

Mason McBride

30 March 2022

Professor Marta Gonzalez

Data Science for City Planning

Assignment 3: Project Preparation

Project Preparation Guidelines

Step 1)

Network: A directed weighted graph of the international imports from each country where imports are partitioned into their country of origin in US Million.

Nodes: [countries], the number of nodes is 212

Edges: [export], the number of edges is 1060

Step 2)

There is a related paper titled “[Scale free is not rare in international trade networks](#)” which describes the same network I constructed above. It describes fine-grained analysis on specific products of total trade that are scale-free networks, like the ones formalized by the Barabási–Albert model. My network analysis focuses on international trade as a whole, and trying to analyze the scale-free properties of the entire network as well as the emergence of sub-clusters in the import-export model

Step 3)

- a) I had to construct by network with data from World Integrated Trade Solutions (WITS) which is software made by the World Bank that allows people to query for international

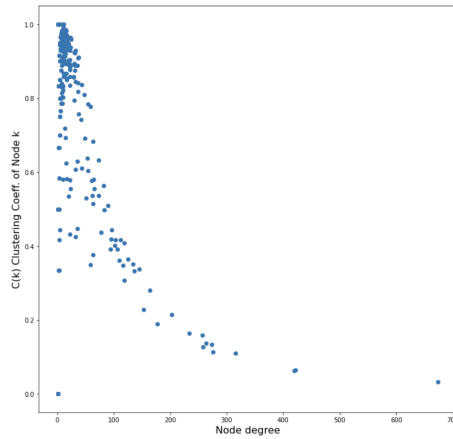
trade data. I then cleaned the data to generate a networkx DiGraph that iterated over a dataset that contains an entry from each country to its listed import countries with the weight on the edge coming from the total import cost in US dollars by million.

- b) The Nodes just have a “country name” attribute, and the edges of each node represent both the existence of a trade agreement between the two node countries with the amount of money included as the weight of the edge as well.
- c) The physical distance of the links do not have meaning. But in a weight directed graph like this, the weighted degree of each node roughly corresponds to the influence a node has on its neighbors, which is a type of distance
- d) I want to incorporate the concepts of ‘scale free networks’ and ‘small world clustering’ analysis onto this dataset because according to the above paper international trade is a network with preferential selection between nodes and edges in the network due to the fact that countries with lots of money can leverage that fact and import more goods compared to smaller countries who don’t have the money to import the same transformative goods.

Analyzing Weighted Networks

1. Out of the 850 counties listed Alameda County, San Francisco County, Santa Clara County, Contra Costa Country, Los Angeles Country, Orange County, Riverside County, San Bernardino County, San Mateo County, and Ventury Country are all counties with over 50,000 commuters. This is a total of 10 California counties.

- The scatter plot of the clustering coefficient of each node versus its degree is graphed



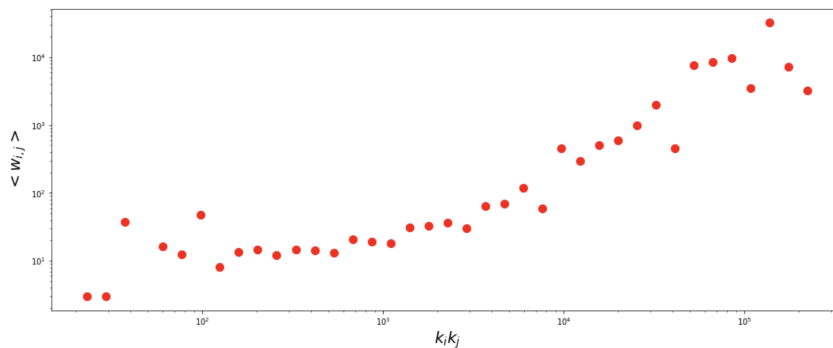
below.

This graph follows a power law as node

degree increases which represents the natural property of this network that has small connected nodes contributing to the clustering coefficient.

- The plot of the loglog plot of the disparity versus the degree of each node is given below.

This graph represents how accurate a given approximation for analyzing clustering properties of networks is.



On a log log scale,

what should be visible is a constant increase in this accuracy (or, disparity) as the degree of the node increases since it's harder to guarantee larger nodes. Small nodes also have a lot disparity implying this model is accurate for nodes in this area.