

TDA & THD

BiRG Meeting 03/07/18

Xiu-Huan Yap

Topological Data Analysis (TDA)

Higher-level, structural
understanding of data shape

Data
visualization

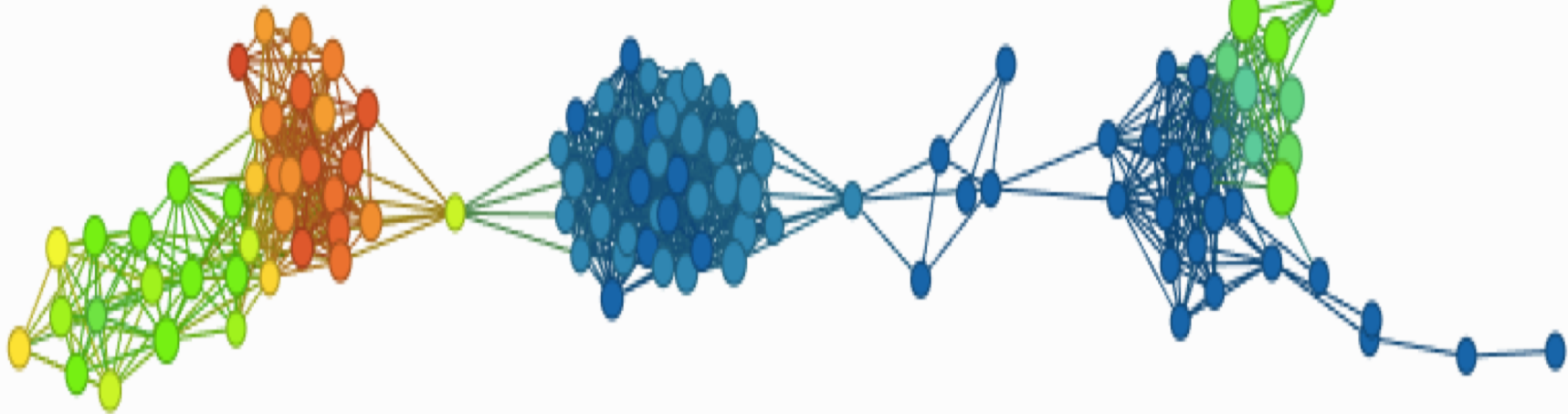
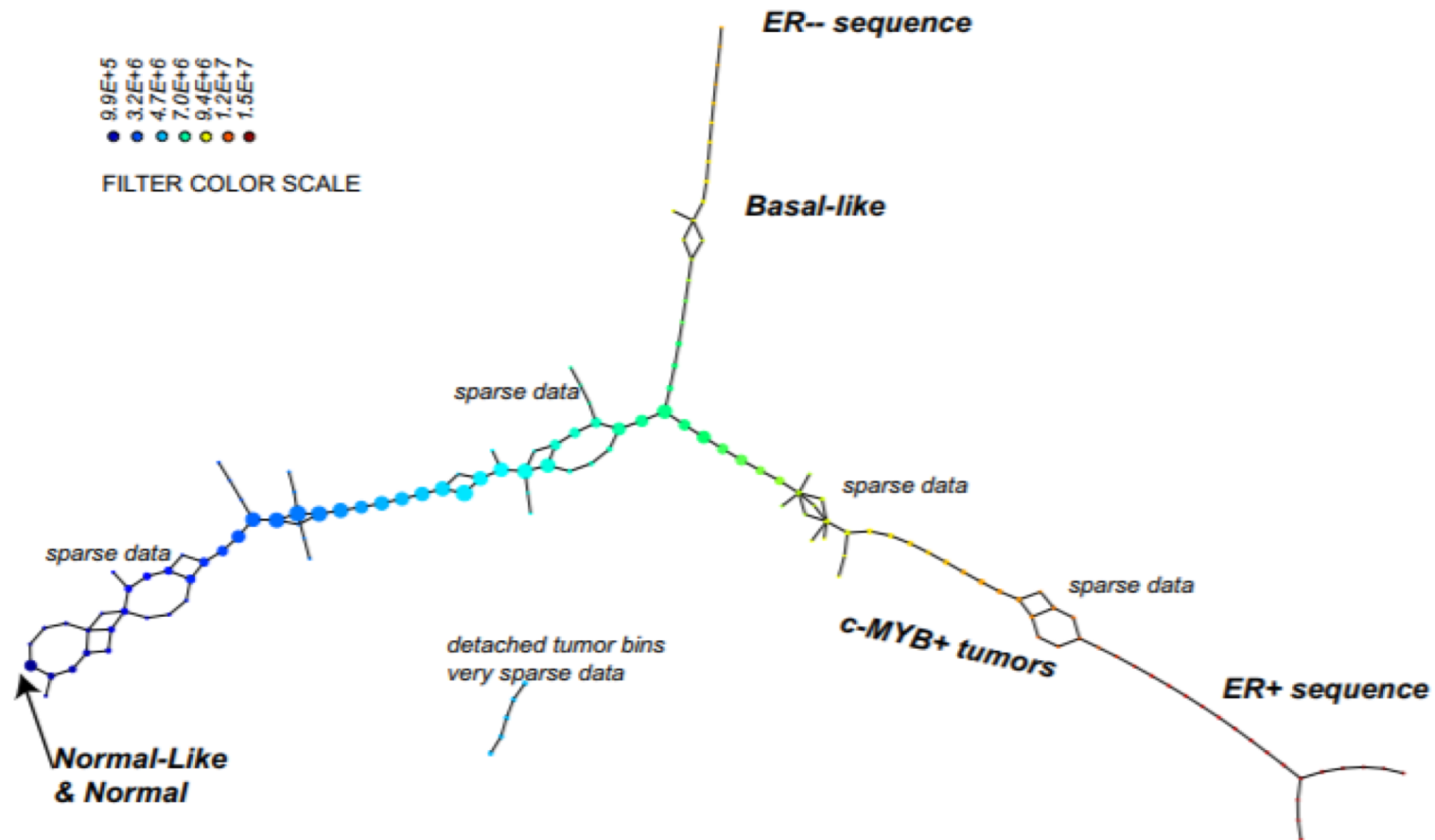


Exhibit relations between
clusters, variables etc

Less computationally
expensive

Breast cancer subtype classification



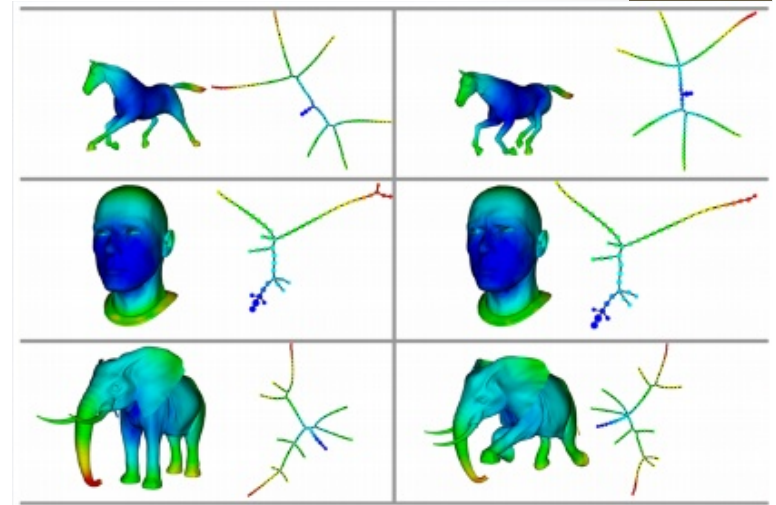
Nicolau et al., 2011. PNAS 108(17):7265-70

General principles

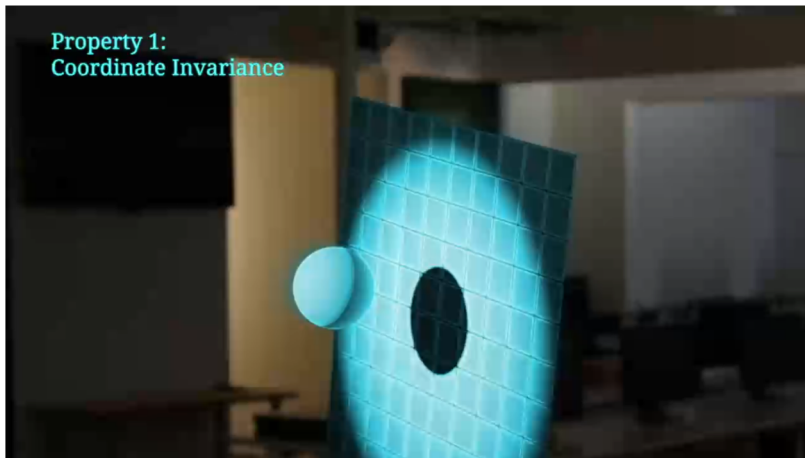
Data has shape and shape has meaning

Properties of topology that enables data shape analysis:

- 1) Co-ordinate invariance
- 2) Deformation invariance
- 3) Compressed representation



Property 1:
Coordinate Invariance



Property 3:
Compressed Representation



TDA workflow

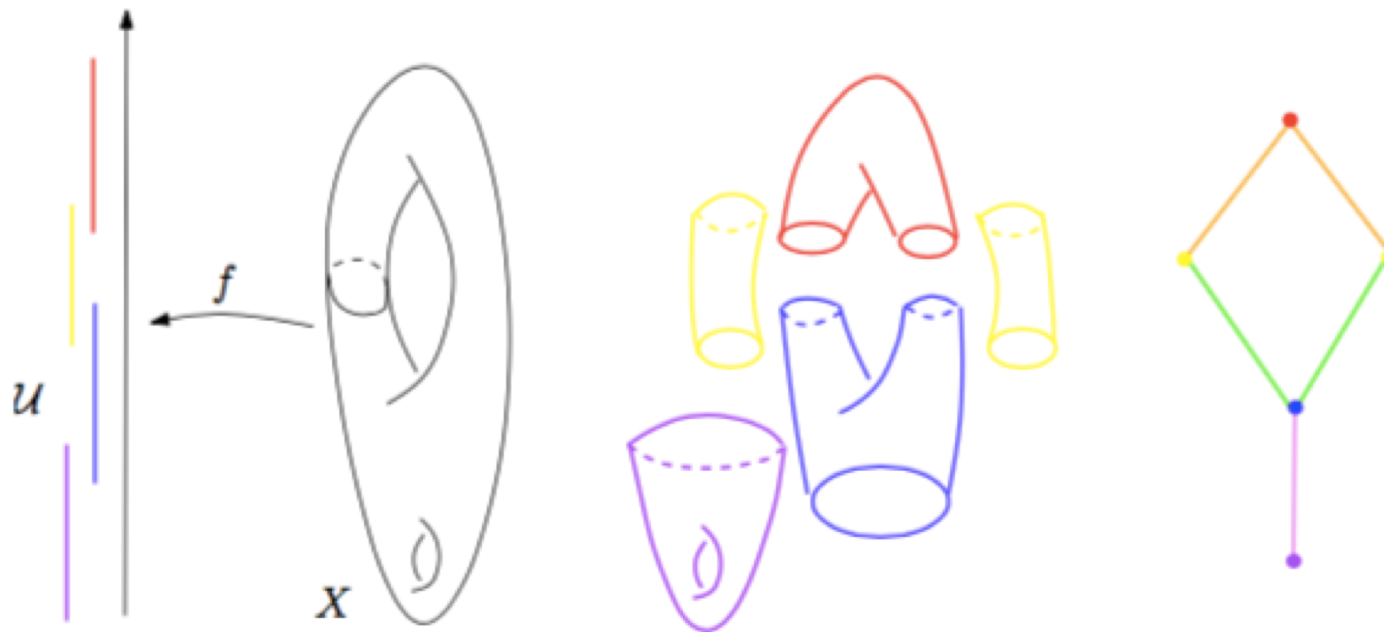


Figure 4: The refined pull back cover of the height function on a surface in $\mathbb{R}^3$ and its nerve

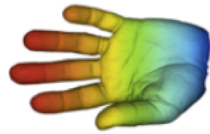
Chazal and Michel, 2017. An Introduction to TDA: fundamental and practical aspects for data scientists. arXiv:1710.04019 [math.ST]

TDA workflow

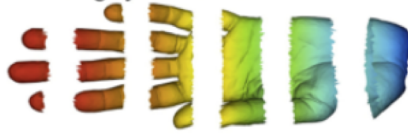
A Original Point Cloud



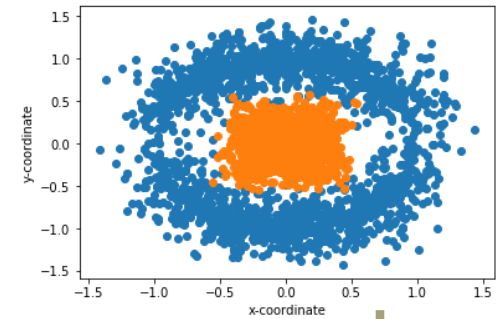
B Coloring by filter value



C Binning by filter value

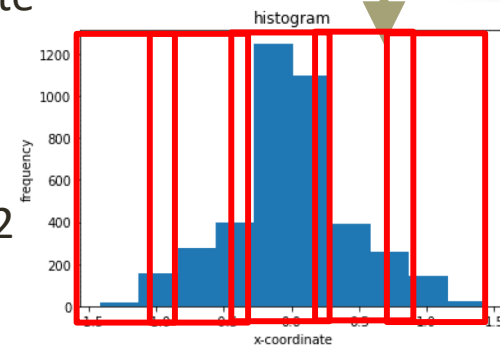


D Clustering and network construction

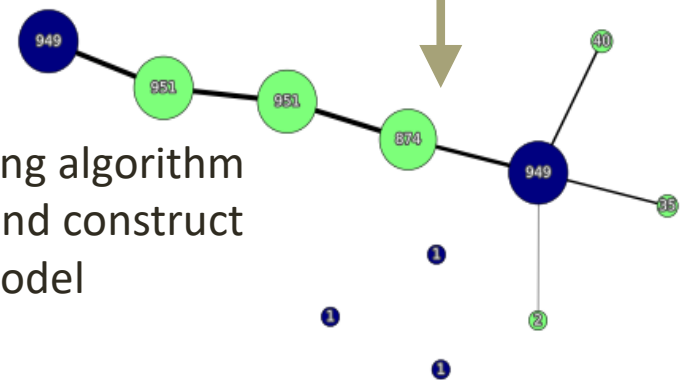


Filter: X-coordinate

Coverage: 5
Degree of overlap: 0.2

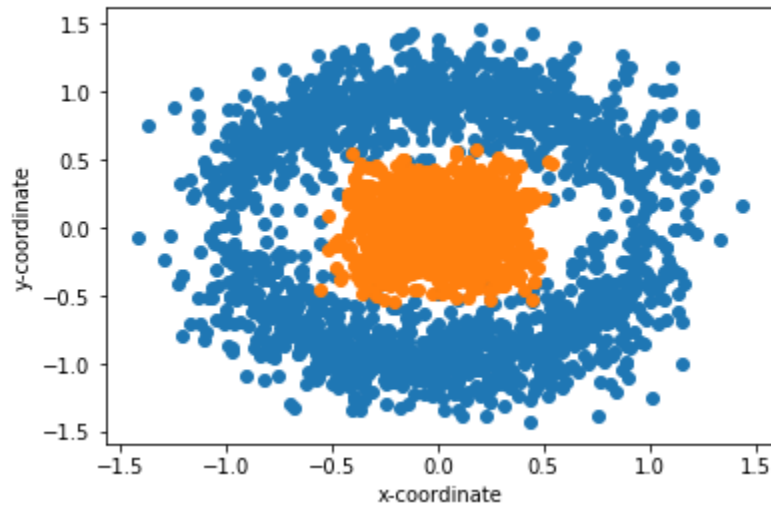


Apply clustering algorithm
for each bin and construct
topological model



Size range: [1,951]

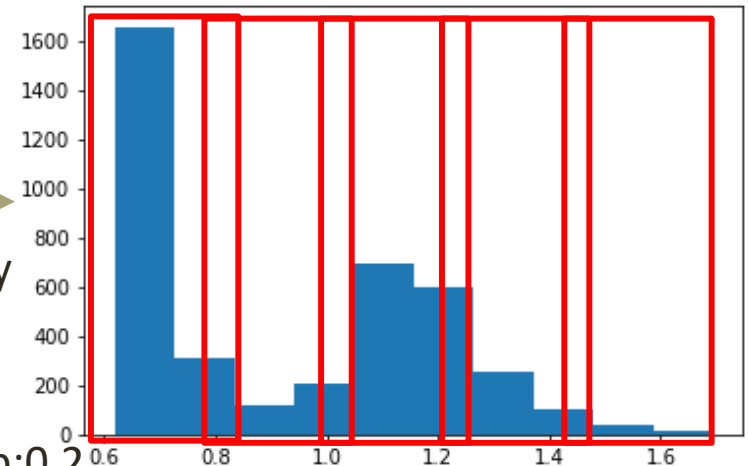
TDA workflow



Filter: Eccentricity

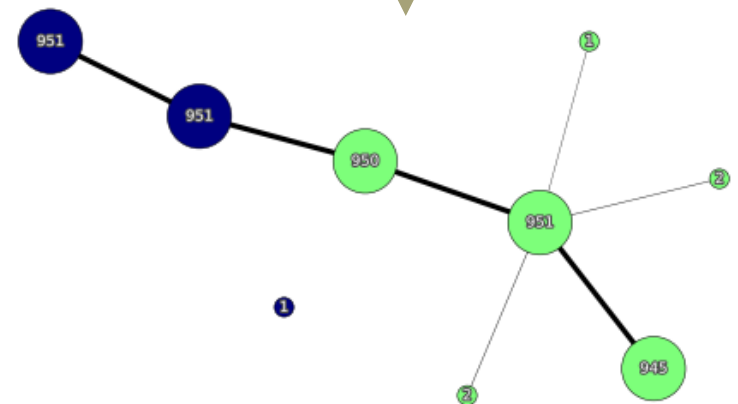
Coverage: 5

Degree of overlap: 0.2



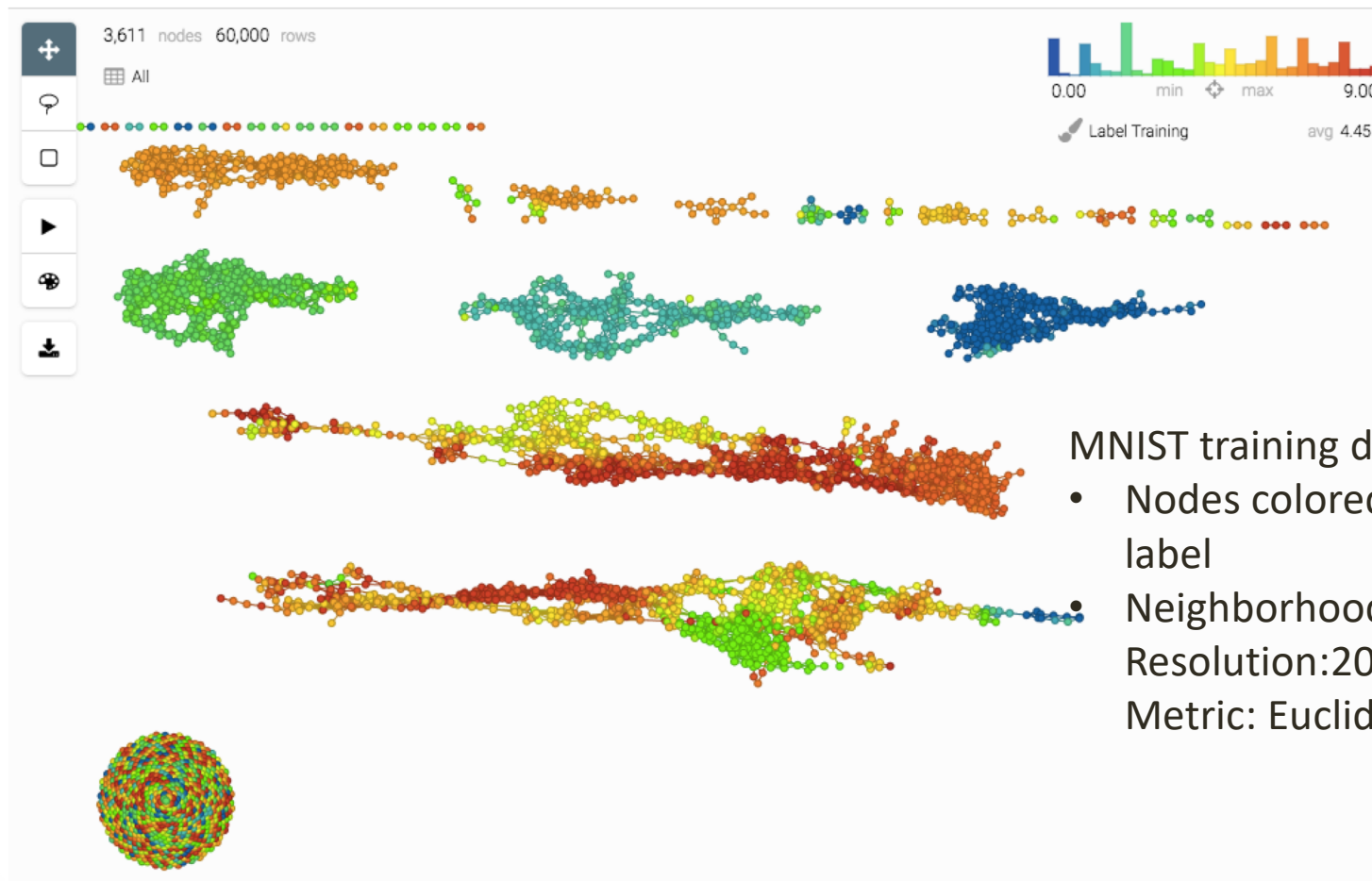
Eccentricity filter, a measure of **Centrality**: high value when a point is far from the 'center' of the data.

Cluster and construct top. model



Size range: [1,951]

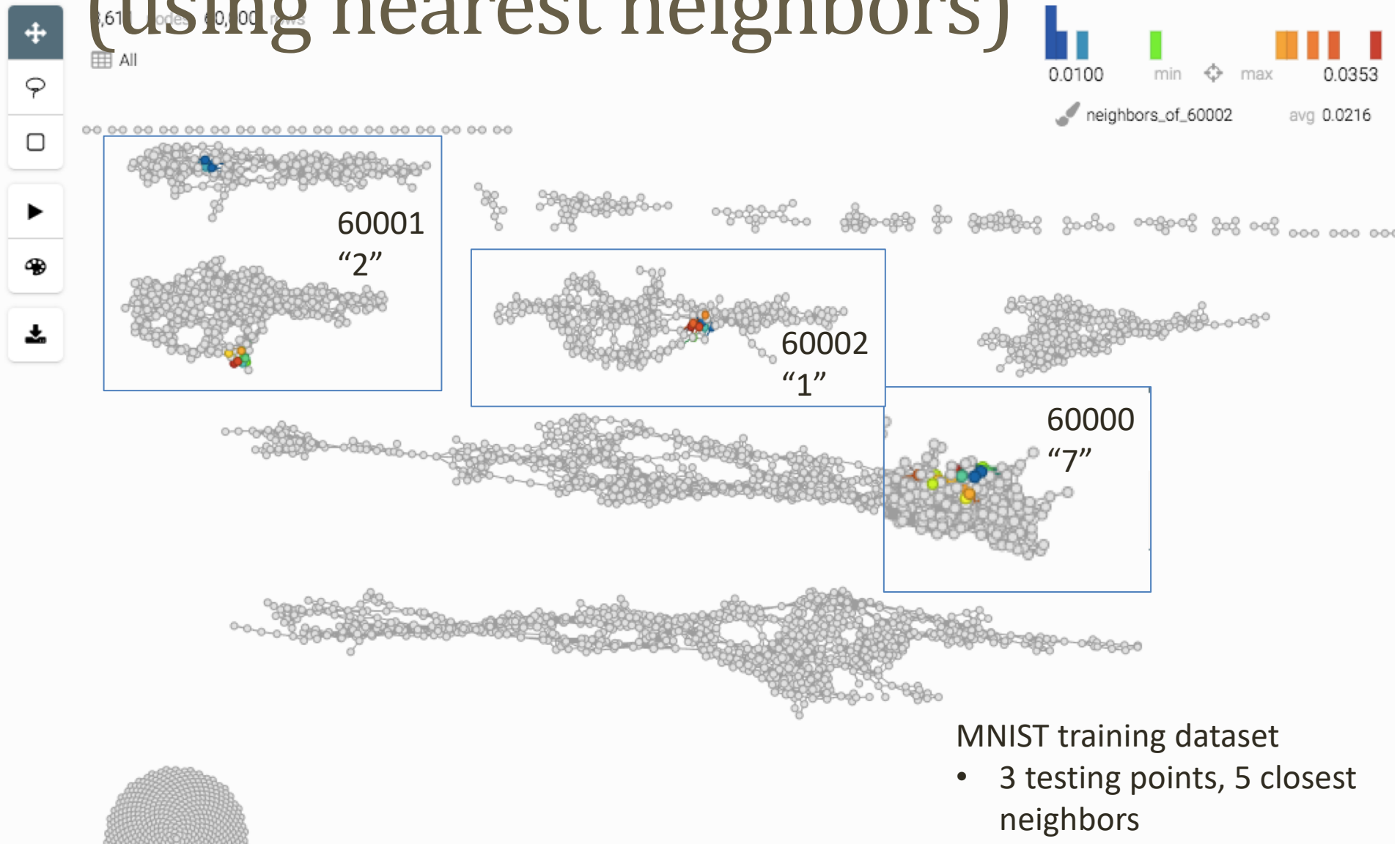
Ayasdi & KS analysis



MNIST training dataset

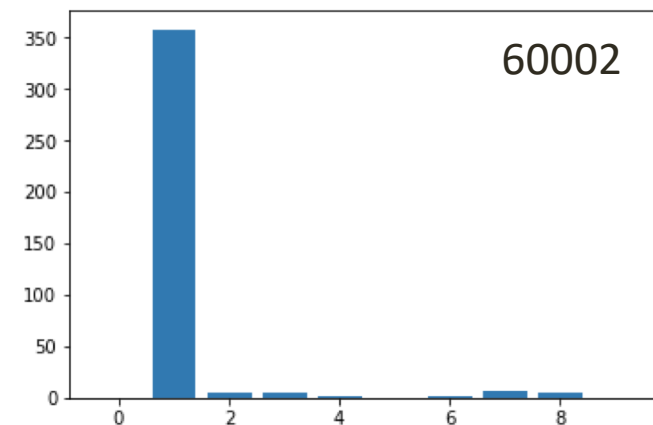
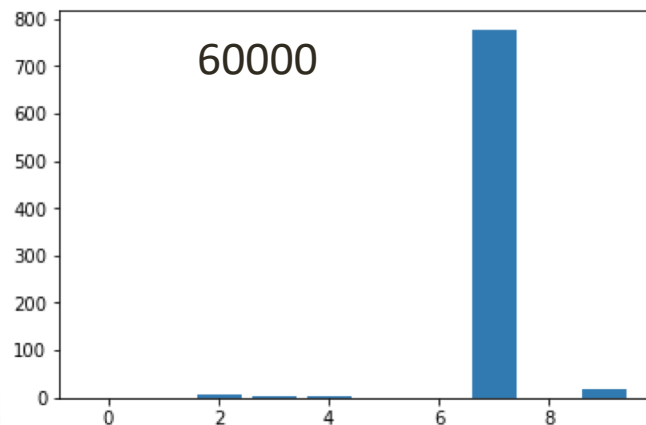
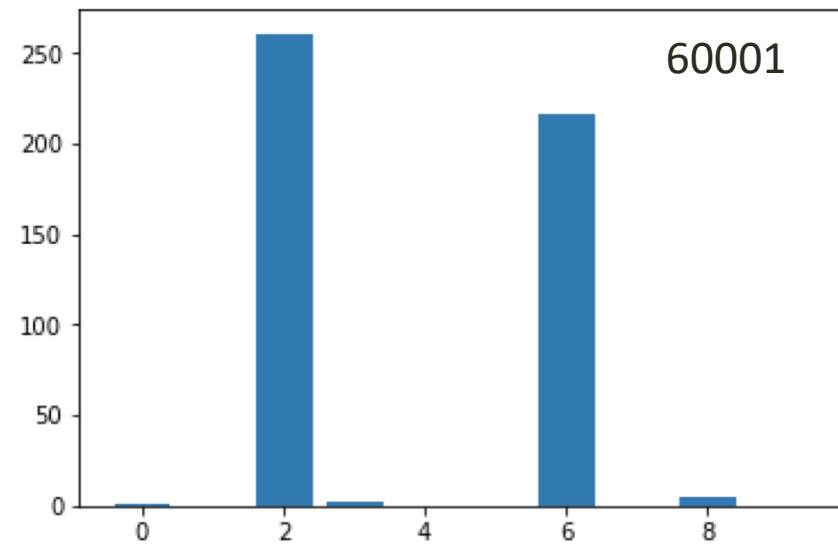
- Nodes colored by training label
- Neighborhood Lens 1&2, Resolution:200, Gain:2.5, Metric: Euclidean

Ayasdi & Topological predict (using nearest neighbors)

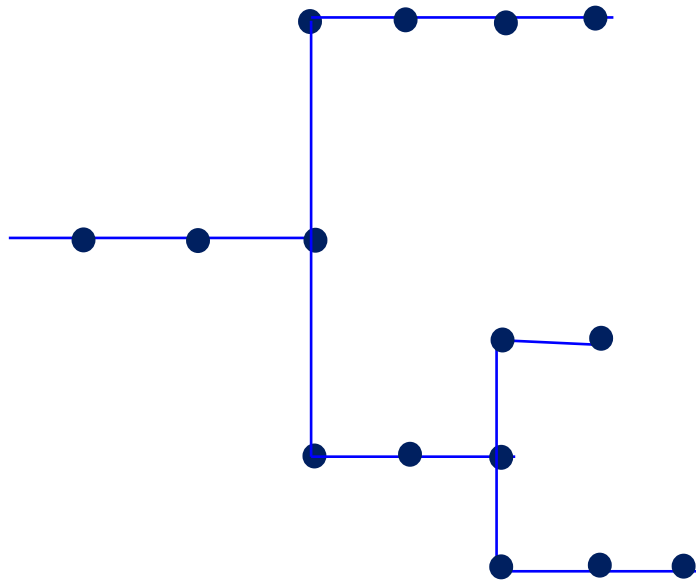


Topological predict (Using containing nodes)

Predicted values	Prevalence (Manual)	Prediction prevalence
7	777/807	160
2	261/485	
6	216/485	48
1	358/380	51

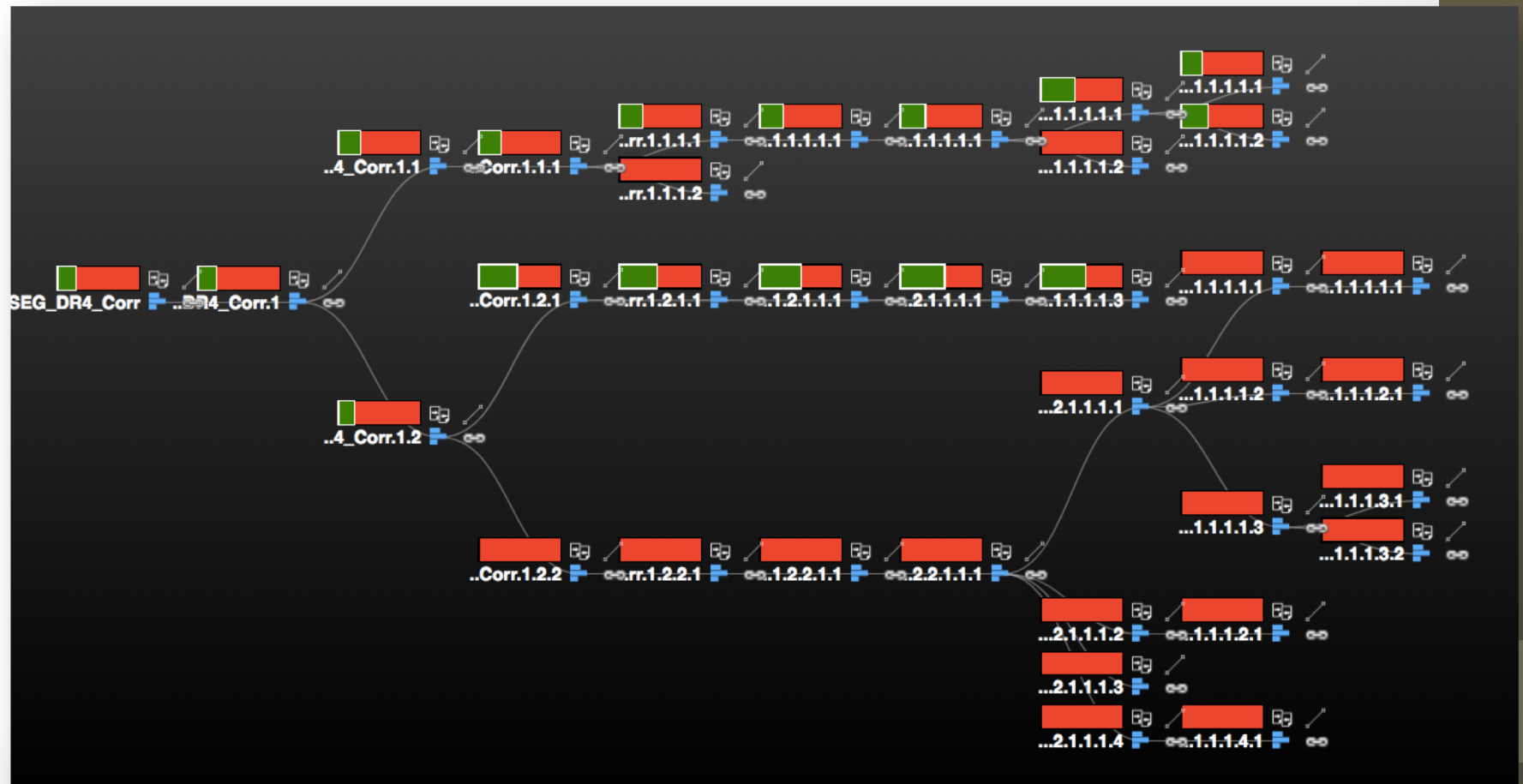


Topological Hierarchical Decomposition (THD)



- Each node represents a TDA network
- Networks formed from connected components of preceding networks
- Resolution is increased in discrete amounts until branches are observed (two or more connected components from 1 network)
- Terminal leaves formed when the group meets size thresholds (min/max # of points)

THD, Classification & Topological Prediction



Code by Brad Reynolds & Ryan Kramer, AFRL

Merged THDs

- Use binary labels to represent if nearest neighbor(s) are in each node
- Combine labels with other THDs

Possible methods of predictions:

- Linear or weighted combinations of THD predictions
- Use similarity comparisons to find closest training points across all THD models
- TDA/THD analysis of the merged THDs (interesting loops!)