# Adaptive Covers for Ball Mapper

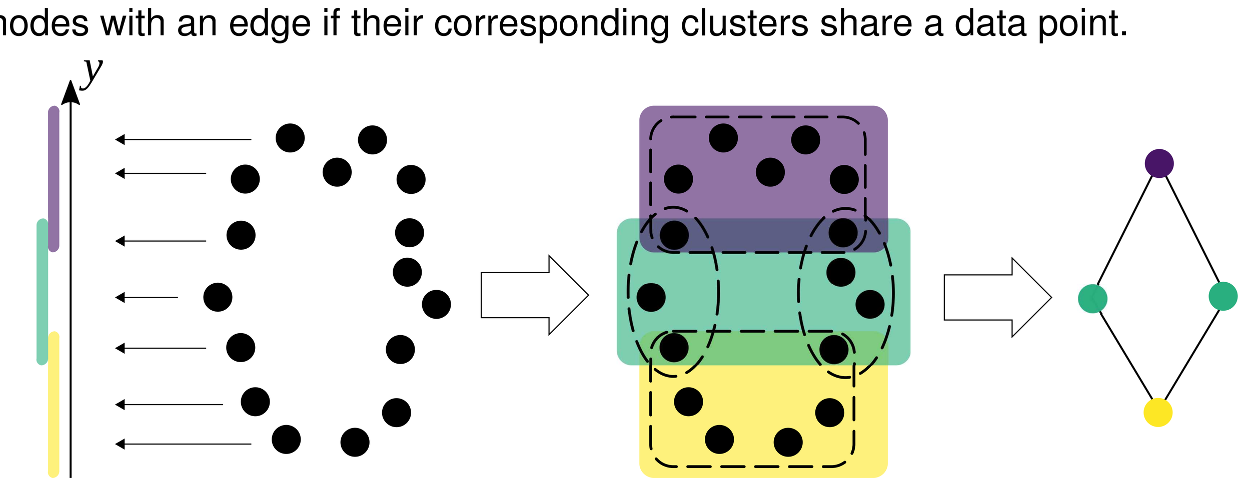
Enrique Alvarado [1], Robin Belton [2], Kang-Ju Lee [3], Sourabh Palande [4], **Sarah Percival** [4], Emilie Purvine [5]

[1] University of California, Davis [2] Smith College [3] Seoul National University [4] Michigan State University
[5] Pacific Northwest National Laboratory

## Mapper

The mapper algorithm is a flexible topological tool that describes the underlying structure of a dataset in a simplified manner. Its strength lies in the ability to capture both physical structure, such as in volumetric data, and abstract structure, such as in datasets that do not necessarily admit an embedding into Euclidean space. Given an input set $X$ of point cloud data, the mapper algorithm works as follows:

1. Choose a lens function $f : X \to \mathbb{R}$ that assigns to each point in $X$ a real number.
2. Cover $\mathbb{R}$ with a set of overlapping intervals $\{U_\alpha\}$.
3. Cluster the points inside each $f^{-1}(U_\alpha)$.
4. Create a node in the mapper graph for each cluster obtained in Step 3, joining two nodes with an edge if their corresponding clusters share a data point.




The Mapper algorithm

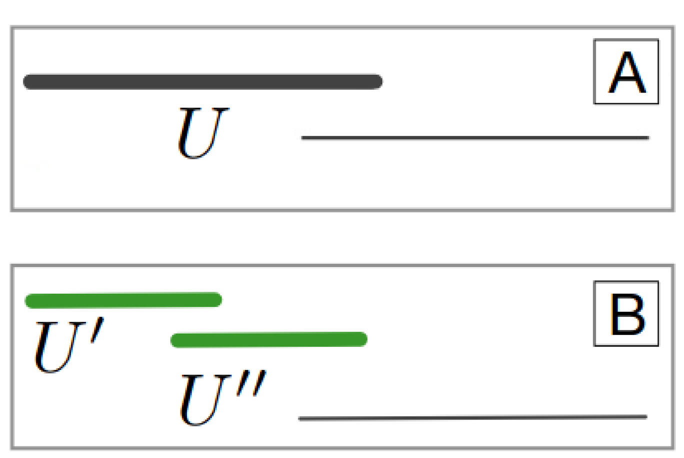
The mapper algorithm depends on the choice of lens function, distance metric, covering, and clustering algorithm.

## Adaptive Covers for Mapper

To reduce the effects of parameter selection, Chalapathi et al. introduced adaptive covers for mapper. Instead of requiring the user to select a cover, this algorithm uses information criteria to iteratively generate a cover for the mapper graph. Their algorithm has the same inputs as the original mapper algorithm, and works as follows:

1. Generate a mapper graph $\mathscr{M}$ with $l$ intervals of uniform length.
2. Select an unprocessed open cover interval $U$.
3. Compute the Bayesian Information Criterion (BIC) of the subgraph of $\mathscr{M}$ induced by $U$.
4. Split $U$ into two overlapping intervals with user-specified overlap percentage $p$ and compute a new mapper graph $\mathscr{M}'$ for the refined cover.
5. Compute the BIC of $\mathscr{M}'$. If the BIC is improved, accept the split.

Repeat steps 2-5 until all intervals have been processed.

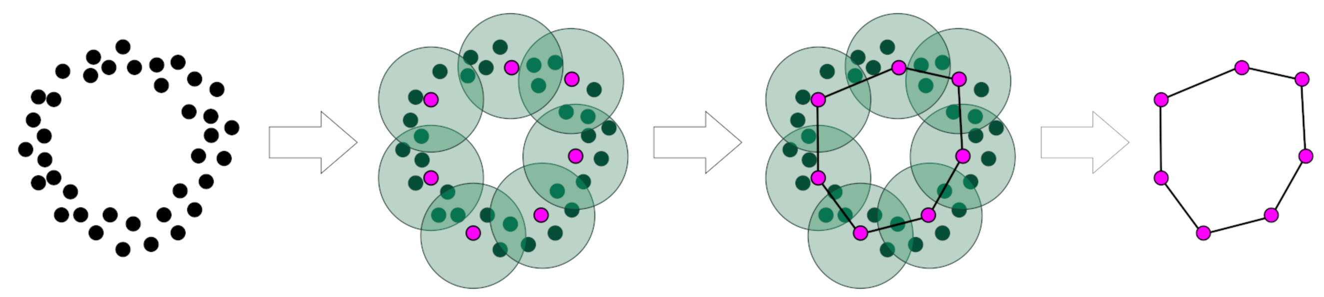



Interval splitting (Chalapathi et al.)

We gratefully acknowledge the Institute for Mathematical and Statistical Innovation for providing travel support.

## Ball Mapper

Similar to the traditional mapper algorithm, ball mapper, introduced by Dlowtko (2019), summarizes the shape of point cloud data into a one-dimensional graph.

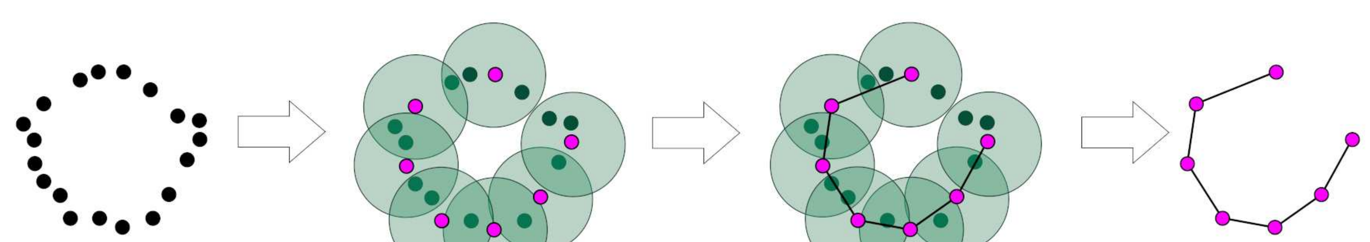


The algorithm works as follows:

1. Construct an $\varepsilon$-net for a user-specified parameter $\varepsilon$.
2. Construct the nerve of this epsilon cover, adding a simplex between centers if there is a point in their overlap.

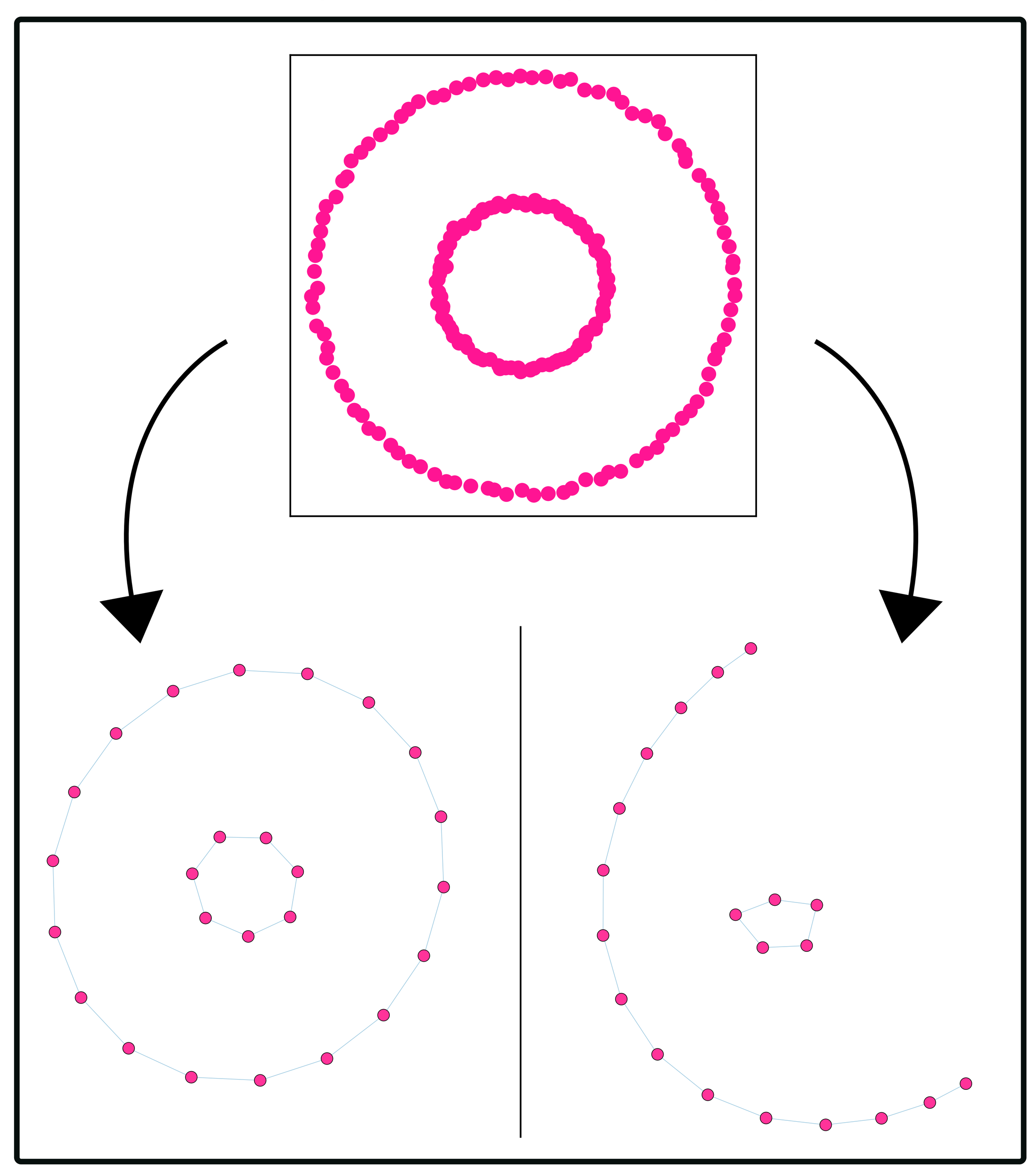
## Example

Not only is the structure of the ball mapper graph dependent on the parameter $\varepsilon$, but the structure of the ball mapper graph depends on the $\varepsilon$-net construction itself.



The ball mapper graph does not capture the circular shape of the data set.

Different $\varepsilon$-nets on the same point cloud can generate graphs with different topology.



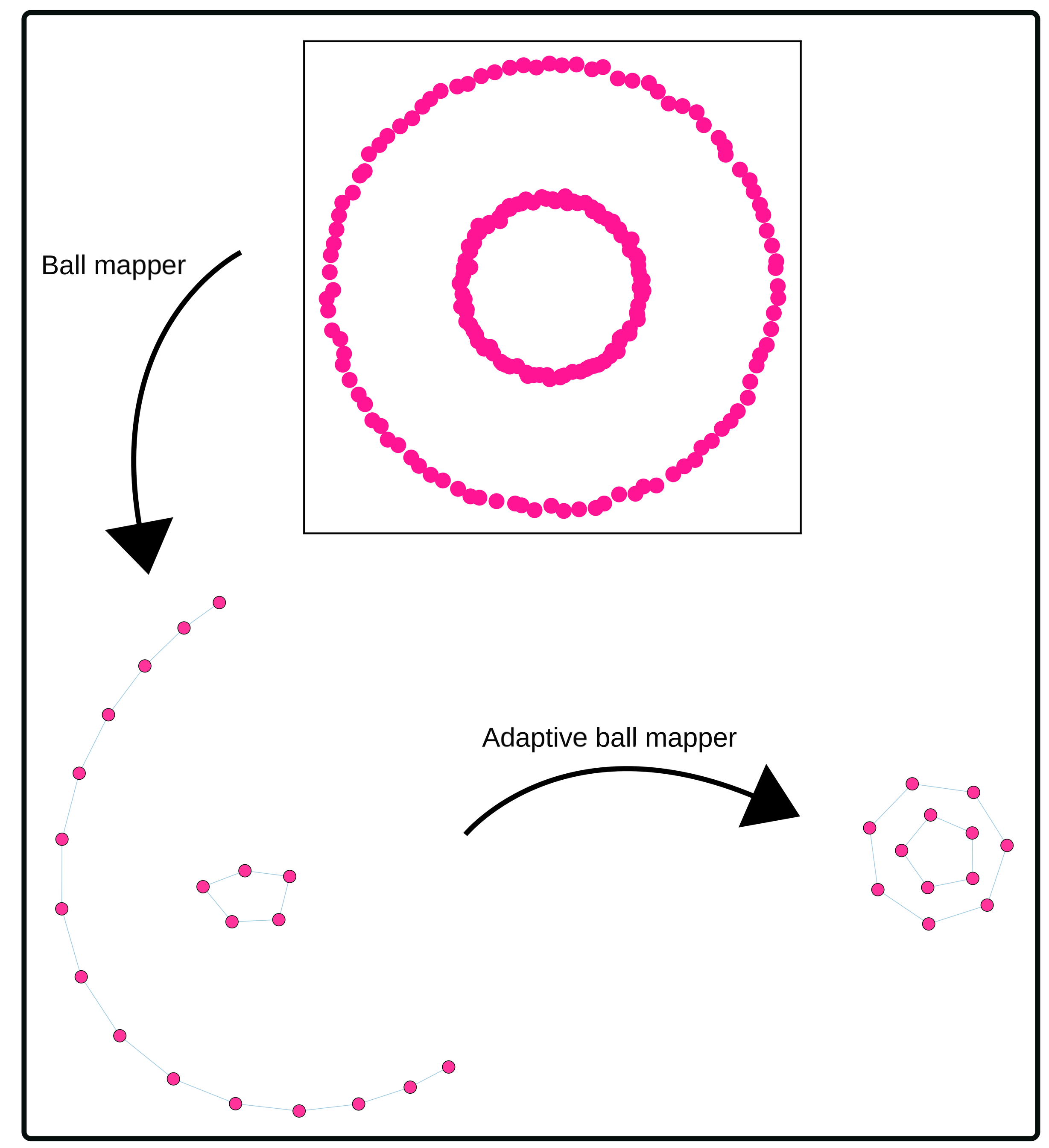
Two different ball mapper graphs originating from the same point cloud.

## Adaptive Ball Mapper

We developed an algorithm for adaptive covers for ball mapper which not only tests splitting the cover balls, but also tests merging them. The algorithm works as follows:

1. Construct a ball mapper graph $\mathscr{M}$ for a given $\varepsilon$.
2. Refine the open cover following Chalapathi et al, splitting balls by constructing an $\varepsilon * \delta$-net in place of the chosen ball, for a user-specified $\delta < 1$.
3. Select an unmerged cover element $B$.
4. For the two closest cover elements $B'$ and $B''$, compute ball mapper graphs $\mathscr{M}'$ and $\mathscr{M}''$, replacing $B$ with $B \cup B'$ and $B \cup B''$, respectively.
5. If the BIC of $\mathscr{M}'$ or $\mathscr{M}''$ is improved, accept the split for the graph that improves BIC the most.

Repeat steps 3-5 until the BIC no longer improves.




The graph that is constructed through this process of splitting and merging the cover balls more closely matches the underlying shape of the data set.

## References

Chalapathi, N., Zhou, Y., & Wang, B., “Adaptive Covers for Mapper Graphs Using Information Criteria,” *2021 IEEE International Conference on Big Data (Big Data)*, IEEE (2021)

Dlowtko, P. “Ball mapper: A shape summary for topological data analysis,” *arXiv preprint arXiv:1901.07410* (2019).

Singh, G., Memoli, F., & Carlsson, G. “Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition,” *Eurographics Symposium on Point-Based Graphics*, Botsch, M., Pajarola, R., Chen, B., and Zwicker, M., eds., The Eurographics Association (2007).