**Food trade network analysis**

By Group R

**Introduction**

Nowadays, people usually evaluate the trade activity of a country by measuring the amount of export and import value. The metrics seem to be intuitive, but it ignores the diversity of a country’s trade network. Developed countries usually go at the top of the Export/Import listing simply due to their economic scale and not necessarily due to the fact that the network is efficient and broad.

Food security is one of the natural strategic focus of a nation. This project makes use of the trade data of 11 agriculture products from the Food and Agriculture Organization of the United Nations (UNFAO) to examinethe connectivity of a country’s food trade network by applying the PageRank method. We also analyze the PageRank between OECD countries and Least developed countries (LDCs).

**Data collection**

All the data is downloaded from FAOSTAT online database[[1]](#footnote-1) while the entire detailed trade datasets has been selected for our analysis.

The dataset contains more than 30 million records (4.2GB) which show the trade data (reporter country, partner countries, import/export quantity & value in USD) of the 425 types of crops or livestock for 255 areas from 1961 to 2017.

|  |  |
| --- | --- |
| Attributes | Description |
| Reporter Countries | The country which report the trade data |
| Partner Countries | The trading partner of the reporter country |
| Item | Types of crops or livestock |
| Element | Import quantity / value or Export quantity / value |
| Year Code | Year |
| Value | The weight, quantity or dollar value of the trade record |
| Unit | The unit of value, e.g. tons, head, USD in thousands |

**Table1** Dataset description

**Data cleansing**

As mentioned above, the original datasets from FAOSTAT is huge (4.2GB). Hence, we cleanse our datasets by using Python Pandas library. Since the import data and export date are equivalent, only export value (USD) has been selected. Besides, our team decide to include the 12 types of crops and livestock, which are **apples, bananas, grapes, maize, cattle meat, chicken, pork, potatoes, rice, soybeans, sugar cane** and **wheat**, that comprise most of our daily meals. We also focus on the trade data of the recent decades, i.e. 1997 to 2017.

Since the original dataset is huge and reading in the whole datasets is memory consuming, we divided the whole dataset into several subsets that each subset includes 4 million records at most. Then we construct the DataFrames that each stores the trade data for each food item of a year.

Our team also compare the PageRank between develop countries (OECD) and poor countries (LDCs). Based on the above processed dataset, we have constructed another dataset in which the trade data of OECD has been grouped and so as the data of LDCs. Our group eliminates the trade within the OECD group since we would like to analyze the inter-trade activity between the OECD and non-OECD members. Similar data processing has also been applied to the trade data of LDCs.

**Data Processing**

The processed datasets are in the form of pandas DataFrame (df). For conducting the PageRank analysis, we have transformed the df to 756 NumPy arrays (Matrix). Each array (S) contains the trade data for one item of a year. For example, the S of soybeans in 2017 stores the trade data between reporter countries and partner countries. The element Sij equals the normalized export USD value of reporter country j to partner country i. Then Matrix M was created by normalizing each element in S[[2]](#footnote-2). For import matrix, we simply transpose the matrix S and perform the normalization.

The following iteration formula is used for calculating the **import PageRank** and **export PageRank**.

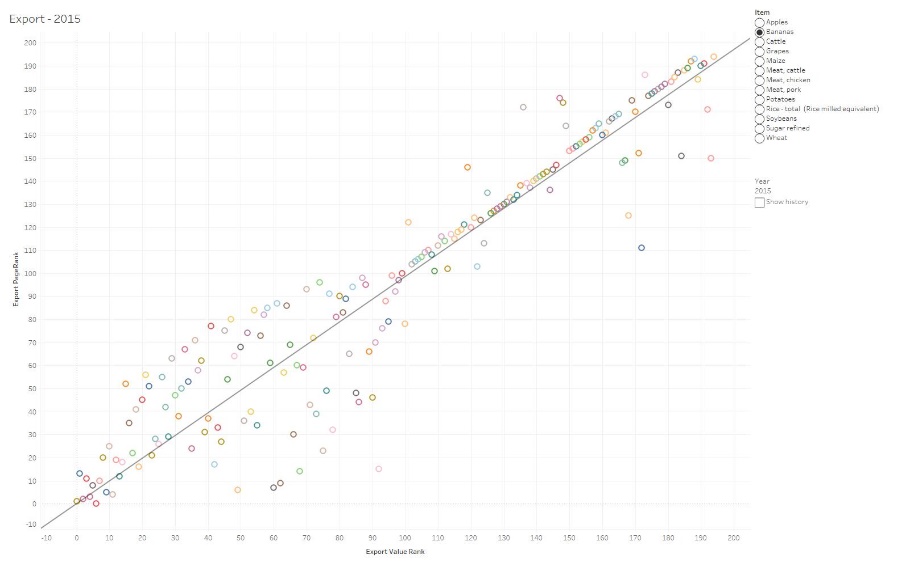
Where α = 0.85; e = all-ones vector; N = number of nodes

After running the PageRank algorithm, we analyze the trade network connectivity by the countries’ ranking.

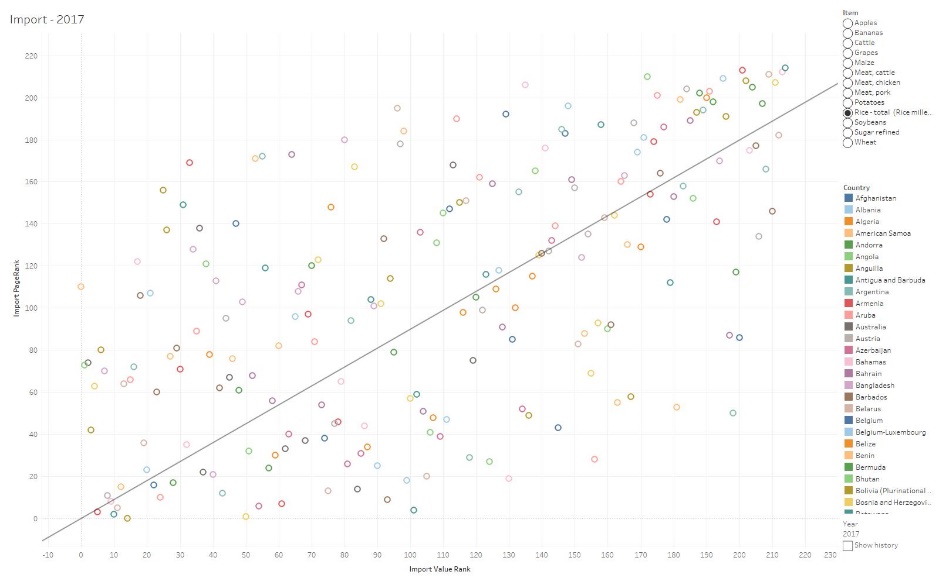
**Key Findings and visualization**

Since our team would like to construct a tableau workbook for visualizing out result, we apply **the glob library** for merging the 756 PageRank result. The glob library allows us to obtain the file name in the target directory and enhance the efficiency of opening each file.

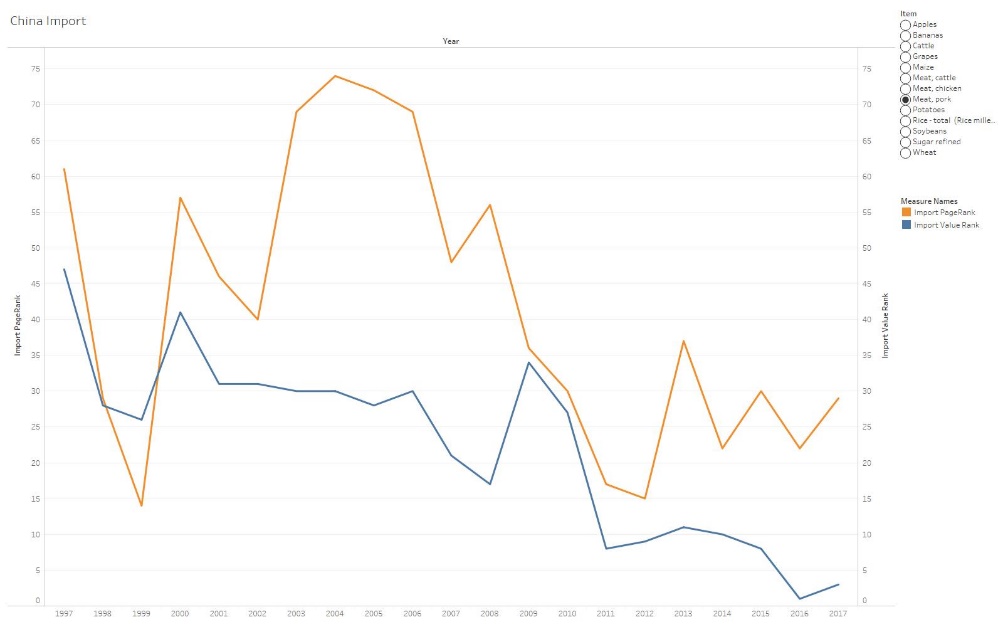
1. The export PageRank and the export value ranking are quite similar. For example, most of the points lie near the 45-degree line in the scatter graph of two types of country trade ranking scatter plot for banana in2015.



1. On the other hand, the points in the scatter plot for import data are more dispersed, which means the import PageRank is not aligned with the import value rank.



1. Developed countries (OECD) always have top ranking in both PageRank and value rank.
2. If we treat poor countries as a group (LDCs), their food trade PageRank is usually above 50 which suggests their trade network connectivity is satisfactory. Besides, the difference between PageRank and trade value rank is generally larger than the difference of OECD.
3. Even though China’s import value rank for pork is quite high, the import PageRank is consistently lower than its import value rank throughout the last twenty years. It may suggest that China’s partner portfolio for importing pork is not diversified enough.



Github Link:

https://github.com/OceanZacharyTam/CISC7210finalProject\_GroupR

1. <http://www.fao.org/faostat/en/#data> [↑](#footnote-ref-1)
2. Each element is equal to the export value from country j to i divided by the total export value of country j. If all element in a column is zero, each element will be substituted by 1/N [↑](#footnote-ref-2)