CS 224w Project Report: International Trade as a Network of Interconnected Nation-states

Roneil Rumburg and Chris Guthrie

Abstract

We give a complete characterization of the distribution of exports and imports across countries for the past 25 years. In addition, we characterize the distribution of the deviation of exports and imports from two simple trade models: one where trade between country A and country B is proportional to the product of their populations, and another where it's proportional to the product of their GDPs.

Project Introduction

International relations consist of many different types of interactions between cohesive states. This leads to a natural model of international politics in terms of a network with countries as nodes and edges representing these interactions. We're analyzing one specific type of interaction, international trade, as a network to identify trends over time in the characteristics of global trade. In this network, each directed edge is weighted proportionally to the exports or imports between the two countries.

The network is useful in understanding relative economic influence between states - the more a country exports, the more global economic power it has. So, we can treat the weights at each node as a meaningful statistic for economic power, and we're interested in how the distribution of these weights changes over time. This distribution, taken as a snapshot, indicates roughly how skewed power is toward certain countries, as well as how globalized world trade is. Insight into the nature of the distribution has led us to a network-based model of trade via a preferential attachment mechanism, because the trade distribution was fit best out of the many distributions we examined by a log-normal distribution, meaning that a small set of nodes has significant influence in the network and strongly suggests that nodes preferentially attach to influential nodes. Looking at how the network changed over time has also led to the discovery of some interesting insights about the nature of international trade, which will be explored further later in this paper.

Prior Work

We discuss two important papers that are relevant to our project:

Paper 1: A Brief History of Generative Models for Power Law and Lognormal Distributions

Michael Mitzenmacher. (2003)

Mitzenmacher's paper on the different uses of power law and log-normal distributions served as a basis for our model of preferential attachment in the network of international trade, allowing us to justify modelling the network we observed using both power-law and log-normal distributions.

Mitzenmacher found that both distributions are inherently very similar to one another, and can be used to model similar situations. Mitzenmacher explained the differing generative models that can lead to the generation of data representable by one or both distributions, and also explained some of the underlying data properties that can lead to data being better represented by one distribution or another. He states that similar generative models can lead to both power law and log-normal distributions depending on "seemingly trivial variations," leading one to reasonably conclude that preferential attachment can be explained by both a power law distribution and a log-normal distribution. We look at the fit of both distributions, and Mitzenmacher's results will be revisited when analyzing the fit of each one to our data set.

Paper 2: The International Trade Network: weighted network analysis and modeling

K Bhattacharya, G Mukherjee, J Saramäki, K Kaski and S S Manna

J. Stat. Mech. (2008)

This paper explores the connection between the GDP of a pair of countries and the amount of trade between the countries. It considers the distribution of trade volumes between countries and shows that total trade volume is exponentially related to a country's GDP. The paper discusses the gravity model of international trade as an explaining factor for this phenomenon. The paper examines a rich-club effect, where a small community of nations controls a large proportion of international trade, and observes that the size of this club is declining. The primary result of the paper is that the distribution of trade volumes is log-normal and invariant over time.

Bhattacharya et al. suggest that the distribution of trade volume over time is log-normal, and does not change over time, and we re-visited Bhattacharya's findings, confirming that a log-normal distribution is a better fit than a power law distribution but finding that the distribution of trade volume over time changed significantly, which will be explained further in the results section. This contradiction may not be meaningful, however, because we analyzed a slightly different object - while Bhattacharya treated trade as symmetrical (i.e. an undirected graph where edge weights are exports plus imports), we treat trade as asymmetrical (a directed graph where an edge from A to B is weighted by A's exports to B).

Model

For a given year we model our network by a directed graph G , leading us to produce 25 distinct graphs, each representing the network of trade for a specific year in the interval [1988, 2012]. Each directed edge E represents a flow of goods between two countries, with edges going from source of goods to destination of goods being exports and edges. This means that incoming edges for a node N are imports and outgoing edges are exports. We compute the weight $W(N)$ by adding together the weights of all outgoing edges from N . For each node N in G, we also save the GDP and population of the country during that year. For a given graph G, we want to determine the probability distribution of W . For all the networks we considered, we tested the following types of distributions:

- Normal
- Log-normal
- Power law

We choose these distributions because they often arrive in the context of highly complex real-world systems. Normal and log-normal distributions arise from properties that are the sum and product (respectively) of a large number of random variables, and power law distributions, as we saw in class, arise from preferential attachment processes. The fit of these distributions will be explained further in the results section.

There is a slight problem with both our theoretical model as described up to this point and the model described by Bhattacharya et al., which led us to devise a novel method of trade network analysis. From a very simple model of the world (where countries have equal economic resources and technology), we would expect trade volume between countries to be related to the population of the countries and/or the GDP of the countries.

Based on our recognition of this problem, our primary contribution in this project was a novel algorithm to re-weight edges in the graph during distribution calculations based on metadata about nodes, allowing us to look at the influence of other parameters on the weight distribution. We looked at two specific pieces of metadata, country population and country GDP, and used these to generate two new graphs alongside of each original graph, each of which had its edges divided by the product of either the source and destination node’s GDP or population, controlling for changes in these variables. These networks turned out to be quite different from the network analyzed by Bhattacharya, as shown below in our results section.

Statistical methods

For each year, we computed six datasets, three each for imports and exports, based on three different edge weighting schemes:

- Total imports/exports
- Population-weighted imports/exports
- GDP-weighted imports/exports

Each data point represented the sum of all outgoing / incoming weights for a given node. For population and GDP weighting, we scaled the total imports and exports inversely by the product of the populations / GDPs associated with the endpoints of each edge.

For each dataset, we used maximum-likelihood estimation to find best fit parameters for a normal distribution, a log-normal distribution, and a power law distribution.

For each fit, we estimated the strength of the fit using the Kolmogorov-Smirnov test to compare the dataset with a random sampling of data points from the best fit distribution.

Tools

All graph manipulation was done through the Python library Snap.py. We implemented maximum-likelihood estimation using NumPy and SciPy, generated a random sample using from the best-fit distribution using NumPy, and executed the Kolmogorov-Smirnov test using a SciPy library function.

We analyzed trends in best-fit parameters and created time-series charts with the statistical language R.

Results

Imports vs exports

We focus our analysis in this section exclusively on export data, for simplicity. In our methods, however, we performed the same analysis on import data. We reached nearly identical conclusions.

Best fit model for weight distribution

We identified the distribution of weights (i.e. exports) across countries as log-normal with strong probability ($p > .08$ [1] for the best fit distribution) for all years in our dataset. This is consistent with Bhattacharya's findings. On top of that, the log-normal fit was just as strong when we

[1] Here, *p* is a statistic output by the Kolmogorov-Smirnov comparison method run on the actual data and the best-fit model. Although it should ideally represent the probability that the dataset is drawn from the best-fit model, since the best-fit model was initially derived from the dataset via MLE, *p* will systematically overestimate this probability.

adjusted exports for the GDP of both countries. We found that a power law distribution did not accurately describe the entire dataset, but appeared to fit the tail end of the export distribution - i.e. the top 60% of companies in terms of total exports or GDP-adjusted exports - with $p > .10$ for most years. Population-adjusted exports had no good fits between normal, log-normal, and power law models, although the tail end of population-adjusted exports fit a power law distribution very well ($p > .20$ for most years).

Strength of Fit

	Model		
Variable	Normal	Log-normal	Power law
Total exports	No fit	Strong fit	Tail-end fit
GDP-adjusted exports	No fit	Strong fit	Tail-end fit
Population-adjusted exports	No fit	No fit	Tail-end fit

For all variables, the tail-end power law fit suggests that trade is based on a preferential attachment or rich-get-richer mechanism. Although traditionally a strong power law fit across the entire dataset is key to establishing a preferential attachment model based on data, Mitzenmacher (2003) argues that log-normal distributions are close enough to power law distributions that a preferential attachment model could easily lead to a strong log-normal fit. Additionally, Mitzenmacher points out that the most important differences between log-normal and power law distributions occur in the long tail, and our strong power law fit on the tail-end gives us more reason to believe that the log-normal fit could be a result of preferential attachment. In the context of international trade, rich-get-richer makes sense, since the most well-developed economies have a more complex economies and therefore a greater need to trade to sustain all components of the economy. Complex economies tend to be higher-growth, leading to a rich-get-richer phenomenon.

That population-adjusted exports do not fit a log-normal distribution may suggest that population-adjusted exports are a less meaningful metric since they don't have a simple real-world model. Due to ease of analysis, from here, we'll focus our analysis on GDP-adjusted exports.

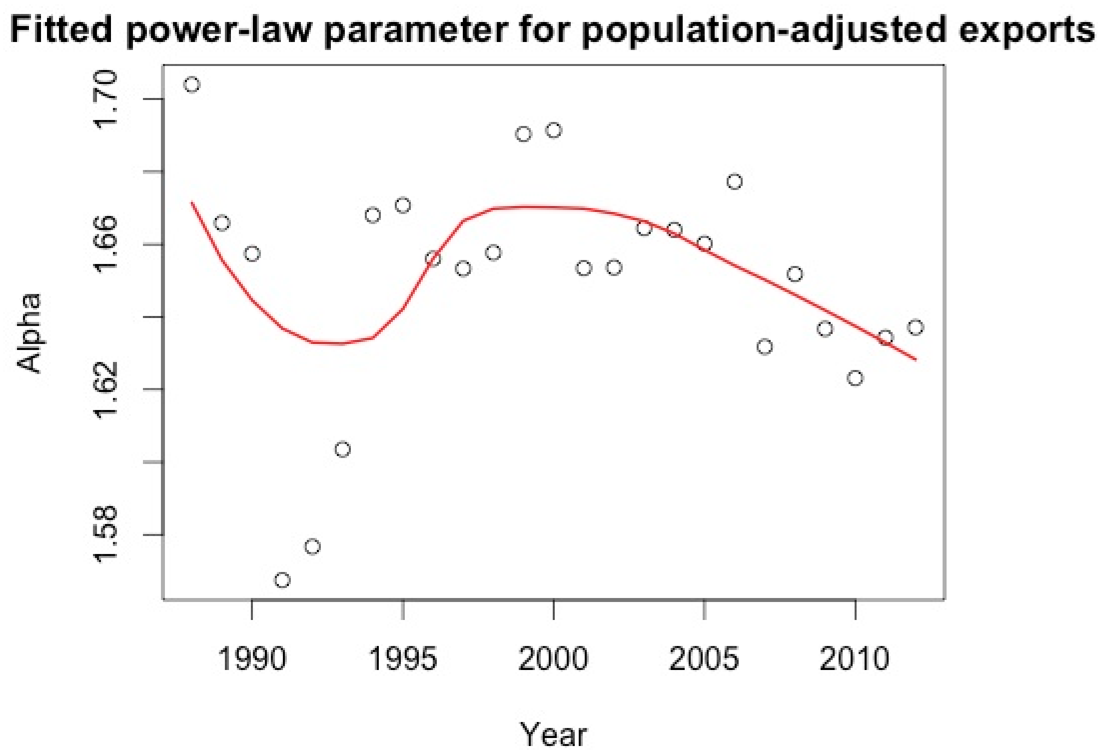
Trends over time

To glean insights into how the global trade system is changing over time, we track trends in how the parameters for our best fit models change over time. A key factor to keep in mind is that the number of countries increases only slightly (by 15%) over the 25 years of analysis.

Population-adjusted exports

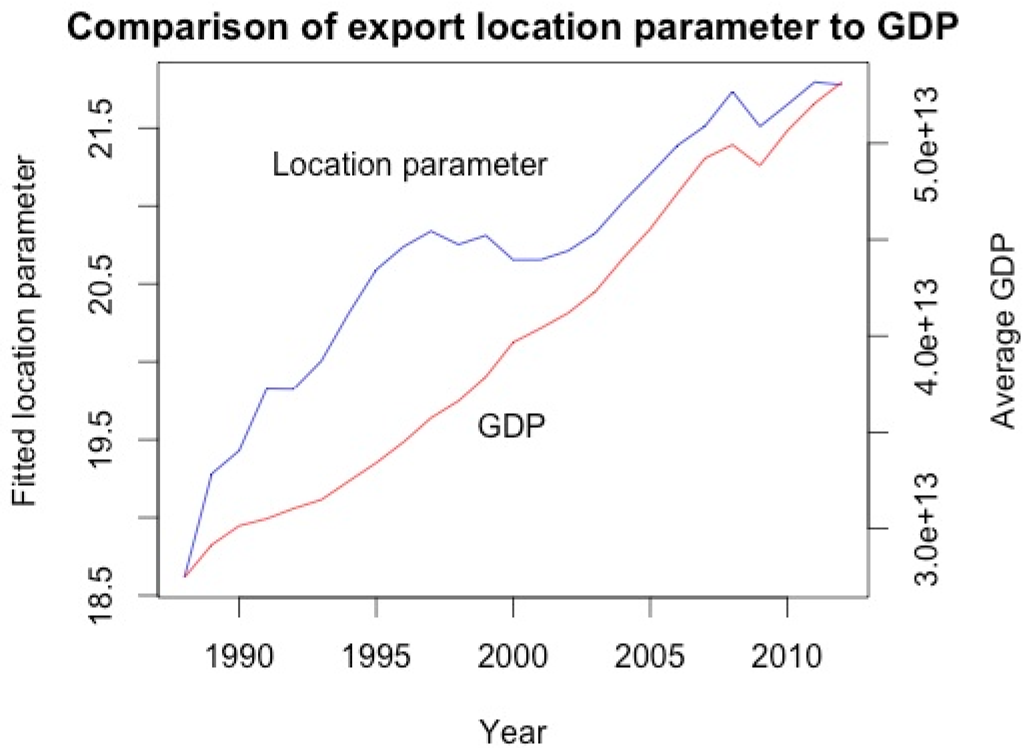
Although we focus mostly on total exports and GDP-adjusted exports, we'll look quickly at how our power law fit for population-adjusted exports changes over time. The change in parameter for the tail-end power law fit follows a weak, decreasing trend. This suggests that over time, the rich-get-richer quality of the international trade network may be decreasing.

From 1991 to 1993, we discovered a huge drop in the parameter alpha. We have not been able to explain these outliers.



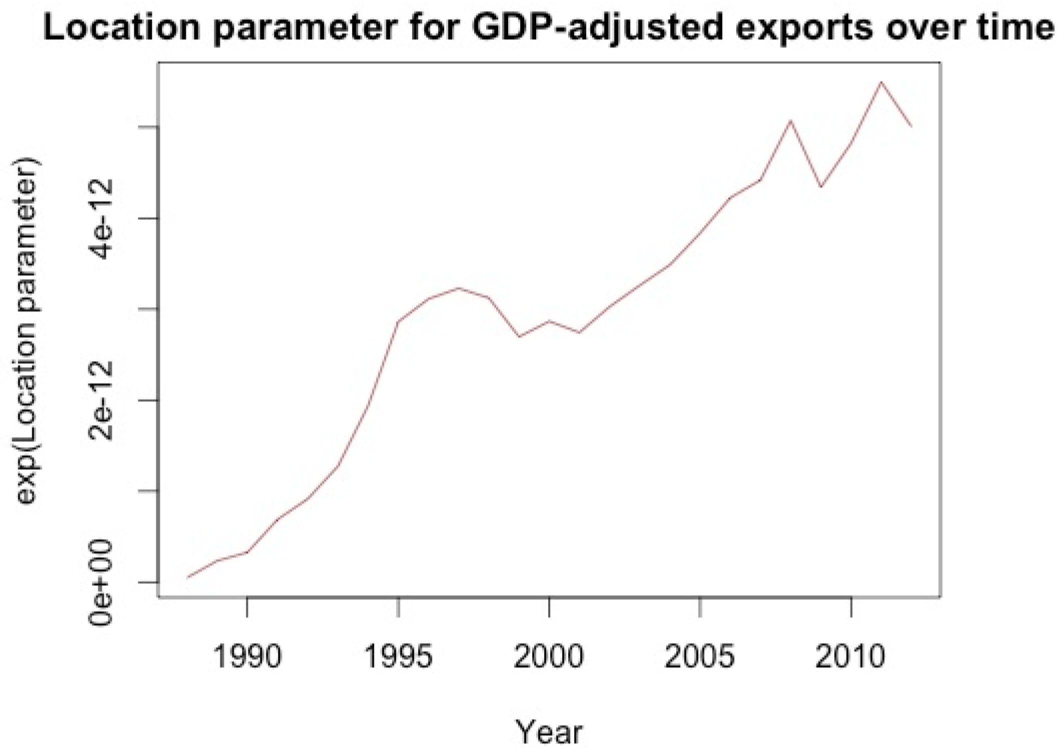
The location parameter for total exports and GDP-adjusted exports

For log-normal distributions, the location parameter μ gives the expected value of the logarithm of samples from the distribution. For total exports, the location parameter increases approximately linearly with the growth of the global economy. We use average GDP as a proxy variable for global economic growth in the graph below.



Observe that a linear relationship between GDP and μ tells us that the expected value of exports per country actually increases exponentially with average global GDP, since we're dealing with a log-normal distribution. This is a non-obvious property of international economics.

If we adjust export numbers for the GDP of both seller and buyer countries, we find another log-normal distribution with a familiar trend in location parameter:

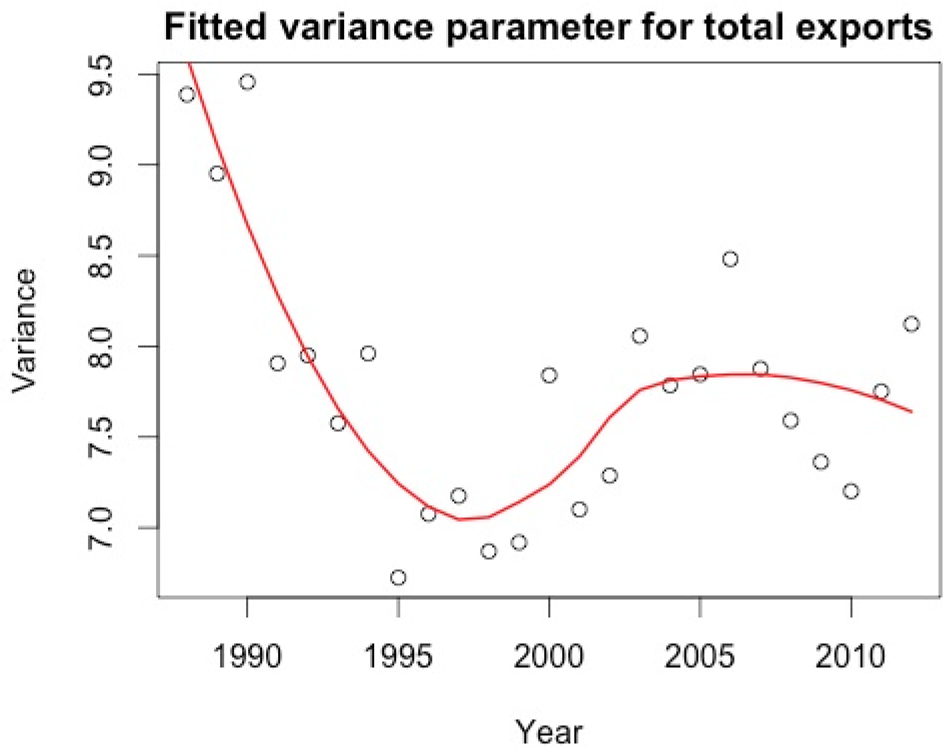


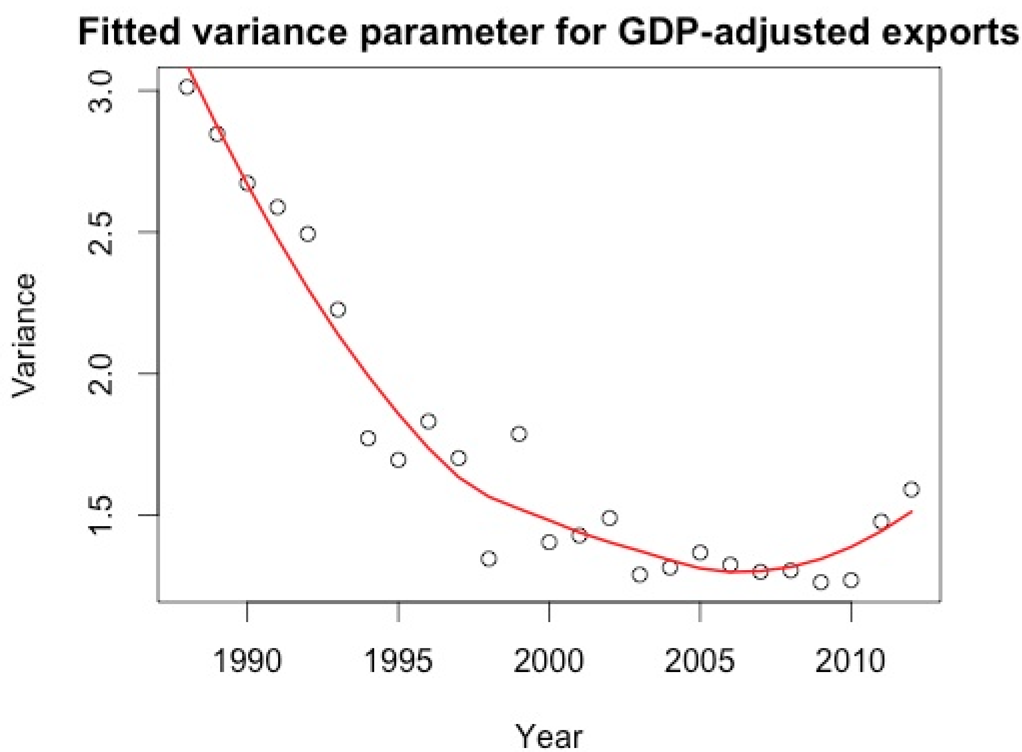
Here, we take the exponent of the location parameter on the y-axis and end up with a trend that appears to be very strongly correlated with the regular location parameter for the non-adjusted exports. So, GDP-adjusted exports, on average, increase in proportion with global GDP growth. This sounds like it might be expected, but remember that we only adjust exports with respect to the GDPs of the buyer and seller countries, not a global GDP metric. In the average case, this means if we model trade volume as:

$exports_{A \to B} = c \cdot GDP(A) \cdot GDP(B)$, the constant c increases linearly with global GDP. This, too, is a non-obvious property of international economics.

Variance for total and GDP-adjusted exports

We find a generally decreasing trend over time in fitted variance in both total and GDP-adjusted exports. This means that differences in total trade in the network are evening out over time. Globalization could help explain this effect. In the globalization model, as the playing field gets flattened, it becomes easier for weaker economies to overcome the rich-get-richer effect. Reduced variance in the log-normal distribution helps confirm this.





Note that the trends over time for these parameters differ significantly. Even though the decline in variance for total exports was arrested around the year 2000, a decline in variance continued for GDP-adjusted exports until 2010. This observation can be explained by variance in GDP. Total trade could become more skewed after 2000 just because GDP becomes more skewed, which does not reflect on properties intrinsic to the trade system. When we take out the GDP component, we get a picture where trade barriers continue to decrease, while the economic system as a whole becomes more unequal across countries.

Future studies (~3-5 years out) should watch the trend we are beginning to observe where trade volume variance is increasing.

Conclusions

- Trade volume can be reliably modeled via a log-normal distribution with time-variant parameters. This is true as well for GDP-adjusted trade volume. This suggests (as per Mitzenmacher 2003) a preferential attachment mechanism in international trade.
- It appears possible to model the expected trade between two countries A and B via $E[exports_{A \to B}] = c \cdot GDP(world) \cdot GDP(A) \cdot GDP(B)$, where c is a time-invariant constant.
- Variance in trade volume, GDP-adjusted trade volume, and population-adjusted trade volume have been decreasing over time (until now). Under a preferential attachment mechanism, this means that attachment is overall less "preferential," suggesting a world where there are fewer barriers to trade or more incentives to trade with a wide variety of partners.