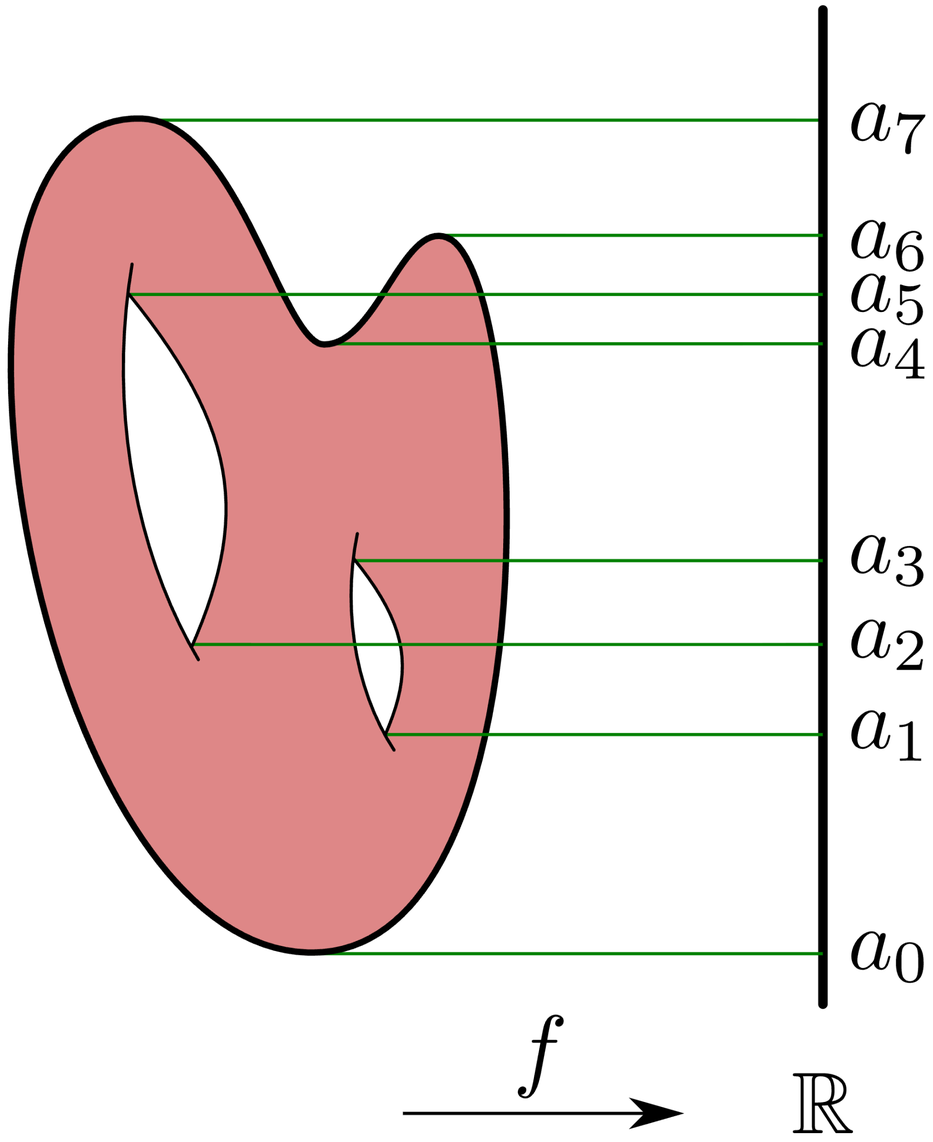
Reeb and Mapper Graphs

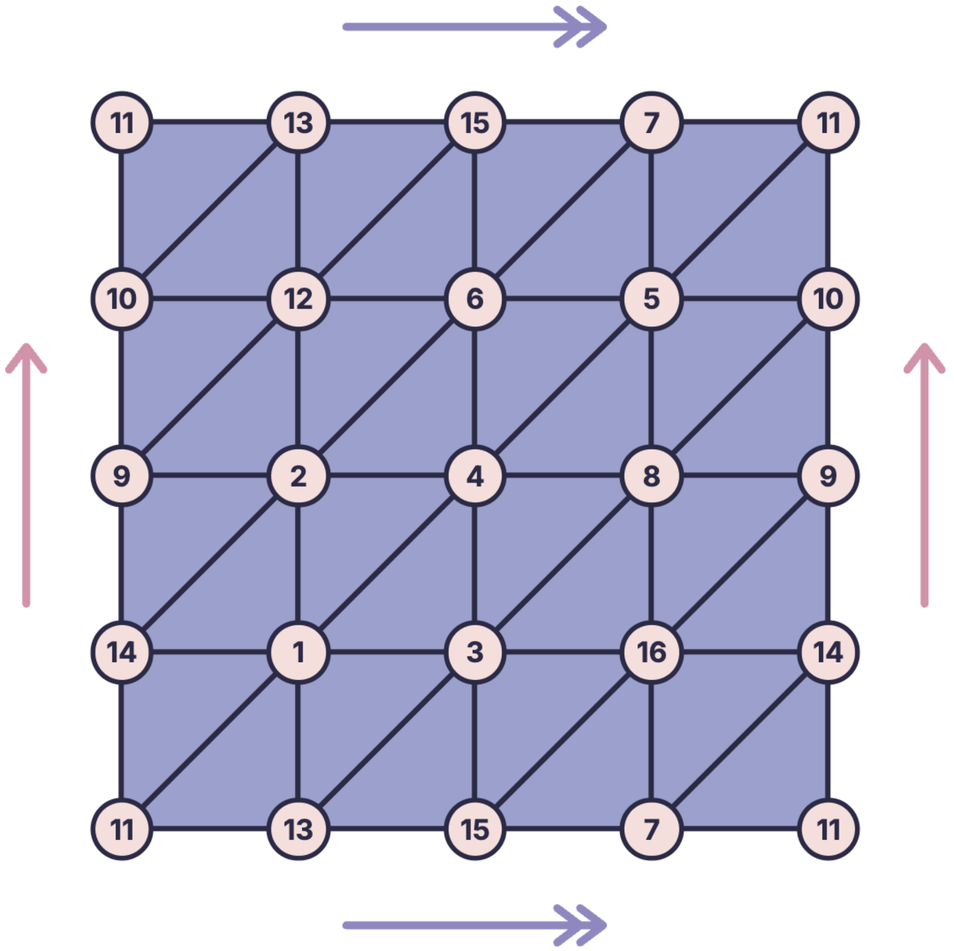
Dr. Elena Xinyi Wang

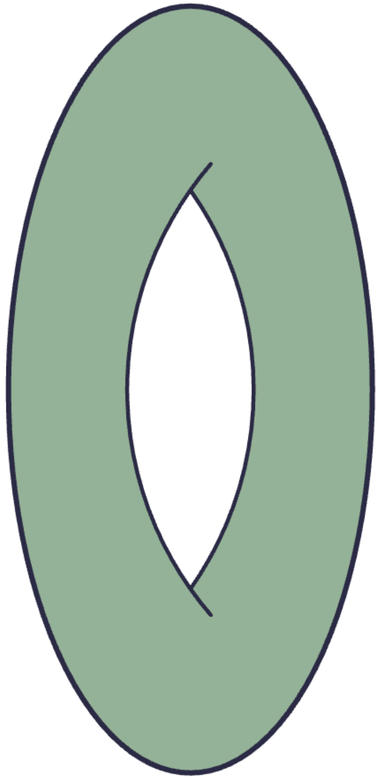
November 25, 2026

Height Function



Triangulation of a Torus





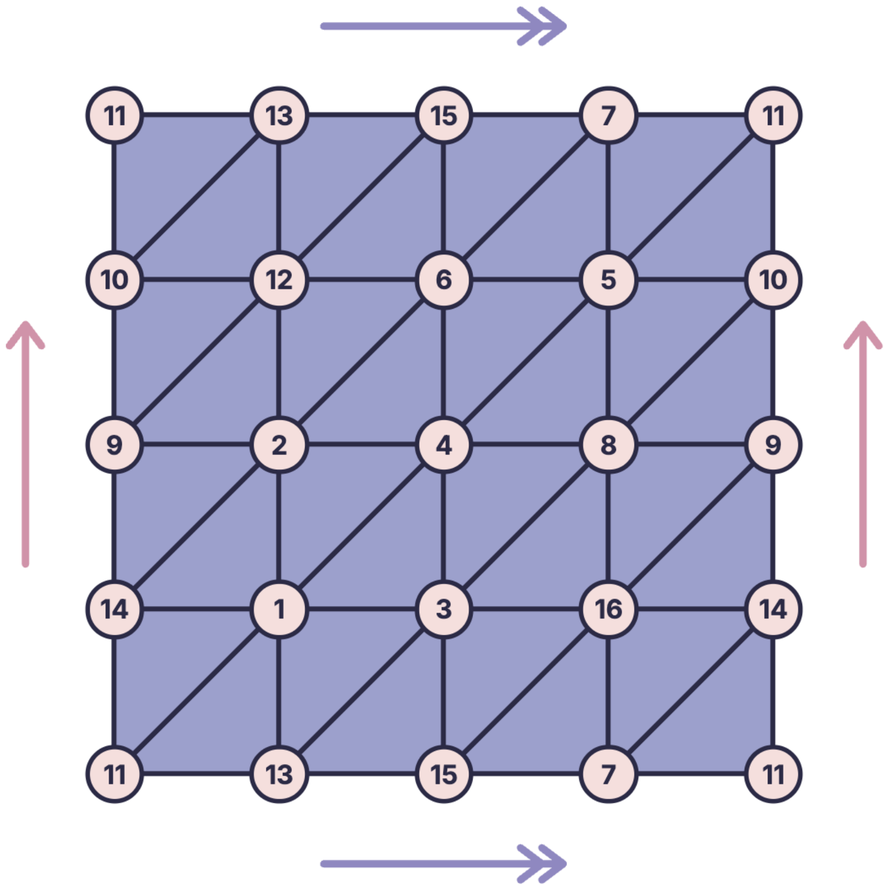
Level set

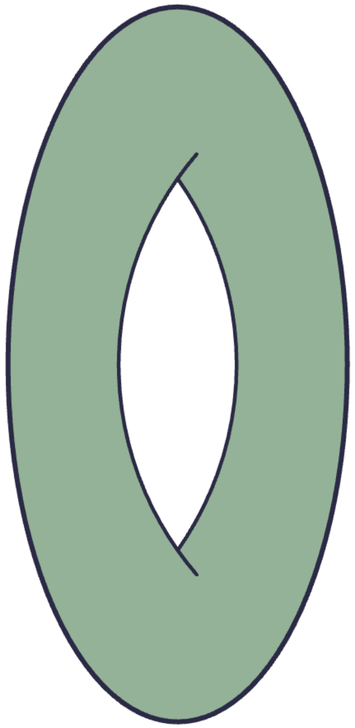
Given $f : |K| \to \mathbb{R}$.

A sublevel set: $f^{-1}(-\infty, a]$

A superlevel set: $f^{-1}[a, -\infty)$

A level set: $f^{-1}(a)$





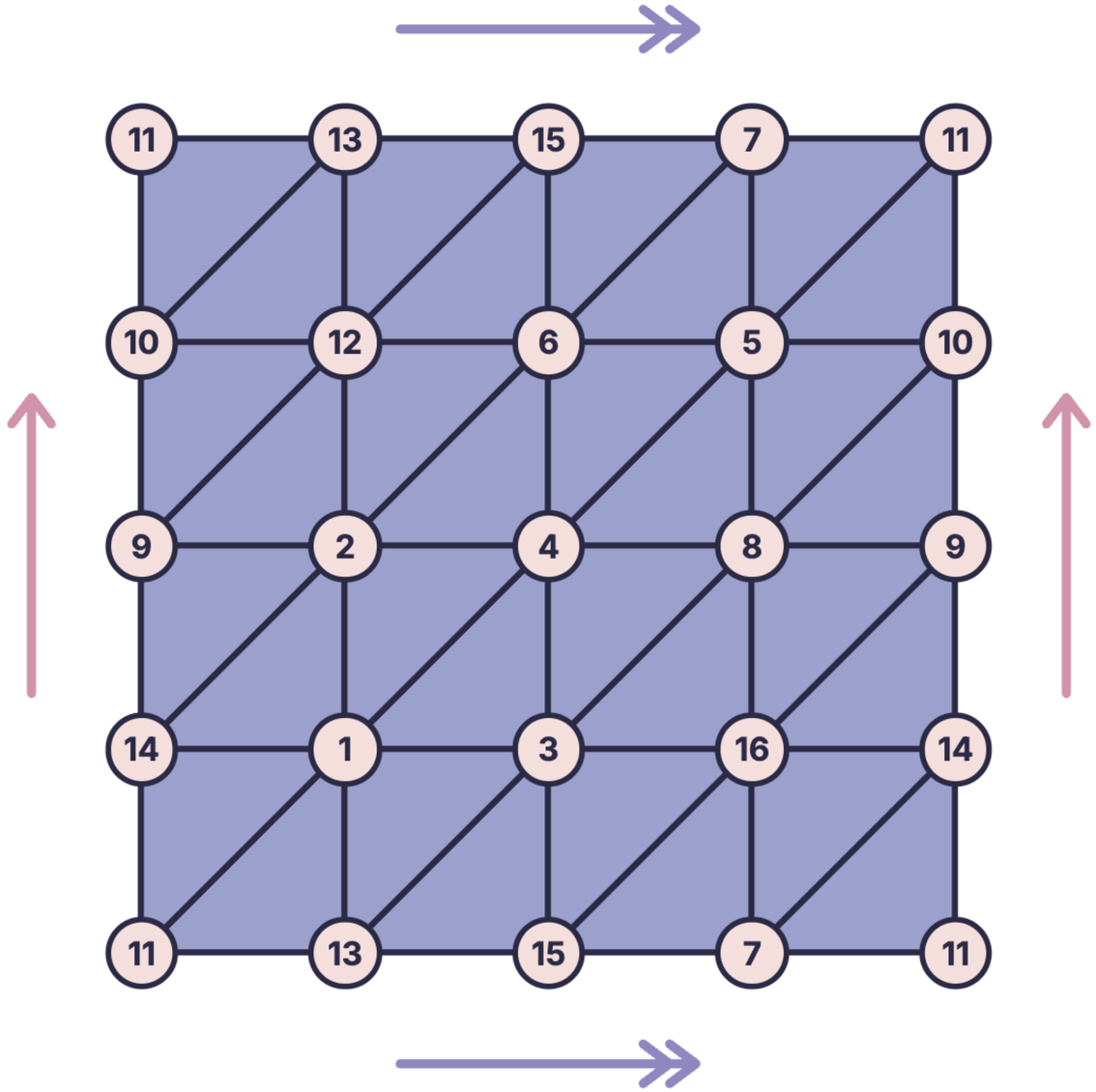
Try it

Draw the following level sets:

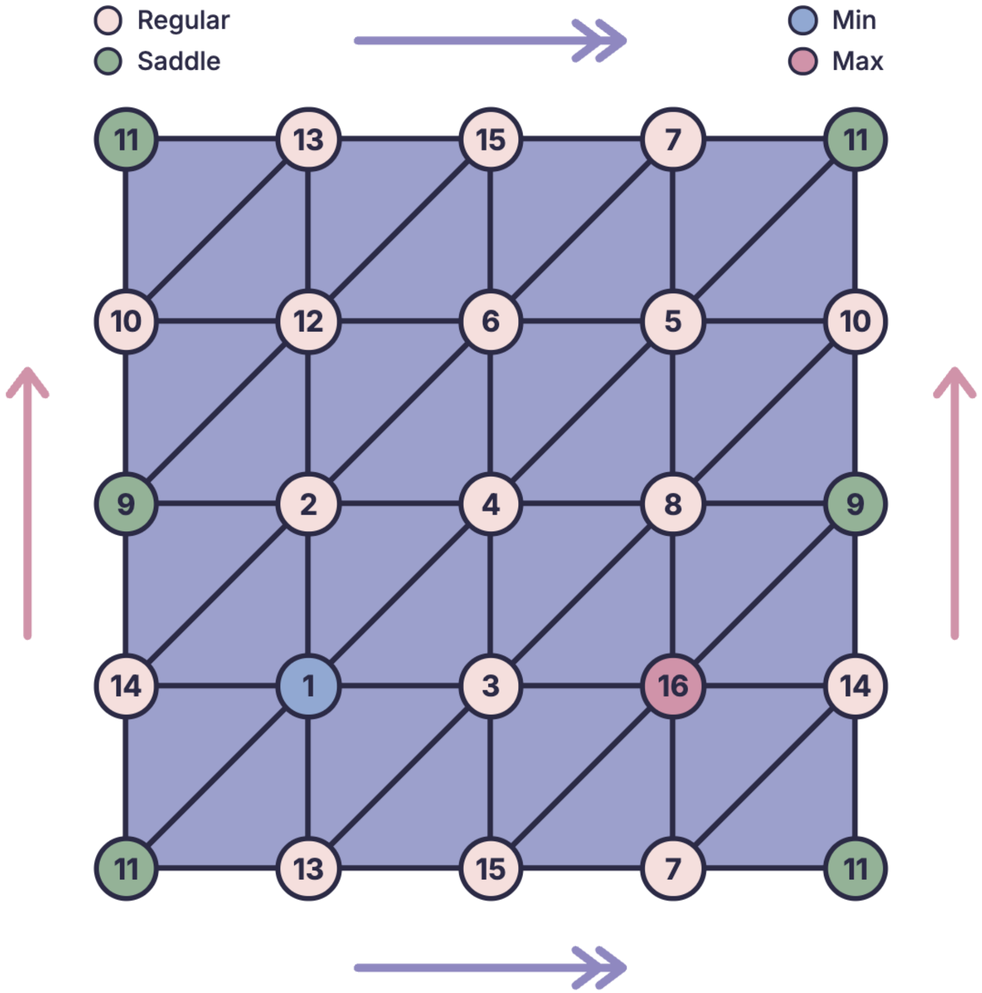
$f^{-1}(4.5)$

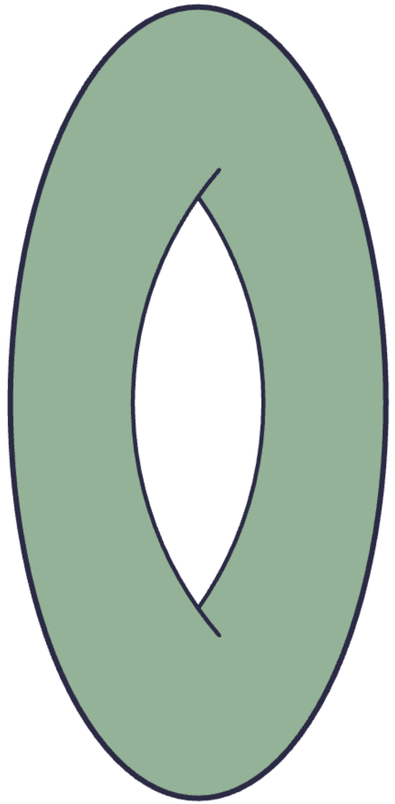
$f^{-1}(10.5)$

$f^{-1}(9)$



Match it to the height function





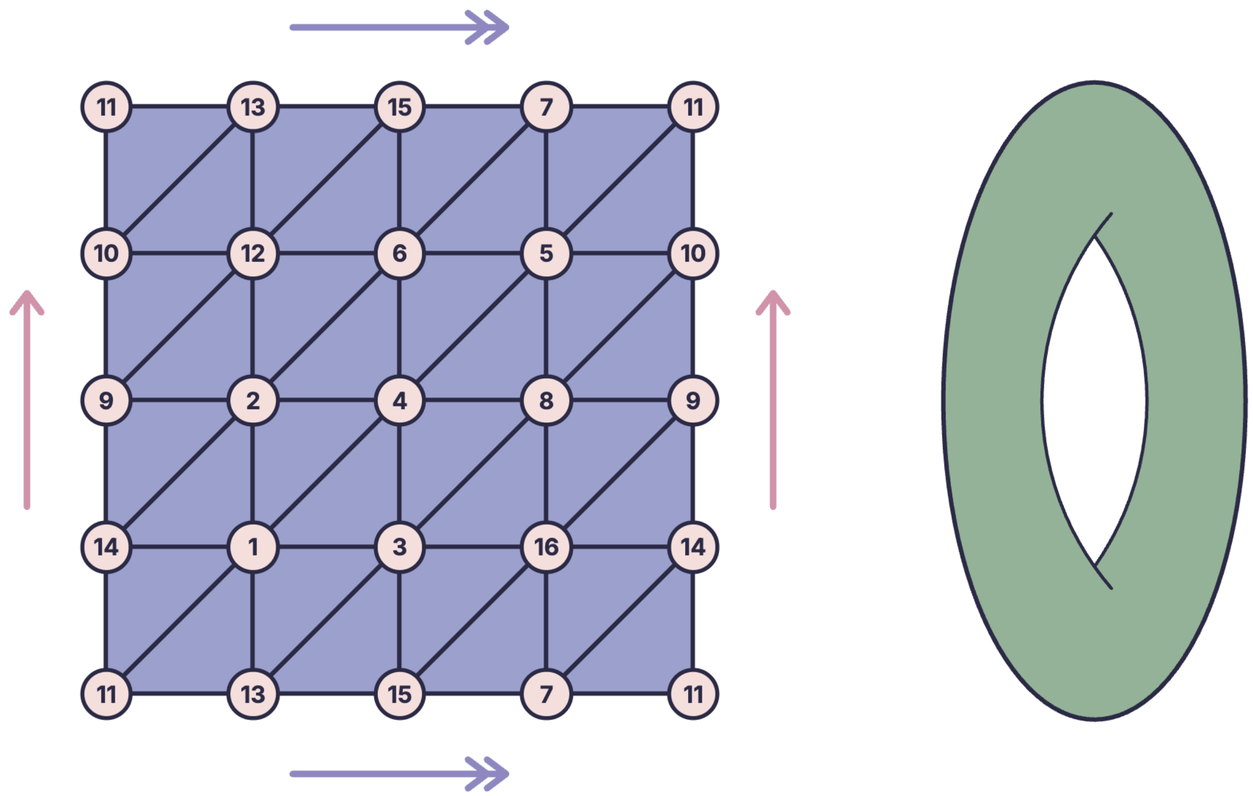
Recall: Connected Components

A space X is disconnected if there are disjoint open sets U and V with $X = U \cup V$.
A space is connected otherwise.

A connected component is a maximal subset that is connected.

Contours as level sets

A contour is a connected component of a level set.



How do the number of contours change as the level set function value changes?

Recall: Equivalence Relation

An equivalence relation $\sim$ on a set S satisfies:

- Reflexivity: $x \sim x$
- Symmetry: if $x \sim y$, then $y \sim x$
- Transitivity: if $x \sim y$ and $y \sim z$, then $x \sim z$

The *equivalence class* of x consists of all elements equivalent to x:

$$[x] = \{y \in S \mid y \sim x\}$$

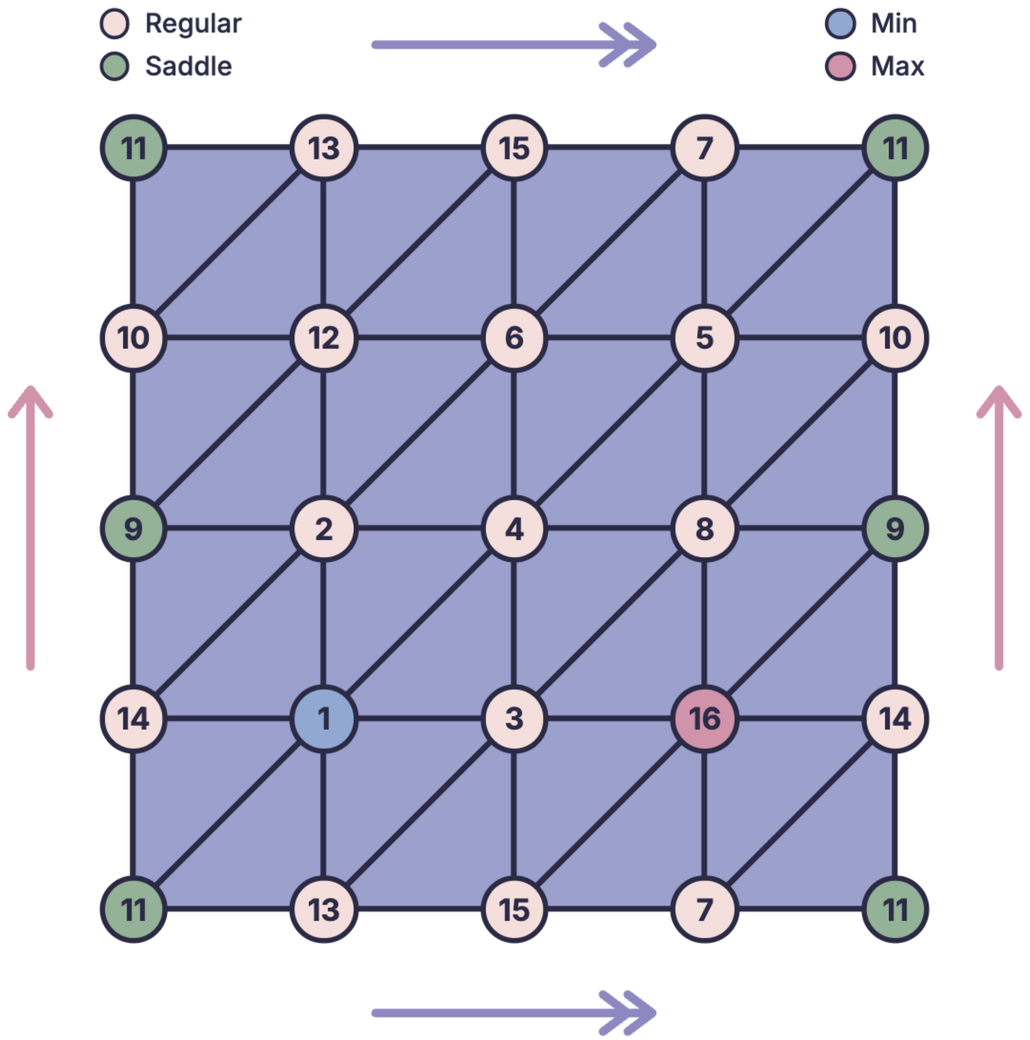
Reeb Graph Equivalence Relation

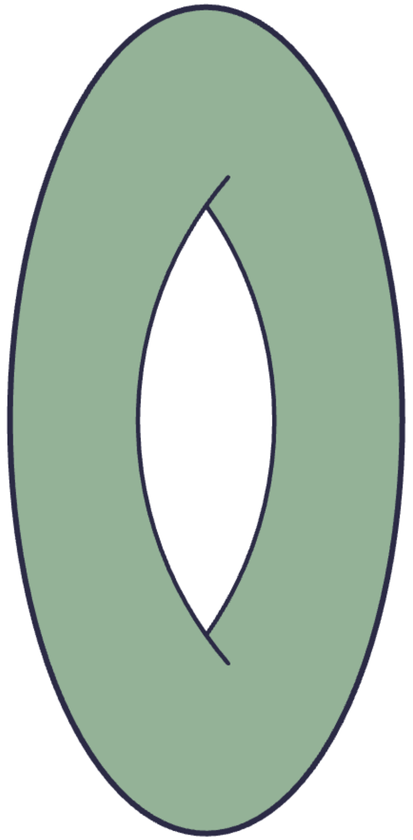
Given a function $f : X \to \mathbb{R}$.

Define an equivalence relation $\sim$ by $x \sim y$
if and only if:

- $f(x) = f(y) = \alpha$
- x and y are in the same connected component of the level set $f^{-1}(a)$.

Equivalence relation example



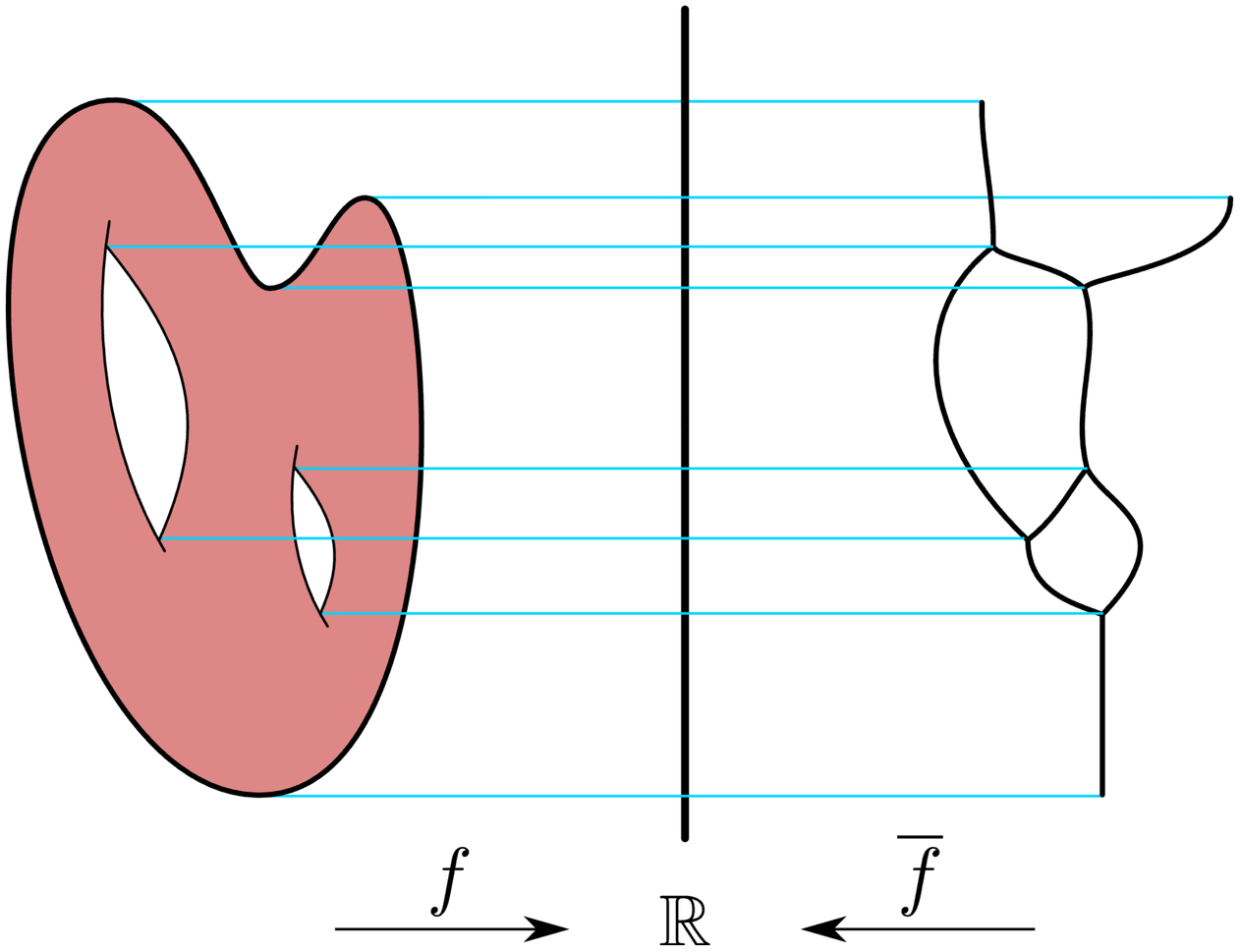


Recall: Quotient Space

Let X be a topological space with an equivalence relation $\sim$

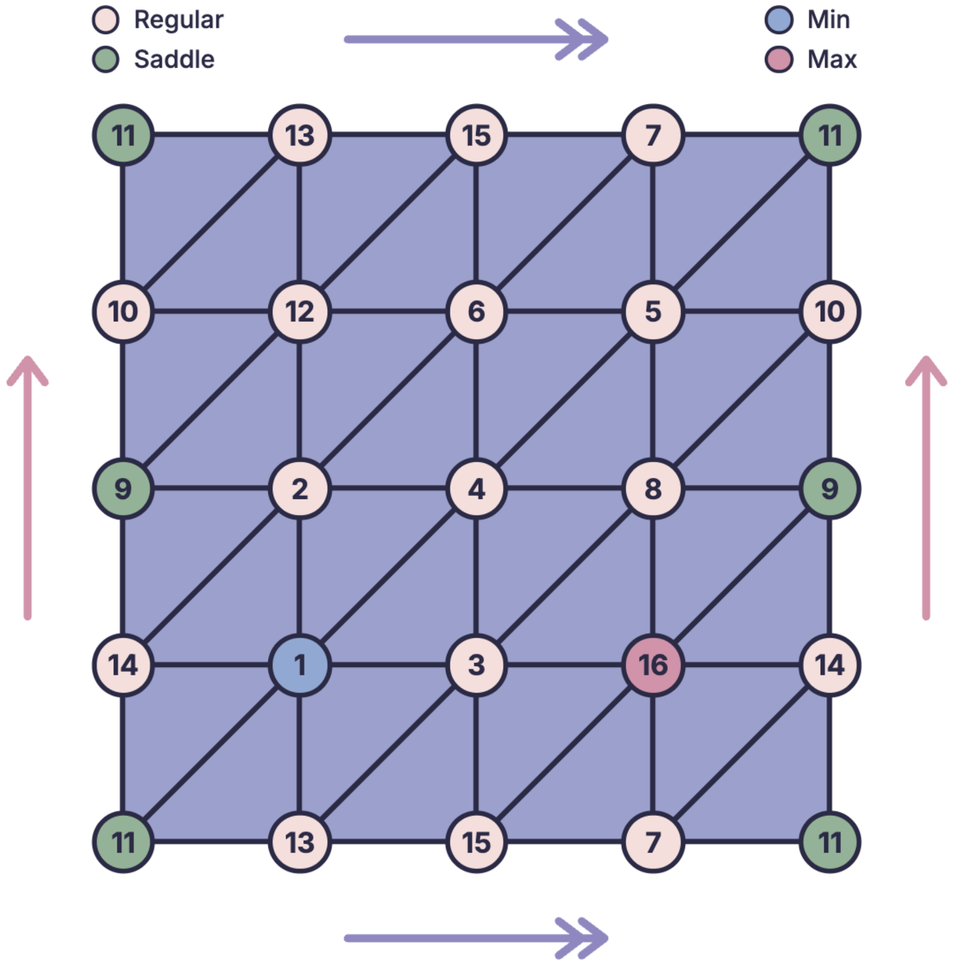
- The quotient space $X/\sim$ has one point for each equivalence class.
- The quotient map $q : X \to X/\sim$ sends each point x to its class $[x]$.
- A subset $U \subseteq X/\sim$ is open when $q^{-1}(U)$ is open in X

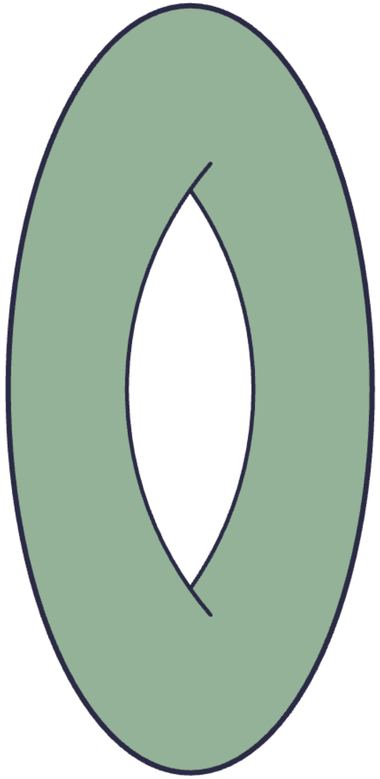
Reeb Graph



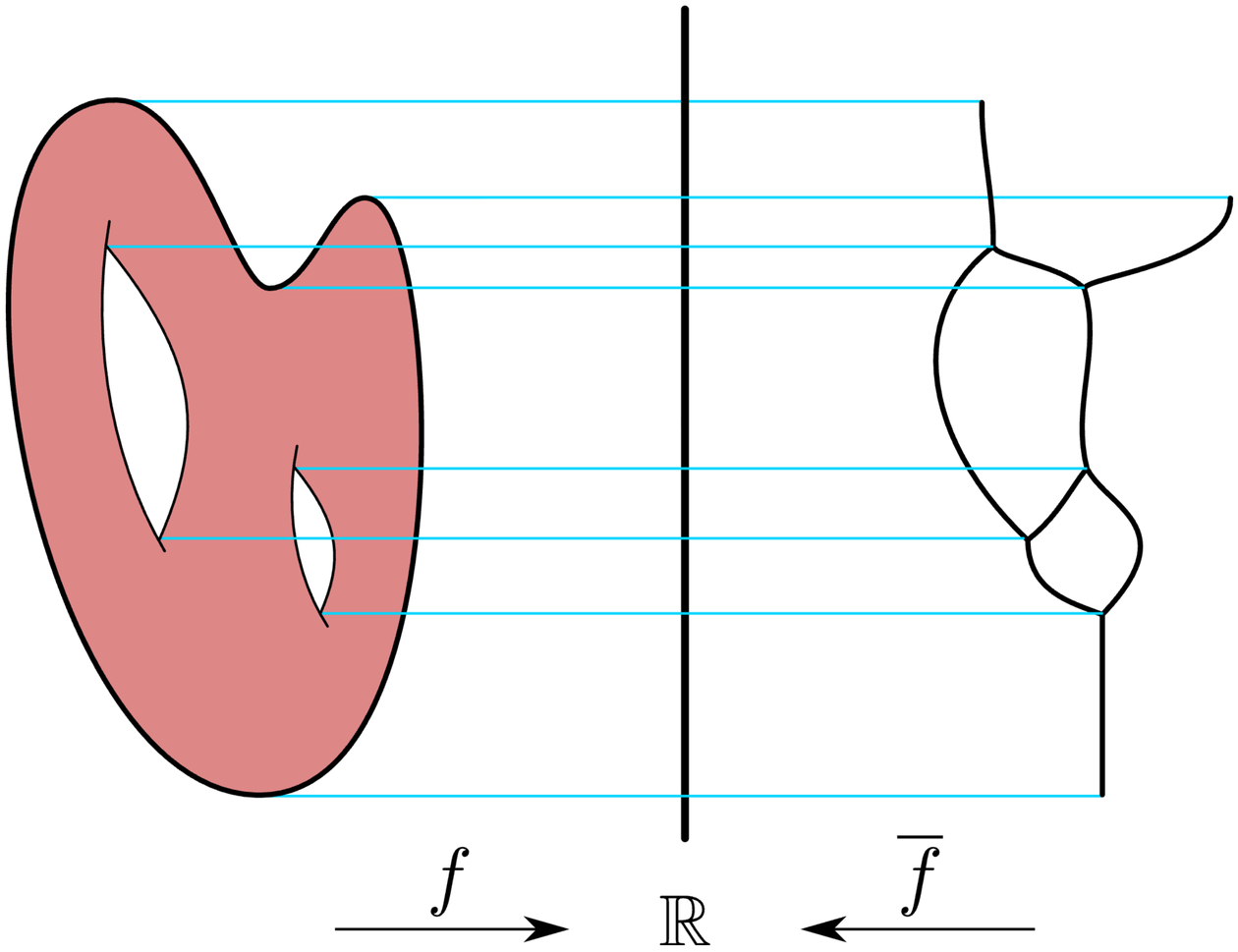
Let $[x]$ denote the equivalence class of $x \in X$. The Reeb graph R_f of $f : X \to \mathbb{R}$ is the quotient space $X/\sim$ with quotient map $q(x) = [x]$.

Computing the Reeb Graph



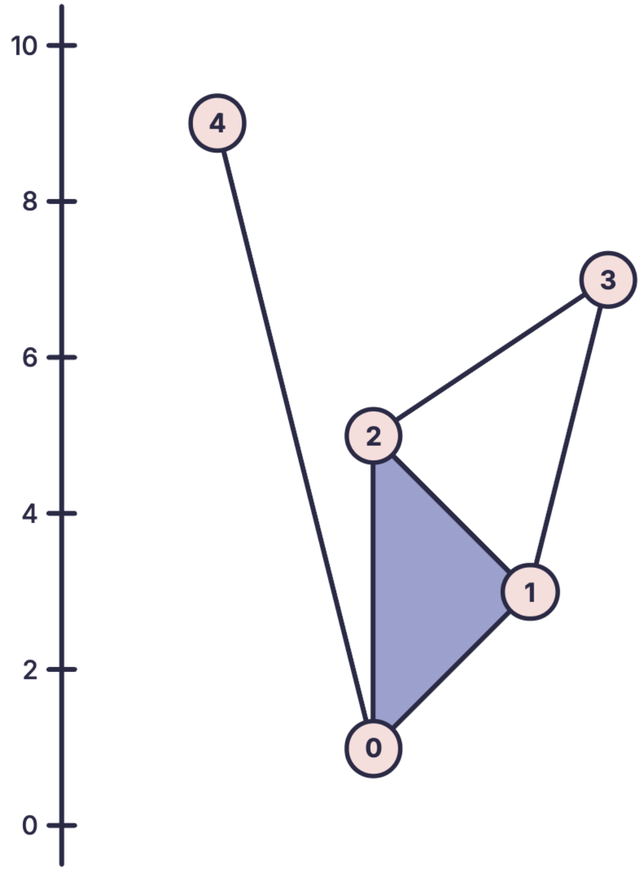


Induced Map $\bar{f} : \boldsymbol{R}_f \to \mathbb{R}$



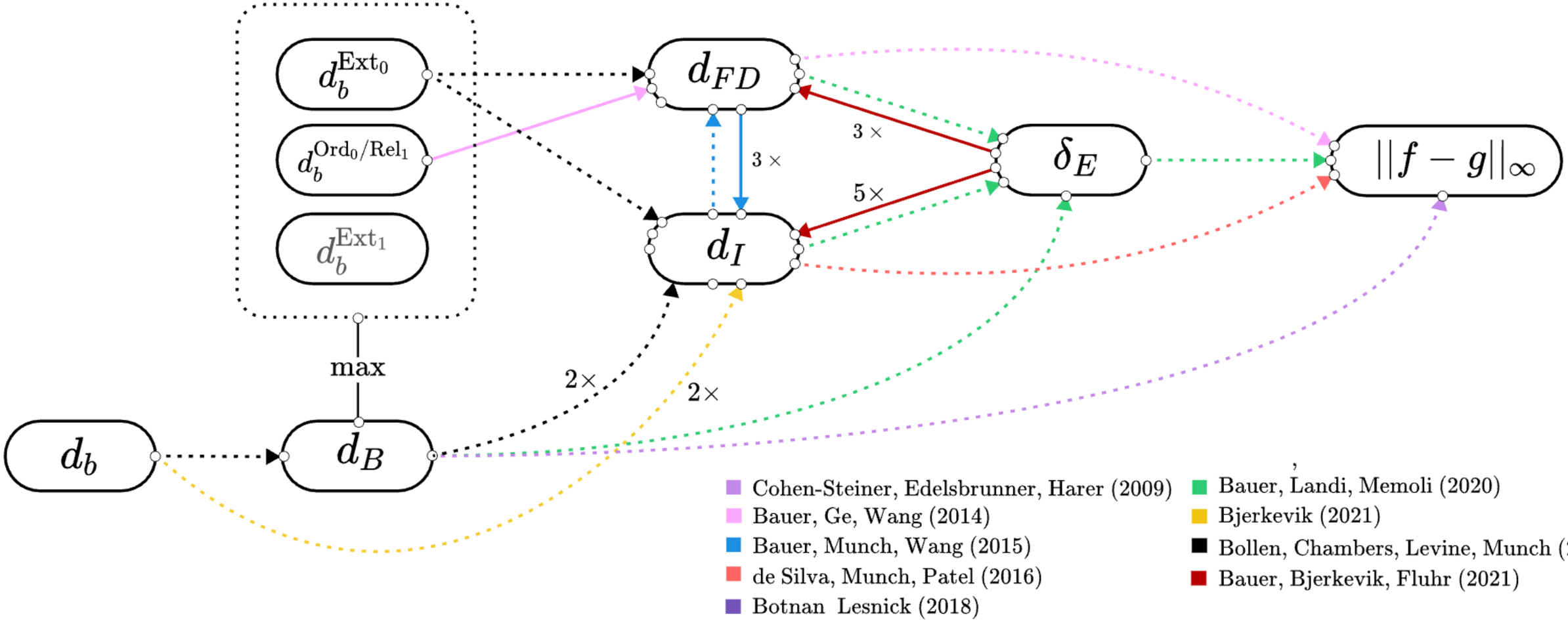
Another Example

Compute the Reeb graph



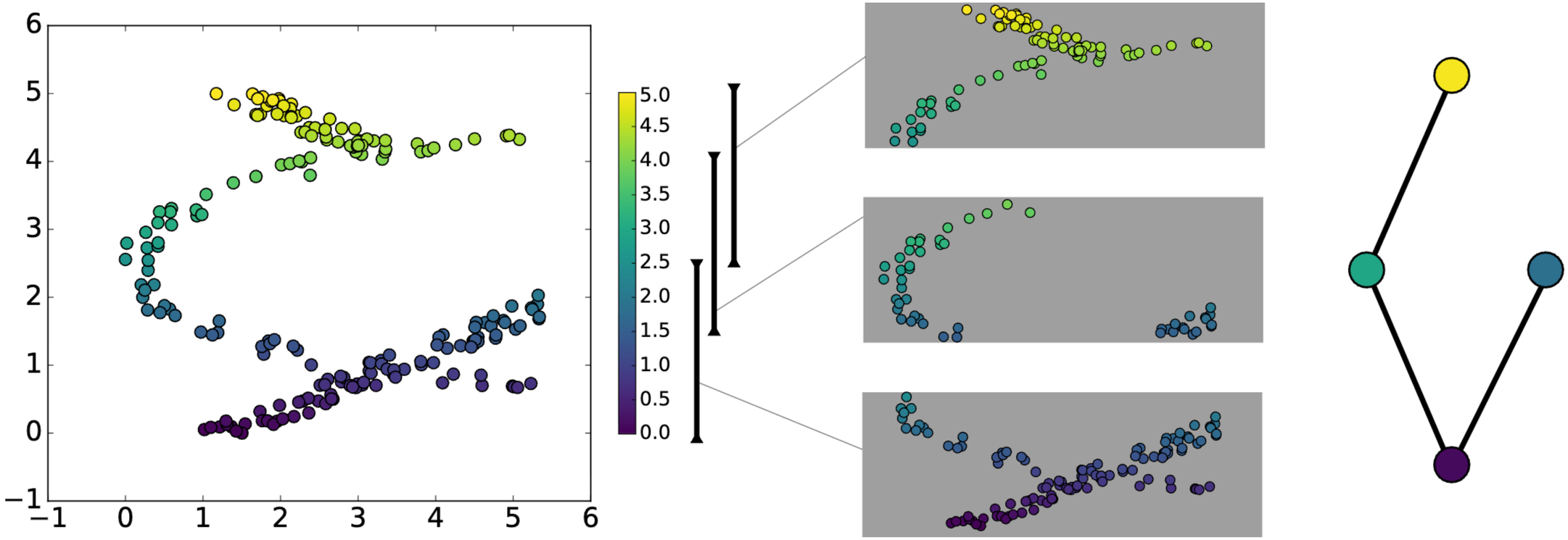
Theoretically Very Nice!

But we won't get into it...



Mapper

Idea: Approximate Reeb graph when we have only a point cloud sample of the space



Recall: Cover

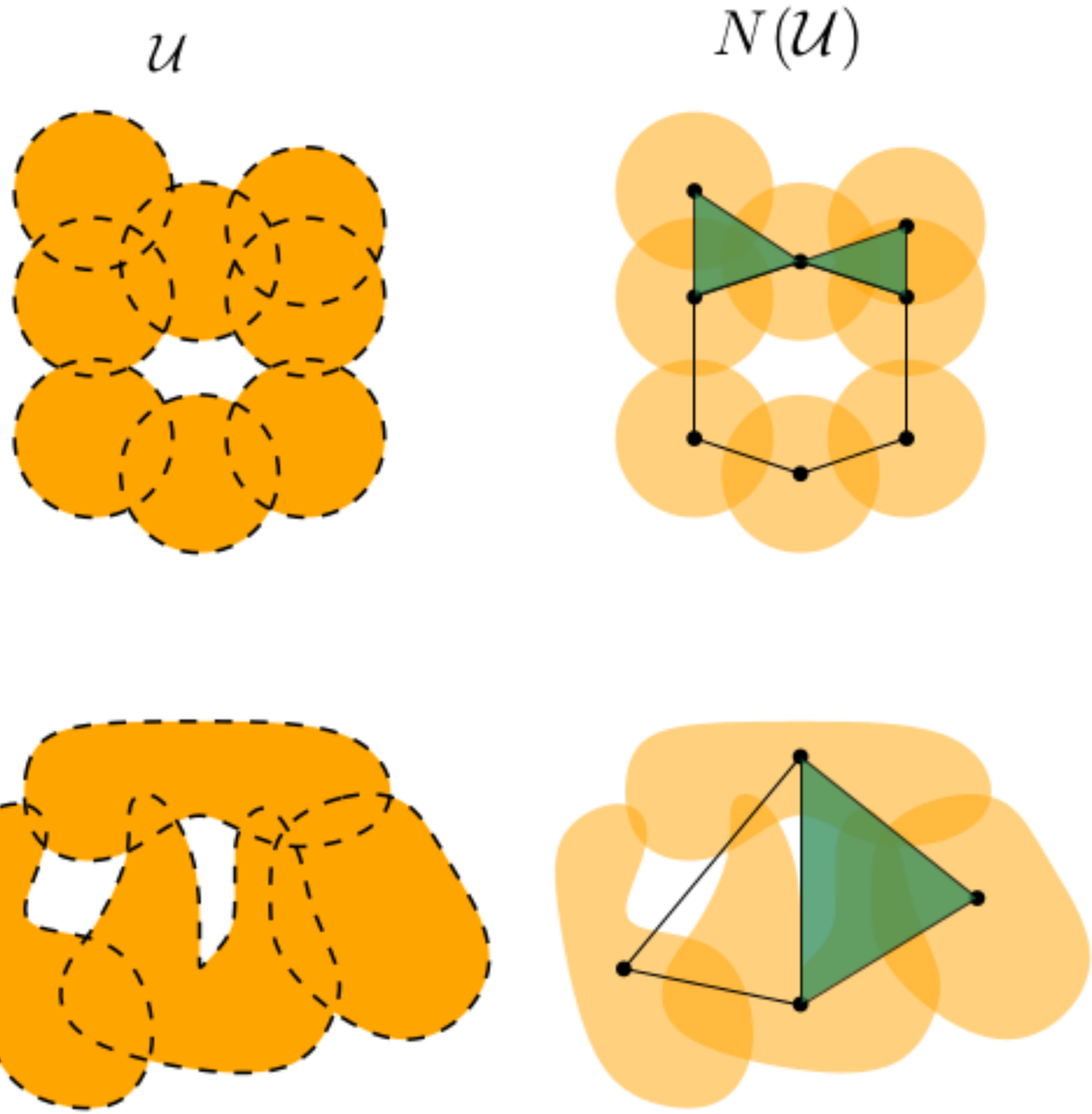
An *open (closed) cover* of a topological space $\mathbb{X}$ is a collection $\mathcal{U}$ of open (closed) sets so that $\mathbb{X} \subseteq \bigcup_i U_i$.

Ex. $\mathbb{R}, C = \{(n-1, n+1) \mid n \in \mathbb{Z}\}$

Recall: Nerve

Given a finite collection of sets $\mathcal{U}$, the *nerve* is the simplicial complex

$$\text{Nrv}(\mathcal{U}) = \{X \subseteq \mathcal{U} \mid \bigcap_{U \in X} U \neq \emptyset\}$$



Mapper Graph Paper

Eurographics Symposium on Point-Based Graphics (2007)
M. Botsch, R. Pajarola (Editors)

Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition

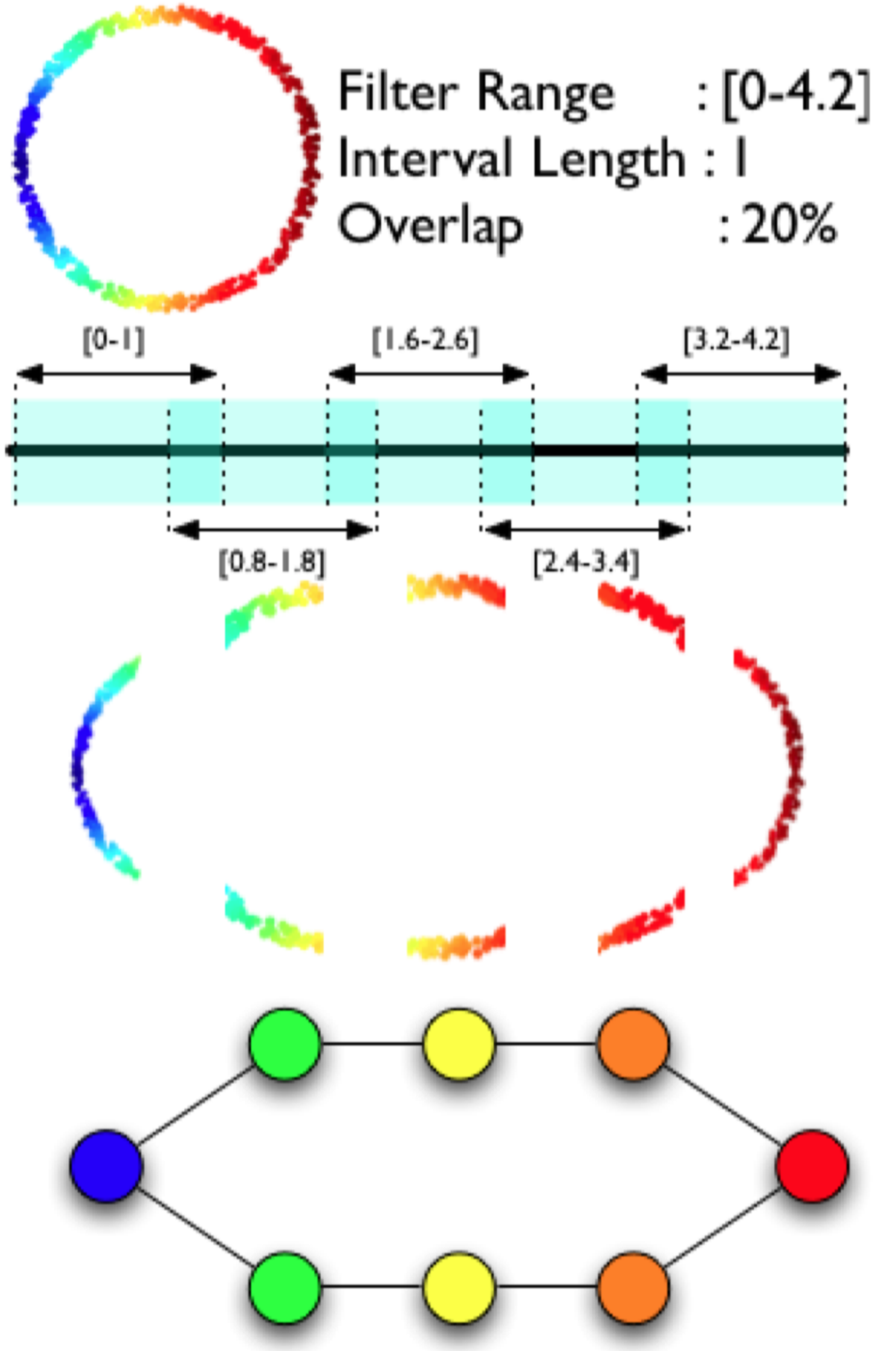
Gurjeet Singh[1] , Facundo Mémoli[2] and Gunnar Carlsson[†2]

[1]Institute for Computational and Mathematical Engineering, Stanford University, California, USA.
[2]Department of Mathematics, Stanford University, California, USA.

Abstract

We present a computational method for extracting simple descriptions of high dimensional data sets in the form of simplicial complexes. Our method, called `Mapper`*, is based on the idea of* partial clustering *of the data guided by a set of functions defined on the data. The proposed method is not dependent on any particular clustering algorithm, i.e. any clustering algorithm may be used with* `Mapper`*. We implement this method and present a few sample applications in which simple descriptions of the data present important information about its structure.*

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modelling.

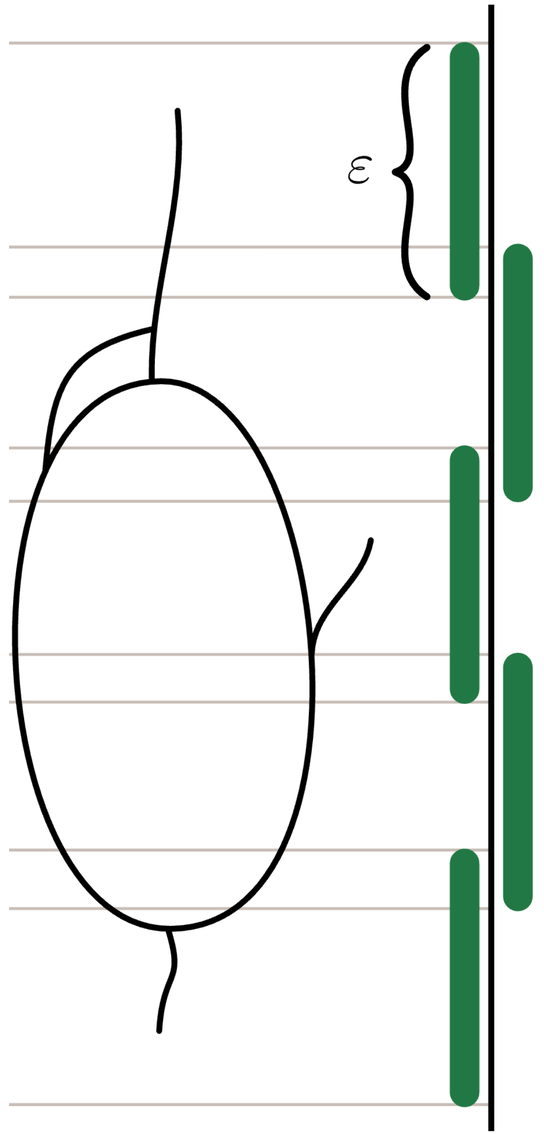


Mapper Definition (continuous input) - I

Given $f : |K| \to \mathbb{R}$.

Fix a cover $\mathcal{U} = \{U_\alpha\}$ of $\mathbb{R}$

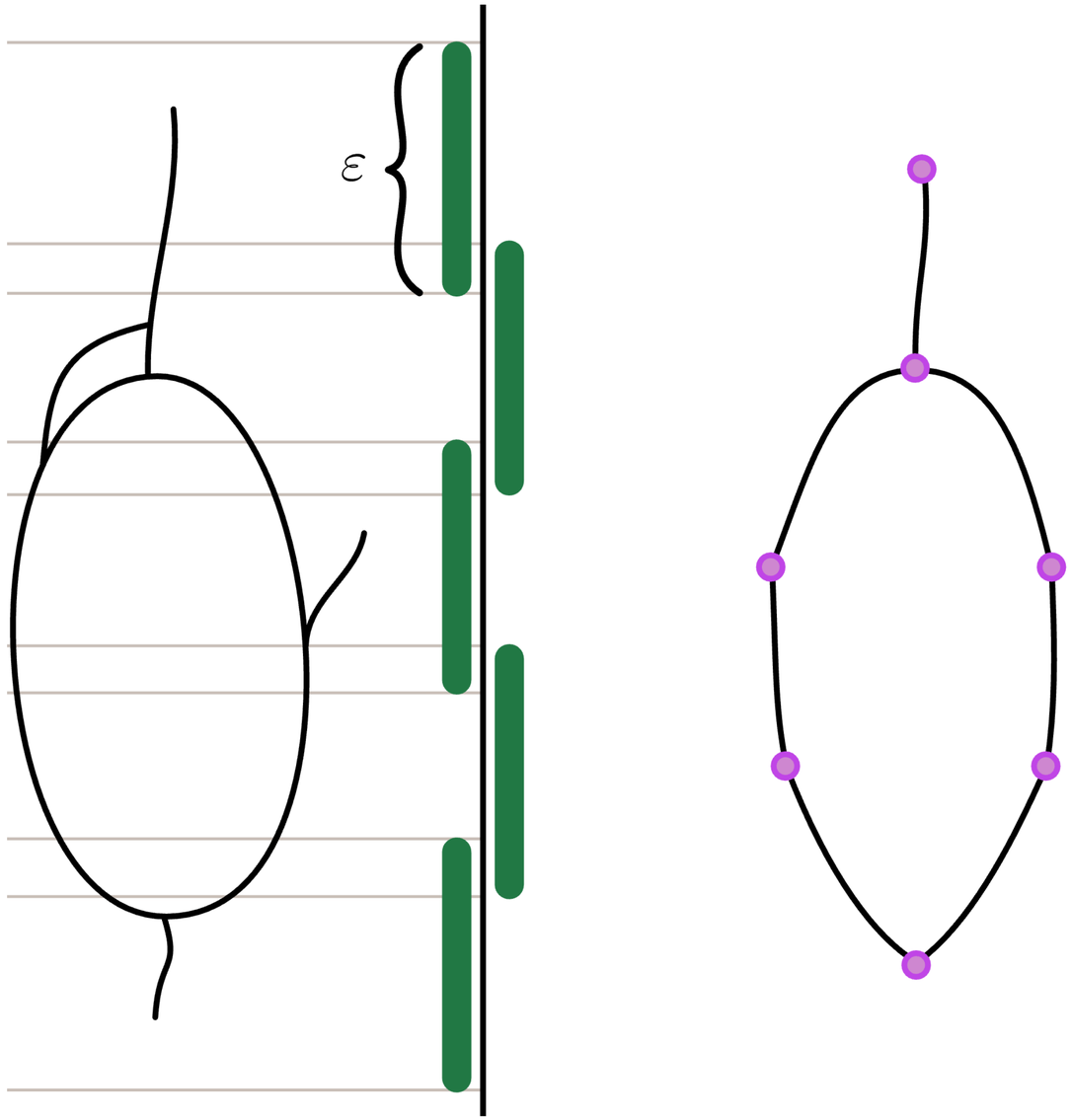
The collection $f^{-1}(\mathcal{U}) = \{f^{-1}(U_\alpha)\}$ is a cover of $\mathbb{X}$



Mapper Definition (continuous input) - II

Let $f^{-1}(\mathcal{U})$ be the cover which splits the sets into connected component.

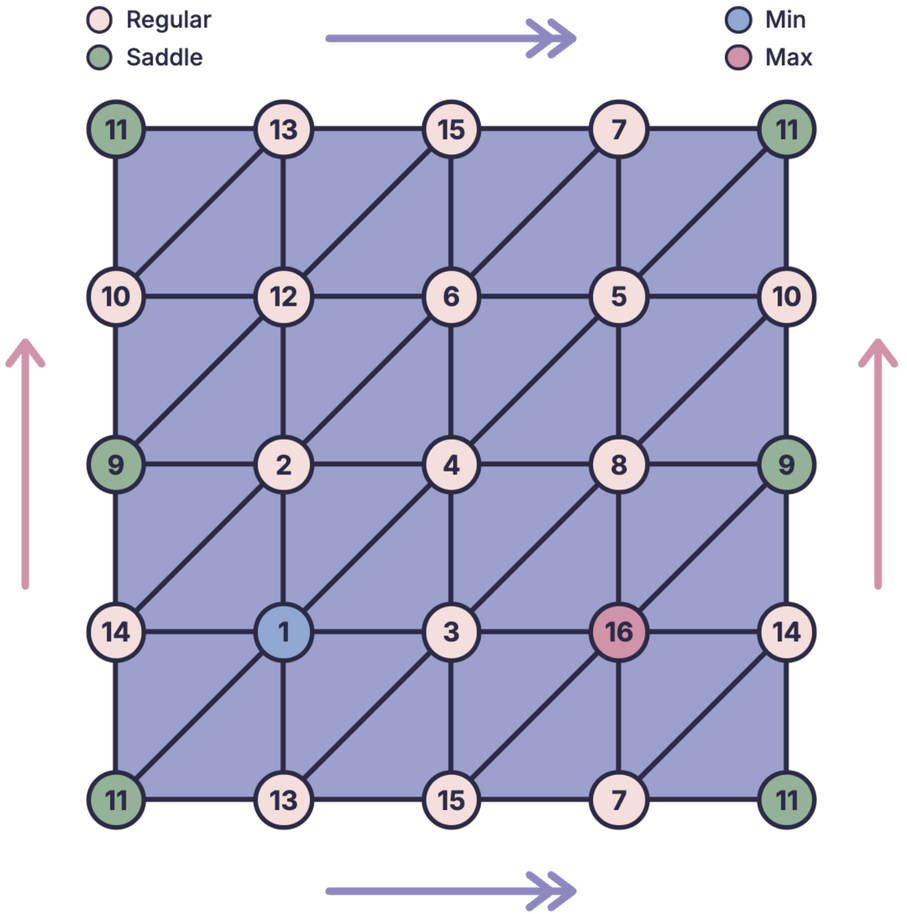
Then Mapper is the nerve of this cover.

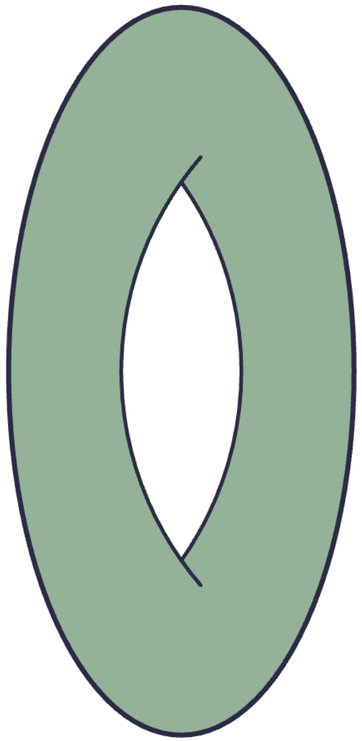


Try it

What is the Mapper graph of the cover

$$\{U_1 = (0,5), U_2 = (4,10), U_3 = (9.5,10.5), U_4 = (10,20)\}$$

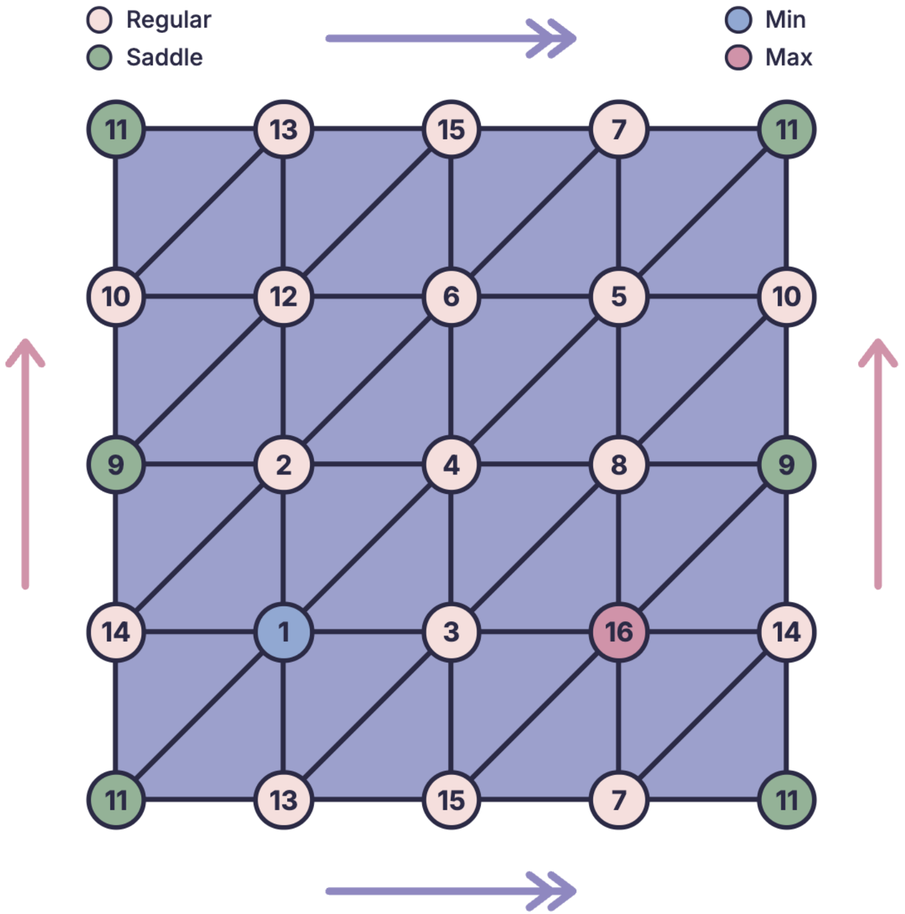


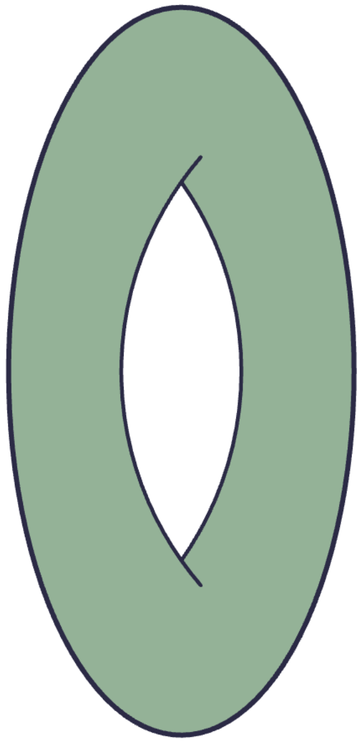


Try it

What is the Mapper graph of the cover

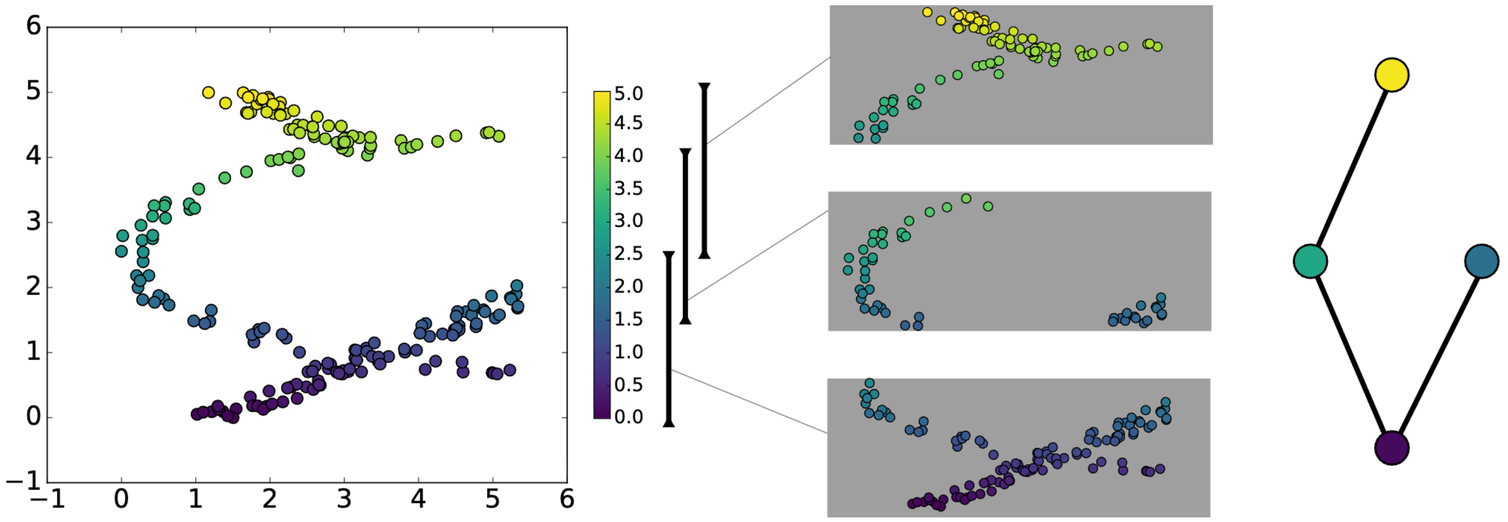
$$\{U_1 = (0,5), U_2 = (4,10), U_3 = (9.5,10.5), U_4 = (10,15), U_5 = (14,20)\}$$





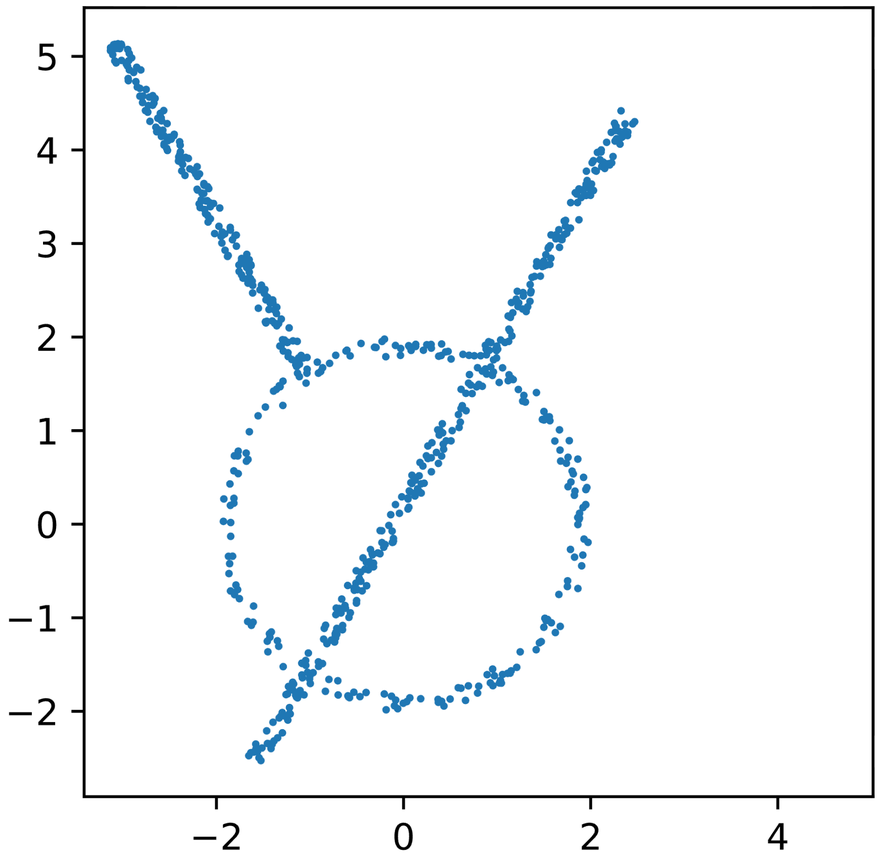
Point Cloud Version

- Choose a $\mathbb{R}$-valued function on the data. (“Lens function”)
- Choose a cover of the range
- Cluster the points with values inside each cover element
- Construct the nerve



Assumptions

Given poin cloud P of N points with distance $d(x, y)$.



Given poin cloud P of N points with distance $d(x, y)$.

Lens Function: Density

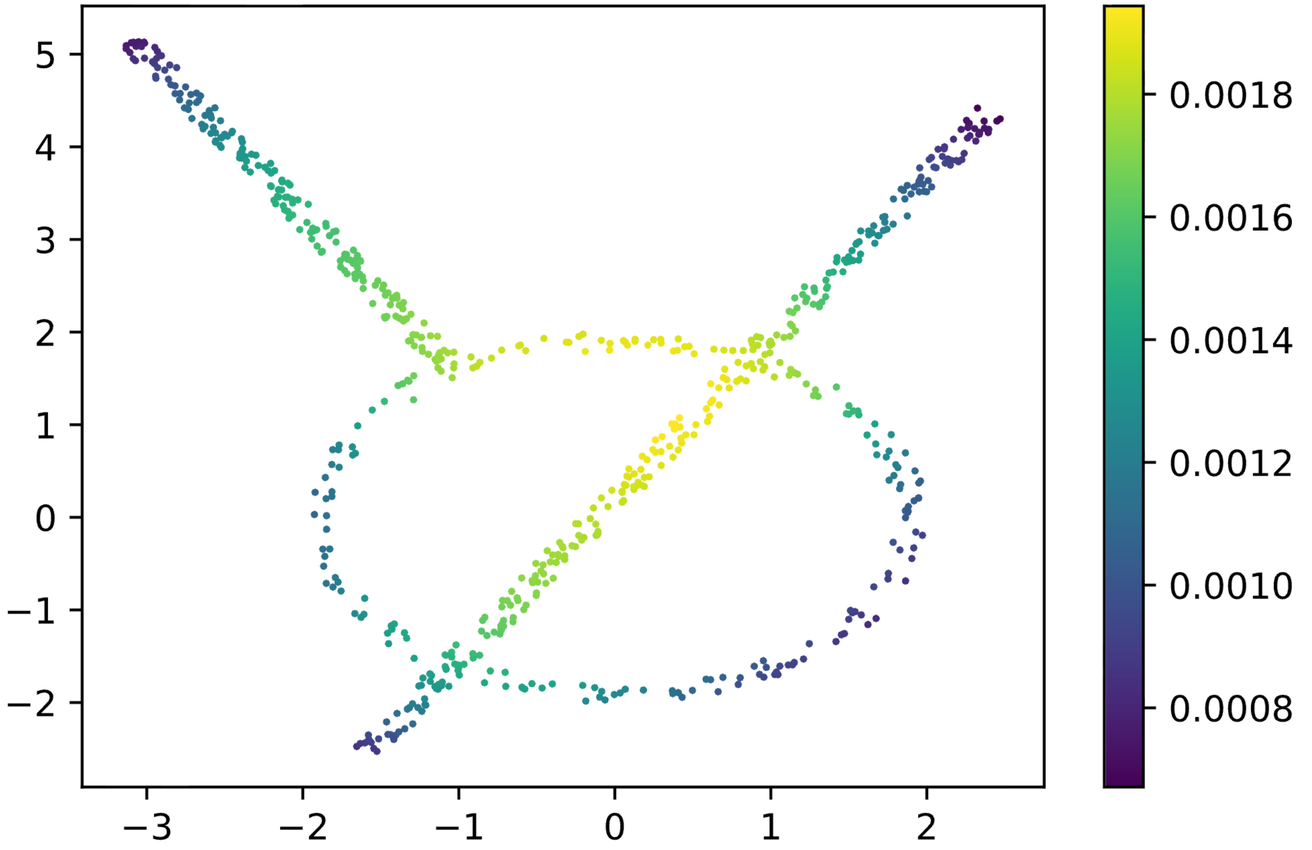
Idea: Choose function based on estimation of density.

Gaussian kernel version: For $x \in P$,

$$f_\epsilon(x) = C_\epsilon \sum_{y \in P} \exp\left(\frac{-d(x,y)^2}{\epsilon}\right)$$

Parameters:

- $x, y \in P$
- $\epsilon > 0$: determines smoothness
- C_ϵ is constant so that $\int f_\epsilon(x)\, dx = 1$



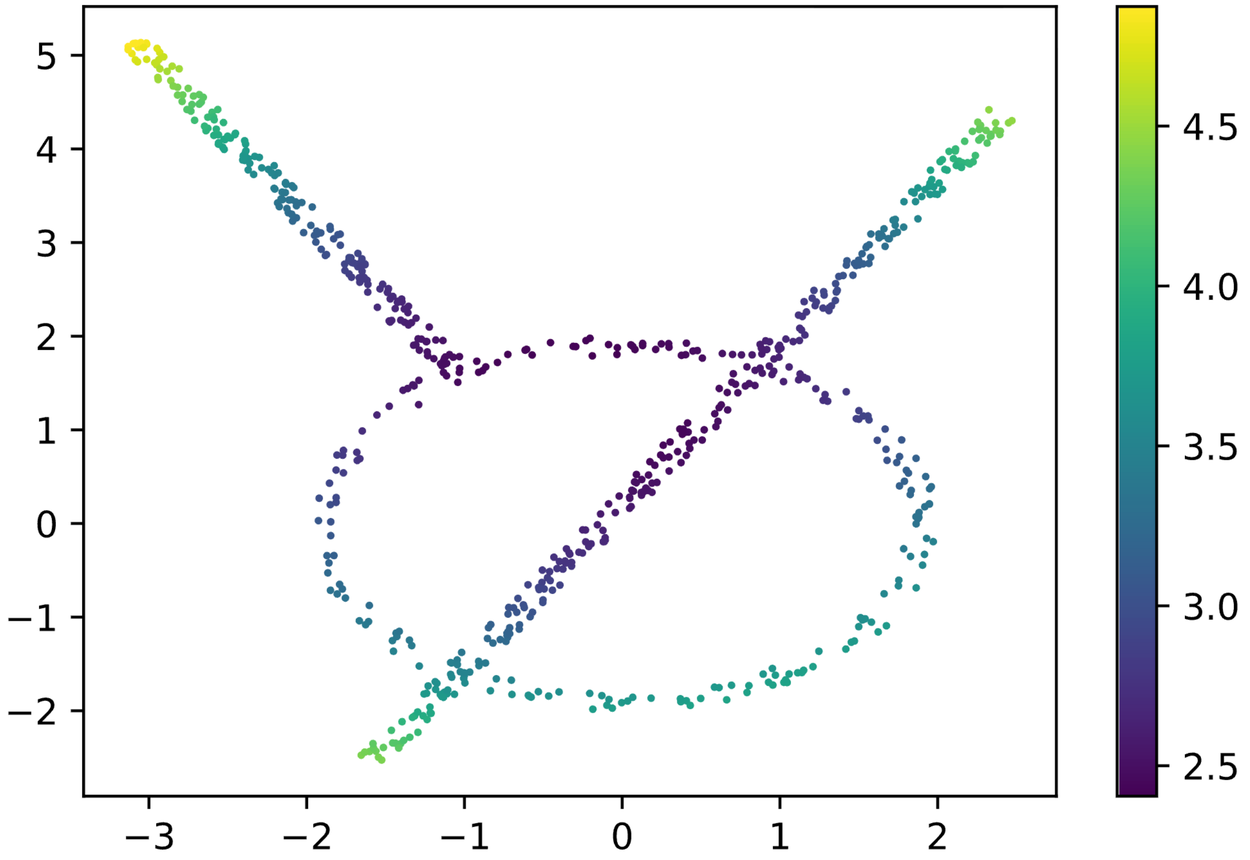
Lens Function: Eccentricity

Idea: low values correspond to points near the “center” of the data set, high correspond to points far from the center.

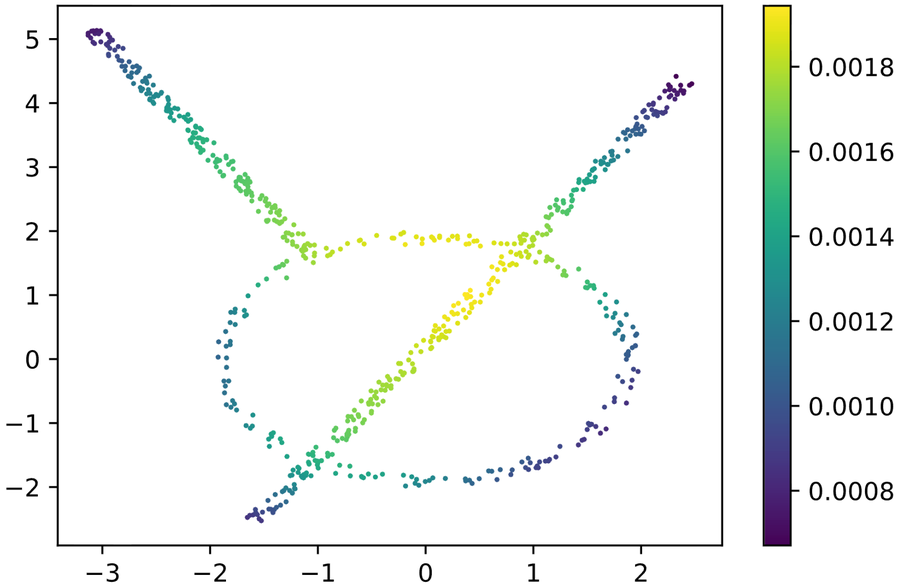
Given $1 \leq p \leq \infty$:

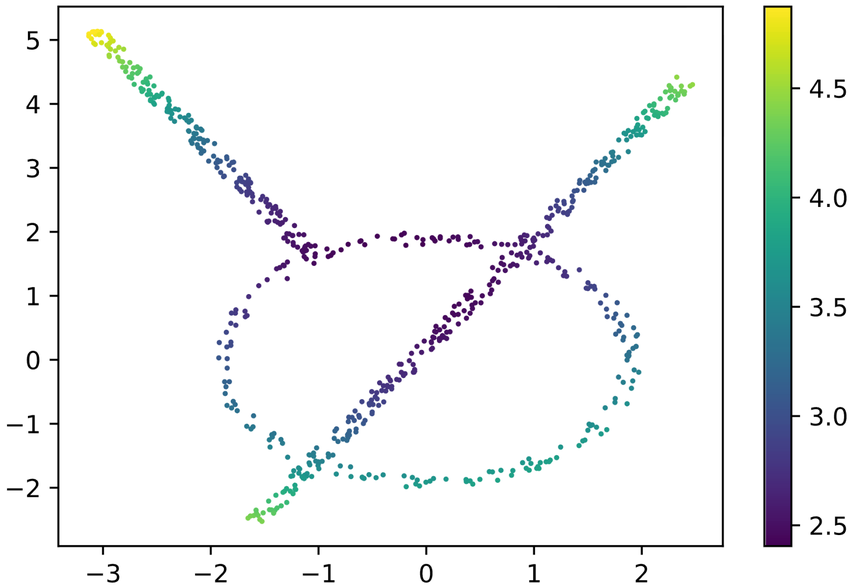
$$E_p(x) = \left(\frac{\sum_{y \in P} d(x,y)^p}{N} \right)^{\frac{1}{p}}$$

Note: no actual center has to be determined.



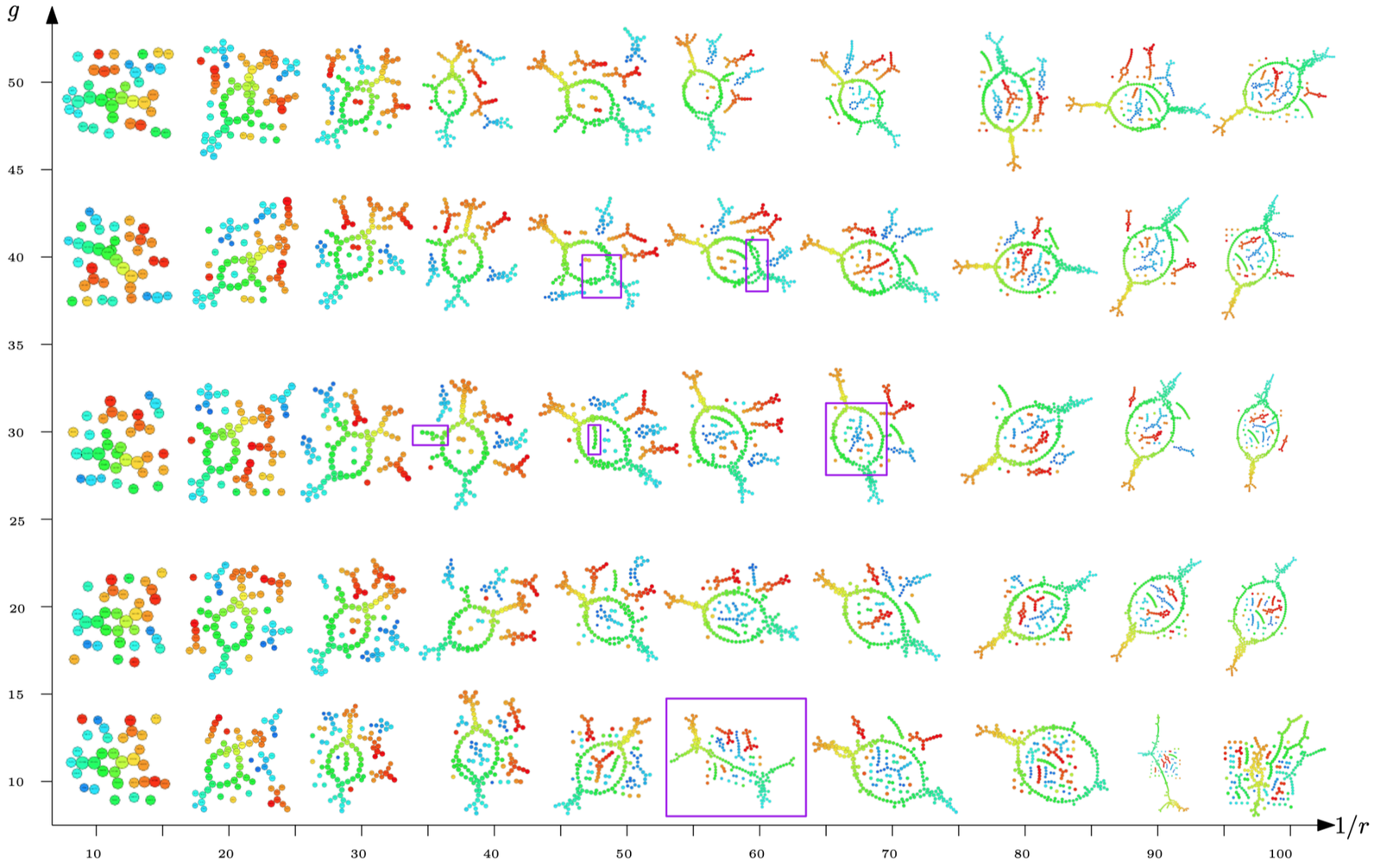
Differences





Cover choice is important!

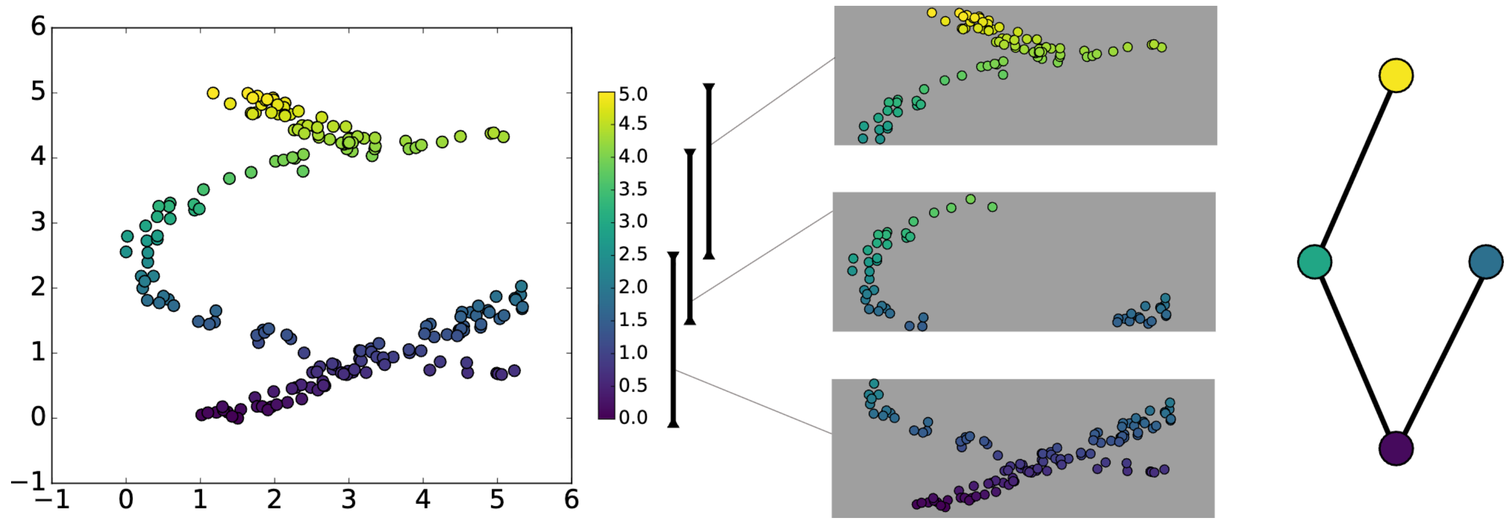




Interval length: r, overlap g

What do we want the clustering to do

- Work on general metric spaces (not just point clouds in $\mathbb{R}^d$
- Don’t need to specify the number of clusters in advance

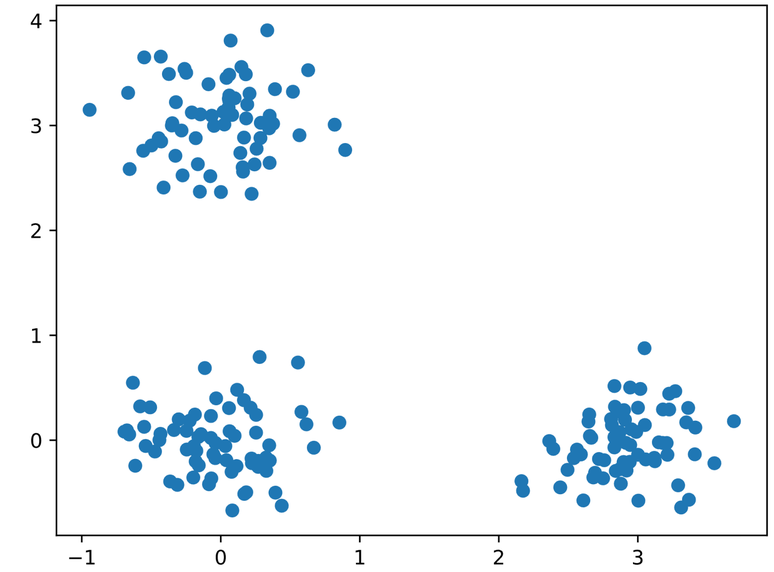


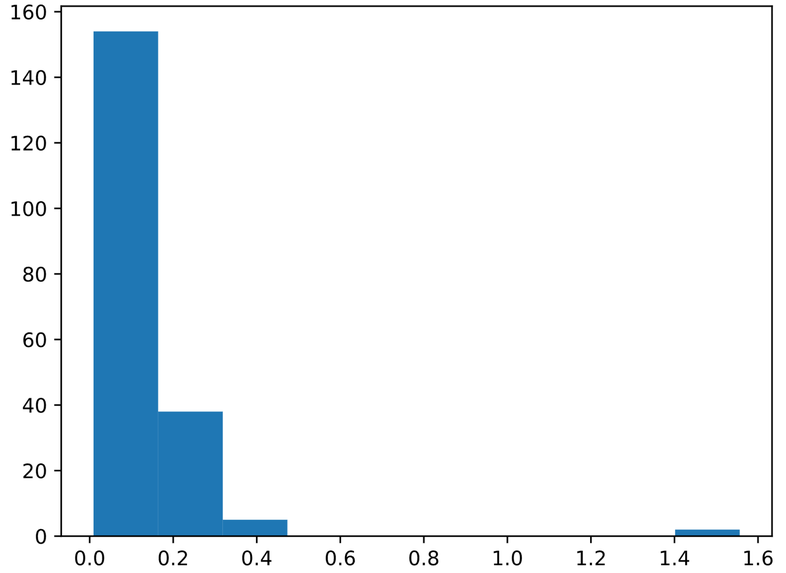
Clustering: Histogram Heuristic

Use the 0-dimentional persistence diagram

- All points are (0, death)
- Draw the output as a histogram
- Determine cutoff where there is an empty bin in the histogram

Sometimes called the *single-linkage clustering*.



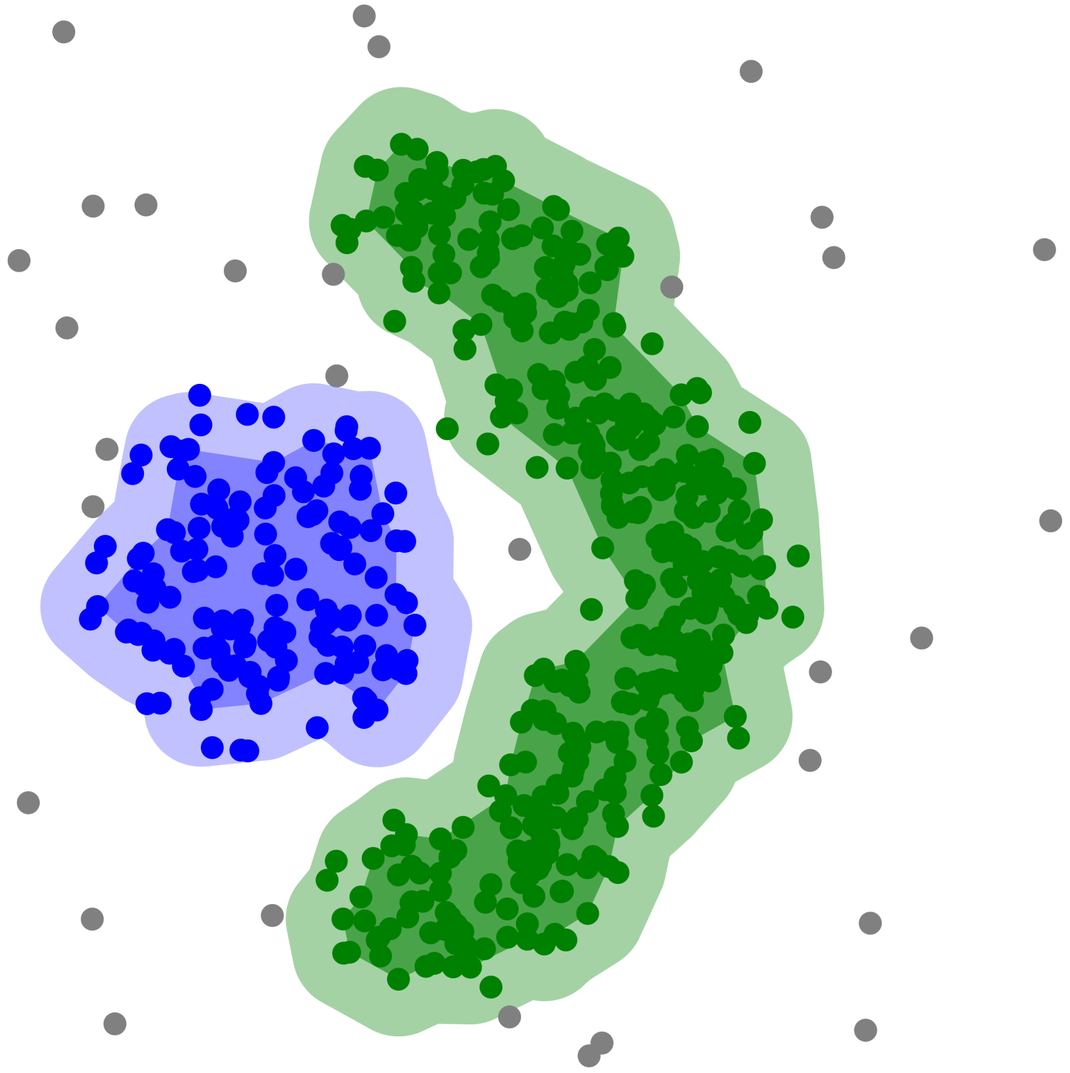


Clustering: DBScan

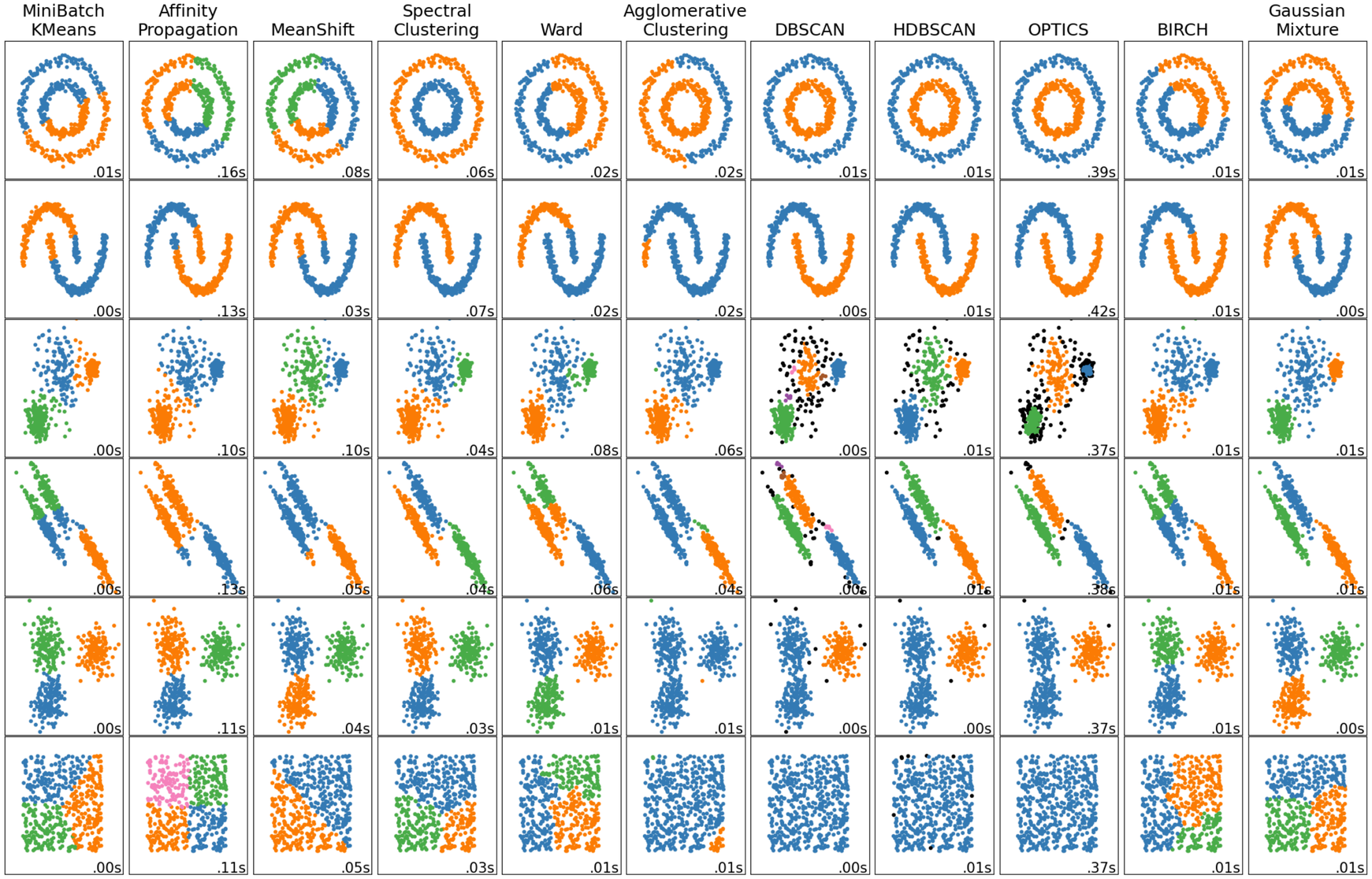
Based on density of points.

Parameters:

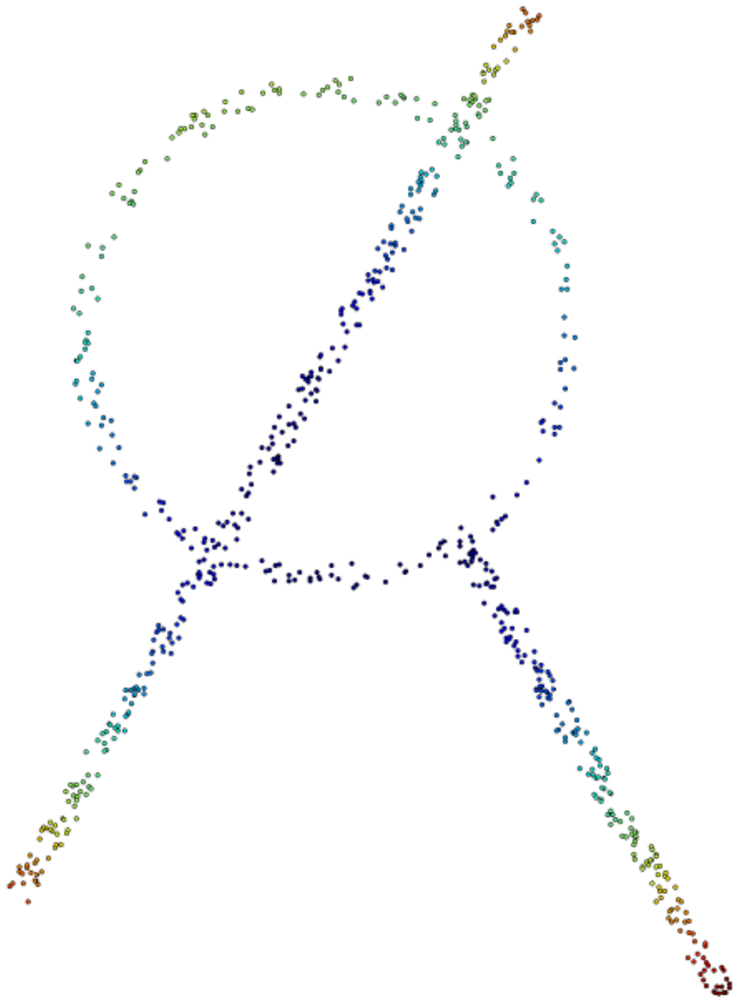
- ϵ: the maximum distance between two points for one to be considered in the neighborhood of the other
- minPts: minimum number of points within an ϵ radius needed to form a dense region

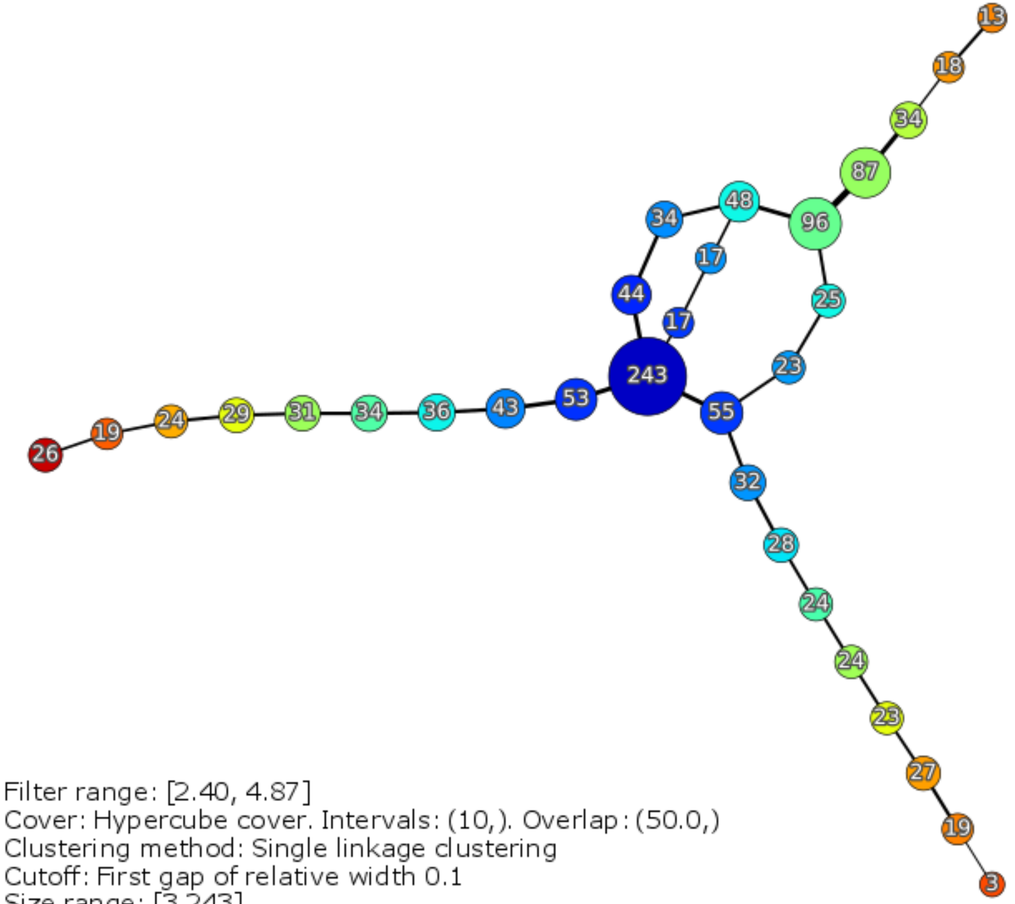


More Clusterings



Put it all together





Filter range: [2.40, 4.87]
Cover: Hypercube cover. Intervals: (10,). Overlap: (50.0,)
Clustering method: Single linkage clustering
Cutoff: First gap of relative width 0.1
Size range: [3,243]

Application: Breast Cancer Data

PNAS

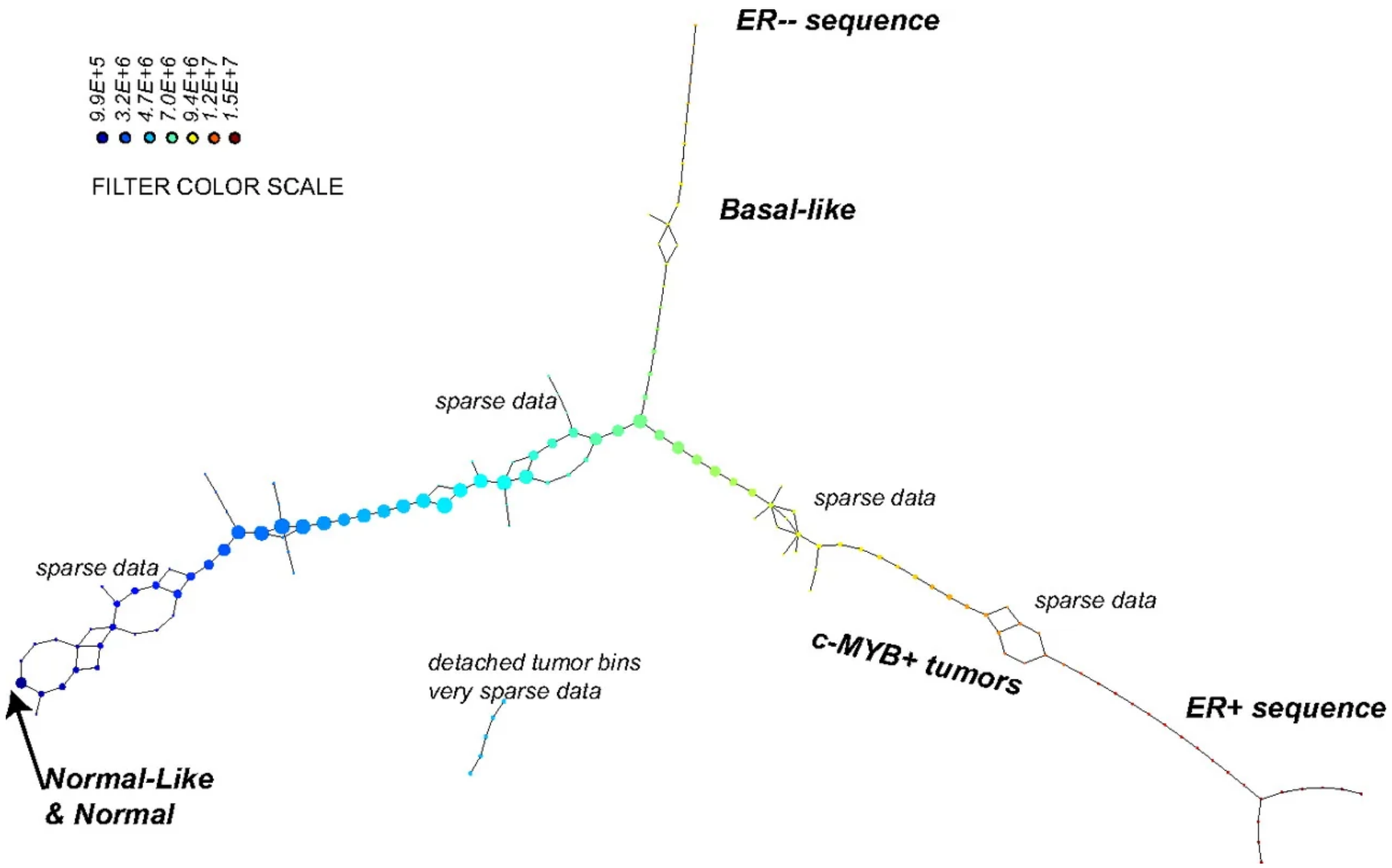
Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival

Monica Nicolau[a], Arnold J. Levine[b,1], and Gunnar Carlsson[a,c]

[a]Department of Mathematics, Stanford University, Stanford, CA 94305; [b]School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540; and [c]Ayasdi, Inc., Palo Alto, CA 94301

Contributed by Arnold J. Levine, February 25, 2011 (sent for review July 23, 2010)

The Mapper Graph



Clustering of data

