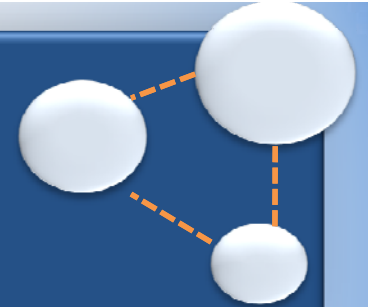




Data Mining and Modeling Of Time-Evolving Graphs



About Me

- Luke Gebauer
 - College:
Allan Hancock
 - Major:
Electrical Engineer

Faculty

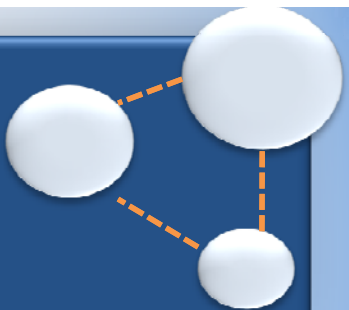
- Mentor:
Kathy Macropol
- Advisor:
Dr. Ambuj Singh

Project

- Computer
Science
- Funded by
the US Army



What Time-Evolving Graphs Are Not



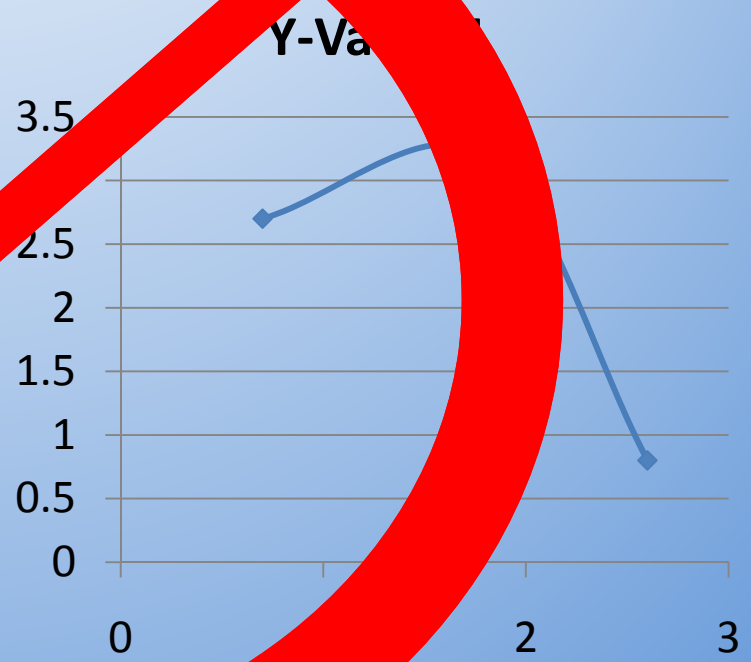
Sales



- 1st Qtr
- 2nd Qtr
- 3rd Qtr
- 4th Qtr

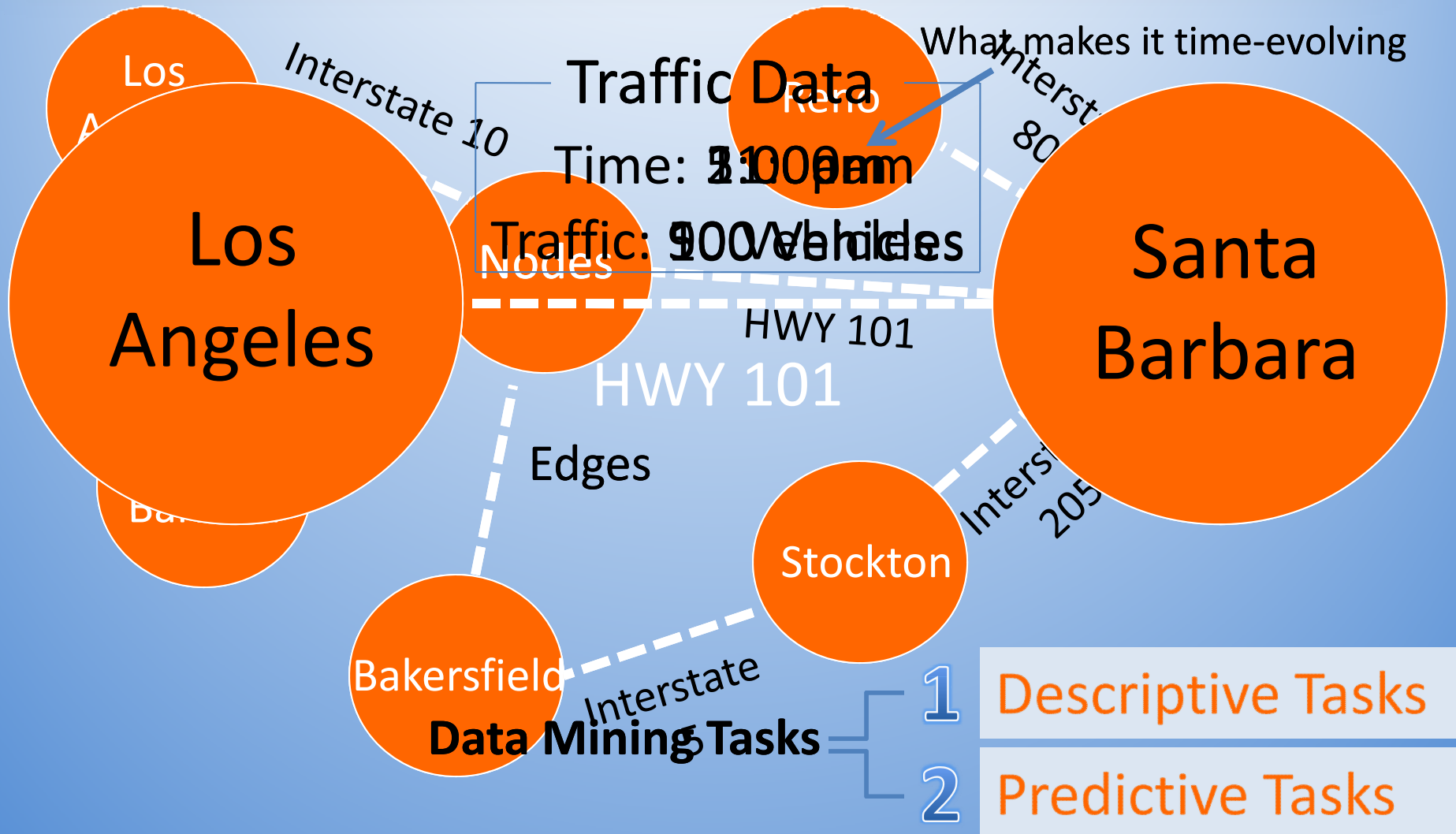
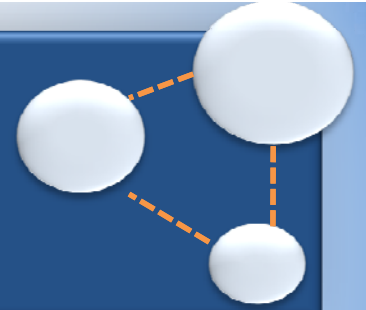


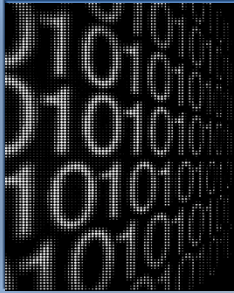
- Series 1
- Series 2
- Series 3



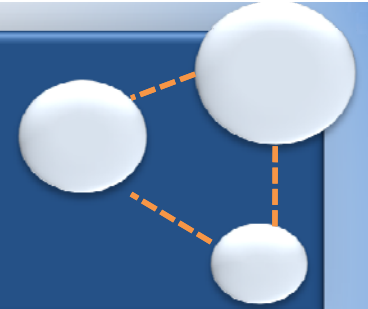


What They Are and What We Do With Them





Research Goals



Research Objectives

Importance

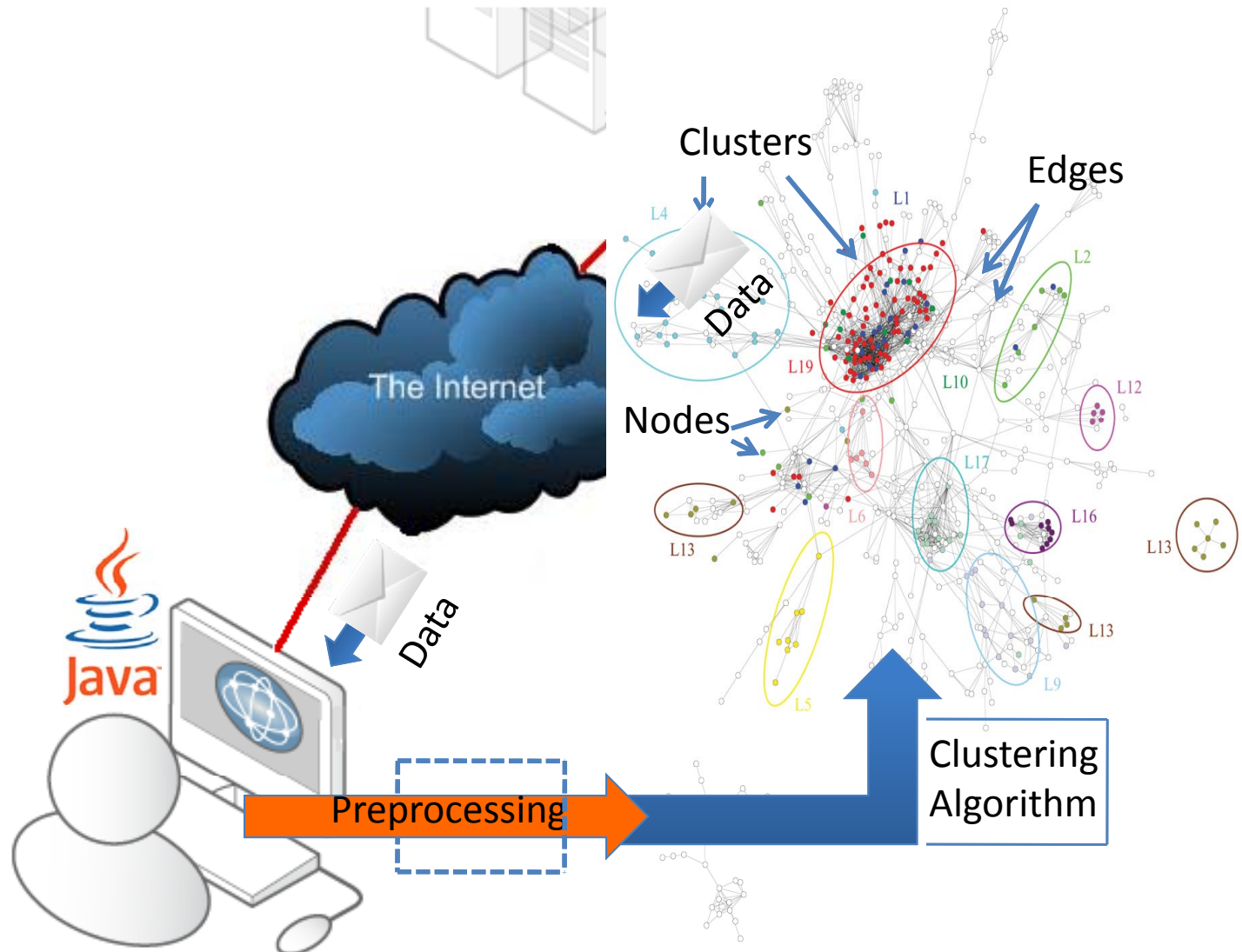
Construct a time evolving graph of countries and their international trade transactions

Provide additional time evolving graphs for future data mining algorithms

Perform clustering algorithms on constructed graph

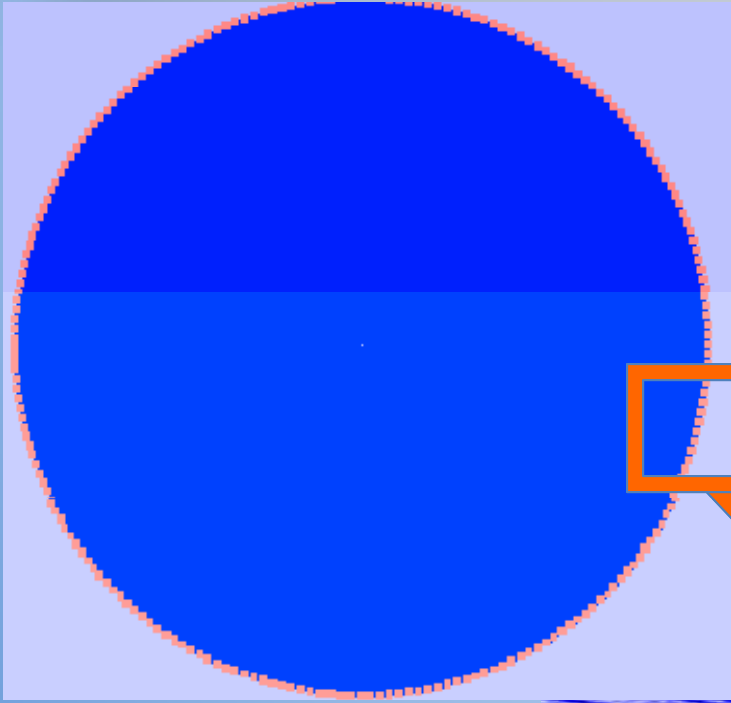
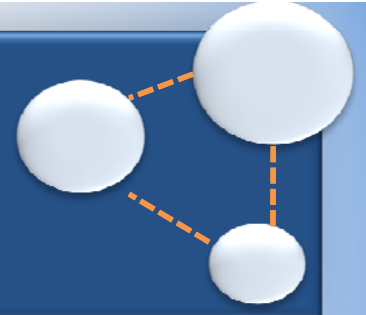
Predict the future socio-economic statuses and international relations of countries

Research Methods

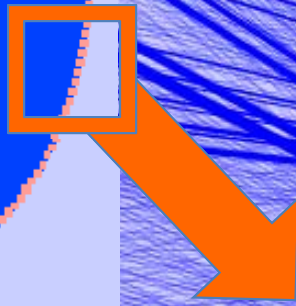




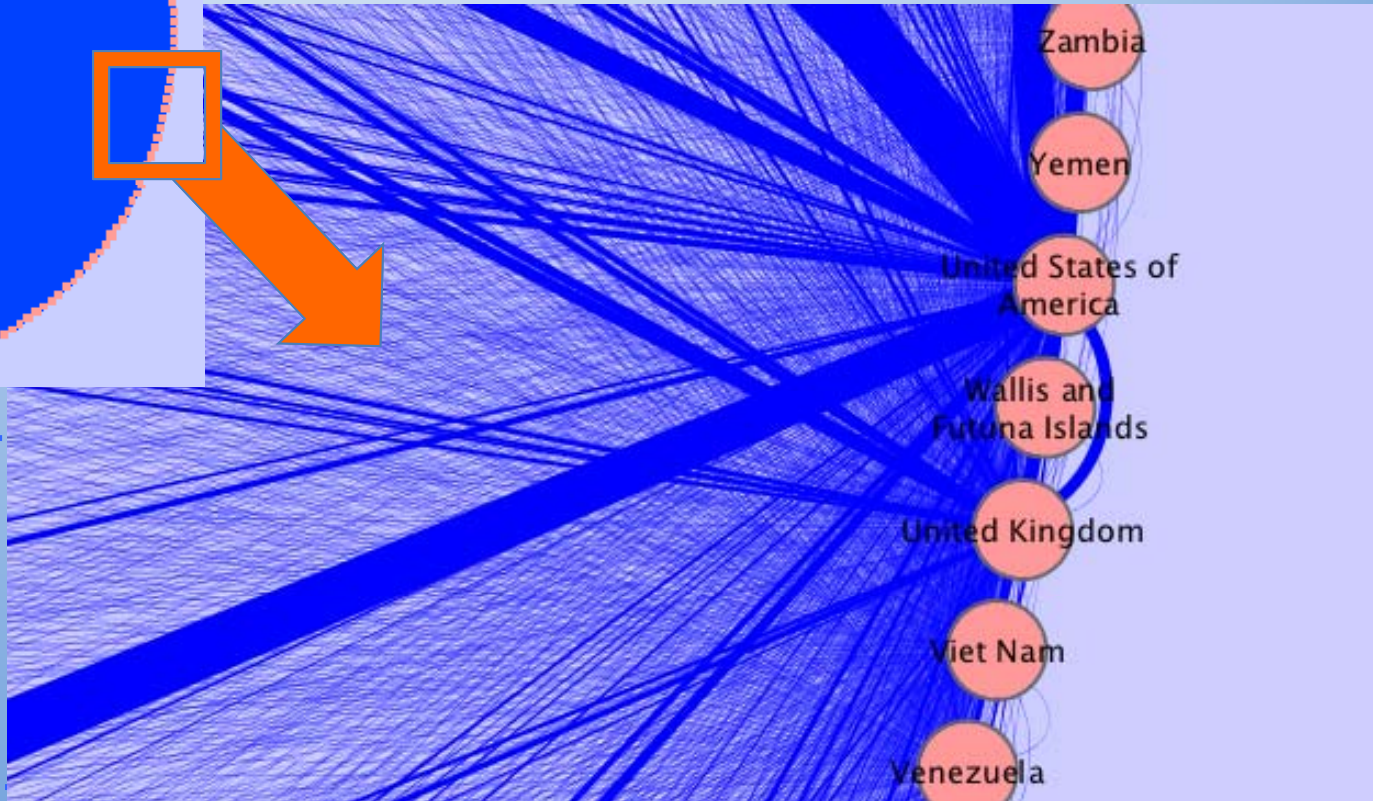
Graph Results



← Large scale image of time-evolving country graph

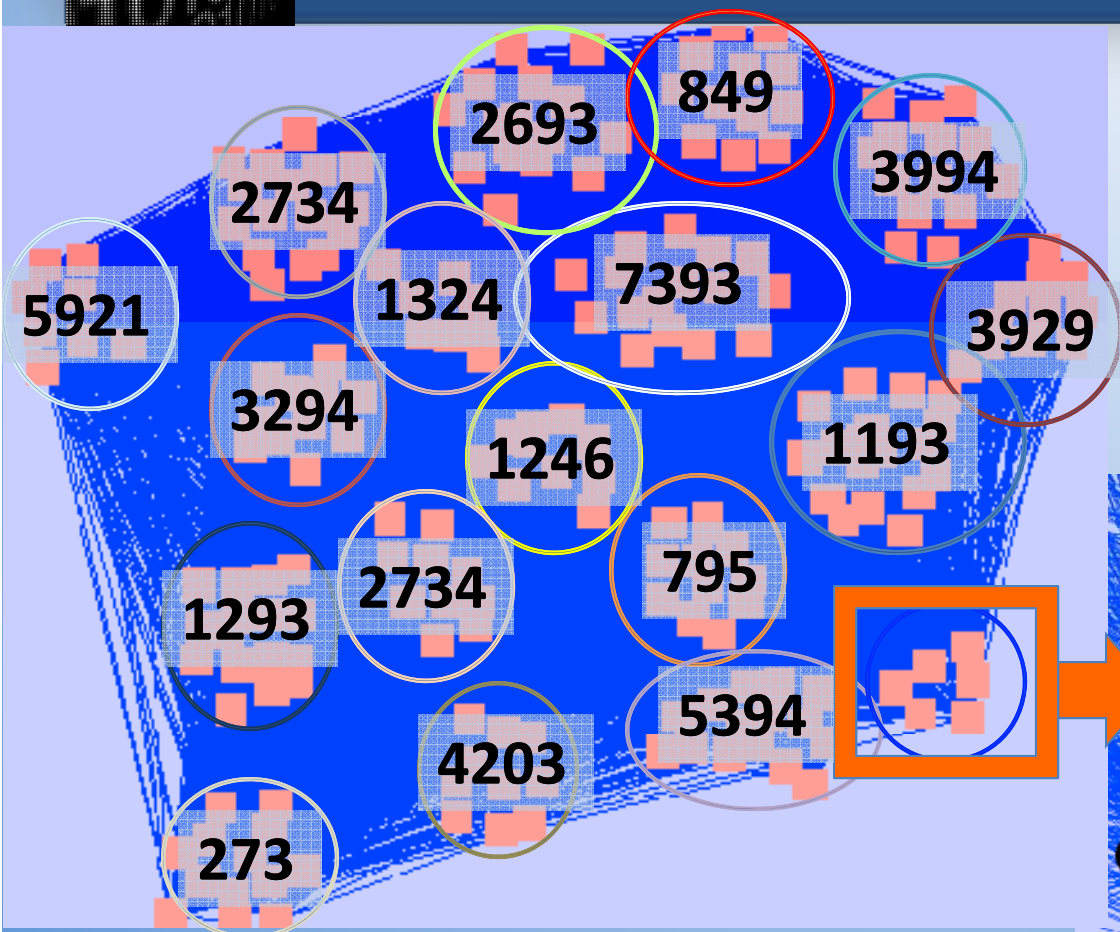
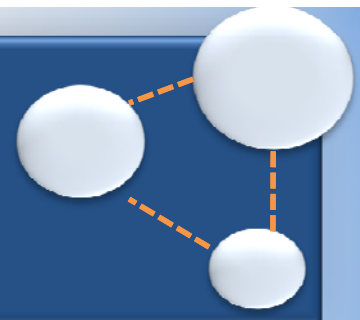


Enlarged image of time-evolving country graph



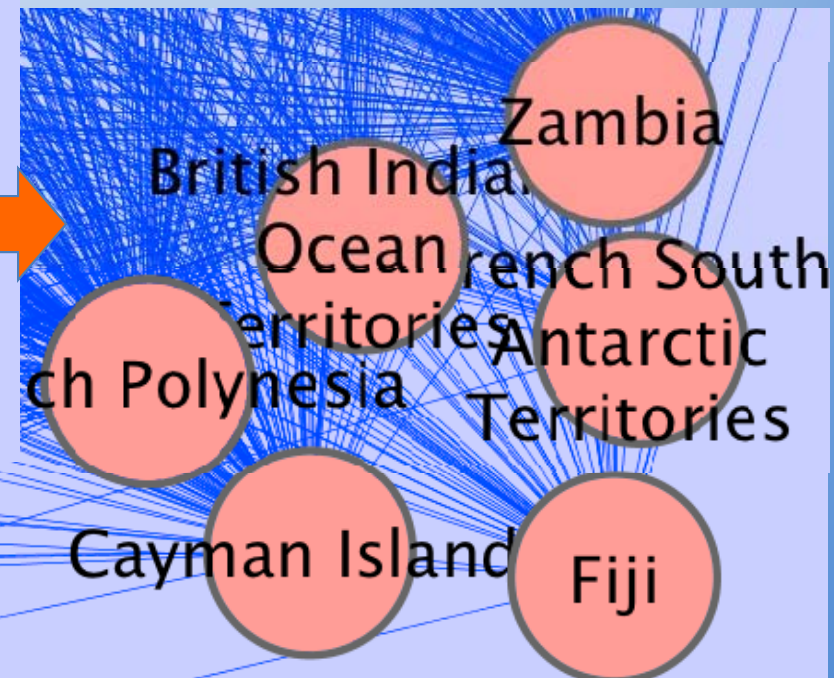


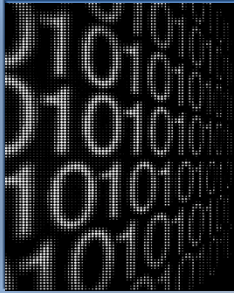
Data Mining Results



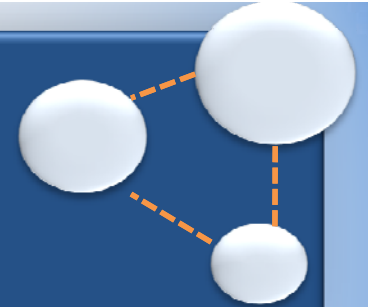
← Clustered time-evolving country graph

Enlarged image of clustered graph →





Research Results



Pearson Correlation Coefficient
(between the average physical distance of
the countries in the clusters, and the
average cluster score)



-0.2



Data is
significant!

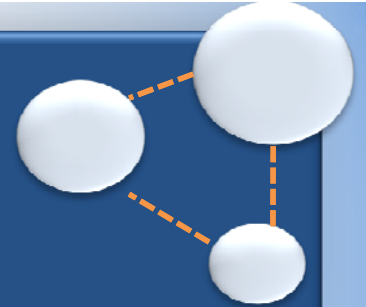
(significance is anything that has an
absolute value greater than ~ 0.07)

Conclusion

Since the cluster score increases as the
average distance between countries
decreases, greater amounts of trading
takes place between countries that are in
close proximity to one another



Future Work



Additional hypotheses to verify through data mining

Do smaller countries tend to trade in close proximity?

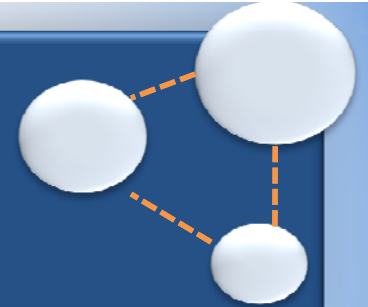
Based upon their current trade values, can a model be built to predict how clusters of countries will form in the future?

What other attributes of a country determine its trading values (i.e. population, geographical placement, etc...)?

Do countries tend to stay within their own clusters?



Summary



Time-Evolving Graphs

Nodes

Data objects that we want to know more information about

Edges

Some common factor between the data objects that changes value over time

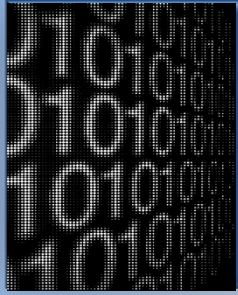
Data Mining Tasks

Predictive Tasks

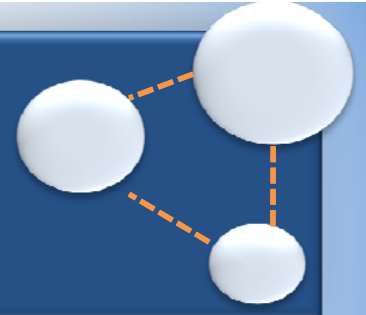
Models that can predict future events within the graph based upon its attributes

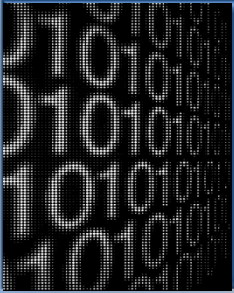
Descriptive Tasks

Deriving patterns that summarize the underlying relationships between the data objects

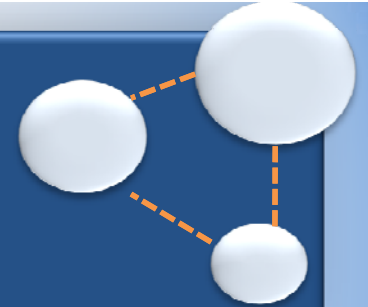


Questions



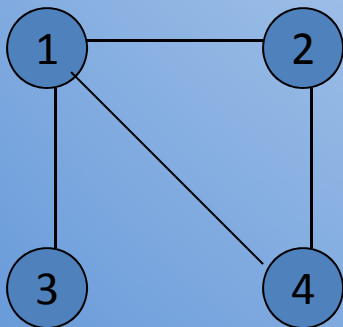
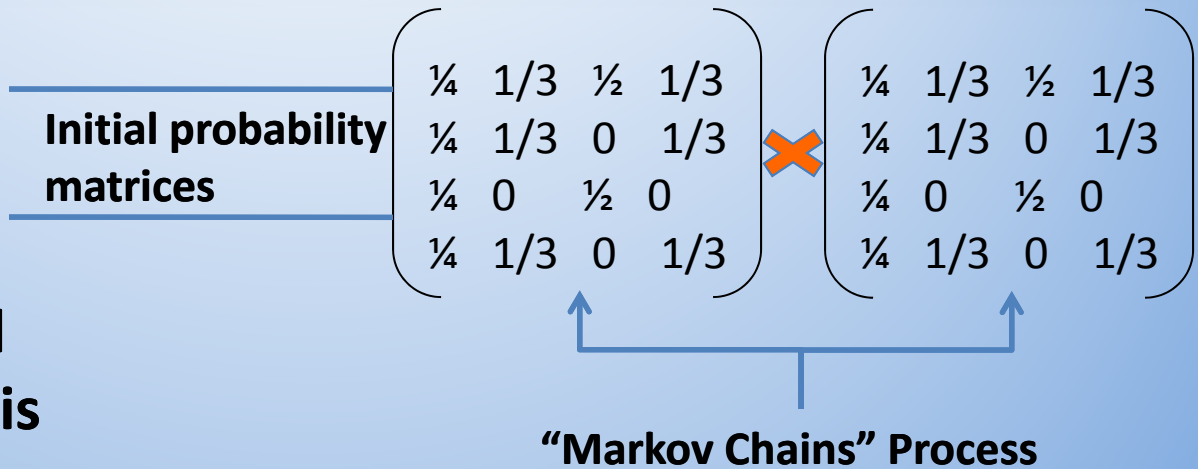


TopGC Clustering

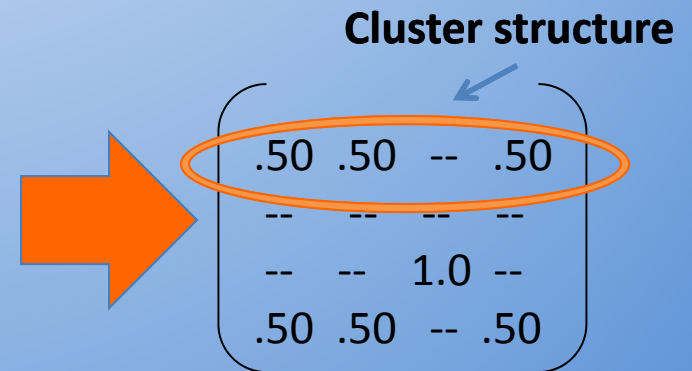


Basic Idea:

Build a probability matrix and perform random walks using "Markov Chains" until the converged matrix is obtained.

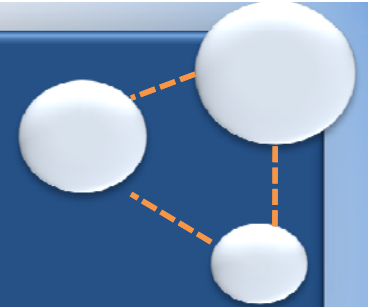


Will eventually equal the converged matrix where possible cluster structures may be seen

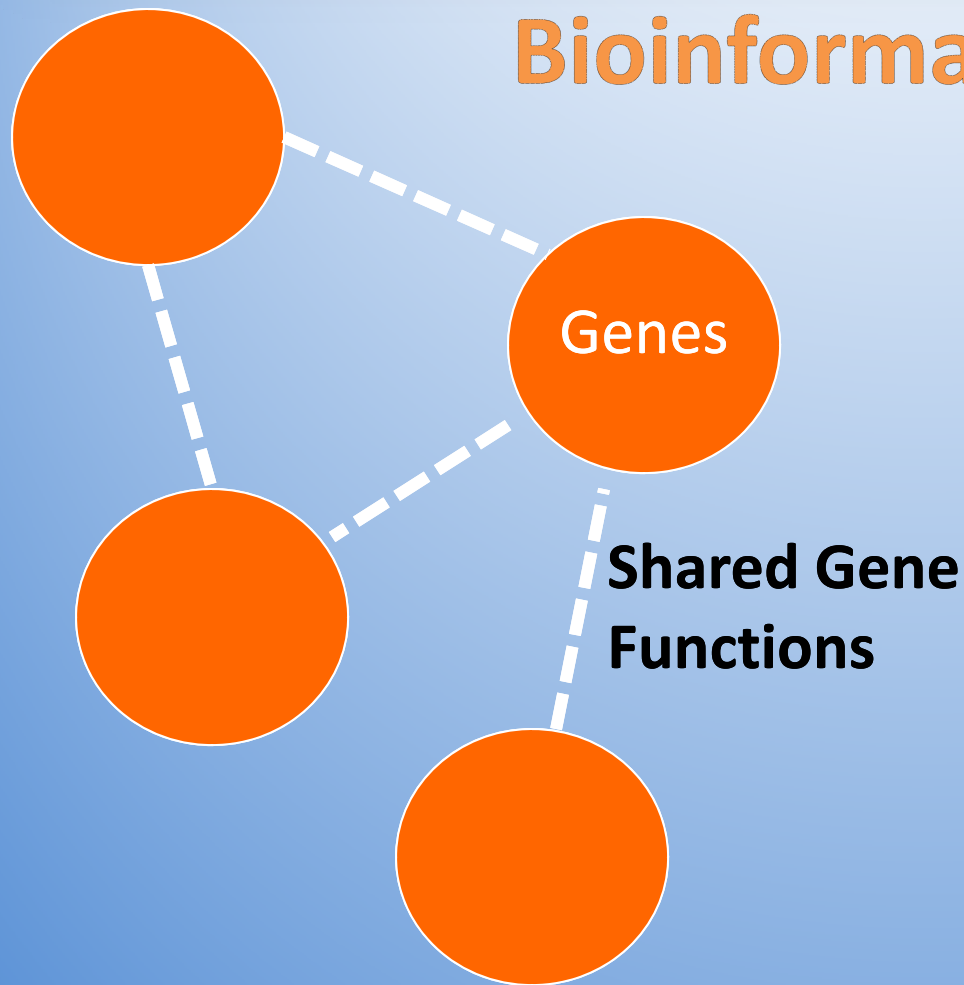




How This Relates To Nanotechnology

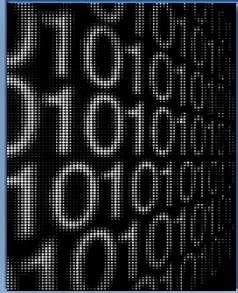


Bioinformatics Graph

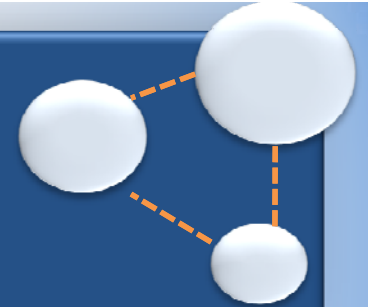


Main Purpose

Predict an unknown gene's function by looking at what other genes were clustered together with it based upon known shared functions



Future Work



Further Data Mining Analysis

Build predictive models for graph based upon its attributes

Test to determine if there are any connections between properties of the countries within the clusters and how those clusters form

Granger Causality Coefficient

Statistical method which measures how likely it is that one property causes another, across time

Do changes in population drive changes in trade or cluster size?