CS 89 Final Project

Alina Chadwick, Isabelle Kitchel, Nick Krivov, Julia Martin

March 2023

1 Abstract

For our final project, we use graph theory as a framework to study international trade dynamics in 2021. Using the UN Comtrade Database, we create an undirected graph of international trade with edges weighted by import/export value. Then, we study the graph properties of the international trade network such as centrality, average shortest path length, largest connected components, and the number of triangles. We also analyze whether the international trade graph follows the power law and we create a visualization of the network. We find that international trade is a highly connected graph and does not follow graph laws such as the power law. Finally, we provide analysis of our findings and discuss potential explanations.

2 Introduction

During the COVID-19 Pandemic, it became evident that international trade is heavily concentrated between certain countries. This resulted from accelerating globalization in recent years and 'just-in-time' manufacturing. With globalization came trade dependencies that allowed companies to become more efficient and source parts of their products from different countries. These trade dependencies caused a lack of diversification in supply chains, resulting in large, long term shortages of goods ranging from clothing to semiconductors. Russia's invasion of Ukraine exacerbated shortages and supply bottlenecks contributed to rising global inflation.

For this study, global trade was modeled as an undirected graph in order to analyze graph theory properties in relation to trade and potentially gain insights into how trade is concentrated. In the graph, nodes are countries and edges are weighted by the amount of trade (imports are multiplied by -1 and exports are positive, in US dollars) between two countries. This graph framework allows for the study of whether the network of global trade follows the graph properties that were taught in class. This project seeks to answer the following questions: does the global trade network follow the power law? What are the main connected components, and which countries are in them? What are the various centrality values for countries and what can they reveal about the structure and interconnectivity of global trade? Is global trade similar to a 'Small-World' graph in terms of path length rule? Answering these questions is a very interesting application of the material covered in class.

Overall, international trade networks are an interesting and relevant application of graph theory. Free trade is important to the world economy so it is important to understand what dependencies exist so that the world economy is robust against economic shocks.

3 Challenges

Our approach to modeling global trade faced a number of challenges. One primary obstacle was the need to access and clean data in order to conduct our analyses. Accessing data from the UN Comtrade database was not completely straightforward, as there are many ways to query this data and create a working dataset. We also had to use regular expressions in a spreadsheet editing software to filter out certain rows where the trade partner was reported as "World." We initially planned to conduct a pre- and post-COVID analysis, but the 2019 data required for this was not available. It was important to consider the quality of the data and the impact it could have on the accuracy of the analysis - the interpretation and validation of results with a large and complex dataset were challenging. Using directed graphs also posed its own set of challenges. Calculating the weight of edges that represented trade volume between countries in relation to imports and exports, for example, required us to think outside of the box and use positive and negative weights. Additionally, there were factors that introduced noise into our data. For example, some entries had multiple trade partners that were not represented in the graph, which could skew the results. We also found that the global trade network consisted of only one connected component, resulting in a completely connected graph - our analysis was limited to reported trade to the United Nations, which naturally cannot reflect all of the trade occurring in the world. Despite these challenges, we put in our best effort to overcome all obstacles and accurately model and analyze global trade patterns using our chosen methods.

4 Related Prior Work and Its Shortcomings

There are a variety of works focused on analyzing nations' import and export patterns. However, the majority of these papers center around a specific country or a specific good, with few focusing on the global trade network as a whole.

Mafakheri et al's 2023 work is an example of a paper that focuses on one specific product in the global trade exchange: petrochemical [1]. Using a computational method based on link prediction, this paper predicted both import and export of global petrochemical trade, including the establishment and cancellation of this good. The results were then represented and analyzed in bipartite network graphs. Mafakheri et al does a strong job establishing prediction mechanisms for petrochemical trade but that paper is restricted in focusing only on bipartite graphs.

An example of a related paper that centered its analysis around specific countries is Lui et al's 2018 paper, "Economic Benefits and Environmental Costs of China's Exports: A Comparison with the USA Based on Network Analysis." The paper utilizes a basic multi-regional input-output model to represent the relationship between countries and sectors as matrices [2]. While this paper does a good job identifying patterns in trade and environmental impacts, it does not extrapolate such findings to determine how trade should be changed or potential supply chain impacts.

5 Data

The United Nations Comtrade Database is a collection of global trade statistics from 200 countries, constituting more than 99% of the world's merchandise trade [3]. The data can be analyzed either annually or monthly, and has various setting options to determine the year(s) of focus, trade partners, the type of trade flow, and even the type of goods. For this project, we focused on world trade in 2021. The dataset is organized by row entries that contain: "reporter" country, "partner" country, "trade flow," and "trade value." The "trade flow" metric will have either "import" or "export" as its entry value. "Trade value" is a numeric entry that signifies the net value of the trade exchange in USD.

6 Proposed Method

Building the Graph:
The graph was built by iterating through the row entries of the database, establishing the countries as nodes and connecting weighted edges between the respective trading countries with the trade value of the exchange. The types of trade flows are incorporated in the graph via the weighted aspect of the connected edges. If the trade flow was an export with respect to the reporter country, the weighted edge was positive based on the trade value. If the trade flow was an import, the trade value was multiplied by -1 to create a negatively weighted edge.

Visualization:
The graph visualization was achieved via the draw networkx function. The blue nodes represent the trading countries. Black edges connect the nodes that participate in trade together. All nodes should have a degree of at least one, since the countries included in this data set participate in global trade and thus have at least one trade exchange.

Centrality:

Various centrality measures were taken on the international trade undirected graph: degree, closeness, and betweenness centrality. The networkx built-in centrality functions were used and then the output was manipulated to sort the countries by highest to lowest centrality values.

Connected Components:
We identified both connected components and strongly connected components in the graph using networkx built-in functions. They are designed to output a list of sets that contain all nodes (countries) in a connected component or strongly connected component. The strongly connected components function only works on directed graphs, so we decided to convert our graph to a directed graph for that metric.

Average Shortest Path Length:
To evaluate how connected world trade is, another metric we decided to use was the average shortest path length between nodes. The average shortest path length gives insight as to how disconnected nodes are from one another. The networkx average shortest path length function was utilized here to determine that value.

Number of Triangles:
The number of triangles in the world trade graph would indicate whether the graph is truly interconnected or whether there just might be one or two key hubs. We used the networkx num triangles function to determine that value.

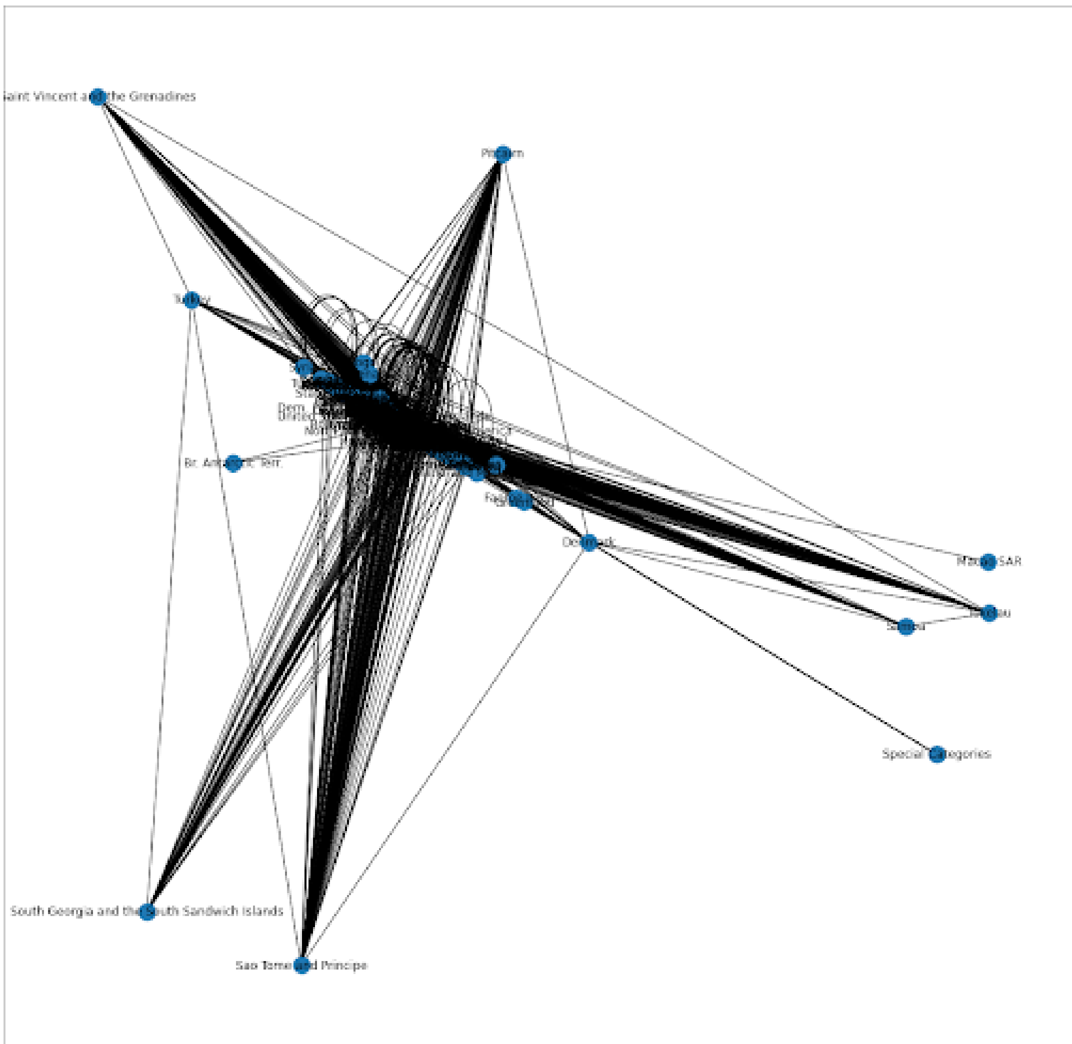
Distribution Plots:
Finally, in order to identify whether the world trade model follows the power law, we aim to create a PDF plot, a CCDF plot and a Zipf plot. There are no networkx built-in functions that do so effectively, so we plan on implementing the code from scratch and analyzing our results.

7 Experiments

Visualization:
Figure 1 shows the basic graph visualization for the 2021 dataset. Every node has an edge connecting it to at least one other country node, which is supported by the data. The location of nodes on the graph is randomized and changes each time the code is run. There is therefore no significance to certain countries being clustered around the center versus the outside of the graph.

Centrality:
See Table 1 for the 10 countries with the highest degree, closeness, and betweenness centrality values. Degree and closeness centrality yielded similar results, giving mostly European Union countries with high degree and closeness centrality values. Here, degree centrality relates to the idea that important countries

Figure 1: Visualization



for the international trade network have many connections. Similarly, closeness centrality measures how important countries will be close to other important countries in the international trade network. Given these interpretations, it makes sense that these measures are dominated by EU countries because they are highly interconnected, making them important to the trade graph. Betweenness centrality shows how important countries are highly connected. This measure differs from degree and closeness because it is more akin to how much of a hub a country is. Here, it makes sense that China is ranked first because China is a major export hub. It also makes sense that Japan is ranked second because Japan is a major importer.

Table 1: Top 10 Centrality Countries, 2021

	Degree		Closeness		Betweenness	
1.	Ireland	0.971311475	Sweden	0.9644268774	China	0.0247239401
2.	Sweden	0.963114754	Spain	0.9644268774	Japan	0.0109348990
3.	Spain	0.963114754	Poland	0.9644268774	Peru	0.0094828961
4.	Poland	0.963114754	Ireland	0.9644268774	Israel	0.0084235038
5.	Belgium	0.963114754	Belgium	0.9644268774	Other Asia	0.0083285954
6.	Thailand	0.959016393	Netherlands	0.9606299212	Ireland	0.0064446114
7.	Netherlands	0.959016393	France	0.9568627450	Belgium	0.0064446114
8.	South Africa	0.954918032	Czechia	0.9568627450	Sweden	0.0062769777
9.	France	0.954918032	Thailand	0.9531250000	Spain	0.0062328962
10.	Czechia	0.954918032	Germany	0.9531250000	Poland	0.0062328962

Connected Components:
In both identifying connected components and strongly connected components, the results showed that the world trade graph is one connected component and one strongly connected component. This implies that all international trade is interconnected, at least in this dataset. The hypothesis that trade is tightly connected is supported, but rather than having subgroups of countries, it is an interconnected web. This is not a surprising finding, as hub trade countries have many partners, and almost every country is connected to a hub.

Shortest Path Length:
When calculating the average shortest path length between nodes, we came to a result of 1.4. In our graph, the implication is that every country is on average 1.4 countries away from another. Given our findings on connectivity as the graph is very interconnected, it is not a surprising finding that the path length between nodes is, on average, relatively small.

Number of Triangles:
When calculating the number of triangles in the graph, we found that there are 1.8544754561311632783653256751480832* 10e22 triangles. This also supports our hypothesis of connectivity, as there are relatively very many triangles, which implies that the graph is very inter-connected.

Distribution Plots:

PDF Plot
Figure 2 does not follow the power law as the plotted points are not aligned with the best fit line.
CCDF Plot
Figure 3 shows that the graph follows the power law.
Zipf Plot
Figure 4 also shows that the graph follows the power law.

8 Conclusion

In conclusion, this project applied a network framework to international trade and found that the international trade graph is highly connected. This is evidenced by the fact that international trade has only one connected component, a very small average path length, and a very large number of triangles. The low shortest average path length shows that countries should be able to work around supply chain bottlenecks when geopolitical issues arise because each country is only 1.4 steps away from any other country. Moreover, the high connectivity of international trade reveals that it is not impossible to diversify trade and avoid supply chain shortages.

Centrality measures revealed which countries are important 'hubs' and which are highly connected. Interestingly, countries in the European Union had the highest centrality measures, likely because the EU is a region with free trade between countries, and in terms of trade, often is compared to operating like the states in the US. China had the highest Betweenness Centrality, which makes sense given how important China is to the global economy as China is a major exporter and is the world's second largest economy by nominal GDP.

This study also found that the international trade graph does not follow the power law. However, this makes intuitive sense because international trade is not random: there are many economic, geographic, and political factors that influence countries to trade with other countries, or not. For example, tariffs and import taxes aimed at political ideologies or in reaction to how a foreign government may hinder trade. In contrast, long term partnerships and alliances may cause some countries to trade disproportionately more than they naturally would without policies in place.

Finally, the high connectivity of international trade makes sense when one considers how certain countries often specialize in exporting certain raw materials or goods depending on the country's natural resources. A country may be

Figure 2: PDF Plot

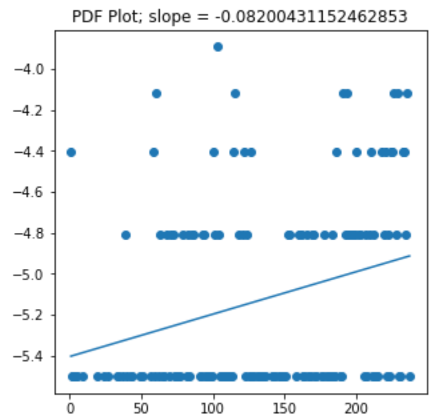


Figure 3: CCDF Plot

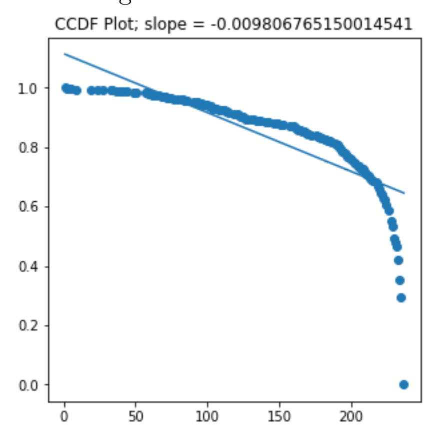
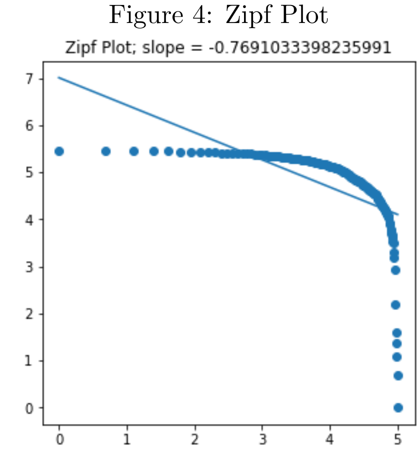


Figure 4: Zipf Plot



a major exporter of a certain product without being a major global exporter. This project confirms the effects of increased specialization and globalization in the past few decades: countries are highly interconnected. While high interdependencies pose risks when geopolitical or economic shocks occur, the interconnectivity of trade also means that countries can be flexible in the long run to reconfigure import and export patterns and mitigate risk.

Overall, this project was a great opportunity to apply the theories we learned in class to a real-world problem. We learned how to work with large data and build a graph from scratch and how to come up with our own analysis. Furthermore, we interpreted our results in context of our initial query.

9 Division of Work

Alina - 30 %, Isabelle - 30 %, Julia - 30%, Nick - 10%

10 Link to Code

https://colab.research.google.com/drive/1bENcDUlZpqPaatSex7X$_5 mOmwNwm7WWW?usp = sharing$

11 References

[1] Aso Mafakheri, Sadegh Sulaimany, Sara Mohammadi. 2023. Predicting the establishment and removal of global trade relations for import and export of petrochemical products. https://doi.org/10.1016/j.energy.2023.126850.
[2] Ya Liu, Yuhuan Zhao, Hao Li, Song Wang, Yongfeng Zhang, Ye Cao. 2018. Economic Benefits and Environmental Costs of China's Exports: A Comparison with the USA Based on Network Analysis. https://doi.org/10.1111/cwe.12251
[3] The United Nations Comtrade Database https://comtrade.un.org/
[4] Lucas Lacasa, Bartolo Luque, Fernando Ballesteros, Jordi Luque, Juan Carlos Nuño. 2008. From time series to complex networks: The visibility graph https://doi.org/10.1073/pnas.0709247105
[5] visibility-graph 0.4.1 https://pypi.org/project/visibility-graph/