# International trade network and generalized version of hubs and authorities centrality

**Osman Akar** *, **Bin Liu** * **Zhijian Li** *

* *University of California, Los Angeles*

Submitted as Math 191 Final Report

International Trade Network | Hubs and Authorities Centrality | Weighted Directed Network Centrality

Abbreviations: ITN, international trade network;

## 1. Introduction

In this era of "globalization", the ITN, composed of countries (and economies) trading goods and services with each other, is one of the most important system illustrating the idea of a interconnected world. Trade could be naturally represented as a network, with countries as nodes and bilateral trade relations as edges. Compared to studies based on traditional economics models, the analysis of the ITN from a network science approach provides us with new understandings of the dynamics and interactions within the very complex patterns of trade activities. For example, De Benedictis et al. [1] studied the sectoral trade networks of several goods including bananas, oil, and engines, and calculated various centrality measures in the networks. It is not hard to understand how network structures could facilitate the study of international trades.

In this paper, we will measure the relative importance of countries in the ITN by "export centrality" and "import centrality", adaptions of the hub and authority centrality measure. Our variant of hub and authority centrality is built on a directed and weighted network, which we believe is most natural to represent the ITN. The design of this variant also incorporates a "bonus" component rewarding fast growth, which is a choice based on economics reasoning rather than network science models. We calculate the export centrality and import centrality and figure out they are consistent with well-known qualitative patterns in the trade data. We further use the calculated centralities to illustrate the dynamics within ITN, in particular the effect of 2009 Trade Crisis and correlation of export growth within regions.

## 2. Notation

• We behave international trade data in each year as a weighted graph with $N = 220$ nodes. For each year $t \in \{2003, 2001, ..., 2014\}$ we define adjacency matrices $\mathbf{A}^t$ where

$$\mathbf{A}^t_{ij} = \text{total export from country } i \text{ to } j \text{ at year } t$$

The values are in USD. These matrices will be called "trade adjacency matrices".

• For each year and country, we define total import($I^t_j$) and total export($E^t_i$) as follows

$$I^t_j = \sum_{k=1}^{N} \mathbf{A}^t_{kj} \qquad E^t_i = \sum_{k=1}^{N} \mathbf{A}^t_{ik}$$

• For each year's trade network , we will define two scores for each country $i$: export centrality $x^t_i$ and import centrality $y^t_i$. The precise definitions will be given at..... Moreover, we define column vectors $\mathbf{x}^t = (x^t_1, x^t_2, ..., x^t_N)^T$ and $\mathbf{y}^t = (y^t_1, y^t_2, ..., y^t_N)^T$ and we name them "export centrality vector" and "import centrality network" of year $t$, respectively.

• We write $\mathbf{B}^T$ for transpose of matrix $\mathbf{B}$ (avoid confusion with $\mathbf{B}^T$ and $\mathbf{B}^t$).

• For given year t, we define the diagonal import matrix $D^t_I$ such that

$$(D^t_I)_{ij} = \begin{cases} I^t_i & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

and similarly the diagonal export matrix $D^t_E$ such that

$$(D^t_E)_{ij} = \begin{cases} E^t_i & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

• In all notation, we may not use the time parameter either when the year is understandable from the text or when we are doing general cases. For example, $\mathbf{A}$ can be used as an adjacency matrix instead of $\mathbf{A}^t$ and $I_j$ and $E_j$ can be used instead of $I^t_j$ and $E^t_j$.

### 3.1 Data Source

We downloaded the data from OEC, The Observatory of Economic Complexity, at MIT. OEC got the data from BACI International Trade Database, a international trade database at product level constructed by CEPII, a French research center, using data from UN Comtrade. Hence, the data is originally from UN comtrade. The data contains international trade flows from 2003 to 2014 among about 230 countries (the number of countries varies from each year) at product level. After cleaning, we end up having 220 countries each year. Also, the data contains both import flows and export flows, which is redundant, so we constructed our adjacency matrices through export flows.

### 3.2 Data Extraction

• Split the data by year and remove the redundant part of data.
• Construct weighted directed graph (flows at product level are merged into total flows at this step), and extract the adjacency matrices and export them to CVS files.
• We cleaned matrices. Since we need to take inverse of adjacency matrices, we removed countries that have zero in-degree

**Reserved for Publication Footnotes**

or out degree in any year. Moreover, to make our matrices consistent, we cleaned up country secession. In particular, we combined Serbia and Montenegro together for years after 2006 to be consistent with Serbia and Montenegro (one country) in previous years. We also combined South Sudan and Sudan together to deal with South Sudans independence of Sudan in 2011.

• Use python to read csv files into numpy arrays.

## 4. Method

### 4.1 Review of Hubs and Authorities

In unweighted directed networks, we can assign two scores for each node: *hub score* and *authority score*. The *authority score* somewhat measures "the importance" of the node in the network, and the *hub score* somewhat measures "how correctly the node directs the important nodes(i.e. the nodes with high authority scores)". For example, for the citation network (where there is a directed edge from paper A to paper B if and only if A cited B). In this network, meaning of having high authority centrality is that the paper contains useful important information, and meaning of having high hubs centrality is that the paper cites important papers (in other words, it directs readers to the important papers).

Let the network contains $N$ nodes, and let its adjacency matrix be $\mathbf{A}$. The node $i$ has authority score $x_i$ and hub score $y_i$. In Kleinberg's approach, we want the scores satisfy the following equations

$$x_i = \alpha \sum_{j=1}^{N} \mathbf{A}_{ij} y_j$$

$$y_i = \beta \sum_{j=1}^{N} \mathbf{A}_{ji} x_j$$

for all $i \in \{1, 2, ..., N\}$, where $\alpha > 0$ and $\beta > 0$. So we want "the authority score of a node is to be proportional to the sum of hub scores of nodes that point to it, and similarly we want the hub score of a node is to be proportional to the sum of authority scores of nodes that it points to. If we write $\vec{\mathbf{x}} = (x_1, x_2, ..., x_N)^T$ and $\vec{\mathbf{y}} = (y_1, y_2, ..., y_N)^T$, we can simplify the equations as

$$\vec{\mathbf{x}} = \alpha \mathbf{A} \vec{\mathbf{y}}$$

$$\vec{\mathbf{x}} = \beta \mathbf{A}^T \vec{\mathbf{y}}$$

### 4.1 Import and Export Centralities

The most important part of this project is the definition of the "import centrality" and "export centrality". They are generalized version of the hubs and authority centralities for weighted directed graphs.

Let's start our formal definition. For fixed year $t$, let $\mathbf{A}$, $x_i$, $y_i$, $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$ be defined as in the notation. We want the equations

$$x_i = \alpha \sum_{k=1}^{N} \frac{\mathbf{A}_{ik}}{I_k} y_k + \beta_i^1 \qquad (1)$$

$$y_i = \alpha \sum_{k=1}^{N} \frac{\mathbf{A}_{ki}}{E_k} x_k + \beta_i^2 \qquad (2)$$

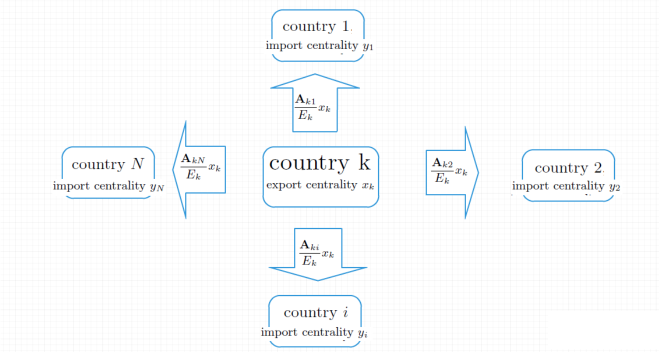


**Fig. 1.** The Distribution of Export Centrality among Other Countries' Import Centralities

satisfy for each $i \in \{1, 2, ..., N\}$. Here $\alpha$ and $\beta_1^1, \beta_2^1, ...\beta_N^1, \beta_1^2, \beta_2^2, ...\beta_N^2$ are constant parameters that do not depend on years. Our precise way of choosing these parameters will be given at **4.4** and **4.5**. Let's comment on the equations. Consider a country $k$ (see Fig 1). The export for this country is import for the other countries. Therefore it is reasonable that the export centrality $x_k$ increases the other countries' import centralities. So we want to distribute $x_k$ among $y_i$'s. In the equations, we chose the distribution so that $y_i$ gets fraction of $x_k$ proportional to $\frac{\mathbf{A}_{ki}}{E_k}$, which is fraction of net export from $k$ to $i$ over total export of $k$.

In the equations, we also see constants $\beta_i^1$ and $\beta_i^2$ are added to $x_i$ and $y_i$, respectively. These constants are meant to give a "bonus" to "some" countries, regarding of their economic growth. The way of choice of them will be given at **4.4**.

### 4.3 Calculating Import and Export Centralities

Firstly, let's define column vectors $\vec{\beta}^1$ and $\vec{\beta}^2$

$$\vec{\beta}^1 = (\beta_1^1, \beta_2^1, ...\beta_N^1)^T \qquad \vec{\beta}^2 = (\beta_1^2, \beta_2^2, ...\beta_N^2)^T$$

The equations in (1) and (2) can be written as

$$\vec{\mathbf{x}} = \alpha \mathbf{A} D_I^{-1} \vec{\mathbf{y}} + \vec{\beta}^1 \qquad (3)$$

$$\vec{\mathbf{y}} = \alpha \mathbf{A}^T D_E^{-1} \vec{\mathbf{x}} + \vec{\beta}^2 \qquad (4)$$

We can solve these equations as

$$\begin{aligned}\vec{\mathbf{y}} &= \alpha \mathbf{A}^T D_E^{-1} \vec{\mathbf{x}} + \vec{\beta}^2 \\ &= \alpha \mathbf{A}^T D_E^{-1} (\alpha \mathbf{A} D_I^{-1} \vec{\mathbf{y}} + \vec{\beta}^1) + \vec{\beta}^2 \\ &= \alpha^2 \mathbf{A}^T D_E^{-1} \mathbf{A} D_I^{-1} \vec{\mathbf{y}} + \alpha \mathbf{A}^T D_E^{-1} \beta^1 + \beta^2\end{aligned}$$

which is equivalent to

$$(Id_N - \alpha^2 \mathbf{A}^T D_E^{-1} \mathbf{A} D_I^{-1}) \vec{\mathbf{y}} = \alpha \mathbf{A}^T D_E^{-1} \vec{\beta}^1 + \vec{\beta}^2 \qquad (5)$$

where $Id_N$ is the $NxN$ identity matrix. Thus for a given year and parameters $\alpha$, $\beta^1$ and $\beta^2$, all the values except $\vec{\mathbf{y}}$ are known in the equation, so we can solve $\vec{\mathbf{y}}$. Then we can simply use (3) to determine $\vec{\mathbf{x}}$. We used computer for calculation of solution of (5).

After finding all of the centralities, we are going to normalize them as follows.

$$\vec{\mathbf{x}} \to \frac{\vec{\mathbf{x}}}{\sum_{i=1}^{N} \mathbf{x}_i} \qquad (6)$$

so that sum of all export centralities will add up 1. We do the same change for import centrality.

### 4.4 Choice of The Vectors $\vec{\beta^1}$ and $\vec{\beta^2}$

$\vec{\beta^1}$ and $\vec{\beta^2}$ in formulas (1) and (2) are designed to give a "bonus" to specific countries, which we believe should have greater importance in the ITN than what is reflected in the trade volumes. In our model, we decide to give such bonus to fast-growing countries, which lead the growth of ITN and global economy. The $\vec{\beta}$'s may look similar to the random teleportation term in the PageRank model, but there are different motives behind them.

We measure "fast-growing" in the following method: to determine $\vec{\beta^1}$ in formula (3), we will rank countries according to their geometric average growth rate of total export over the period of study, $g_i^E = \left(\frac{E_i^{2014}}{E_i^{2003}}\right)^{1/11} - 1$, and assign $\beta_i^1 = \beta$ for the top countries in the ranking, and $\beta_i^1 = 0$ for others(Here, $\beta$ is a constant parameter as $\alpha$). To minimize the influence of noise from countries with very small trade volume (whose growth rates are often very high only due to size effect), only those countries that exceed certain threshold are ranked. In particular, let $\max E^{2014}$ denote the total export of largest exporter in 2014 (in our data it is China), only those countries with $E_i^{2014} \geq \Theta_1 \cdot \max E^{2014}, \Theta_1 \in (0,1)$ are included in the ranking. Then these countries are ranked according to $g_i^E$, and the top $\Theta_2 \in (0,1)$ fraction of eligible countries get assigned $\beta_i^1 = \beta$.

For $\vec{\beta_2}$ in formula (4), we use the same method but rank according to growth rate of total import $g_i^I = \left(\frac{I_i^{2014}}{I_i^{2003}}\right)^{1/11} - 1$. We will still use the thresholds $\Theta_1$ and $\Theta_2$. Thus, the two vector parameters $\vec{\beta^1}$ and $\vec{\beta^2}$ are reduced into two scalar parameters $\Theta_1$ and $\Theta_2$.

Let's focus on choice of $\beta$. Assume that we multiply the equations (3) and (4) by some positive constant $C > 0$. Then we would get

$$C\vec{\mathbf{x}} = \alpha \mathbf{A} D_I^{-1} C\vec{\mathbf{y}} + C\vec{\beta_1}$$

$$C\vec{\mathbf{y}} = \alpha \mathbf{A}^T D_E^{-1} C\vec{\mathbf{x}} + C\vec{\beta_2}$$

so that choice of $C\beta$ instead of $\beta$ gives use new bonus vectors $C\vec{\beta_1}$ and $C\vec{\beta_2}$, and this values gives us export and import centralities $C\vec{\mathbf{x}}$ and $C\vec{\mathbf{y}}$

$$C\vec{\beta_1} \text{ and } C\vec{\beta_2} \quad \to \quad C\vec{\mathbf{x}} \text{ and } C\vec{\mathbf{y}}$$

but in the normalization process of vector centralities (see (6)), we would get the same normalized export and import vectors for $C\vec{\mathbf{x}}$ and $C\vec{\mathbf{y}}$. Therefore, WLOG, we can choose $\beta$ to be 1.

### 4.5 Choice of $\alpha$, $\Theta_1$ and $\Theta_2$

As we have seen, once parameters $\alpha$ and $\Theta_1$, $\Theta_2$ are determined, we could calculate centralities $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{y}}$ for each year. The remaining part of the model thus turns out to be how to choose these parameters.

Our main idea is that, since the export and import centralities measure the relative importance as a "share" in the entire ITN, if we distribute exports $E_i^{t+1}$ and imports $I_j^{t+1}$ according to the centralities, the results should be predictative for trade volume matrix $A^{t+1}$. Hence for different choices of parameters we calculate a prediction error and choose the lowest among them.

Therefore, we assume parameters $\alpha$ and $\Theta_1$, $\Theta_2$ are fixed, and calculate the prediction error in the following method:

1. Use the solving method in previous section to calculate $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{y}}$ in year 2003-2013
2. For $t = 2004, 2005, \ldots, 2013$, we make the prediction for following year

$$\widehat{A}_{ij}^{t+1} = \frac{1}{2}\widehat{E}_i^{t+1}\frac{\widehat{y}_j^{t+1}}{\sum_l \widehat{y}_l^{t+1}} + \frac{1}{2}\widehat{I}_j^{t+1}\frac{\widehat{x}_i^{t+1}}{\sum_k \widehat{x}_k^{t+1}}$$

where $\widehat{E}^{t+1}, \widehat{I}^{t+1}, \widehat{\vec{\boldsymbol{x}}}^{t+1}, \widehat{\vec{\boldsymbol{y}}}^{t+1}$ are computed from a linear projection using actual and realized data. It is worth mentioning that the usage of a linear projection seems somehow arbitrary and there may exist better estimates for $\widehat{E}^{t+1}, \widehat{I}^{t+1}, \widehat{\vec{\boldsymbol{x}}}^{t+1}, \widehat{\vec{\boldsymbol{y}}}^{t+1}$. However our goal is not to provide as accurate prediction for $A_{ij}^{t+1}$ as possible, but to choose parameters $\alpha$ and $\Theta_1$, $\Theta_2$ that makes most economic sense. Therefore, we believe the imperfect in linear projection could be tolerated.

3. We do have not only the predicted trade volumes $\widehat{A}_{ij}$ but also the realized value from data, so we could calculate the mean absolute error across all entries in the matrix:

$$\langle err^{t+1}\rangle = \frac{1}{N^2 - N}\sum_{i \neq j}\left(\left|A_{ij}^{t+1} - \widehat{A}_{ij}^{t+1}\right|\right)$$

4. Calculate the mean absolute error for each year, and we want the parameters that minimize the average error across all years:

$$\langle err\rangle = \frac{1}{10}\sum_{t=2004}^{2013}\langle err^{t+1}\rangle$$

The prediction formula could be seen as the average of a "projected exports distributed according to import centralities" and a "projected imports distributed according to export centralities". It makes sense economic-wise. Countries do not have direct control of bilateral trade volumes $A_{ij}$ as they wish, and such amount could be volatile from one year to another. On the contrary, total exports $E_i$ and imports $I_j$ are more tractable and predicatable. Moreover, the distribution rule using centralities is consistent with the calculate of export and import centralities themselves, where $x_i$ is composed of $y_j$'s distributed according to trade volume $A_{ij}$.

We computed the average error for $\alpha$ from 0.6 to 1.2 with an increment of 0.05, $\Theta_1$ from 0 to 0.1 with an increment of 0.01, $\Theta_2$ from 0 to 0.5 with an increment of 0.05. The range of parameters and increments are chosen to make economic-wise sense and ease computational workload. During the computation we notice that prediciton errors are most sensitive to change in $\alpha$. Finally, we concluded that $\alpha = 1, \Theta_1 = 0.05, \Theta_2 = 0.4$ will produce the lowest error for our model.

## 5. Results

We use the parameters determined in the presious section, and calculated the export and import centrality for all 220 countries from 2003 to 2014. The results are shown in Appendix 1.

Below are the plot of top 20 countries with highest export centrality and import centrality in 2014, respectively(Fig 2 and 3). Noticeably, the top 20 exporters in 2014 takes 71% of total export centralities, and top 20 importers also takes 71% of total import centralities. The 71-29 division indicates that ITN is dominated by a number of major partners in terms of relative important.

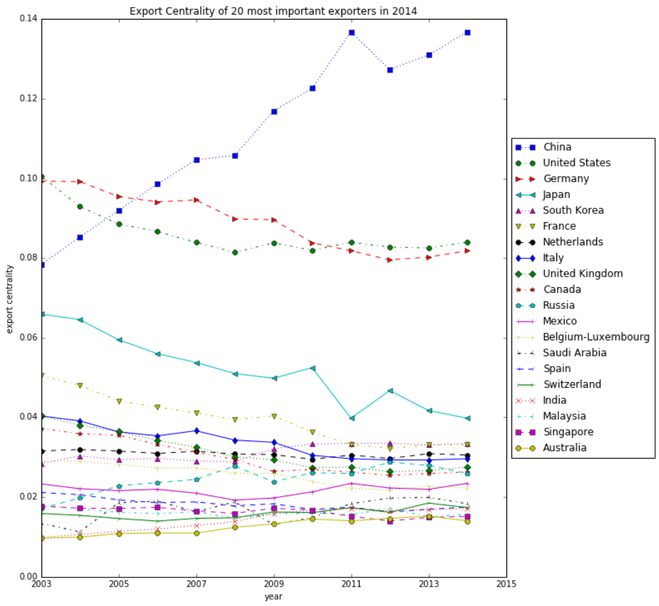


**Fig. 2.**

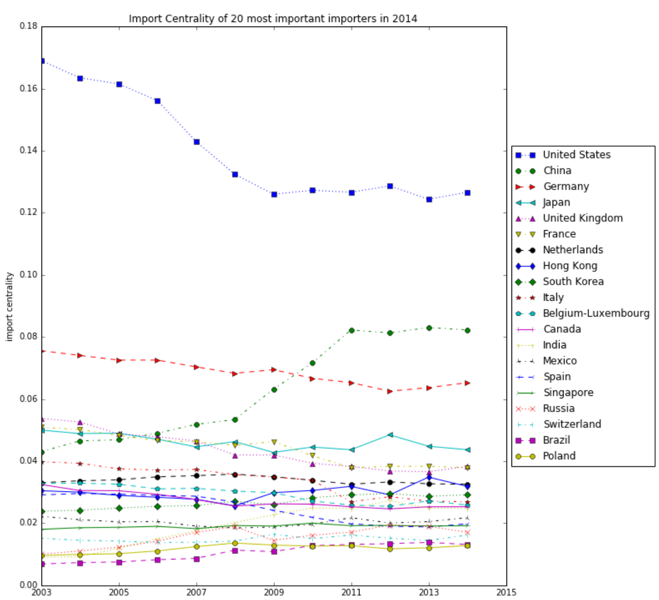


**Fig. 3.**

## 5.1 Result Discussion

As we can see from the plots above (Fig 2 and 3), the export and import centralities are consistent with some "ground truth" of ITN in classical macroeconomic studies: rise of China in exports, Germany standing out among European Union (EU), declining of Japan in exports, dominating role of United States, etc. Moreover, using the calculated centrality measures, we are able to observe the dynamics and interactions behind the ITN more closely than from classical macroeconomic approaches.

One observation is that the 2009 Trade Crisis had little affect on the trends of centralities. Below we plot four countries' trade volume and centralities: Japan, Netherlands, India, and Brazil (see Fig 4, 5, 6, and 7 ). All of them experienced a significant "bump" in both export and import volume in 2009, as clearly seen on the plots. In fact, the same shock happens to nearly all countries in the ITN, referred to as "2009 Trade Crisis" which took place right after the financial crisis in 2007-2008. Although the trade crisis had greatly striked trade volumes, the centrality measures are unaffected.

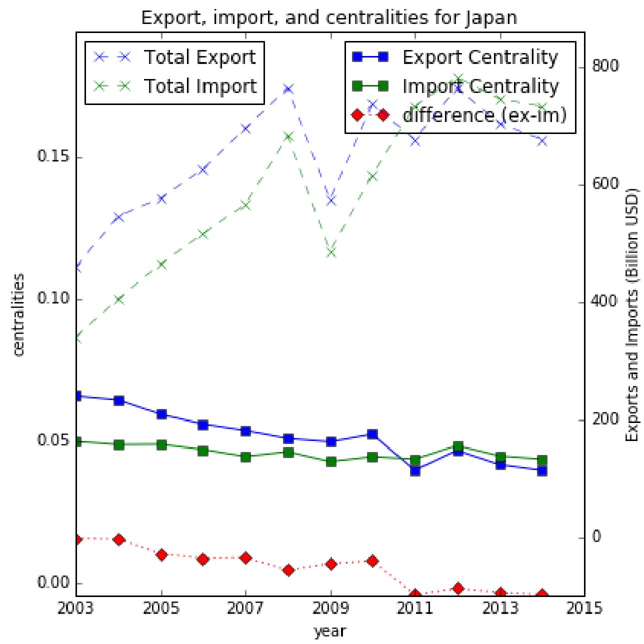


**Fig. 4.**

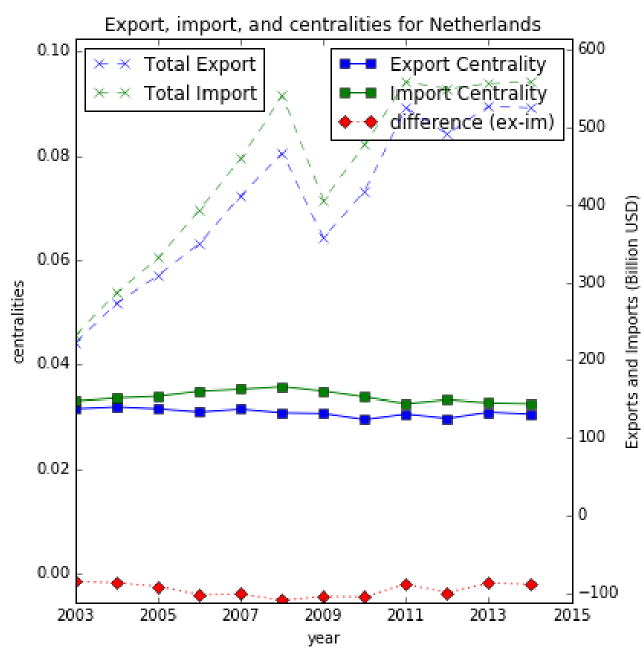


**Fig. 5.**

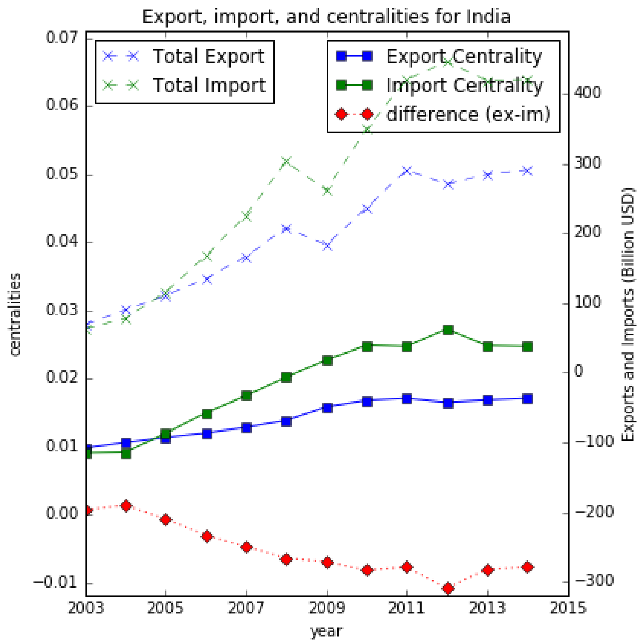


**Fig. 6.**

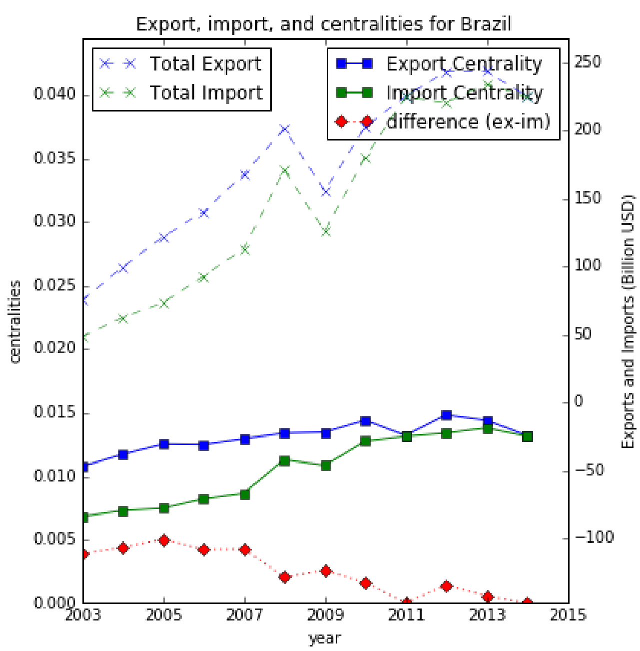


**Fig. 7.**

Another observation is the lack of correlation of export centralities for countries in the same region. We plot export centralities of countries in the same regions in the same plot (see Fig 8, 9, 10, and 11), although a weak common downward trend could be observed in Europe, we fail to see similar trends in North America, South Asia, or Southeast Asia. According to our centrality analysis, there is little evidence supporting the claim that regional correlation of exports are evident.

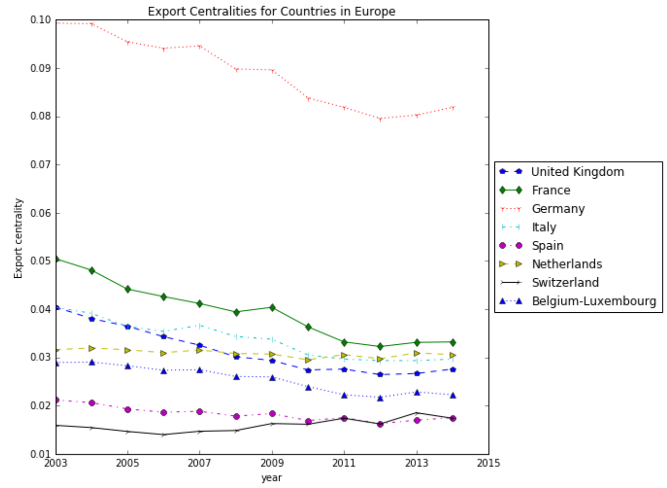


**Fig. 8.**

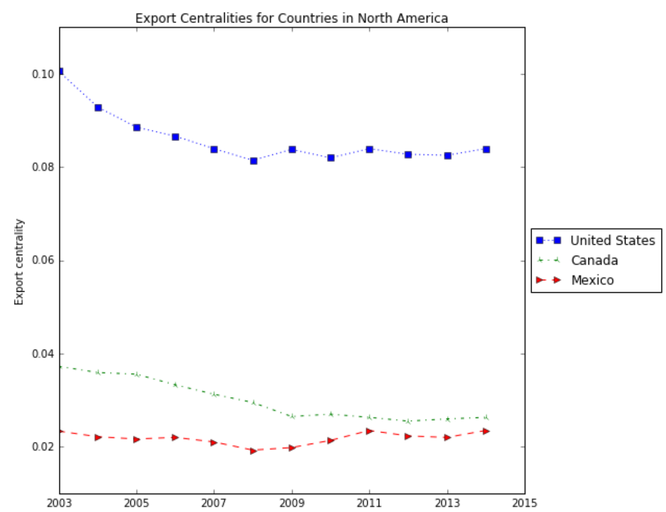


**Fig. 9.**

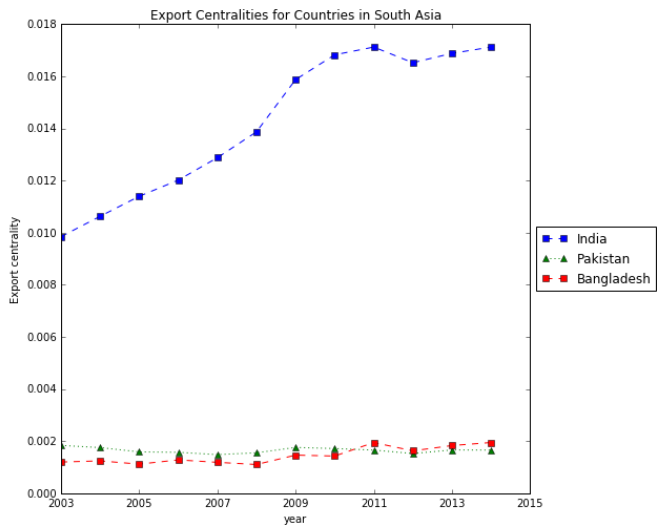


**Fig. 10.**

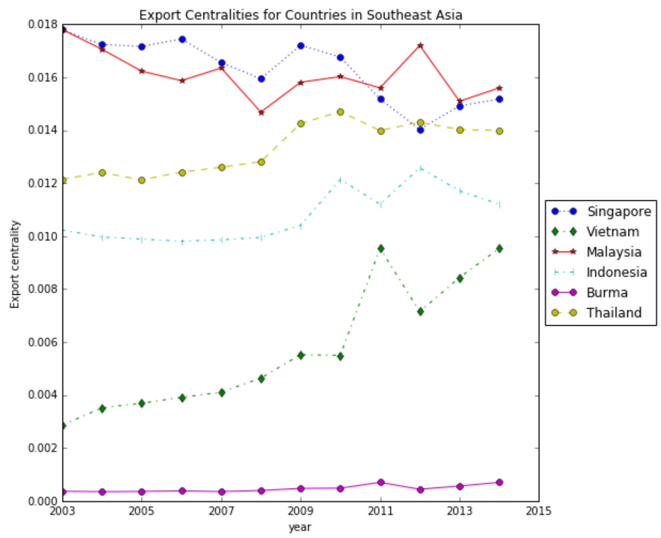


**Fig. 11.**

Another intersting question is that what trend in export and import centralities could we observe before the 2003-2014 period of our study. How did the collapse of USSR affect european countries? Was NAFTA more beneficial or not for United States? These questions could be answered using the same model as ours, on another dataset. Due to limitations on data availability and computation workload, we will not conduct the study here, but these are interesting and constructive areas to look at for future studies.

## 6. Conclusion

We adapted the hub and authority centrality measure to accomodate a directed and weighted network. The variant centrality applied on the International Trade Network is a proper and meaningful measure of relative importance of countries in global trades. We are able to use the result of the model to illustrate the dynamics behind ITN, which further contributes to future studies in both network science and macroeconomics.

## 7. Notes for Further Development

• In our model, we chosen $\vec{\beta^1}$ and $\vec{\beta^1}$ year independent(we only used 2014 data). To improve the model, these vectors can be chosen as $\beta^{\vec{t},1}$ and $\beta^{\vec{t},1}$ using each year's data.

• Our error function $\langle err^{\vec{t}+1}\rangle$ can be improved.

## Distribution of Work

Here the distribution of the project.

• The parts 1, 4.4, 4.5, 5, 5.1, 6 are written by Bin Liu
• The parts 2, 4.1, 4.2, 4.3, 7 are written by Osman Akar
• The parts 3.1, 3.2 are written by Zhijian Li
• The ITN research was offered by Bin Liu
• The export and import centrality model is built by Osman Akar
•The error function is built by Bin Liu
• The data extraction was done by Zhijian Li
• The optimization calculation in **4.5** was done by Bin Liu and Zhijian Li

1. De Benedictis, Luca, Network Analysis of World Trade using the BACI-CEPII dataset., Global Economy Journal, 3 (2014).
2. Hausmann, Ricardo, Hidalgo, Cesar A., Klinger, B., and Barabasi, A.The Product Space Conditions the Development of Nations