11/5/2018

Analysing the World Trade Network

SUBMITTED TO

Dr. Li Xiaoli

INSITITUTE OF SYSTEM SCIENCE
NATIONAL UNIVERSITY OF SINGAPORE

PREPARED BY

AASTHA ARORA (A0178188L)
AISHWARYA SUBASH CHANDRA BOSE (A0178277M)
ANJALI SINHA (A0178476L)
CHETNA GUPTA (A0178260A)
MISHA SINGH (A0178309W)
ZAIRA HOSSAIN (A0178331E)

Table of Contents

1. INTRODUCTION

Even before the emergence of topology sciences or graph theory, international trade had always been conceived as a network where trade flow was represented as a straight line between two points representing trading partners. Mutually intertwined supply chains in today's contemporary economy have resulted in a complex network of trade relationships with a highly non-trivial topology that varies with time. In order to understand the complex interrelationships among different countries and economic sectors, as well as their dynamics, a holistic view on the underlying structural properties of this network is necessary. This study aims to uncover certain key traits that characterize this network while identifying dominant nodes and how they govern the network.

1.1. OBJECTIVE

Use Graph/Network mining techniques to:

- Identify Key Players in the Global Trade Network
- Identify sub-networks or communities with dense intra trade flow
- Identify rising stars or evolution of key-players over time
- Identify if & how internal disturbances impact the overall trade flow and global trading prominence of a country or its partner nations.

2. RELATED WORKS

Some studies have been done in the area of network analysis and graph mining to study and analyse the trade routes. We have made use of trade export datasets from 2 different sources to serve our analysis. The UN ComTrade is the largest repository available online and has been used in trade analysis by various institutes to study the emergence and growth of trade. This has been used to study things like commodity contamination while exporting leading to food safety issues, fragmentation of trade by intermediary and final goods, studying intra-region trade dynamics etc.
The BACI-CEPII dataset has been quoted in over 451 papers on google scholars. This has been used to study emergence of new players in the world trade network and their rapidly changing roles, study the evolution, do a vulnerability analysis, comparison of different regions locally and globally etc.

3. DATASET

3.1. DATASET DESCRIPTION

3.1.1. UN TRADE STATISTICS DATASET

The International Merchandise Trade Statistics Section (IMTSS) of the United Nations Statistics Division (UNSD), Department of Economic and Social Affairs produces analytical merchandise trade tables containing trade values and indices for countries (areas) and regions. These tables are then published in the United Nations Monthly Bulletin of Statistics (MBS), from where they have been fetched.
Website: https://unstats.un.org/unsd/trade/data/tables.asp#quarterlyindices
Dataset: http://comtrade.un.org/api/getmbsdata?series_type=T41.A.V.$&fmt=CSV

3.1.2. BACI-CEPII DATASET

There are different organizations which maintain and publish world trade and data and related statistics, such as the United Nations, World Bank etc. The French Research Centre CEPII has come up with Trade data based on its own research efforts, and this is called the BACI dataset. This dataset gives us a rich

insight into ComTrade spanning over multiple countries and over largest period of time. It has trade data based on various commodity categories too.
Website: http://www.cepii.fr/CEPII/en/welcome.asp
Dataset: https://db.nomics.world

3.2. DATA PREPROCESSING

The network graph has the countries/regions as the nodes and the directed edge denotes the flow of trade from one region to the other. The edge weight represents the total export value in USD.

3.1.1. UN TRADE STATISTICS DATASET

To bring the data into Gephi compatible format, certain columns had to be extracted and relabelled as follows:

Nodes Table

Original columns	Renamed Columns
country_code	Id
Country_english_name	Label

Edges Table

Original columns	Renamed Columns
country_code	Source
partner_country_code	Target
value	Weight
year	year

3.1.2. BACI-CEPII DATASET

The data was pulled through an API multiple times, as the data volume to be pulled at once is limited to 1000 entries/API call. This data was processed to the following format(GEPHI compatible) through a python script. Data was pulled for three commodities, namely, Agriculture, Crude Oil and Engine.

Nodes Table

Original columns	Renamed Columns
exporter/importer	Id
Exporter/Importer	Label

Edges Table

Original columns	Renamed Columns
exporter	Source
importer	Target
value	Weight

4. IMPLEMENTATION

For the assignment we have utilized the **Gephi Software** for network analysis.

Advantages of Gephi

- Platform robust (PC, Mac, Linux)
- Good User Interface
- Open source
- Usable for both simple and complex network
- Attractive visualizations
- Comprehensive outputs
- No coding required

Disadvantages of Gephi

- Works best for single nodal/edge types
- Less flexible and nuanced than coding languages such as R

We also tried using R programming but faced the following issues:-

- The R programming takes a long time to process the code as compared to the Gephi software and cannot handle large networks
- Each node of the network needs to be programmed which is a tedious task, also take a long time and needs high level programming skills
- It is less user-friendly than a point and click program
- Visualization needs to be beautified and not as visually attractive as Gephi

5. NETWORK STRUCTURE

The network can be characterized by the following key attributes:

5.1. INTER NODE DISTANCES

- Network Diameter = 2
  This indicates that every country on the network is connected to every other country through a maximum of 1 intermediate node.
- Average Path Length = 1.177
  This indicates that in order to get to any 1 country from any other country, the Average no of nodes to be traversed is 1.177

These values indicate that it is a **highly connected dense network** which is in resonance with the concept of Globalization.

5.2. DEGREE DISTRIBUTION

The distribution of degrees within the network is highly skewed as seen in fig 51.1 such that out of a given total of 240 nodes:

- Majority of the nodes have degrees between 1 to 16
- Remaining 16 nodes have degrees greater than 150
- There exist no nodes with degrees between 20 to 150

This proves that the network is **not entirely homogeneous** in its connections and though, as per the Avg Path Length it seems to be a dense network, it is only so due to the connectedness of the top 16 players, whose absence would render the network sparse and disjoint.

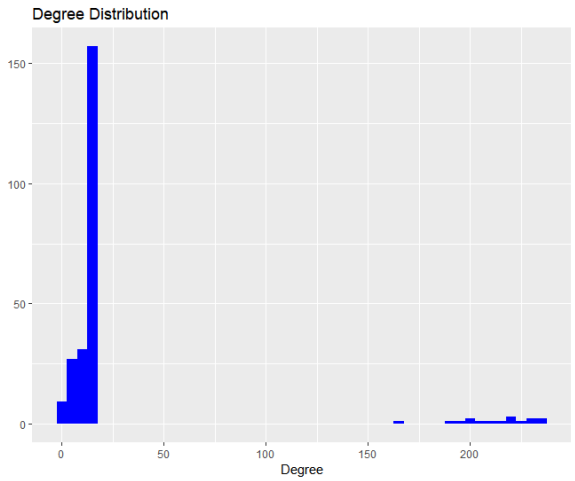

Fig 5.1: Frequency Distribution of Degree Nodes

6. KEY PLAYERS & COMMUNITIES

6.1. KEY PLAYERS

Identifying Key Players within the world trade network would help identify those nations whose economies govern most of the global trade by virtue of being large consumers/producers or being centrally located. Hence any impact to their economies can drastically impact global trade flow and in turn the economies of small nations which are dependent on these nations for either resources or revenue.

Since it is a directed & weighted network, major importers and exporters have been identified as follows:

6.1.1. Key Importers

Identified based on Weighted In-Degree to yield the following result as shown in fig 6.1:

- USA is the biggest importer
- Closely followed by China & other countries

6.1.2. Key Exporters

Identified based on Weighted Out-Degree to yield the following result as shown in fig 6.2:

- China is the biggest exporter
- Closely followed by USA & then Germany

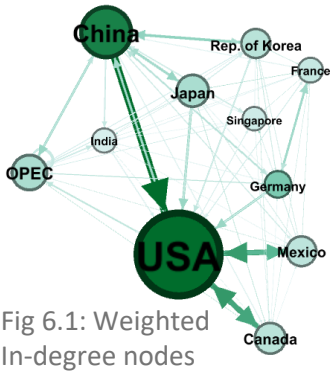

Fig 6.1: Weighted In-degree nodes

Fig 6.2: Weighted Out-degree nodes

6.2. COMMUNITY DETECTION

Louvain's modularity community detection algorithm was used to identify subsets of densely connected communities within the global network. The principal concept of the algorithm is:
Optimization of the "weighted modularity" of a partition of the network based on the following 2 step approach:
i) Locate "small" communities by optimizing modularity locally.
ii) Aggregate nodes belonging to the same community and build a new network with communities as nodes.
Post calculation of the modularity, the following settings were applied to visualize the results:

Settings - Node Size = Weighted Degree
Node Colour = Partitioned by Modularity
Edge Size = Weight ($ value of trade volume)
Layout = Fruchterman Reingold + No-overlap

Hence the following set of 3 communities is obtained as shown in fig 6.3:

6.2.1. Communities Identified

1. Community 1 – **Asian Cluster**
   Community 1 primarily comprises of Asian countries with China, Oil Producing Economies, Korea and Japan as prominent key players. The quality or strength of interconnections within this community is high as per the following modularity level:
   Modularity = 0.45
2. Community 2 – **European Cluster**
   Community 2 primarily comprises of European countries and Russia with Germany, France and Russia as prominent key players. The quality of this community is not as good as the Asian community with a lower modularity indicating that the strength of connections within the community is not very high compared to its connections with other communities.
   Modularity = 0.21
3. Community 3 – **American Cluster**
   Community 3 primarily comprises of countries from the American subcontinent with USA, Canada and Mexico as prominent key players. It is the weakest community with the lowest modularity.
   Modularity = 0.16

The generated communities have a clear geographic distinction with mostly countries of 1 continent being clustered together. This indicates that trade density is higher between countries of the same continent which could be due to the fact that close proximity of neighbouring countries allows for cheap transportation, hence encouraging trade.

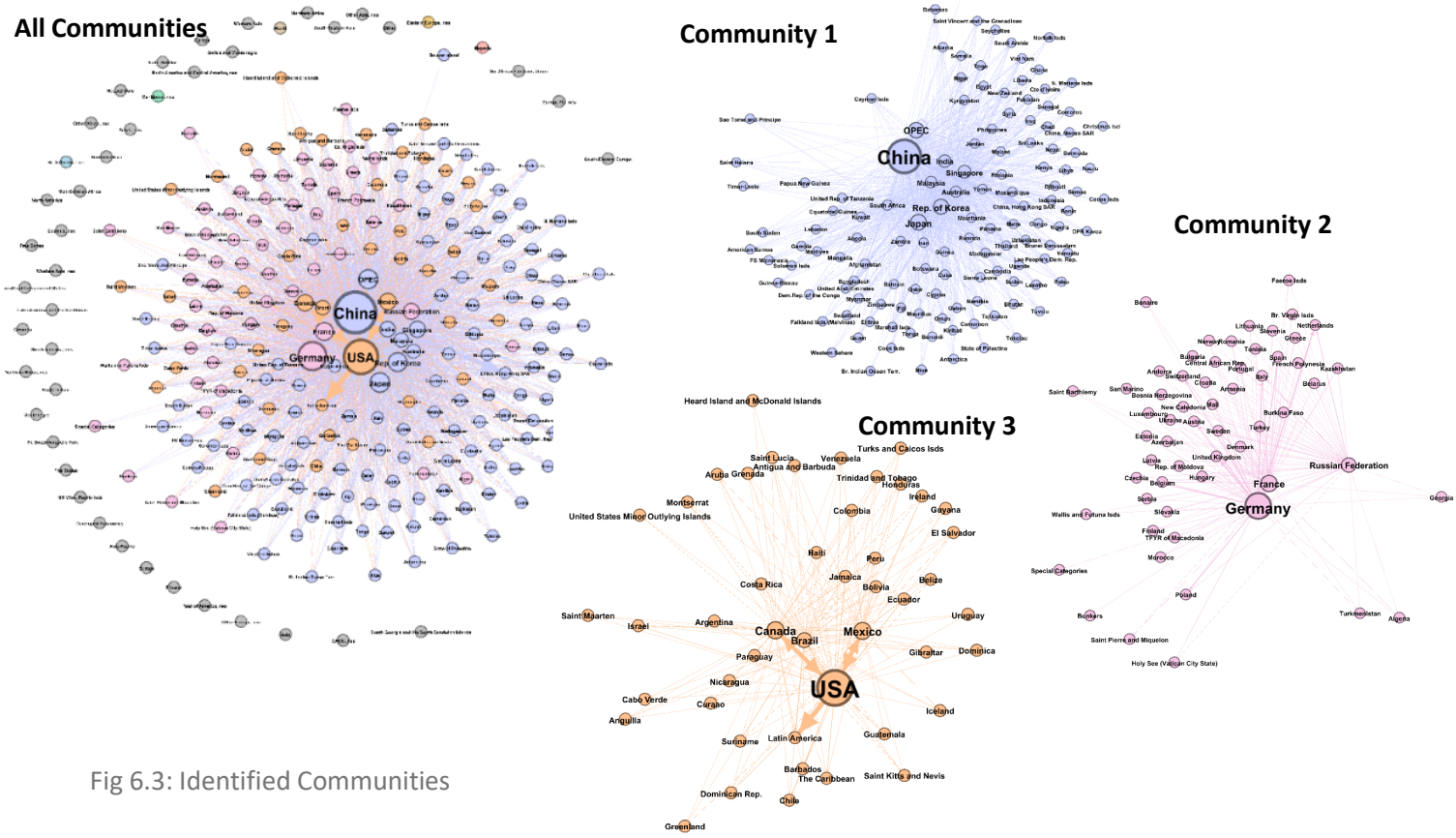

Fig 6.3: Identified Communities

6.2.2. Exceptions to Community Classification

Since Africa does not have a dense trade network within itself, the African countries mostly get assigned to either the Asian cluster or the European cluster. Close analysis of these countries reveals that most of the eastern African countries like Egypt, Ethiopia and Kenya get assigned to the Asian cluster while any of the north western countries like Morocco or Algeria get assigned to the European cluster. This implies that the north western countries have denser trade with Europe located adjacent to their north while the eastern countries have denser trade with Asia located adjacent to their east.

Hence, our above hypothesis is confirmed that **closely located countries** have **denser trade** owing to ease of transport.

7. DYNAMIC ANALYSIS

Analysing the Trade network over time can help uncover evolving trade alliances, rising key players and the impact of various internal socio-economic factors on a nation's trade as well as that of its smaller trading partners. Thus, as a part of this study the following 2 aspects have been analysed at a dynamic level:

7.1. EVOLUTION OF KEY PLAYERS

A study of the world trade network over a period of 15 years from 2000 to 2015 reveals the following notable changes in the positions of key players as indicated by fig 7.1:

- China has grown massively in the past 15 years to come at par with USA as a key trading centre,

by gaining prominence not only in terms of trade value but also in terms of no. of trading partners.

- European countries like Germany and France have retained their importance as central to the global trade network, owing more to their interconnectedness than size.
- Korea & India have shown a steady rise in prominence as global trading centres.

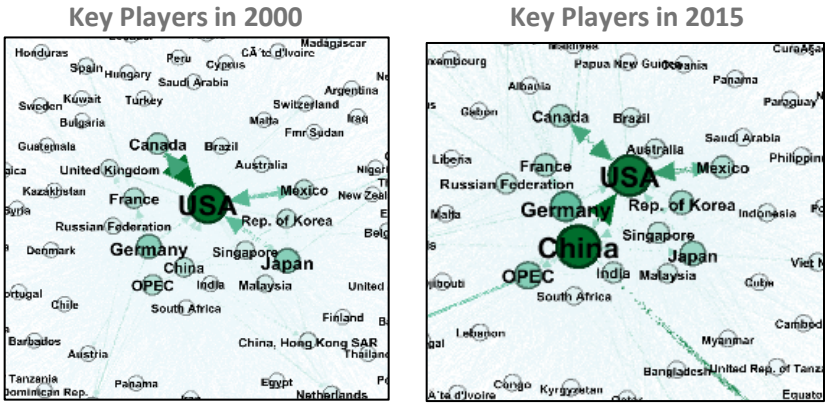

Fig 7.1: Key Players of 2000 and 2015

7.2. IMPACT OF INTERNAL SOCIO-ECONOMIC FACTORS ON A NATIONS'S TRADING PROMINENCE

To study how any internal disturbances within a nation can impact its position as a key player and in-turn affect its smaller partner nations, a case study of Japan has been selected.

Japan's trading connections and volumes were analysed in context of the rest of the world for the years 2010 and 2011 in order to study the impact of Japan's 2011 Triple Disaster that was triggered by an Earthquake culminating in a Tsunami and a Nuclear Meltdown.

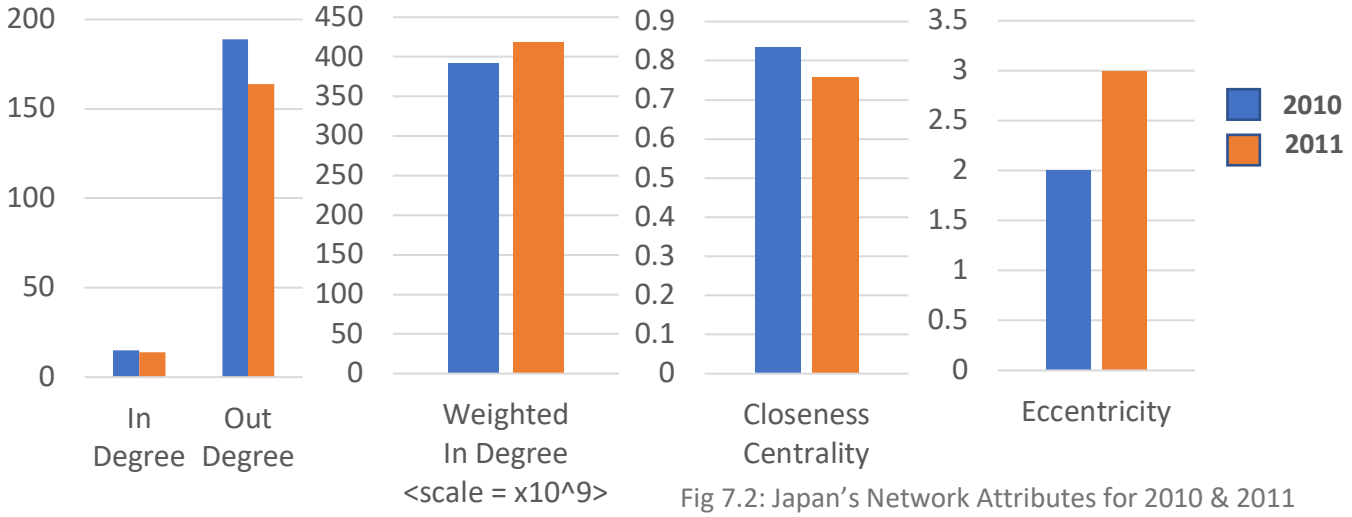

Fig 7.2: Japan's Network Attributes for 2010 & 2011

The following notable observations were made based on the parameters indicated in fig 7.2:

1. **Japan's Export Partners decreased**

As indicated by the decline in Out-Degree while In-Degree remained almost the same.

A probable caused could be Japan's inability to produce quality goods (specially agriculture) at the same pace as before due to the disaster. This could also have been aggravated by countries like USA intentionally barring imports from Japan for fear of Nuclear contamination.

2. **Net imports increased**

As indicated by the rise in Weighted In-Degree. This could be a result of increase in Oil imports to obtain an alternate source of power generation during the nuclear shutdown.

3. **Decline in prominence as a Trading Centre**

As indicated by the decline in Closeness Centrality implying that the avg length of the shortest path between Japan and any other nation declined. This is also supported by the increased eccentricity value from 2 to 3 implying that the maximum no. of intermediate nodes to be traversed to reach any other country from Japan, increased from 1 to 2.

It was hence concluded that any internal disturbances do adversely impact a nation's position in the global trade network as can be seen by Japan's **decline in connectivity** and **incurrence of a trade deficit,** which in turn is likely to impact smaller trading nations with low degrees that trade only with Japan and maybe another one or two countries.

8. COMMODITY SPECIFIC ANALYSIS

8.1 AGRICULTURE TRADE NETWORK

Agriculture is a commodity that depends on the land area of a country and labour available thus the key players in this sector are the following countries:

- **USA** : It has a large land area and low population density. It is a large exporter of agricultural bulk commodities and high like wheat, rice, coarse grains, oilseeds, cotton, vegetables, etc. The major imports of horticulture products such as Coffee, Sugar Cocoa and rubber.
- **China** : It has a high population density and low labour cost. China's tends to import land-intensive commodities (soybeans, cotton, barley, rubber, and oils), and exports labour -intensive commodities and processed agricultural goods)
- **Germany** : Germany is the second largest importer(eggs, vegetables, cheese, cocoa) third largest exporter of consumer oriented agricultural products(milk, cereals, potatoes, wheat, barley, and sugar beets), and by far the most important European market for foreign producers
- **Netherlands** : Netherlands serves as the main grower of fresh products to European market because of its climate(tomatoes, sweet peppers, cucumbers, meat and dairy products).Netherlands also acts the hub for receiving goods at its ports and exporting them to the European countries.
- **India** : India has a high population density and low labour cost and a tropical country which makes it an exporter of products like rice, vegetables, fruits, groundnuts and imported high value customer goods such as coffee, vegetable oil etc.

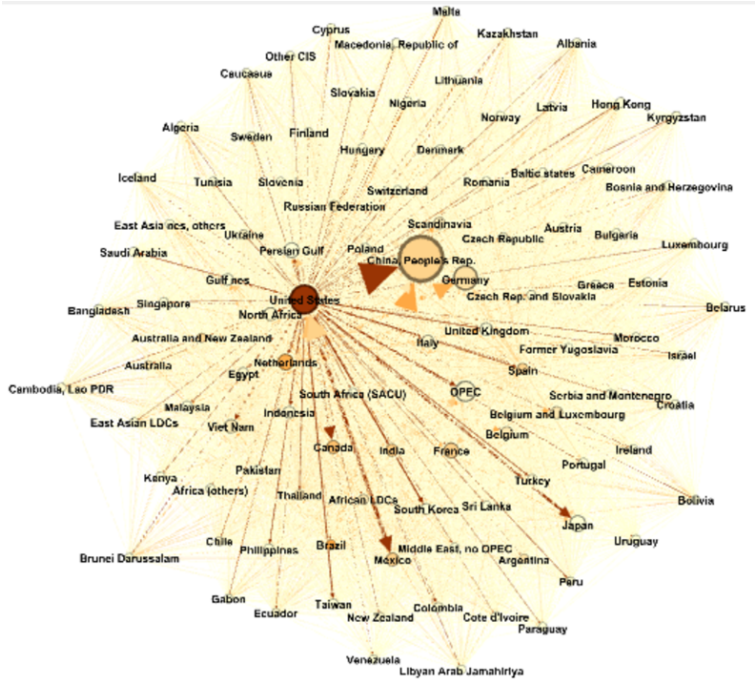

Fig 8.1: Agriculture Key Players

The node colour intensity in graph depicts the volume of trade and shows the same trend.

KEY INSIGHTS:

It is also noted that the strongly connected component analysis for the agriculture commodity in the trade network gives us the value 1, which mean that the trade network for agriculture is very strong and all the countries/ regions are connected. Removal of some key players does not affect the trade network so badly and is quite self -sustaining. This however is not observed of Crude oil and Engine as commodities due to some regions having dominance over the others in terms of raw materials, technology and availability.

On applying Louvian's modularity Community Detection Algorithm to the Agriculture Trade Network, the following 3 communities emerged:

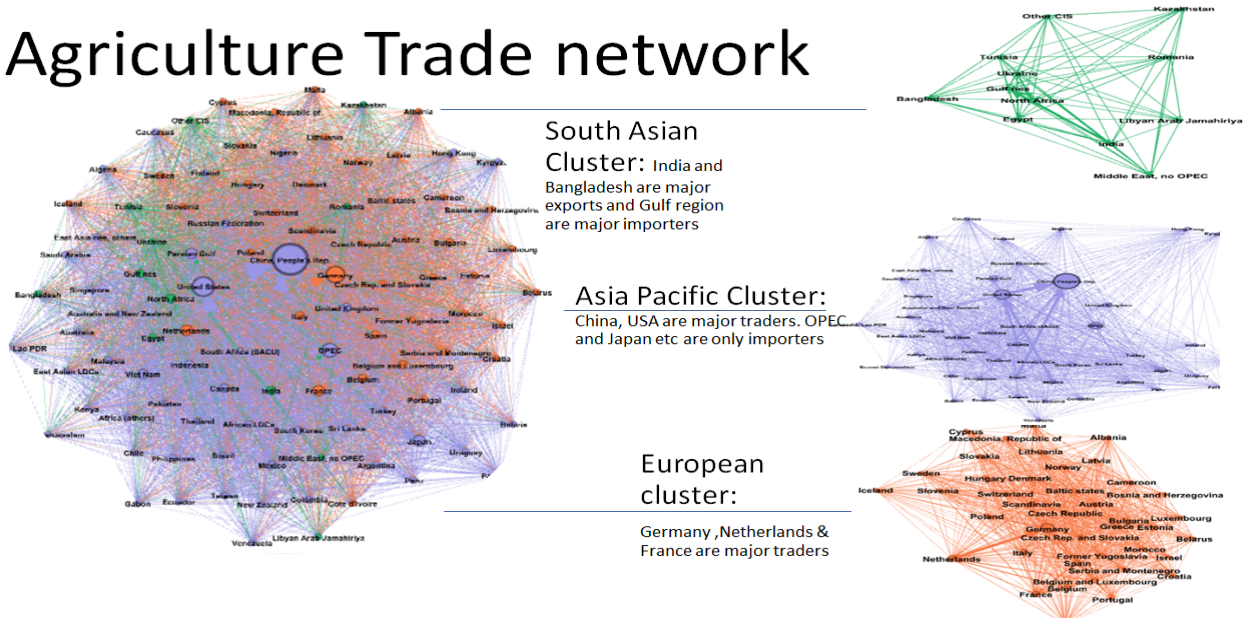

Fig 8.2: Agriculture Trade Communities

8.2 OIL TRADE NETWORK

The oil export network has key players like the **OPEC Nations**, **Persian Gulf Africa** and **Middle east**. Meanwhile the major Importers are **Asia and Oceania**, **North East Asia**, **India**, **Japan etc.**

- **Russia** plays a central role in the European trade system and Kazakhstan acts as a secondary hub in Central Asia. The country acts as a bridge between Europe and China. China has been investing heavily in Kazakhstan's oil and in the country's infrastructure, including pipelines, highways and rail roads, and refineries. China's purchases of Kazakhstan's oil, gas and minerals has booming, and Hong Kong is the gateway to China.
- **Latin American** countries appears as composed of two separate cohesive sub-groups. On one side, Venezuela, Mexico, Ecuador coordinate the northern cluster. Brazil, Argentina and Colombia coordinate the southern one, both groups of leading countries keep the intra-continent trade connected and they also link the continent to the rest of the world.

The clustering of the networks as per the modularity forms the three following clusters:-

0. Orange Cluster : Cluster of the oil imported by African countries, BRIC nation, China from a number of small nations.
1. Purple Cluster : Cluster of Middle size players that import and export among each other these include the European nations and Russia predominantly.
3. Green Cluster : This is a cluster of major traders and contains exporters like OPEC, Persian Gulf, Middle east and Africa and importers like Asia and Oceania, India, Japan, Pakistan , ASEAN and hubs like Singapore.

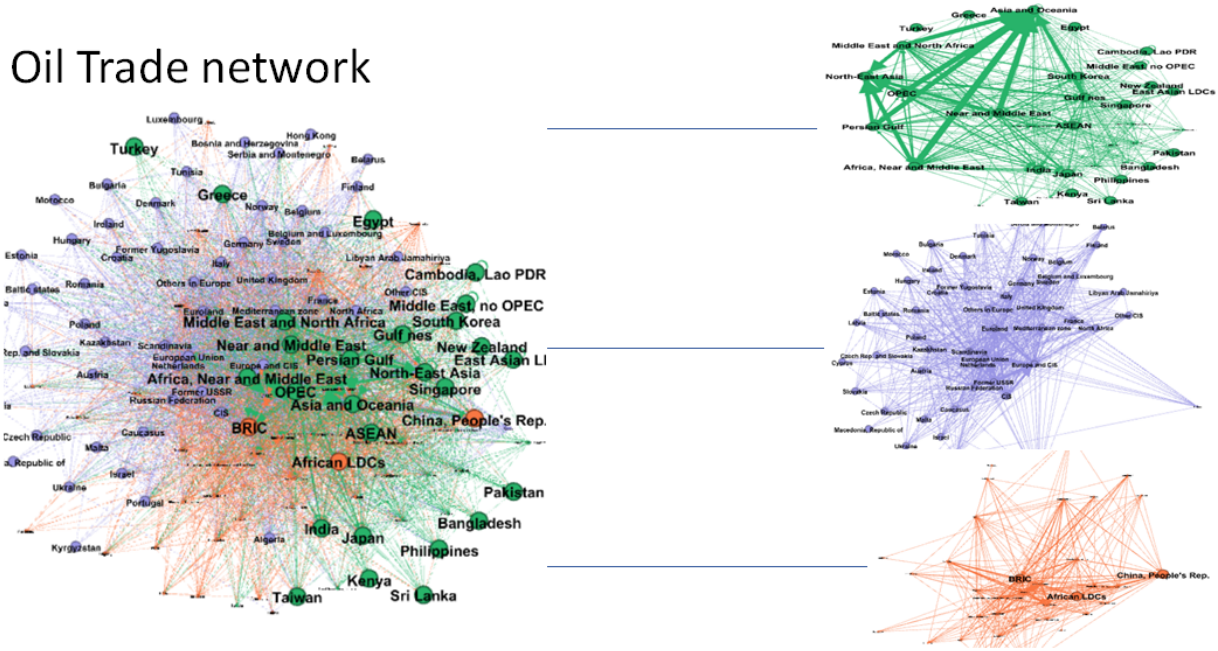

Fig 8.3: Oil Trade Communities

8.3 ENGINE TRADE NETWORK

The Automobile Engine export network has Germany, Japan and USA as the key players, in the same order.

- It is observed that China does not maintain its stand as a major producer in this sector.
- We also see South Korea, Thailand and Singapore as influencers in the Asia region.
- In the European trade system, we can notice the role of the United Kingdom, France and Spain and of some other European countries, such as the Czech Republic ,Hungary and Lithuania, exploiting their convenient geographical location and lower labour costs to boost international fragmentation of production from old European member countries.

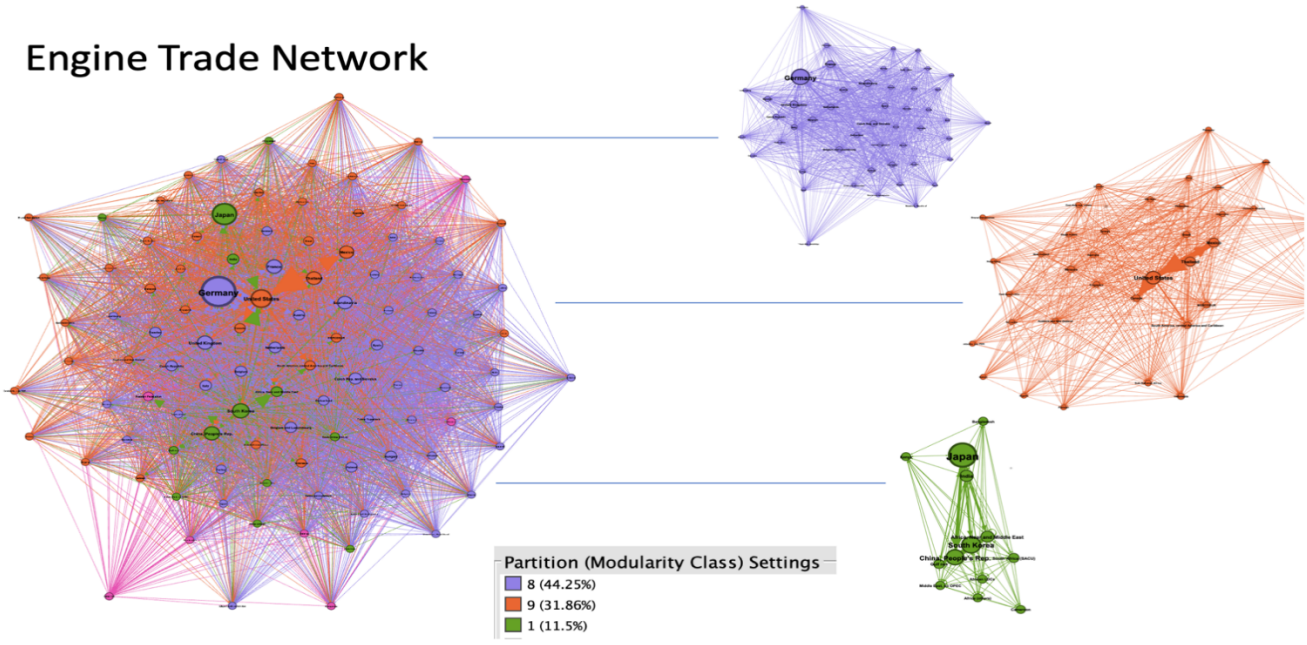

Fig 8.3: Engine Trade Communities

9. KEY FINDINGS & INSIGHTS

9.1. BUSINESS INSIGHTS

- The Global Trade Network is a densely connected network of countries dominated by mostly 16 countries.
- USA has continued to retain its prominence as a Key Player globally.
- In the past decade China has dramatically risen to dominate the global trade markets, outshining USA in terms of exports.
- Since it is a highly connected network, not many communities can be separated out.
- Trade quantities & connections are higher between closely located countries, hence communities formed have a geographic distinction at the continent level.
- Any socio-economic disturbances in a country such as occurrence of a disaster or even social unrest can impact its position as a key-player. This, in-turn can further impact the trade and the economy of very small partner countries.

9.2. TOOL INSIGHTS

- Gephi provides only 1 weighted community detection algorithm that is based on Louvian Modularity
- Gephi provides a dynamic network feature which is meant to update the network parameters and animate its appearance based on an time interval parameter. However, this dynamic feature is incapable of computing weighted network attributes at a dynamic level and only produces aggregate weighted values for the entire timeseries
- There exists no layout feature to separate out communities based on Modularity.

10. FUTURE WORK

We have the following tasks as the future scope of our World Trade Network analysis :-

- Creation of a dynamic timeline that covers the changes in the volume of trade monthwise over a long period of time, to study the growth of trade over the years
- Carrying out network analysis for other important commodities and sub commodities
- Conducting analysis after normalizing trade value by GDP for each nation, in order to deduce a nation's reliance on trade.
- Using other tools such as Pajek, Neo4j etc. to overcome the analysis handicaps of Gephi and R.
- If possible, to do a comparative study of various graph analysis tools available.
- We also aspire to create a machine learning model to predict the export value over the coming years based on the historic data