Food Supply Networks

Daniel Heath, Yuga Iguchi, Liam Riordan, Steven Robertson, Victor Shirandami, Jackie Voros.

Supervised by Alexandra Brintrup and Jack Foster.

August 11, 2022

Main Goals

- To analyse impact on countries in the food and ingredient supply chain in the event of vertices dropout.
    - If we remove country X from the supply chain, what is the effect on country Y?
    - If we remove country X from the supply chain, what is the effect on the supply of ingredient Z in country Y?
    - If we remove country X from the supply chain, what is the effect on the supply of food W in country Y?
    - What are the critical countries in the global supply of ingredient Z and food W?
- To determine a good measure of ‘impact’ and ‘effect’.
- To create a visualisation tool to aid with analysis.

Comtrade Data

- The original data set was not restricted to food and consequently it was over 6GB in size.
- A program was written in Python to remove all non-food items. This was done using the commodity code unique to each trade good.
- However, the data is still imperfect. Namely,
    - There are still over 250000 individual trades reported.
    - The imports and exports are not mirrored.
    - Countries are inconsistent reporters.
    - Some trades are subsets of other trades.
    - These ‘sub-trades’ do not sum to the correct total.

Datasets

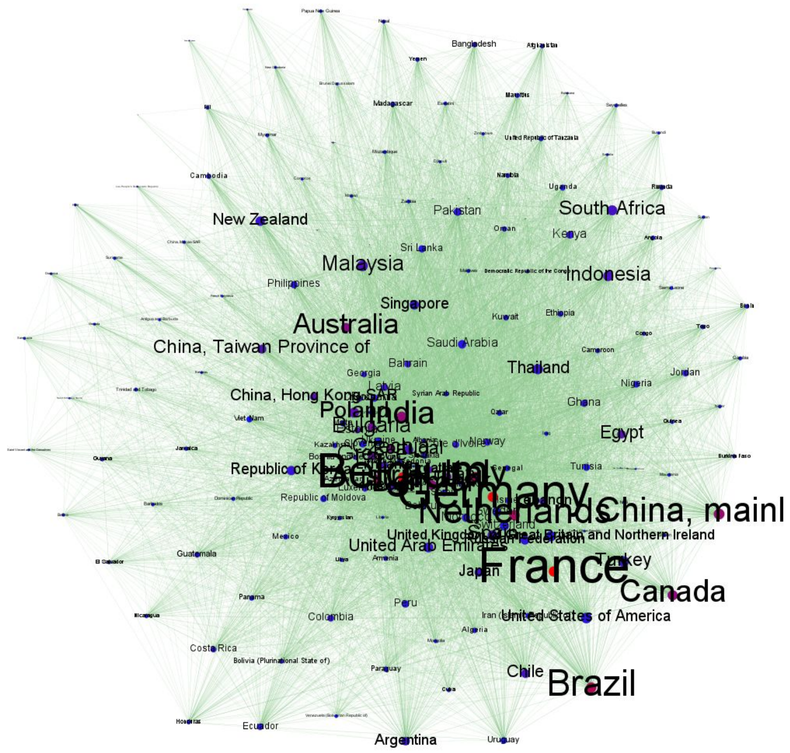
- Two subsets of the data have been created
- The first has had all ‘sub-trades’ removed, so that there are no duplicates.
    - The data is now split into roughly 20 basic types of food, making it ideal to analyse to see the ‘bigger picture’.
    - Some detail and nuance has been lost in the data.
- The second has only the ingredients for specific meals.
    - In particular, there is a dataset that contains only the ingredients to make mac and cheese, roast dinner, margarita pizza and apple crumble.
    - These meals were picked because they have very standardised recipes.
    - These datasets are small enough that the graphs created when analysing the networks are comprehensible to humans.

How can we analyse the data?

The data is best represented as a directed graph. We have done this through a specialised program called Gephi.

Each country is a node, each edge represents an export or an import.

We have chosen a few measures to analyse our graph data, although there are many measures available to use. These measures give a broad overview of the questions we are trying to answer.

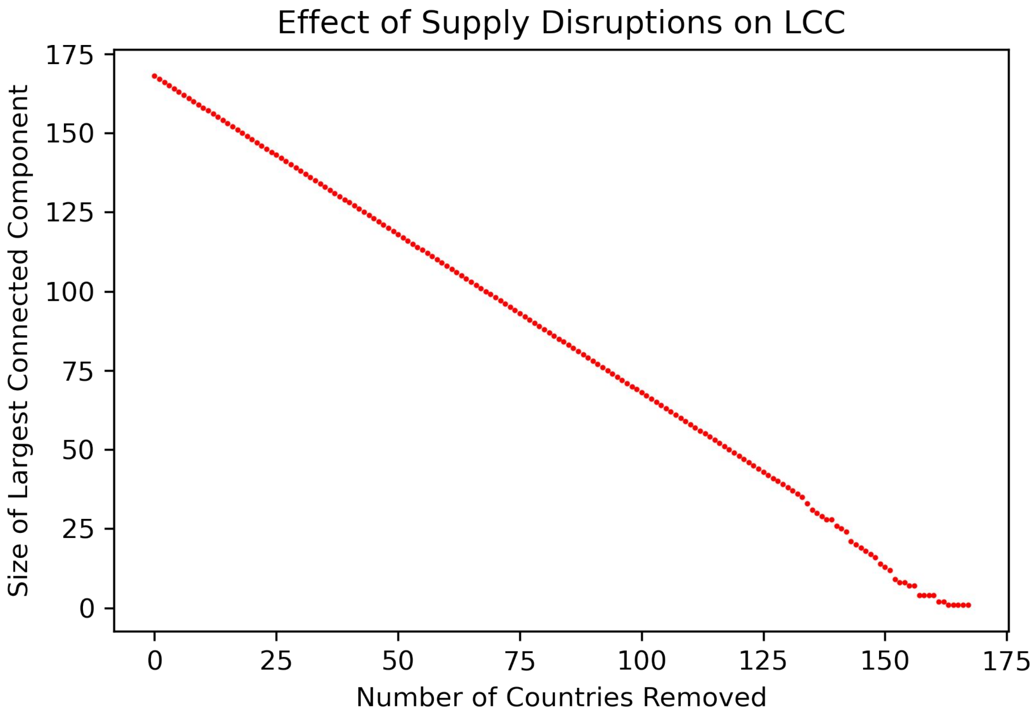


Graph terminology

- A graph G = (V,E)
  - V the set of vertices (or nodes)
  - E the set of edges

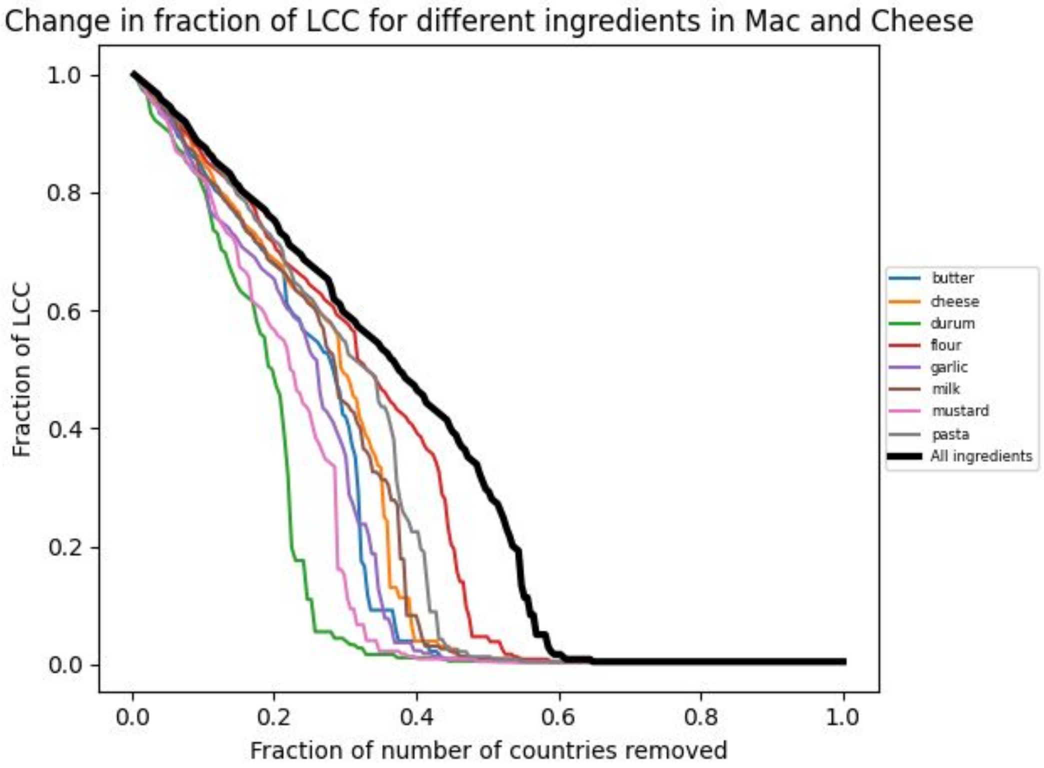
- The edges are weighted and directed
- Degree centrality shows the degree of a vertex - how many edges are attached to a node
  - In-degree
  - Out-degree
  - Total degree
- A component is a connected subgraph that is not part of any larger connected subgraph
  - For a digraph this is called ‘weakly connected’

Largest connected component



- Assess the vulnerability of a supply network by progressively removing nodes with the highest degree centrality and then measure the size of the LCC
- In this example, the graph is completely connected
- This makes it difficult to perform specific analysis

Largest connected component



- Gives impression of how quickly a supply network may become disconnected for a specific type of food
- The steeper the curve, the more necessary the ingredient is in the food production

Further analysis from LCC

```
Top 10 critical
exporters for
Mac and Cheese:
1.  France
2.  USA
3.  Germany
4.  Netherlands
5.  Italy
6.  United Arab Emirates
7.  United Kingdom
8.  Belgium
9.  Spain
10. Turkey
```

- Within the program, we list the first ten countries removed from the graph
- We can also analyse the specific impact for this food

Choose a food to analyse: Mac and Cheese

Choose a country to remove: France

Choose a country to analyse: United Kingdom

```
Mac and Cheese consists of the following ingredients:
Flour, Butter, Cheese, Mustard, Milk, Garlic, Pasta, Durum

###################################################################################################

France provides 21.185868960844598% of flour imported to United Kingdom
France provides 9.133726981951312% of butter imported to United Kingdom
France provides 15.907068343977745% of cheese imported to United Kingdom
France provides 30.274918699039617% of mustard imported to United Kingdom
France provides 6.826154715399556% of milk imported to United Kingdom
France provides 2.550498856058073% of garlic imported to United Kingdom
France provides 0.1711574813026039% of pasta imported to United Kingdom
France provides 63.065718832887618% of durum imported to United Kingdom
```

Degree distribution

The degree distribution acts as a probability distribution. We let n_k denote the number of vertices with degree k. Then P(k) is the fraction of vertices in V with degree k.

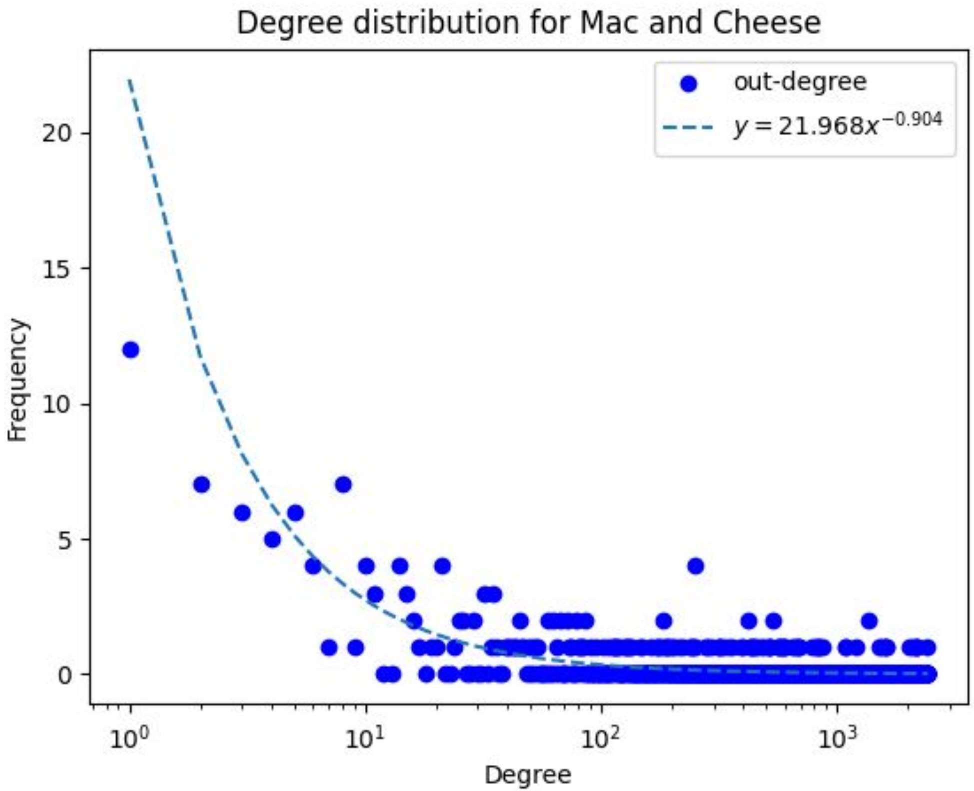
$$P(k) = \frac{n_k}{|V|}$$

For a random graph, the distribution is binomial. As |V| tends to infinity, the distribution tends to a Poisson distribution.

In real life examples, we get the following distribution

$$P(k) \sim k^{-\gamma}$$

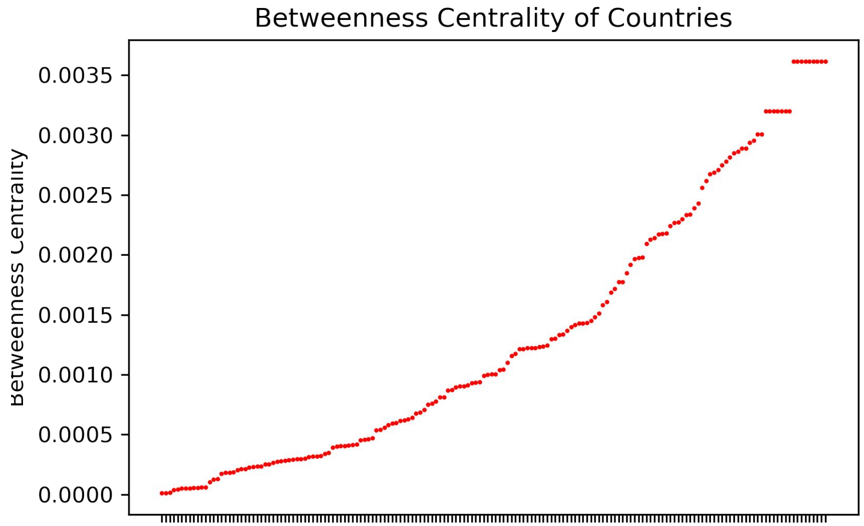
Degree distribution



- Such networks are considered as ‘scale free’
- These networks have the majority of vertices with low degree centrality
- Typically very few ‘hubs’ with very high degree centrality
- This analysis gives information on the presence of ‘hub’ countries

Betweenness Centrality

This measures how often a node will sit on paths that connect different nodes to each other in the network. Firms with high betweenness centrality may act as bottlenecks.




```
Malaysia 0.003201207573328647З
Brazil 0.003613956856138804
Canada 0.003613956856138804
Denmark 0.003613956856138804
France 0.003613956856138804
Germany 0.003613956856138804
Netherlands 0.003613956856138804
United Kingdom 0.003613956856138804
USA 0.003613956856138804
India 0.003613956856138804
```

Food Analysis Program Demonstration

ML tools

- Wanted to fill in missing data. This is because we had a lot of import data however the corresponding exports were often not listed .
- In particular we wanted to find the export value based on the input data.
- Without the context, this essentially means that we want to use machine learning to guess a multivariate function. Export = f(rep,par,comm,inp)
- Our two options were Neural networks and XGBoost.
- So how do they work?

ML decision trees

- For XGBoost we take a table where each row is a set of inputs and the last entry is the output.
- XGBoost's first guess is the mean of the outputs. Not too interesting.
- Then XGBoost chooses a split point. These are values which split each input data and we can consider the total loss of the tree as a function of these split points.
- By using gradient descent we can find split points which get us closer.
- We then add the outputs of these two trees with the new tree hopefully helping to cancel out the old trees errors.
- Final normalised error rate of 0.30
    - Sounds large, but our labels range from e^2 to e^12 so accurate predictions are very, very hard.

Next Steps and Future Research Directions

- Further consideration of LCC
  - Which order do we remove nodes of the same degree?
- Addition of more foods
  - Adding foods manually…
  - …or using ML to predict the ingredients of a given food and add automatically.
- Application of graph neural network (GNN) for predicting country substitution
  - Transforming the dataset into “graph” and learning the structure via NN.
  - predict a country substitution when a country stops trading (export/import).