

Topological Data Analysis

Hendri Karisma
Sr. Software Development Engineer
bilibli.com

What we think if we
heard the word “data”?

Provide Large Data

- Complicated dataset to analyse
- How to make sense of large array of numbers, of column and rows

Topological Data Analysis???

- Analogy a group of data is a node
- Each node have relationship and we call it edge
- We can see unstructured massive number, we have network and shape. Shape of network representing the shape of our data
- Use visual system to look the data and identify the feature, in the network which corespon with the data.
- Extracting knowledge of data.

Why TDA ?

- Data has shape
- Shape has meaning

TDA History

- Topology sub field in math the concern with the study of the shape
- The origin in 18 century with swiss mathematician, Leonhard Euler.
- Euler become aware of a challenge problem, concerning the seven bridges crossing the river.
- He took all the information about the bridges and the river, and the island, and the land mass. Then converted it all into simple network, and he found that in fact that is not possible to cross the bridge exactly one.

- The last 15 years, there has application to many different real problem.
- One of those, the analysis and understanding of high dimensional and complex dataset.
- This area study is called topological data analysis, it's changing the way that people are able to understand and analyze their data

Complexity of the data

- Large data set can be simple in extractor
- Small data set can be complex

The Properties of TDA

three big concept that give it's big power for analysing and understanding shape

- Coordinate invariance
- Deformation Invariance
- Compressed Representations



Line



Clusters

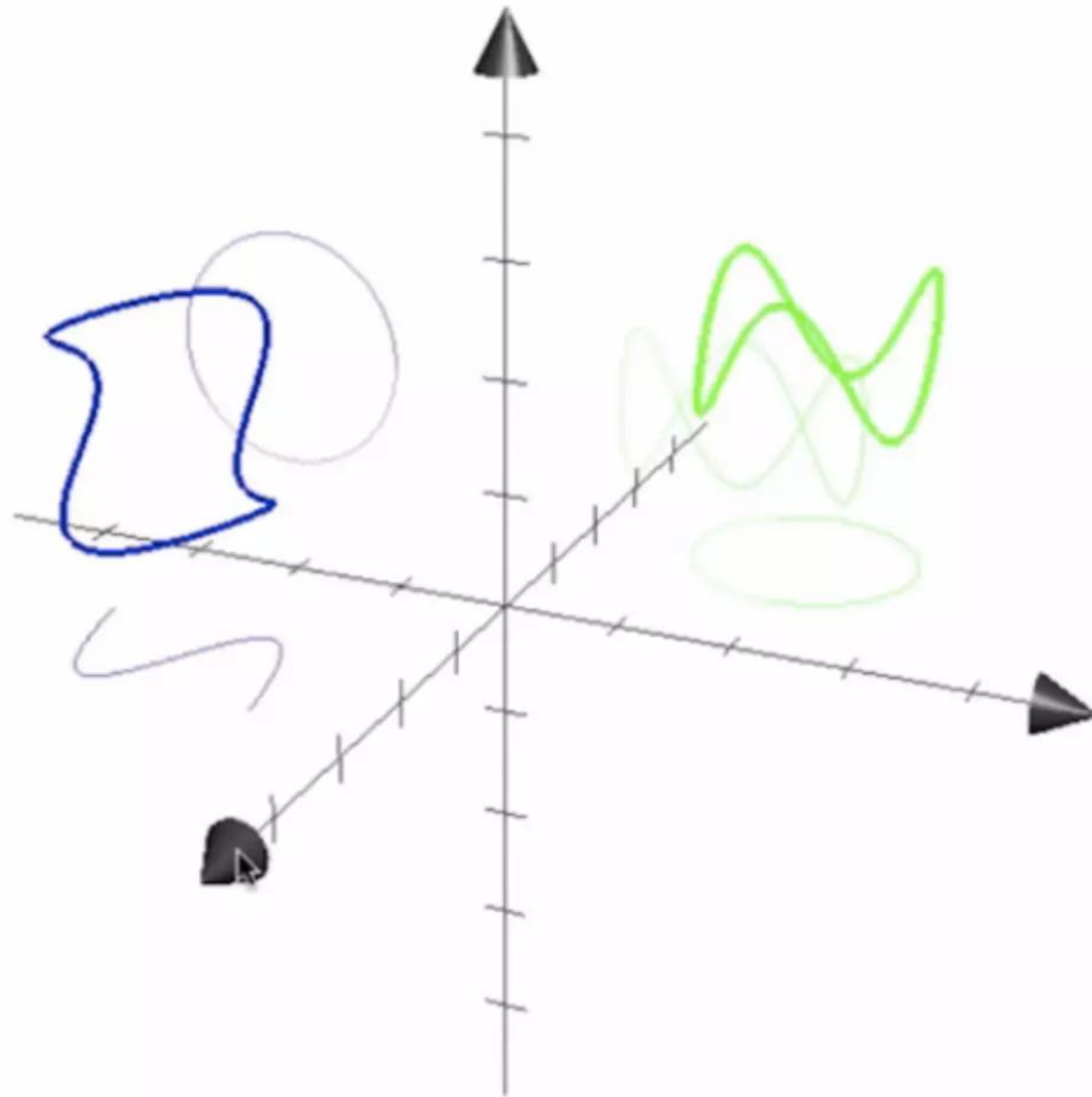


Loop

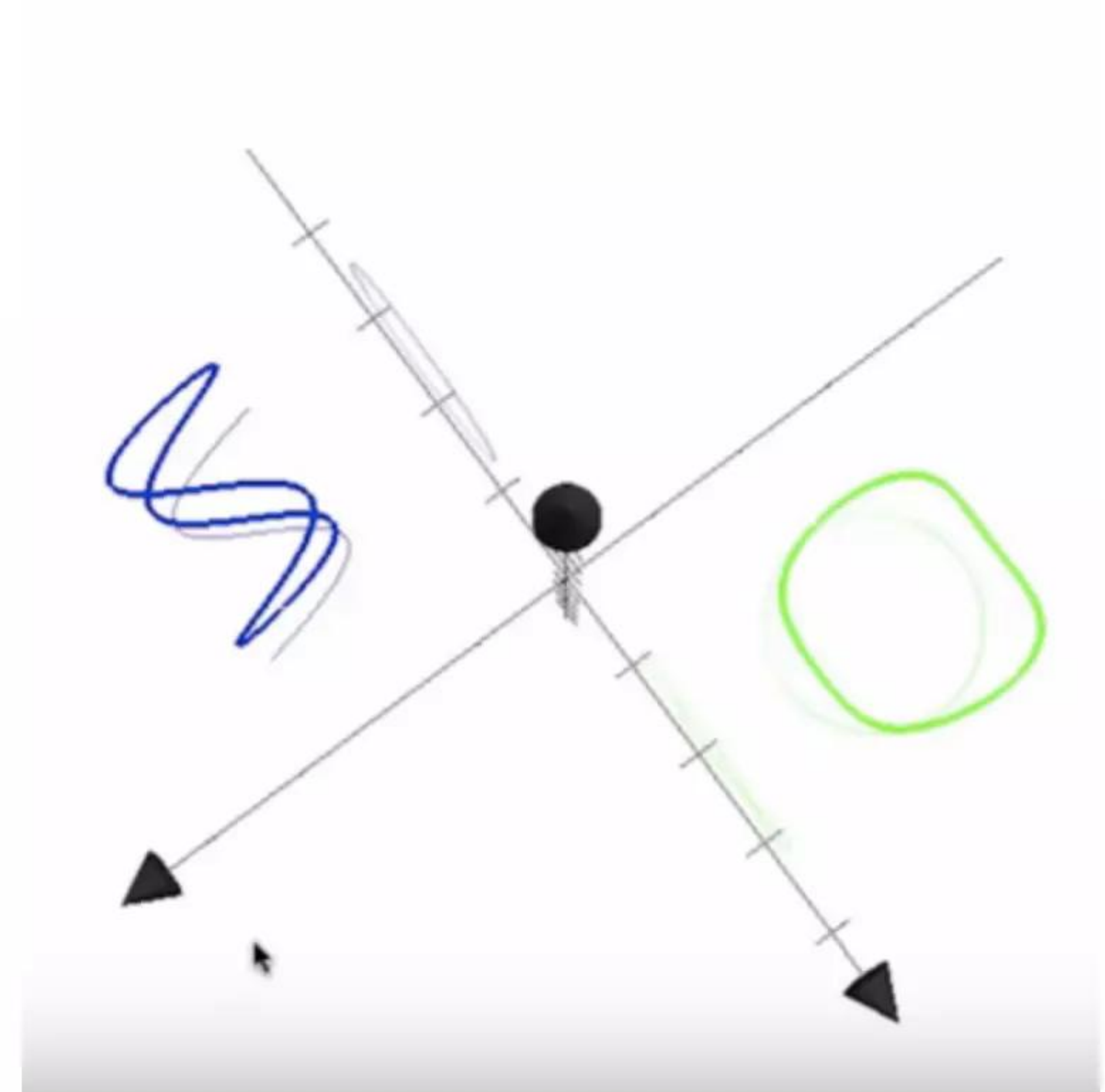
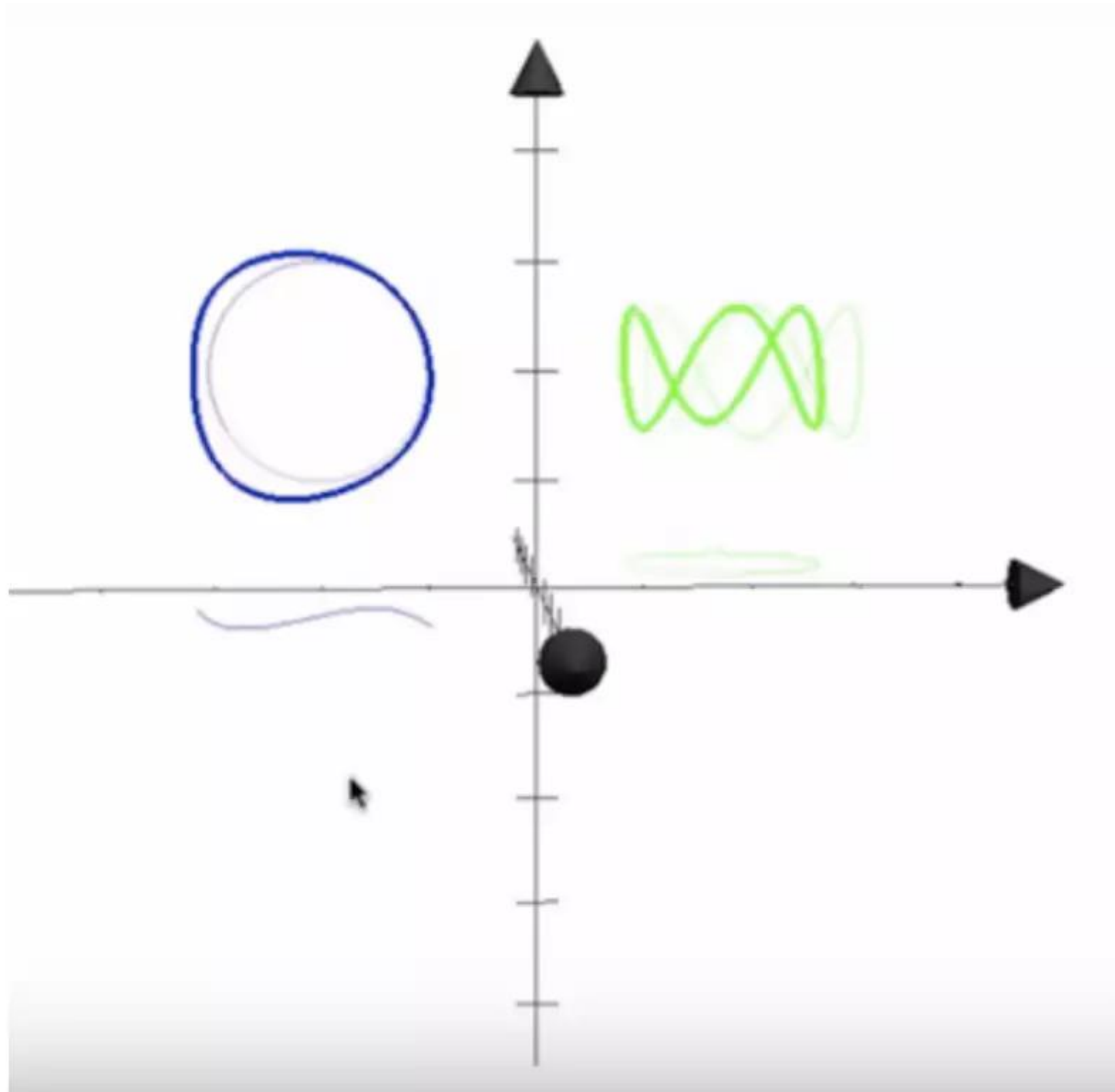


Flares

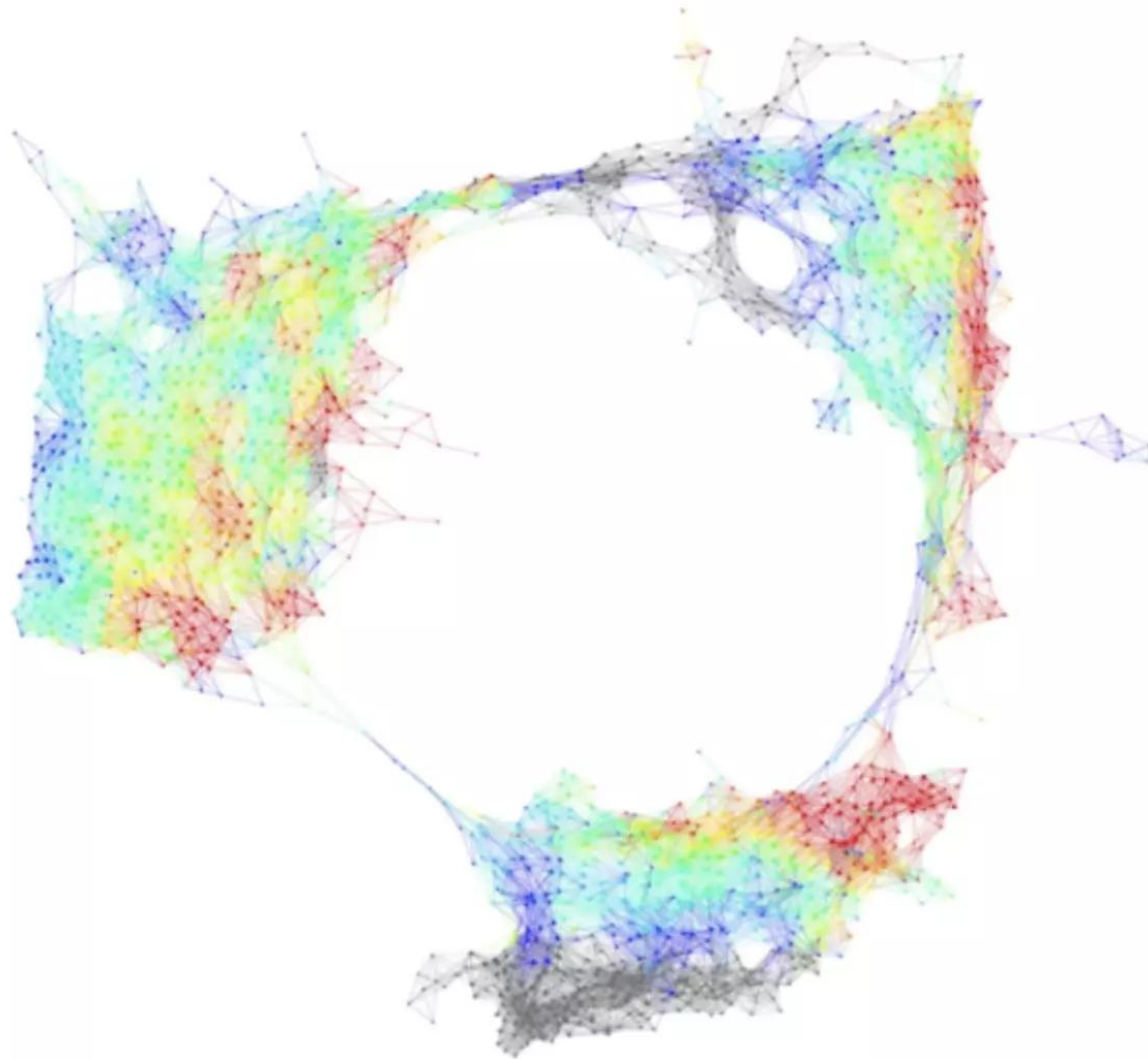
Data Comes with The Wrong Candidates



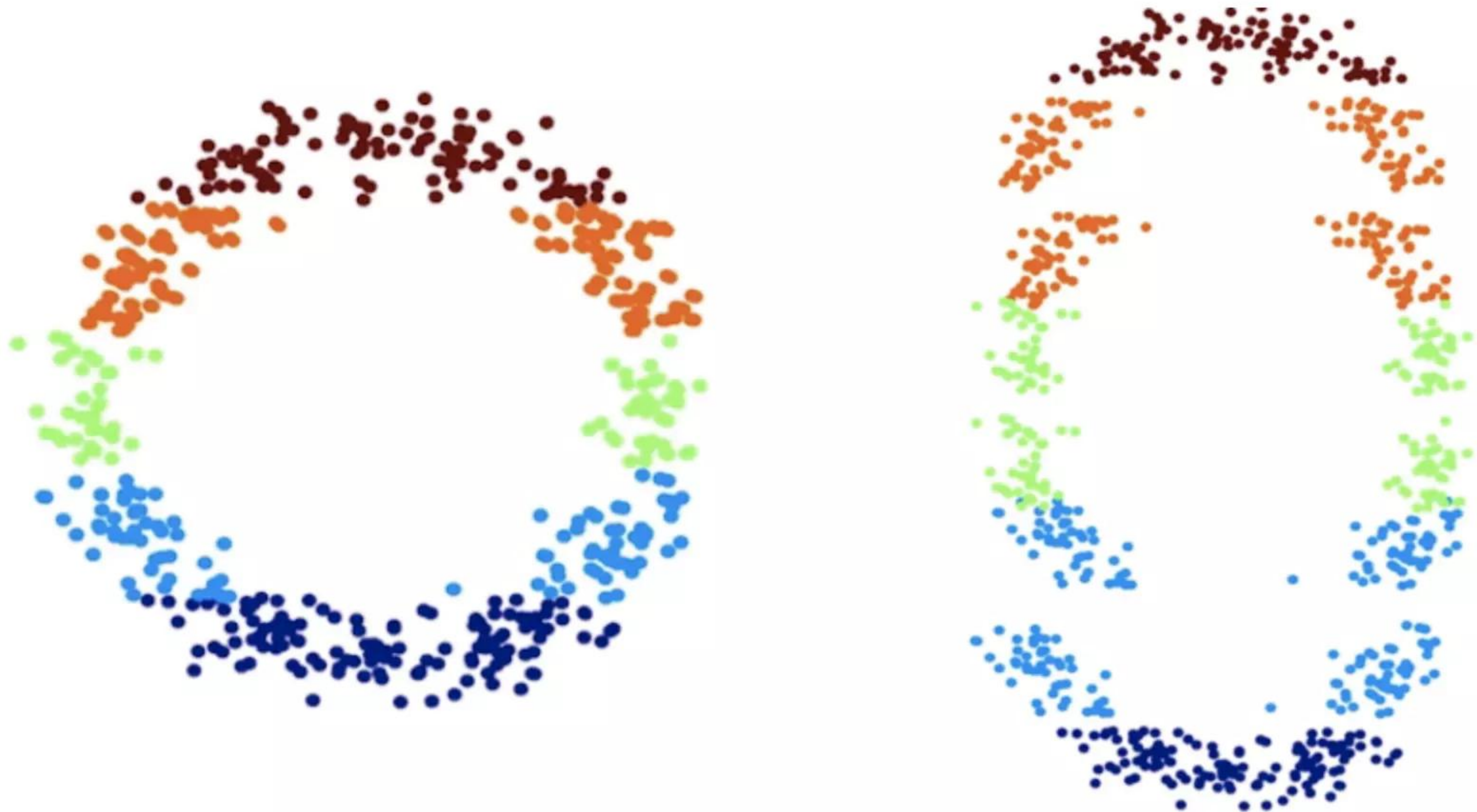
Data Comes with The Wrong Candidates



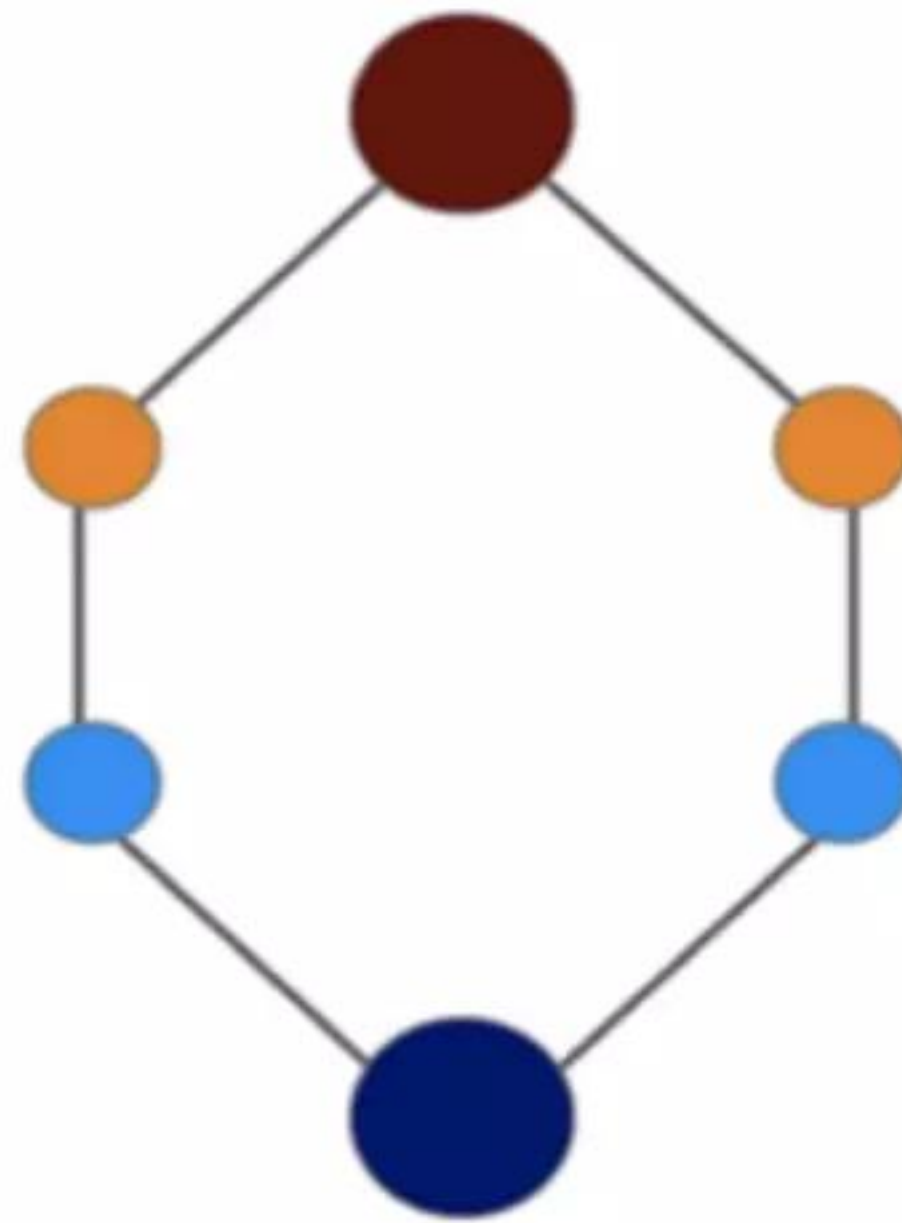
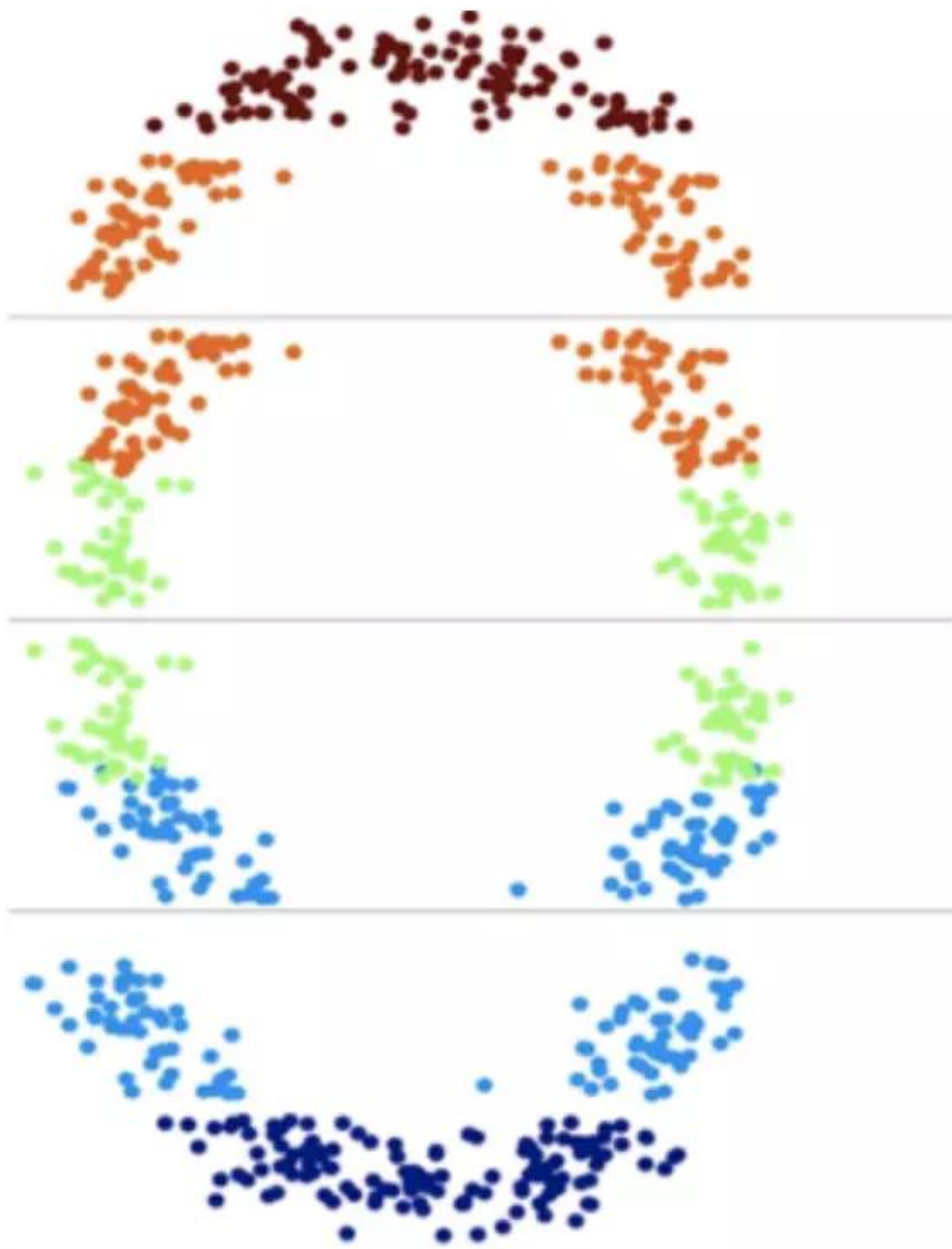
Complex Data



Topological Network Capture Shape

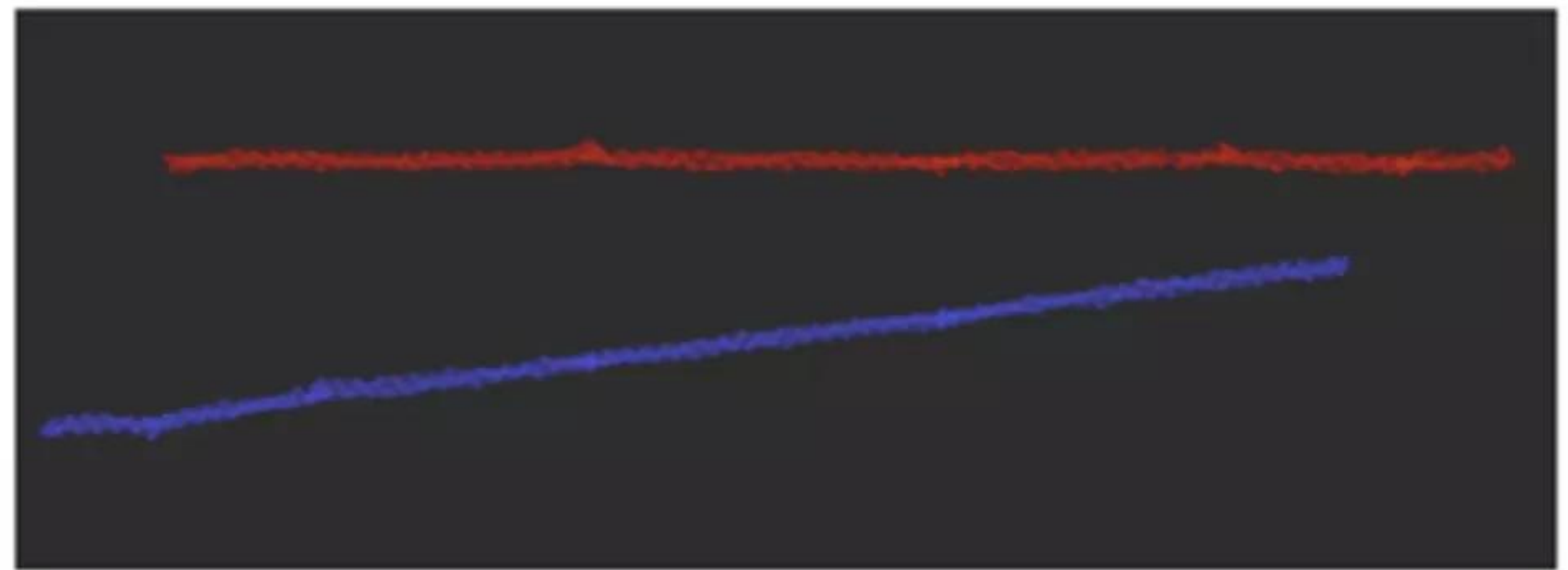
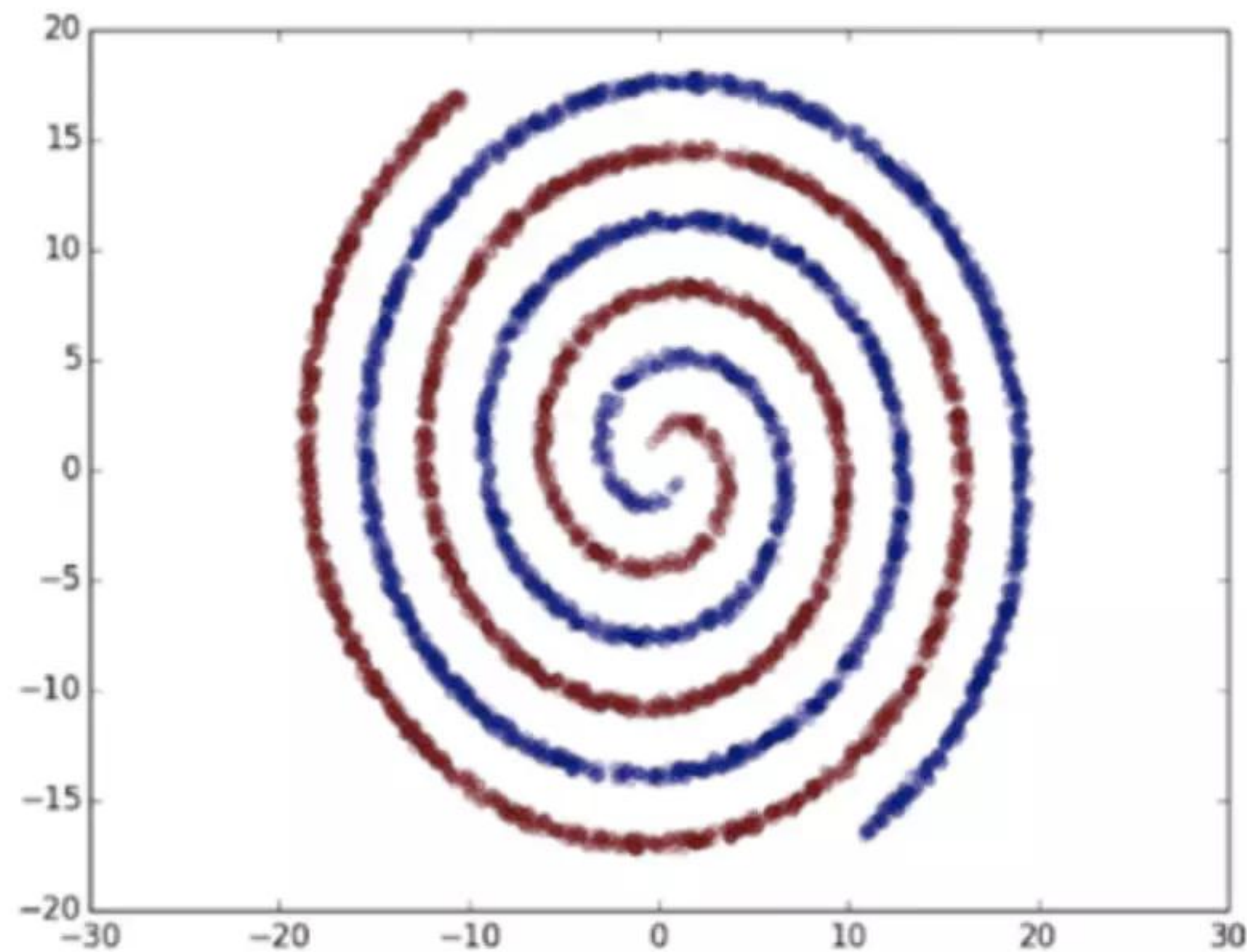


Topological Network Capture Shape

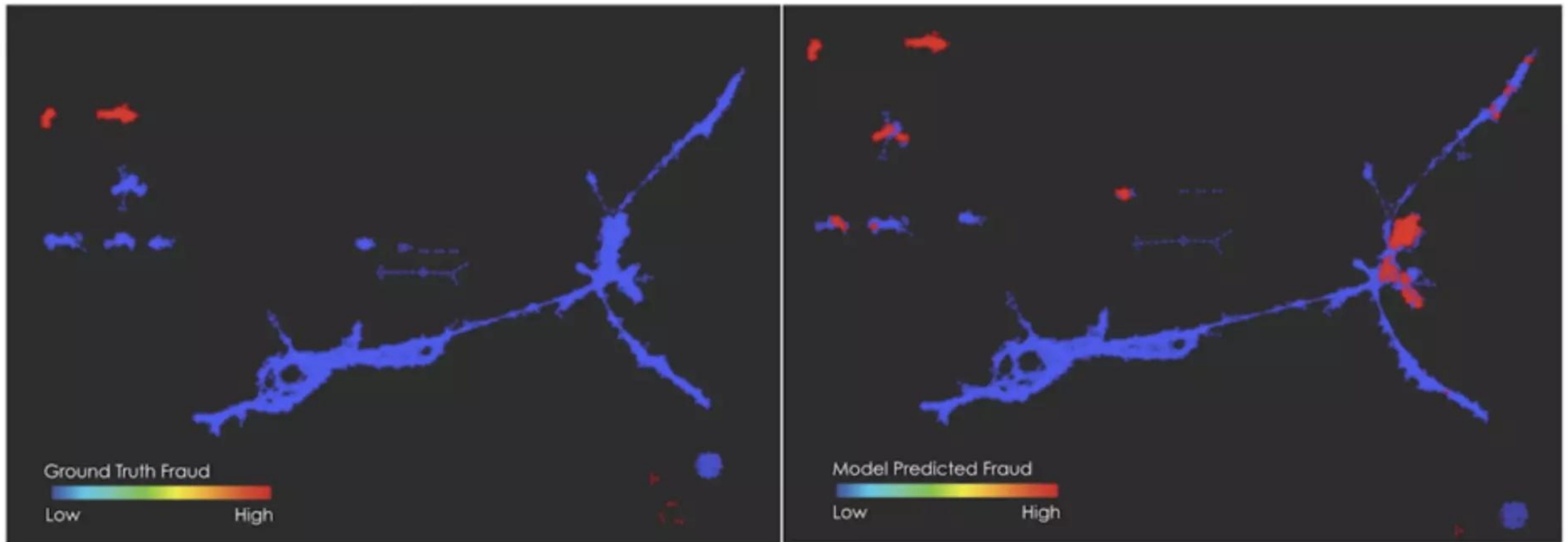


- Color by a real-valued function
- Bin in the image of that function
- Cluster within bins, in the original space
- Connect clusters that share data points

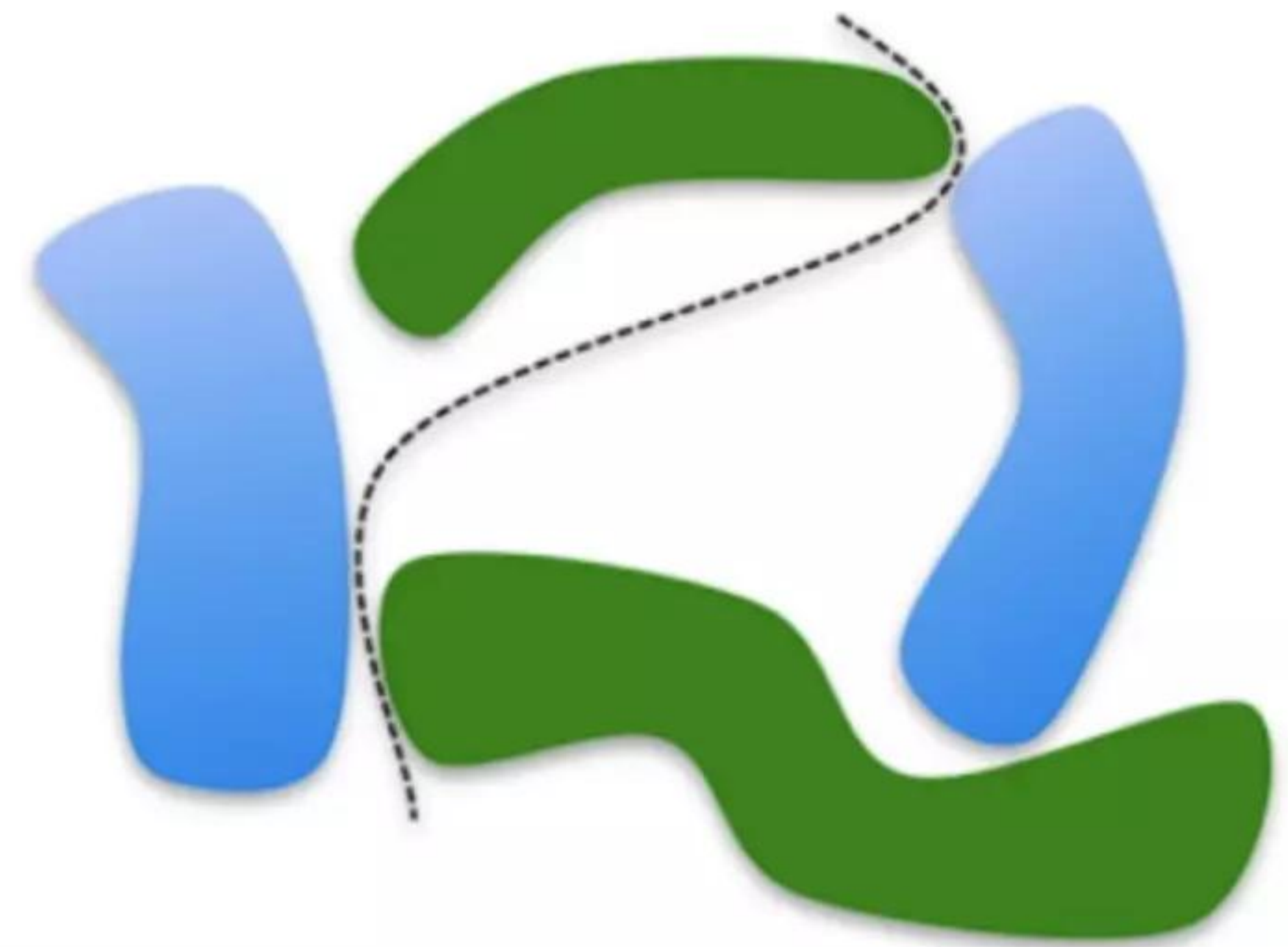
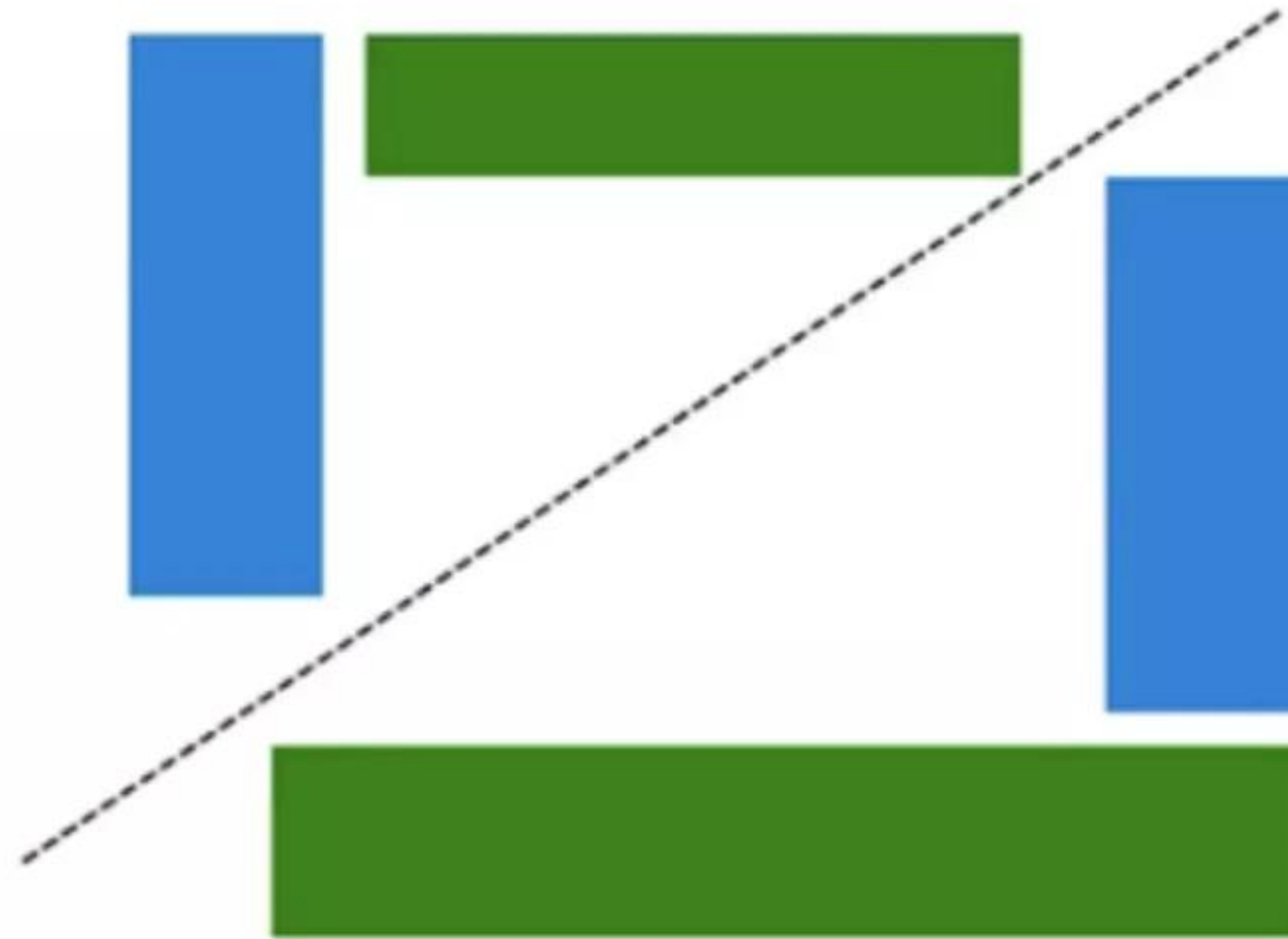
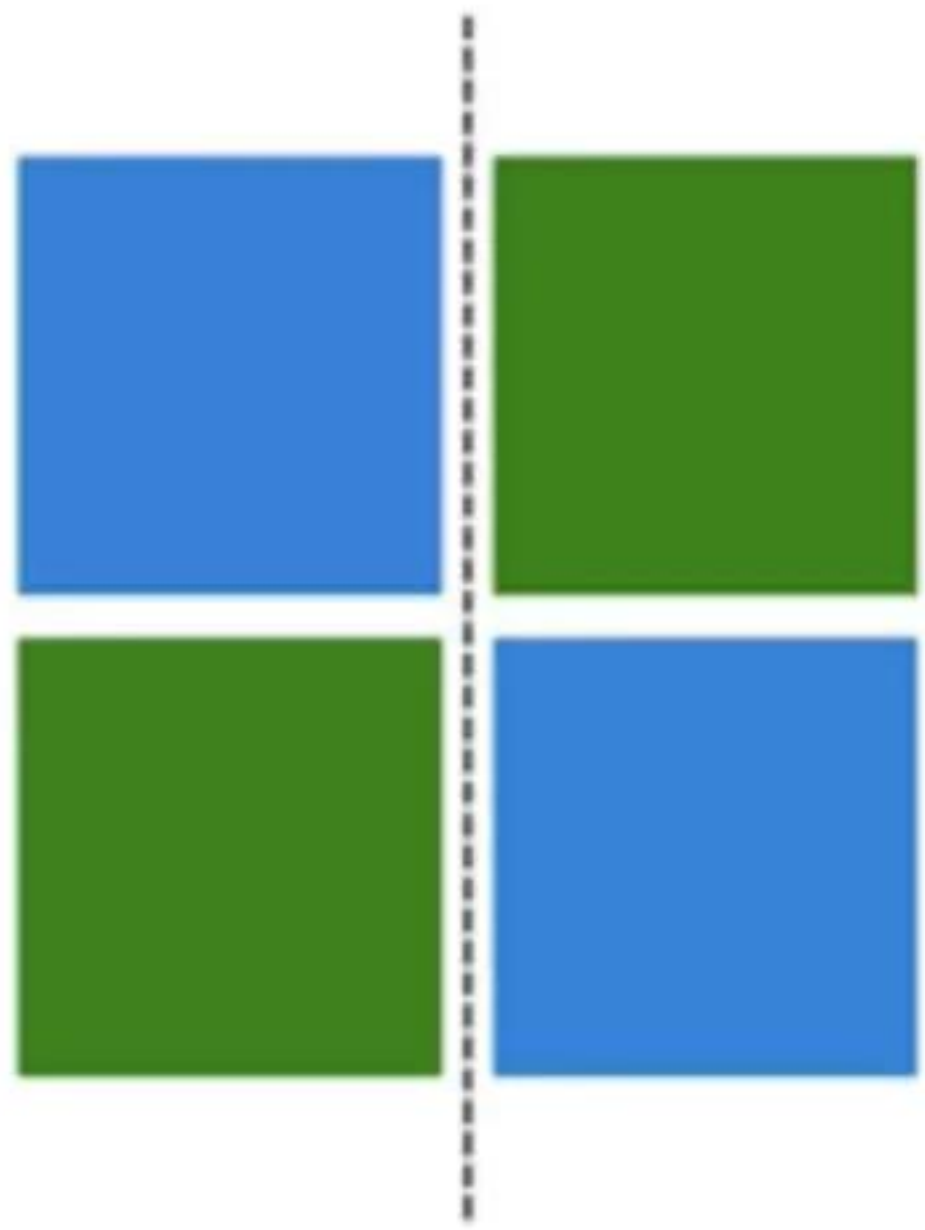
Capturing Complex Shape



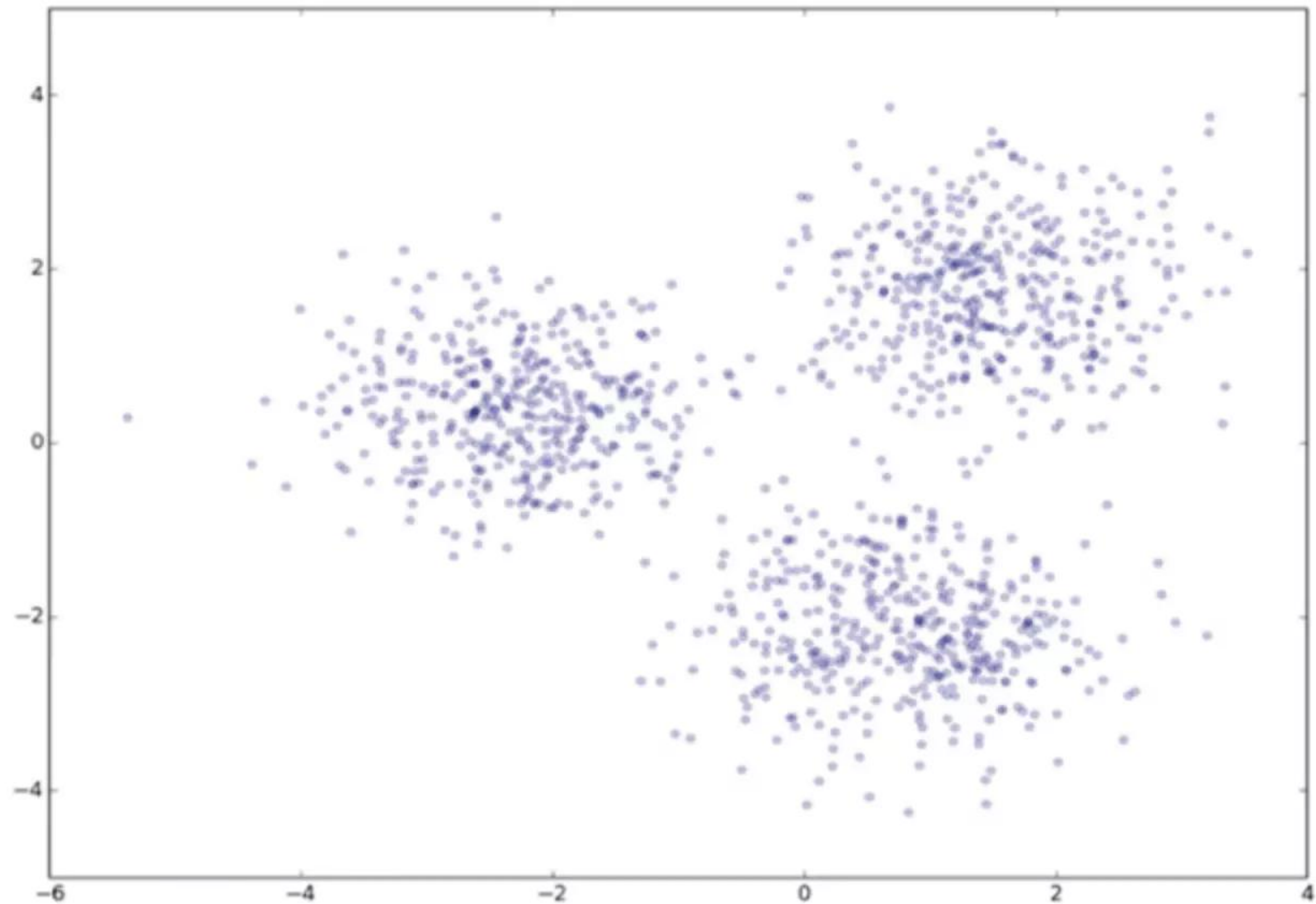
Example Fraud



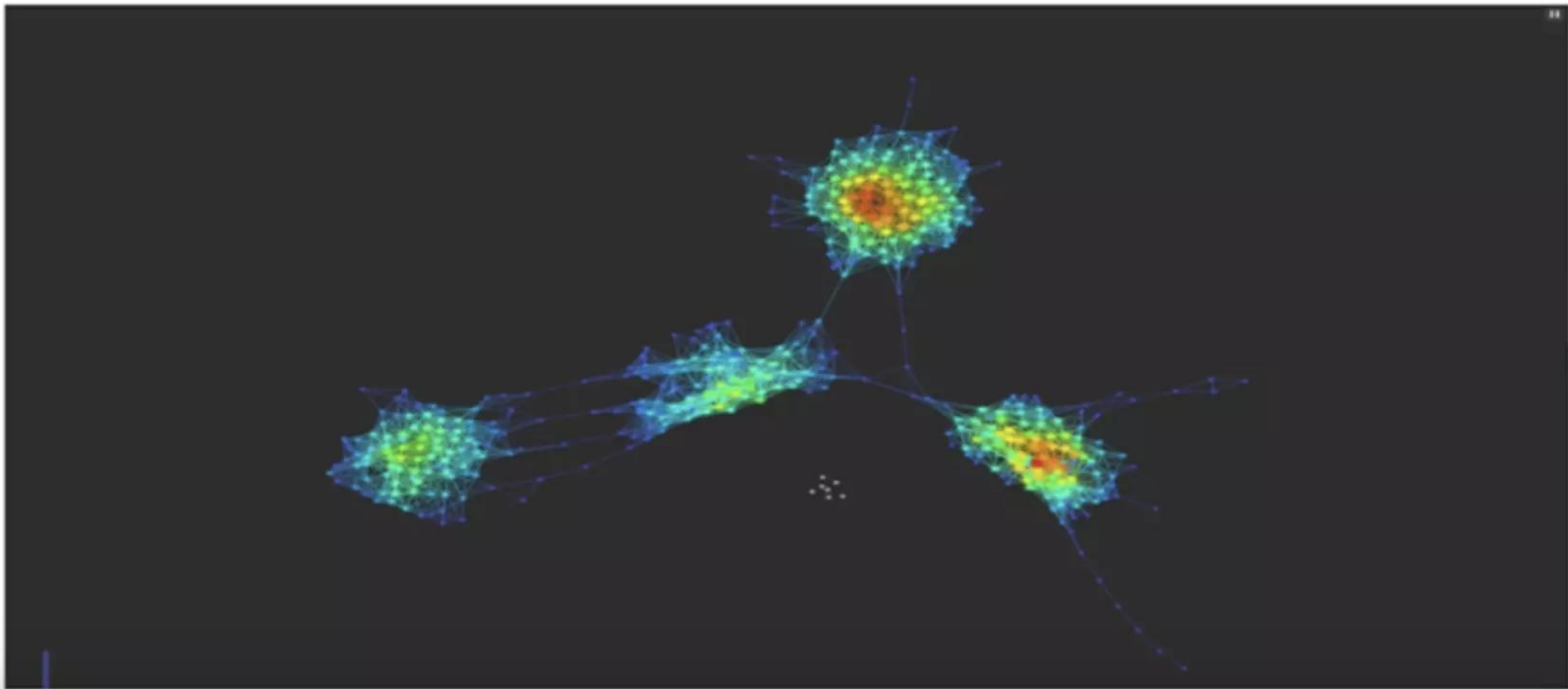
Topology Yields Principled Local Models



PCA Find 3 Cluster



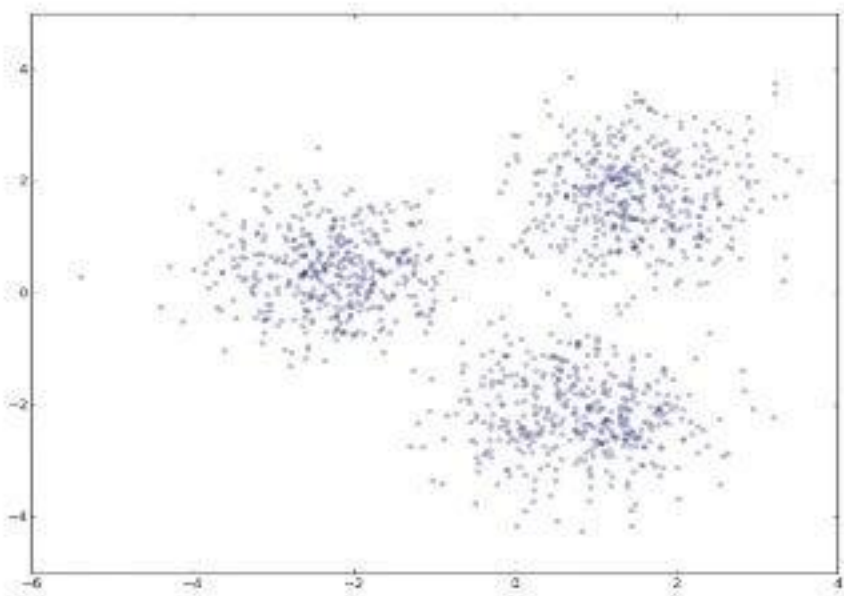
Actual is 4



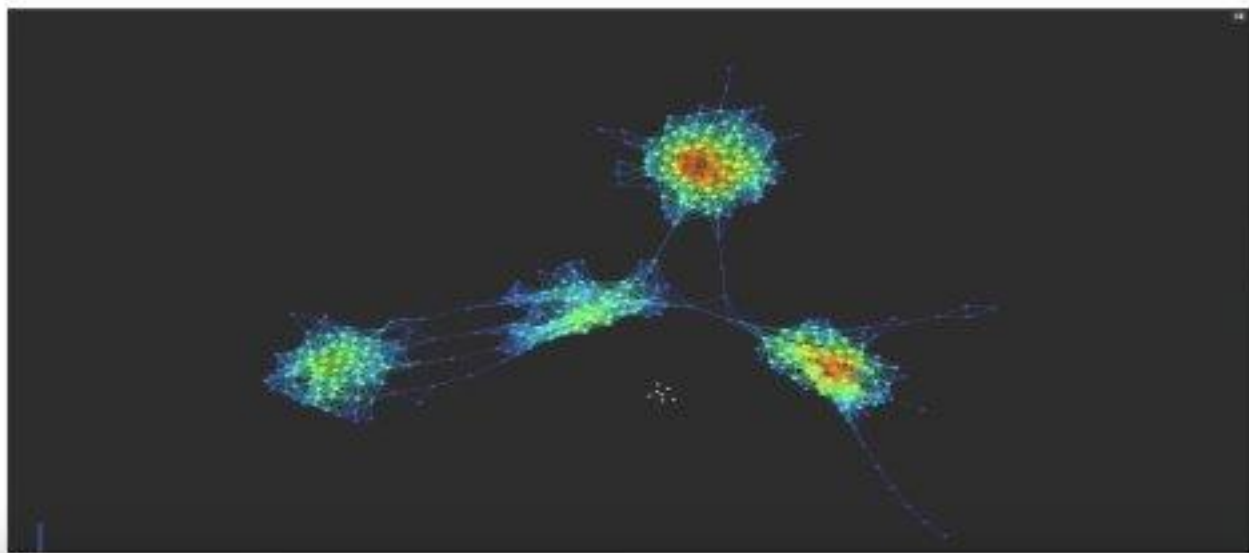
Summary

- The three properties combine and very striking ways to allow one to analyse and understand very large and complicated data sets
- Topological data analysis represent a fundamental advances in machine learning
- In the near future machine will help humans organise simplify and understand their very large and complicated data sets

PCA FIND 3 CLUSTER



THE ACTUAL IS 4



SUMMARY

- ▶ The three properties combine and very striking ways to allow one to analyse and understand very large and complicated data sets
- ▶ Topological data analysis represent a fundamental advances in machine learning
- ▶ In the near future machine will help humans organise simplify and understand their very large and complicated data sets

THANKS