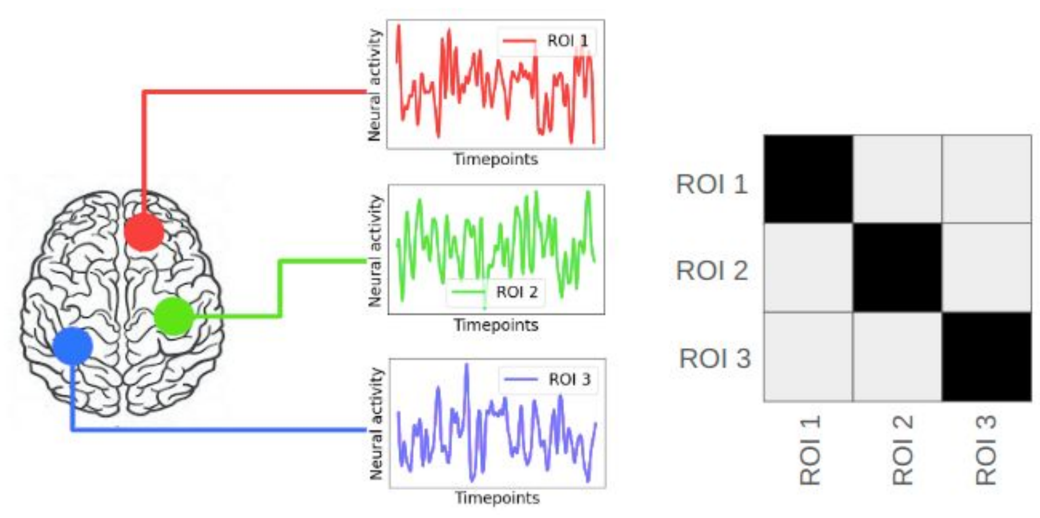
Neurological Diseases Spectrum Mapping using Topological Data Analysis

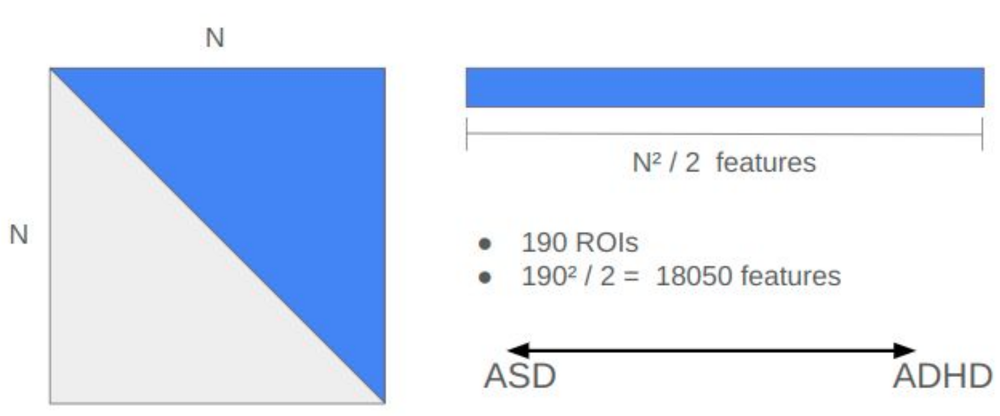
Charis Christopher Hulu (chulu@uic.edu)
Sidharth Kumar (sidharth@uic.edu)
Gopikrishna Deshpande (gopi@auburn.edu)

Introduction

Background: While behavioral studies have explored the spectrum between ASD and ADHD, there is a lack of research examining this continuum through fMRI-based brain connectivity. Our motivation is to fill this gap by leveraging topological data analysis to reveal transitional patterns in functional networks underlying these conditions.

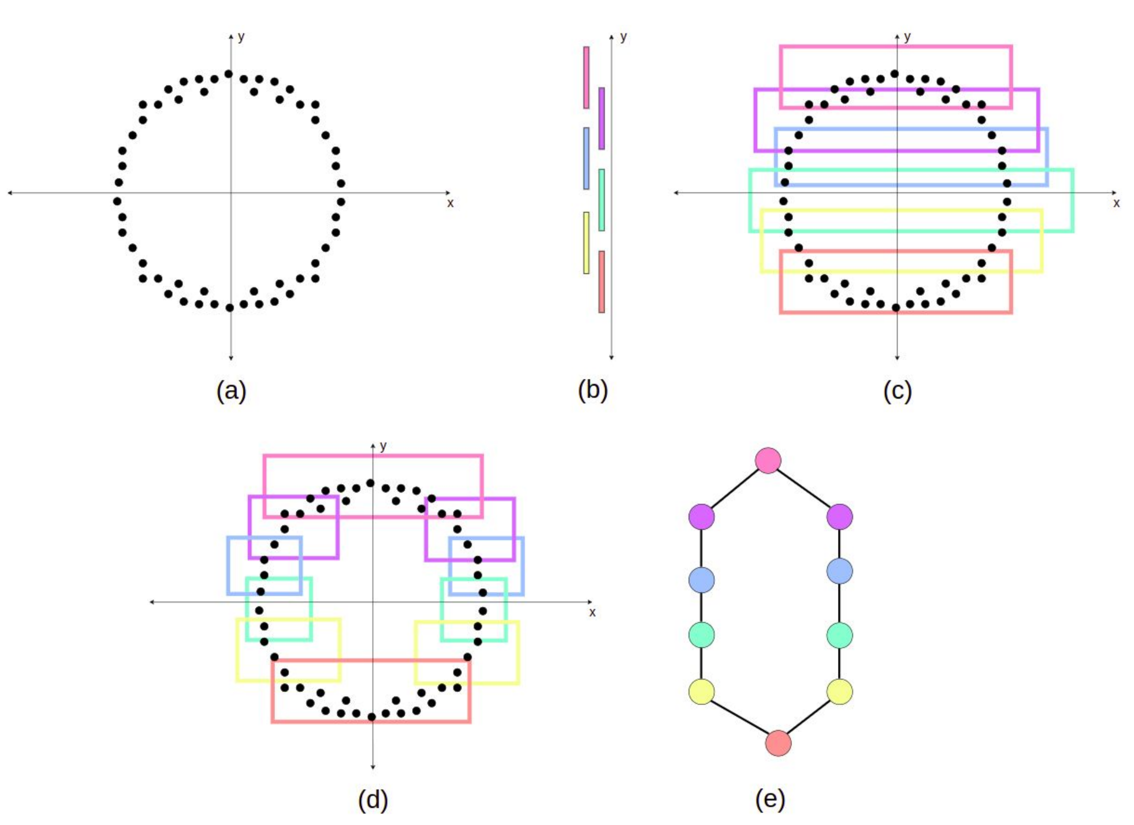
fMRI data





Mapper Algorithm

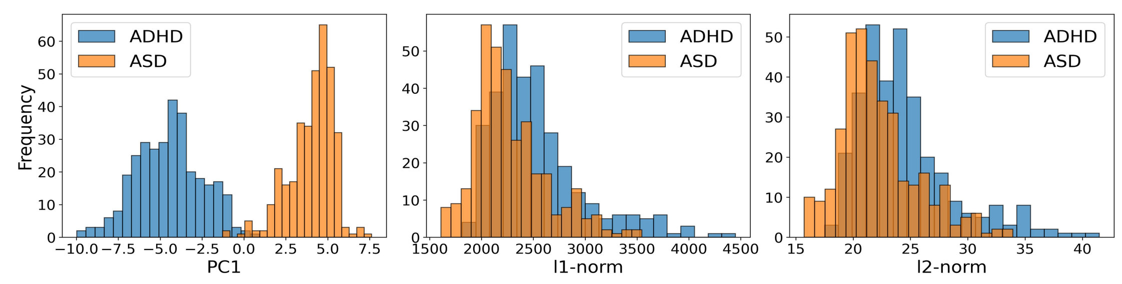
Mapper algorithm: a tool from topological data analysis that can extract meaningful structure from high-dimensional data. At a high level, it simplifies complex data by creating a graph that captures the shape of the data in a lower-dimensional representation.



a. Start with a dataset that has d features and apply the projection function f(x) to reduce each data point to a single value.
b. Divide the range of projection values into n overlapping bins, with each bin overlapping the next by p%.
c. Assign each data point to one or more bins depending on its projection value.
d. Within each bin, apply DBSCAN clustering to group data points.
e. Generate **mapper graph** by representing each resulting cluster as a node in a graph. If two clusters from adjacent bins share at least one data point, connect their nodes with an edge.

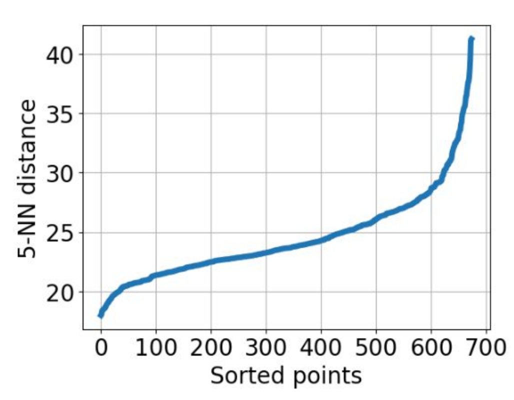
Generating Mapper Graph

Selecting projection function



We consider **PC1**, **L1 norm**, and **L2 norm** as candidate scalar projection functions, and select the one that provides **moderate overlap** between ASD and ADHD to ensure connectivity **without losing distinction** in the mapper graph.

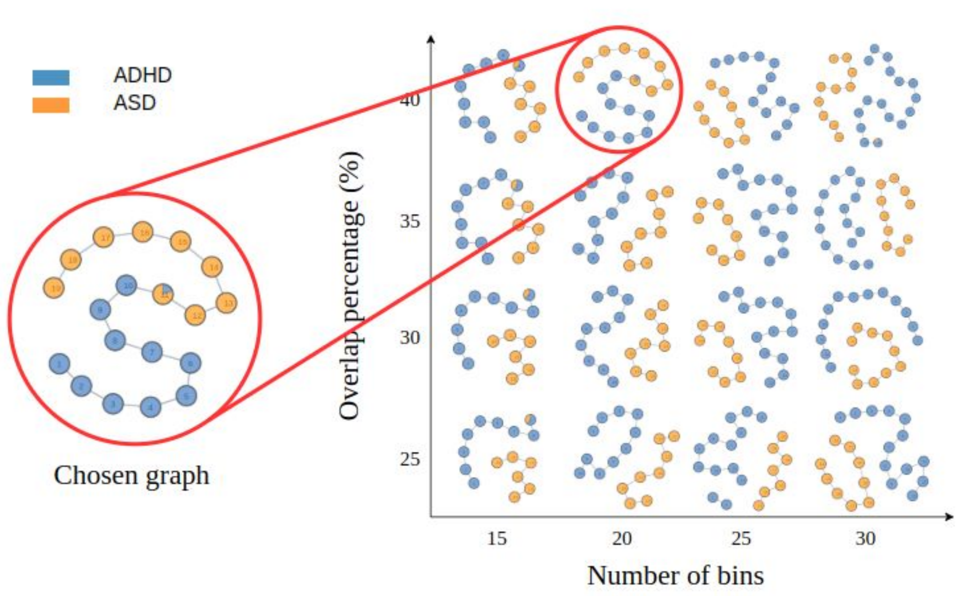
Selecting DBSCAN parameters



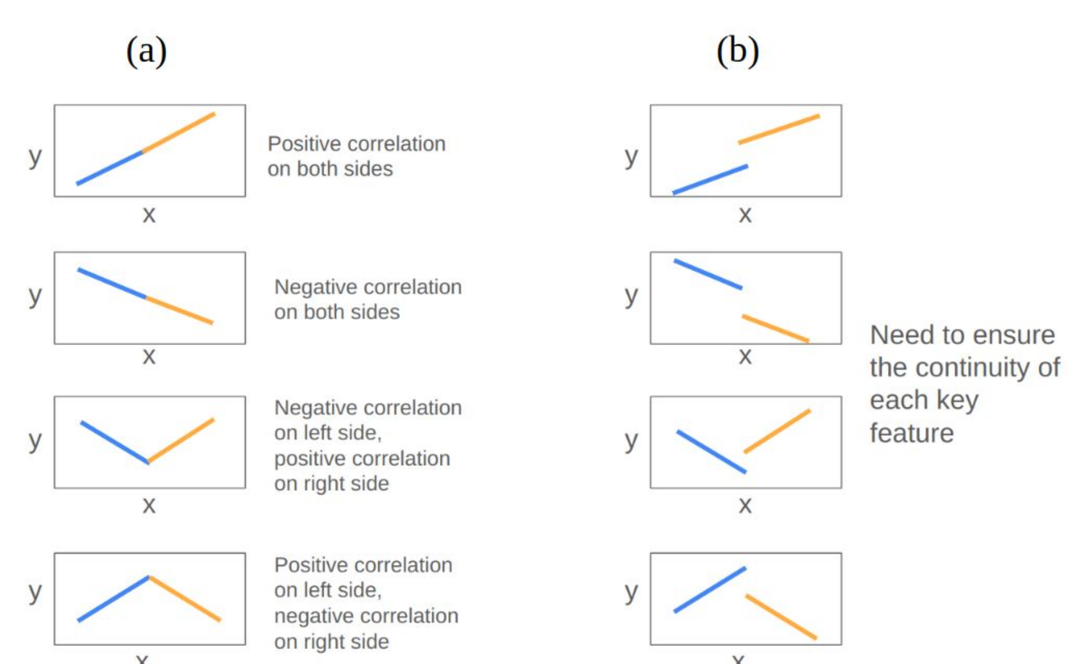
minPts = 5: Minimum number of neighbors to form a cluster.
ε: Distance threshold for neighborhood, chosen via elbow method.

Selecting mapper parameters

Number of bins | Overlap percentage



Extracting Transition Features

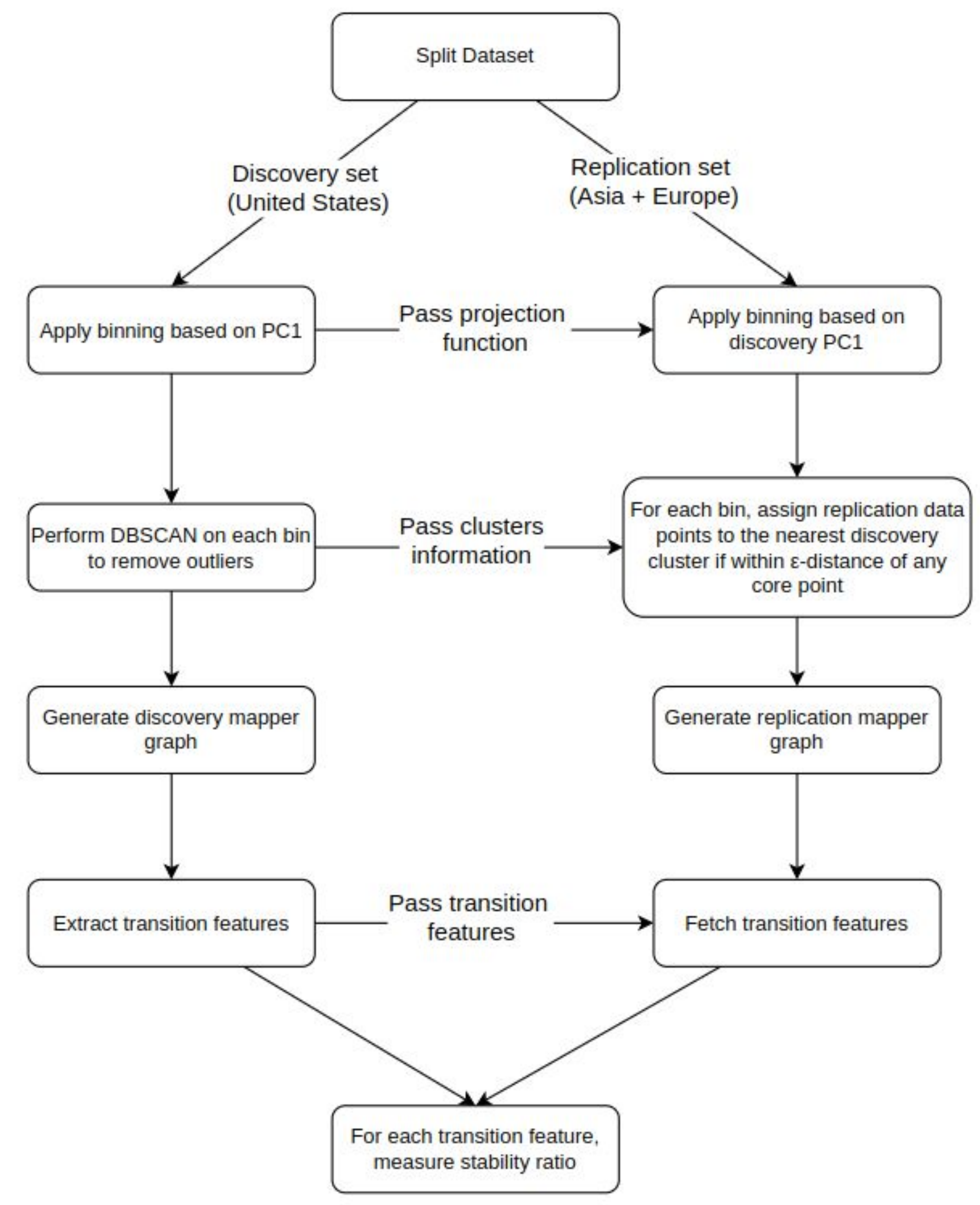


First pattern: values increase with node position
Second pattern: values decrease with node position
Third pattern: values dip at the mixed node.
Fourth pattern: values peak at the mixed node.

A valid transition feature must meet two criteria:

Correlated with position	No difference at the mixed node
It must show significant correlation (|r| > 0.15, p ≤ 0.05) with the first principal component.	It must show no significant difference between ADHD and ASD within the mixed node (t-test p ≥ 0.5).

Validation

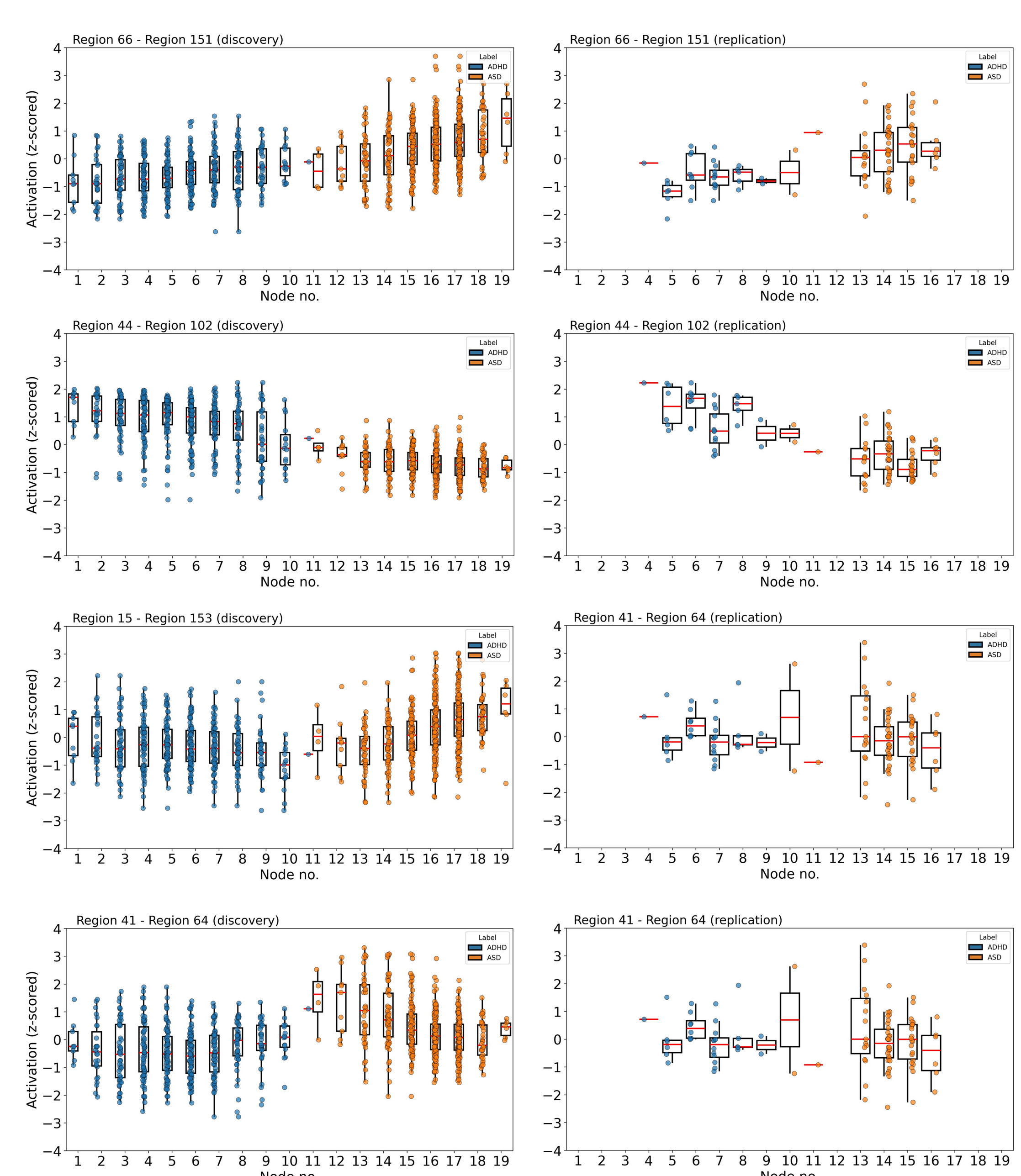


$$Stability(f) = \frac{S_f}{|K_f|}$$

S_f is the number of nodes where the statistical test show that the feature f is consistent between the discovery and replication sets, with p-value greater than 0.05.

$|K_f|$ is the total number of bins where both discovery and replication nodes exist, meaning the feature f can be compared between the two sets.

Results



Pattern	Number of Features	Average Stability Ratio
First pattern	249	0.827
Second pattern	395	0.801
Third pattern	75	0.770
Fourth pattern	52	0.825

Conclusion

- Mapper algorithm with PCA and DBSCAN effectively clustering ADHD and ASD disorders on a spectrum.
- Transitional features of the spectrum can be extracted with some statistical techniques.
- By nature, the mapper algorithm is capable of producing intuitive visualization in displaying the spectrum.