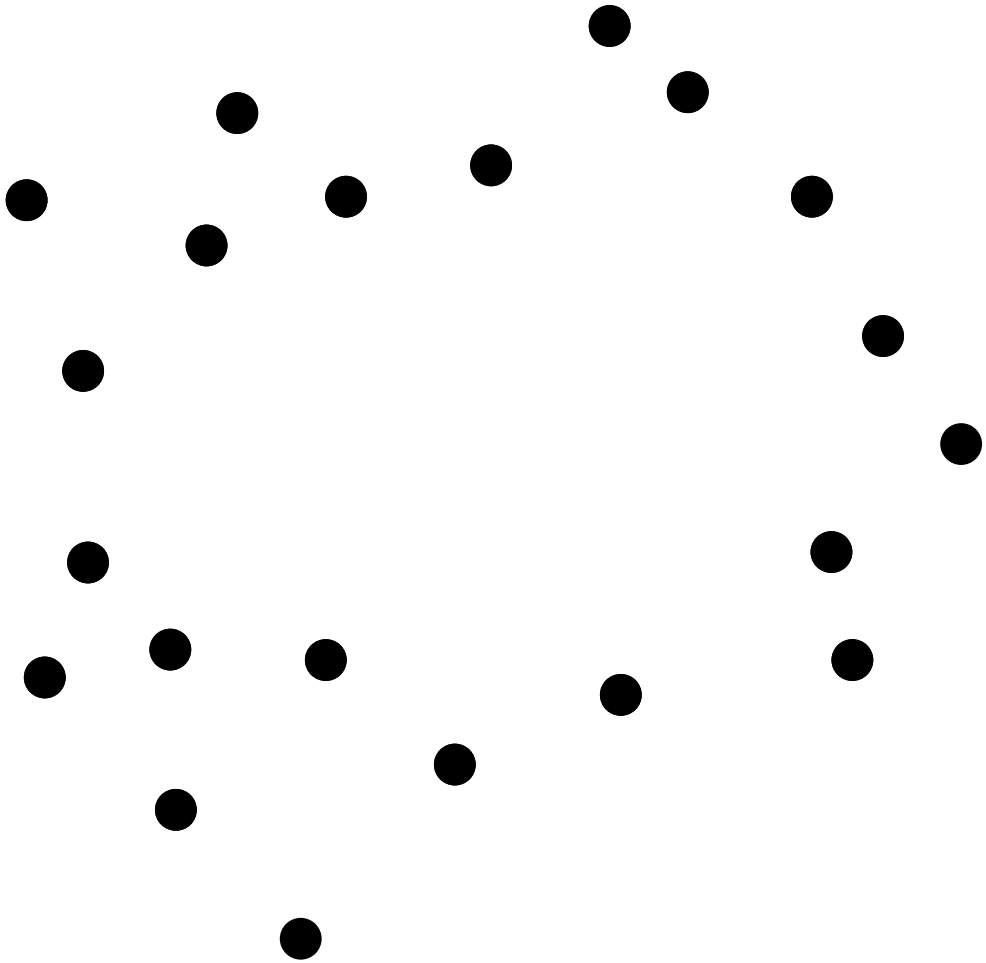
Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

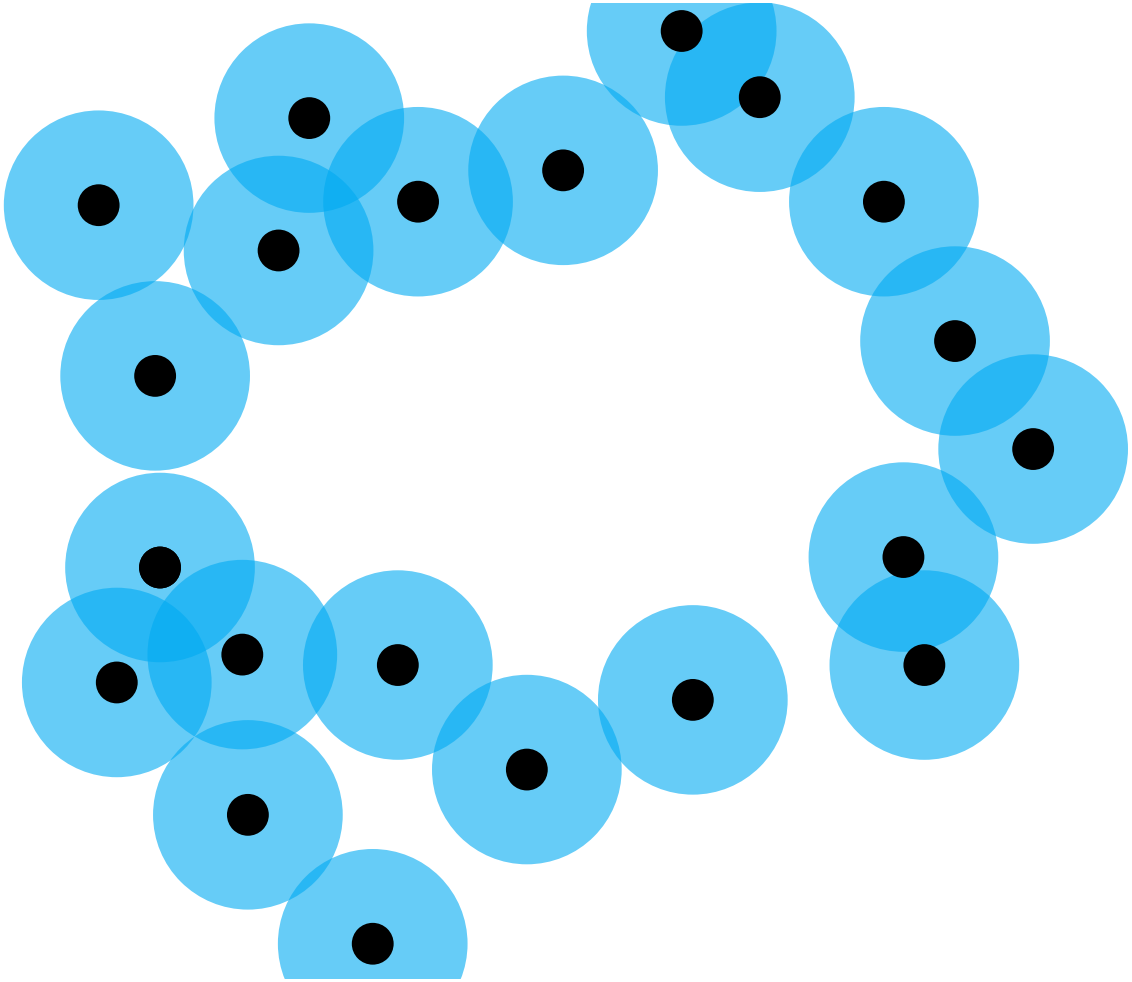
Datasets have shapes

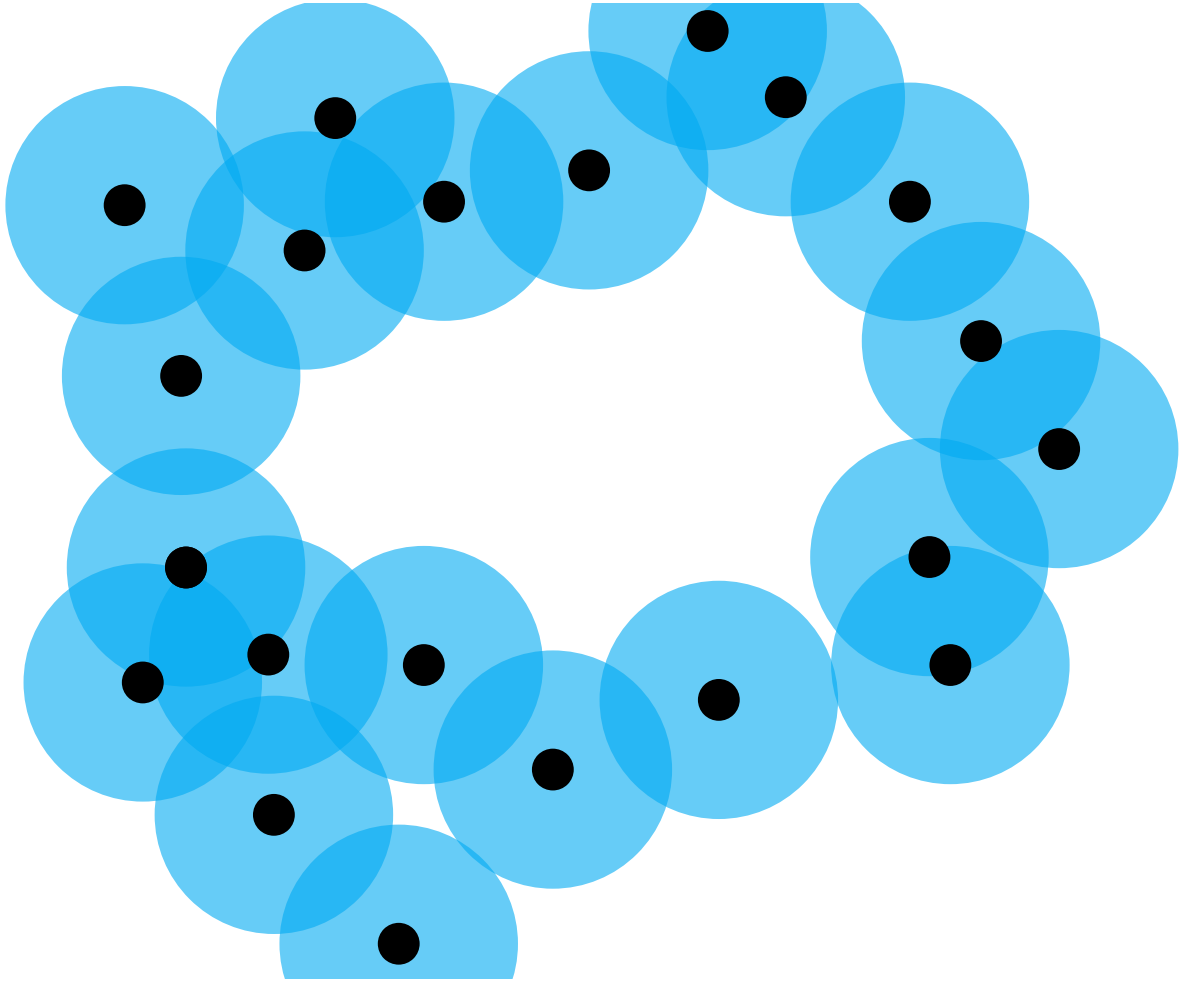
Datasets have shapes

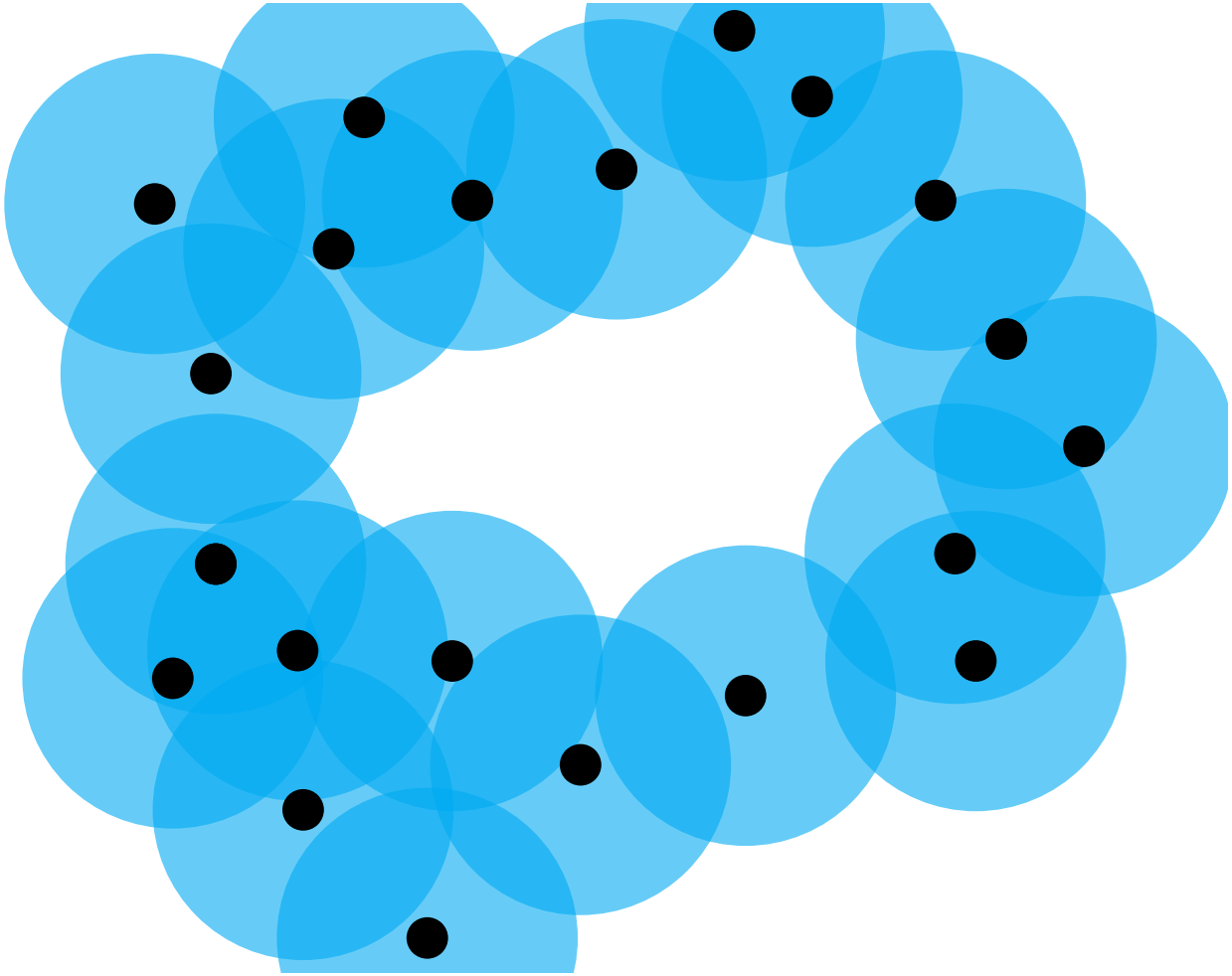
What shape is this?

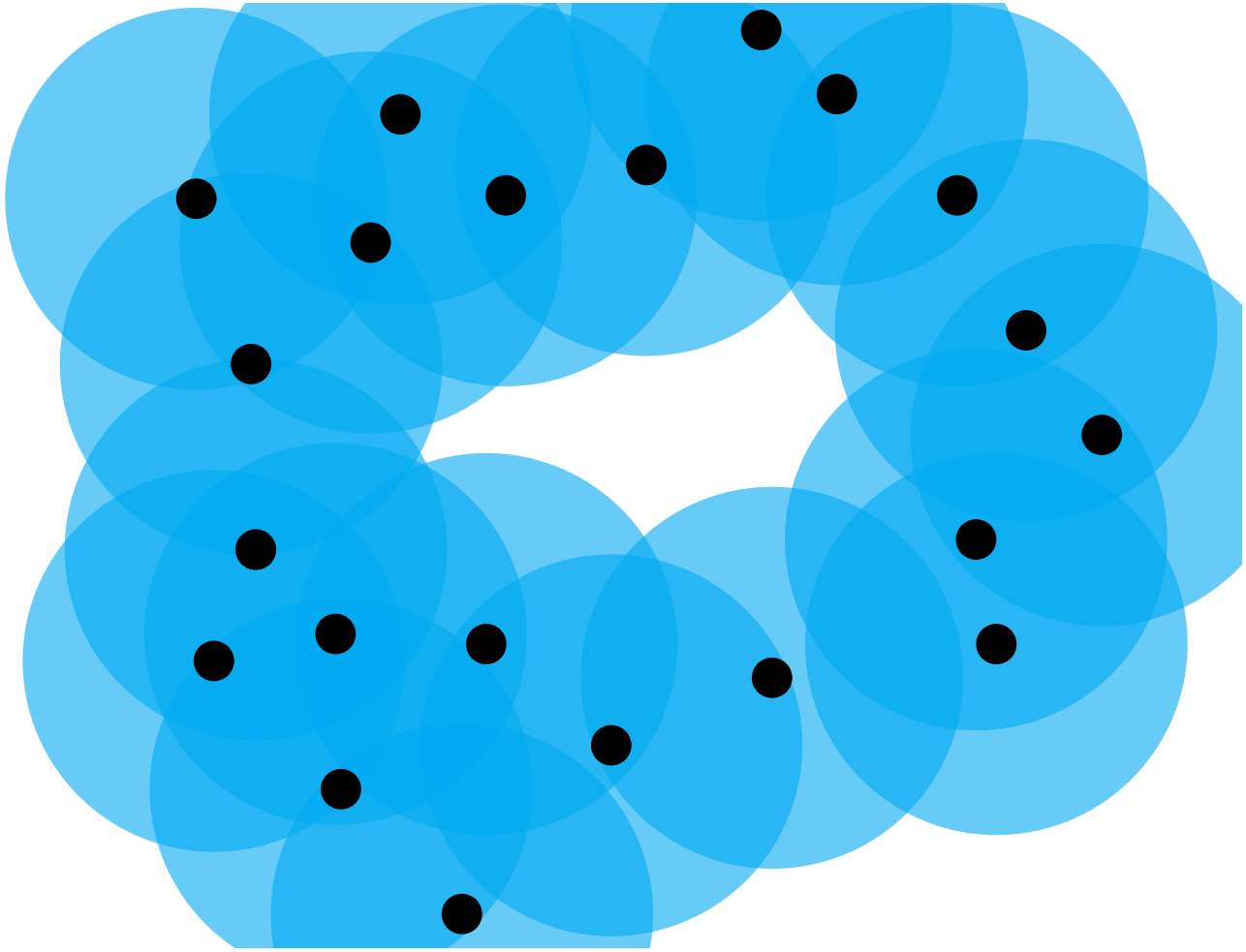


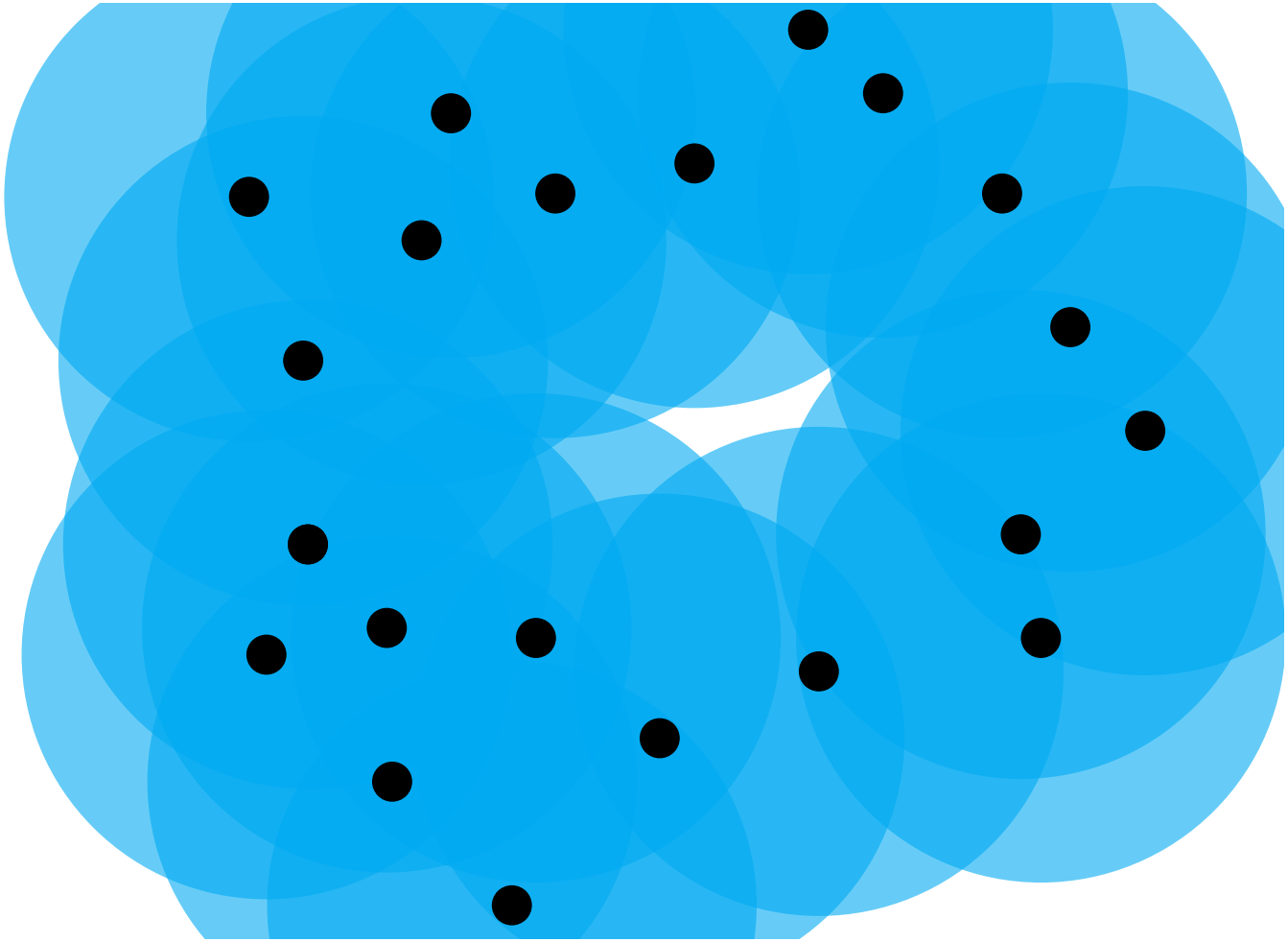


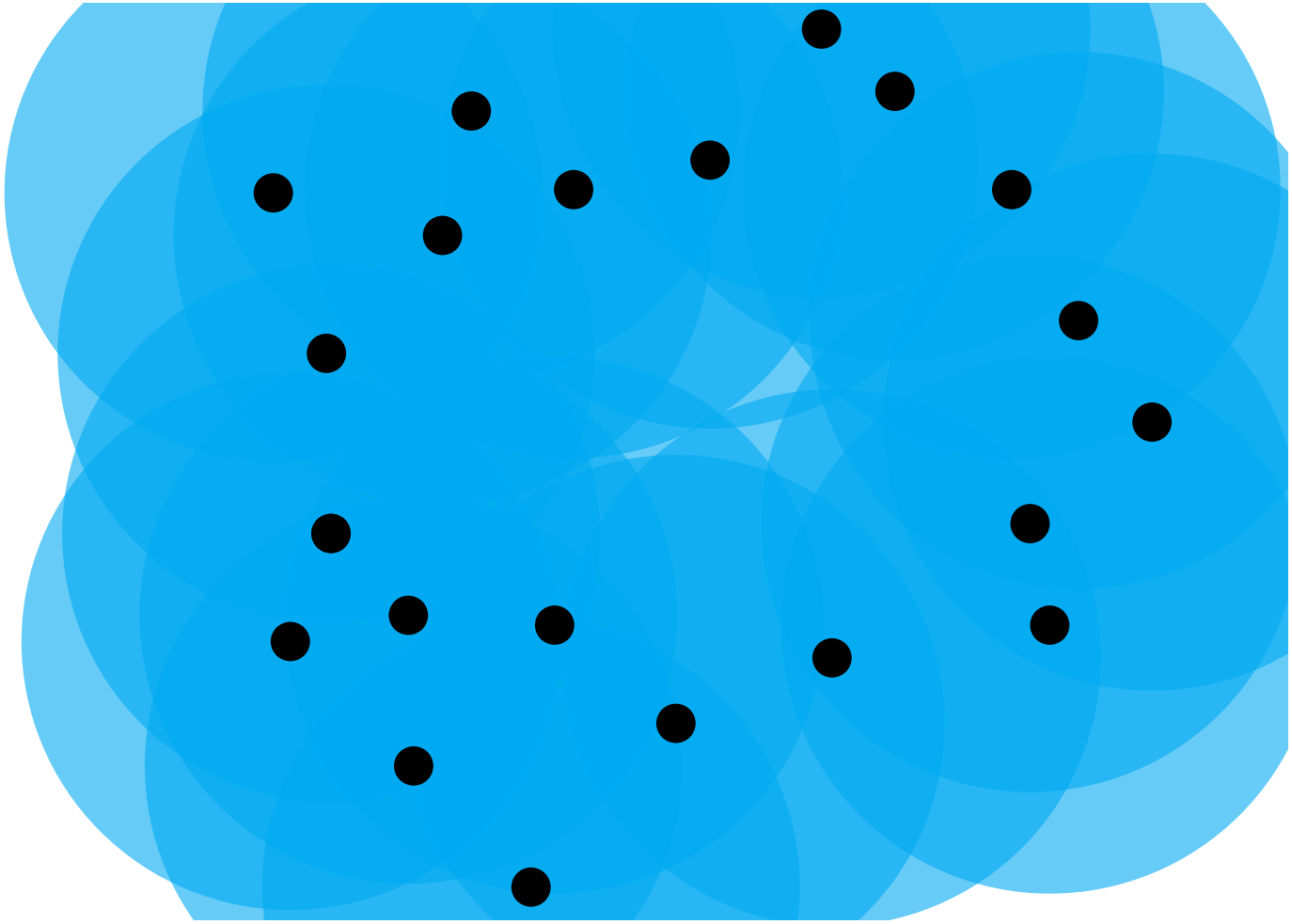


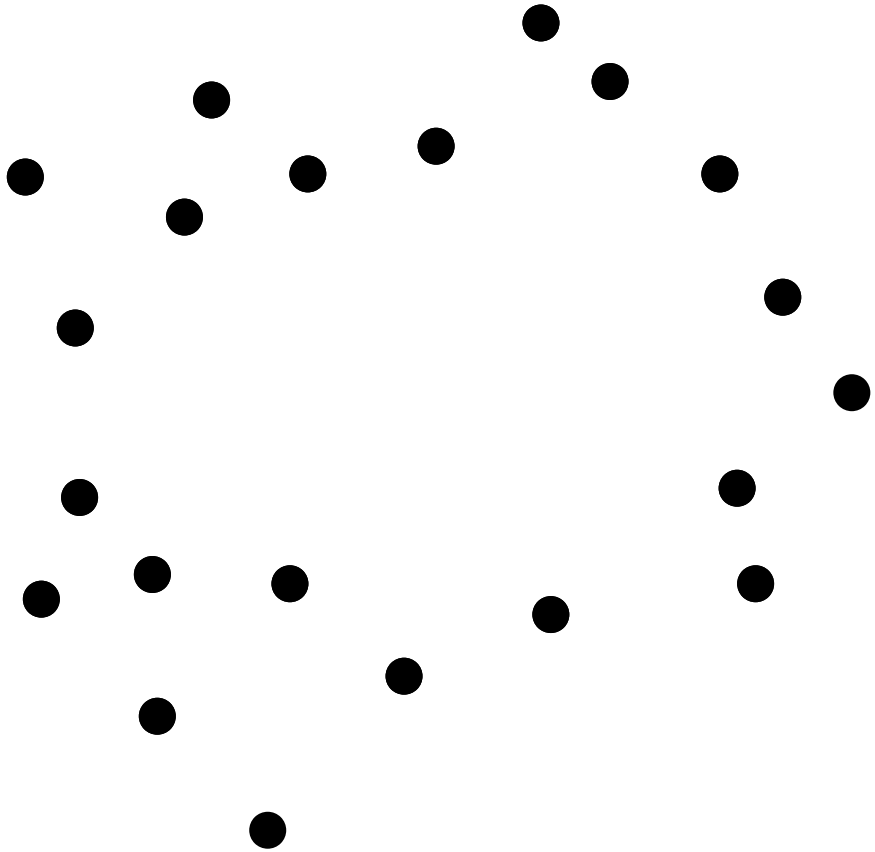








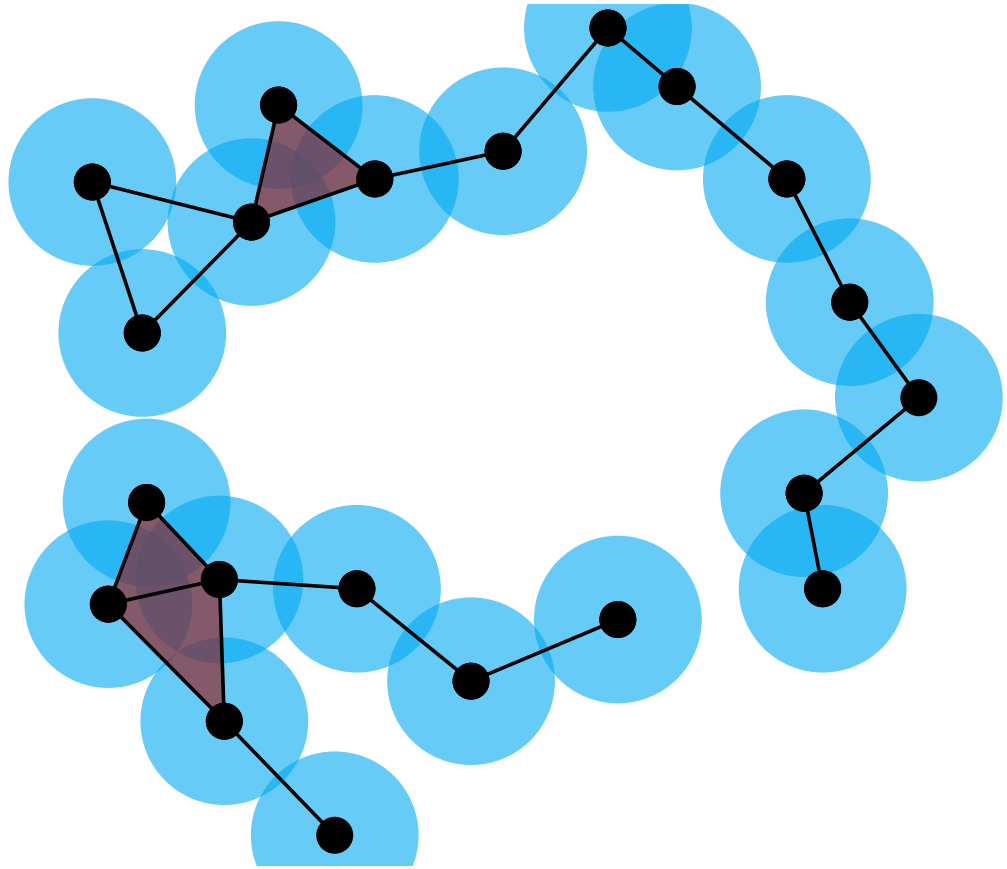




Definition

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

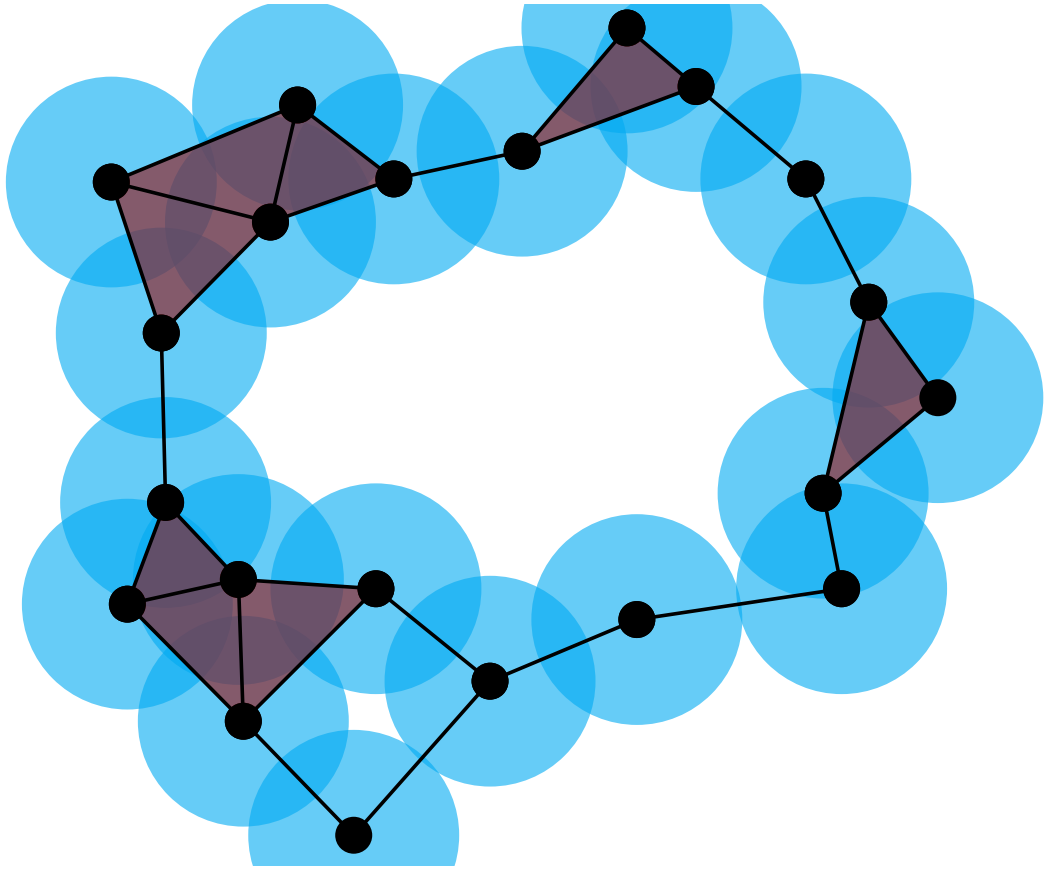
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

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

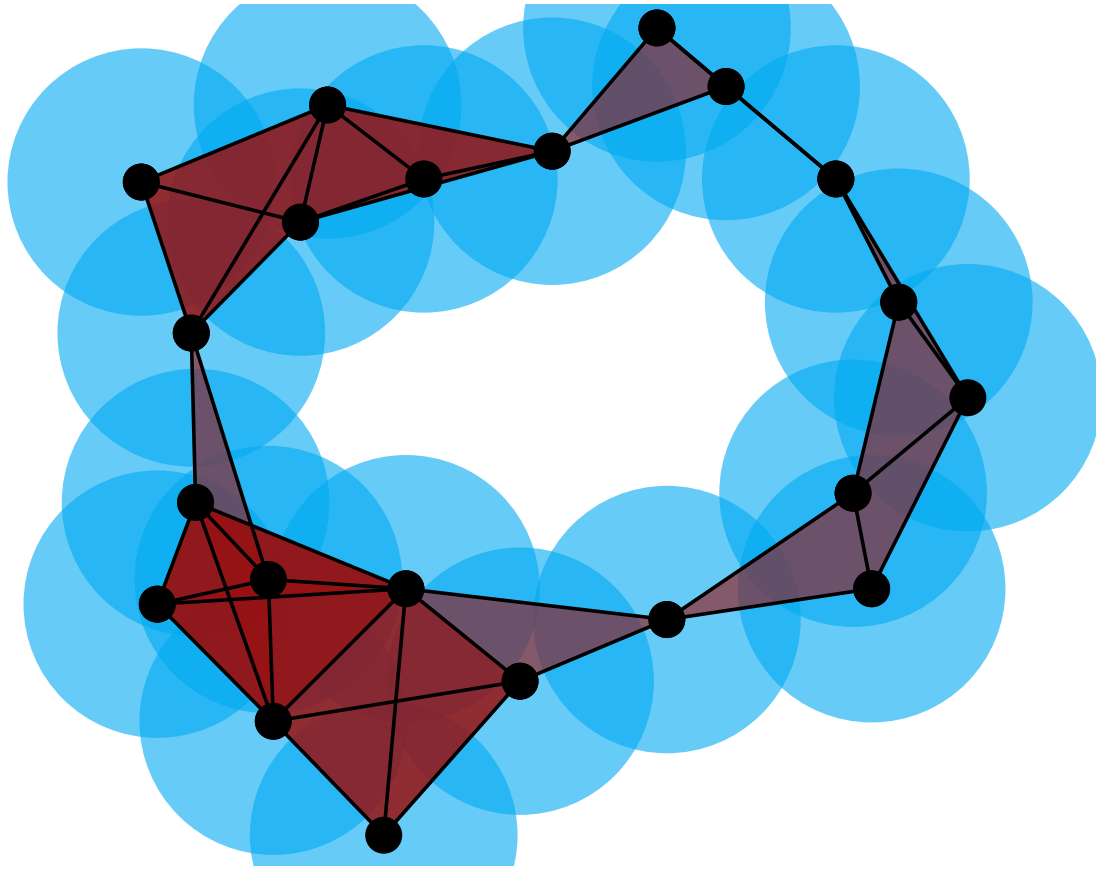
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

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

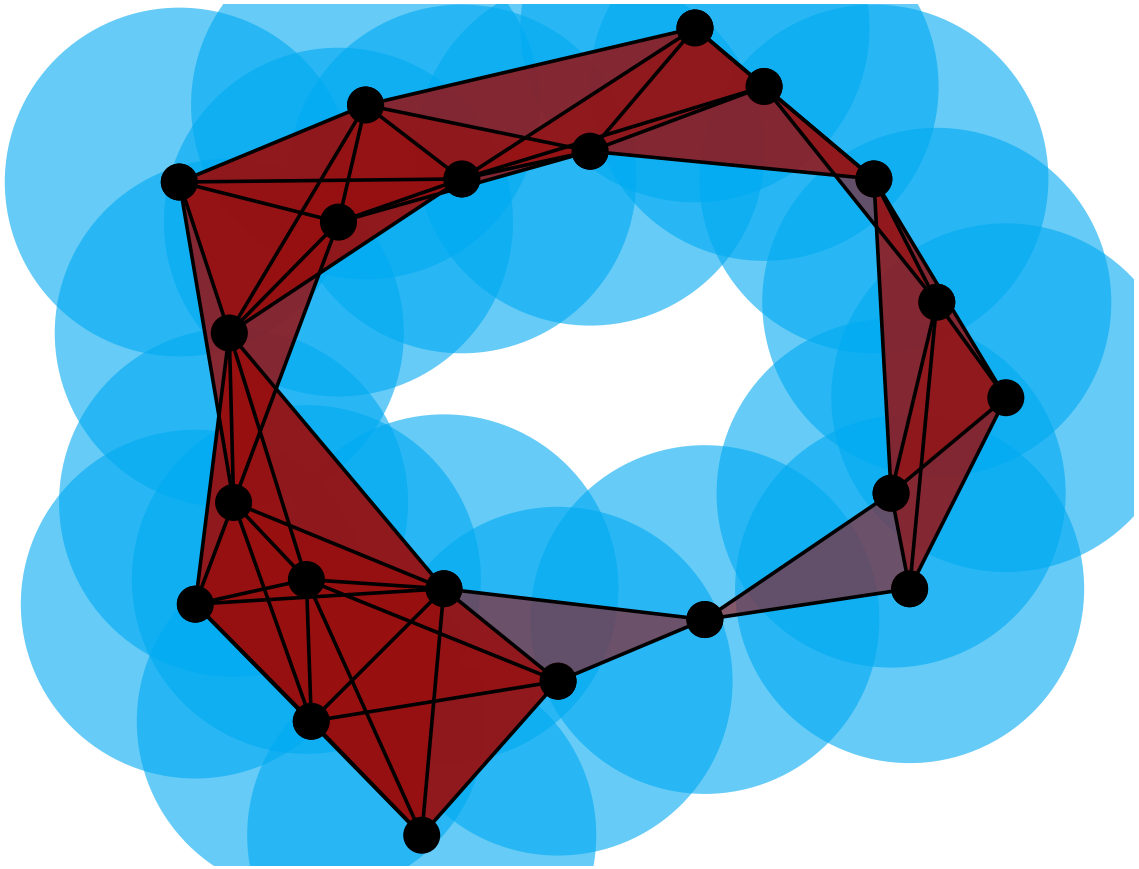
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

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

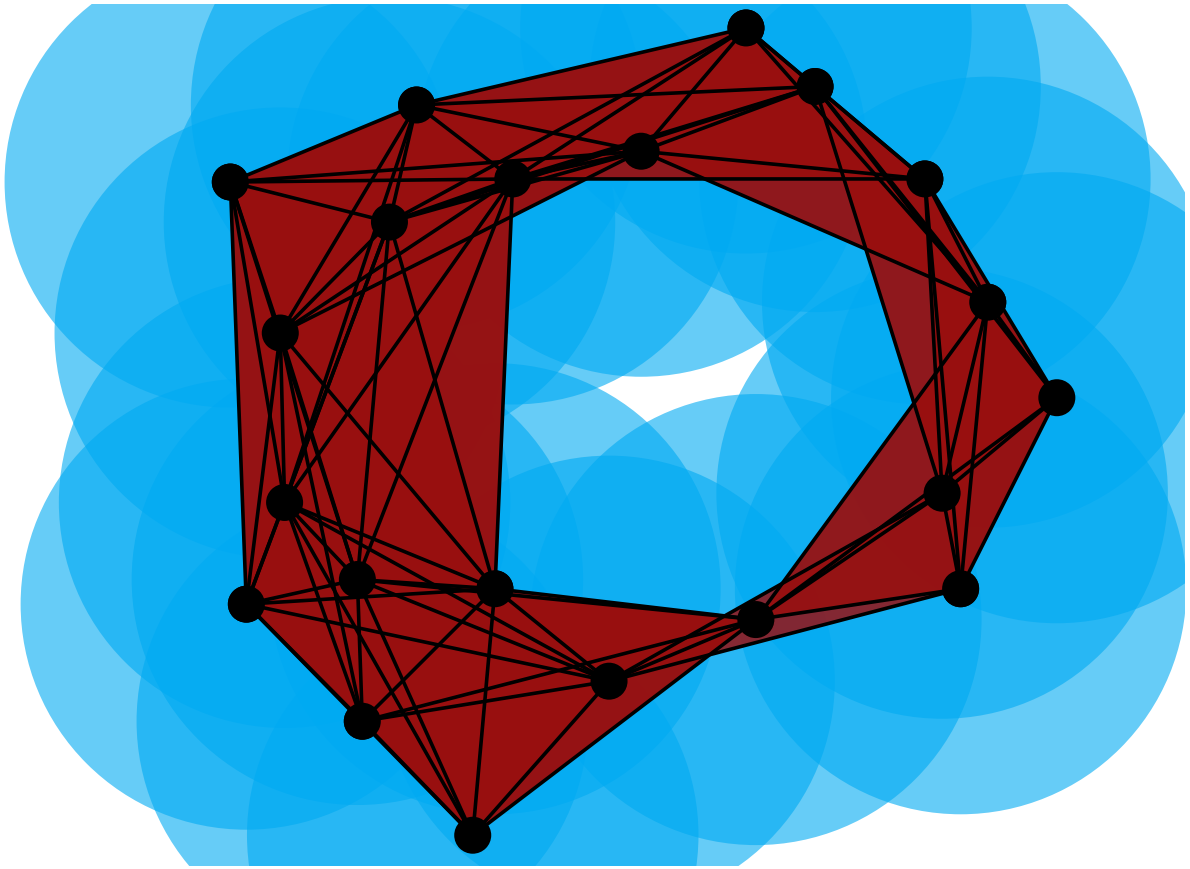
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

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

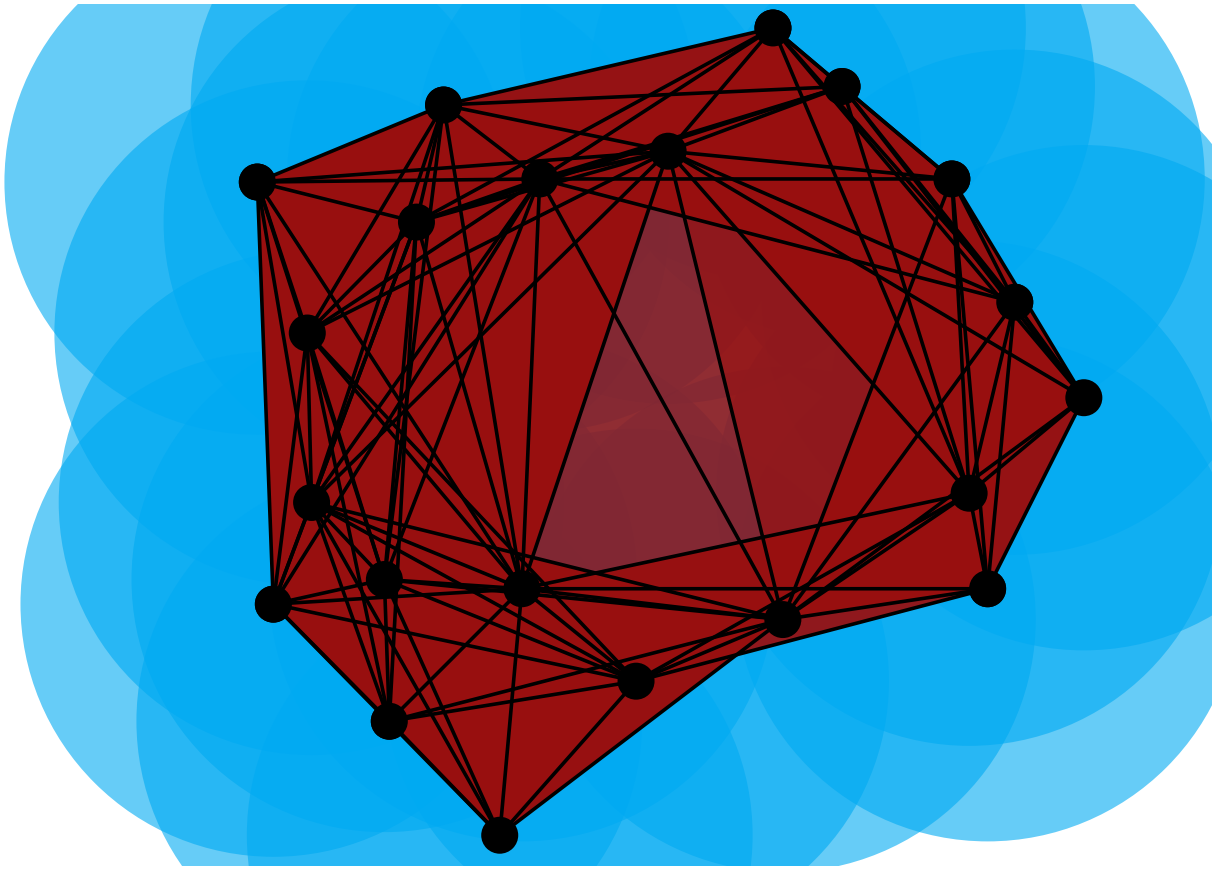
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

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

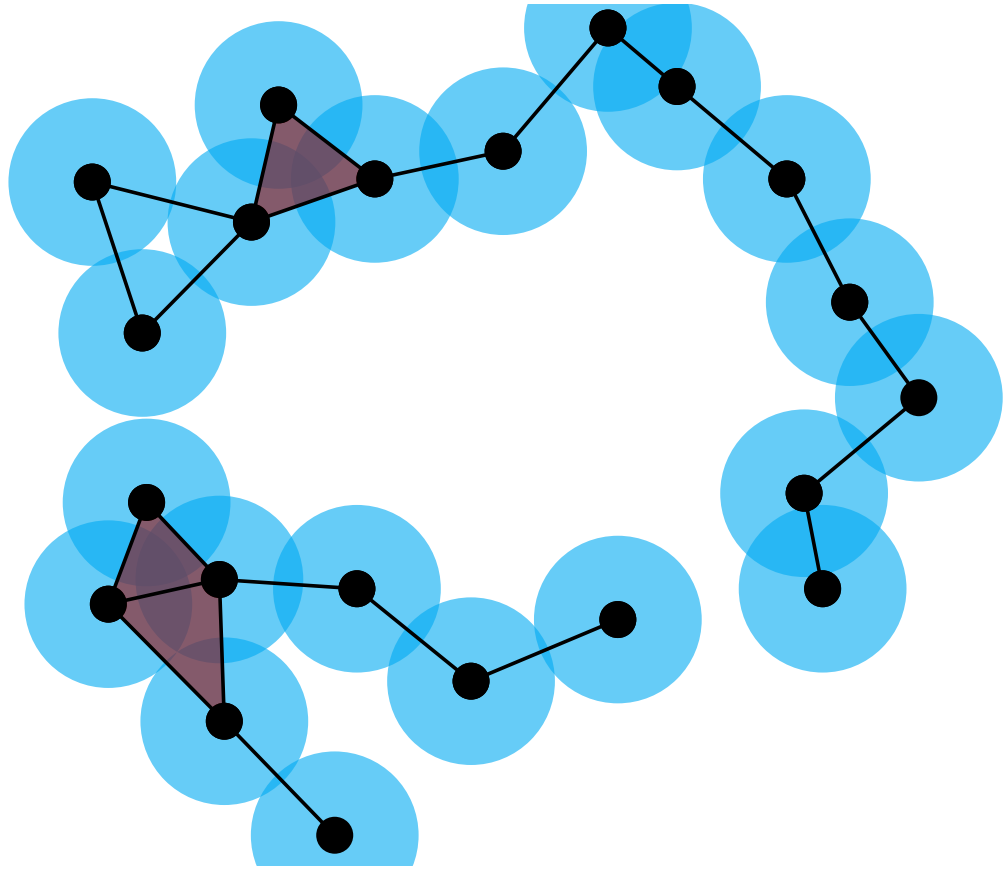
- vertex set X
- finite simplex $\{x_0, x_1, \ldots, x_k\}$ when $\cap_{i=0}^{k} B(x_i, r) \neq \emptyset$.



Definition

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

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



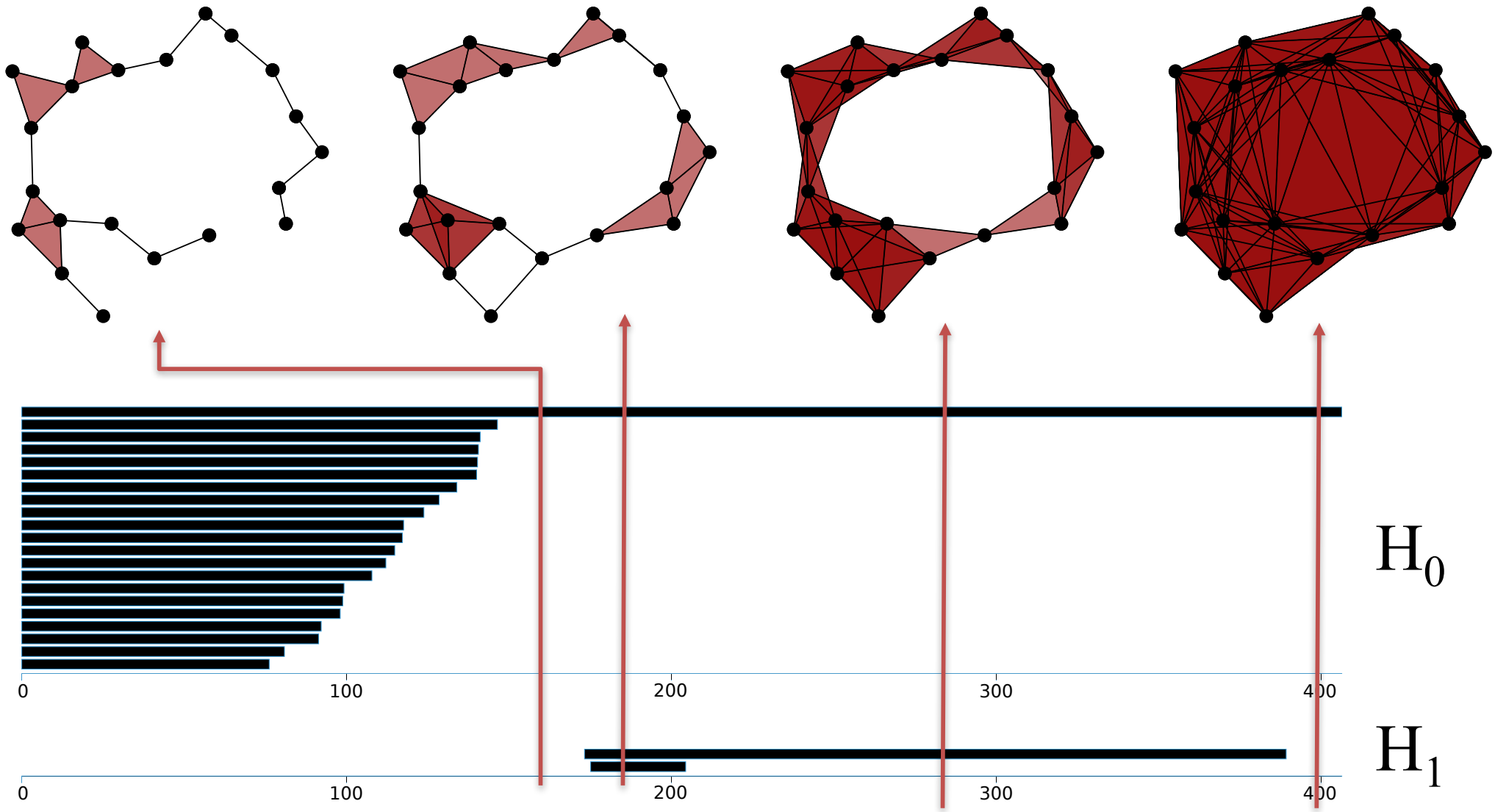
<u>Nerve Lemma.</u> $\check{\mathrm{C}}\mathrm{ech}(X;r) \simeq$ union of balls

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 applied to data

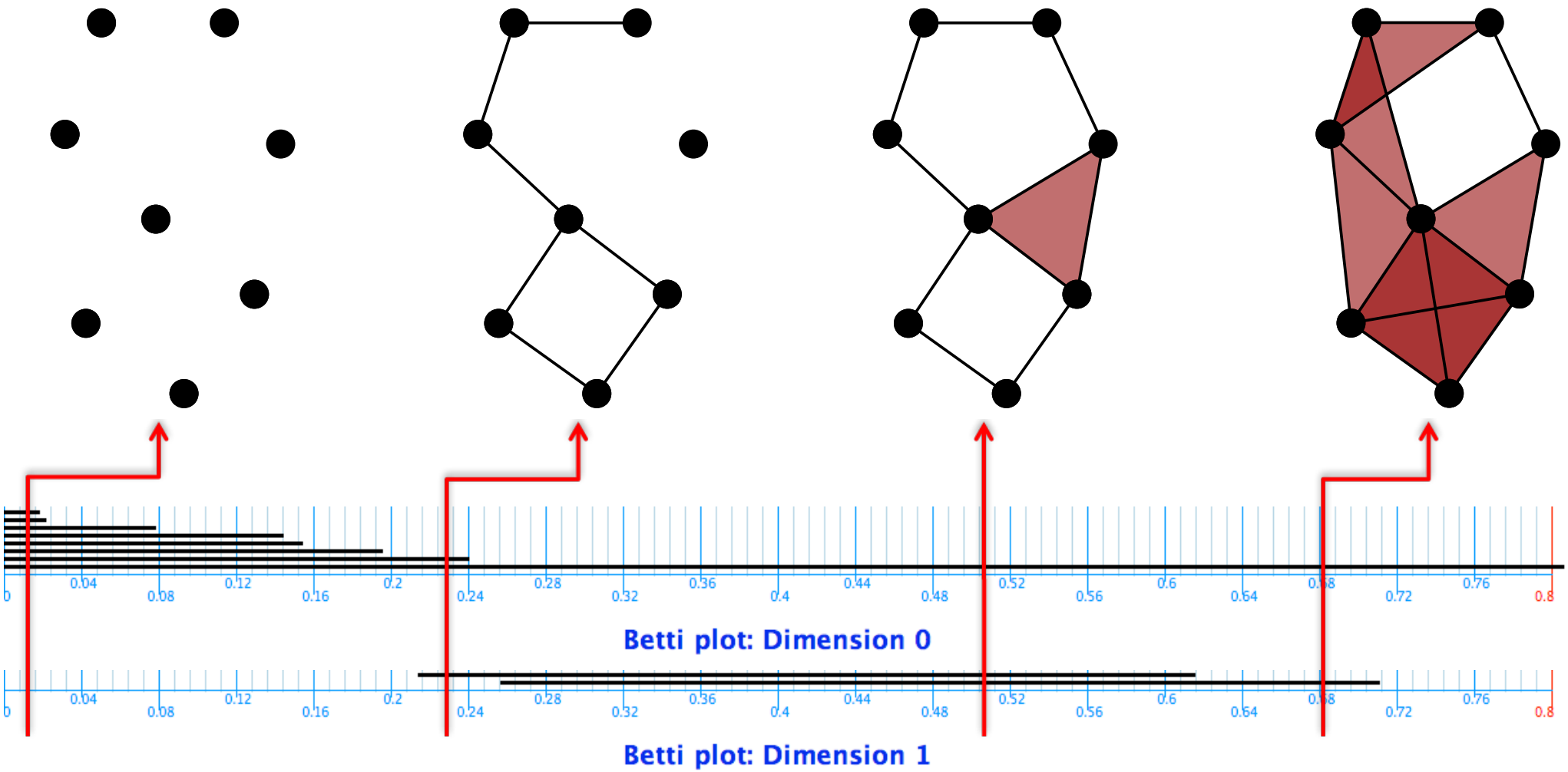


Input: Increasing spaces. Output: barcode.

Significant features persist.

Cubic computation time in the number of simplices.

Persistent homology applied to data



Input: Increasing spaces. Output: barcode.

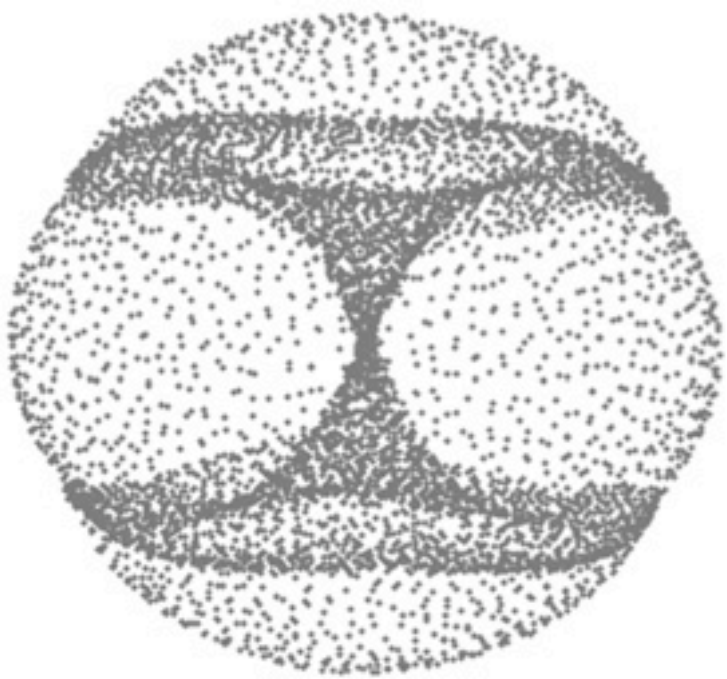
Significant features persist.

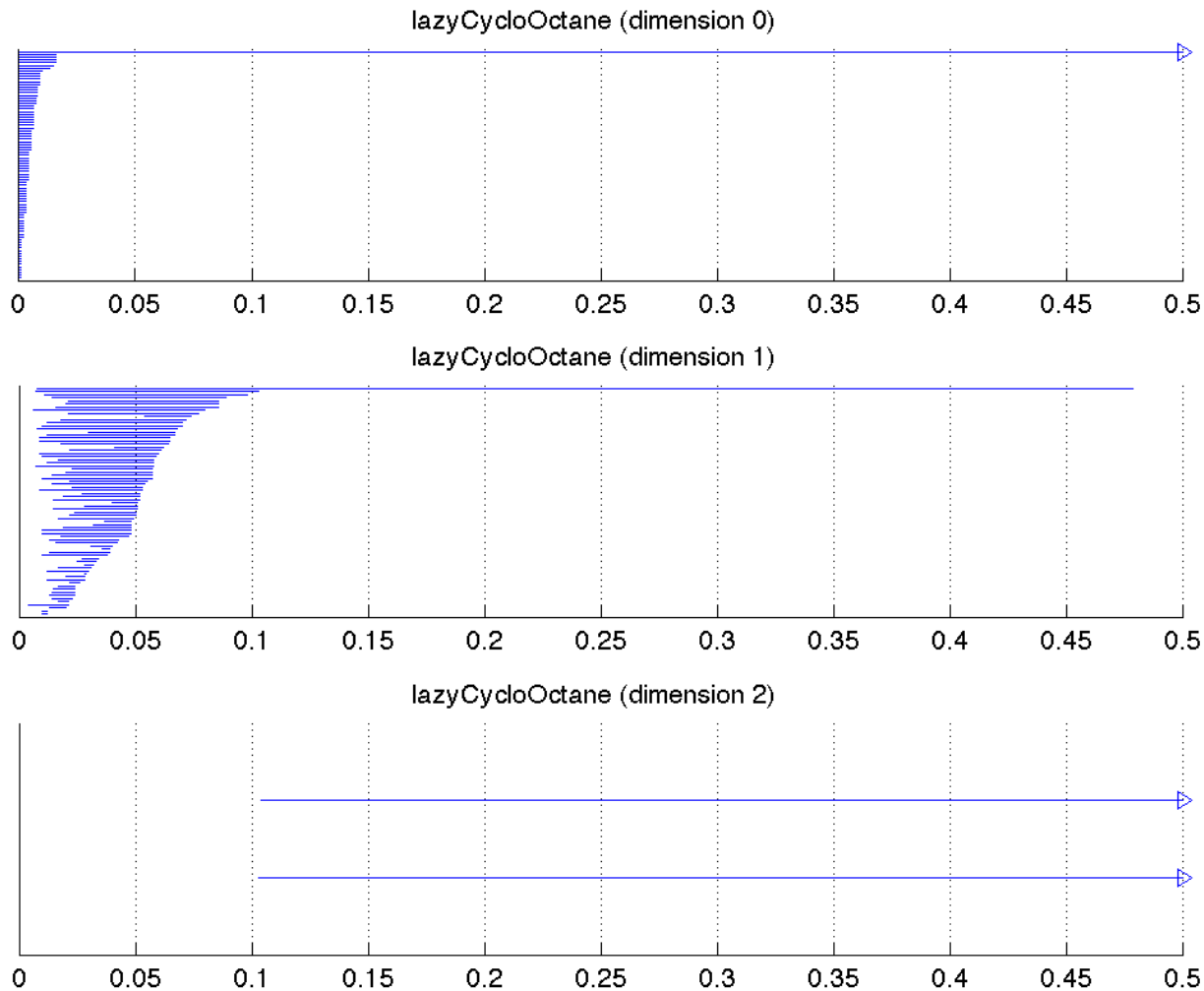
Cubic computation time in the number of simplices.

Persistent homology applied to data

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

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



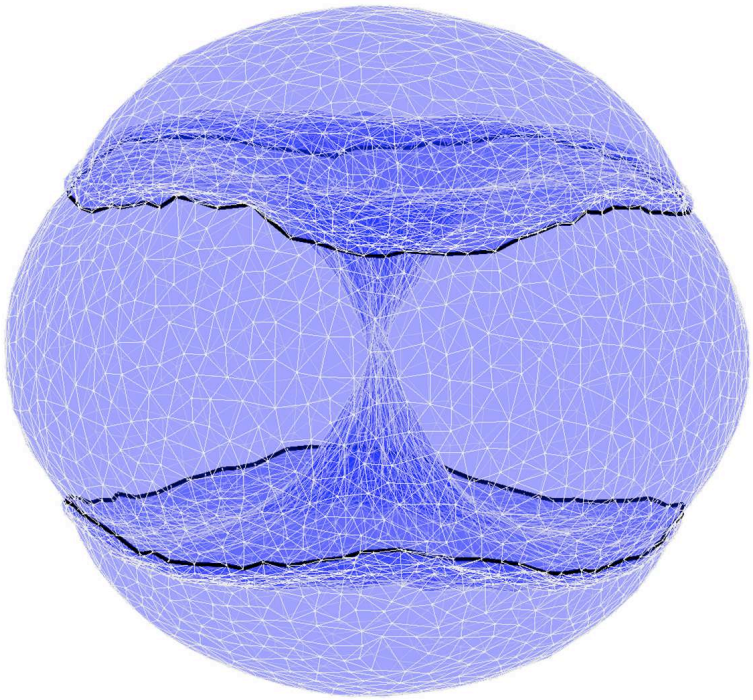


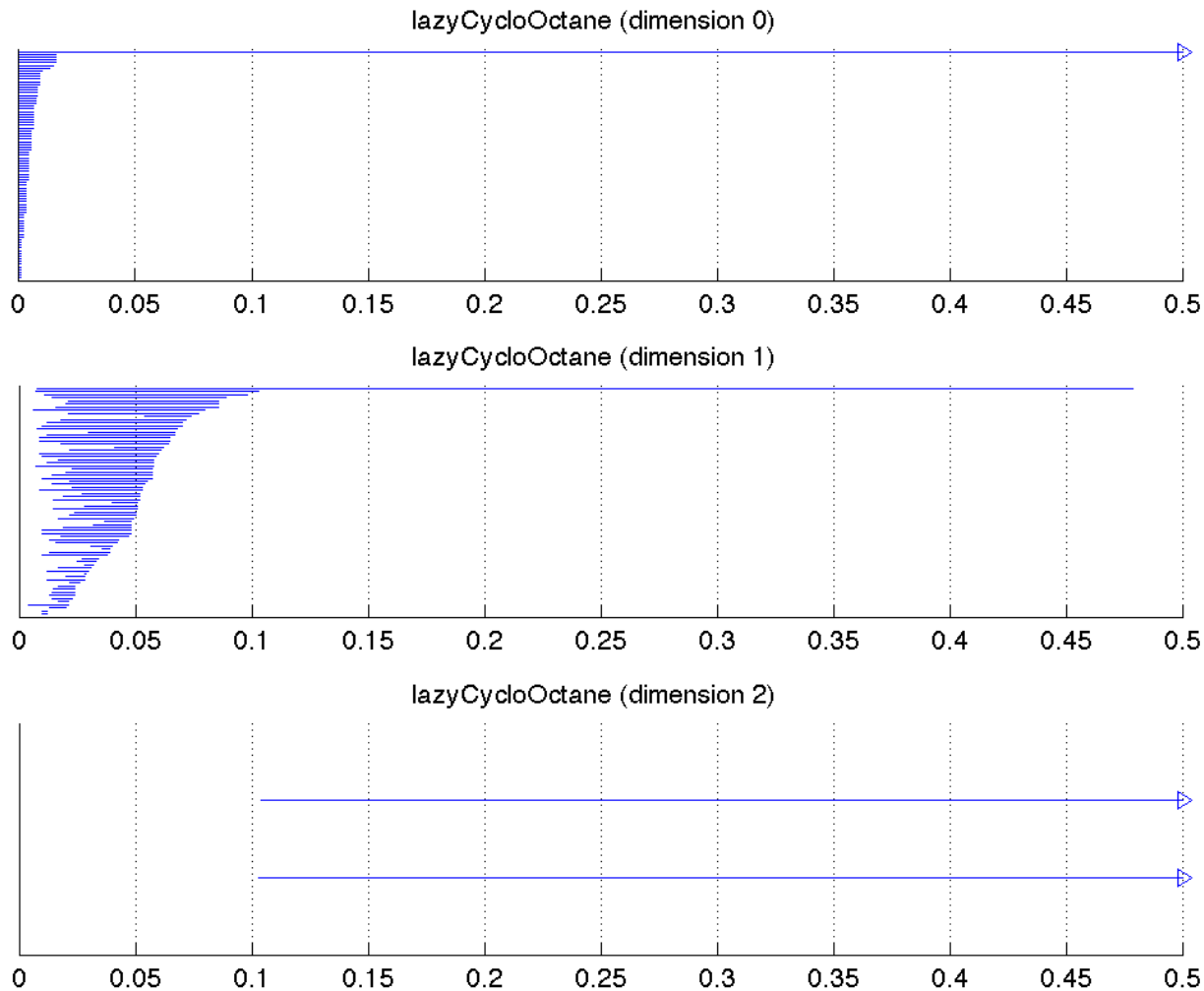
Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

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

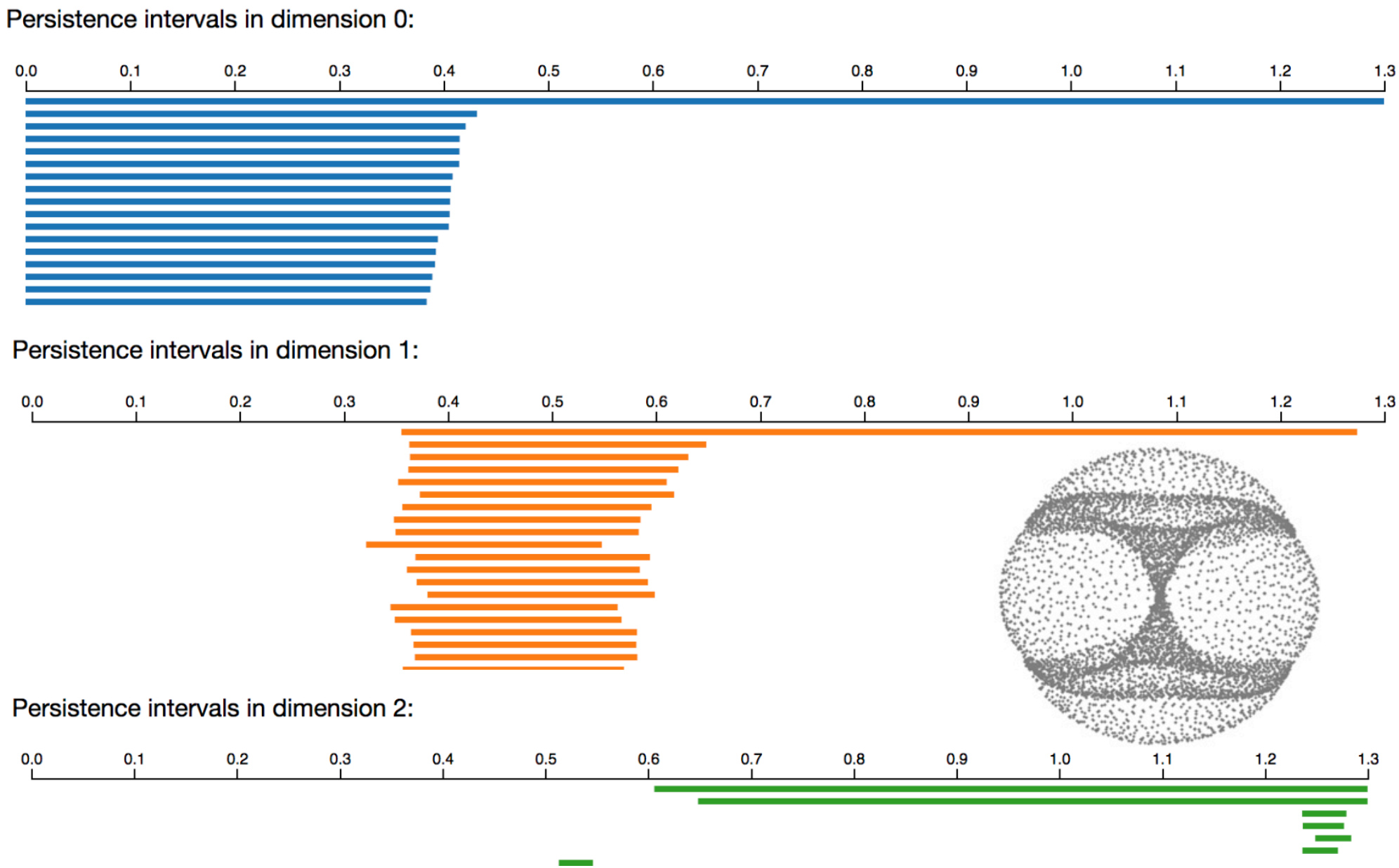
1,000,000+ points in 24-dimensional space





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

Persistent homology applied to data

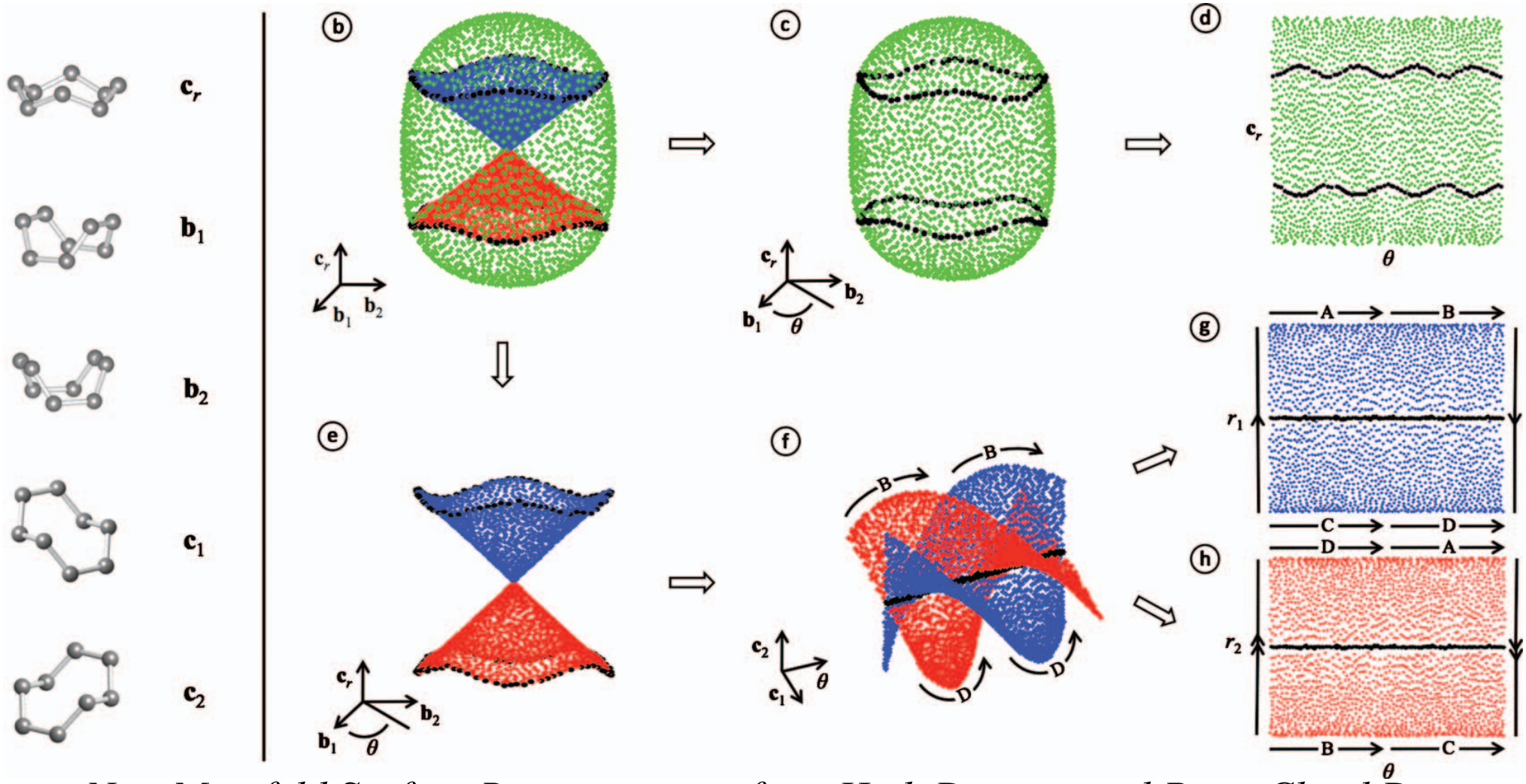


Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

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

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

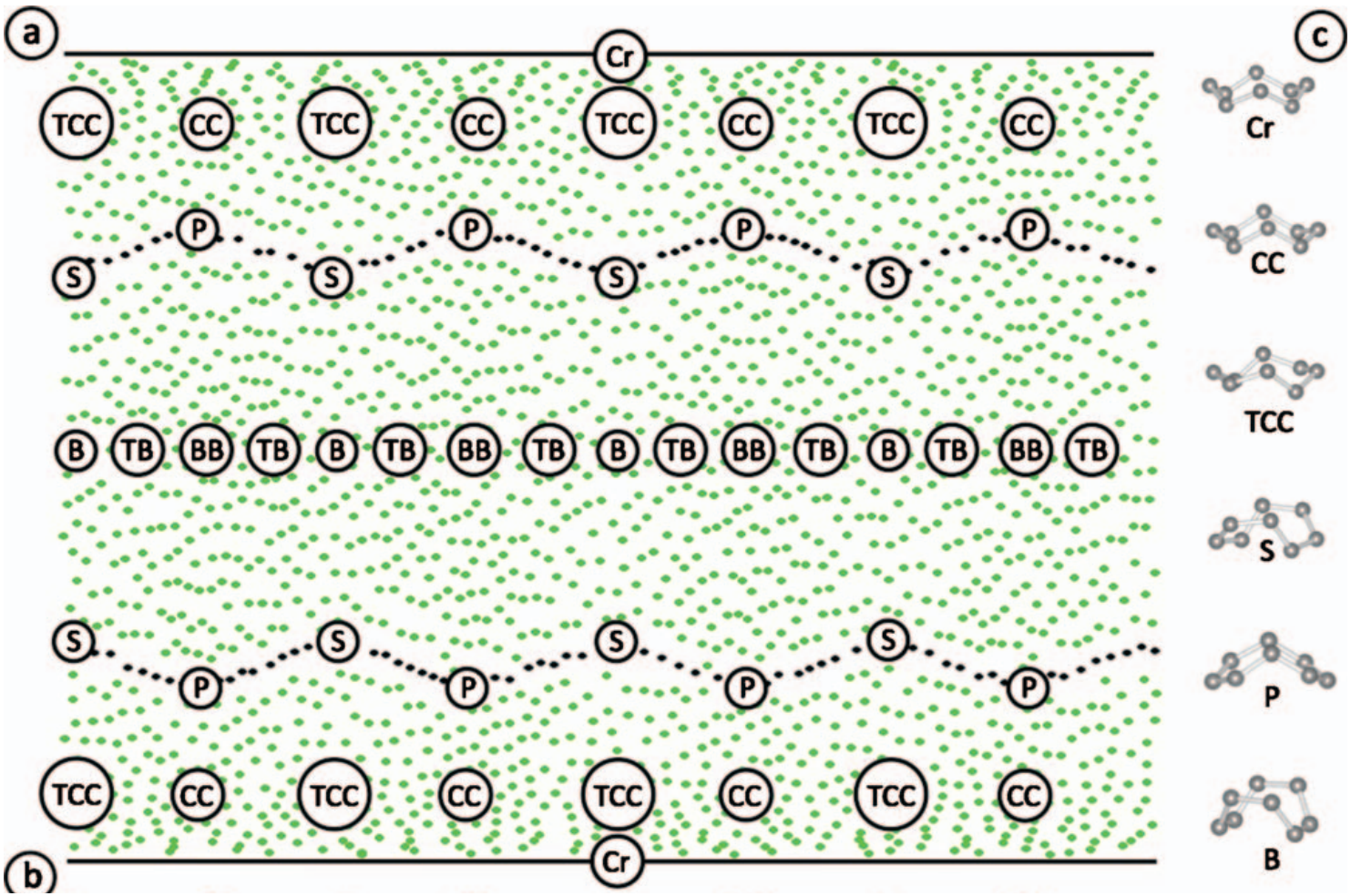


Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

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

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

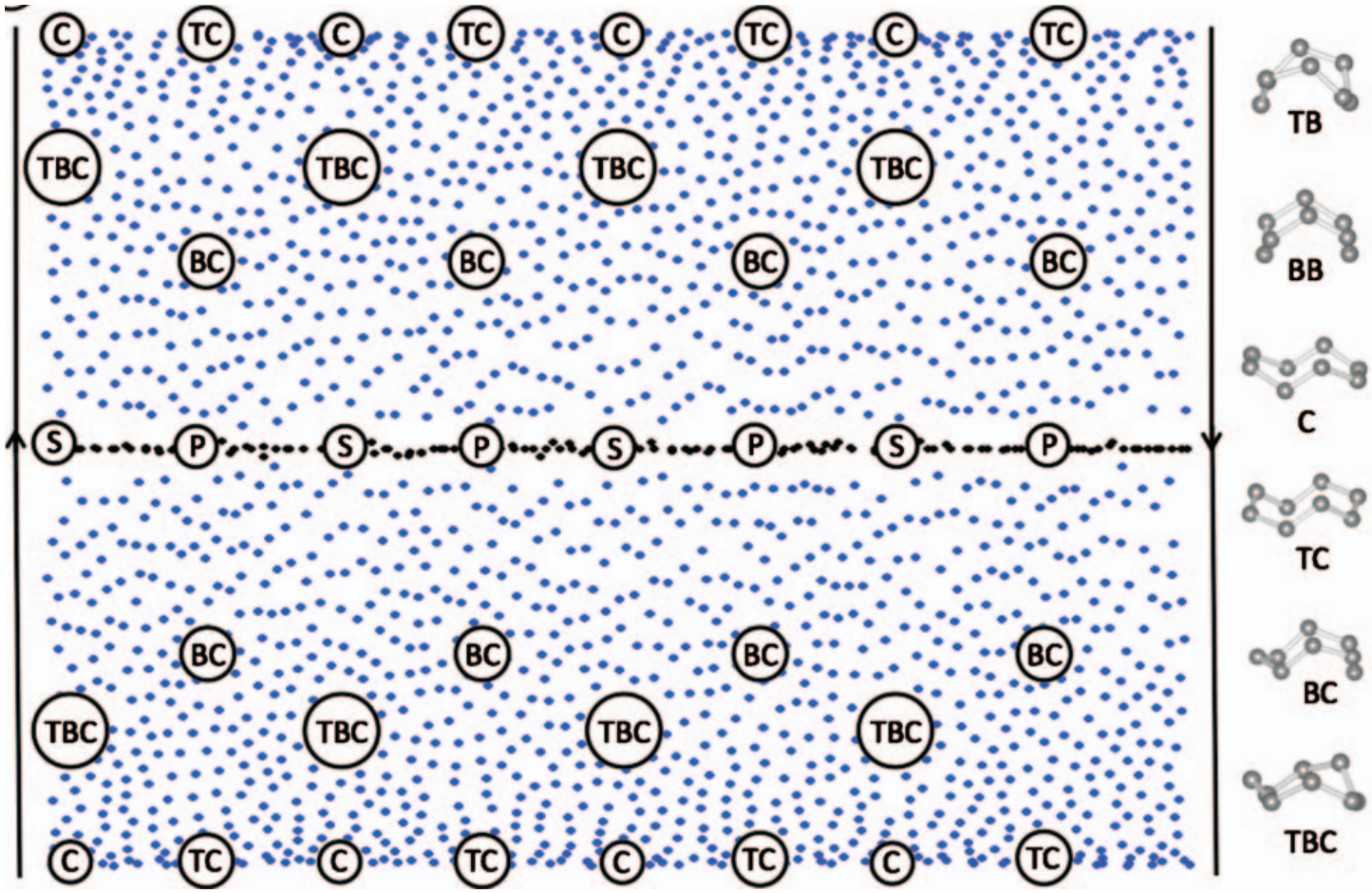


Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

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

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

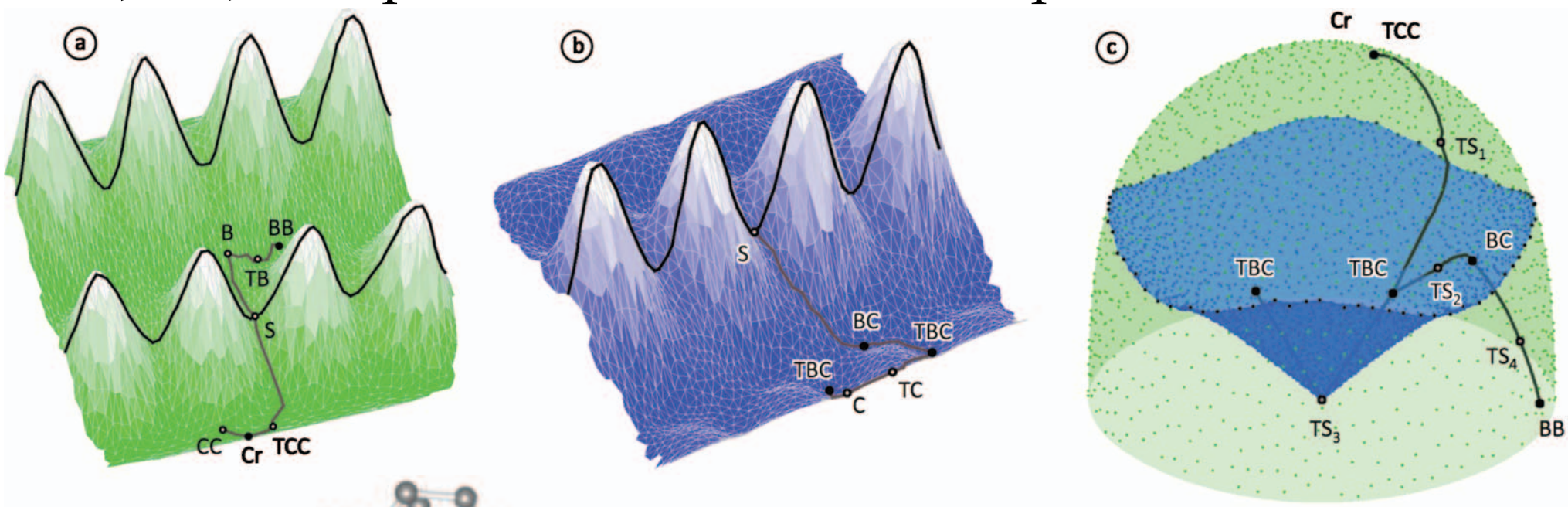


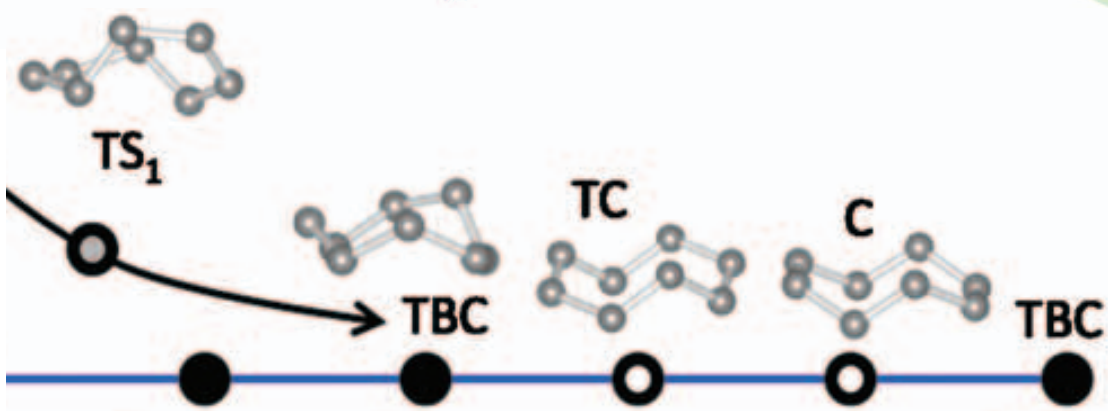
Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

Persistent homology applied to data

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

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





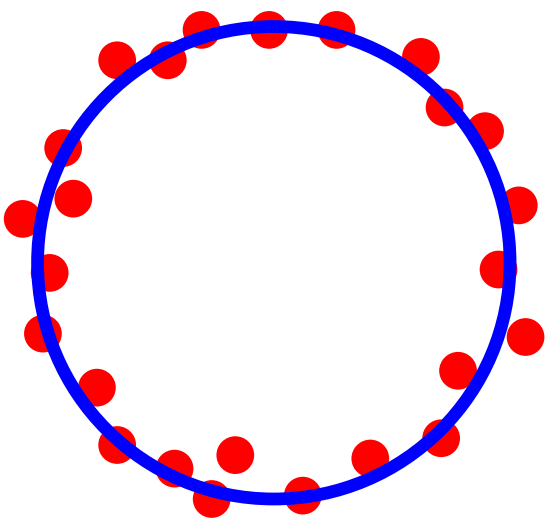
Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.

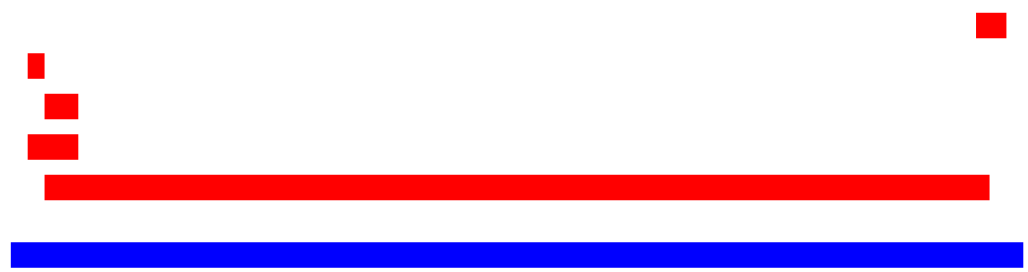
Persistent homology applied to data

- <u>Stability Theorem.</u>

  If X and Y are metric spaces, then

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

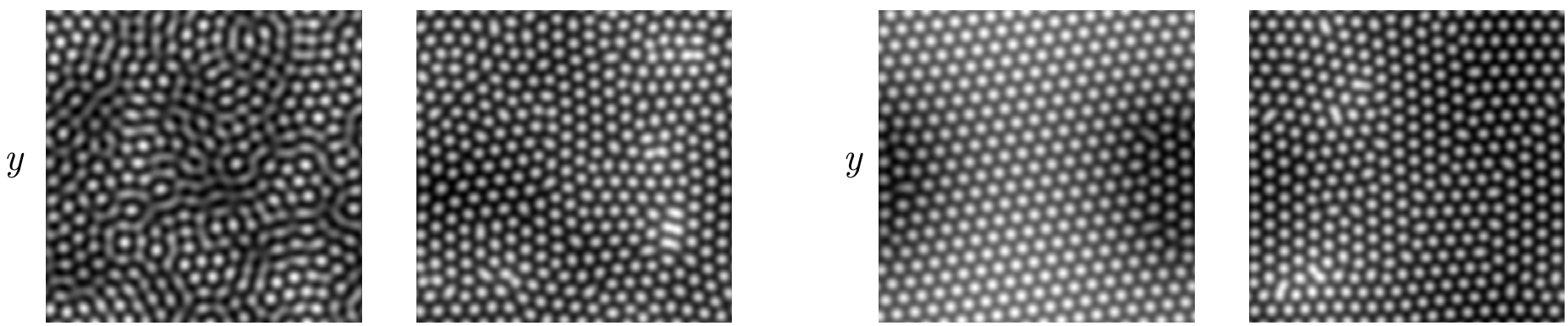




Why is applied topology popular when few datasets have Klein bottles?

- 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.

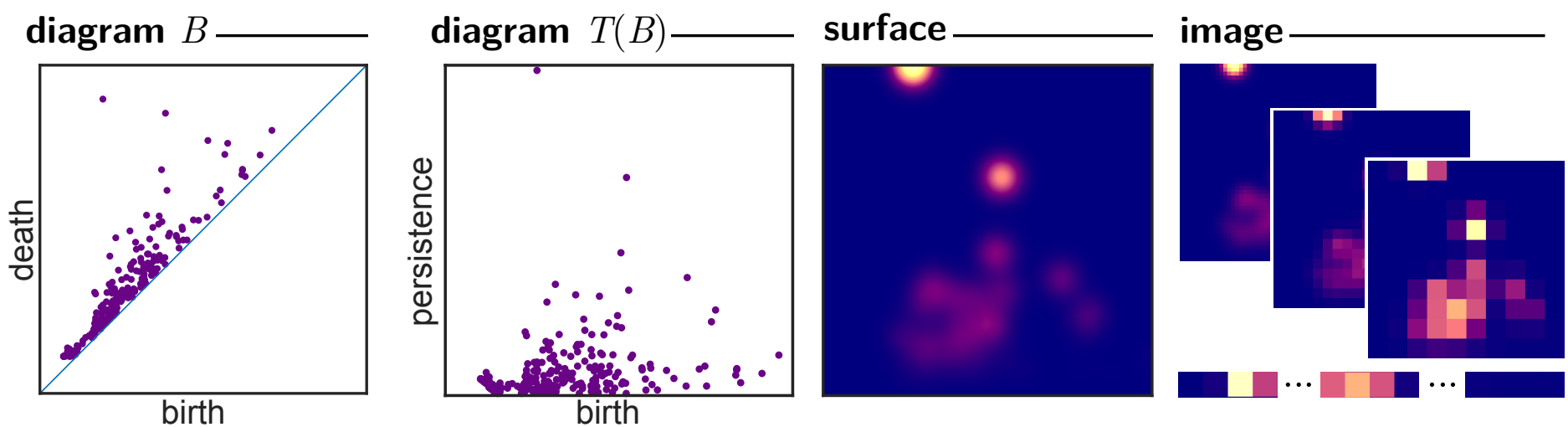


Measures of Order for nearly hexagonal lattices by Francis Motta, Rachel Neville, Patrick Shipman, Daniel Pearson, and Mark Bradley, 2018.

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

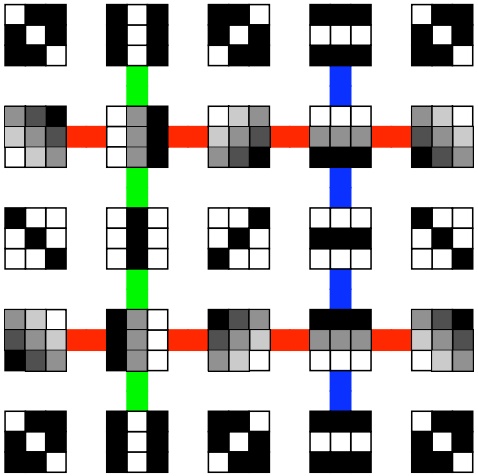
- Many datasets have clusters & flares (as in the diabetes example)
- Motivates interesting questions in many pure disciplines: mathematics, computer science (computational geometry), statistics
- Interest from domain experts in biology, neuroscience, computer vision, dynamical systems, sensor networks, ...
- 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.



Conclusions

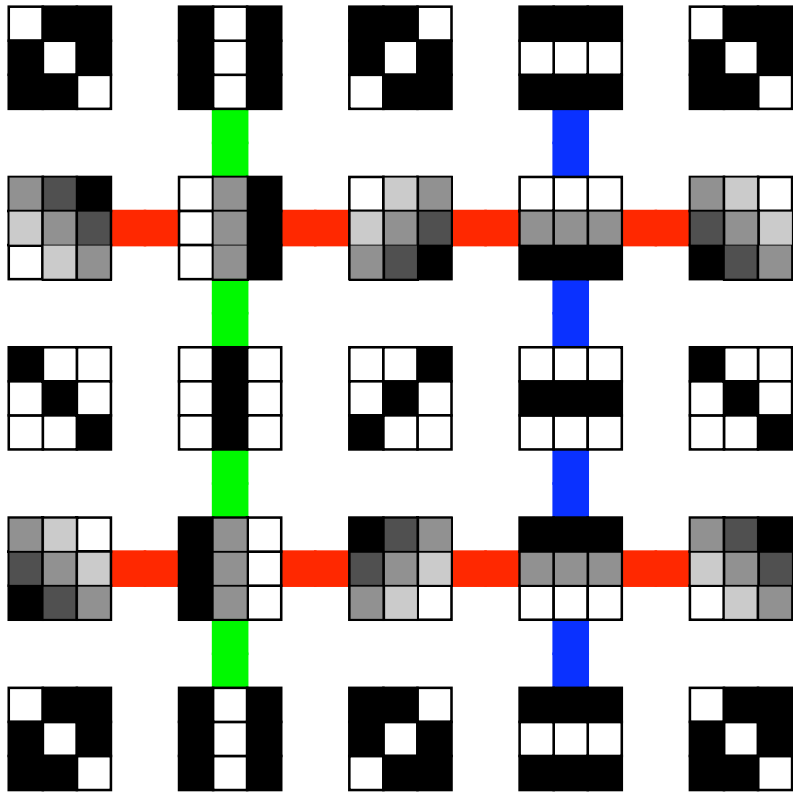
- Persistent homology can be applied to any increasing sequence of spaces, measuring some of the local and global topology.
- The two most common versions of persistent homology are sublevelset persistence and persistence applied to point cloud data.



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

- Aleksandr Solzhenitsyn, *The First Circle*

Topology applied to image data



Persistent homology applied to data

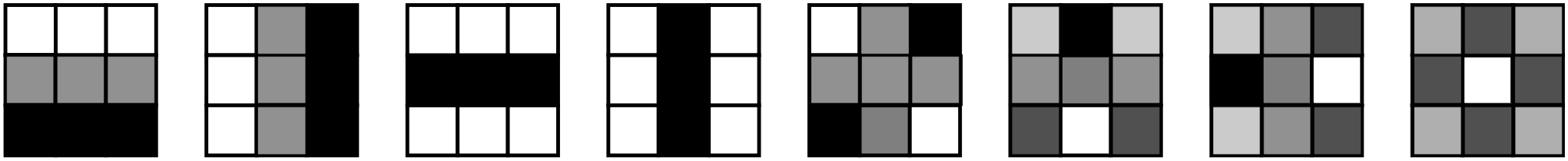
The receptive fields of cells in our primary visual cortex (V1) are related to the statistics natural images.

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

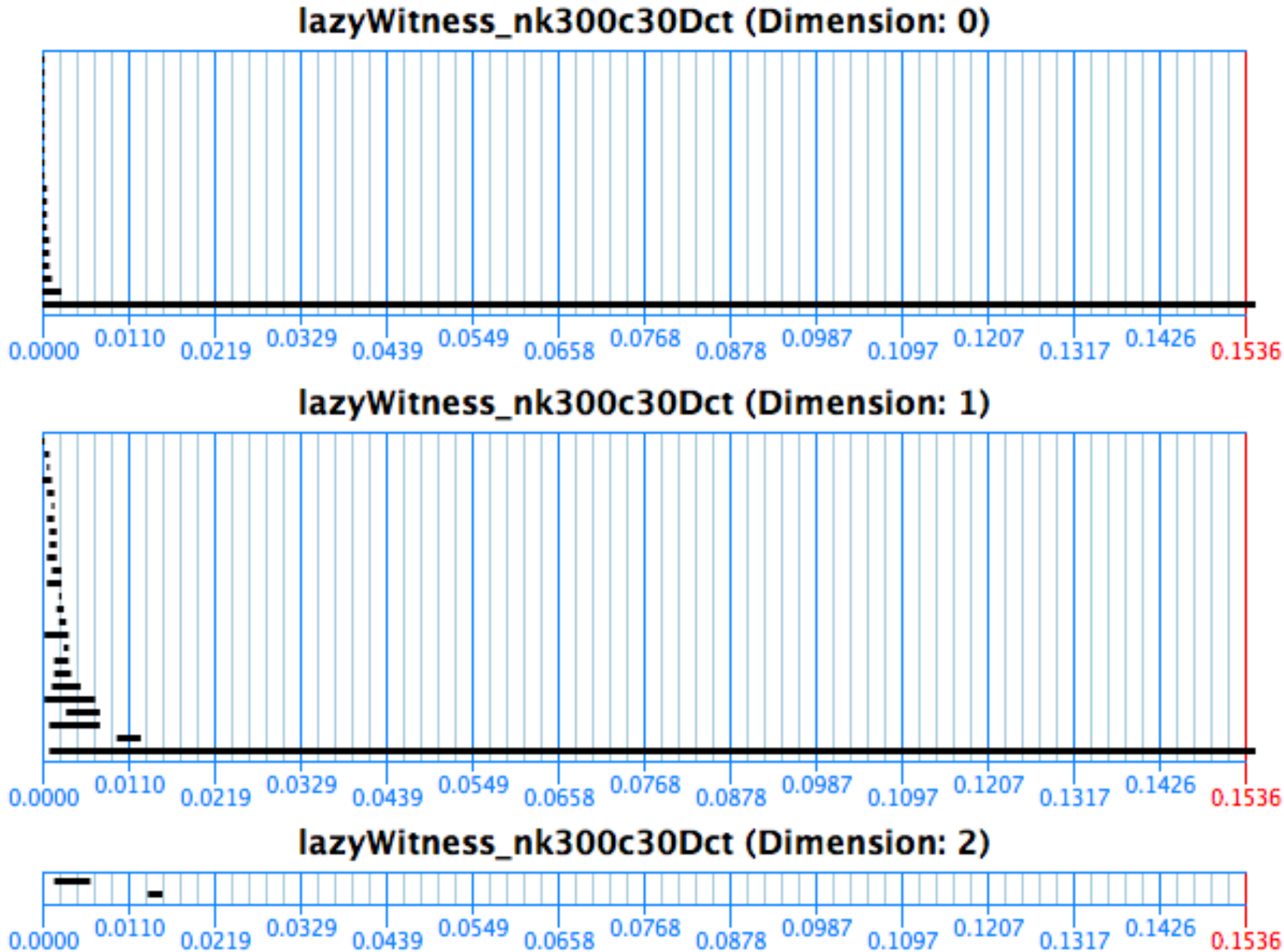
Points in 9-dimensional space, normalized to have average color gray and contrast norm one (on 7-sphere).



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

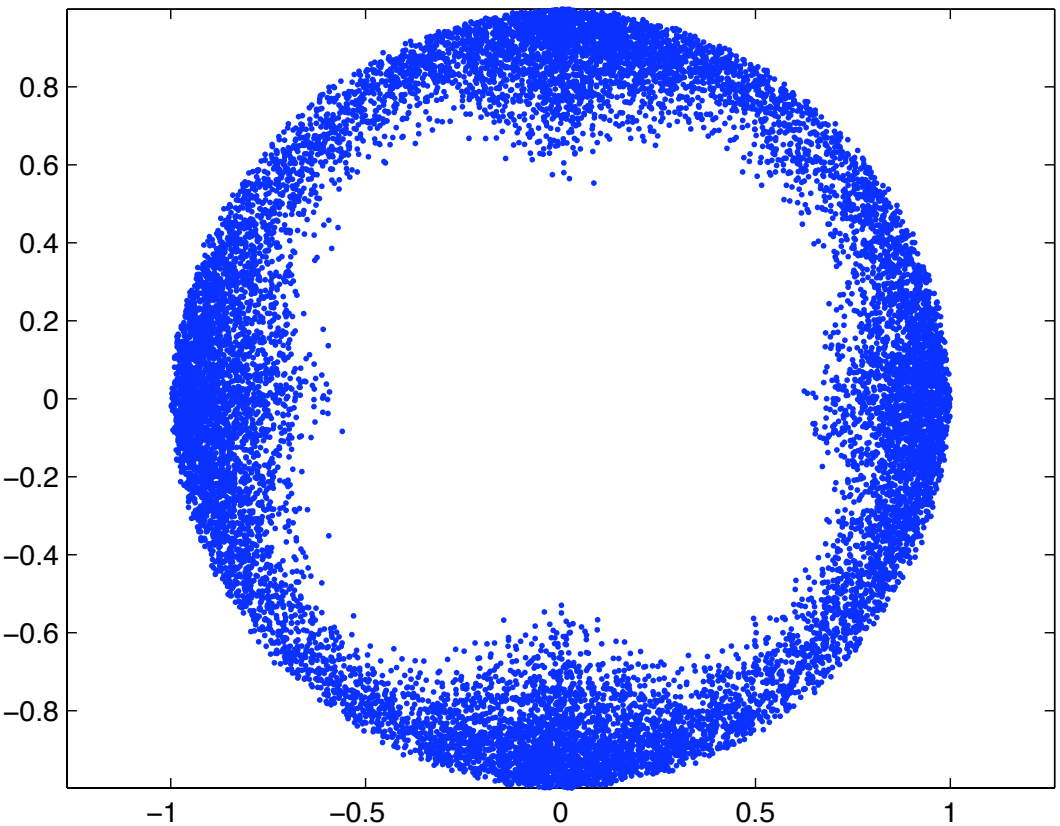
Persistent homology applied to data

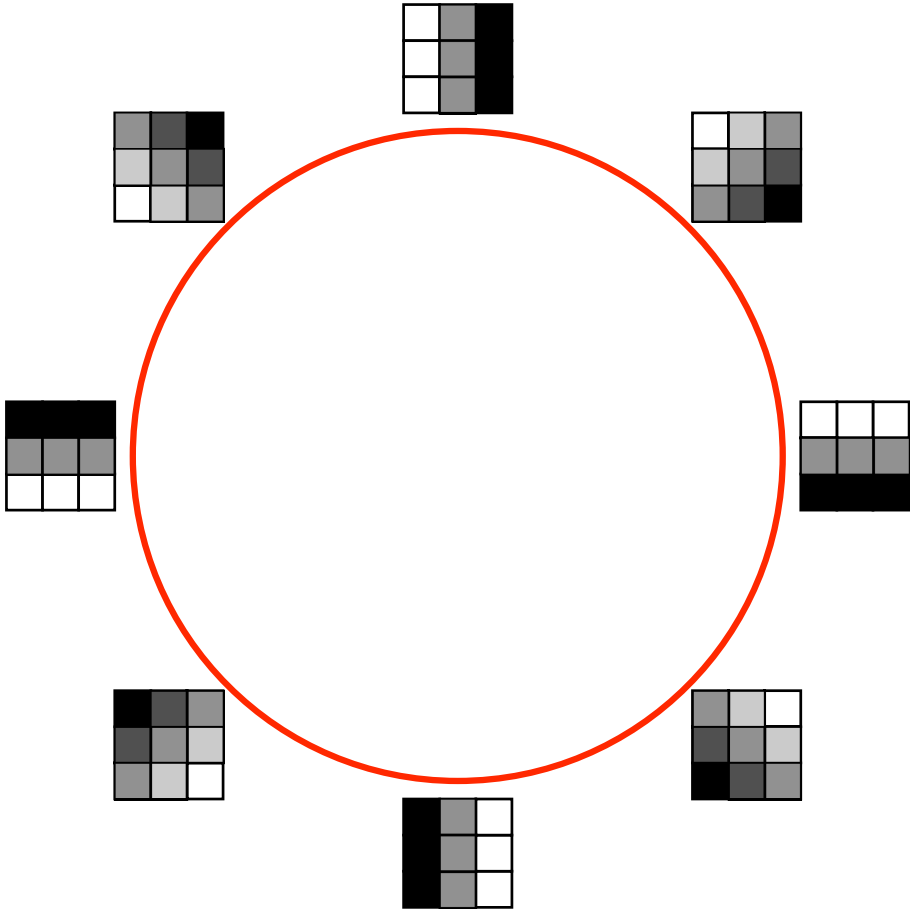
1. Densest patches according to a global estimate



Persistent homology applied to data

1. Densest patches according to a global estimate

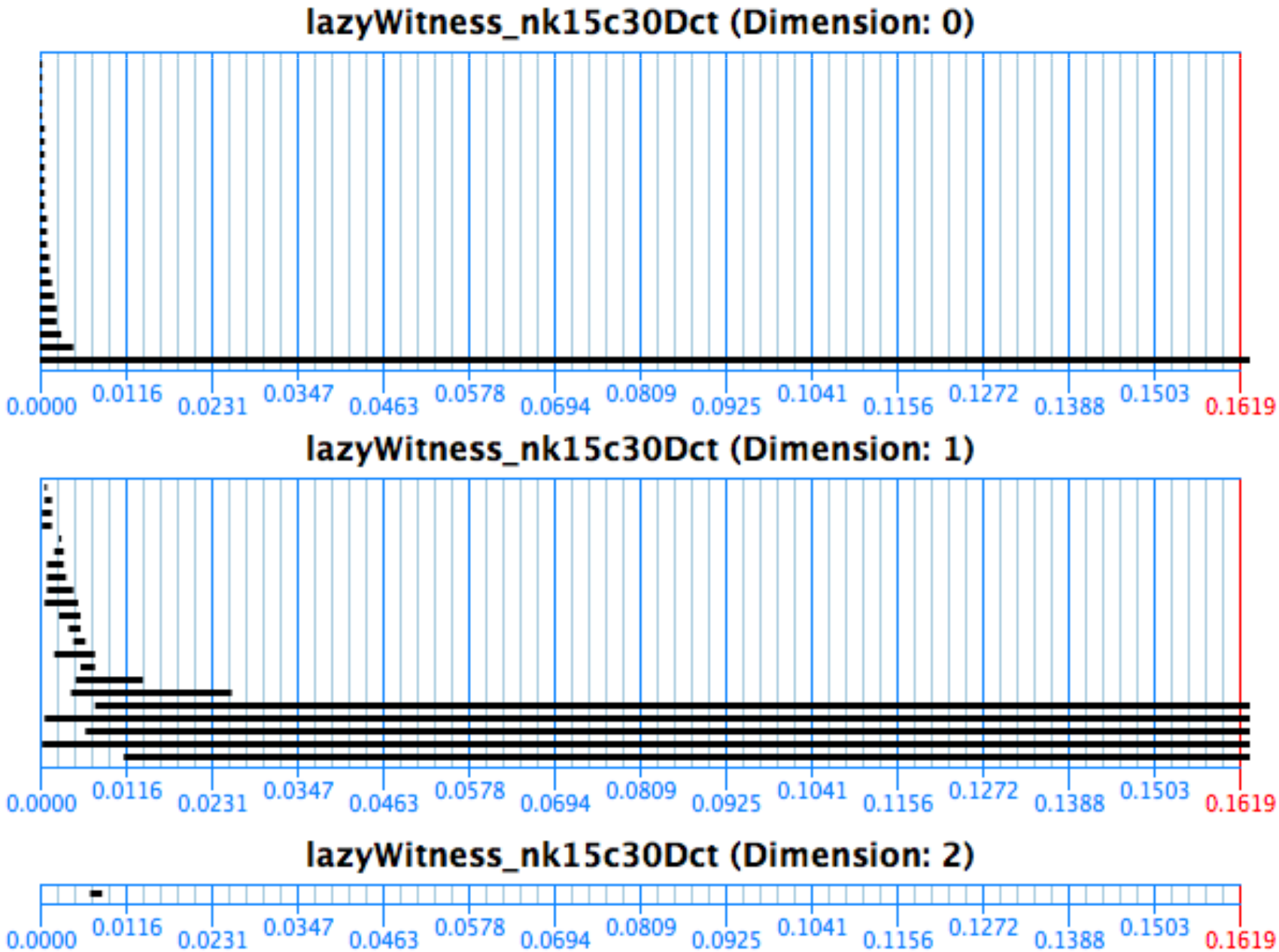




Interpretation: nature prefers linearity

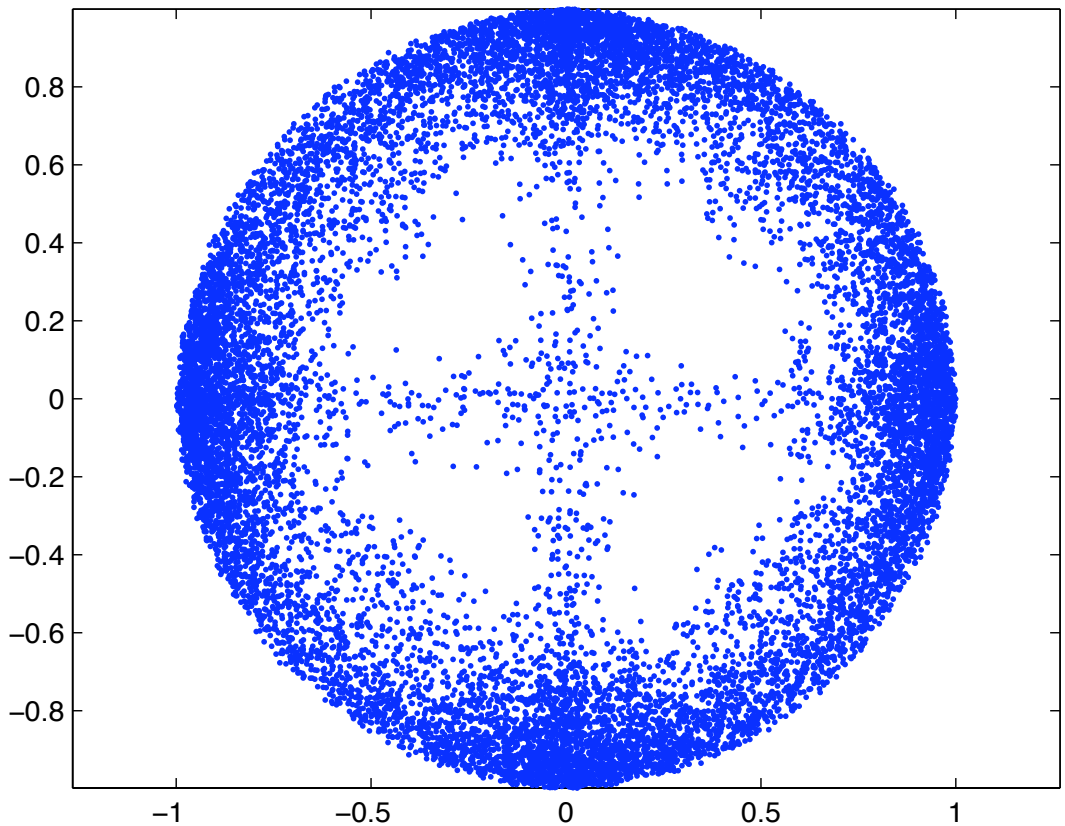
Persistent homology applied to data

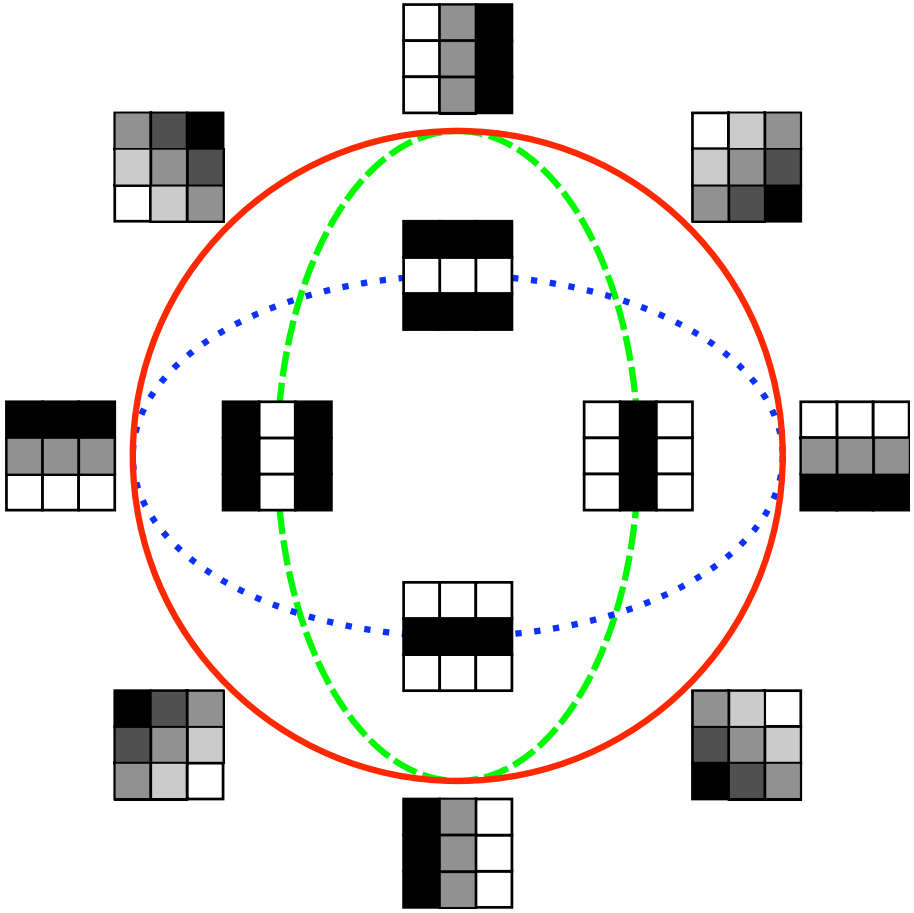
2. Densest patches according to an intermediate estimate



Persistent homology applied to data

2. Densest patches according to an intermediate estimate

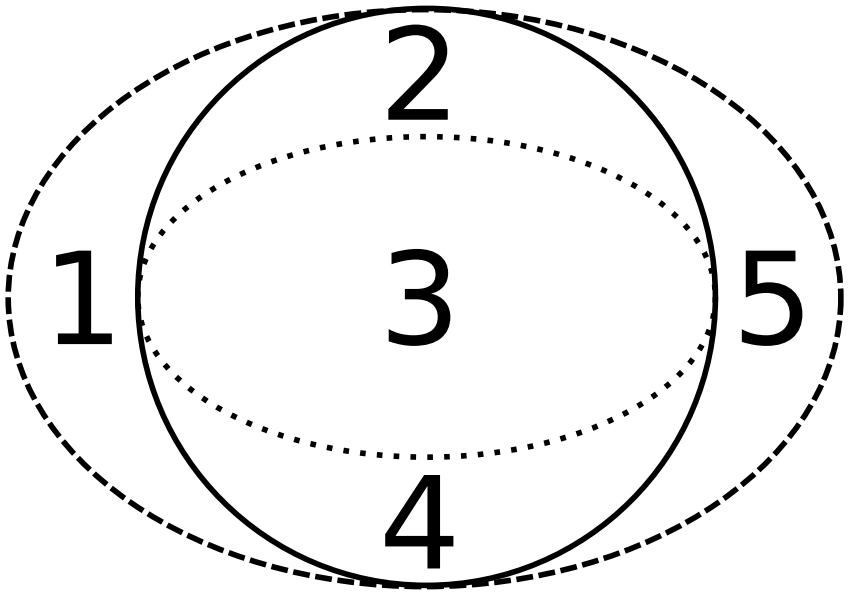


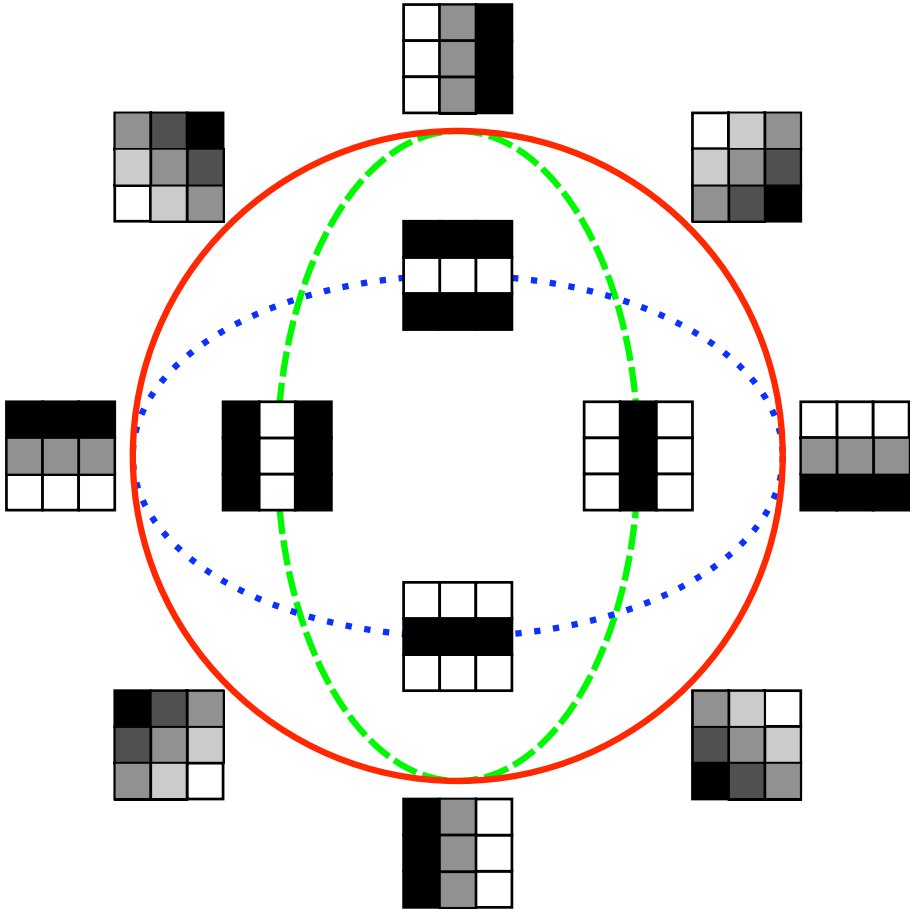


Interpretation: nature prefers horizontal and vertical directions

Persistent homology applied to data

2. Densest patches according to an intermediate estimate

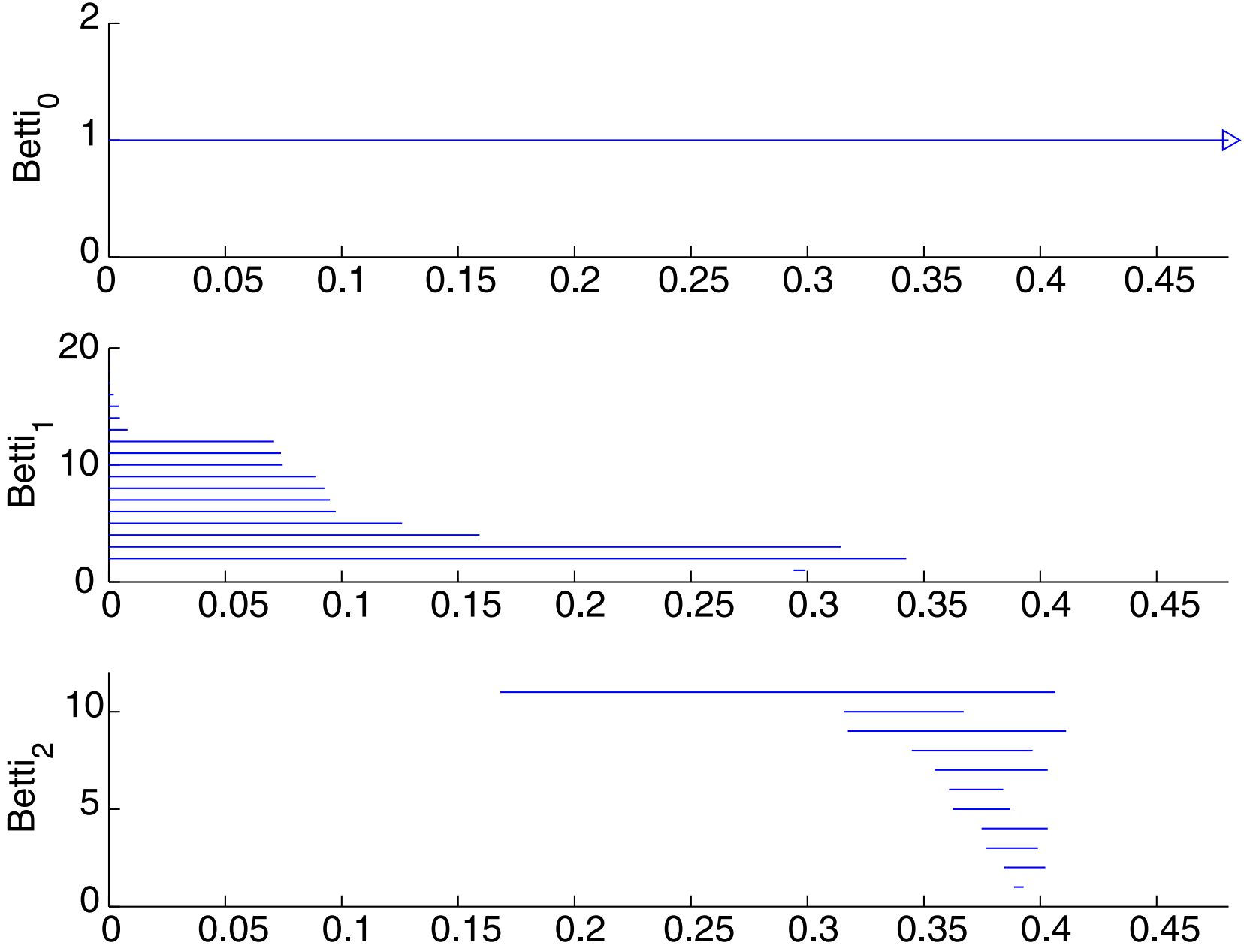




Interpretation: nature prefers horizontal and vertical directions

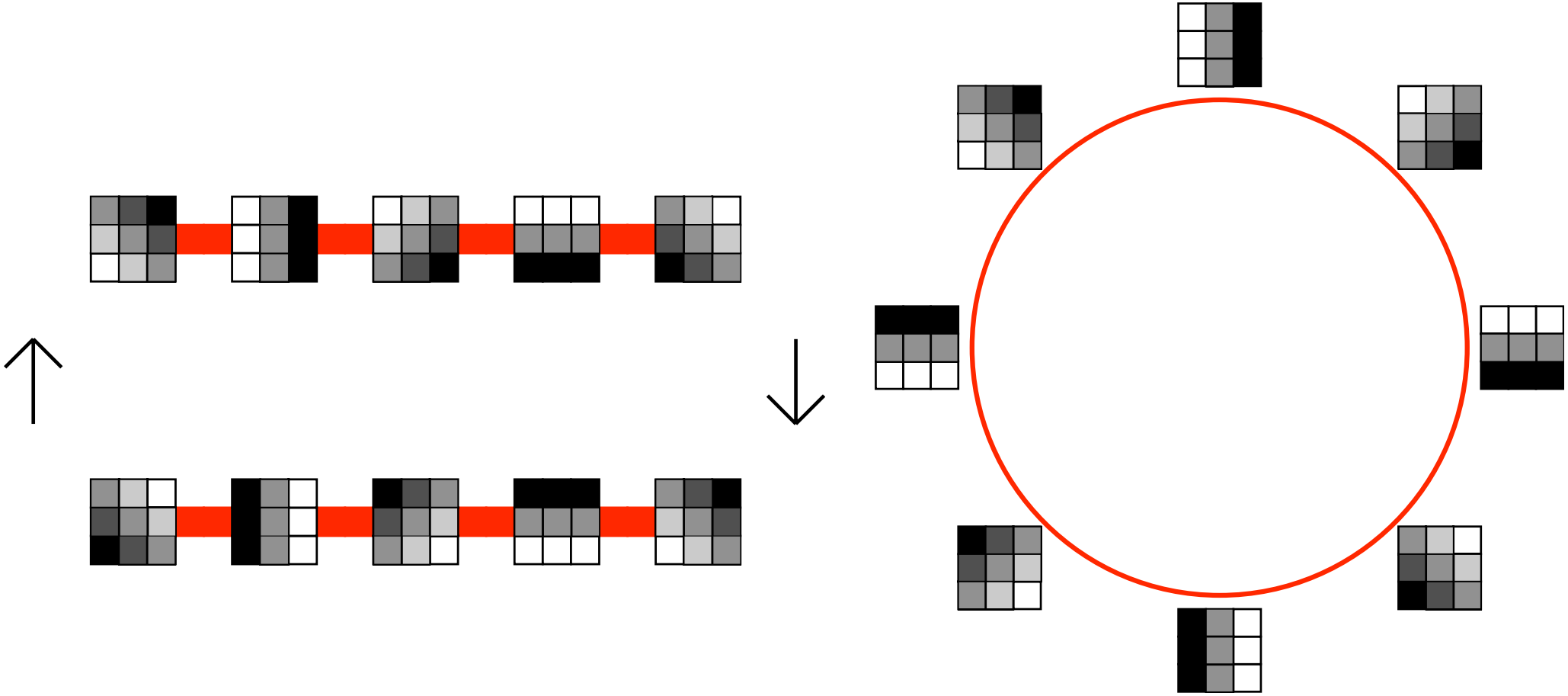
Persistent homology applied to data

3. Densest patches according to a local estimate



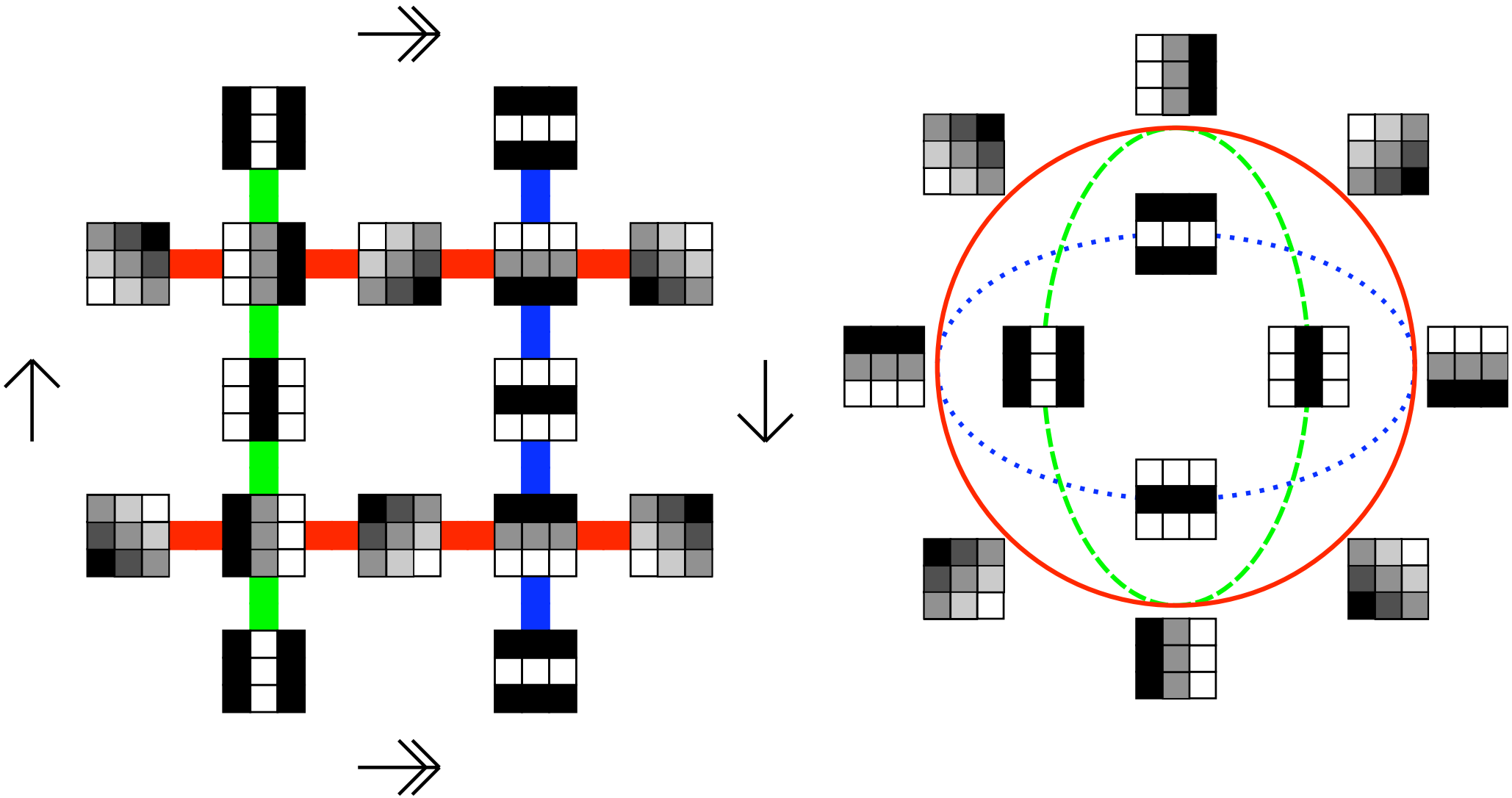
Persistent homology applied to data

3. Densest patches according to a local estimate



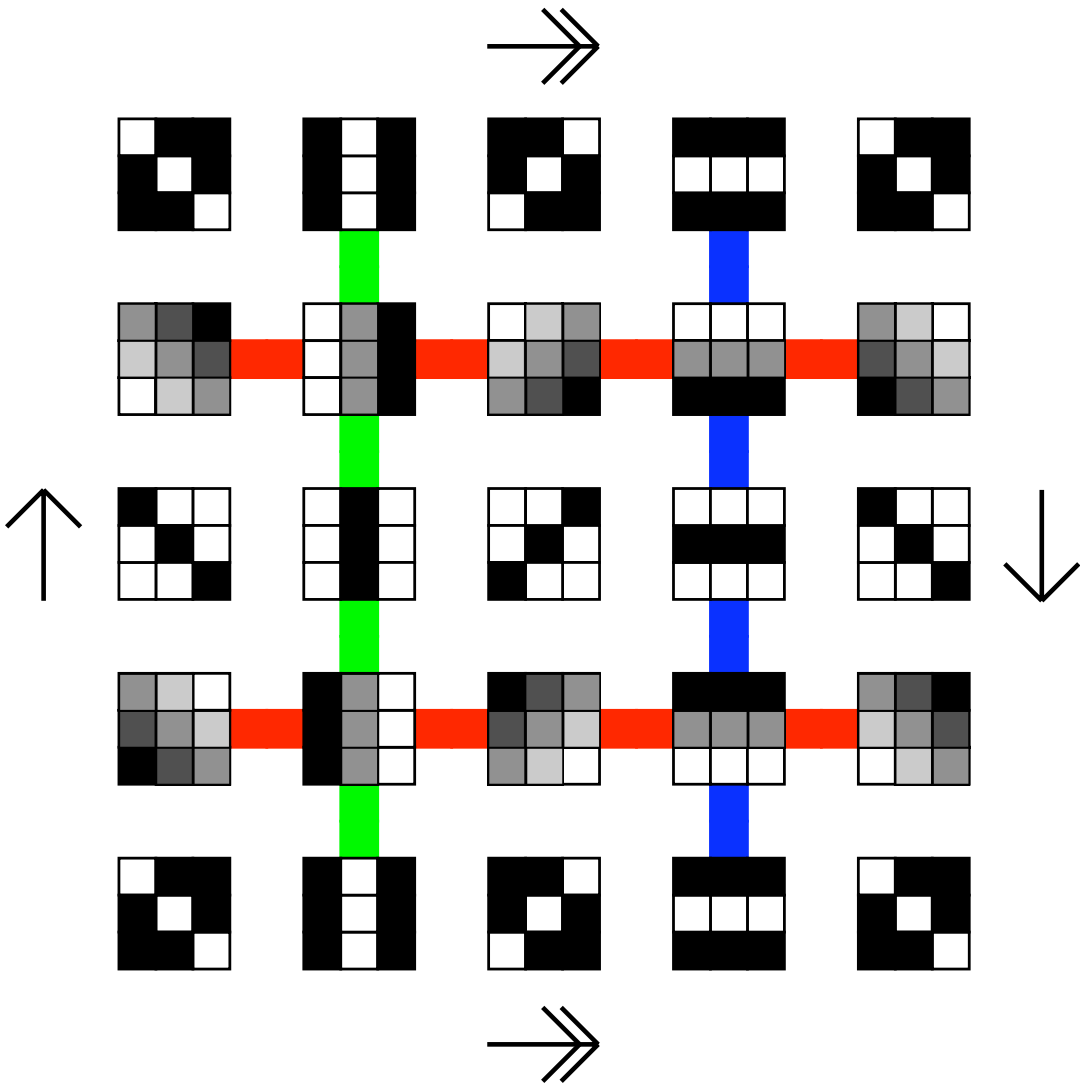
Persistent homology applied to data

3. Densest patches according to a local estimate



Persistent homology applied to data

3. Densest patches according to a local estimate



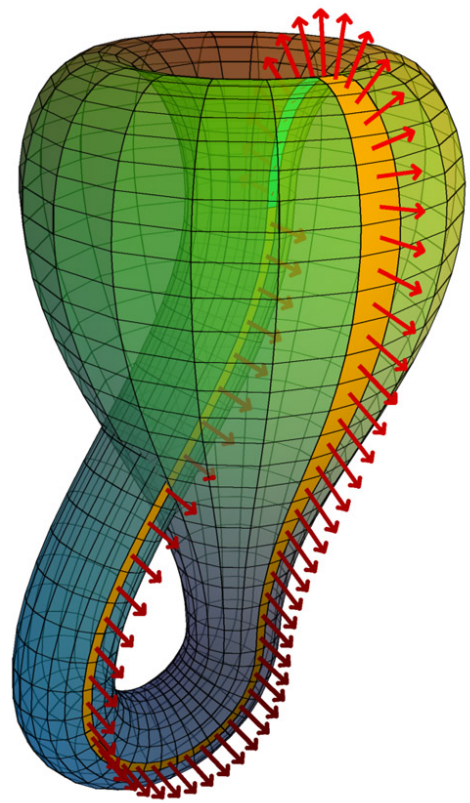
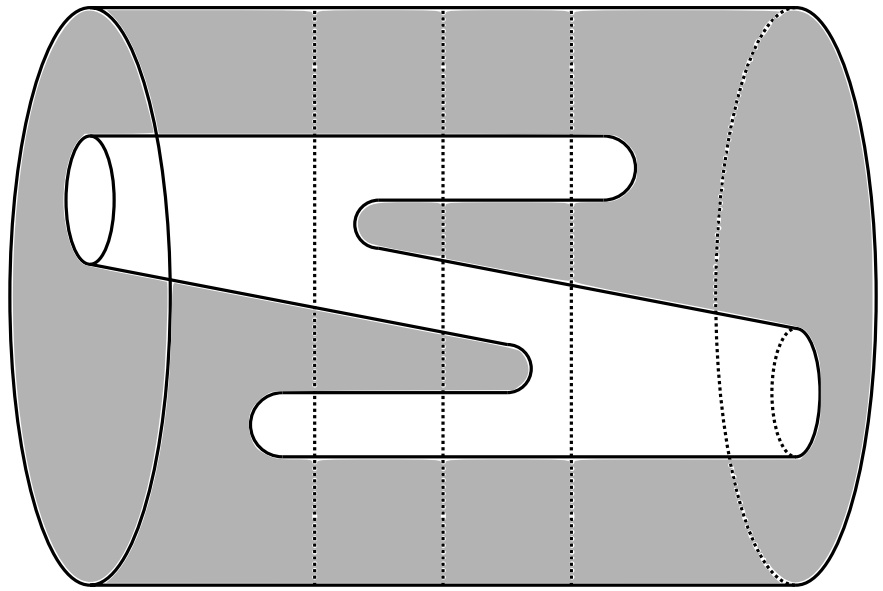


Image credit: https://plus.maths.org/content/imaging-maths-inside-klein-bottle

Interpretation: nature prefers linear and quadratic patches at all angles

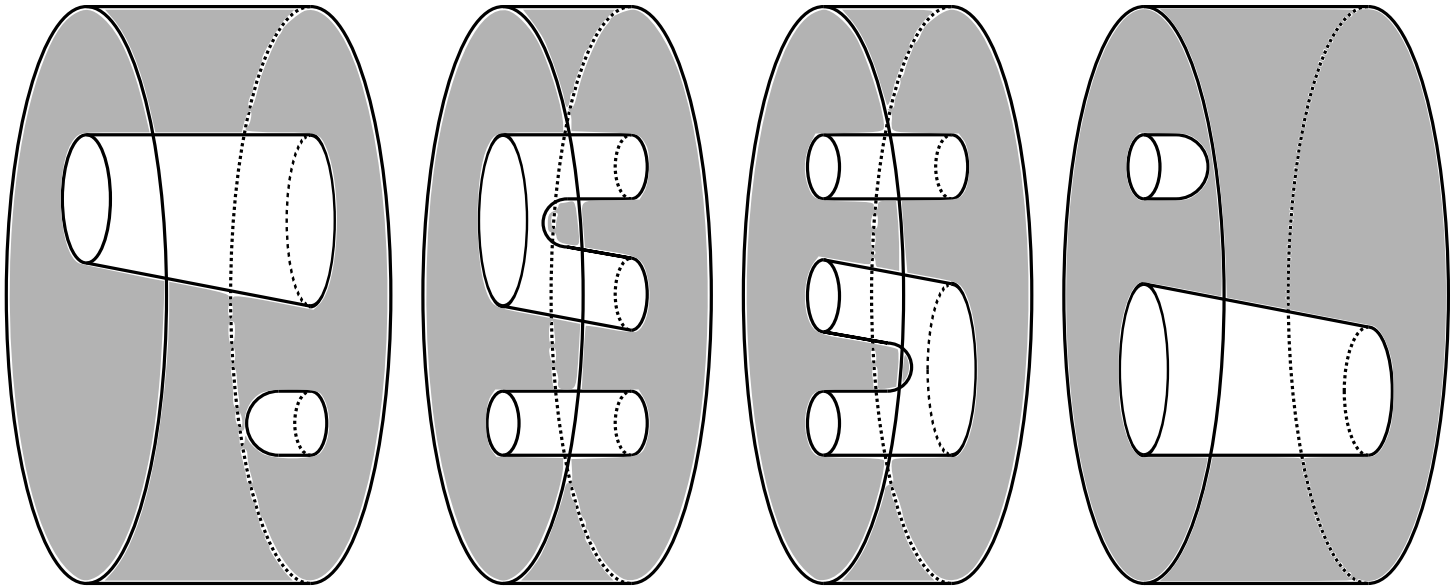
Zigzag persistent homology

Form zigzag module for $X \to I$ with $(d-1)$–dimensional homology.



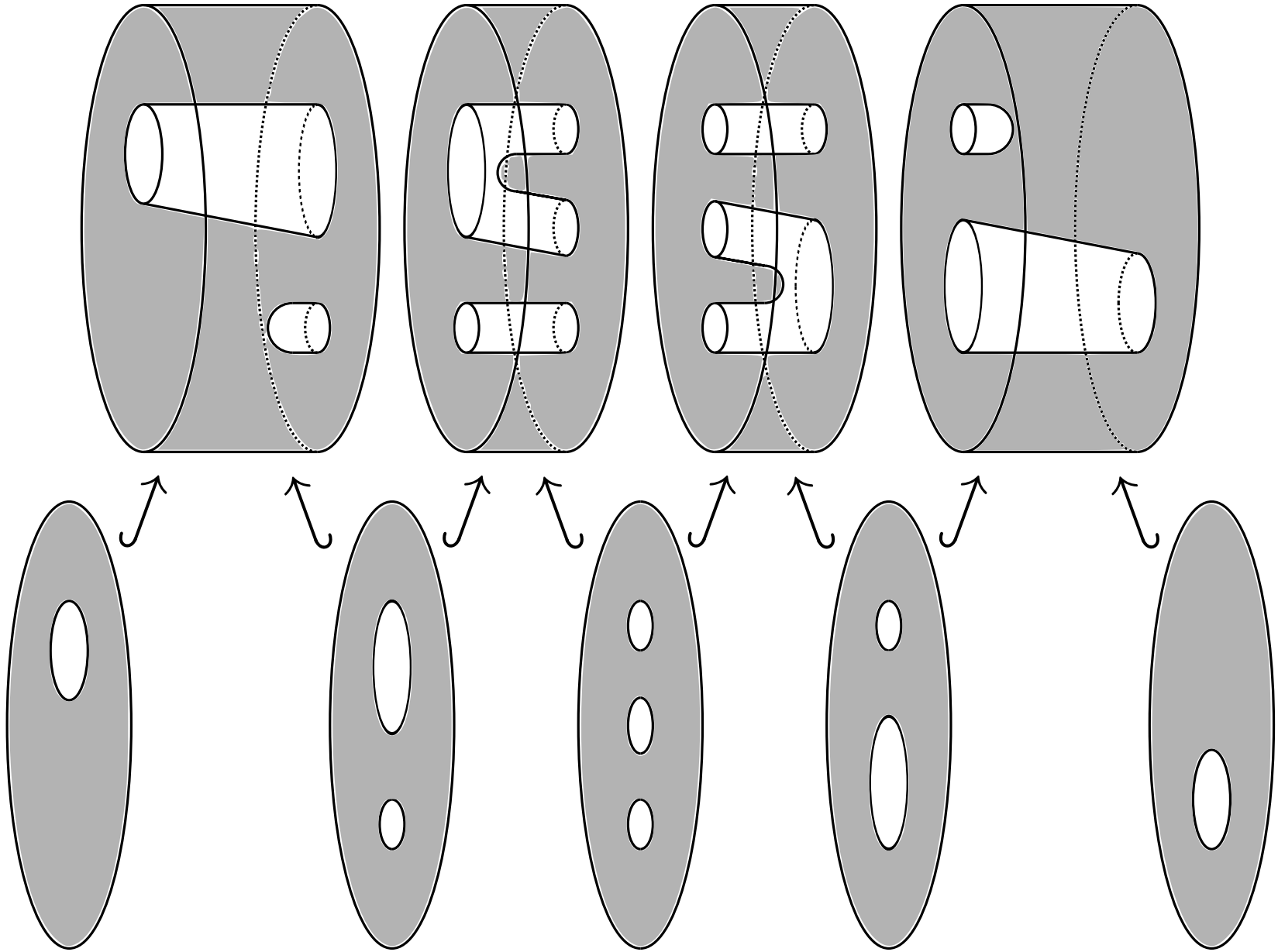
Zigzag persistent homology

Form zigzag module for $X \longrightarrow I$ with $(d-1)$–dimensional homology.



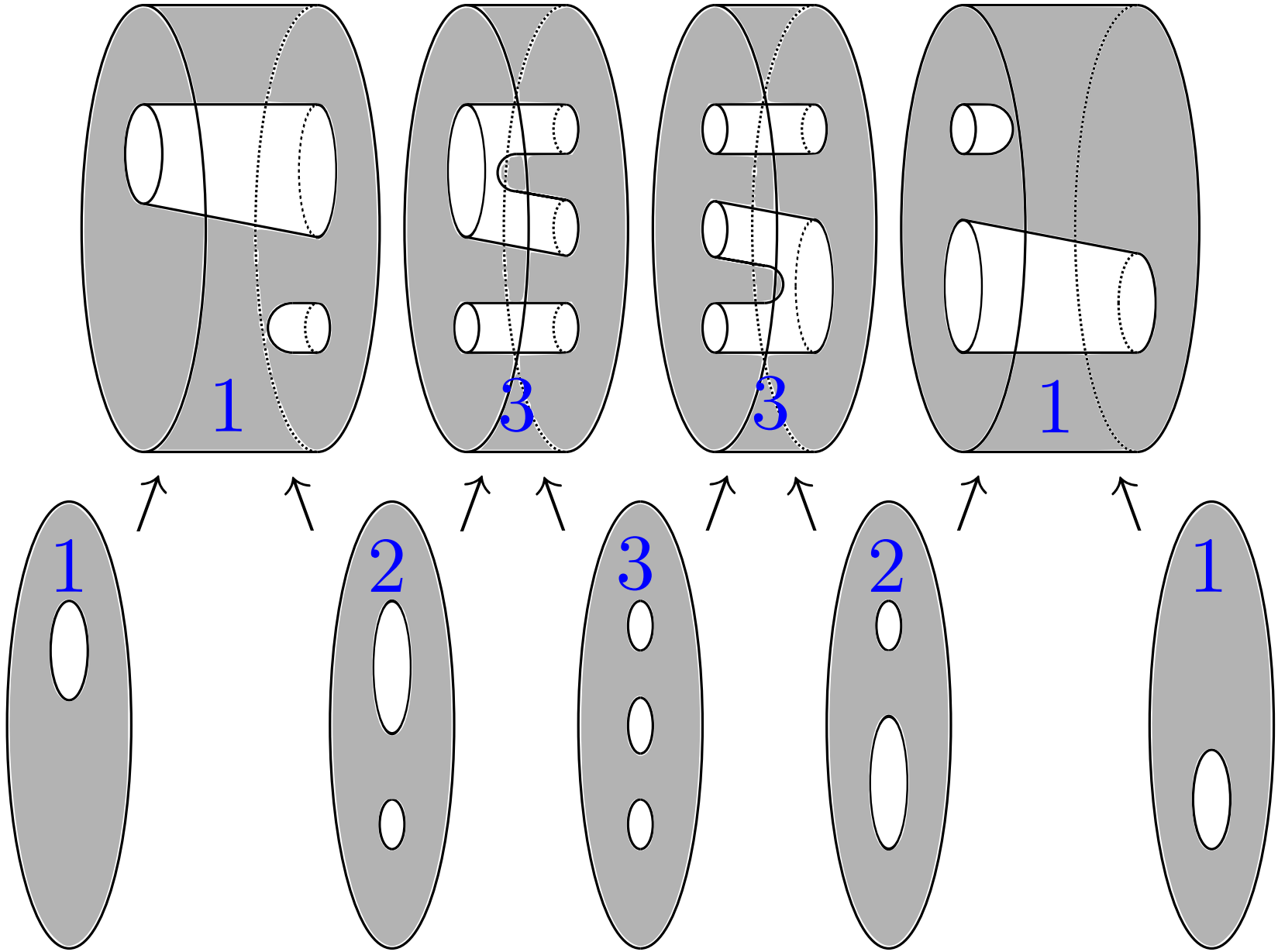
Zigzag persistent homology

Form zigzag module for $X \longrightarrow I$ with $(d-1)$–dimensional homology.



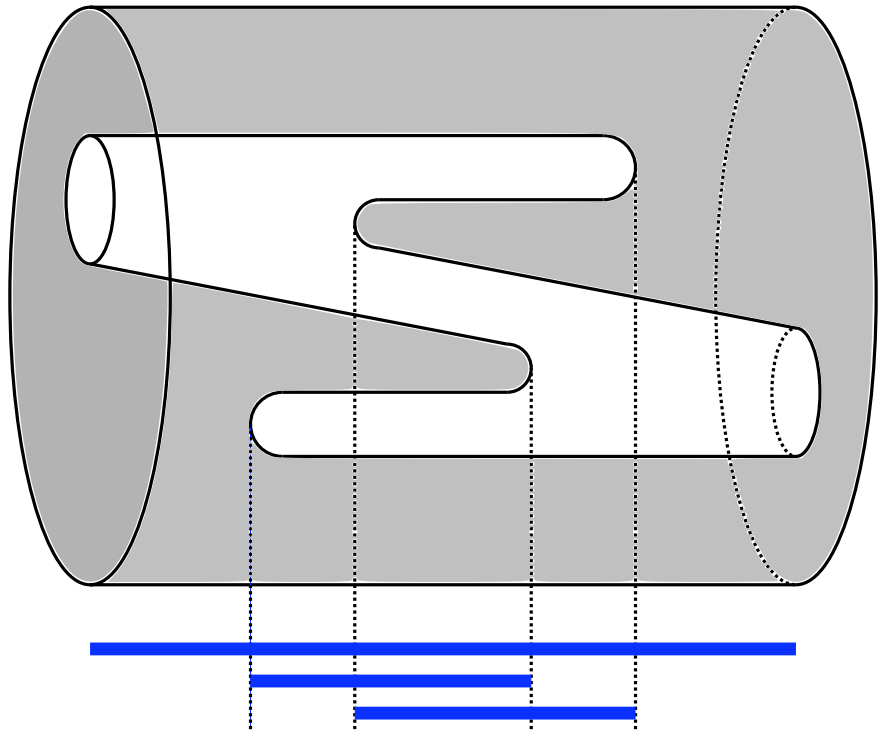
Zigzag persistent homology

Form zigzag module for $X \longrightarrow I$ with $(d-1)$–dimensional homology.



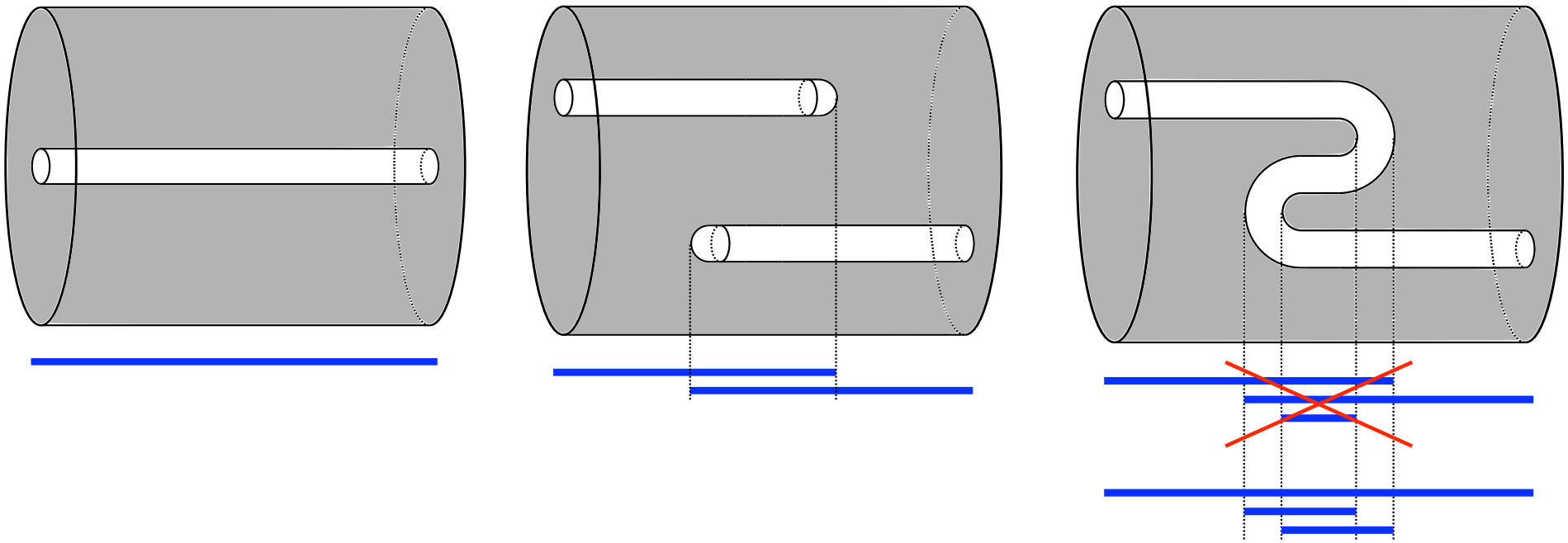
Zigzag persistent homology

Form zigzag module for $X \longrightarrow I$ with $(d-1)$–dimensional homology.



Zigzag persistent homology

Form zigzag module for $X \longrightarrow I$ with $(d-1)$–dimensional homology.



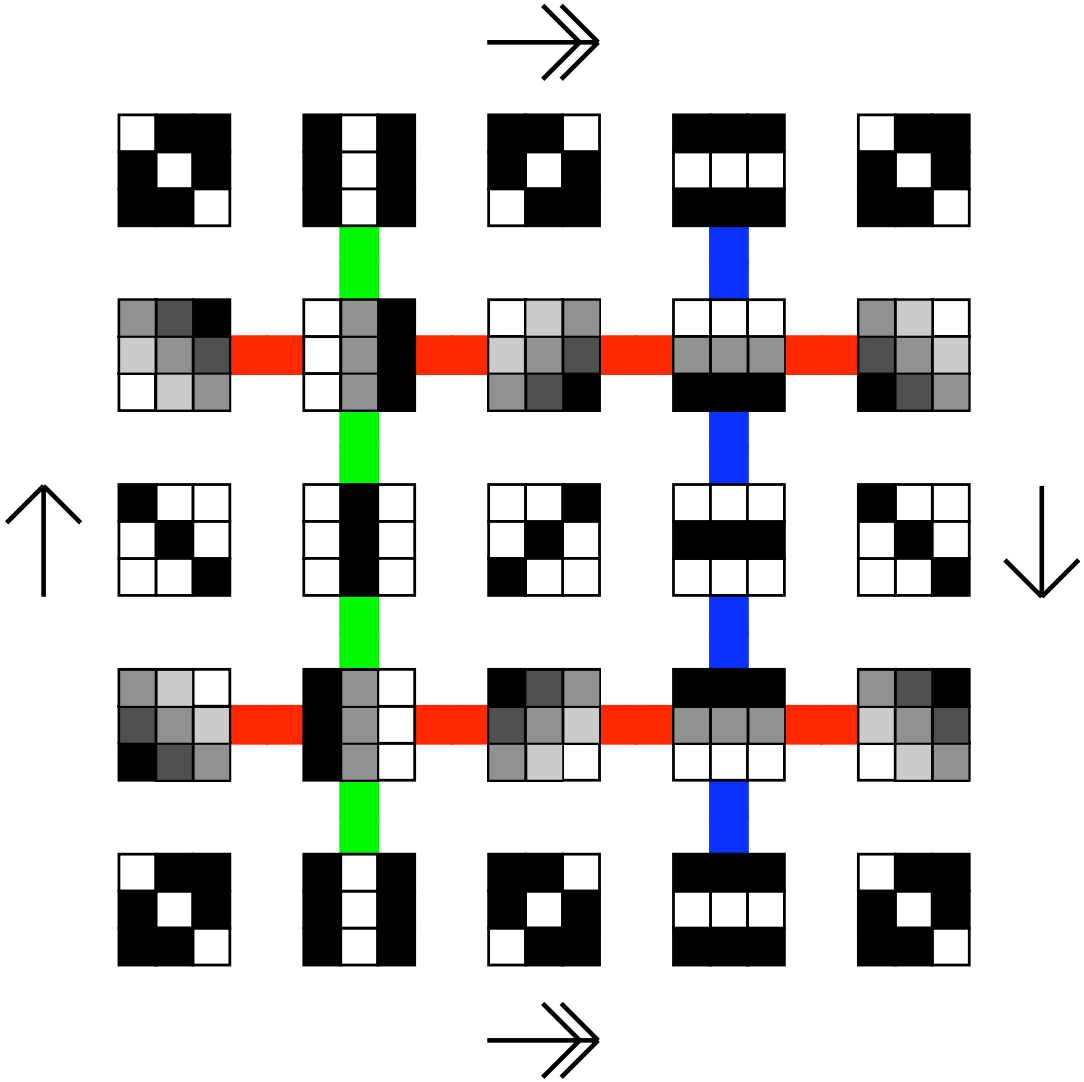
- <u>Theorem.</u>

  If there is an evasion path then there is a full-length bar.

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!