Understanding Distance Effects in Global Food Trade Through Network Analysis

Amy Xu
Stanford Univeristy
xuamyj@stanford.edu

Jordan Duprey
Stanford Univeristy
jduprey@stanford.edu

Abstract

As the global food trade industry continues to grow, researchers have observed that despite lower shipping costs, trade flow continues to inversely correlate with geographical distance (ie. the ”distance effect”). Our paper investigates the global trade networks of specific foods and agricultural commodities using graph-based algorithms such as weighted global clustering coefficient, community detection, and weighted k-core decomposition. In particular, we model the trade flows for bananas and soybeans as weighted undirected graphs. We find that the banana trade network has lower global clustering coefficients (averages of 0.55 for imports and 0.38 for exports, compared to soybeans' averages of 0.81 for imports and 0.57 for exports), more visible and geographically determined clusters, and a lower maximum weighted k-core decomposition value (1518, compared to 4260 for soybeans) shared amongst more countries (7, compared to 2 for soybeans). Furthermore, we find that the clustering of the import and export graphs of both crops appeared to be correlated with the distance effect.These differences support our claim that distance effects impact different foods/commodities very differently.

1 Introduction

Global trade in food and agricultural products has grown rapidly in recent decades[1]. At the same time, the cost of shipping across large distances has decreased. However, economists have found that distance continues to inversely correlate with trade flow amounts.

This pattern is usually known as the ”distance effect,” and can be calculated using the economic gravity equation: ”Bilateral trade between two countries is proportional to size, measured by GDP, and inversely proportional to the geographic distance between them” (Chaney, 2013). These distance effects ”almost certainly do not arise solely from transport costs,” since for 80% of all shipments transport costs take up less than 4% of the total value. In addition, researchers have found ”a distance effect of 1.1 for 'digital goods' consumed over the Internet” (Disdier and Head, 2008).

Disdier and Head (2008) adds that distance effects have been found to be persistent in multiple senses: they appear throughout papers with a wide range of methodologies, and they ”are not declining in studies employing more recent data.” Given that this pattern exists for trade *in general*, how do distance effects impact the trade flows for *specific* foods and agricultural commodities? More broadly, how do the trade flows for different foods/commodities vary from each other?

In this project, we address these questions through a network analysis lens. We investigate the results of several graph-based algorithms, including weighted global clustering coefficient, community detection, and weighted k-core decomposition. Then, we discuss what these results reveal about trade flow structures for different foods/commodities. Through our analysis, we also point out that distance effects are not uniform in the way they affect different foods/commodities.

2 Related Work

In the literature, there is a large body of research that addresses either global food trade as a whole or global food trade in the context of virtual water trade. However, there are relatively few papers that look at global food trade for specific food/commodities. Currently, the papers that are

[1]Food and Agriculture Organization of the United Nations, http://www.fao.org/3/a-i5090e.pdf

available look at maize (Wu and Guclu, 2013) and wine (Cassi et al., 2012).

Wu and Guclu (2013) models global maize trade as a weighted, directed graph and focuses on community detection-based insights. They find three main clusters, which are based around Europe, Brazil/Argentina, and the US. They find that European countries trade much of their maize amongst themselves, and that they trade very minimally with the US. Furthermore, the US exports maize to a large number of Asian, Middle Eastern, and American countries, many of which only import from the US. Finally, Brazil and Argentina "seem to have largely segregated their patterns of export."

Cass et al. (2012) focuses on the relationship between globalization in trade and globalization in scientific knowledge. As such, it takes a different approach in its analysis of the global wine trade. It examines the evolution of the wine trade network and the wine scientific collaboration network over time, focusing on the measure of "coreness," or "the degree of closeness of each node to a core of densely connected nodes in the network." The paper finds that "Old World" wine producers have higher coreness in both networks than "New World" producers.

Our approach is based on community detection and core-based analysis, taking inspiration from Wu and Guclu (2013) and Cass et al. (2012). For algorithms, our paper draws upon Opsahl and Panzarasa (2009)'s definition of the weighted global clustering coefficient, Martin et al. (2013)'s OpenOrd community detection toolbox, and Garas et al. (2012)'s method for weighted k-core decomposition.

Opsahl and Panzarasa (2009) introduces a weighted variant of the global clustering coefficient. It defines the clustering coefficient of a node as the ratio between the total value of closed triplets rooted in the node to the total value of all triplets rooted in the node, where the value can be aggregated using arithmetic mean, geometric mean, maximum, or minimum. The paper compares this variant to the standard clustering coefficient as well as weighted local clustering coefficients in a number of weighted networks, and finds that it performs as well or better, since it is less sensitive to local edge weight distribution.

Martin et al. (2013) describes the algorithm for OpenOrd, a graph visualization toolbox that is designed for surfacing clusters in directed graphs. According to the paper, OpenOrd uses a force-directed algorithm with an objective function that is the summmation of pairs of attractive terms and repulsive terms. The algorithm attempts to minimize the function using a greedy procedure similar to simulated annealing. OpenOrd is parallelizable and can be run on multilevel graphs.

Garas et al. (2012) presents an algorithm for k-core decomposition on weighted networks. In unweighted k-core decomposition, nodes are recursively removed from a network until the network is partitioned; the resulting partitions are directly linked to centrality. The paper's weighted k-core method provides an alternative measure for the node degree, which is then used in the same recursive removal process. It finds that the weighted k-core method is able to "to split the cores obtained by the unweighted method further and to identify which are most central of the central nodes."

3 Methods

3.1 Data and Model

Our data comes from FAOSTAT's Detailed Trade Matrix[2]. The data is collected by the Food and Agriculture Organization of the United Nations (FAO) Statistics Division, using official reports from 186 countries for up to 422 subcategories of items. The data includes country-to-country imports and exports from 1986-2013, with measurements provided both in terms of quantity (tons) and value ($1000 USD).

Using this raw data, we built a parser that generates weighted directed graphs representing the trade flows for a single food/commodity. In the graphs, an edge's source node is the exporting country, its destination node is the importing country, and the weight is the amount traded. Our parser can filter by the desired year and the desired unit of measurement. We chose 2013, since it is the most recent year, and the value-based measurements ($1000 USD) so that it is easier to compare different foods/commodities, especially in context of the trading countries' GDPs.

Our first two graphs encapsulate the global trade flows for soybeans and bananas. Soybeans are a leading US export[3], particularly in recent years as

[2] http://faostat3.fao.org/download/T/TM/E

[3] http://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/agricultural-trade/

a result of strong demand from China[4]. Bananas, on the other hand, are a tropical product, and therefore the US import share is near zero [5].

Occasionally, two countries' reports of their trade will disagree; that is, Country A's reported export value to Country B will differ from Country B's reported import value from Country A. According to the FAO, there are several reasons for these discrepancies, which include

- Time lag: an export in December may arrive as an import in January, and be counted in the following year
- Misclassification of items
- Exported quantities being destroyed by external circumstances, such as accidents or weather conditions
- Data confidentiality by one of the reporters
- Customs tax or embargo avoidance

In these situations, our parser averages the two countries' reports.

3.2 Algorithms

Degree Distributions: After parsing the raw data into weighted directed graphs, we plotted the degree distributions of the graphs. For the unweighted degree distributions, we used the Snap.py library. For the weighted degree distributions, we considered the weighted in(out)-degree of a node to be the sum of the weights of its in(out) edges.

Weighted Global Clustering Coefficient: Continuing our analysis of graph properties, we calculated weighted global clustering coefficient using an algorithm defined by Opsahl and Panzarasa (2009):

$$C = \frac{\sum triplet_{closed}}{\sum triplet}$$

A $triplet$ is defined as a set of three nodes where two nodes, x and y, have a directed edge into node z. A closed triplet is defined as a triplet where at least one additional edge exists between nodes x and y. An open triplet occurs when no such additional edge exists.

Opsahl and Panzarasa (2009) describes four techniques to get the value of the triplet. In each technique, the total value of the triplet is based on the weight of the directed edges that ran from the two neighbors into the central node of the triplet (node z in the prior example). However, they differ in that the value of the triplet can be the arithmetic mean, geometric mean, maximum, or minimum of the weights. Since Opsahl and Panzarasa (2009) concludes that each technique has its advantages and disadvantages, we elected to use and report the results from all four definitions.

Community Detection: The weighted global clustering coefficient results suggested that different foods/commodities have different graph properties, so we investigated further using community detection. For this, we used the OpenOrd layout program on the Gephi platform. We chose OpenOrd for two reasons. First, we wanted to quickly visualize the clustered graphs in order to gain an intuitive understanding. Second, OpenOrd is designed to distinguish better clusters in weighted graphs.

Weighted k-core decomposition: Since the OpenOrd visualizations indicated that

1. there appears a relationship between clusters and geographic/distance effects, and
2. different foods/commodities exhibit different clustering structures,

we wanted to find a way to determine the optimal number of clusters for each trade network. We hypothesized that applying weighted k-core decomposition would remove the outer nodes, thus leaving cores representing each geographical cluster.

We implemented the weighted k-core decomposition algorithm defined by Garas et al. (2012). In the paper, the weighted degree of a node i is defined by

$$d_i' = \left[d_i^{\alpha} \left(\sum_{j}^{d_i} w_{ij} \right)^{\beta} \right]^{\frac{1}{\alpha+\beta}}$$

where d_i is the degree of node i, and $\sum_j^{d_i} w_{ij}$ is the sum of the weights of i's outgoing edges. Garas et al. (2012) assumes $\alpha = \beta = 1$, so we made the

[4]http://www.ers.usda.gov/topics/international-markets-trade/us-agricultural-trade/export-share-of-production/

[5]http://www.ers.usda.gov/topics/international-markets-trade/us-agricultural-trade/import-share-of-consumption.aspx

same assumption for our implementation:

$$d_i' = \sqrt{d_i \sum_j^{d_i} w_{ij}}$$

For the procedure, let us define k_{curr} to be an incrementing index that starts at $k_{curr} = 1$. The first step of the procedure is to choose the node with the lowest d_i' (ties can be arbitrarily broken). If $d_i' \leq k_{curr}$, label node i with k_{curr} (we will refer to this as node i's "k-value"), remove node i and all of its edges from the graph, recalculate k' for all remaining nodes, and repeat. If $k_i' > k_{curr}$, increment k_{curr} until $k_i' = k_{curr}$. Then, continue the procedure with the new k_{curr}.

4 Results

4.1 Degree Distribution

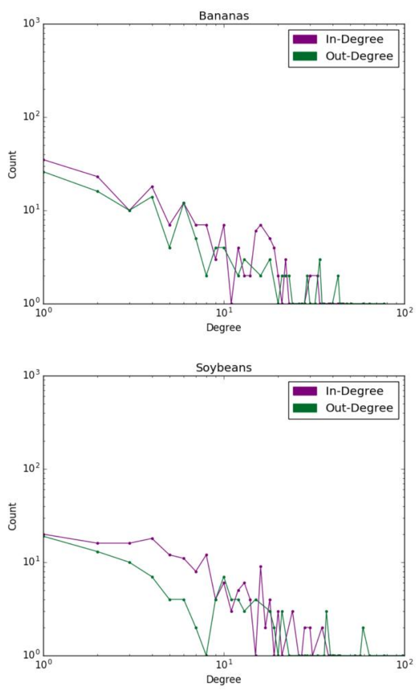


Figure 1: Unweighted degree distributions (both in-degree and out-degree) of banana trade network (above) and soybean trade network (below)

We began our analysis by plotting the unweighted and weighted degree distributions of the two trade networks. Since the networks are directed graphs, in-degrees (imports) are always represented as purple, and out-degrees (exports) are always represented as green.

For both the banana trade network and soybean trade network, there does not appear to be a clear power law, small-world, or random graph pattern in the unweighted degree distributions (see Figure 1). This may be because the networks have relatively few nodes, which increases the impact of noise and/or outliers on the distribution. We note that the unweighted degree distributions appear to be very similar to each other.

The weighted degree distributions provide very little information. There are too many possible weighted degrees and too few nodes for the charts to aggregate them in a reasonable way (see Figure 2).

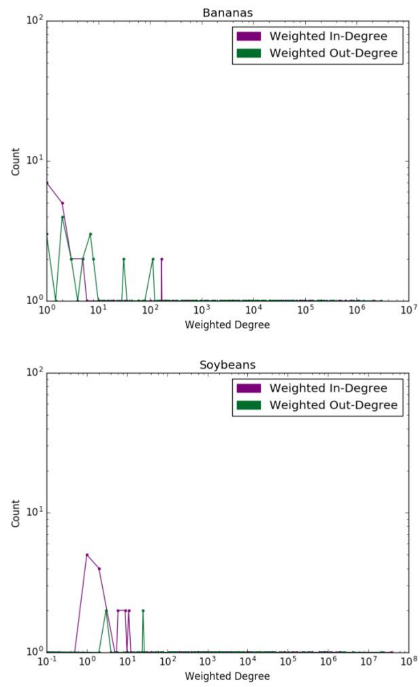


Figure 2: Weighted degree distributions (both in-degree and out-degree) of banana trade network (above) and soybean trade network (below)

4.2 Weighted Global Clustering Coefficient

Next, we calculated the weighted global clustering coefficient over the bananas and soybean trade data.

The initial results indicated that the soybean trade network was clustered more densely, with a higher clustering coefficient, while the banana trade network was more segmented, with a lower clustering coefficient (see Table 1). Additionally, the import graphs' weighted global clustering coefficients were consistently higher than the export graphs' weighted global clustering coefficients.

Bananas		
	Import	Export
Arithmetic mean	0.519008	0.359137
Geometric mean	0.576040	0.410663
Max	0.515398	0.353698
Min	0.577858	0.434700

Soybeans		
	Import	Export
Arithmetic mean	0.769061	0.594078
Geometric mean	0.835612	0.571988
Max	0.764189	0.597177
Min	0.880124	0.515560

Table 1: Weighted global clustering coefficients of banana trade network and soybean trade network, calculated using all four variations of the Opsahl and Panzarasa (2009) method

4.3 Community Detection

Using Gephi visualization software and the OpenOrd layout (Martin et al., 2013), we visualized the banana and soybean trade networks. The banana trade network appears to have four main clusters, with a small faint cluster on the far left.

We labeled and colored the banana trade network based on the weighted in-degree (total imports) of each node (see Figure 3; larger figure provided in the Appendix). A node with more total imports will have a larger label, as well as a dark purple color for its label and outgoing edges. A node with fewer total exports will have a smaller and lighter-colored label, and fainter pink edges.

Figure 3: Major importers of the banana trade network

As one might expect, the largest importers are the more developed countries that have bigger GDPs. Additionally, we note that the clusters have a strong basis in geographical region. The bottommost cluster mostly consists of American countries; the top right cluster mostly consists of Asian and Middle Eastern countries; the faint bottom left cluster mostly consists of African countries. As for the left and middle clusters, both consist mainly of European countries, with more Eastern European countries in the left cluster and more Western European countries in the middle cluster.

Next, we labeled and colored the banana trade network based on the weighted out-degree (total exports) of each node (see Figure 4; larger figure provided in the Appendix). Here, larger/darker green nodes and edges represent countries with more banana exports, and smaller/lighter green nodes and edges represent countries with fewer or no banana exports.

Based on the graph, Ecuador is the largest exporter of bananas. Though it mostly exports to the left (Eastern European) cluster, it also trades with countries from other clusters. Colombia and Costa Rica are major exporters of the middle (Western European) cluster, and Guatemala is a major exporter of the bottommost (American) cluster.

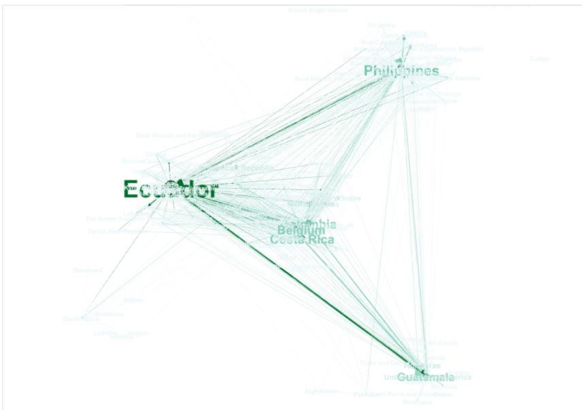


Figure 4: Major exporters of the banana trade network

Here, distance effects may play a small role in determining trade routes (for example, Guatemala is closer to the United States than any other major exporter). However, distance likely matters less than historical relationships between countries, since all of the aforementioned countries are located in tropical Central America and therefore have similar distances to various banana importers.

The largest exporter in the top right (Asian/Middle Eastern) cluster is the Phillipines, which is consistent with the distance effects theory. Additionally, Belgium appears prominently in the middle cluster of both graphs; this is due to Belgium's role as key re-exporter of bananas and other fruits to Europe[6].

After analyzing the banana trade network, we graphed the soybean trade network and again ran the OpenOrd layout. The resulting graph is very different from the banana trade network graph, with a single large cluster.

We applied the coloring/labeling technique that we used earlier to the soybean trade network (see Figure 5). Since there is only one large cluster, the countries are not separated geographically. These results suggest that distance effects play a more significant role in the banana trade network than the soybean trade network. In addition, these results are consistent with our findings in 4.2 that the soybean trade network is clustered more densely while the banana trade network is more segmented.

Figure 5: Importers (purple) and exporters (green) of the soybean trade network. China (mainland) is the largest importer, and the US is the largest exporter

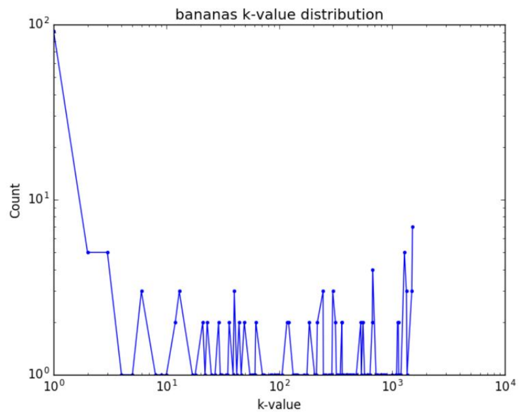
[6]http://www.nationsencyclopedia.com/economies/Europe /Belgium-AGRICULTURE.html

4.4 k-core Decomposition

Finally, we ran weighted k-core decomposition on the soybean and banana trade networks, in which each node is assigned a "k-value" k_s. According to Garas et al. (2012), nodes with a high k_s are located in the center of the network, and belong to one of the nuclei of the network; nodes with a low k_s are located in the periphery. Because of this, we originally hypothesized that k-core decomposition would reveal the underlying clustering structure of our trade networks in a numerical (rather than visual) way.

At first, our hypothesis appeared false when we plotted and compared the k_s distributions for both networks (see figure 6. The two charts appear to be very similar; both graphs have 100 nodes for the k-value of 1, 1-2 nodes for many k-values between 10^2 and 10^3, and more nodes for the highest k-values.

w log.png



w log.png

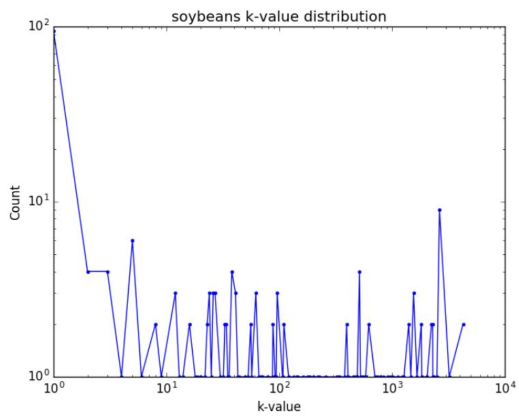


Figure 6: Weighted k-value distributions for banana trade network (above) and soybean trade network (below), plotted on a log-log axis

However, we note two important differences between the banana and soybean trade networks' k_s distributions. First, the highest possible k_s for the soybean trade network is much greater than the highest possible k_s for the banana trade network. Second, fewer countries have the highest possible k_s in the soybean trade network compared to the number of countries that have the highest possible k_s for the banana trade network.

This distinction becomes particularly visible when we look at the top 20 countries with the highest k-values for both trade networks. In the banana trade network (see Table 2), there are 7 countries with the maximum k_s of 1518. Each of these countries also shows up as a one of the "major" (ie. dark green or dark purple) countries in the banana trade network visualization.

On the other hand, only Brazil and China (mainland) have the maximum k_s in the soybean trade network, but this value is 4260. To explain this intuitively, we can think of a node having a larger k_s as having a greater "nucleus status." Since countries in the banana trade network are part of different nuclei, they have to "share" the total "nucleus status." However, in the soybean network, there is only one cluster. China and Brazil are part of the only nucleus and do not have to "share" their "nucleus status" with any other nuclei.

Bananas	
k-value	**Country**
1518	Colombia, Costa Rica, Ecuador, Guatemala, Russian Federation, United States of America, Belgium
1497	Germany, Italy, United Kingdom
1359	Netherlands
1341	Canada, Dominican Republic, Honduras
1288	France, Mexico, Panama, Spain, Ukraine
1201	Sweden

Table 2: Countries with the top 20 weighted k-values in the banana trade network

5 Analysis/Findings

The experiments have shown that the distance effect appears to definitely be correlated to various structures of the graph. To quantify the distance

Soybeans: Top 20 k-values	
k-value	**Country**
4260	Brazil, China (mainland)
3212	Argentina
2633	Canada, Germany, Japan, Netherlands, Paraguay, Spain, Thailand, United States of America, Uruguay
2501	China (Taiwan Province of)
2374	Viet Nam
2328	Egypt
2298	Italy
2289	Russian Federation, Ukraine
2236	Turkey, United Kingdom

Table 3: Countries with the top 20 weighted k-values in the soybeans trade network

effect on our data set we calculated the variance of the distance effect on the imports and exports respectively for the various crops.

$$distanceeffect = \frac{tradeamount * distance}{gdp_1 * gdp_2}$$

We found that for weather dependent crops the distance effect was definitely present since the variance for soybeans exports and imports were $8.174e^{-28}$ and $1.463e^{-29}$ respectively while the variance for banana exports and imports were $1.463e^{-27}$ and $1.816e^{-28}$ respectively. For the bananas, the exports had a much stronger effect due to distance since the variance was two orders of magnitude larger than that of soybean imports. This is consistent with our findings with the network structure of those two graphs. The weighted global clustering coefficient of the banana export graph was much smaller than that of the import graph for soybeans. Likewise, the visualization created with the OpenOrd algorithm showed much more distinct, separate clusters in the banana graph. Our findings support the idea that weather sensitive crops will be traded in higher volume over large distances while still favoring closer sources than othan farther ones.

6 Conclusion

From our analysis of banana and soybean trade networks, we can conclude that trade flows for different types of foods/commodities have very different structure. These differences in structure manifest themselves through

1. The weighted global clustering coefficient of the entire graph being lower vs. higher

2. The existence vs. lack of distinguishable clusters after applying community detection algorithms

3. The maximum weighted k-core decomposition value being lower vs. higher, and having more vs. fewer nodes that have this maximum k-value.

When we consider these results in conjunction with the results of our mathematical analysis, we find that distance effects have a different impact on the trade flows of different foods/commodities, and this variation of impact is related to the clustering properties of the trade networks.

Based on our current results, we hypothesize that these structural differences are due to the widely different climate requirements of bananas. vs soybeans. For future work, we would like to extend our analysis to more foods/commodities, and develop a mathematical model for why different foods behave the way they do. Possible factors we would like to investigate include differences in climate requirements, perishability, and shipping logistics.

References

Cassi, L., Morrison, A., & Ter Wal, A. L. 2012. The evolution of trade and scientific collaboration networks in the global wine sector: a longitudinal study using network analysis. *Economic geography*. 88(3): 311–334.

Chaney, Thomas. 2013. The gravity equation in international trade: An explanation *National Bureau of Economic Research*. No. w19285.

Martin, S., Brown, W. M., Klavans, R., & Boyack, K. W. 2011. OpenOrd: an open-source toolbox for large graph layout. *IS&T/SPIE Electronic Imaging*. No. 786806.

Disdier, A. C., & Head, K. 2008. The puzzling persistence of the distance effect on bilateral trade. *The review of economics and statistics*. 90(1): 37–48.

Garas, A., Schweitzer, F., & Havlin, S. 2012. A k-shell decomposition method for weighted networks. *New Journal of Physics*. 14(8): 083030.

Leskovec, J., Lang, K. J., & Mahoney, M. 2010. Empirical comparison of algorithms for network community detection. *Proceedings of the 19th international conference on world wide web*, 631-640

Opsahl, T., & Panzarasa, P. 2009. Clustering in weighted networks. *Social networks*, 31(2): 155–163.

Wu, F., & Guclu, H. 2013. Global maize trade and food security: Implications from a social network model. *Risk analysis*. 33(12): 2168–2178.

7 Appendix (for larger figures)

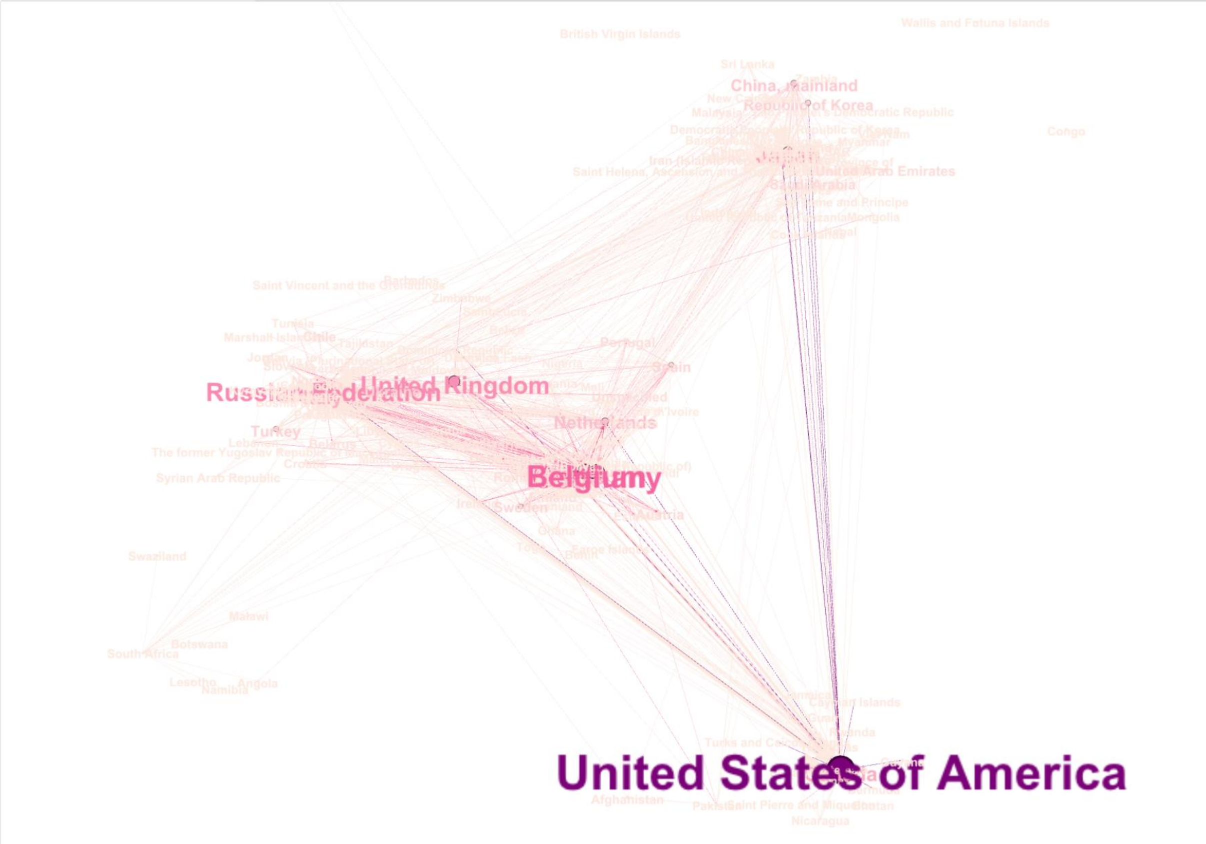


Figure 7. Major importers of the banana trade network

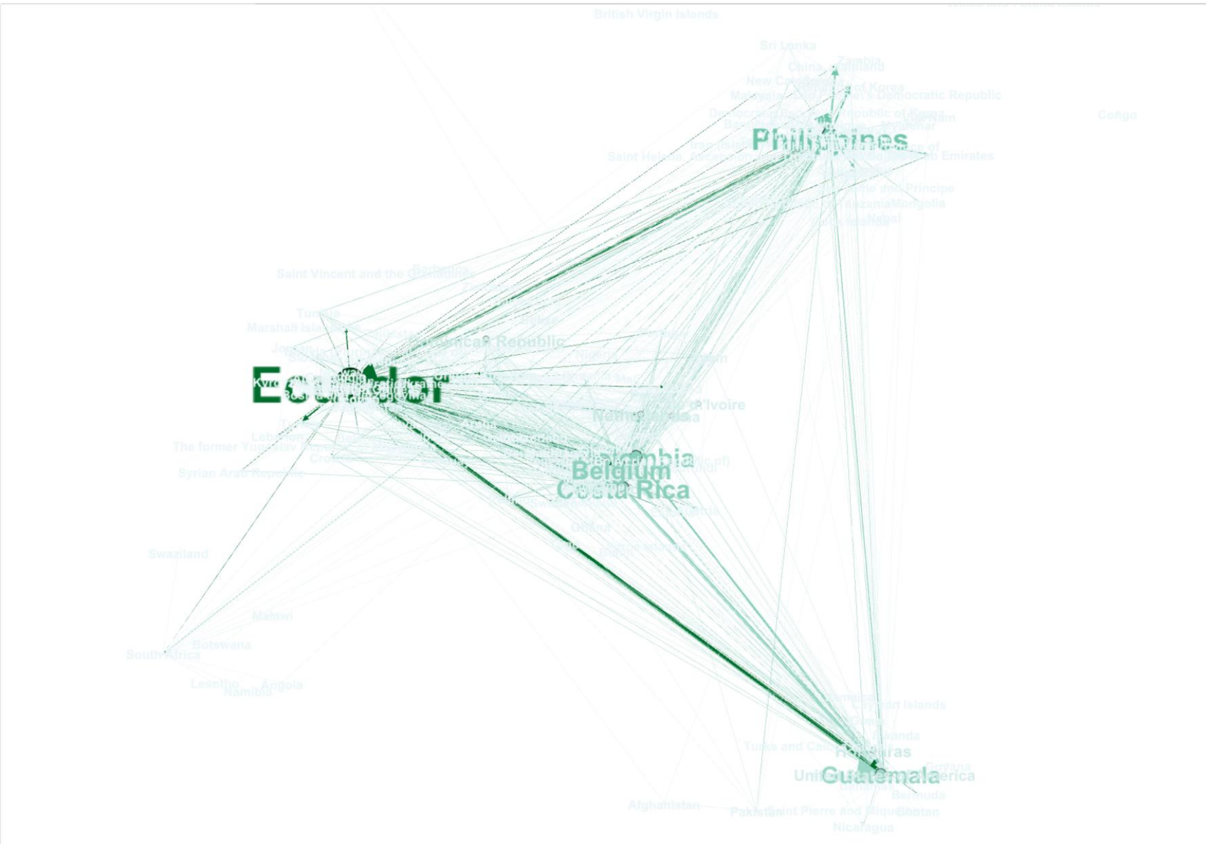


Figure 8. Major exporters of the banana trade network