# Modified Mapper: Estimating Reeb Graphs through Topological Changepoint Detection

Padma Ragaleena Tanikella

Advisor: Prof. Nicole Lazar

## Topological Data Analysis (TDA)

- Topology studies properties of objects that are invariant under any **continuous deformation**.
- Topological data analysis (TDA) is a branch of statistics that offers tools to extract **topological information** from data.

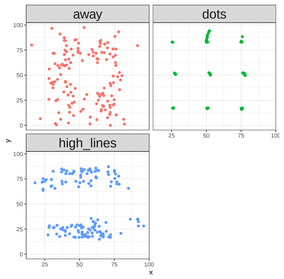



Table 1. Summary statistics for three `datasauRus` datasets

| | away | dots | high_lines |
|---|---|---|---|
| $\mathbb{E}[X]$ | 54.266 | 54.260 | 54.269 |
| $\mathbb{E}[Y]$ | 47.835 | 47.840 | 47.835 |
| $SD(X)$ | 16.770 | 16.768 | 16.767 |
| $SD(Y)$ | 26.940 | 26.930 | 26.940 |
| $\rho(X,Y)$ | -0.064 | -0.060 | -0.069 |

The topological information of interest are usually of two types:

1. **Shape information**, i.e., what is the "shape" of the (unknown) space from which the data is sampled?
2. **Connectivity information**, i.e., how the individual parts interact to form the whole.

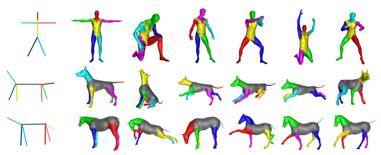


Figure 1. Goal: Given a shape, segment it into a small number of meaningful components (Skraba et al., 2010).

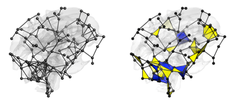


Figure 2. Goal: Model higher order interactions (Anand and Chung, 2021).

Cell complexes like **Rips** and **cubical complexes** are used to model shape or connectivity information.

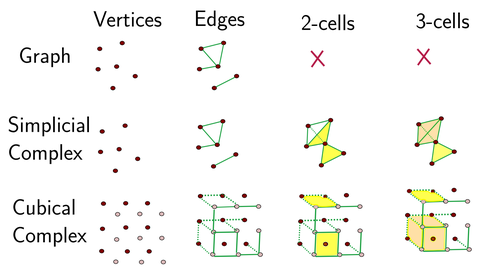
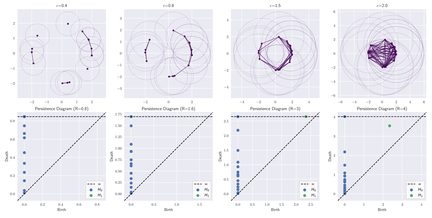
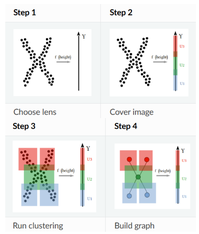



Figure 3. Cell complexes generalize the idea of mathematical graphs.

Persistence Diagrams and Mapper graphs are the most commonly used topological summary statistics.






## Challenges with Mapper and proposed modifications.

**Mapper** is a statistical version of **Reeb graph**.

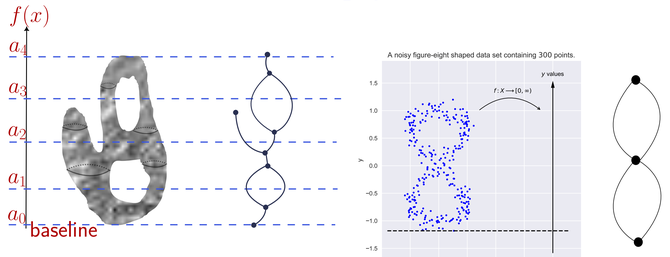
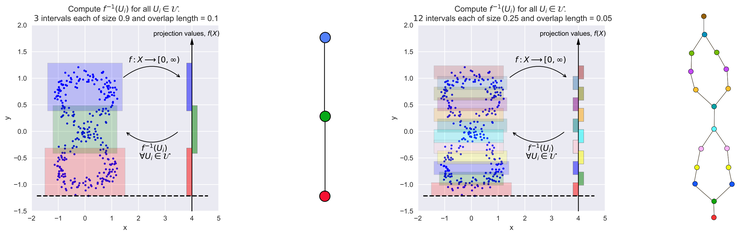



Figure 4. A manifold $\mathcal{M}$ and its Reeb graph w.r.t. the projection function.

Challenges with Mapper:

- Mapper output depends on how we choose to cover the filter function values.
- Mapper algorithm does not place nodes in the output based on how the topology of preimages changes.




Estimating the locations of **topological changepoints** can improve the accuracy of Reeb graph estimation.

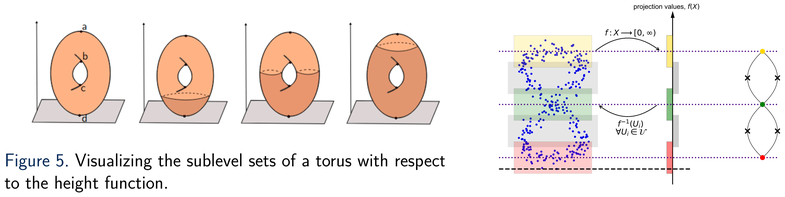



Figure 5. Visualizing the sublevel sets of a torus with respect to the height function.

Extended Persistence Diagrams are defined such that all critical points of the object get paired.

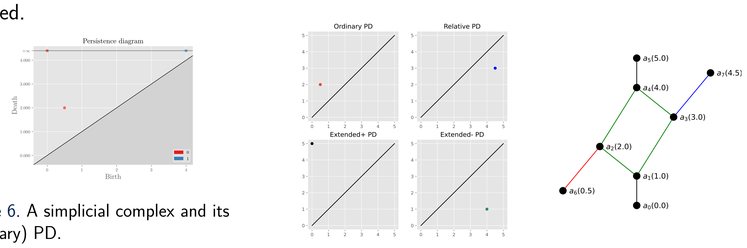



Figure 6. A simplicial complex and its (ordinary) PD.

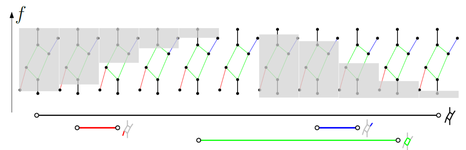
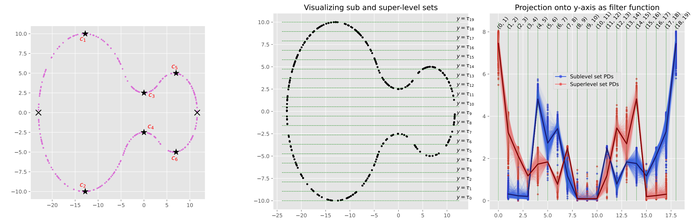



Figure 8. Visualizing the sublevel (left) and superlevel sets (right) of the complex. Note that the unshaded region represents the sub/superlevel set (M Carriere et al., 2020).

## Implementation and future work

Dataset 1: Visualizing the critical points with respect to projections onto $x$ and $y$-axis and the distance between consecutive sublevel and superlevel sets when filter function is projection onto $y$-axis.




Dataset 2: Visualizing the swiss-roll dataset in different directions.

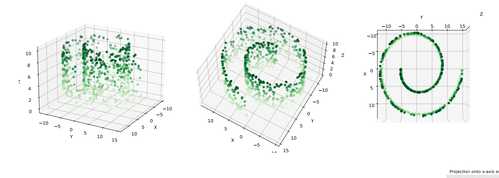





Figure 9. Changepoints w.r.t projection onto x-axis.

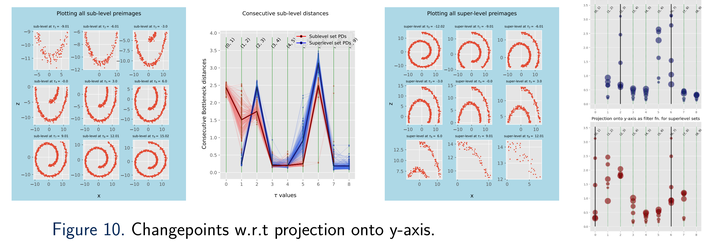



Figure 10. Changepoints w.r.t projection onto y-axis.

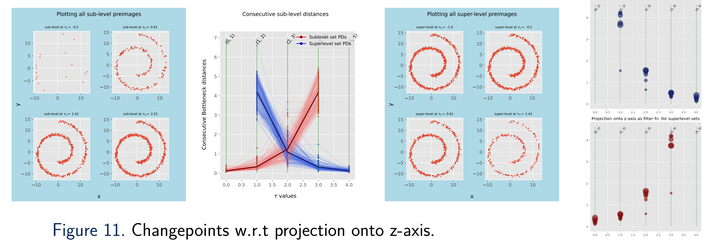



Figure 11. Changepoints w.r.t projection onto z-axis.

**Future work**:

- In how many directions should we estimate the Reeb graph to get "complete" topological information?
- We are working on applying these methods to fMRI data to gain better insights into how individuals perceive the physical and mental states of others.

Pennsylvania State University, University Park

pbt5137@psu.edu

Department of Statistics, Eberly College of Science