TDA - Fall 2025

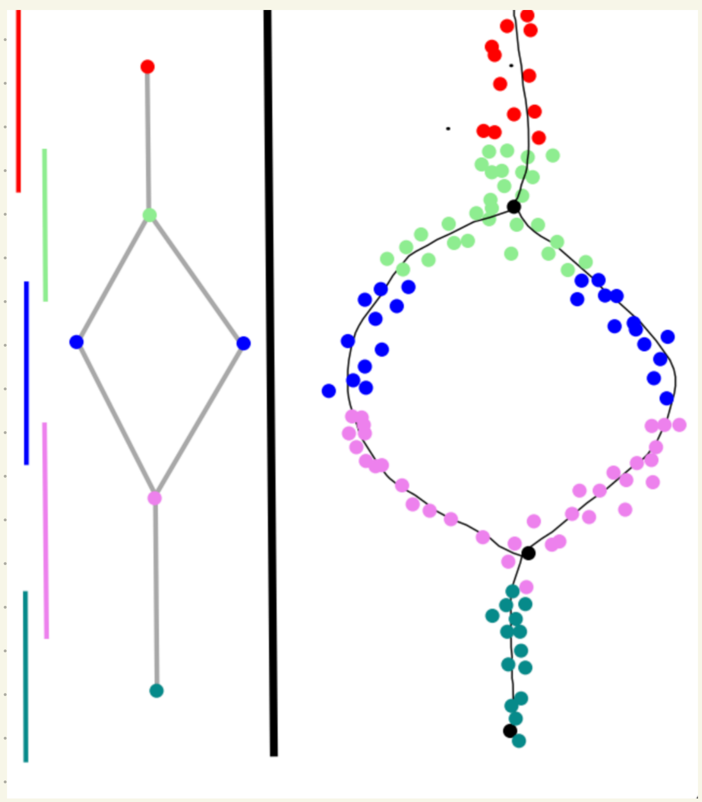
Mapper graphs

Mapper graphs

Idea! Approximate Reeb graphs

Won't always have a PL-space!

What about point clouds?



Idea!

- Give $\mathbb{R}$-values to dots
- Use a cover of $\mathbb{R}$
- Cluster into components + build a graph

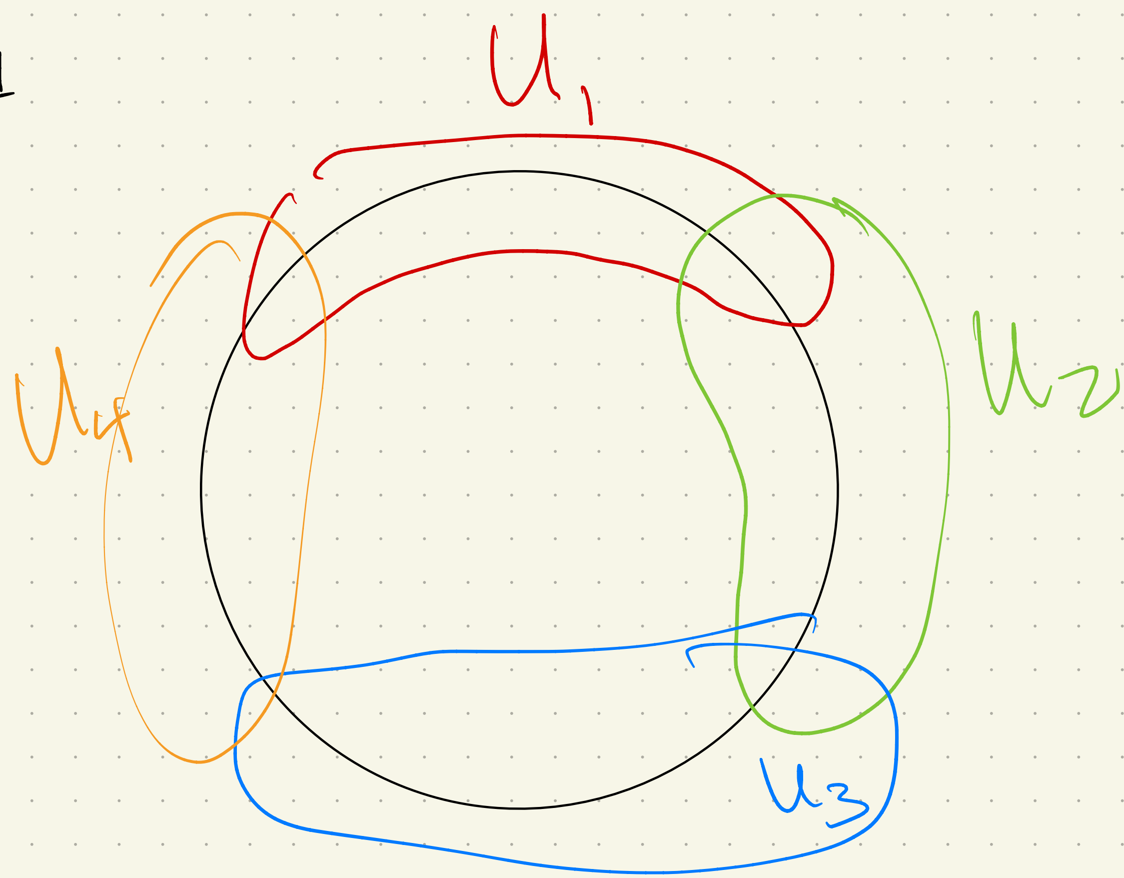
More details:

A cover of a set X is a collection of sets $\mathcal{U} = \{U_1, \dots, U_k\}$ s.t. $X \subseteq \bigcup_i U_i$

open cover $\to$ each U_i open

Ex: S^1

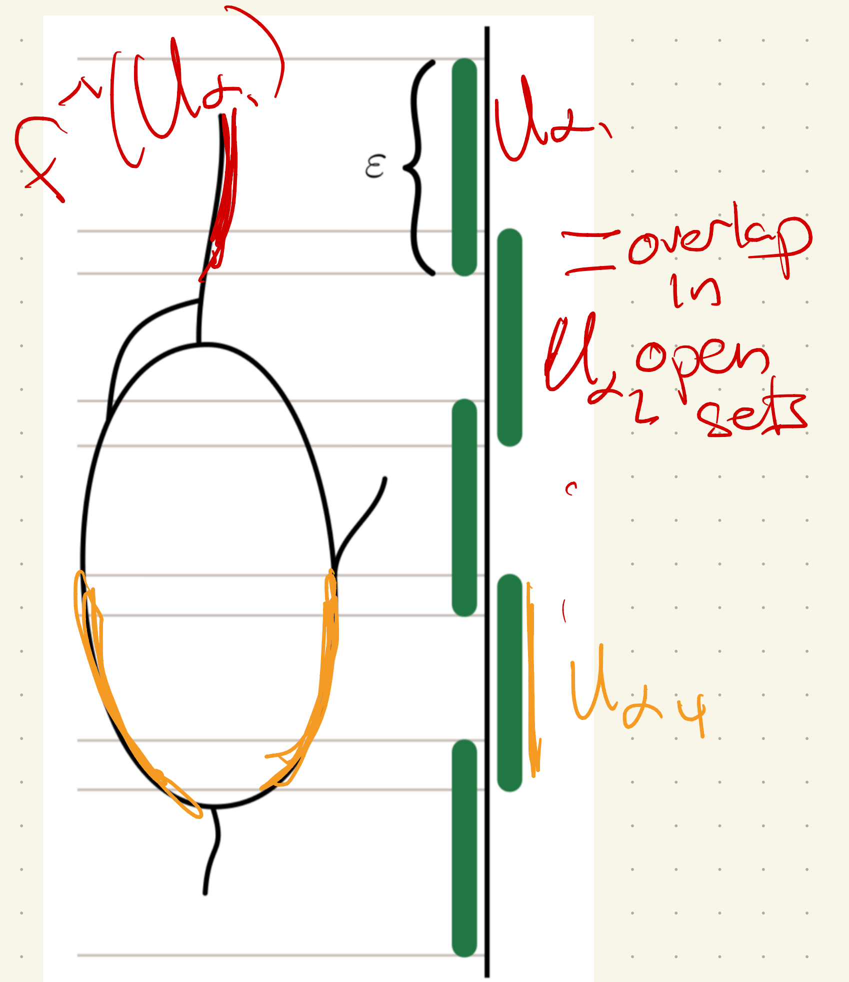
Cover:



Let's start on a simplicial complex:

$\mathbb{R}$ contained in $\bigcup U_\alpha$

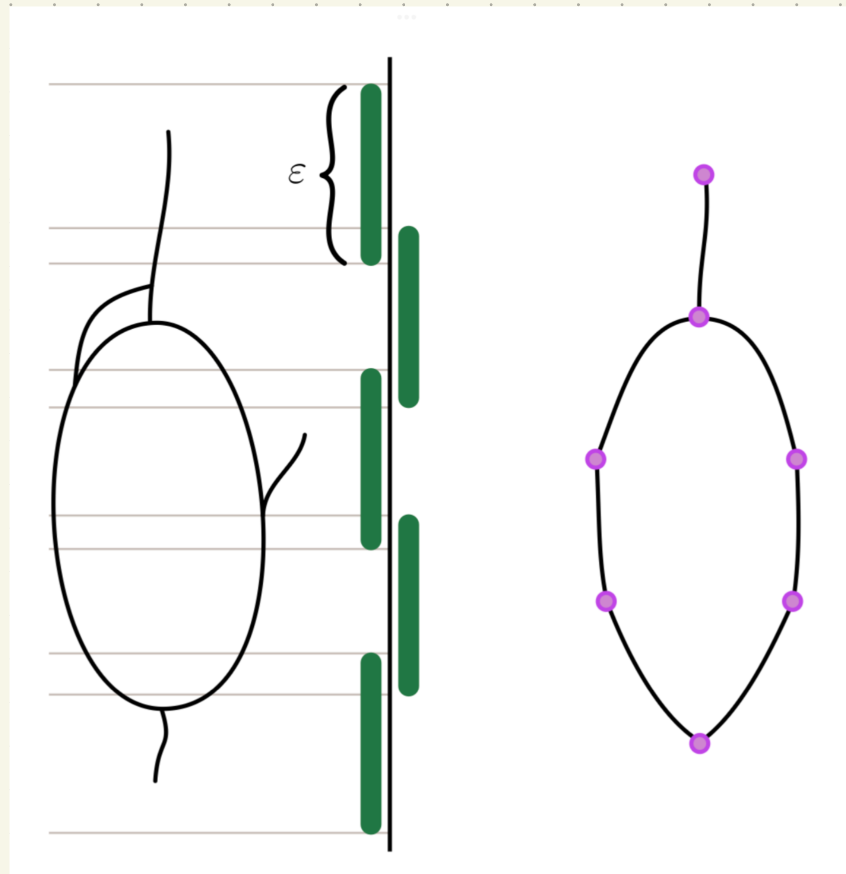
- Given $f: |K| \to \mathbb{R}$
- Fix a cover $\mathcal{U} = \{U_\alpha\}$ of $\mathbb{R}$
- The collection $f^{-1}(\mathcal{U}) = \{f^{-1}(U_\alpha)\}$ is a cover of X.



Mapper for simplicial complex (part 2)

- Let $f^{-1}(U)$ be the cover, then split sets in each preimage into connected components
- Then, Mapper is the nerve of this cover

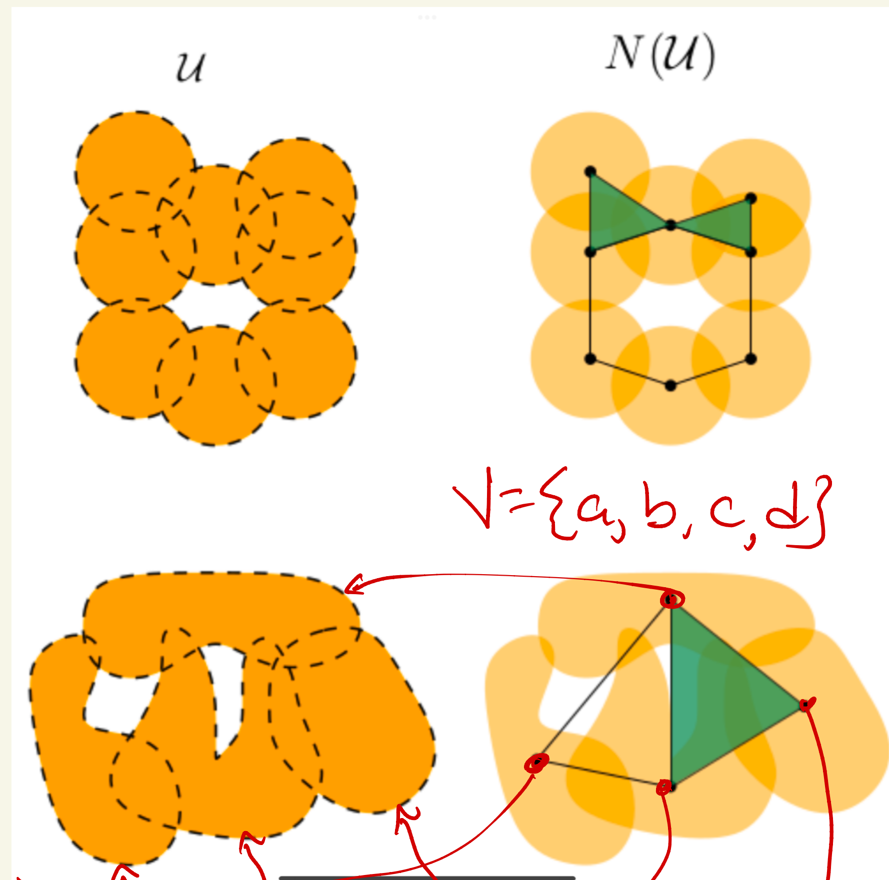
(Remember nerves?)



Given a finite collection of sets $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$, the nerve of $\mathcal{U}$, $N(\mathcal{U})$, is the simplicial complex with vertex set A, where $\{\alpha_0, \ldots, \alpha_k\} \subseteq A$ is a k-simplex $\in N(\mathcal{U})$

$$\iff$$

$$U_{\alpha_0} \cap \cdots \cap U_{\alpha(k)} \neq \emptyset$$

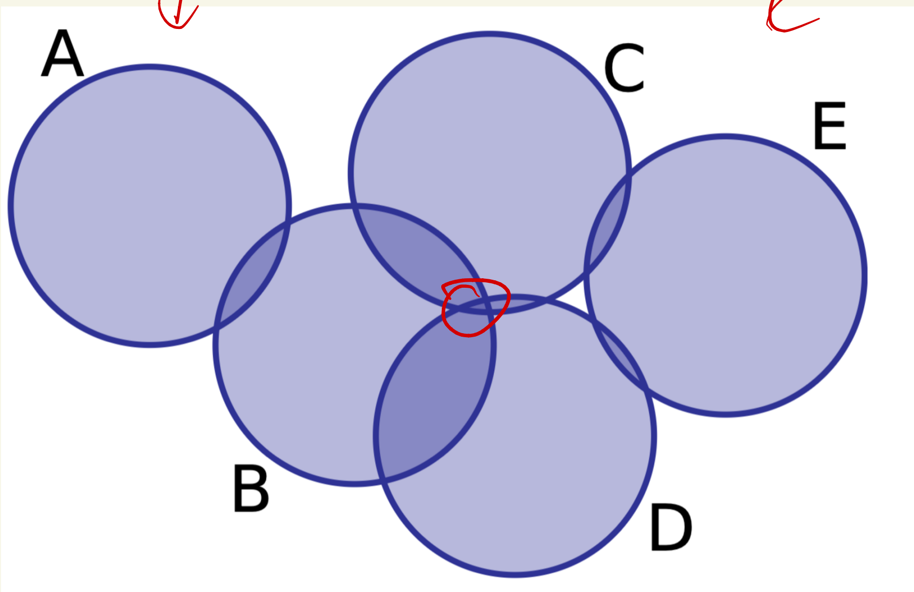


3 sets intersecting $\hookrightarrow \emptyset$

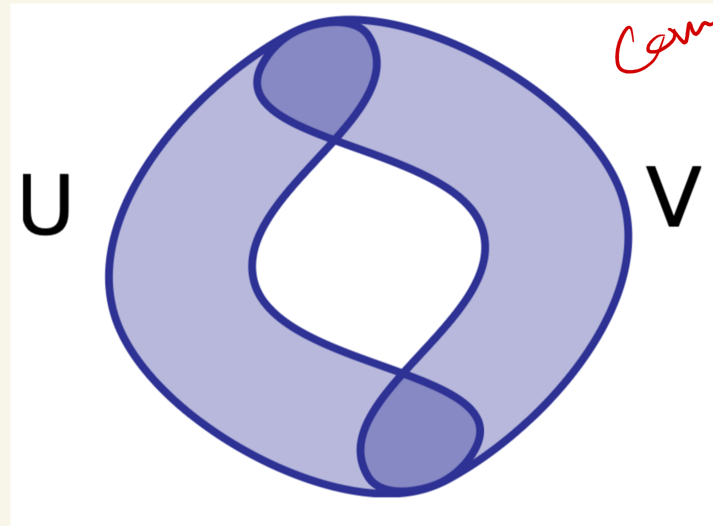
$\{ab\} \in N(\mathcal{U})$?

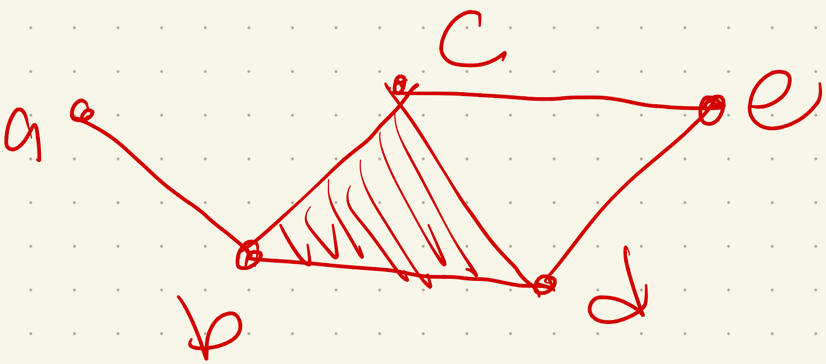
Some examples to try:

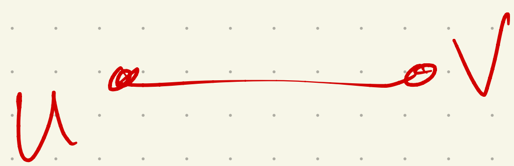
x_i's



$U \cap V$ has 2 components





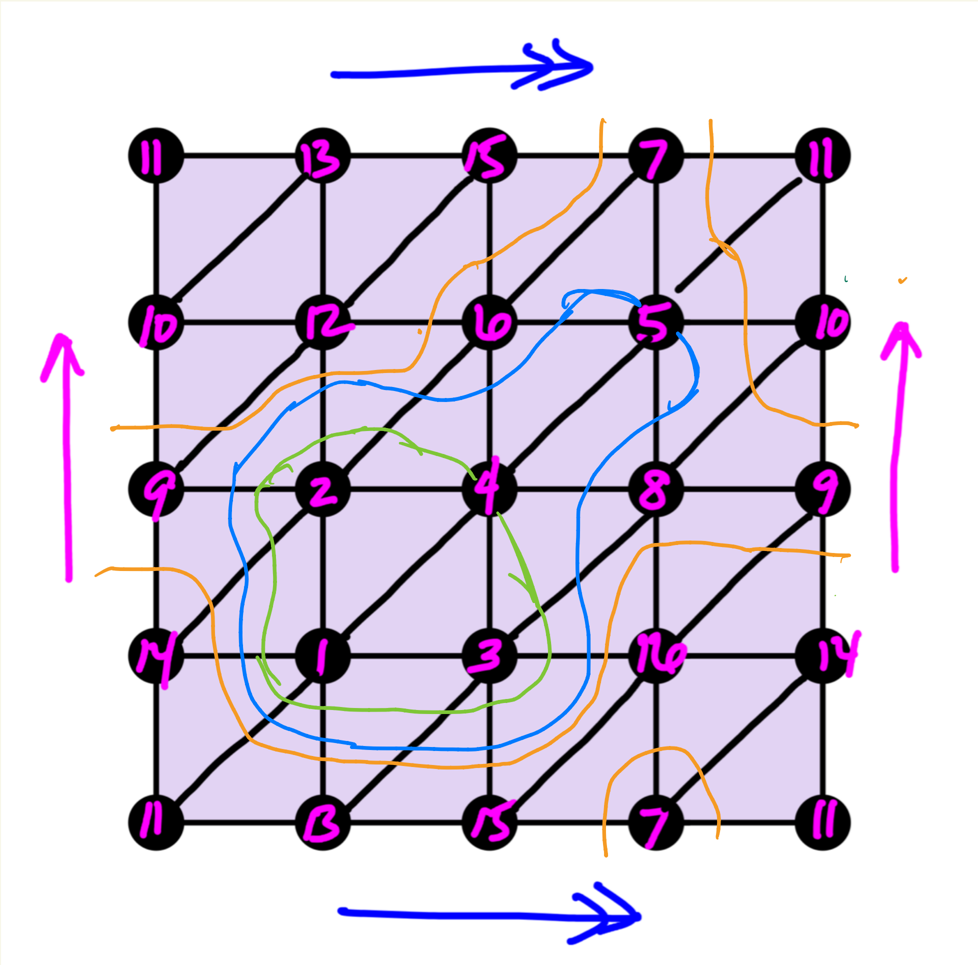


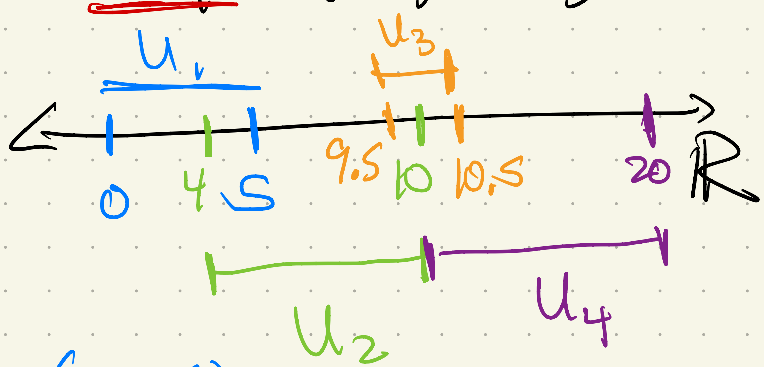
Difference: balls somehow reflect topology

Let's try: What is Mapper graph for cover

$U_1 = (0,5)$, $U_2 = (4,10)$, $U_3 = (9.5, 10.5)$,

$U_4 = (10, 20)$





$(0,5)$

$(0,4)$

$(0,9.5)$

$(0,10)$

$(0,10.5)$

Note: Cover is the key!
A "poor" choice of cover might yield a disconnected mapper graph on a torus!

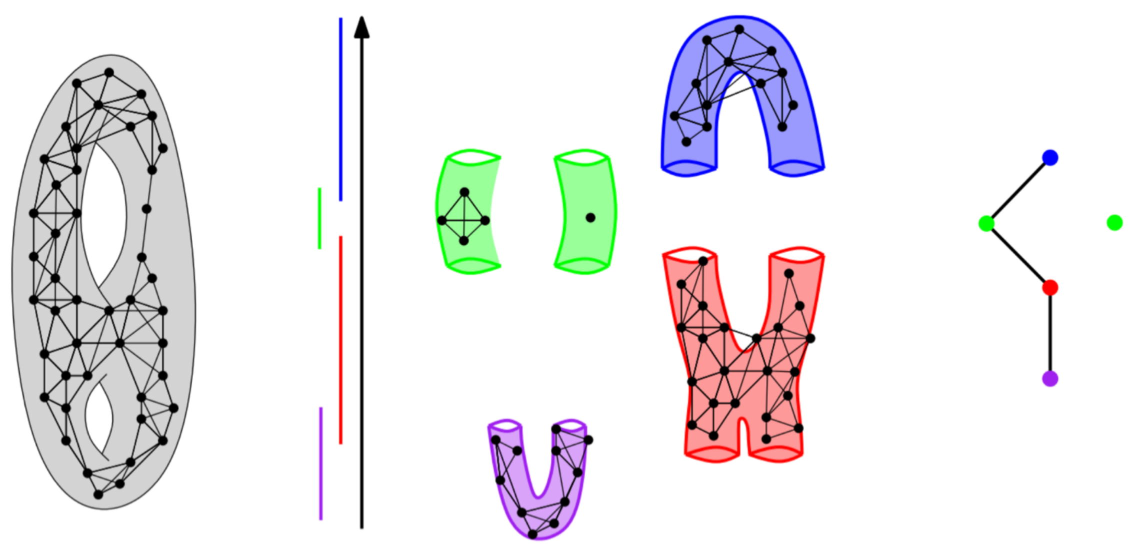


Figure 3: Example of Mapper computed on a sampling of the double torus with the height function f and a cover $\mathcal{I}$ of its range with four open intervals. Clusters are given by a neighborhood graph built on the sampling. Note that the rightmost green vertex is not connected to the other vertices of the Mapper since its corresponding cluster (which contains only one point) has no common points with the others.

In book:

Lots of technicalities for why maps from K to $|N(\mathcal{U})|$ travel through homology.

Some technicalities!

- Singular versus simplicial homology
- Nerve construction can collapse things

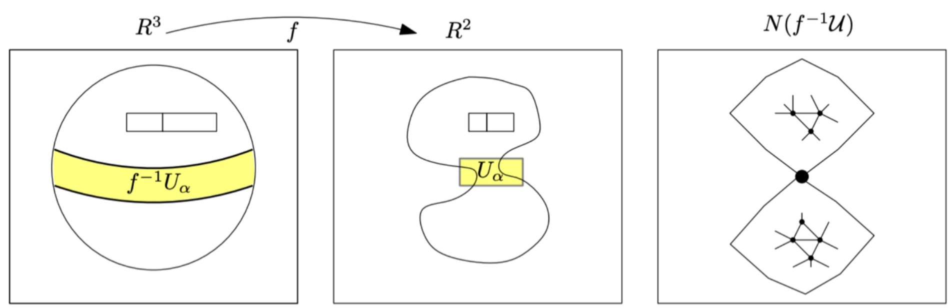


Figure 9.3: The map $f : \mathbb{S}^2 \subset \mathbb{R}^3 \to \mathbb{R}^2$ takes the sphere to $\mathbb{R}^2$. The pullback of the cover element U_α makes a band surrounding the equator which causes the nerve $N(f^{-1}\mathcal{U})$ to pinch in the middle creating two 2-cycles. This shows that the map $\phi_{\mathcal{U}} : X \to N(\mathcal{U})$ may not induce a surjection in H_2.

H_1 is special

- Every one-cycle γ in $|N(\mathcal{U})|$ has a 1 cycle γ' in $N(\mathcal{U})$ s.t. $\gamma = [|\gamma'|]$

$\Rightarrow$ If $\mathcal{U}$ is path connected,
$\phi_{\mathcal{U}*} : H_1(X) \to H_1(|N(\mathcal{U})|)$ is a surjection.

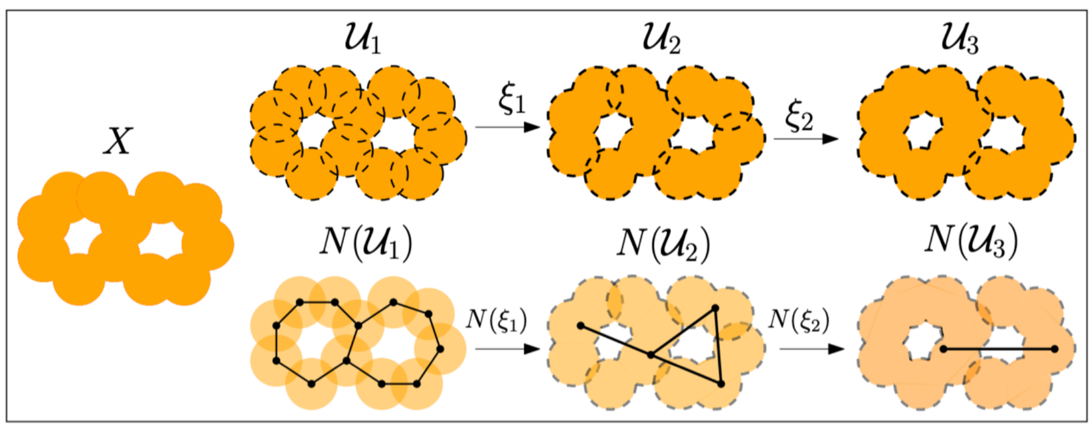
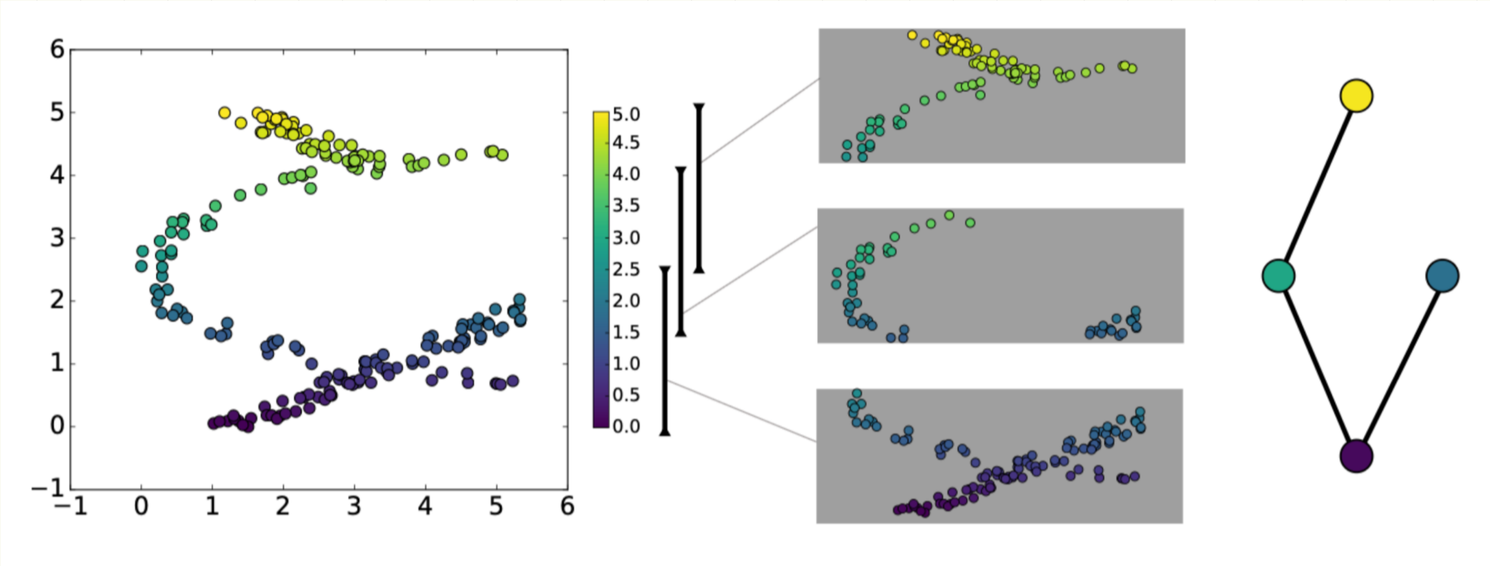


Figure 9.5: Sequence of cover maps induce a simplicial tower and hence a persistence module: classes in H_1 can only die.

Point cloud mapper

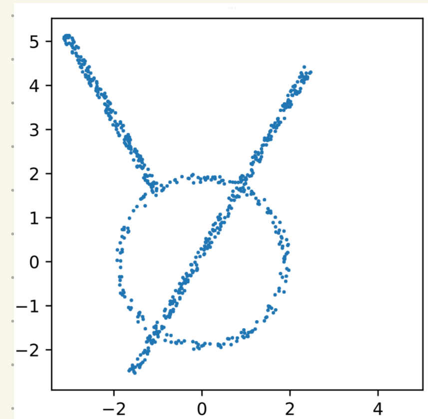
Idea:

- Choose $\mathbb{R}$-valued function on data
- Choose a cover of the range
- Cluster points inside each cover element
- Construct a nerve



Lots of choice!

What if we have points, but no function?



$f: X \to \mathbb{R}$ can be:

the "lens"

- density estimate
- centrality $f(x) = \sum_{y \in X} d(x,y)$
- eccentricity $f(x) = \max_{y \in X} d(x,y)$
- PCA coordinates
- distance to some "root point" (ie filamentary structures)
- functions detecting outlier behavior

⋮

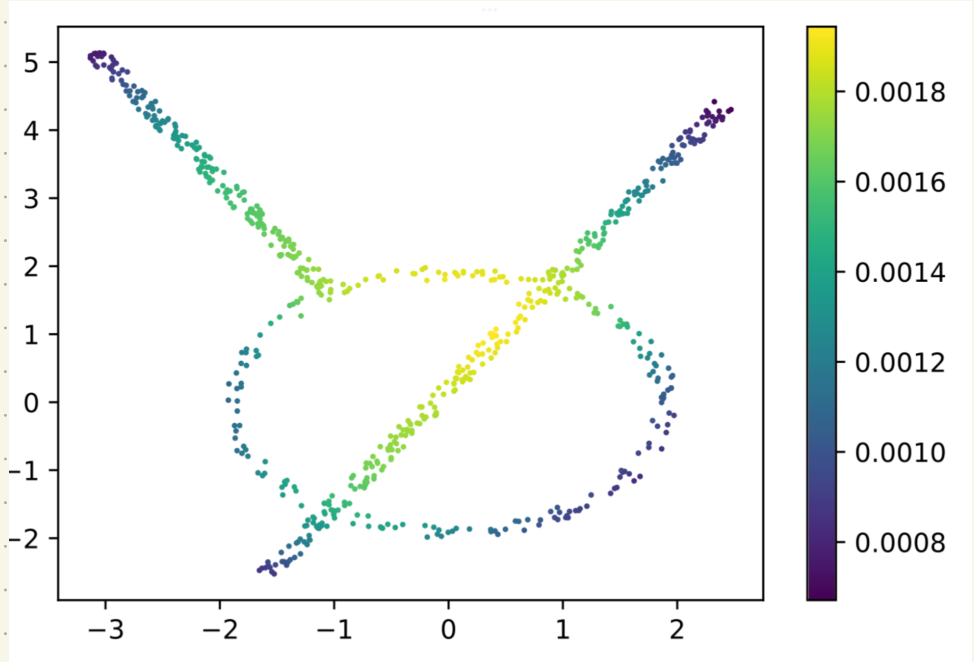
Density estimates

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

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

Parameters:

- $x, y \in P$
- $\varepsilon > 0$: determines smoothness
- C_ε constant s.t. $\int f_\varepsilon(x)\,dx = 1$



Density f_3

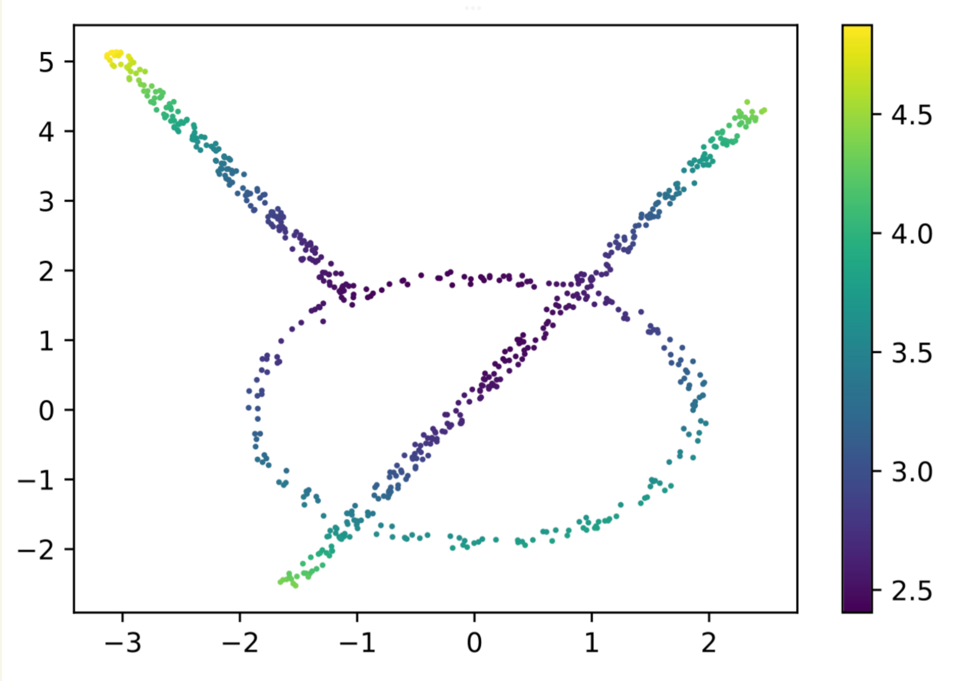
Eccentricity

Idea: low values are points near center of data, high values are far away

Given $1 \leq p < \infty$,

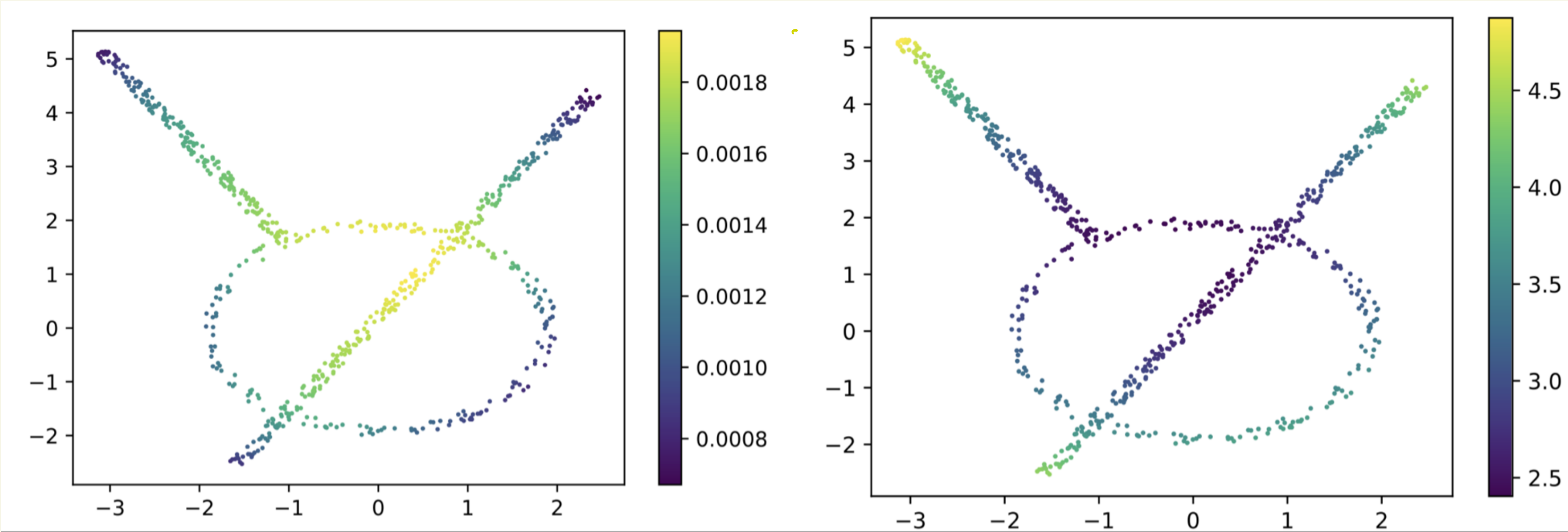
$$E_p(x) = \left(\frac{\sum_{y \in P} d(x,y)^p}{N} \right)^{1/p}$$

where N = total points



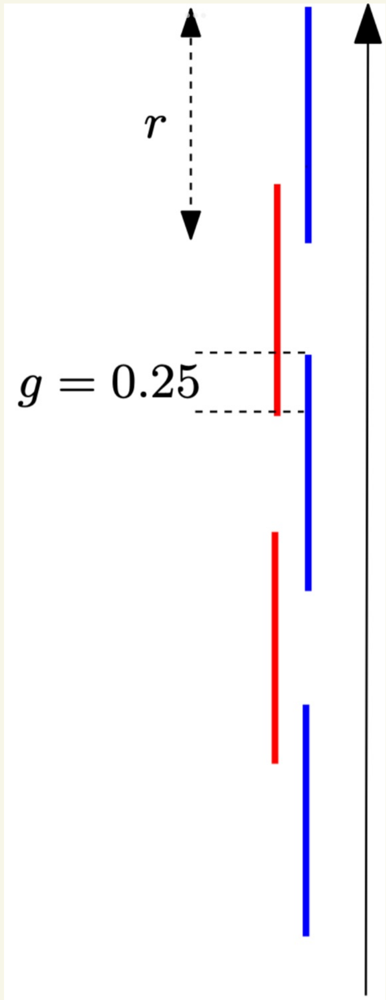
Eccentricity E_1

Whap Mappers would these give?
(need cover + cluster still!)



Choice of covers

- Resolution r is max diameter of an interval in $\mathcal{U}$
  (can also use N for # of intervals)
- Gain g is % overlap between intervals



Intuition:

- small r or large N
  ↳ finer resolution, more nodes
- large r (or small N)
  ↳ rougher resolution, few nodes
- small g → less connected
- large g → more connected

Choice of clustering

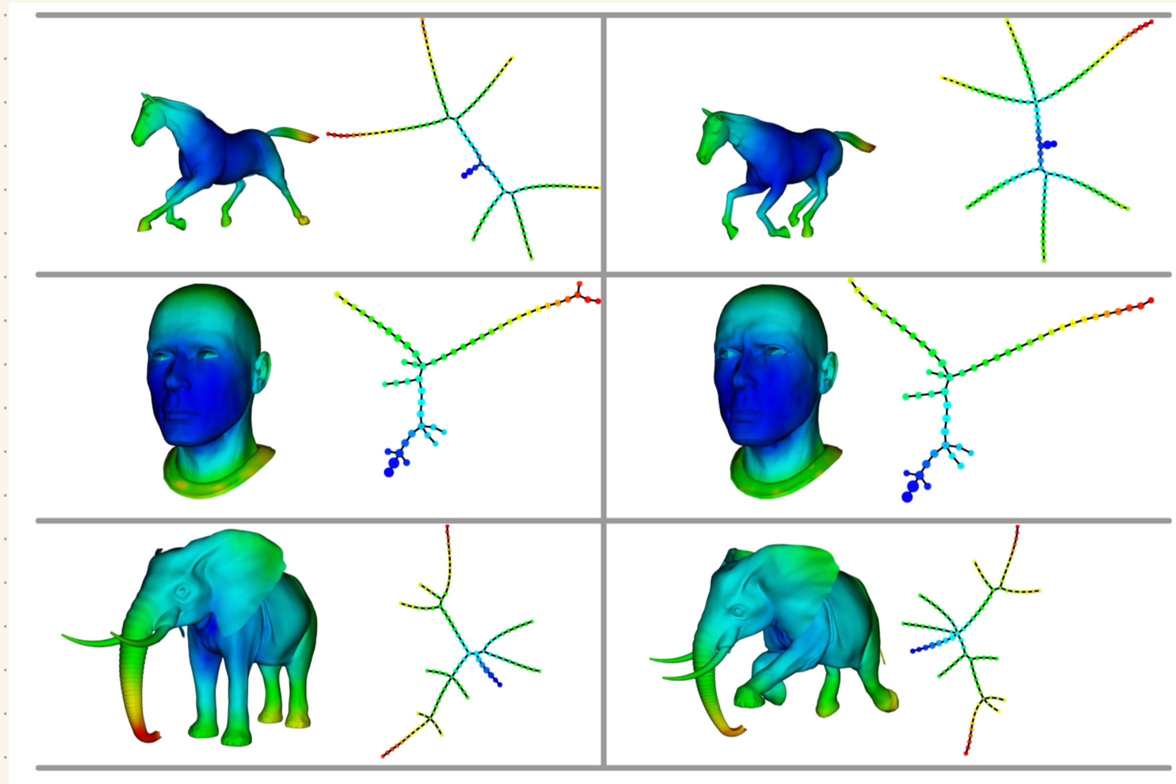
1. Build a graph: select # neighbors for kNN graph, or r for Rips graph
   ↳ then take connected components of subgraph spanned by vertices in $f^{-1}(U)$

2. Take points in bin, + choose your favorite clustering algorithm!
   ↳ more adaptive: can vary between the bins

Shape analysis

Singh et al, Eurographics 2007

Build mapper "skeletons": Find landmarks on mesh, & scale f by eccentricity.



15 intervals
50% overlap

Then: Use these for shape matching

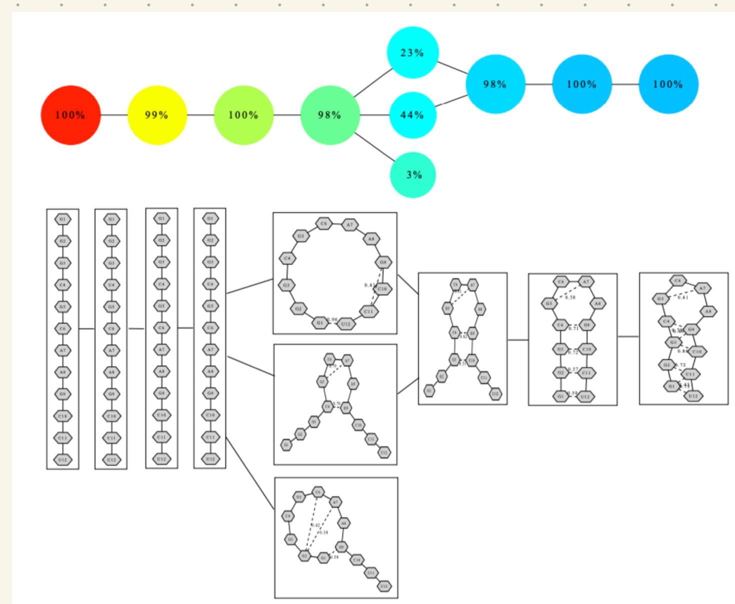
In applications: noisy & explorative!

Yao et al, J. Chemical Physics, 2009

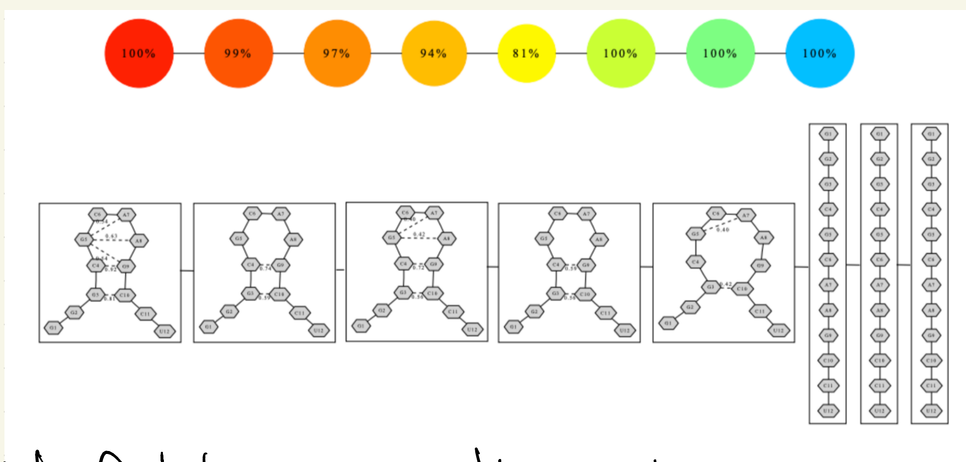
Data: conformations of molecules

Goal: Detect different folding pathways

Clustering on density level sets identifies & separates metastates from intermediate states



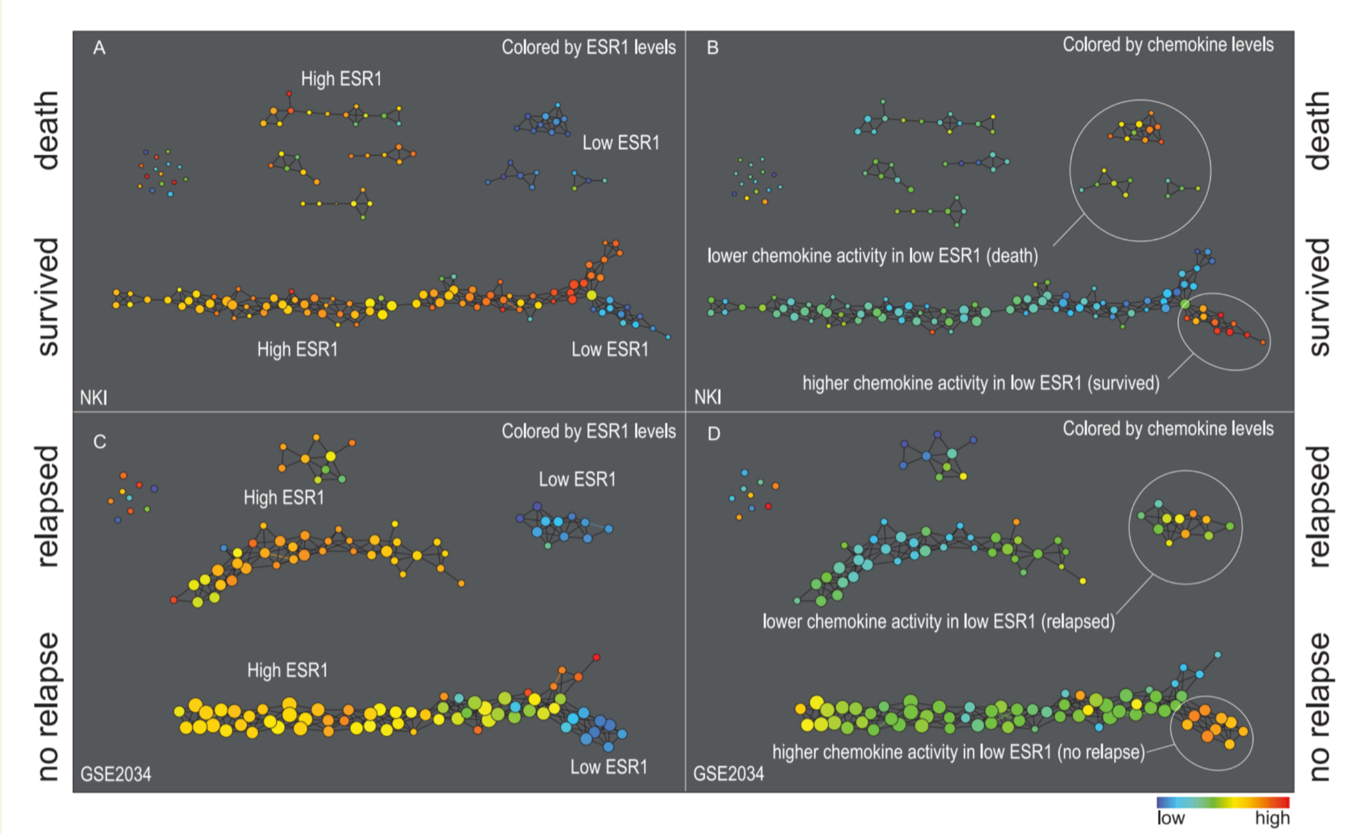
Refolding pathway



Unfolding pathway

Another: Breast cancer patients

Goal:
Detect variables that influence survival after therapy in breast cancer patients



Lum et al, Nature, 2013

- ESR1: estrogen receptor gene, positively correlated w/ survival

Result: identified subgroups w/ low ESR1 who survive across 2 studies

Post-Nature: lots of usage

A Comprehensive Review of the Mapper Algorithm, a Topological Data Analysis Technique, and Its Applications Across Various Fields (2007-2025)

Vine Nwabuisi Madukpe[1,3], Bright Chukwuma Ugoala[2,3], Nur Fariha Syaqina Zulkepli[1]*

[1]School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia.
[2]Department of Biostatistics, Augusta University,1120 15th Street, Rains Hall, Augusta, GA 30912, United States.
[3]Department of Mathematics, Abia State University, P.M.B. 2000, Uturu, Abia State, Nigeria.

*Correspondence: farihasyaqina@usm.my Tel.: +6012-543-2006

Used for visualization in biology, finance, social sciences, & many other applications.

Implementations:

(plus more!)

Software	Advantages	Disadvantages
KeplerMapper Van Veen et al. (2019)	Open-source and actively maintained. User-friendly with comprehensive documentation. Integrates well with Python data science libraries.	Scalability Issues Requires familiarity with Python. Subjectivity in Visualization
Python Mapper Müllner (2013)	Integration with Python Ecosystem Flexible and Customizable	Not Always Scalable for Big Data Limited Documentation and Community Support
tda-mapper Simi, L. (2024).	Optimized for large-scale data.	Relatively newer, smaller user base.
Mapper Interactive Zhou et al. (2020)	Interactive visualization capabilities. Scalable and extendable for large datasets.	Web-based interfaces may require specific setups. Web-based interfaces may require specific setups.
TDA Mapper Pearson (2013)	Integrates seamlessly with R	It may not be as actively maintained.
Sakmapper Szairis (2016)	Automated Selection of Parameters It handles large and complex datasets Faster computation	Potential Loss of Detail Less Control for Experts Limited Adoption and Documentation

Some take-aways:

2 key pieces:

- local clustering (guided by f)
- global connectivity between clusters (covers + nerve)

In general, used for exploration & visualization of data

- powerful in many applications
- not many theoretical guarantees

Warning:

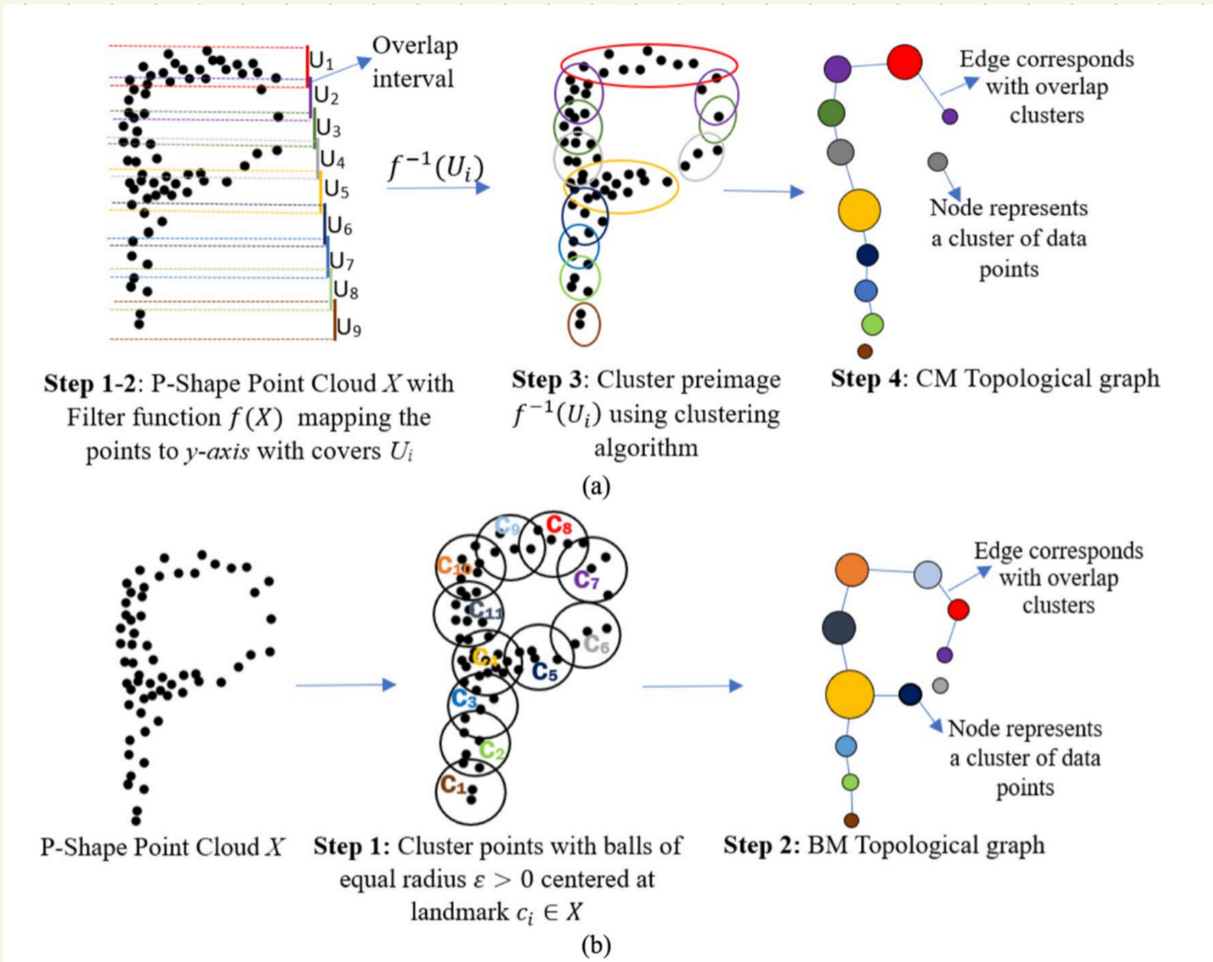
VERY sensitive to parameters!

Ball Mapper $\boxed{\text{Dlotko 2019}}$

An alternate (parameter-free) way to get a cover: build balls that cover the data (an ε-net)

Mapper:

Ball Mapper:



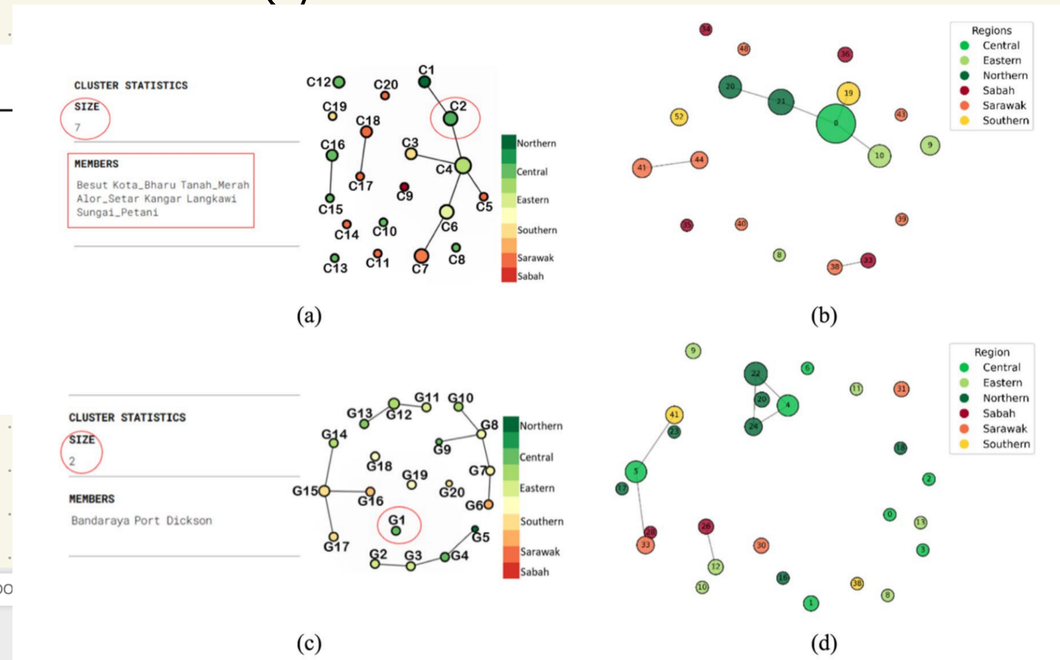
New, but seeing more usage

Environ Monit Assess (2025) 197:136
https://doi.org/10.1007/s10661-024-13477-2

RESEARCH

Comparative analysis of Ball Mapper and conventional Mapper in investigating air pollutants' behavior

Vine Nwabuisi Madukpe ·
Nur Fariha Syaqina Zulkepli ·
Mohd Salmi Md Noorani · R. U. Gobithaasan



Journal of Computational and Graphical Statistics ›
Volume 33, 2024 - Issue 4

Enter keywords, authors, DO

Submit an article | Journal homepage

Listen

Complex Data Analysis

Mapper–Type Algorithms for Complex Data and Relations

Paweł Dłotko, Davide Gurnari & Radmila Sazdanovic

Pages 1383-1396 | Received 29 Mar 2023, Accepted 09 Apr 2024, Accepted author version posted online: 22 Apr 2024, Published online: 07 Jun 2024

Cite this article https://doi.org/10.1080/10618600.2024.2343321 Check for updates

late

Figure 6 Ball Mapper applied to Jones polynomial data of knots up to 17 crossings with (a) just one choice of a mirror, (b) knots and their mirrors with standard Ball Mapper, and (c) Equivariant Ball Mapper. Color reflects the average signature of knots in each cluster. Note that the graph in (c) is symmetric, although this fact is not accurately represented in this image due to the chosen graph plotting subroutine.