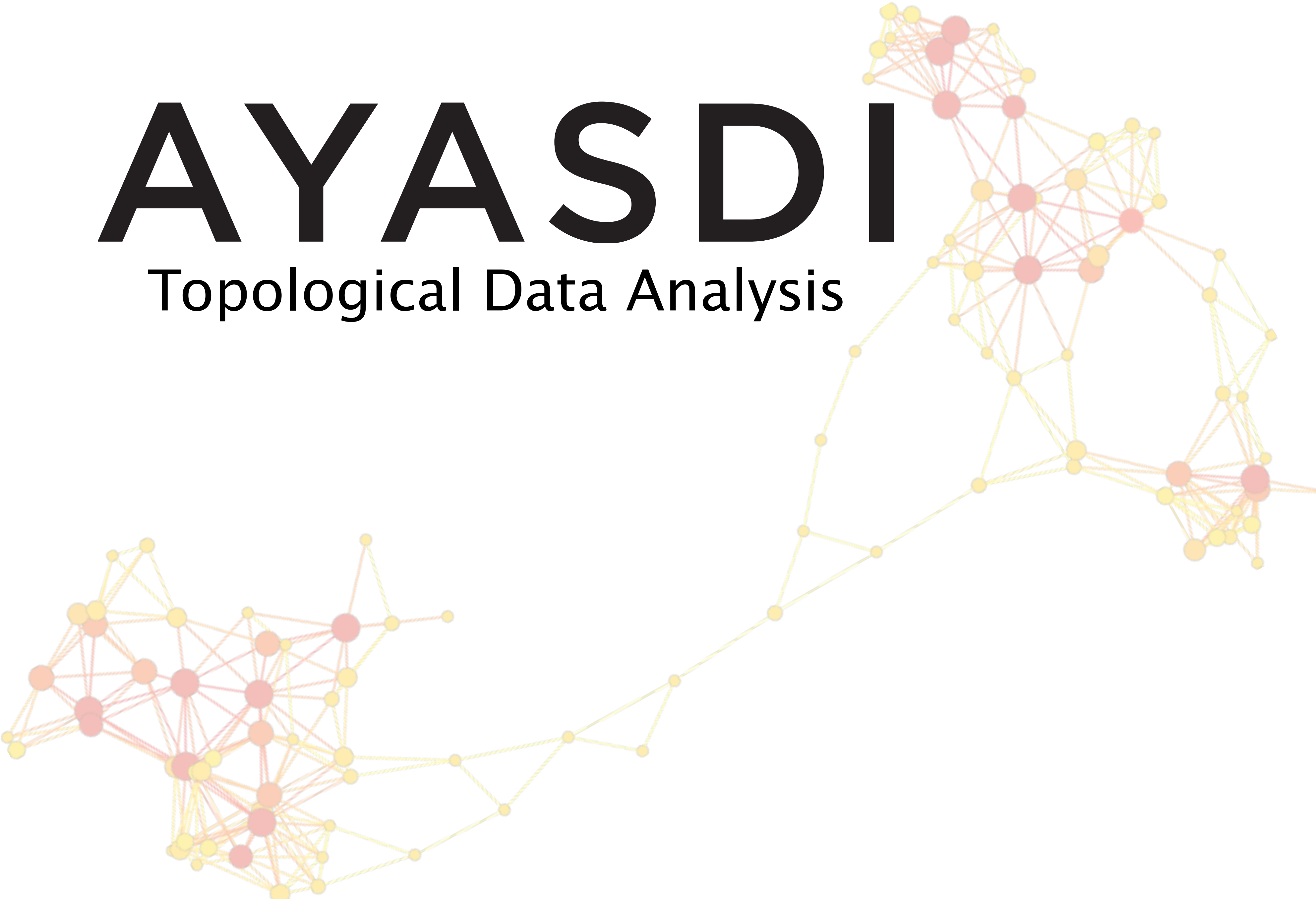
AYASDI

Topological Data Analysis



Data has **Shape**, Shape has **Meaning**, Meaning drives **Value**

Data Basics

AYASDI

Data Basics

Q: What is the fundamental assumption when working with data?

Data Basics

Q: What is the fundamental assumption when working with data?

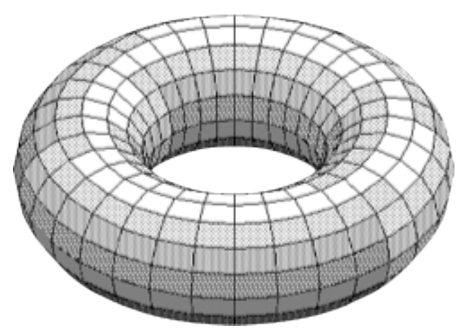
A: Distance between two data points

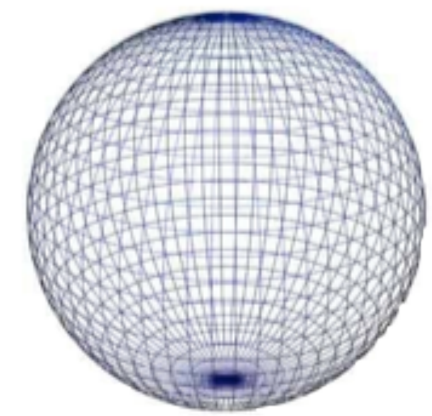
AYASDI

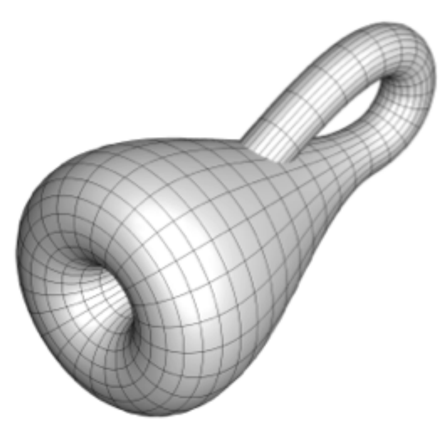
Why Topology?

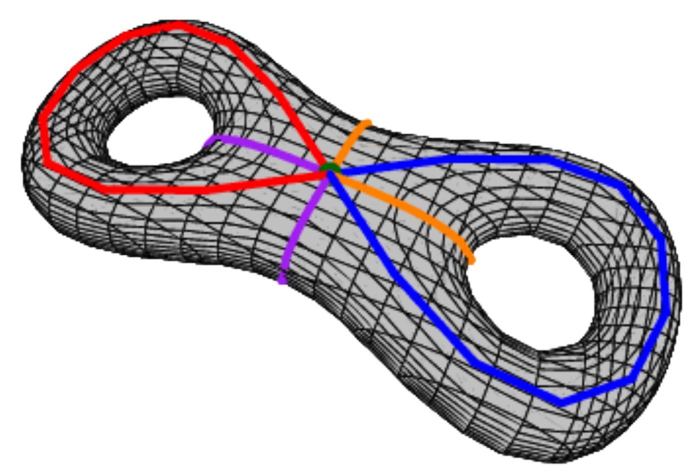
Topology

The branch of mathematics concerned with characterizing the geometric properties of shape







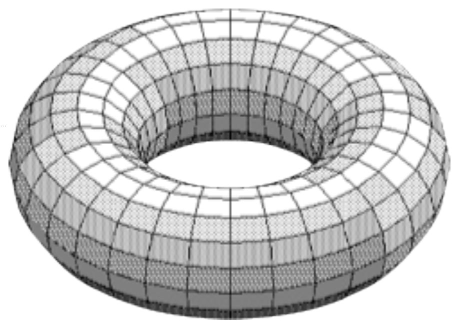


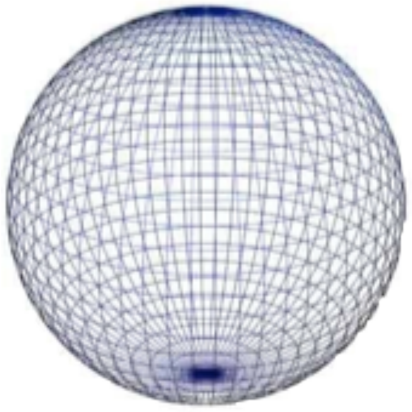
AYASDI

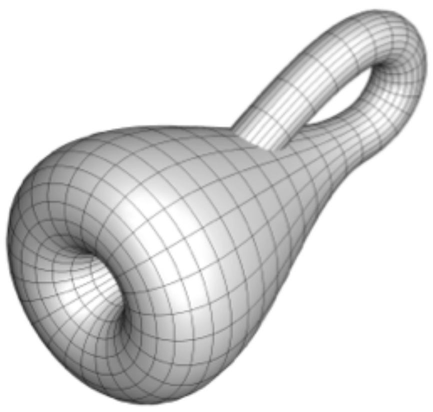
Why Topology?

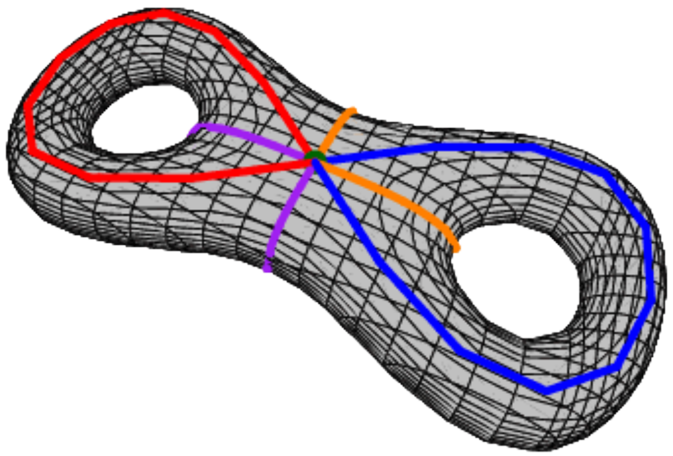
Topology
The branch of mathematics concerned with characterizing the geometric properties of shape

- Coordinate Invariant
- Deformation Invariant
- Compressed Representations









AYASDI

Drop the Assumptions

Let the data tell you what information it holds

Traditional Analytics

- Assumes linearity or normal distributions
- Assume a model of behavior
- Low Dimensionality

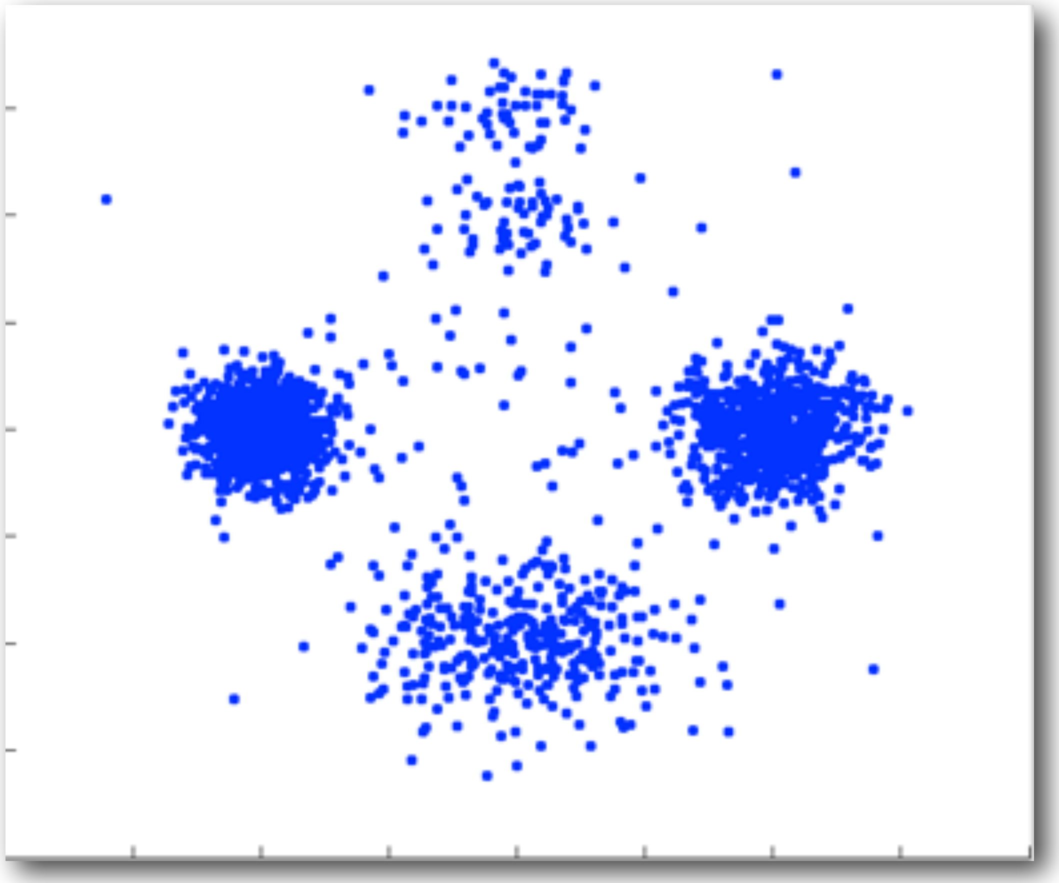
Topological Data Analysis

- Assumes a measure of similarity
- Evaluate and correct models
- High Dimensionality

AYASDI

Topological Data Analysis

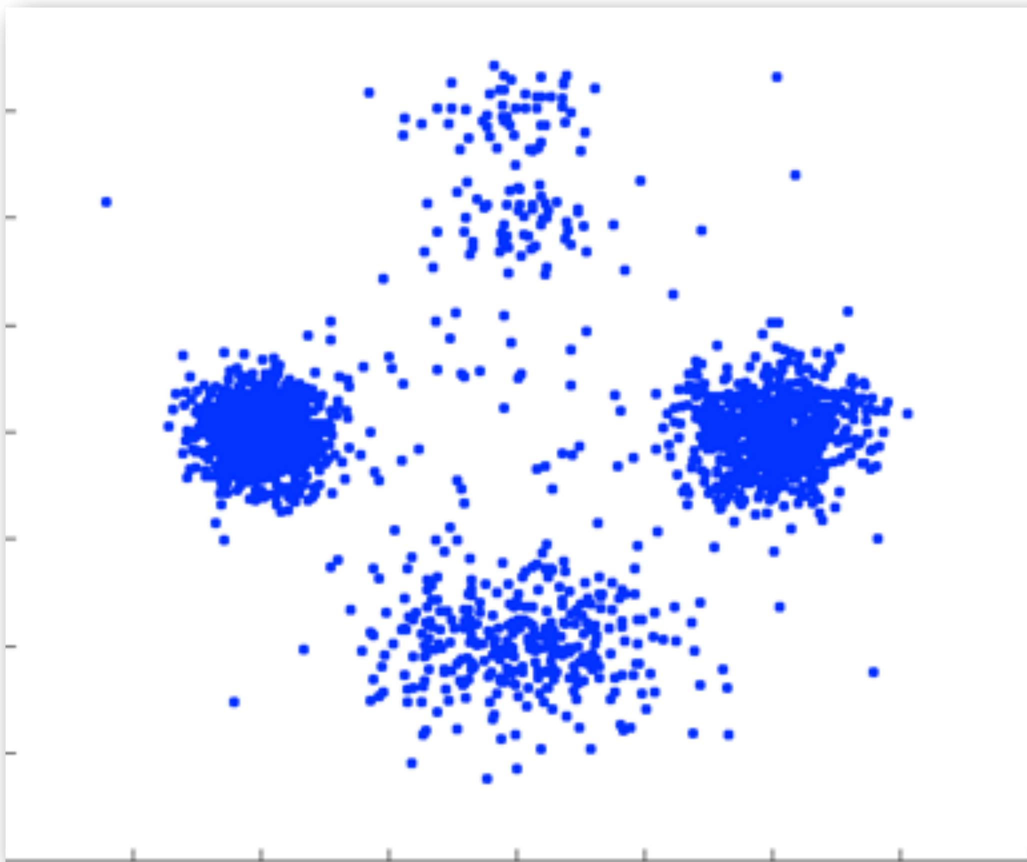
Using local information to gain global knowledge



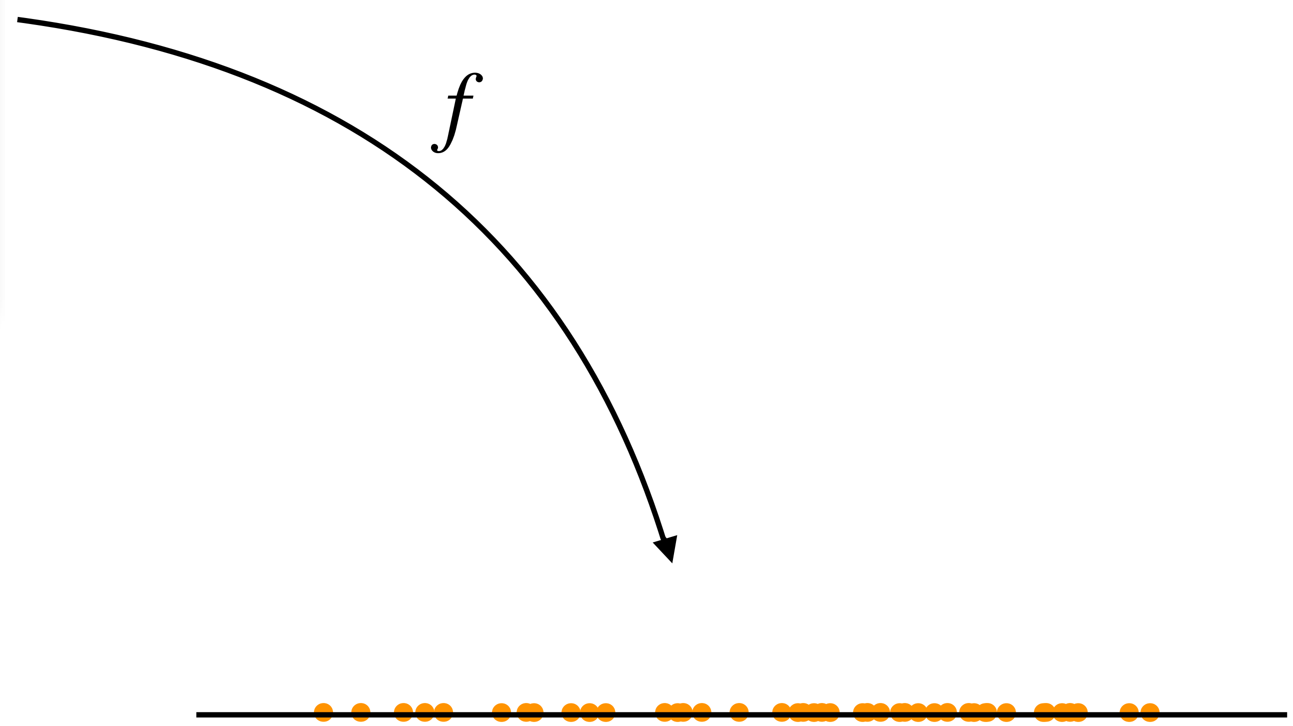
AYASDI

Topological Data Analysis

Using local information to gain global knowledge



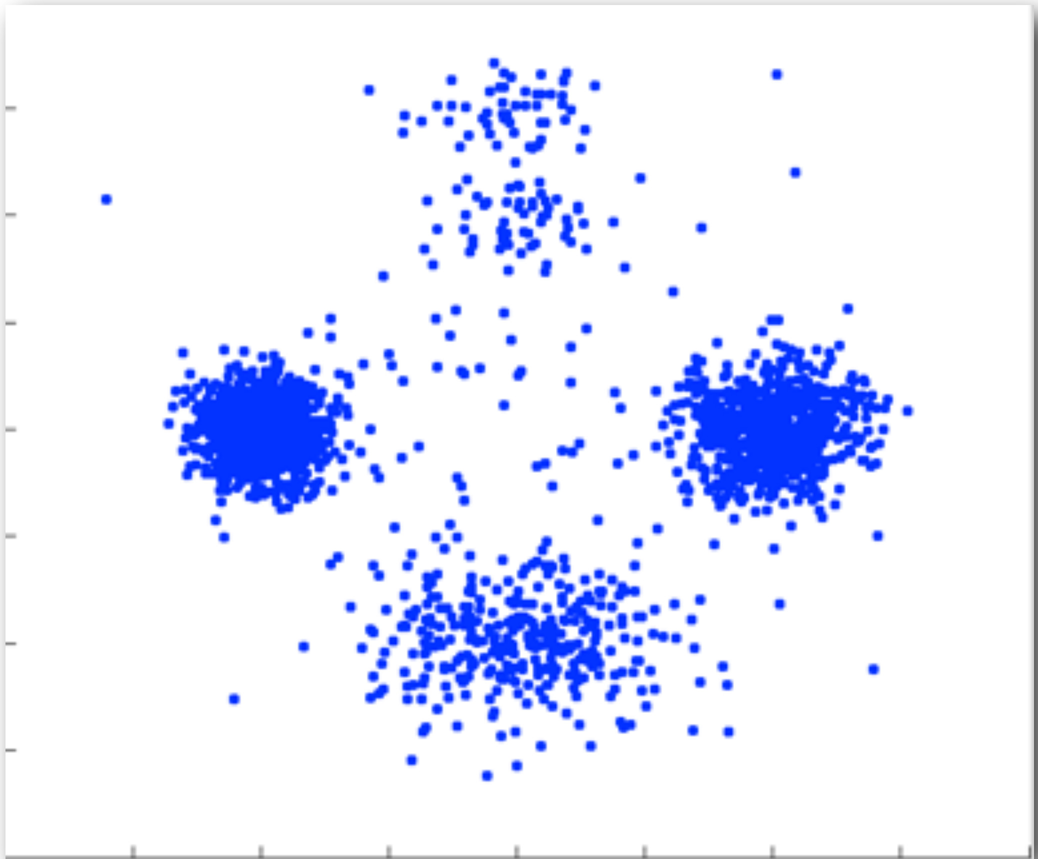
f is a function from the data to some other space (e.g. the real line)



AYASDI

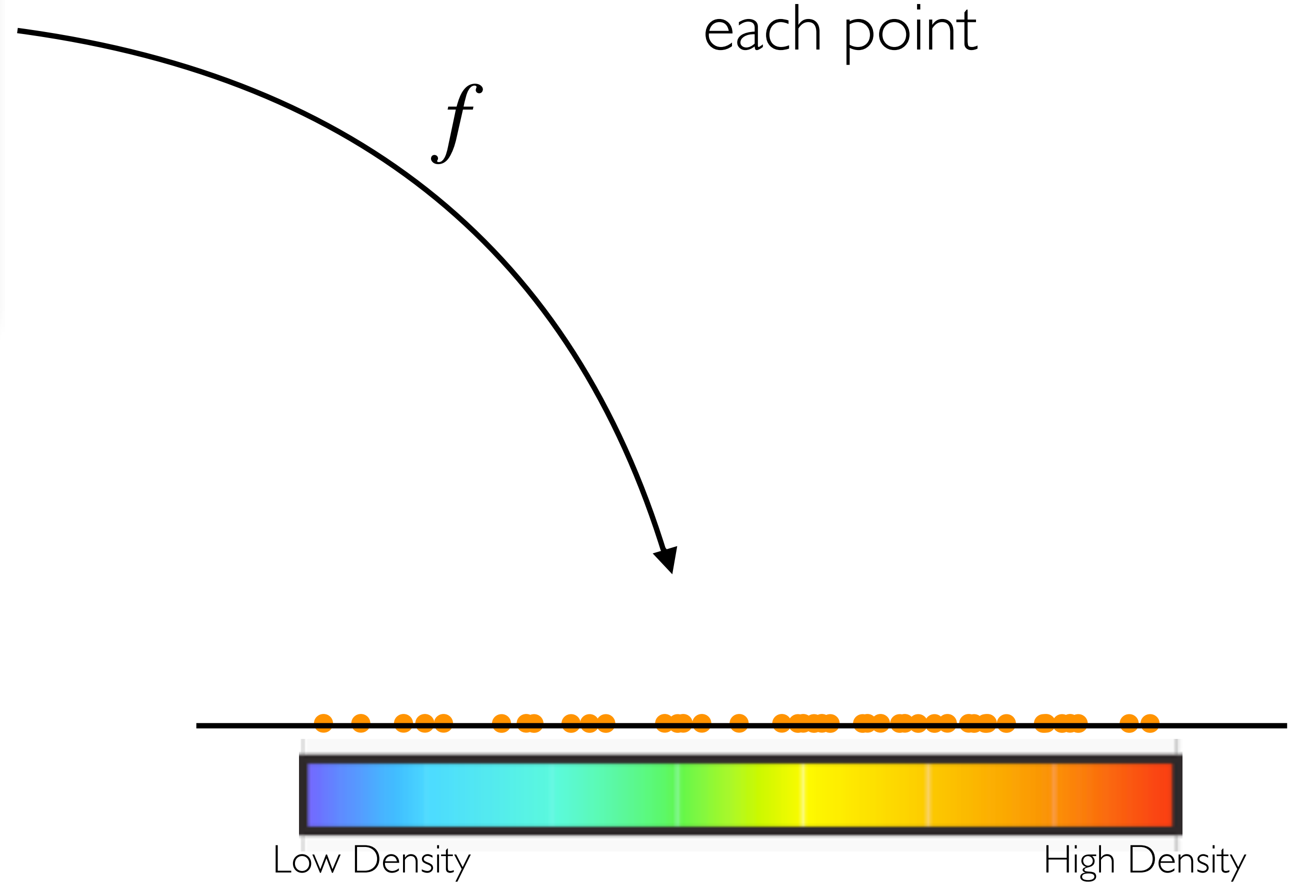
Topological Data Analysis

Using local information to gain global knowledge



f is a function from the data to some other space (e.g. the real line)

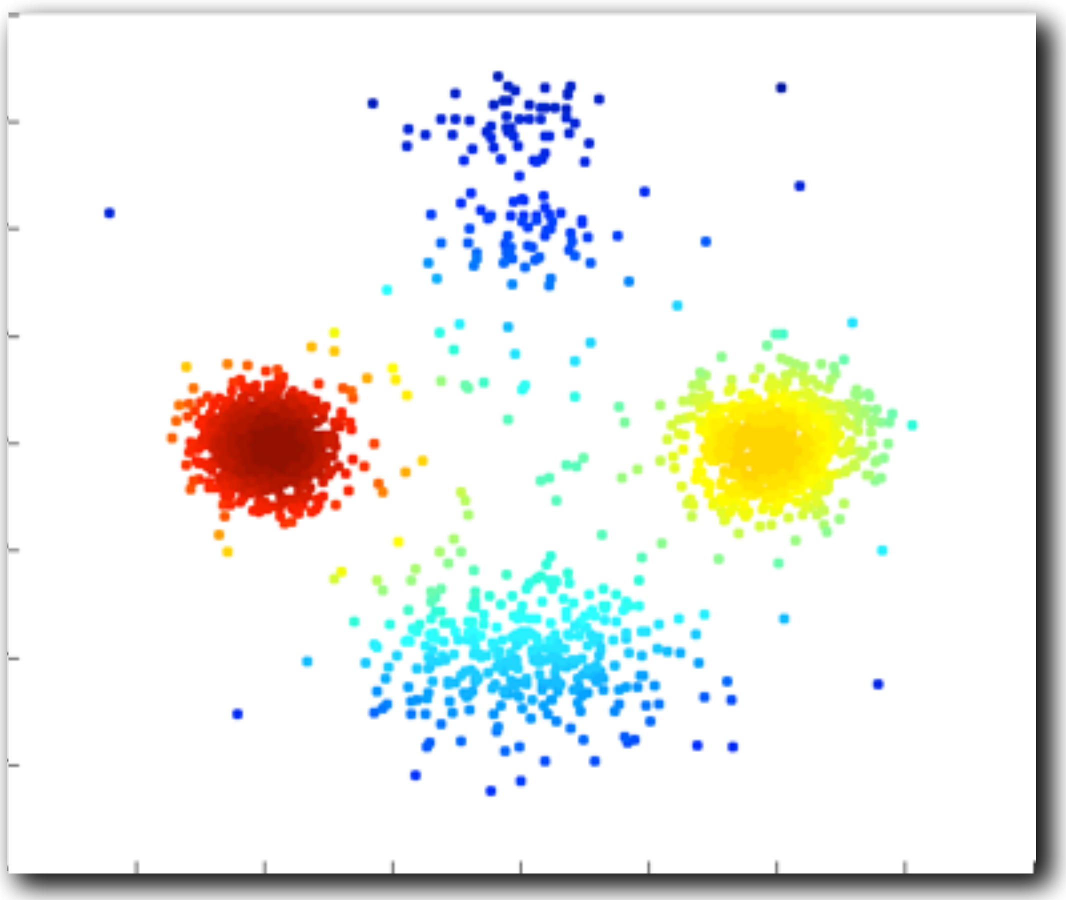
In this example, f is a density estimator at each point



AYASDI

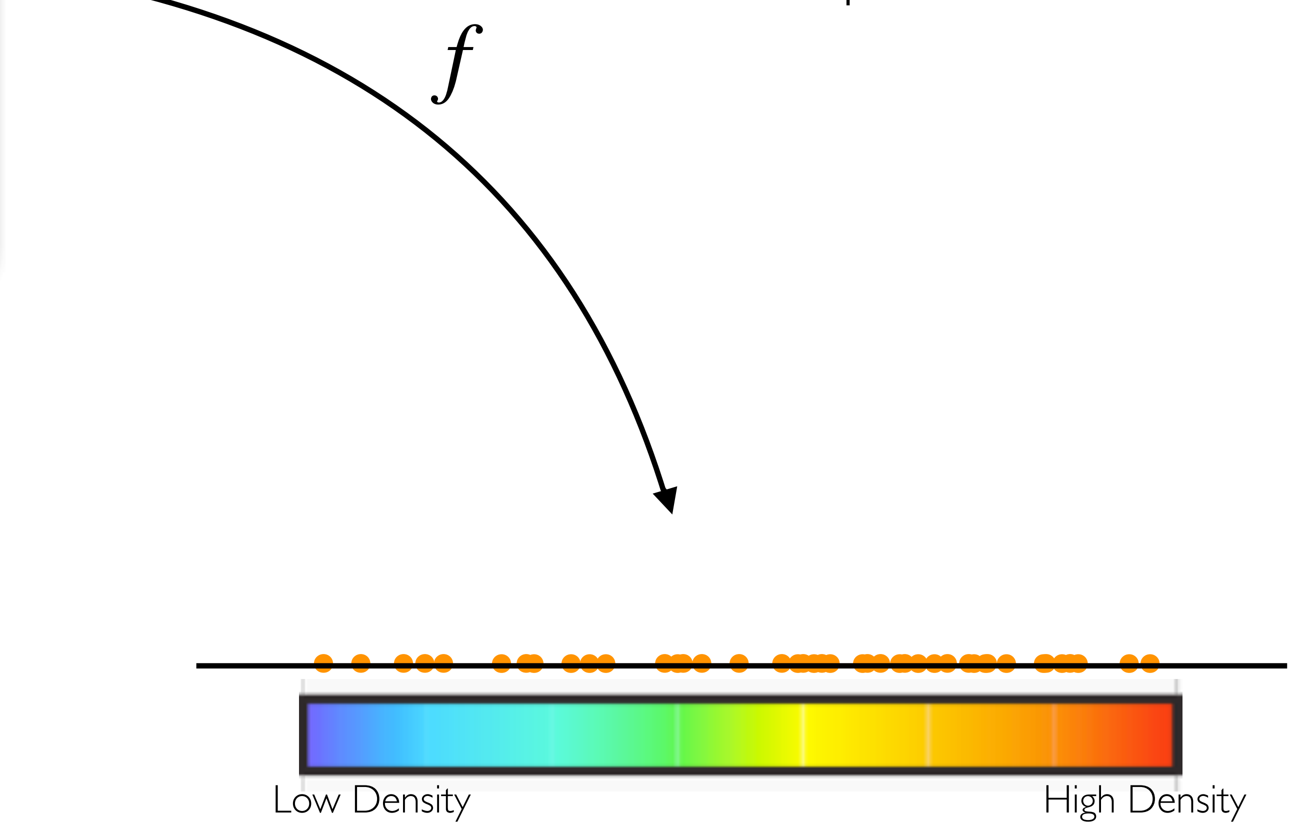
Topological Data Analysis

Using local information to gain global knowledge



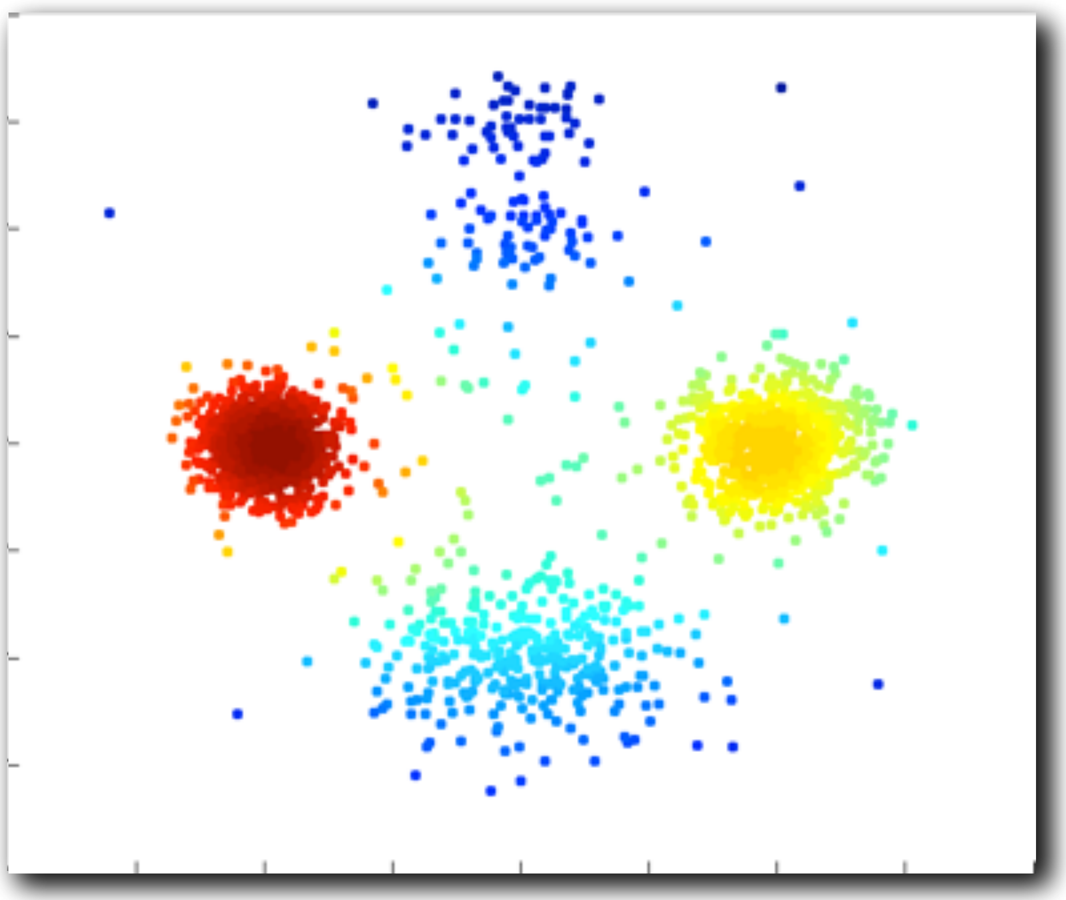
f is a function from the data to some other space (e.g. the real line)

In this example, f is a density estimator at each point

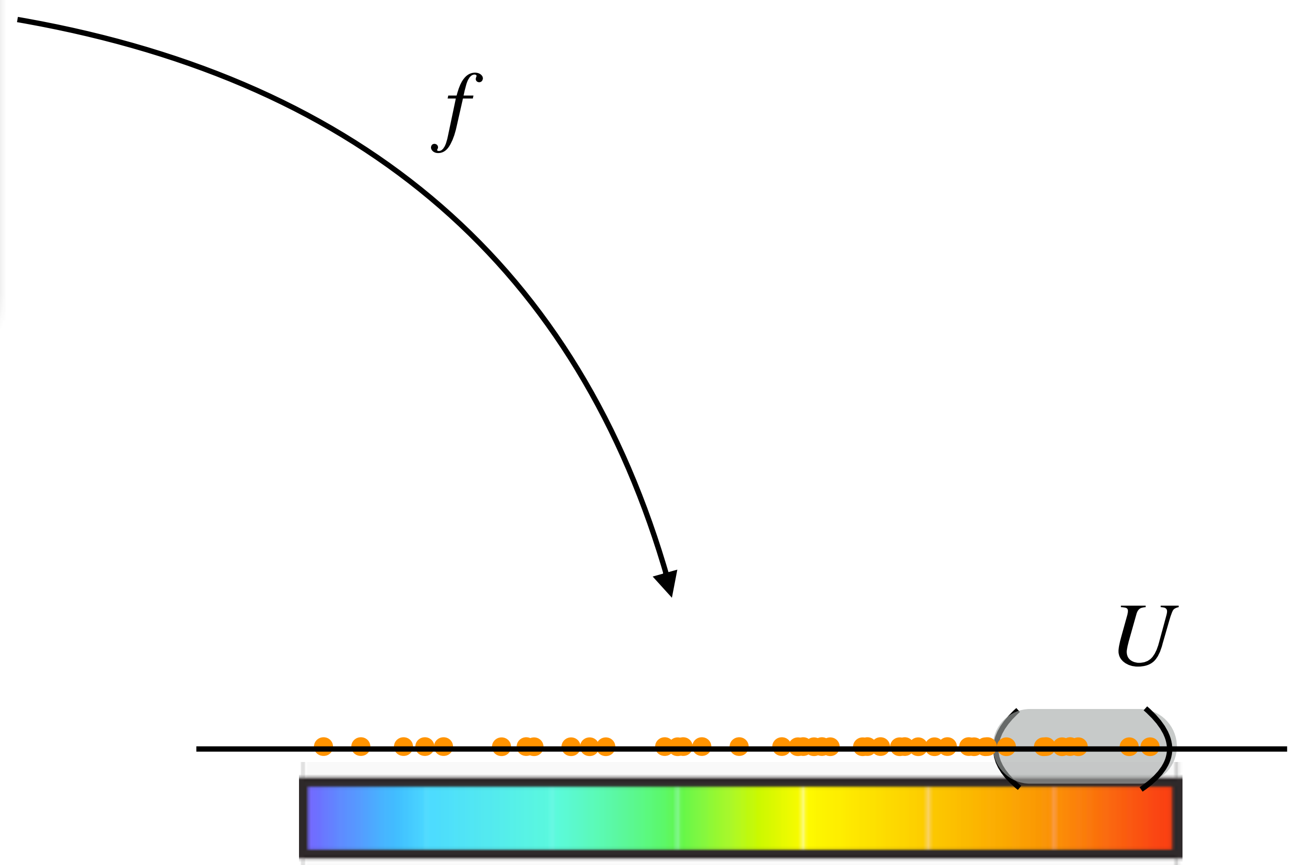


Topological Data Analysis

Using local information to gain global knowledge



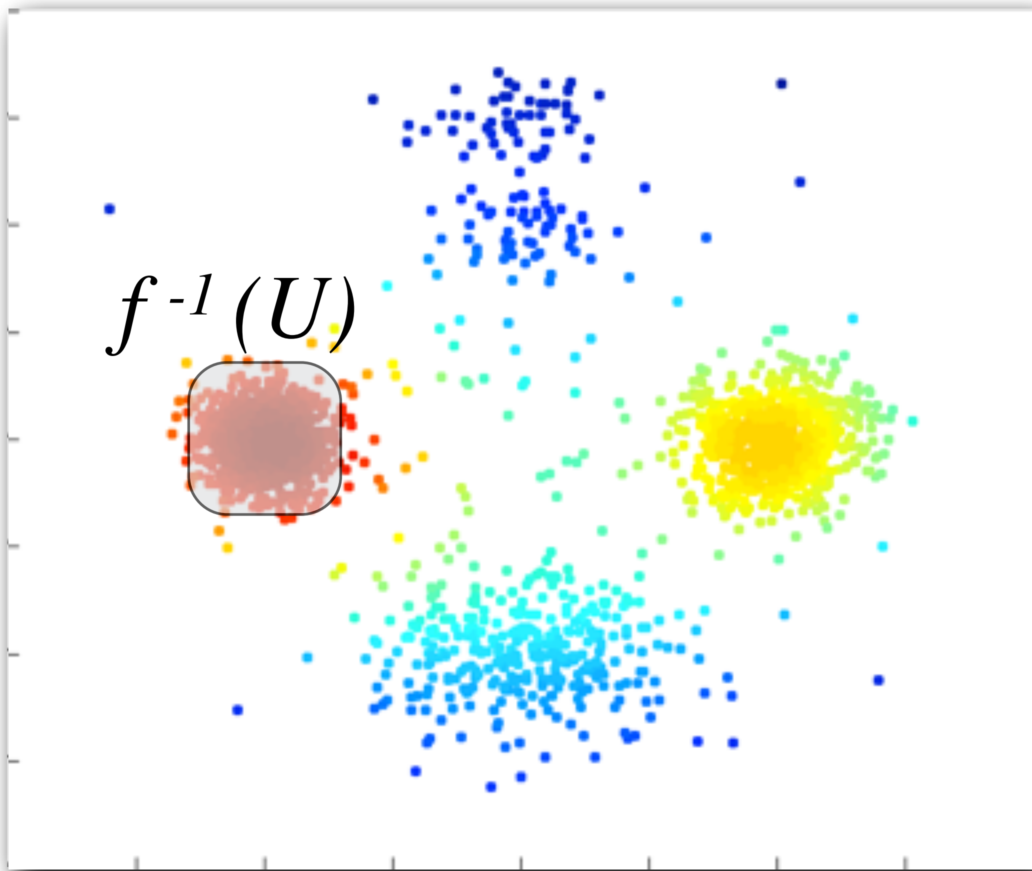
U defines a set of similar points in the image of f



AYASDI

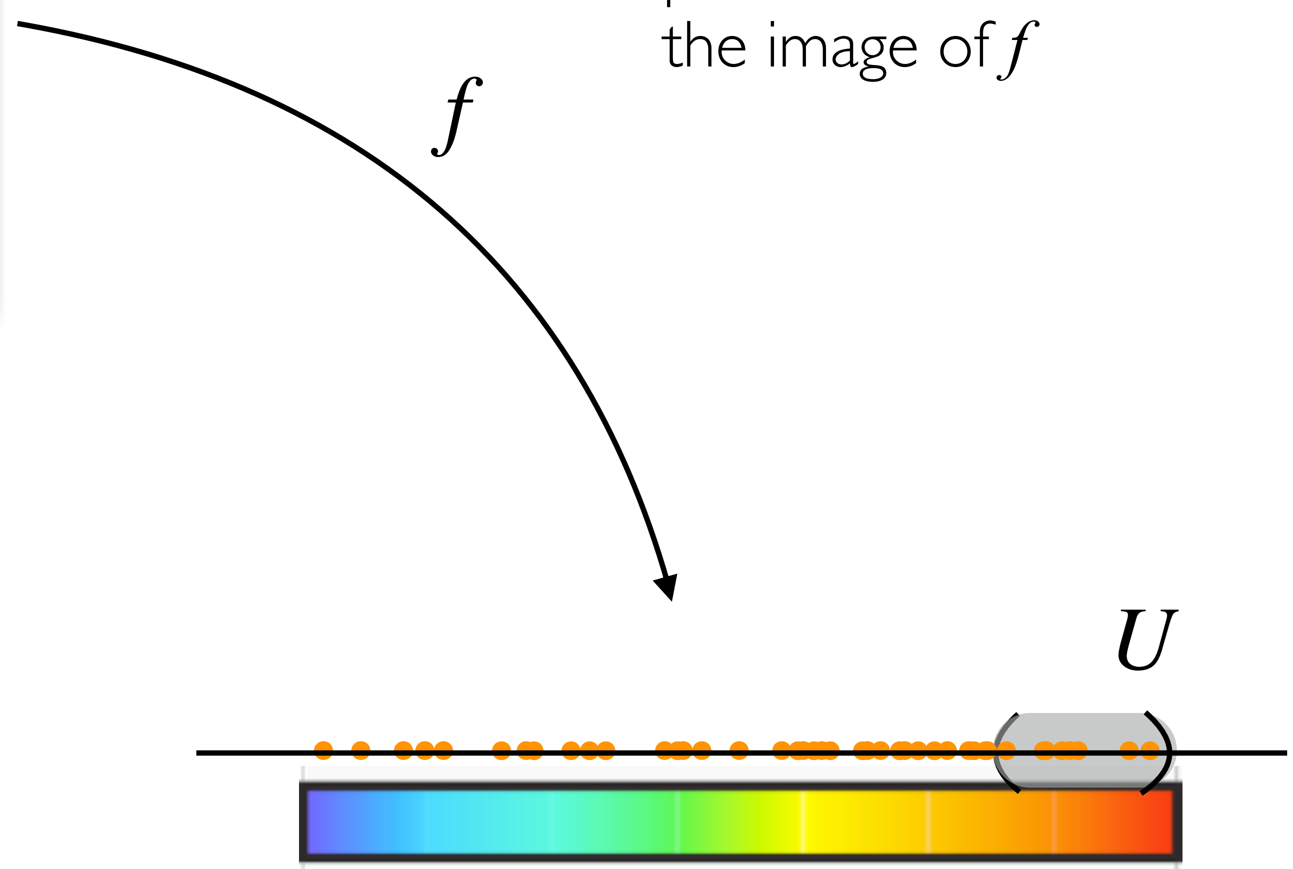
Topological Data Analysis

Using local information to gain global knowledge



U defines a set of similar points in the image of f

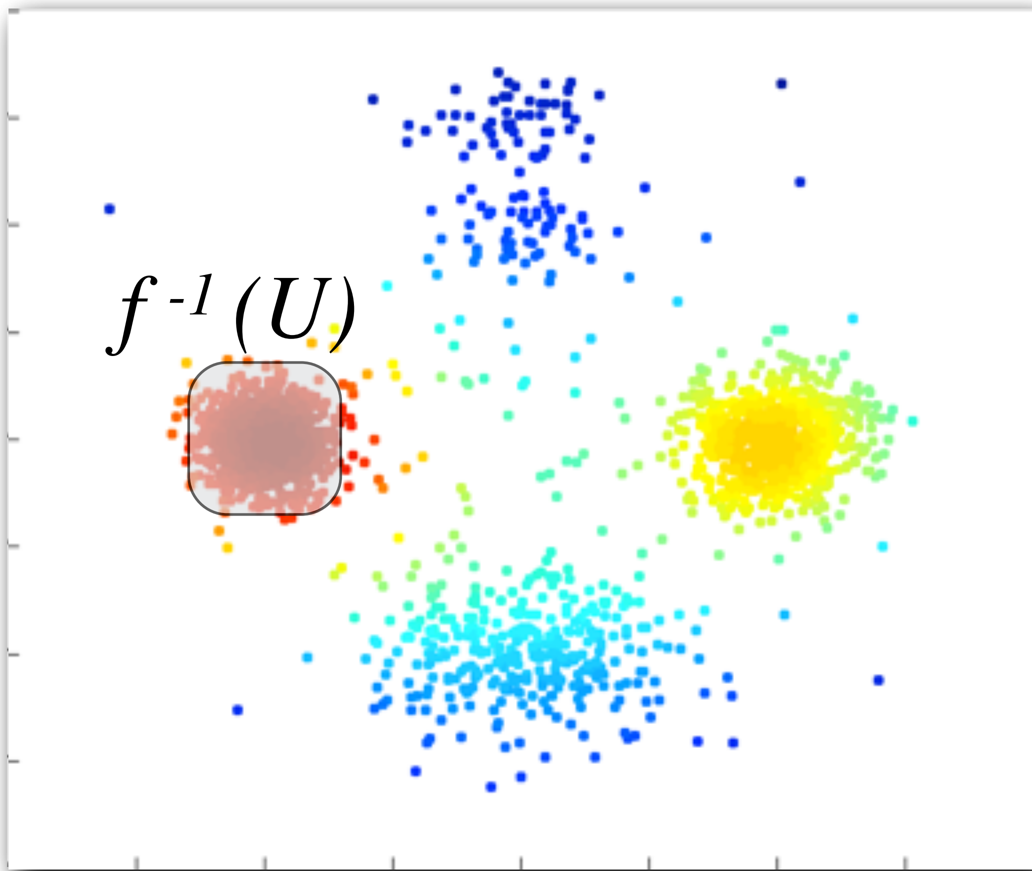
$f^{-1}(U)$ is a set of data points that are similar in the image of f



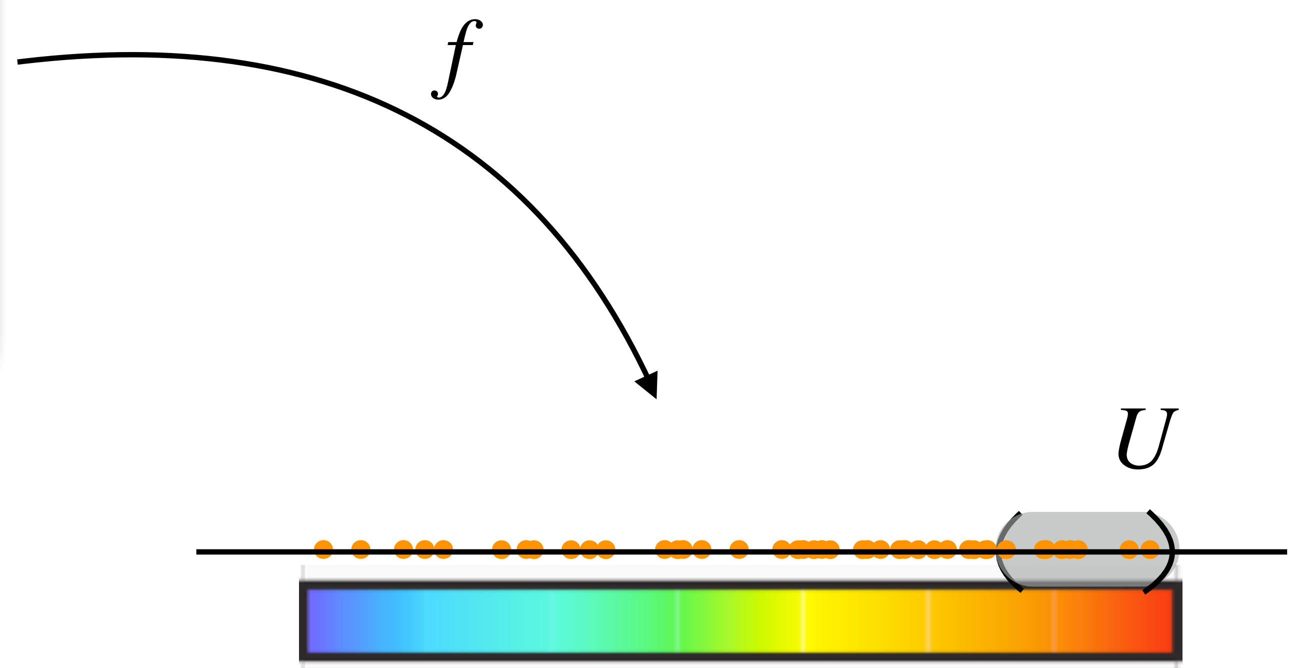
AYASDI

Topological Data Analysis

Using local information to gain global knowledge



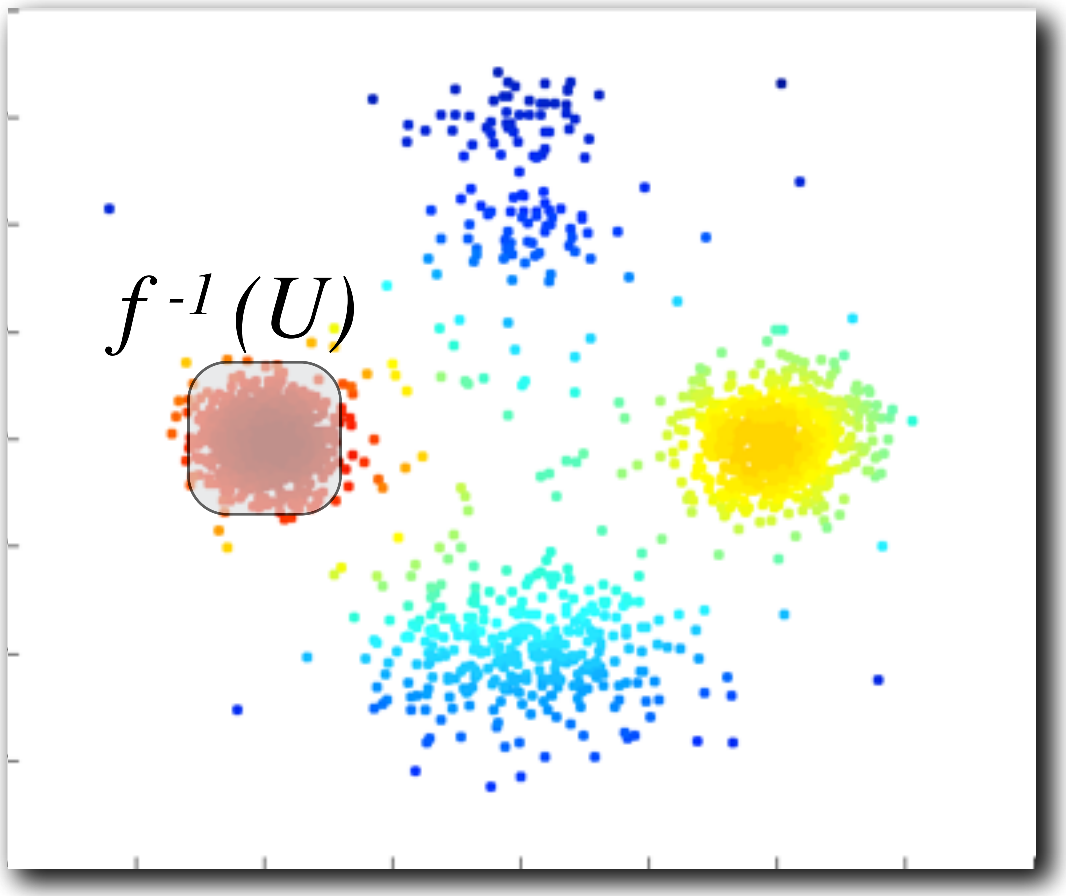
Using the metric, perform clustering to determine the sets of similar points in $f^{-1}(U)$



AYASDI

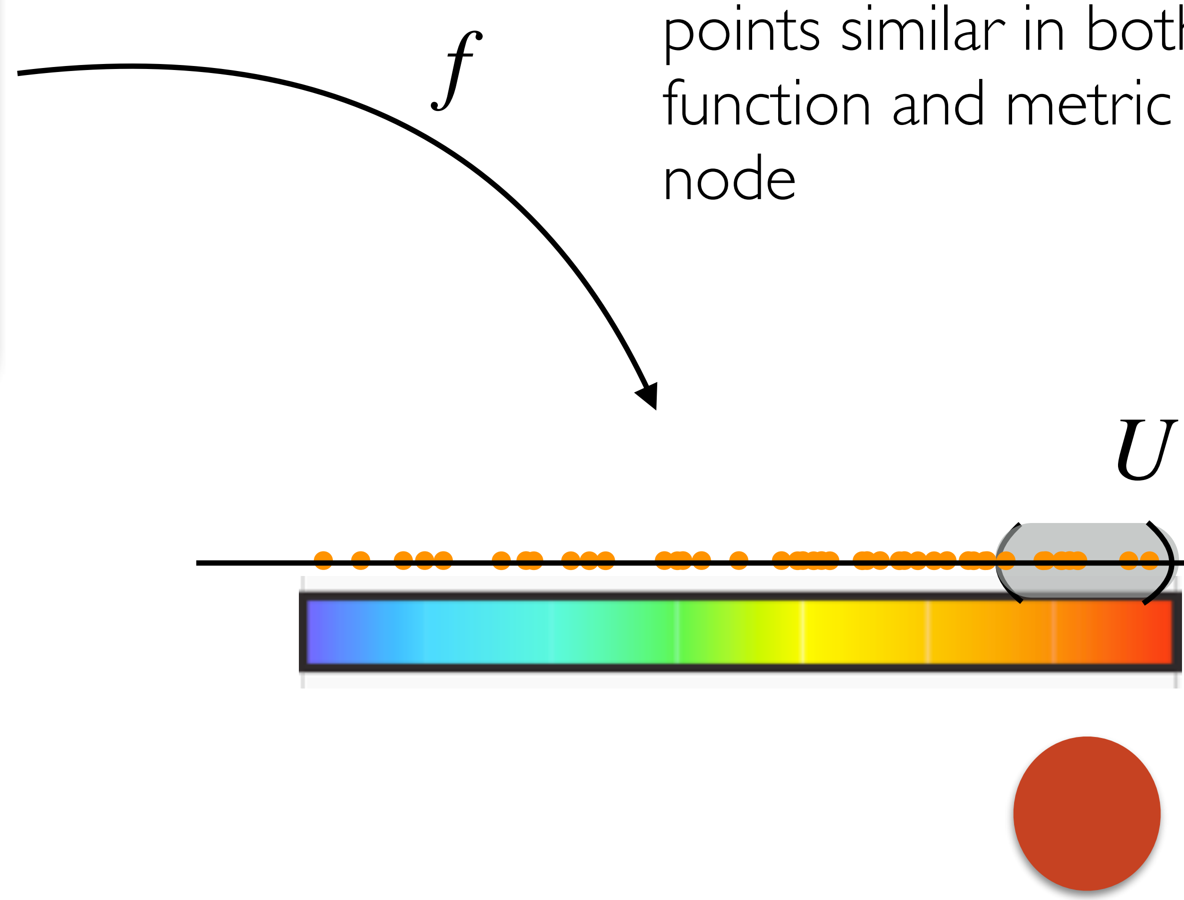
Topological Data Analysis

Using local information to gain global knowledge



Using the metric, perform clustering to determine the sets of similar points in $f^{-1}(U)$

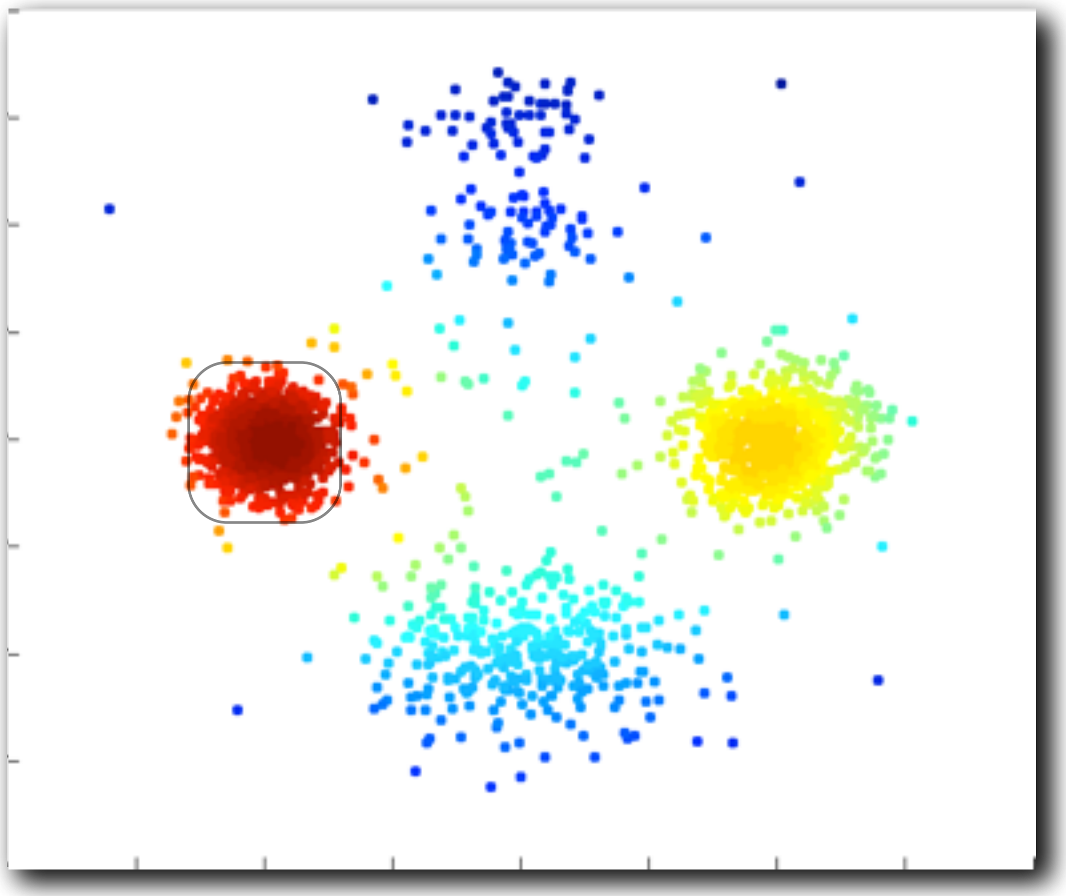
Represent each set of points similar in both function and metric as node



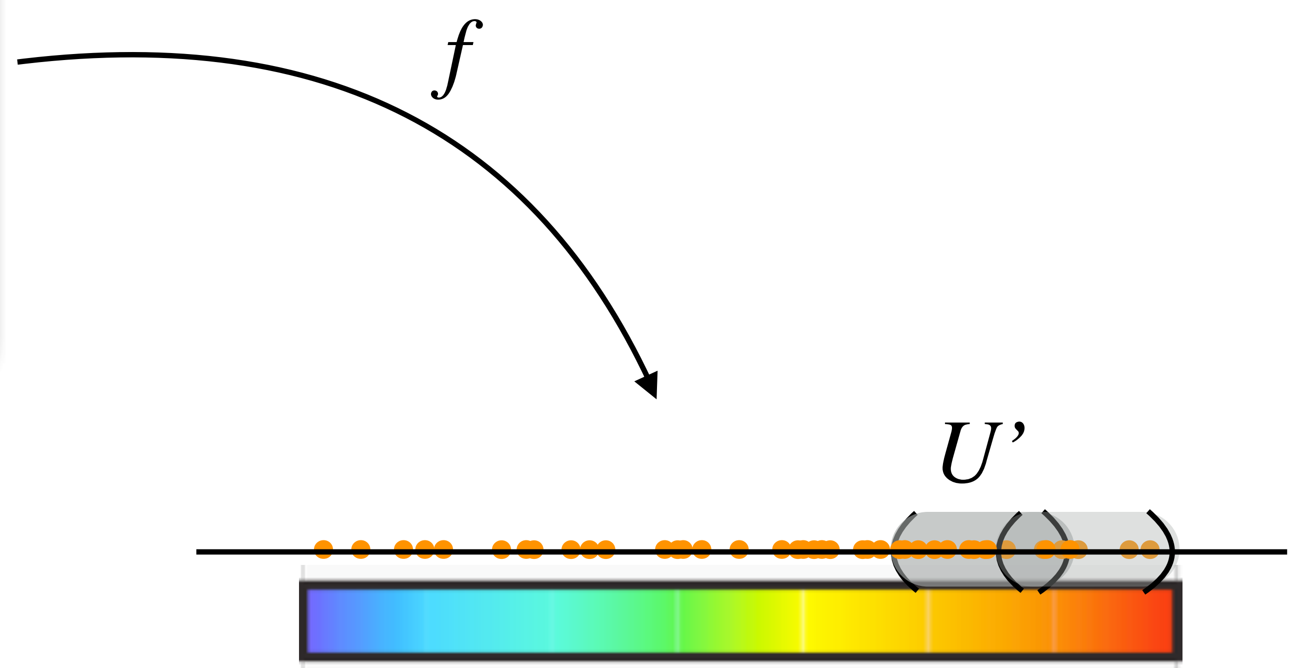
AYASDI

Topological Data Analysis

Using local information to gain global knowledge



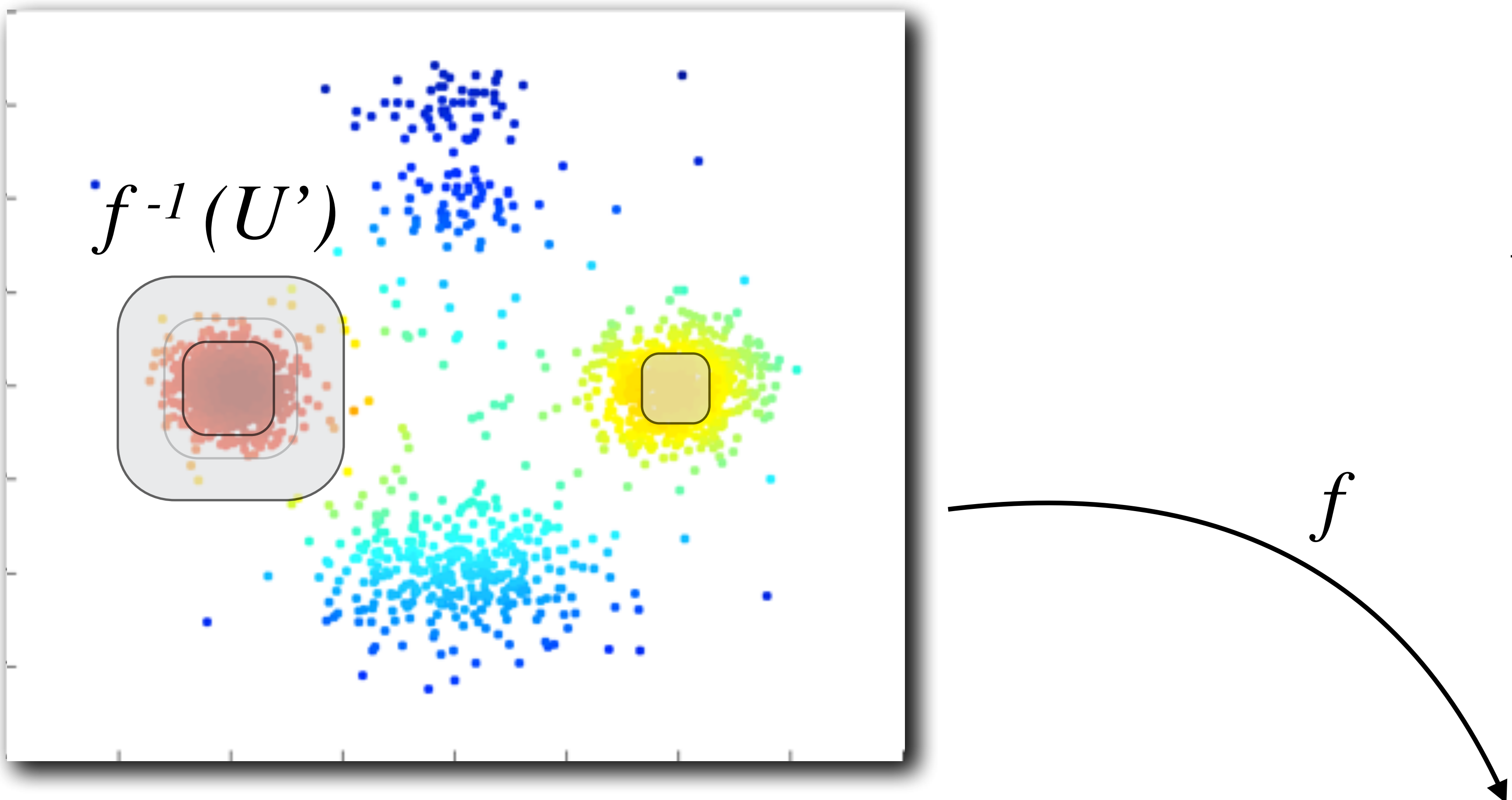
Repeat process with a different set of similar points in the image of the function



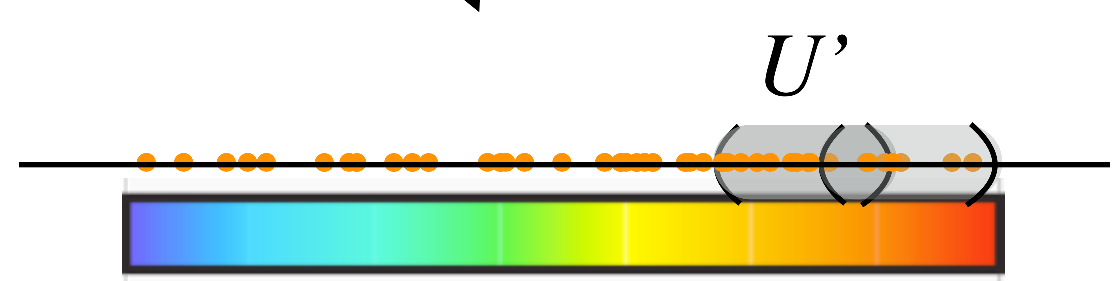
AYASDI

Topological Data Analysis

Using local information to gain global knowledge

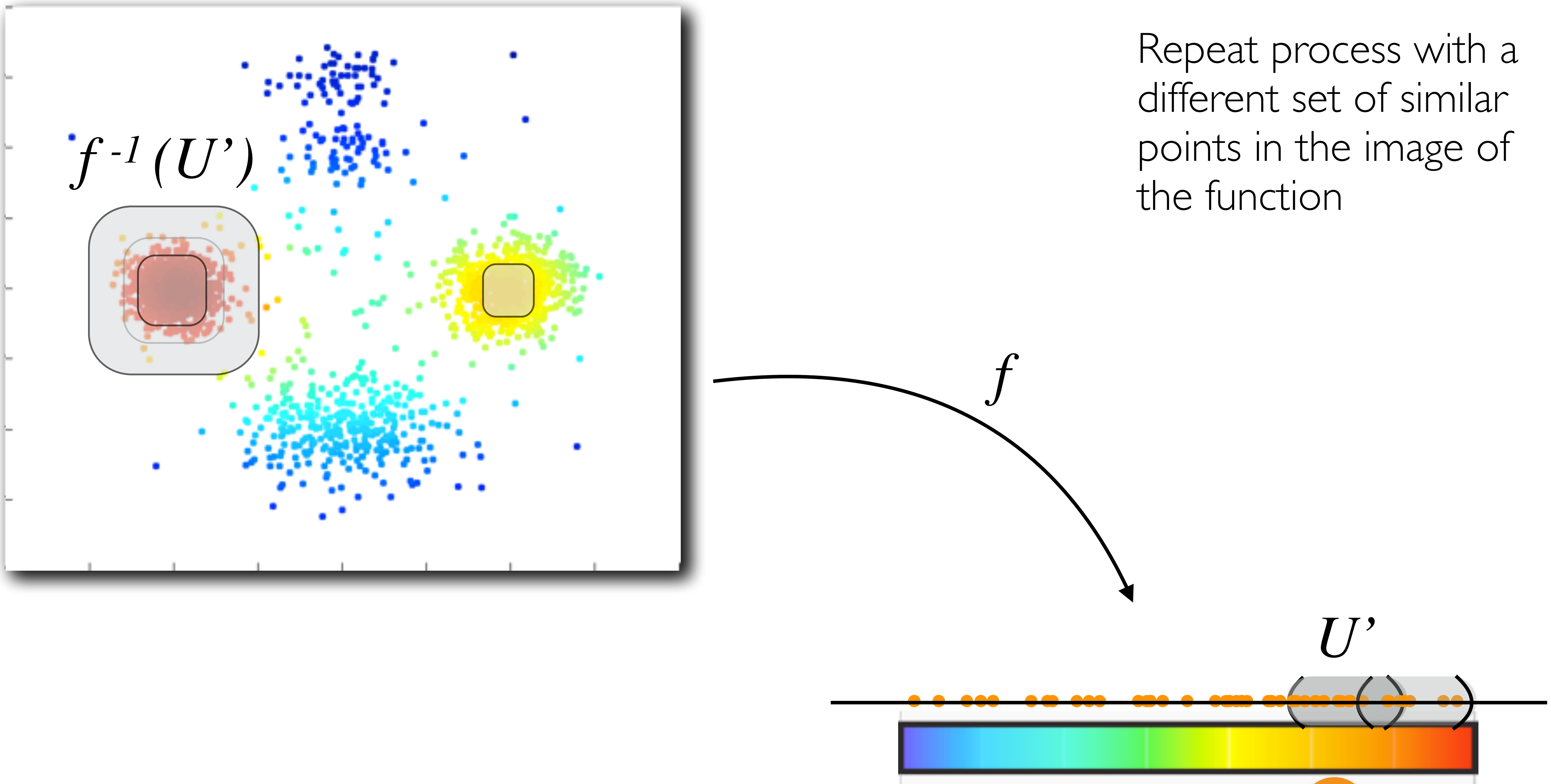


Repeat process with a different set of similar points in the image of the function



Topological Data Analysis

Using local information to gain global knowledge

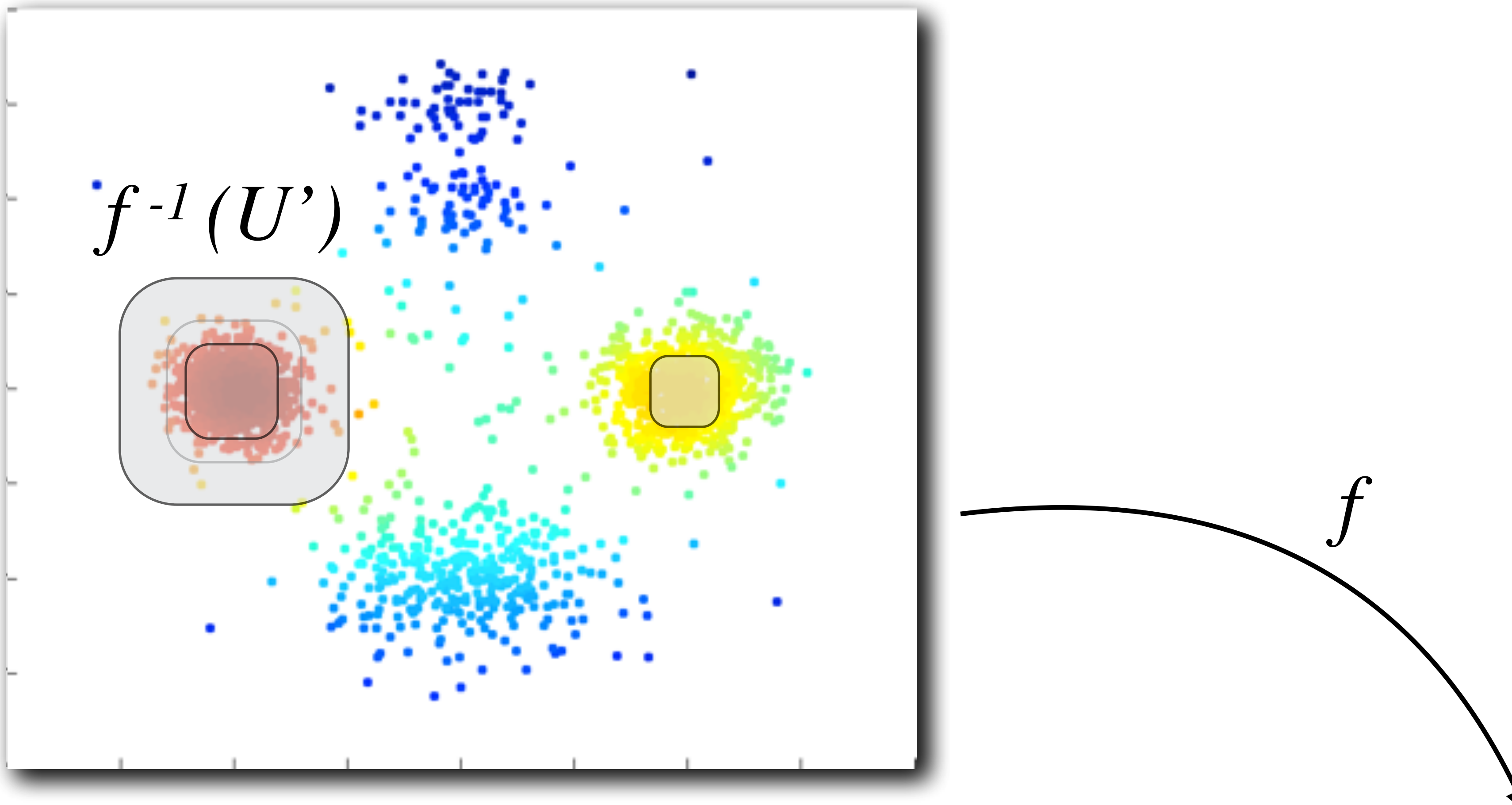


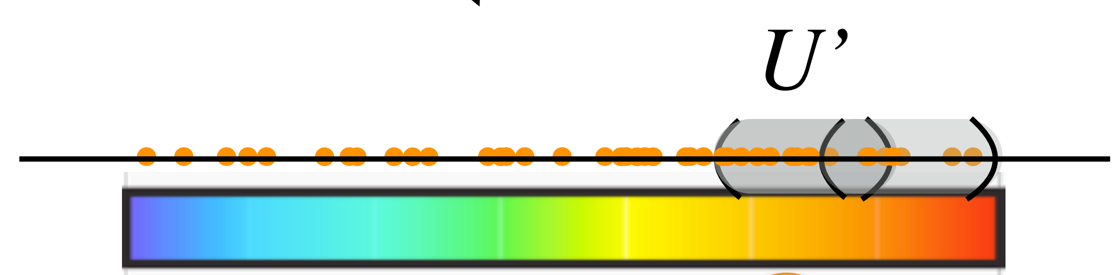
Repeat process with a different set of similar points in the image of the function

AYASDI

Topological Data Analysis

Using local information to gain global knowledge





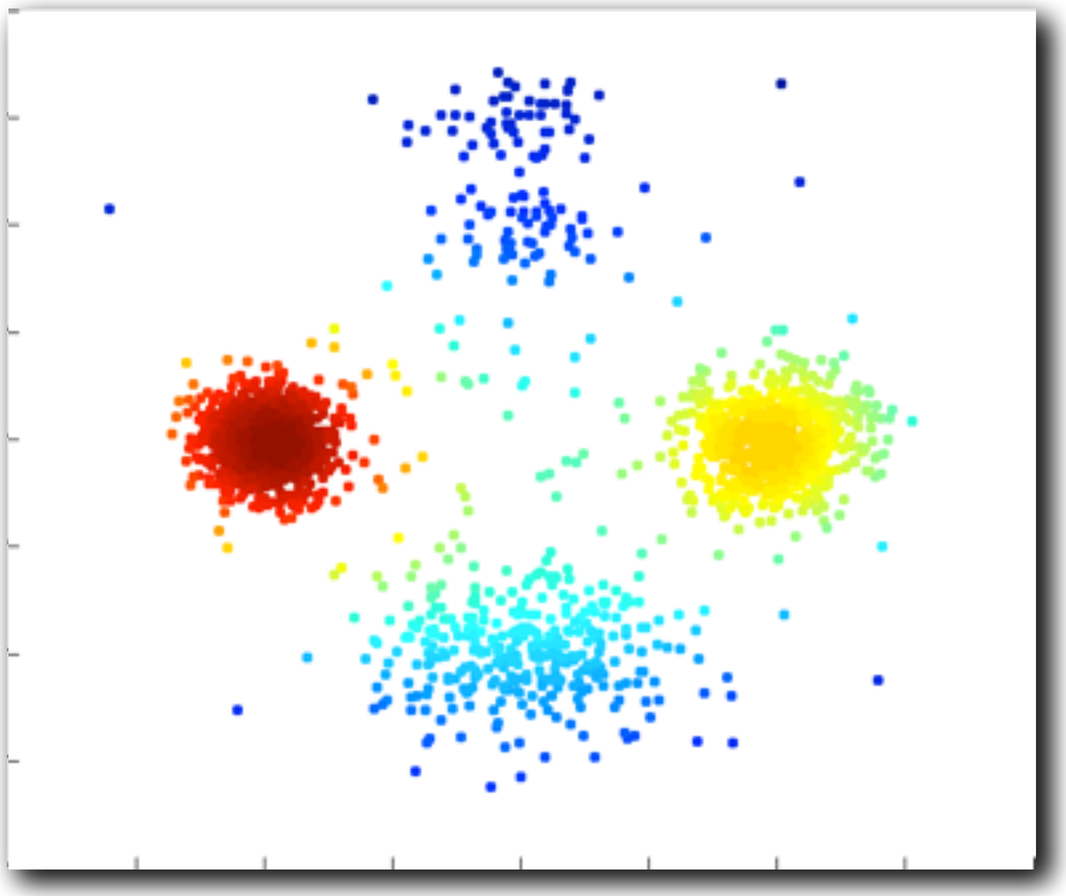
Repeat process with a different set of similar points in the image of the function

Edges between nodes indicate overlapping points. They capture the continuous nature of the data when viewed through the function

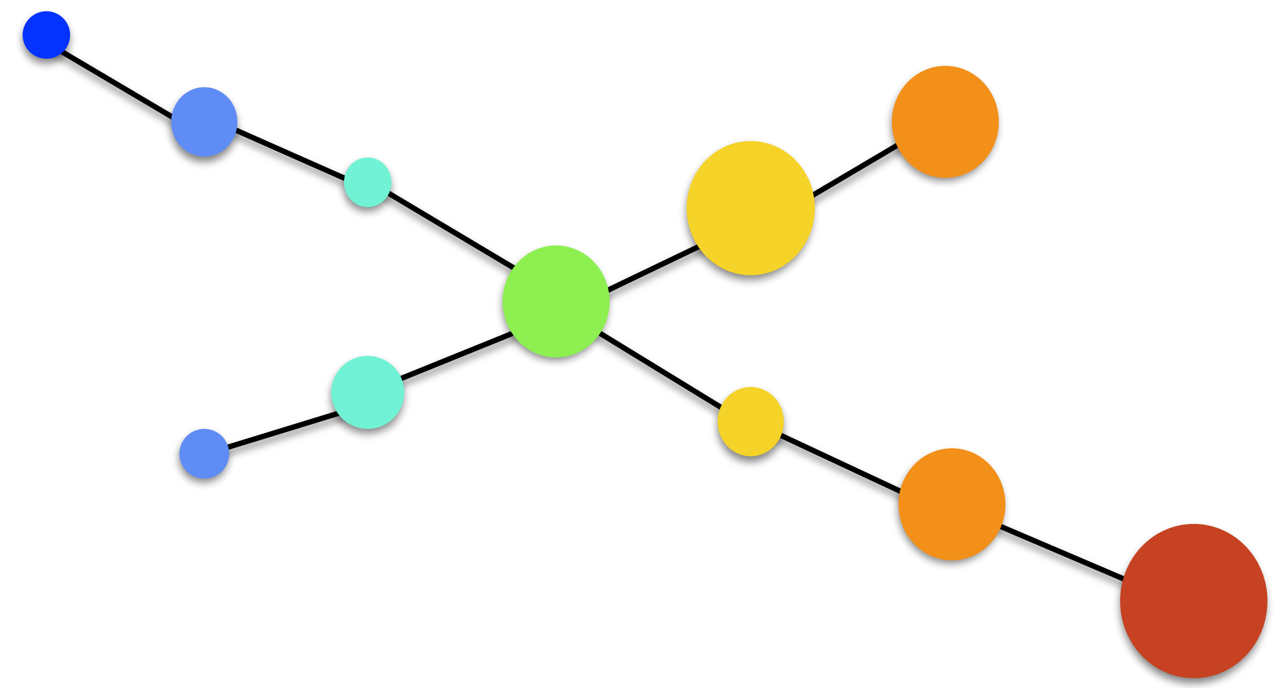
AYASDI

Topological Data Analysis

Powerful geometric summaries of your data



The resulting graph is a geometric summary of the data.



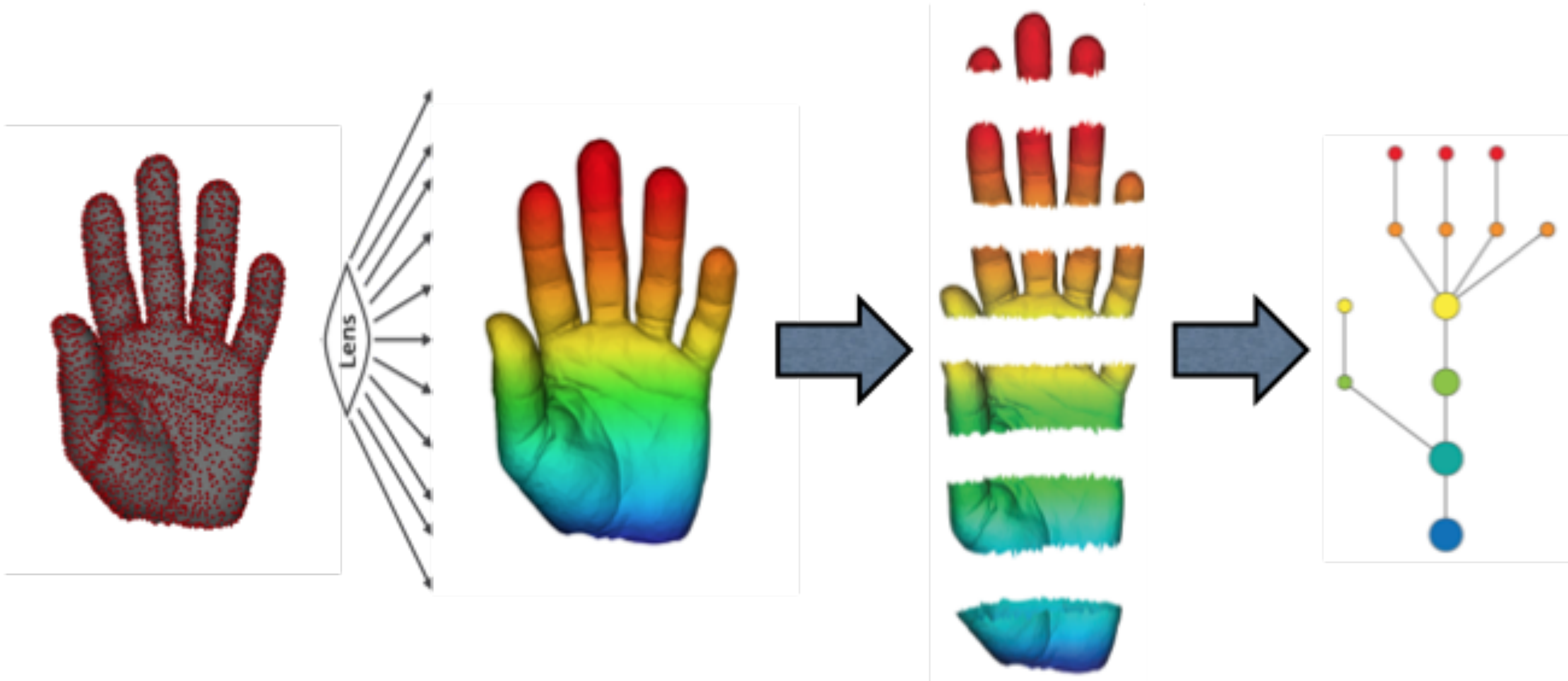
Nodes represent a set of points similar in both function and metric

Edges between nodes indicate overlapping points.

AYASDI

Basic Example

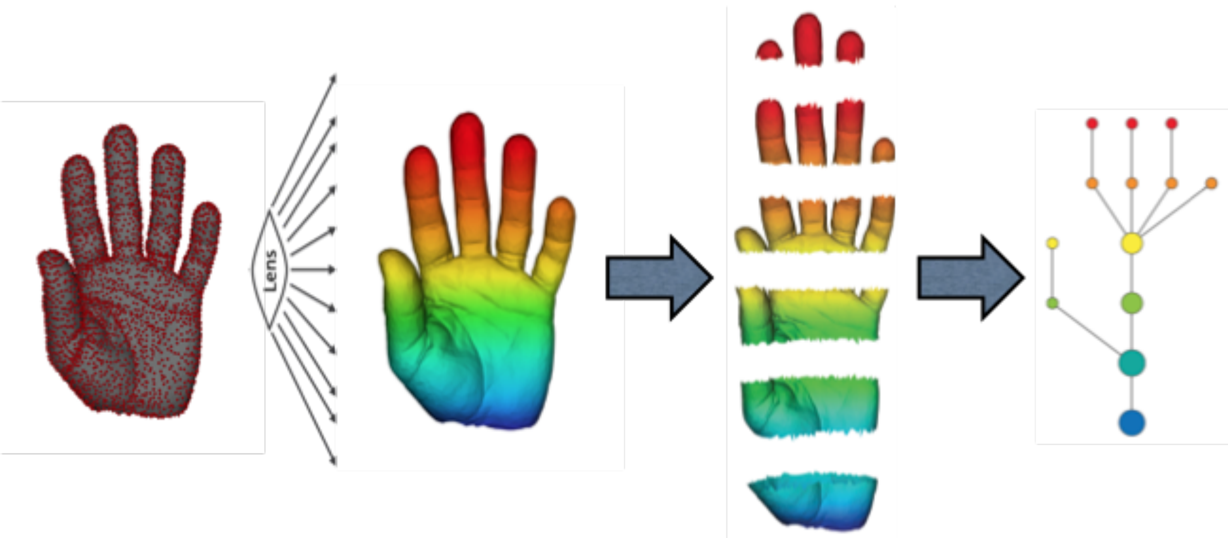
Characterizing shape in 3 dimensions

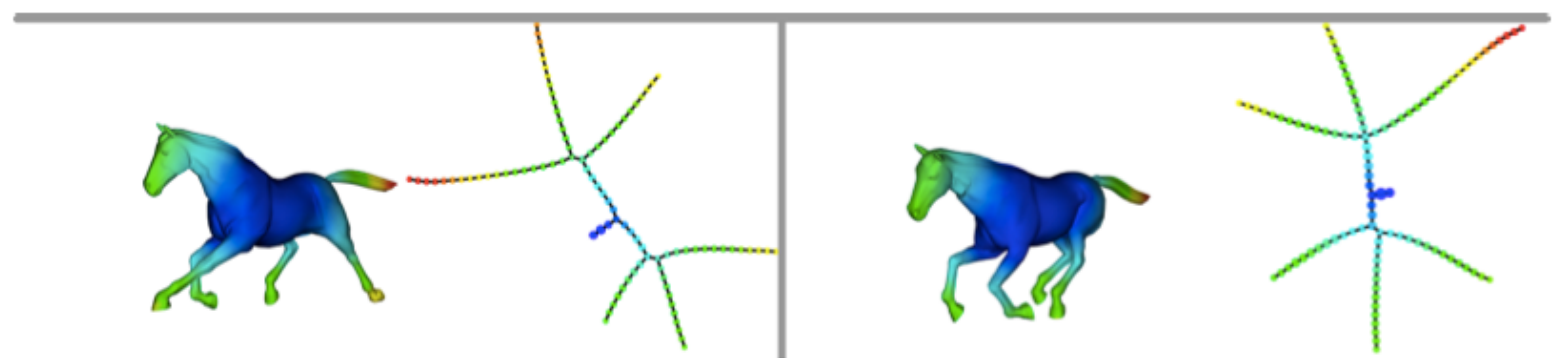


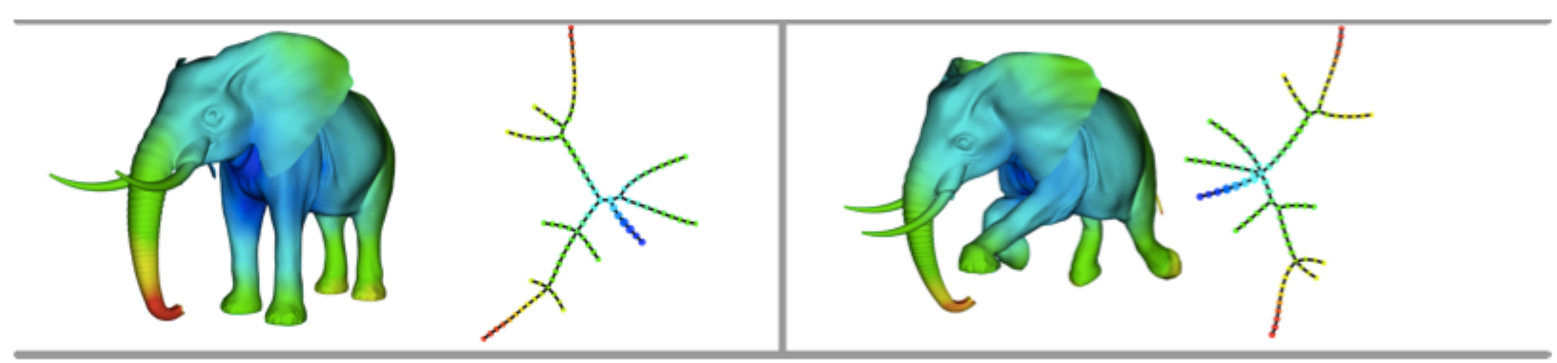
AYASDI

Basic Example

Characterizing shape in 3 dimensions







AYASDI

Machine Learning and TDA

AYASDI

Machine Learning and TDA

Incorporate traditional analytics through the function f

AYASDI

Machine Learning and TDA

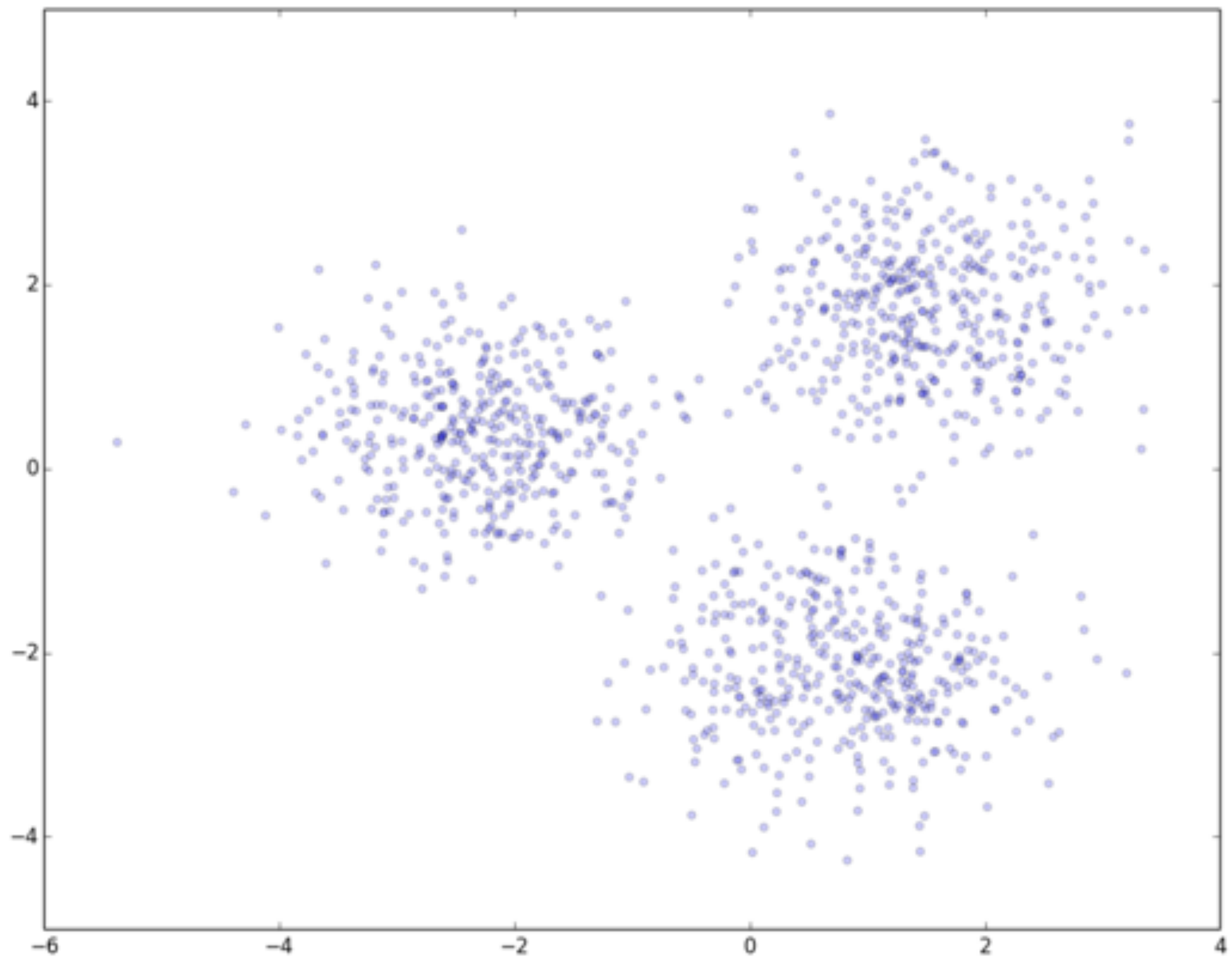
Incorporate traditional analytics through the function f

Statistics	Geometry/ Topology	Machine Learning	Data Driven
Mean/Max/Min/ Variance	Centrality	PCA/SVD	Age
n-Moment	Curvature	Autoencoders	Dates
Density	Harmonic Cycles	Isomap/MDS/TSNE	User Models
…	…	SVM Distance from Hyperplane	
		Error/Debugging Info	

AYASDI

Machine Learning and TDA
Example: PCA

Provides unsupervised dimensionality reduction.
Easy to interpret.

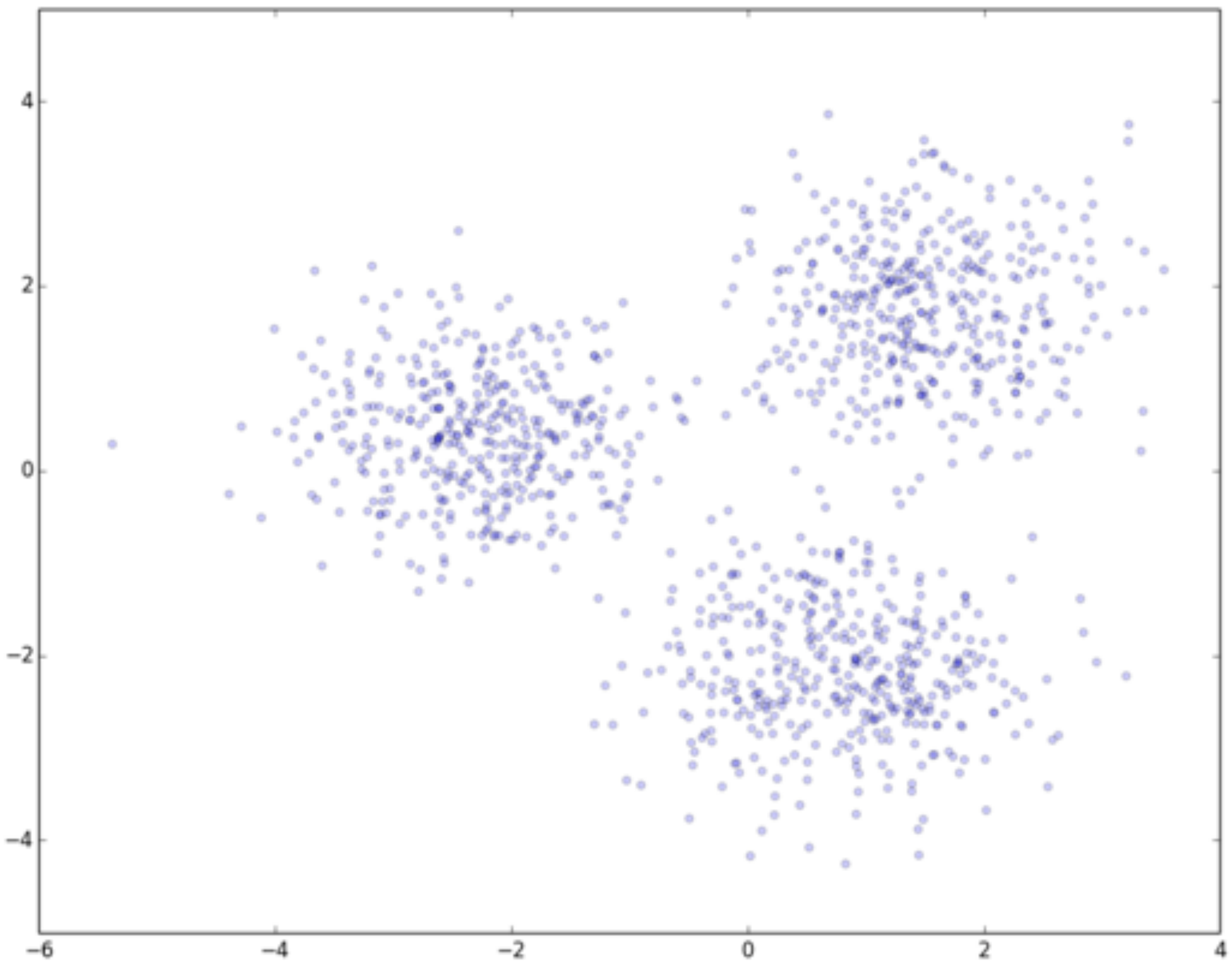


AYASDI

Machine Learning and TDA

Example: PCA

Provides unsupervised dimensionality reduction.
Easy to interpret.



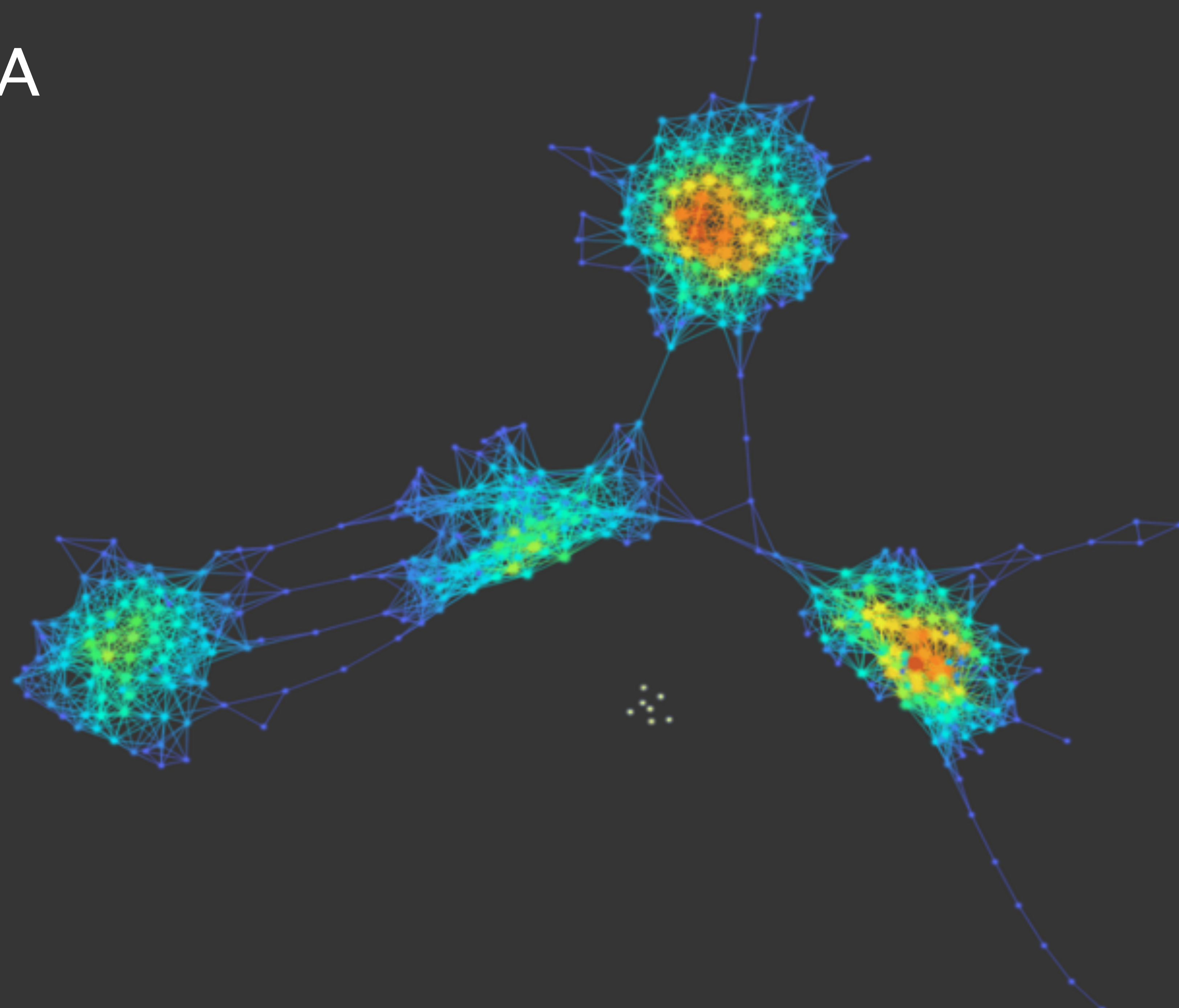
PCA captured 98.4% of the variance

AYASDI

Unsupervised Learning: PCA

AYASDI

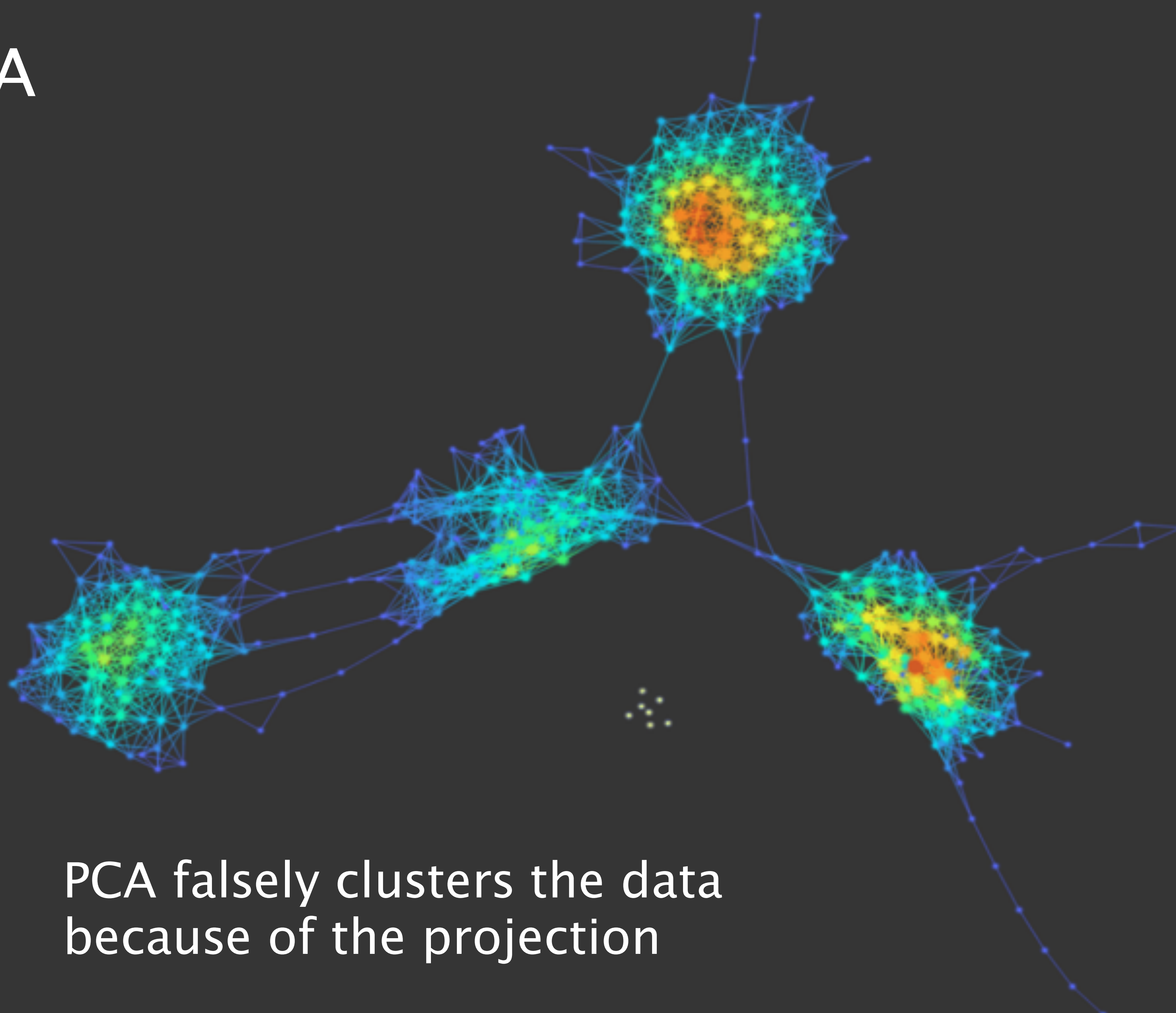
TDA + PCA



Unsupervised Learning: PCA

AYASDI

TDA + PCA



PCA falsely clusters the data
because of the projection

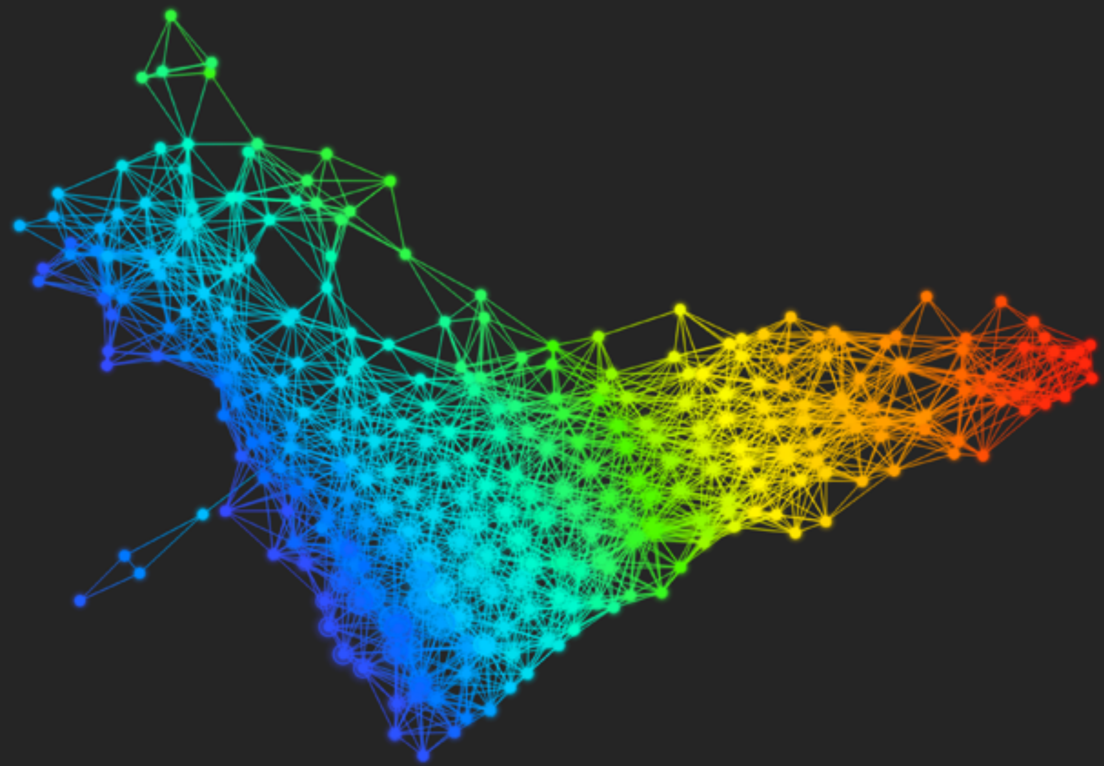
TDA clusters in the inverse image

Emergency room triage model

AYASDI

Predicted mortality

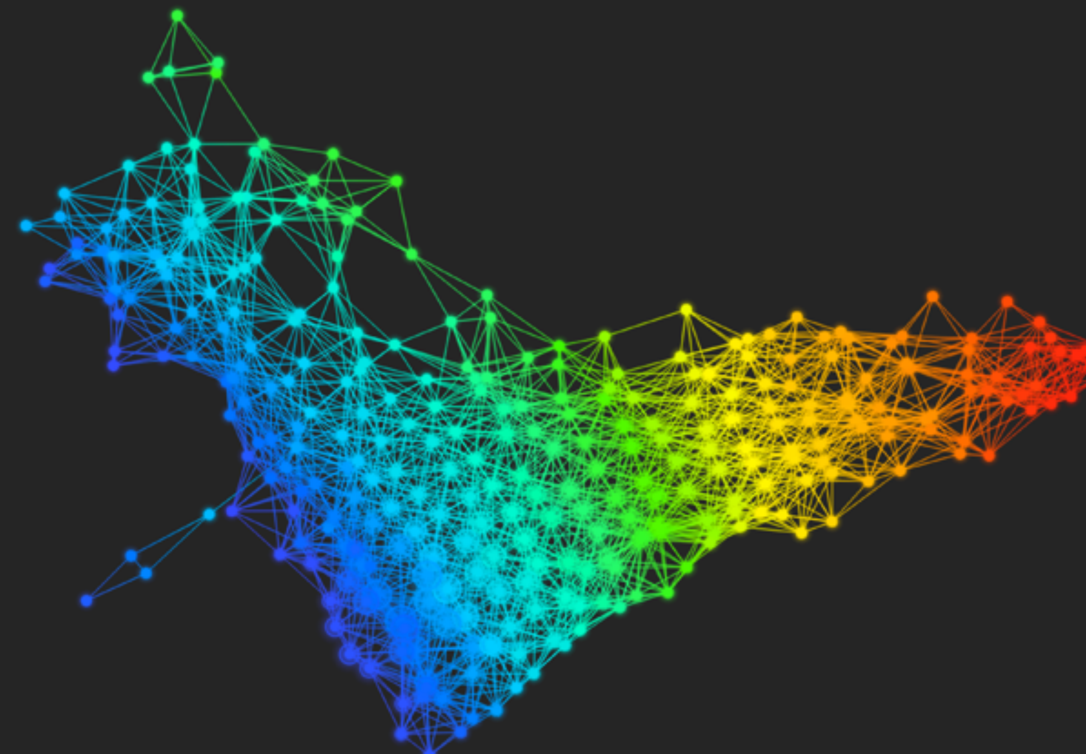
Low High



Emergency room triage model

AYASDI

Predicted mortality Low High

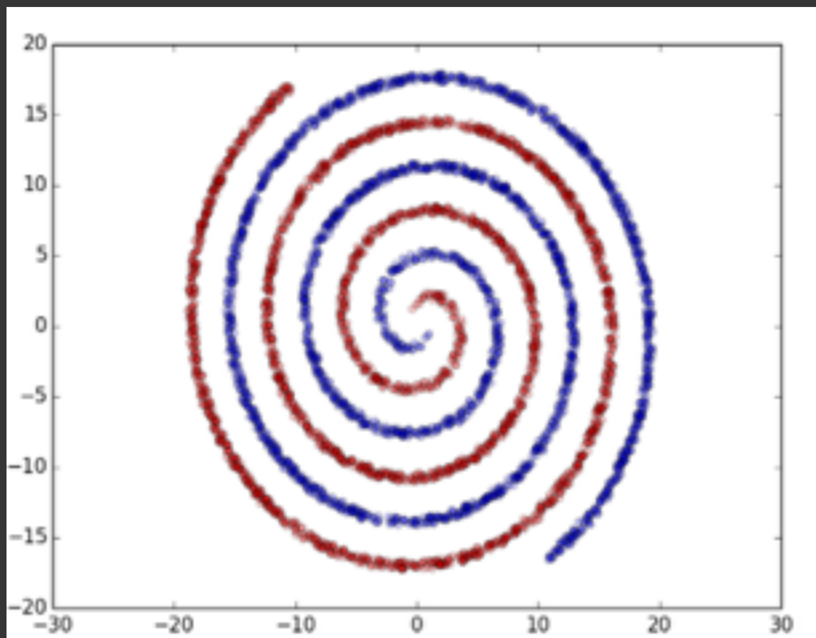


Actual mortality Low High

Missing responses to particular questions caused model to fail for these patients

TDA tells us where to look for problems and questions

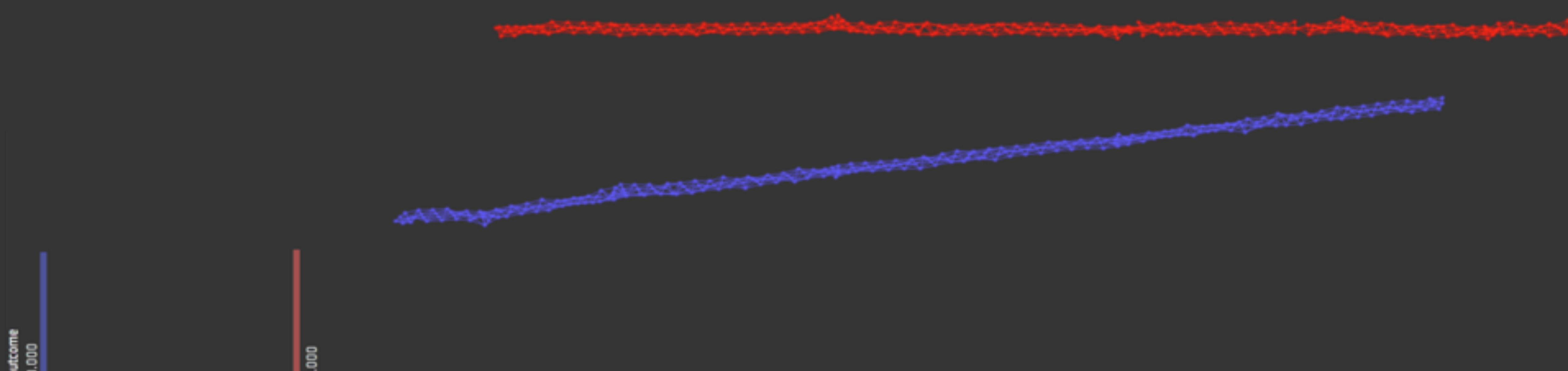
TDA: Beyond Machine Learning



Working locally means features don't change when stretched or distorted.

TDA is

- resistant to noise
- requires less preprocessing of data
- robust/stable in its answers



AYASDI

Customer Use Cases

Predictive Maintenance & Machine Uptime

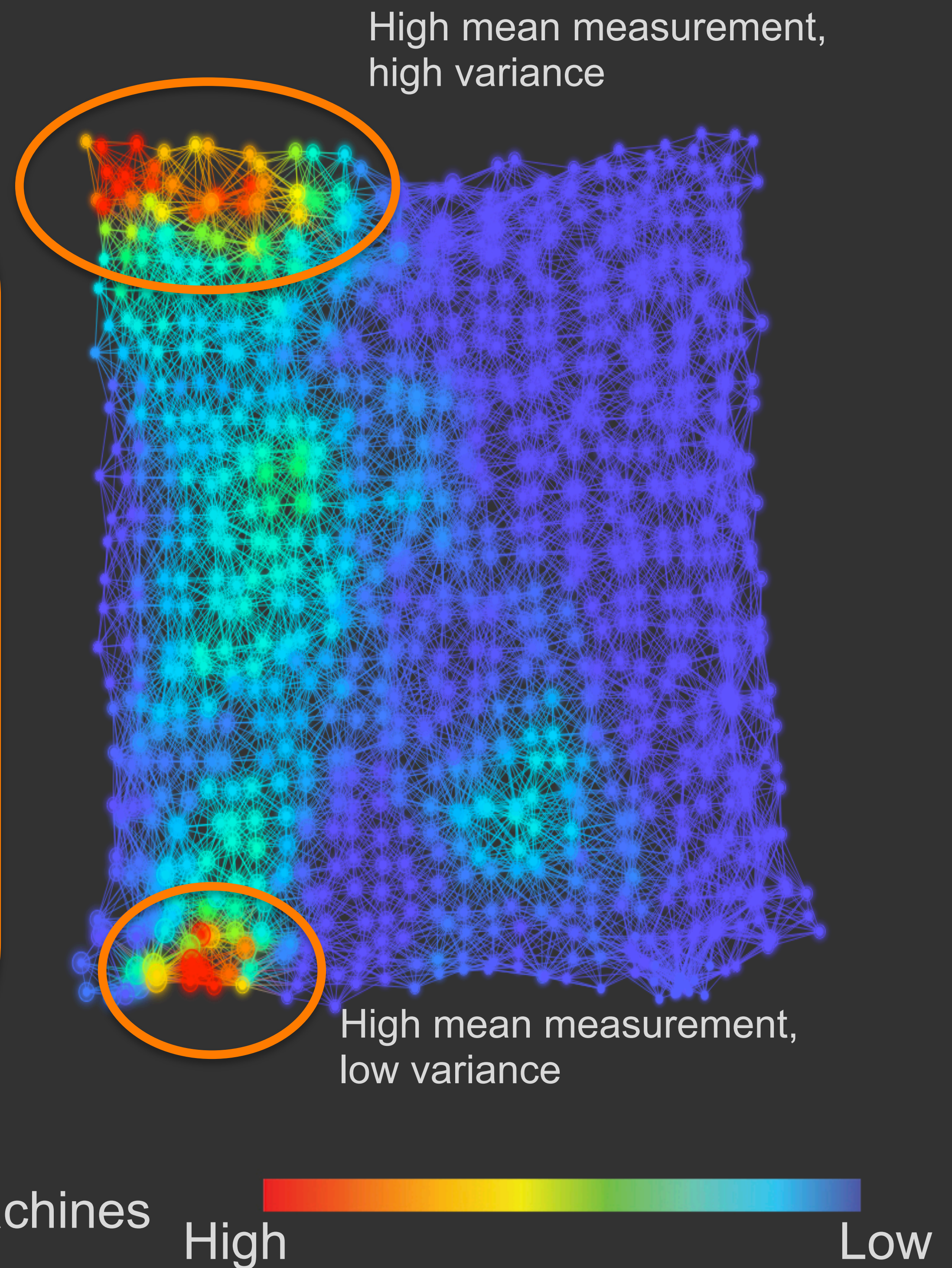
Challenge: identify indicators in sensor data that help indicate machinery failure

Data: system measurements capturing machine characteristics

Result: identification of the key machine attribute that reveals impending failure

- High mean, high variance – failure
- High mean, low variance – failure
- High mean, medium variance – no failure

The key attribute could be identified using standard methods, but that information was insufficient to predicting failure.



AYASDI

AYASDI

Carbon Capture: Searching Zeolite Structures

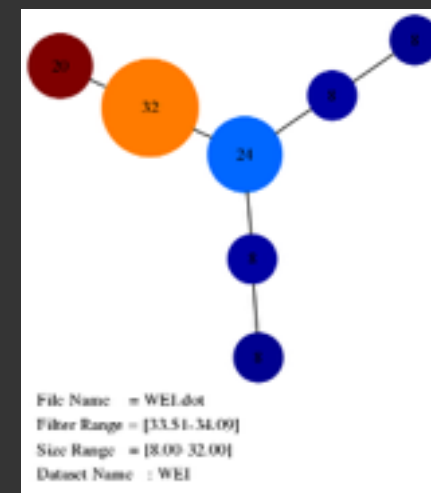
AYASDI

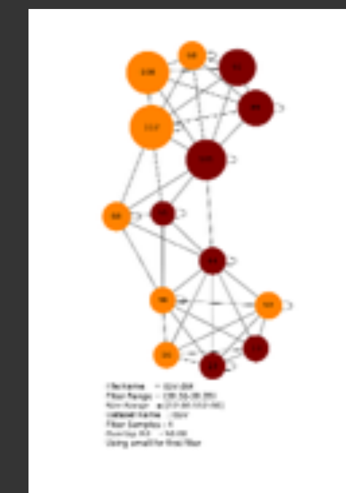
Problem: Search database of compounds to find structures with good carbon capture properties

Data: 30,000 3D crystalline structures of theoretic zeolite compounds

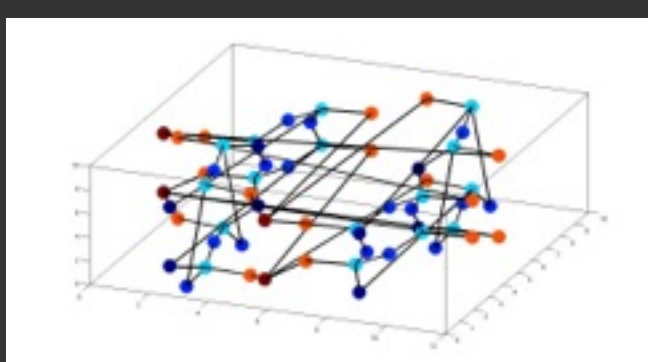
Result: Identified compound that had 10x selectivity for CO_2 over CH_4

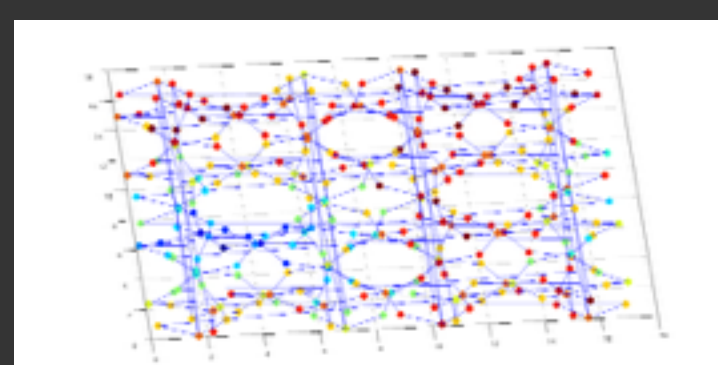
Example geometric summaries





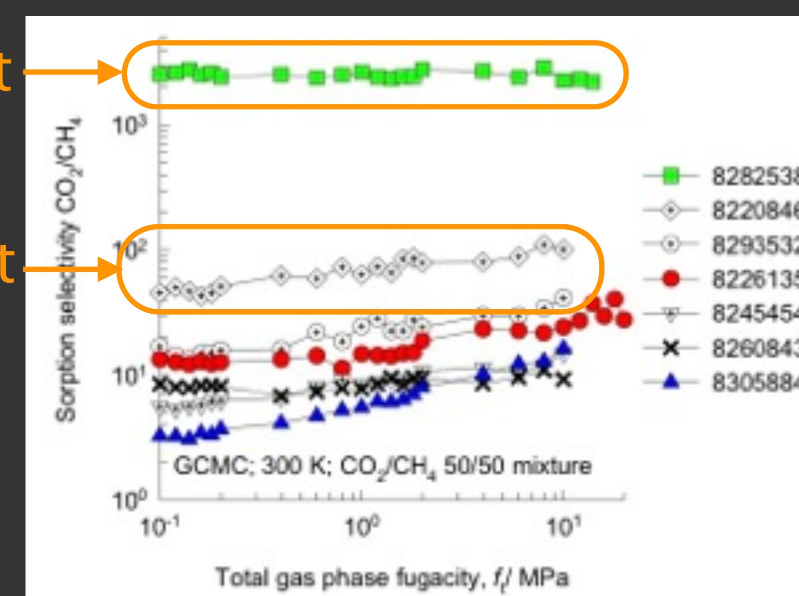
Example data





Newly discovered best

Previous best



TDA + Random Forest: Credit Card Fraud

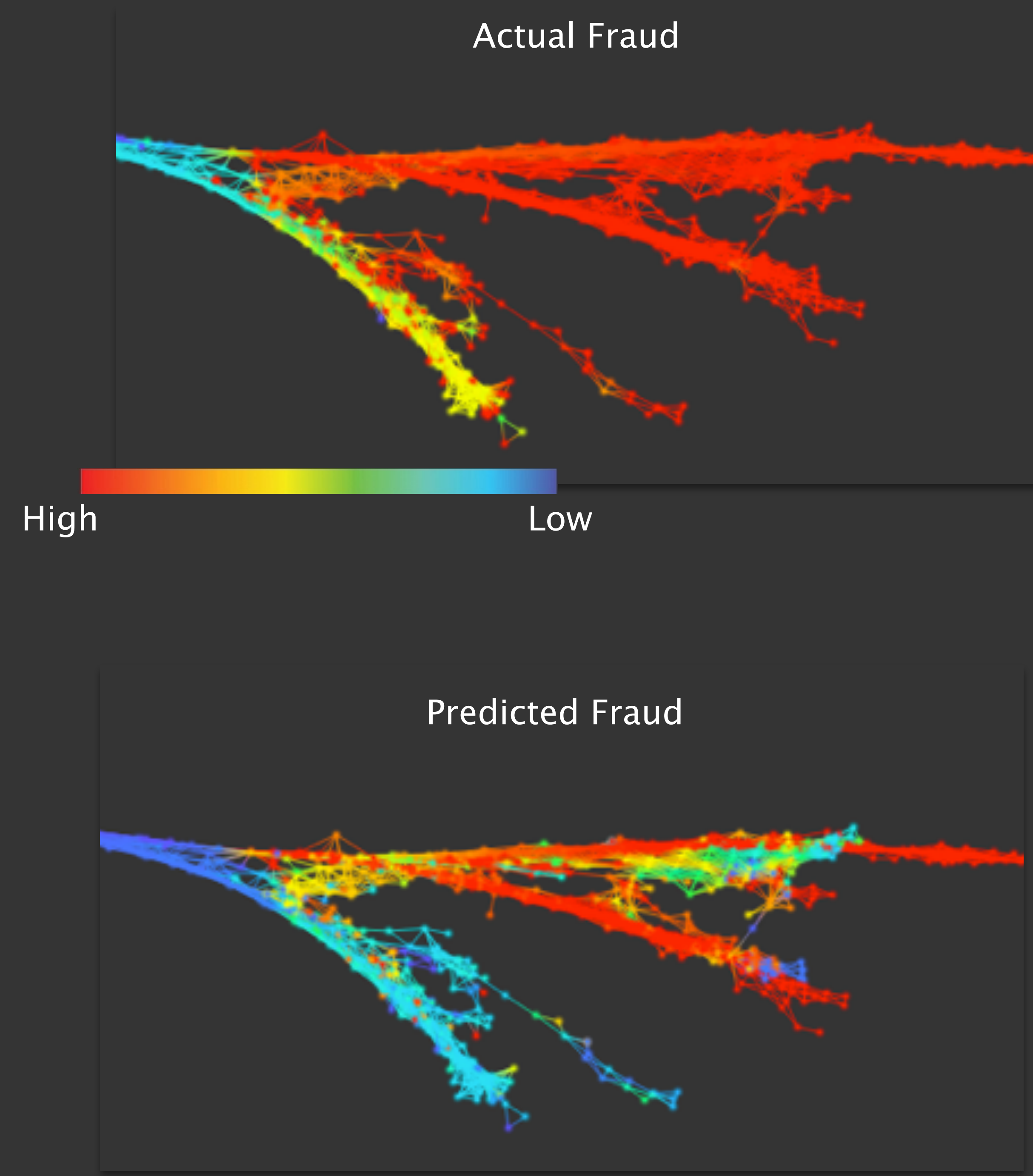
AYASDI

<u>Problem</u>: Identify credit card fraud

<u>Data</u>: Predictions from 200 random forest trees on several hundred features on credit card transactions

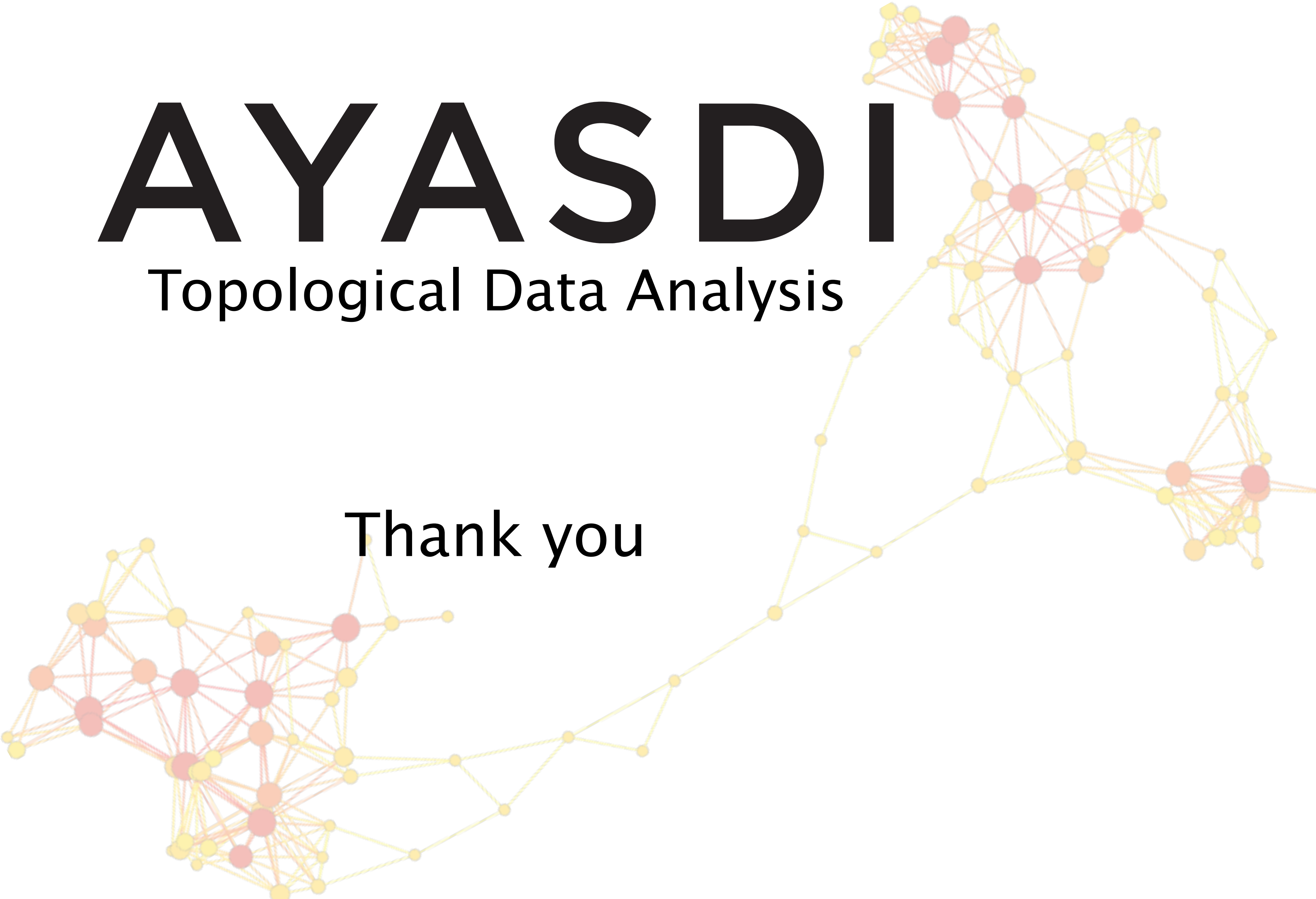
<u>Result</u>: Identified large subset of mis-categorized fraud. Uniquely identify the characteristics of this group and improved accuracy from 28% to 99.3%

Geometric Summary built with Random Forest Metric



AYASDI

Topological Data Analysis



Thank you

Data has **Shape**, Shape has **Meaning**, Meaning drives **Value**