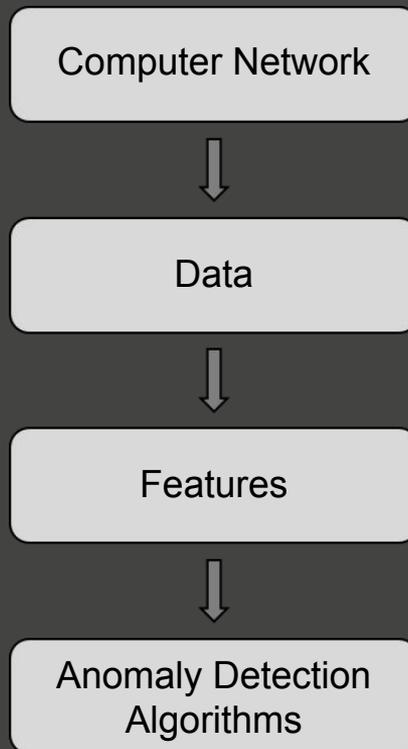


Geometry and Topology for Data

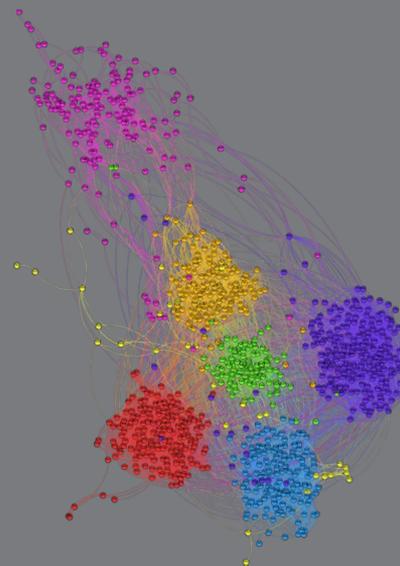
Leveraging topological and statistical tools to uncover structure in data

Problem:

- Detect abnormal behaviour in communications networks using principles of statistics, topological data analysis, and machine learning.
- Capture both local and global anomalies using a comprehensive combination of mathematical methods.

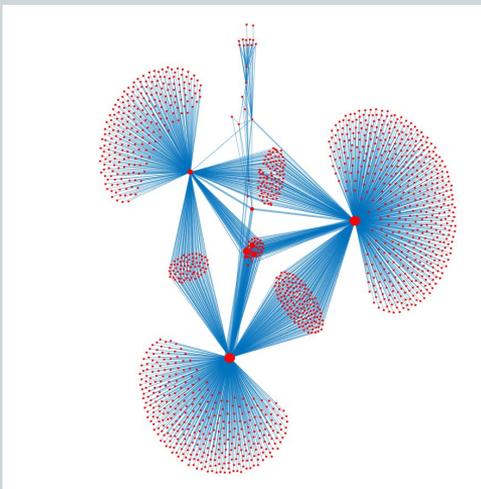


Erin Taylor
Ian Frankenburg
Eric Peshkin
Mentor: Joe Marion
Client: Dr. John Harer,
Geometric Data Analytics

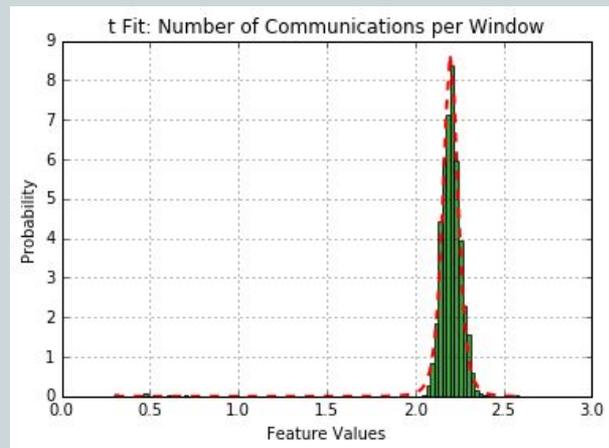


Methods

We built a network with nodes representing different IP addresses and edges representing communications between them. Topological tools allowed us to study how the connectivity within the network changed as edges were added between nodes.

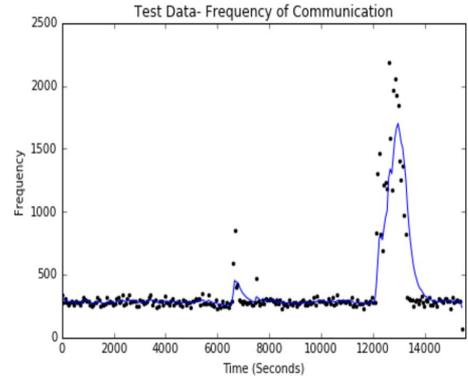
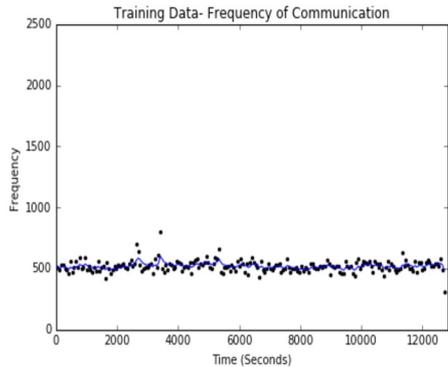
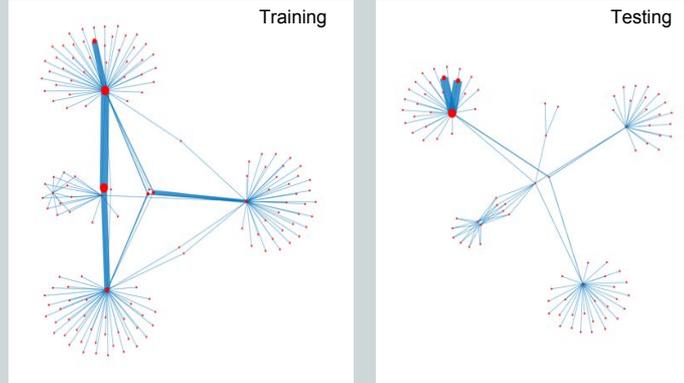


We automated the process of fitting a probability distribution to our data. Our utility fit many potential probability distributions and used AIC to determine the best one. This gave us a probabilistic basis for separating improbable (potentially anomalous) events from typical events.



Results

When compared with the training network, communication within the testing network appears more sparse, and some nodes exhibit unusual communication.



Using a moving average of the frequency of communication between two nodes, we found many anomalous communications.

Conclusion:

We identified many potentially anomalous communications using both the topological and statistical methods. Our analysis indicates that further exploration on more comprehensive data sets could result in interesting findings, especially on the network level.