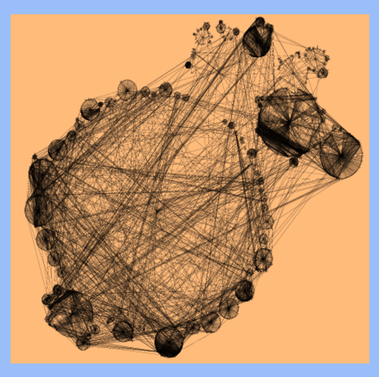
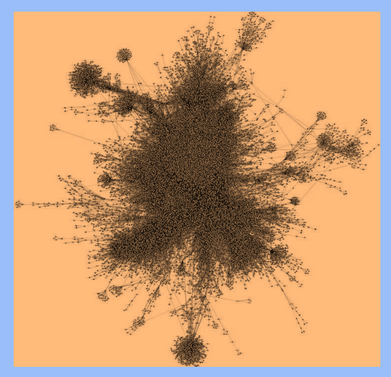


Data Mining and Modeling Of Time-Evolving Graphs

Name: Luke Gebauer
College: Allan Hancock
Major/Minor: EE/Computer Science
Mentor: Kathy Macropol
Advisor: Dr. Ambuj Singh
Department: Computer Science



Introduction

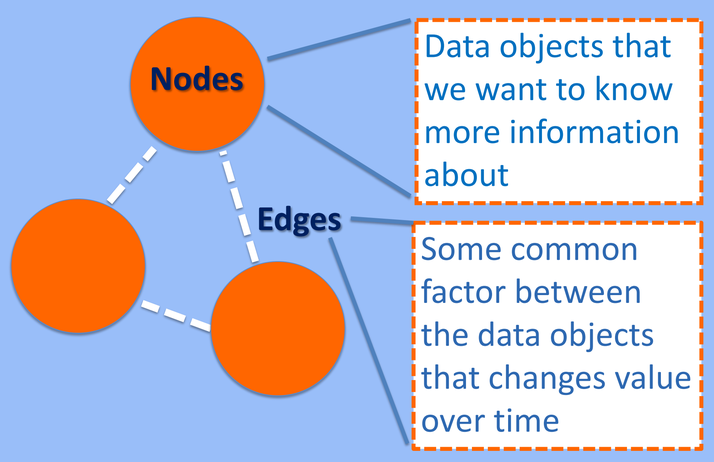
Being able to effectively store and analyze large and complex datasets is key a component in many of today's scientific and industrial processes and can often lead to startling new discoveries. However, traditional forms of data analysis are often inadequate to handle such datasets. Therefore, the processes of data mining and modeling of time-evolving graphs was developed in order to efficiently store and analyze complex datasets that change over time.

Objectives

- Construct a time-evolving graph using countries as its nodes and international trade values as its edges in order to provide the field of Computer Science with additional resources in this new and limited area of research.
- Perform clustering algorithms on the constructed graph in order to analyze and gain information about how countries relate to one another through international trading.

What is a time-evolving graph and data mining?

Time-Evolving Graphs



Data objects that we want to know more information about

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

Data Mining Tasks

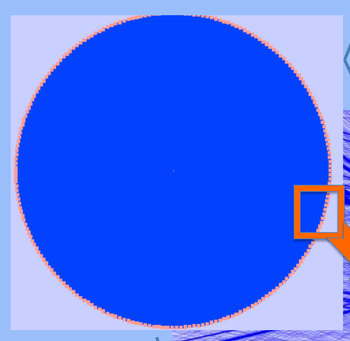
Predictive Tasks

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

Descriptive Tasks

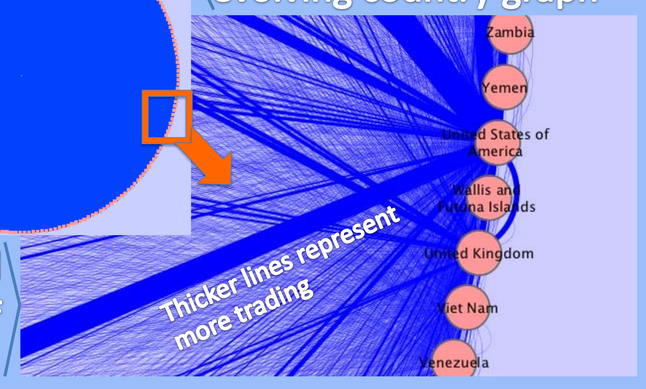
Deriving patterns that summarize the underlying relationships between the data objects

Graph Results



Large scale image of time-evolving country graph

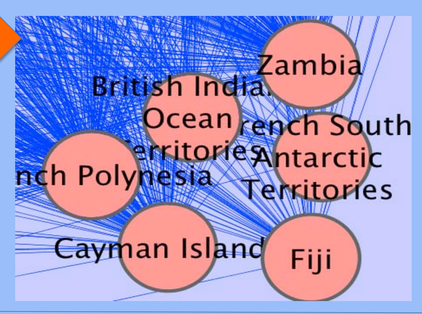
Enlarged image of graph



Clustering Results

(Using TopGC clustering algorithm)

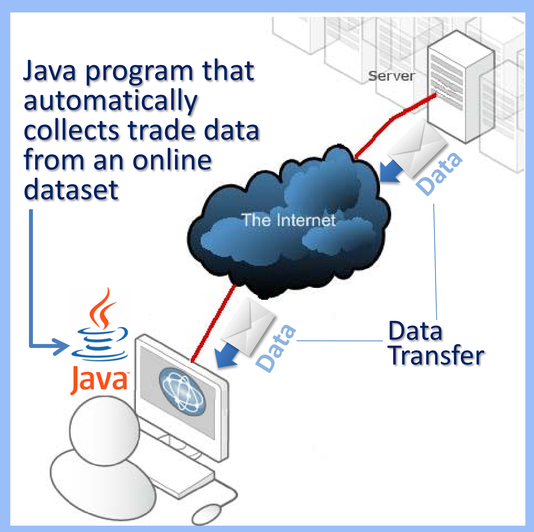
Representation of clusters and cluster scores (i.e. the average trading amount within the cluster)

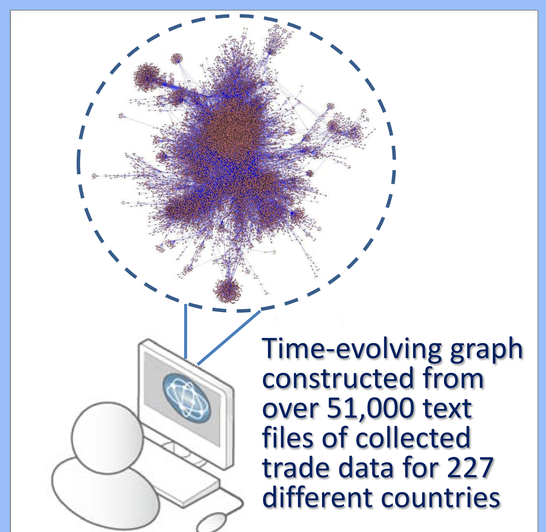


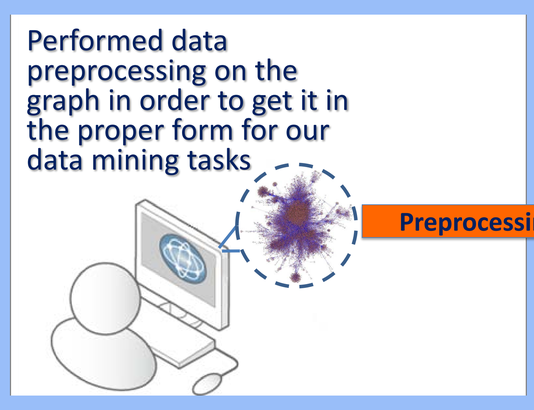
Clustering was based upon node edge weights and connectivity

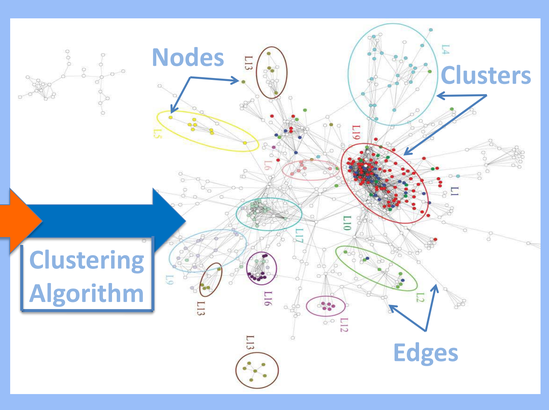
Research Methods

(What we did and why we did it)









Research Results

(Two different ways to interpret cluster results)

1

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

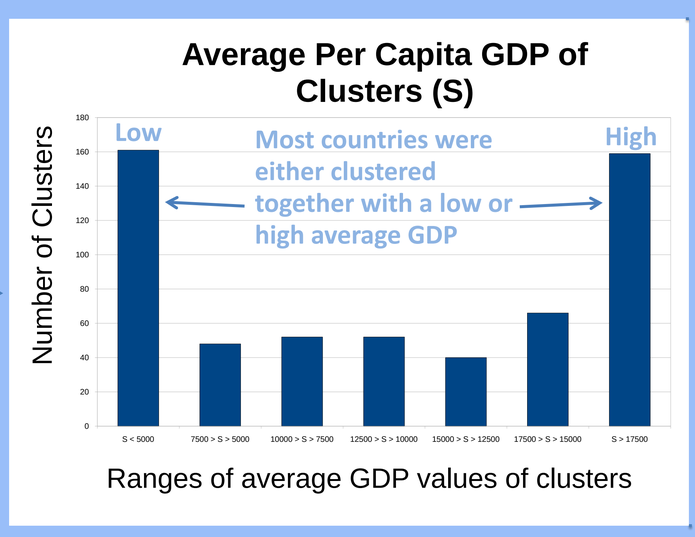
=

-0.2
(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

2



Conclusion

Since most of the clusters have either a low or high average GDP as shown on the graph, this indicates that successful countries tend to trade with other successful countries while less successful countries tend to trade with other less successful countries.

Future Work

Additional hypotheses to verify through data mining

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