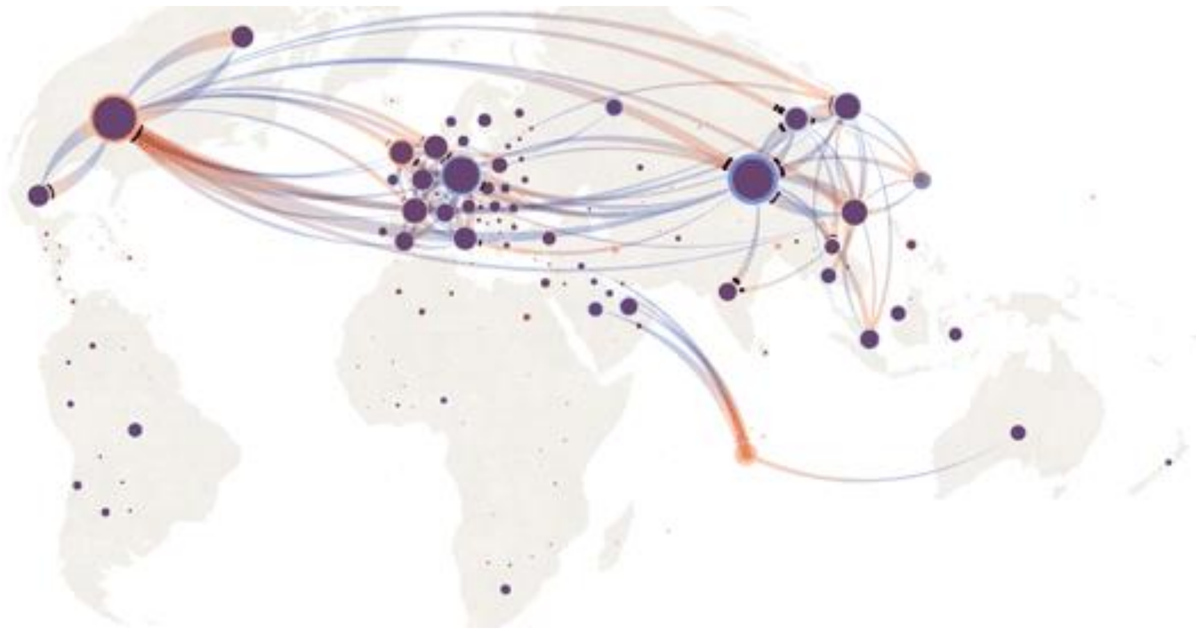


Analyzing Global Export Trade Values

Κωνσταντίνος Κωνσταντόπουλος

Μάθημα: Ανάλυση Κοινωνικών Δικτύων
Διδάσκων: Δρ. Δημήτριος Πουρναράκης

Φεβρουάριος 2021

Index

0. Introduction

The **world economy** or the **global economy** is the economy of all humans of the world, referring to the global economic system which includes all economic activities which are conducted both within and between nations, including production, consumption, economic management, exchange of financial values and trade of goods and services.

One of the features of each country's economic profile, is the import-export activity. The import activity characterizes and describes the needs of the country, be that food or prime materials for the country's industrial/manufacturing/agricultural or any other activities.

On the other hand, the export activity describes the importance of this country's products to the rest of the world. In that respect, if the countries to which each country exports are what we call powerhouses (such as U.S.A., Germany, Canada etc.), that means that this country's products are of the utmost importance to the powerhouse-countries, and that being the case there is a good indication that the economy of this country has solid foundations, since its products are essential to the "bigger" countries.

In this project, we focus on the export activities and therefore we will study and analyze the total export trade value of each country to another.

Through the analysis that follows, we wish to see whether the process of data through the given software will result in facts we already know. There is a number of issues we need to address:

- is U.S.A. or China the major powerhouse in the world?
- What are the specific countries that each distinct country focuses on exporting? In other words, what is the product orientation of each country?
- How does the geographical location of a country affects its export activity?

1. Dataset Extraction

The data was extracted from the official webpage of World Integrated Trade Solution (https://wits.worldbank.org).

The site offers an option to bulk download export trade values for all countries. It should be noted however, that for each country we only have data and information for the top 5 countries to which this specific country exports its goods. This does not constitute a limitation, since the top 5 countries appear to be the ones that make a big impact in the world's exporting activities.

So what we have is a zip file that contains all CSVs, each one corresponding to a country's top 5 list export value in dollars for the year 2018.

```
import glob
import pandas as pd

path = r'wits_en_trade_summary_allcountries_allyears' # path that cointains all files
all_files = glob.glob(path + "/*.csv")

li = []

for filename in all_files:
    df = pd.read_csv(filename,encoding="ISO-8859-1", index_col=None, header=0)
    li.append(df)

#put all CSVs together
frame = pd.concat(li, axis=0, ignore_index=True)

#apply specific filters
frame=frame[(frame["Indicator Type"]=="Export")&(frame["Product categories"]=="All Products")]
frame=frame[~frame['2018'].isna()]
frame=frame[(frame['Partner']!="Unspecified")&(frame['Indicator']=="Trade (US$ Mil)-Top 5 Export Partner")]
frame=frame[(frame['Partner']!=" World")]
frame=frame.loc[frame['Reporter']!=" World"]
frame=frame[['Reporter','Partner','2018']]


dff=frame.groupby('Reporter',as_index=False).sum().rename(columns={'Reporter':'Label'})
dff=dff['Label']

#rename columns so Gephi understand what each column corresponds to
frame=frame.rename(columns={'Reporter':'Source','Partner':'Target','2018':'Weight'})

dff.to_csv('nodes.csv', index=True) #our nodes file

frame.to_csv('edges.csv', index=False) #our edges file
```

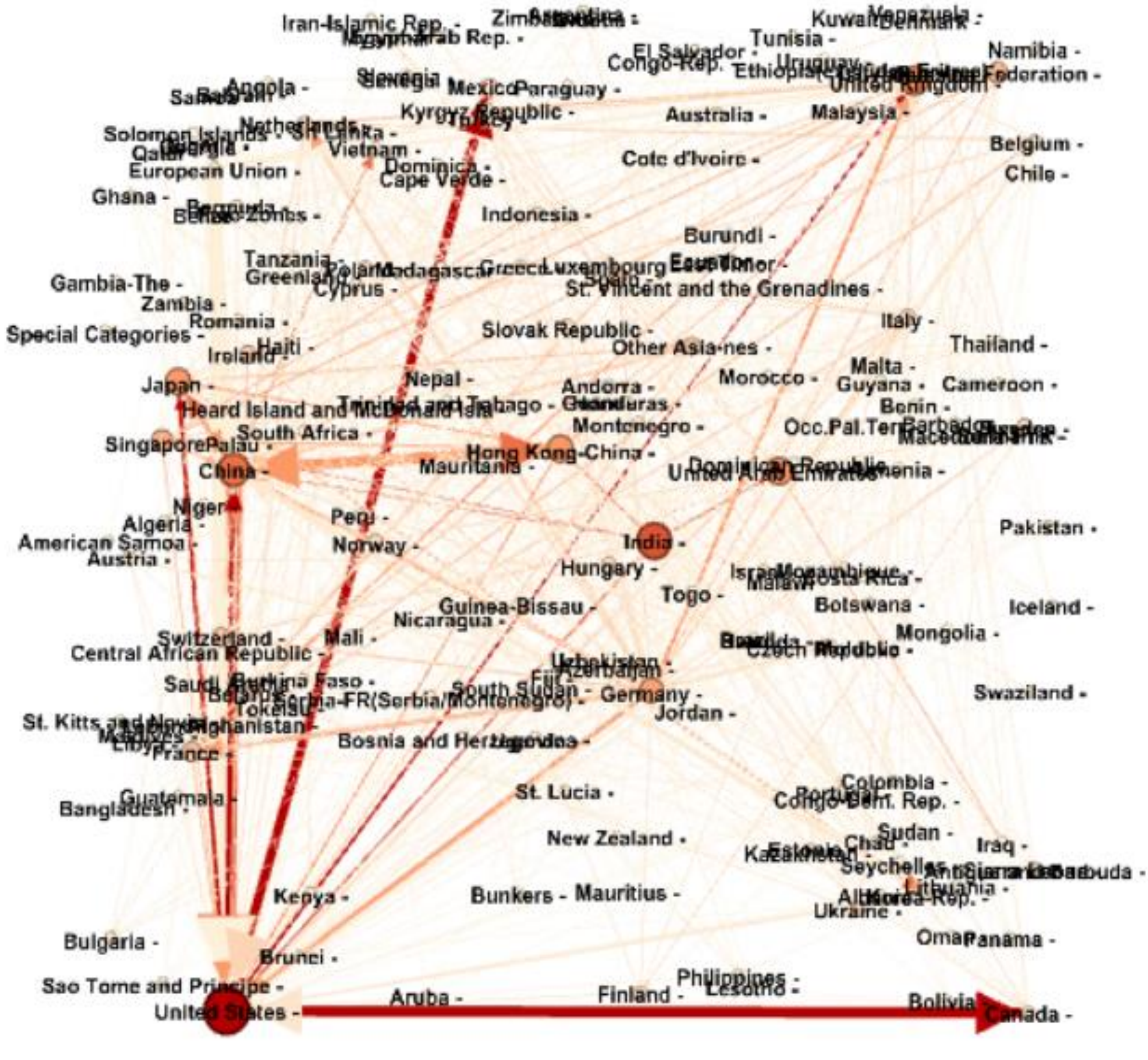
With the help of ***python*** we are able to combine the data that is included in all those excel spreadsheets into one big dataset.

Our dataset contains information on the economic and export activity of all countries in the past 40 years. We put emphasis on data relevant to exports that took place in 2018, because for that specific year we seem to a comprehensive picture for all countries involved. The other years appear not to contain information on the exports of some smaller countries, preventing us from making a full in-depth analysis.

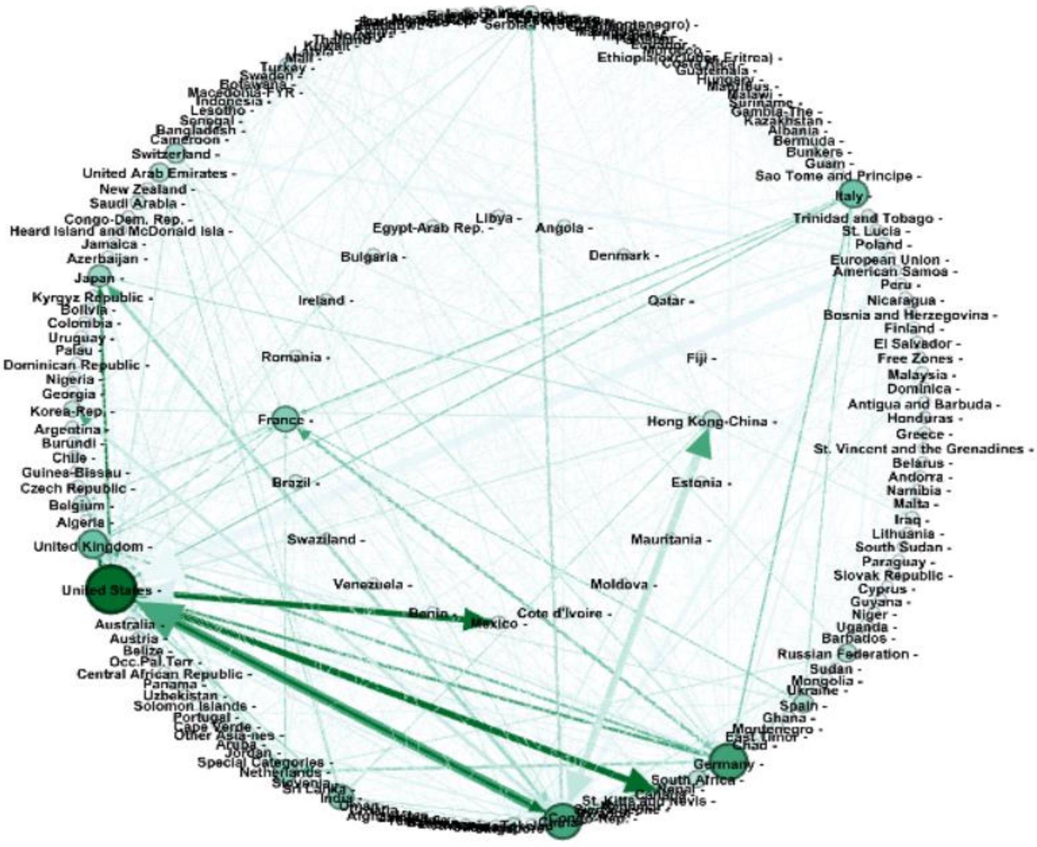
By filtering out some information that is of no use to us, we now have created our dataset. Two files are created and formatted in such a way that Gephi will be able to recognize which are the nodes (countries) and how they are connected with edges. Let’s import it into Gephi!

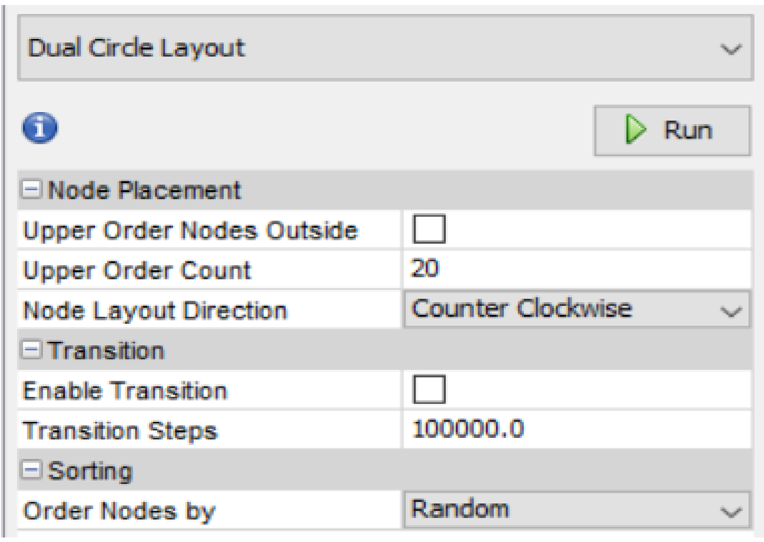
2. Visualizing the Network

The plain visualization of the above mentioned data as done by Gephi itself results in the graph below. All countries are depicted in a graph randomly located in the map but with the strength of export values clearly demonstrated. We also apply an edge weight ranking, showing how much larger an export is by how darker the color of the edge is.



For this specific project, it became apparent that the best visualization was made possible by the selection of the Dual Circle layout, as shown below.





3. Basic topological properties

Results:

Diameter: 13
Radius: 0
Average Path length: 2.9056603773584904

Nodes: |V|=177. Each node is a country that either exports or imports goods from and to other countries. Please note that there are some countries such as South Sudan, Ethiopia that only import goods but do not export (at least according to the given data by WITS).

Edges: |E|=744. Ideally, it should be 5*177=885. Nevertheless, for some countries such as Aruba the countries to which it exports to is less than 5.

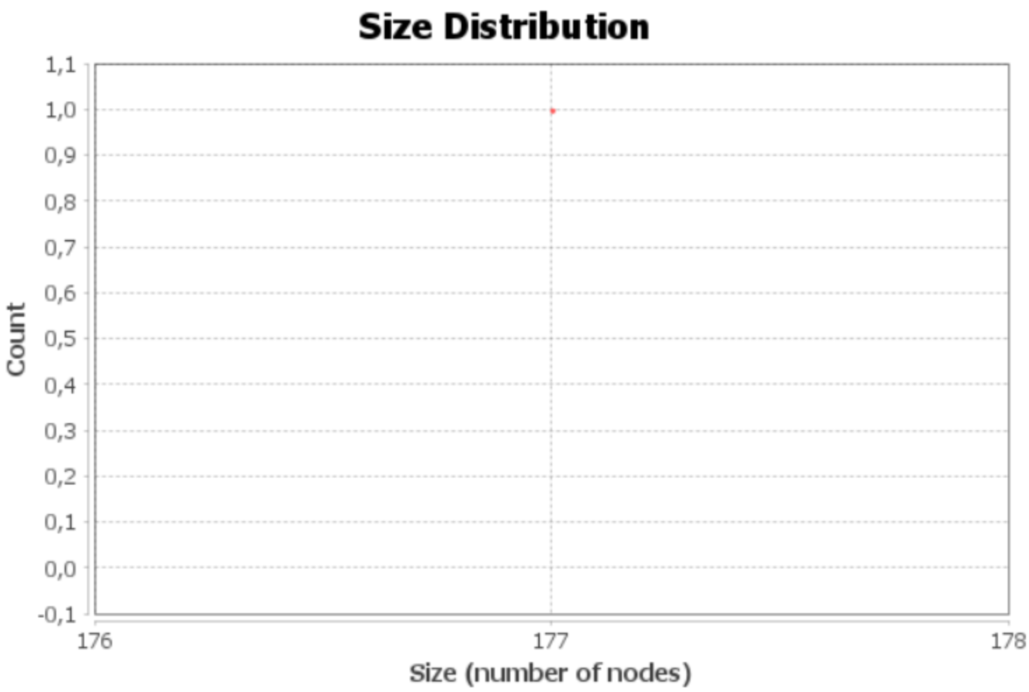
Diameter: In our topology the diameter, that is the longest shortest path between 177 nodes is equal to 13. In our study, the diameter is the longest chain possible involving countries exporting goods to each other. In other words, there is a longest chain of 13 countries which influence each other's economy by exporting their own goods. Practically, we could picture a small union of 13 countries constituting a small circle within which their goods circulate.

Average Path Length: It is not surprising that the average path length is 2.90, that is significantly smaller than the diameter. The reason being, out the 177 countries there are quite a few whose impact in the world trade is insignificant. Therefore, the path lengths involving these countries are of very small value (eg 1 or 2).

4. Component Measures

Results:

Number of Weakly Connected Components: 1
Number of Strongly Connected Components: 126



Connected components in a graph refer to a set of vertices that are connected to each other by direct or indirect paths. In other words, a set of vertices in a graph is a connected component if every node in the graph can be reached from every other node in the graph.

Here, we can see that our graph contains 126 strongly connected components and 1 weakly connected component. The component containing the largest number of countries is component #2; these countries constitute the countries that play the most drastic role in the export business: USA, China, UK, Japan, India, Germany, Italy, UAE, France and 17 more (see Table I below).

For the remaining components, we see that the classification in components is based on the geographical location; that is neighbouring are grouped in the same component. As we see in Table II below, component #11 contains countries of the former Soviet Union, component #12 contains countries of the former

Yugoslavia, component #14 contains the countries the Scandinavian countries and component #17 contains Greece with Cyprus and Bulgaria.

Netherlands	2
United States	2
Canada	2
China	2
India	2
Other Asia-nes	2
Germany	2
Italy	2
Spain	2
France	2
United Arab Emirates	2
Iran-Islamic Rep.	2
Saudi Arabia	2

Vietnam	2
Switzerland	2
United Kingdom	2
Japan	2
Korea-Rep.	2
Belgium	2
Malaysia	2
Singapore	2
Thailand	2
Mexico	2
Hong Kong-China	2
Portugal	2
Ireland	2

Table I: Component #2 including the most influential countries in the commercial world

Russian Federation	11
Belarus	11
Ukraine	11
Serbia-FR(Serbia/Montenegro)	12
Bosnia and Herzegovina	12
Croatia	12
Slovenia	12
Montenegro	12

Norway	14
Denmark	14
Sweden	14
Finland	14

Greece	17
Bulgaria	17
Cyprus	17

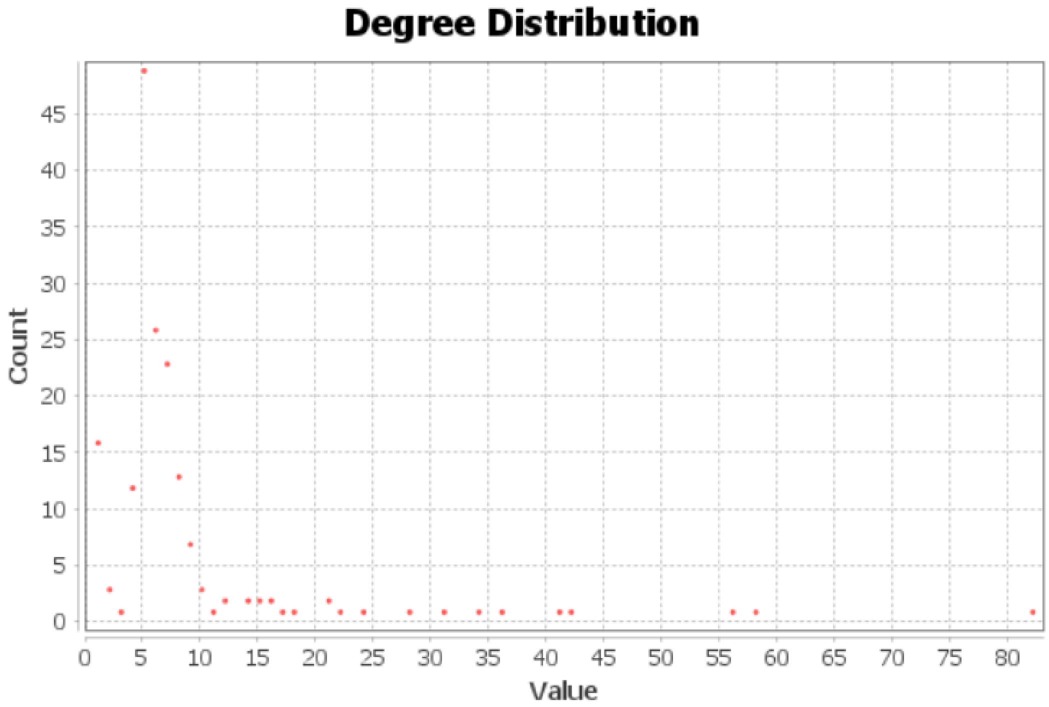
Table II: Components containing neighbouring countries

5. Degree measures

Degree Report

Results:

Average Degree: 4,203

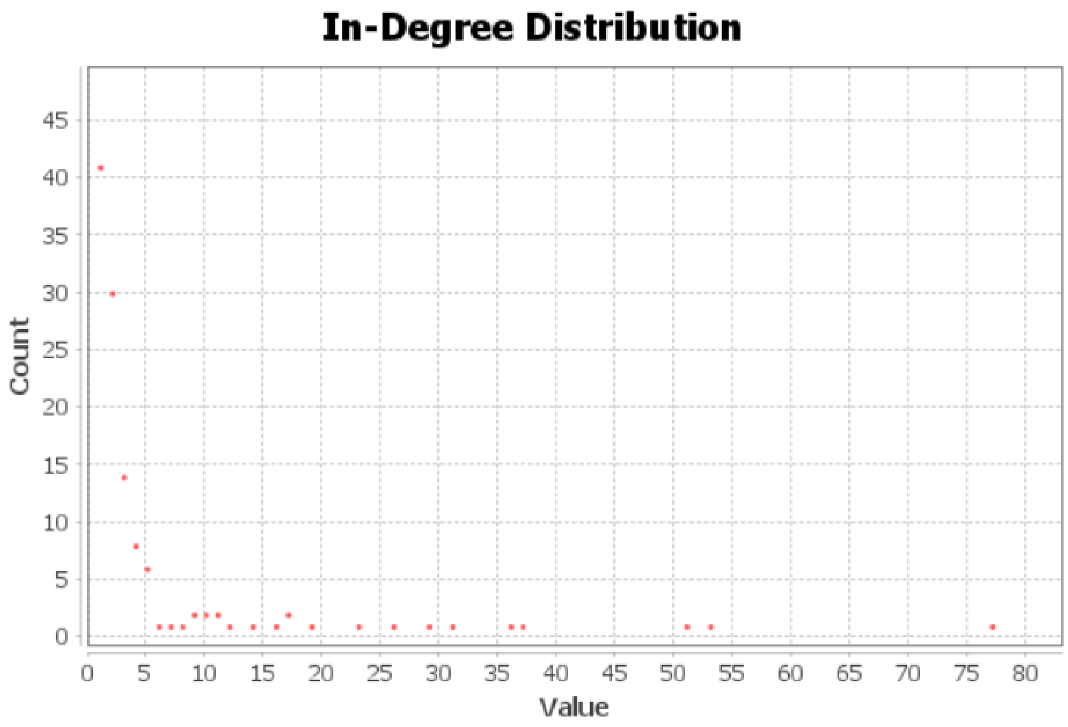


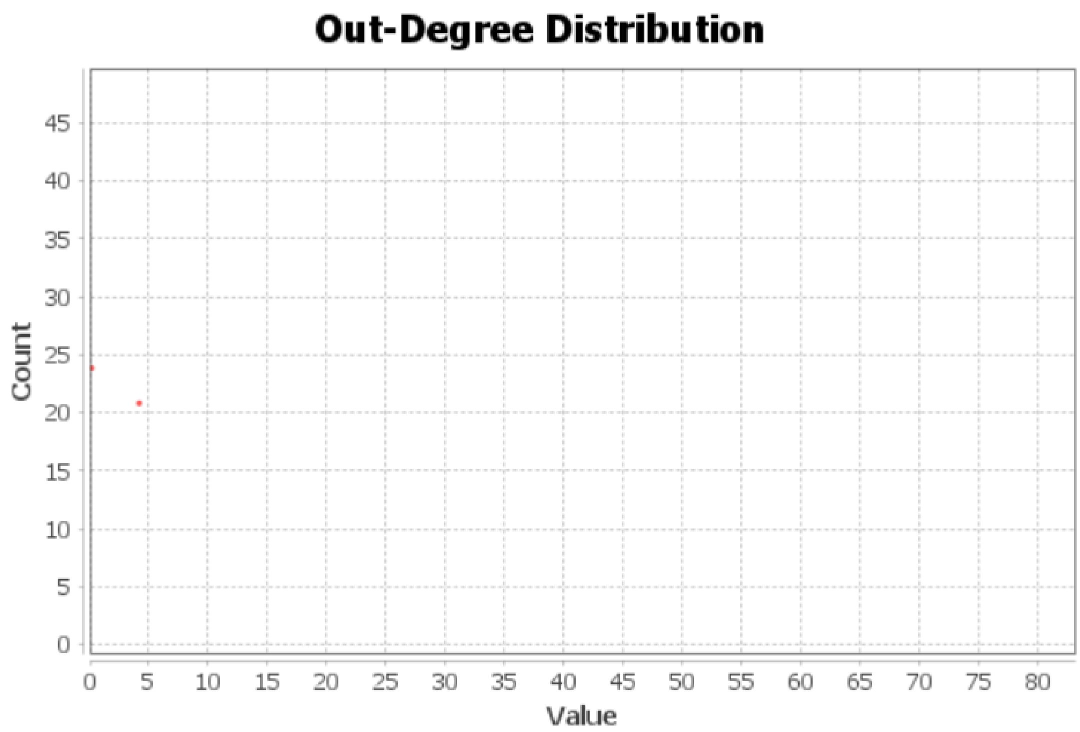
Maximum degree: USA manages to take first place holding the largest total degree of 82 (in-degree:77, out-degree:5) ! We will see later on how USA always manages to climb to the first place of any chart, such as this one.

Average degree: 4,203. This result should not come to us as a surprise, since, as mentioned earlier, the maximum out-degree of each node is 5 because our data is restricted to each country's top 5 importers.

Degree distribution: We should focus on the in-degree distribution because the out-degree distribution does not give us much information. As seen on the chart below, the biggest proportion of countries are in the 0-5 in-degree scale, meaning that there is a certain amount of countries which will become apparent later on in our analysis that actually control the export business.

Specifically, it's only 20 countries that have an in-degree distribution higher than 5, meaning it is only 20 countries that attract the products of all the countries worldwide. That shows us that the industrialization and the productivity, in general, is limited to a very small number of countries which seem to govern the world.



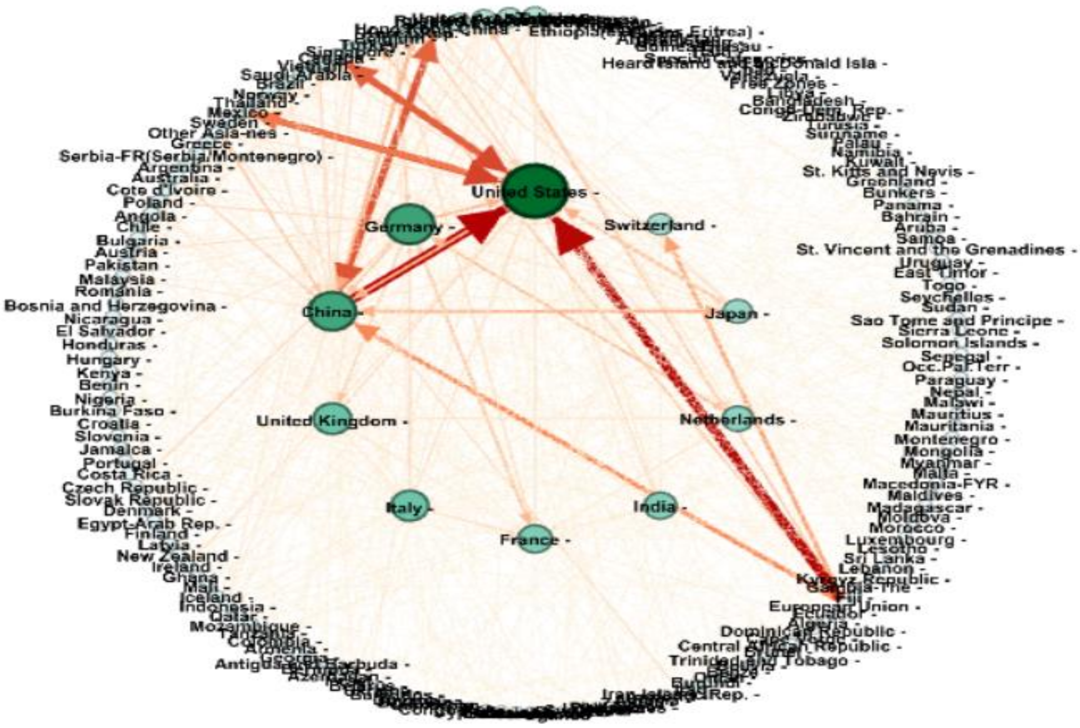


6.1 Degree centrality

Id	In-Degree
United States	77
Germany	53
China	51
United Kingdom	37
Italy	36
France	31
India	29
Netherlands	26
Japan	23
Switzerland	19

As previously mentioned, we focus on the in-degree metric rather than the degree metric because the out-degree does not give us enough information.

It is not surprising that the known countries which we consider powerhouses top the list of the in-degree matrix that follows. In other words, we see that the United States, Germany, China, United Kingdom appear to be importing from a large number of



countries having as a result the variety of products manufactured.

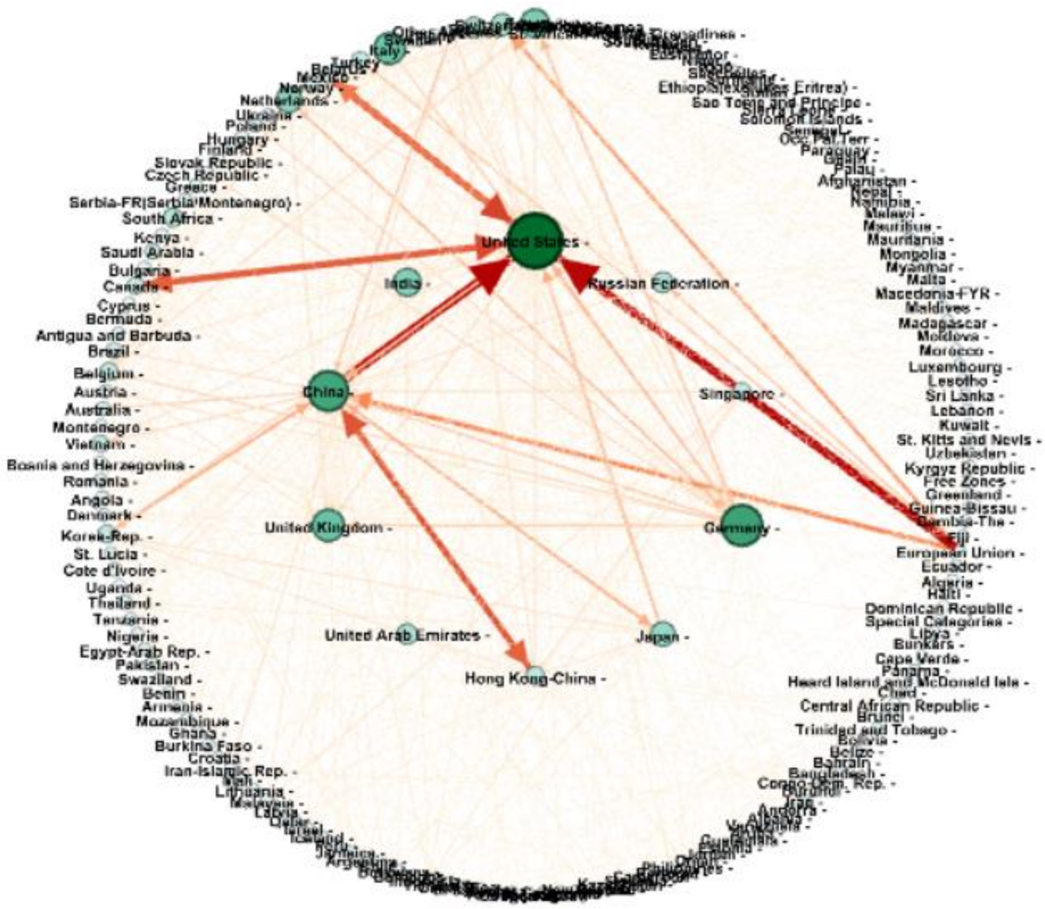
We also notice the fast decline in the degree value as we go further down the list. This gives us an idea of what is about to follow, meaning how a small subset of all countries actually controls and monitors a large part of the bigger picture by being the main importers and buying goods from many other countries.

6.2 Betweenness centrality

Id	Betweenness Centrality
United States	1065.024933
India	742.238043
China	638.473285
United Kingdom	525.593218
United Arab Emirates	484.70689
Hong Kong-China	477.63184
Japan	454.271859
Germany	405.669012
Singapore	321.442491
Russian Federation	286.748557

Betweenness centrality gives us an indication of how the economy is shaped through the circulation of products. Meaning, it gives us an idea of which countries are central in the exchange of goods for the advancement of each country's production and development. For instance, it is not surprising that the USA has the highest betweenness centrality, nevertheless we see that India comes second in place, meaning that its products exert a very significant influence in the economy of many countries.

Similarly, United Arab Emirates though being a main importer for a limited number of countries (17), its products appear to have a significant commercial interaction with many countries: hence the rank of 5 in the betweenness centrality matrix above.



On the other hand, it may be surprising that Germany though being 2nd in the in-degree list, meaning it imports from a vast number of countries, seems to be producing distinct products that appear themselves in less significant routes of product circulation: hence the reduced betweenness centrality.

It is noteworthy that the United States of America, being the largest degree-central node, does not export to any of the other major countries in the inner ring.

6.3 Closeness centrality

Closeness centrality indicates how close a node is to all other nodes in the network. It is calculated as the average of the shortest path length from the node to every other node in the network; therefore, lower values of closeness centrality indicate central nodes.

In our study, the lower the value of closeness centrality, the more central the location of the country is with respect to all the paths of interchanging goods in our graph.

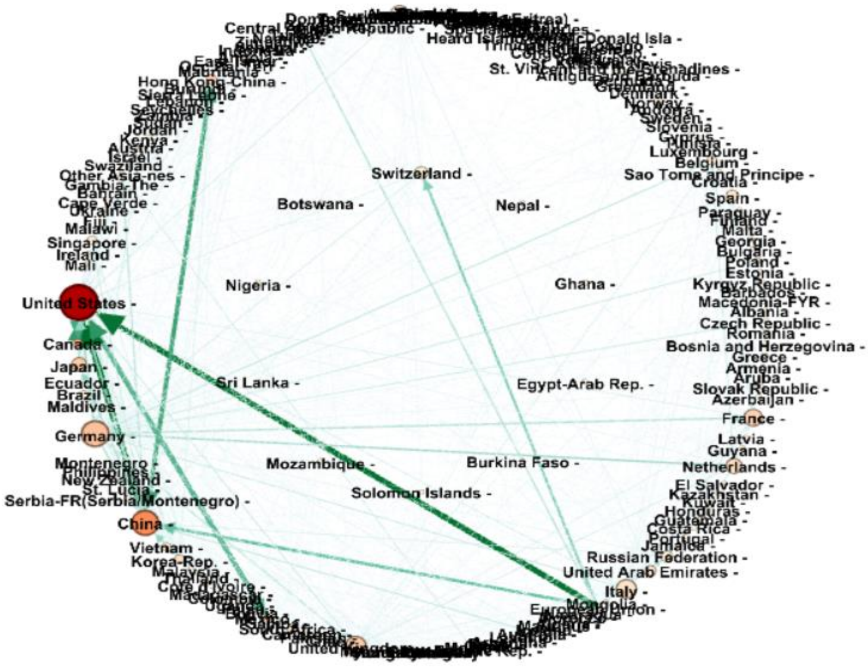
On the other hand, the larger the value of closeness centrality, the least central the location of the country is with respect to all the paths of interchanging goods.

In Table III below, we list the countries with the highest closeness centrality. Except from Switzerland, all the other countries included in this table are countries with very limited impact in the export business: Botswana, Nigeria, Sri Lanka, Mozambique, Solomon Islands etc.

Id	In-Degree	Closeness Centrality	Betweenness Centrality
Switzerland	19	0.465517	214.582526
Botswana	1	0.446154	9.666667
Nigeria	2	0.444444	20.235714
Sri Lanka	0	0.444444	0.0
Mozambique	2	0.439394	16.02381
Solomon Islands	0	0.439394	0.0
Burkina Faso	2	0.432836	15.494048
Egypt-Arab Rep.	2	0.432836	20.083333
Ghana	2	0.432836	15.869048

Table III: Countries with the largest closeness centrality

In the graph below, we illustrate the above grouping of countries with respect to the largest values of closeness centrality.



In Table IV below, we include the countries with the lowest (non-zero) closeness centrality. It is not surprising to see countries with very limited contribution to the export business: St. Kitts, St. Vincent, Antigua, Bermuda and Greenland. These countries have very limited interaction with the rest of the world, very short paths in their export business and very limited amount of nodes-countries with which they exchange goods. Hence, the small value for the closeness centrality.

St. Kitts and Nevis	0.213636
St. Vincent and the Grenadines	0.223629
Antigua and Barbuda	0.230366
Bermuda	0.245714
Greenland	0.247191

Table IV: Countries with the lowest closeness centrality

6.4 Harmonic Closeness

Id	Closeness Centrality	Harmonic Closeness Centrality
Switzerland	0.465517	0.537037
Botswana	0.446154	0.522989
Nigeria	0.444444	0.520833
India	0.428571	0.518519
Sri Lanka	0.444444	0.517857
Mozambique	0.439394	0.514368
Burkina Faso	0.432836	0.511494
Egypt-Arab Rep.	0.432836	0.511494
Solomon Islands	0.439394	0.511494

Table VI: Top 10 countries based on their harmonic closeness centrality

Harmonic centrality (also known as valued centrality) is a variant of closeness centrality, that was invented to solve the problem the original formula had when dealing with unconnected graphs. It is a useful measure that estimates how fast the flow of information **would** be through a given node to other nodes.

As a first remark, we see that the top-10 countries involving the closeness and the harmonic closeness centrality, remain more or less the same, with some slight changes in the order.

The values of the harmonic closeness centrality are within the range of the interval [0.32, 0.53]. From the Table VII below, where we have included the harmonic closeness centrality for the most influential countries, we notice that the value of the index for all these countries is slightly less than the upper bound of the range interval. That confirms the significance of these countries and the importance of their trade transactions in the world commerce.

Id	Closeness Centrality	Harmonic Closeness Centrality
United States	0.380282	0.488272
China	0.364865	0.466049
Canada	0.380282	0.488272
Mexico	0.36	0.478395
Germany	0.369863	0.487654
Japan	0.380282	0.478395
United Kingdom	0.355263	0.472222
Hong Kong-China	0.402985	0.493827
France	0.325301	0.454321

Table VII: Harmonic closeness centrality for the most influential

6.5 Eigenvector centrality

Eigenvector centrality is used to measure the level of influence of a node within a network. Each node within the network is given a score or value: the higher the score, the greater the level of influence within the network.

In our study, eigenvector centrality is a measure of the influence a country has on the international trade world. If a country is pointed to by many countries, that is she imports products from other countries (which countries also have high eigenvector centrality), then that country will have high eigenvector centrality.

In Table VIII below, we demonstrate the top 10 countries based on the eigenvector centrality.

Id	Eigenvector Centrality
United States	1.0
China	0.725775
United Kingdom	0.548011
Germany	0.539908
Japan	0.500109
France	0.423363
Hong Kong-China	0.336726
Netherlands	0.292862
Korea-Rep.	0.238817
Italy	0.21433

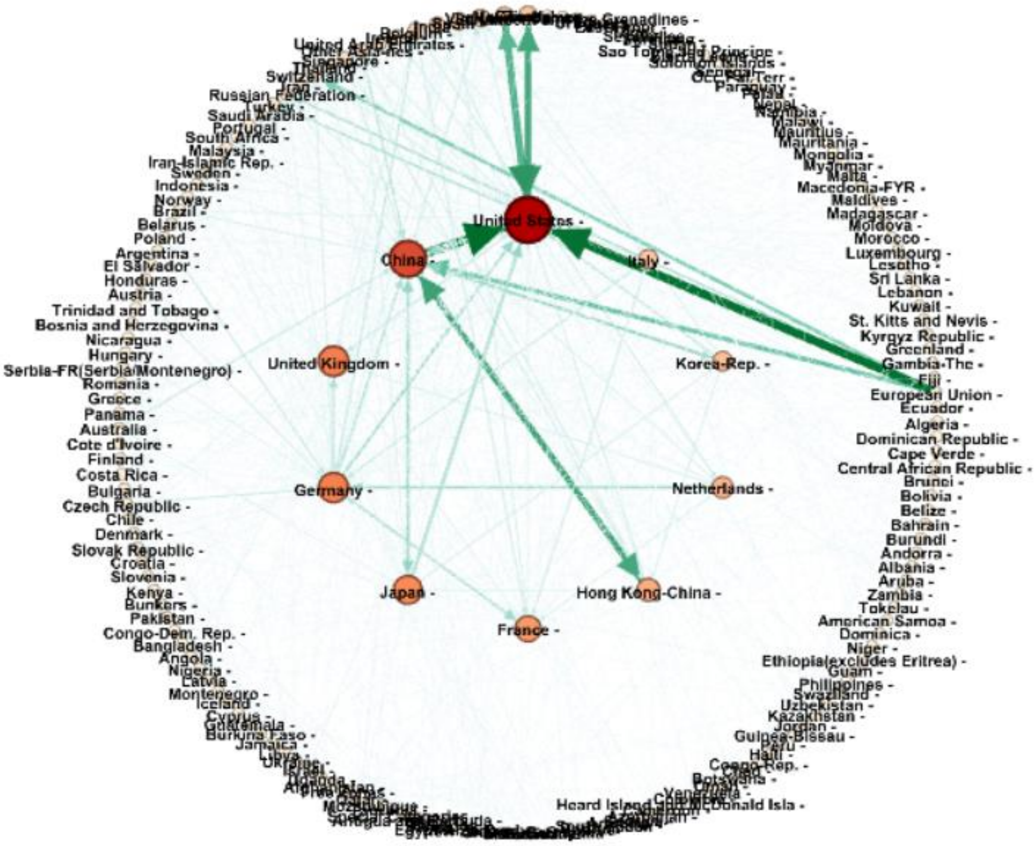
Table VIII: Countries with the highest eigenvector centrality

From the above Table, we note the following:

- As anticipated, countries as USA, China, United Kingdom and Germany, which are known international forces and have a lot of influence in our world, top the table.

- It is quite impressive that USA and China, being the 2 most powerful countries in the world, have such a big difference in their eigenvalue, indicating that the former was significantly more influential than the latter in the year 2018. One would follow this difference in order to see how smaller this gap comes by the year, knowing that through the last 10 years or so, China has been continuously strengthening her position in the world map.

- Looking at the first and the last country on the Table, we notice that the significant difference in the eigenvector centrality, which is equal to 0,79. That implies that the impact of any other country in the world is definitely smaller and less influential.



7.1 Clustering effects-Gephi's algorithm

Results:

Average Clustering Coefficient: 0,225
The Average Clustering Coefficient is the mean value of individual coefficients.

Id	Clustering Coefficient
Tunisia	0.833333
Luxembourg	0.7
Sao Tome and Principe	0.65
Philippines	0.6
Czech Republic	0.55
Guatemala	0.55
Madagascar	0.55
Slovenia	0.5
Bolivia	0.5
Portugal	0.5
Kuwait	0.5
Maldives	0.5

Table IX: Top-10 Countries based on clustering coefficient

A clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. In our study, countries-nodes tend to create tightly knit groups characterized by a relatively high density of ties; this likelihood tends to be greater than the average probability of a tie randomly established between two country-nodes.

We notice that the average clustering coefficient is equal to 0.225, meaning that our network has a low tendency to form clusters. Nevertheless, the countries that top the list with the highest clustering coefficient are countries with not so significant contribution to the world trade: Tunisia, Luxembourg, Philippines, Czech Republic, Guatemala etc. Therefore, for these countries, the limited trade relations to the other countries, makes it easier to form and be a part of a distinct cluster.

On the other hand, there are 25 countries that have a value of 0 in the clustering coefficient attribute. That has a big impact on the average clustering coefficient, resulting in a small value.

After further analysis in the data laboratory, we see that these 25 countries have an out degree of zero, meaning that they do not export goods to any countries. Since our graph is directed, such a result should not surprise us at all. Some of these countries are included in Table X below.

Id	Out-Degree	Clustering Coefficient
Venezuela	0	0.0
Congo-Dem. Rep.	0	0.0
Bangladesh	0	0.0
Chad	0	0.0
Heard Island and McDonald Isla	0	0.0
Panama	0	0.0
Bunkers	0	0.0
Libya	0	0.0

Table X: Countries with zero clustering coefficient

Finally, in order to get a feeling of the clustering capability for the most influential countries, we illustrate in Table XI below, the clustering coefficient for these countries. We readily see that the clustering coefficient for these countries is very low. That was anticipated, since the involvement of these countries with many other countries in the trade business makes it more difficult to include them in clusters.

For instance, USA has the very low value of 0.072, which denotes the difficulty to assign the country to a cluster; this was expected, if we take into account that the in-degree measure for USA is 77, as discussed in section 5 of our study.

Id	Eigenvector Centrality	Clustering Coefficient
United States	1.0	0.072112
China	0.725775	0.101176
United Kingdom	0.548011	0.129445
Germany	0.539908	0.109764
Japan	0.500109	0.221344
France	0.423363	0.179435
Hong Kong-China	0.336726	0.371795
Netherlands	0.292862	0.150794
Korea-Rep.	0.238817	0.384615
Italy	0.21433	0.110661

Table XI: Clustering coefficients for the countries with the highest eigenvector centrality

7.2 Clustering effects-The Clustering Coefficient Plugin with Triangle method

The local clustering coefficient of a country-node is the likelihood that the countries from which it imports goods also have trade relations. In this section, the computation of this index involves triangle counting. Triangle counting is used to detect communities and measure the cohesiveness of those communities.

In other words, the triangle method explores the possibility of future trade relations between countries which at present, have no such relation, but nevertheless have a common denominator, that is a country to which they both export their goods.

In Table XII below, we have included the countries with the highest clustering coefficient value based on the triangle method.

By comparing Tables XI and XII, we see that it is more or less the same countries included in both tables; that shows that both methodologies lead to the same practical results concerning the countries with the highest values in clustering coefficients.

We see that the number of triangles is 695. That shows that there is practically a dense web of interrelations between these countries and a significant potential for future trade business between them. It also demonstrates a huge potential for the circulation of goods and the development of new technologies, novel products and possible advances that may result in innovations and inventions.

Number of triangles: 695
Number of paths (Length 2): 12275
Value of Clustering Coefficient: 0.1698574274778366

Id	Clustering Coefficient
Tunisia	1.0
Croatia	0.9
Czech Republic	0.9
Luxembourg	0.9
Slovenia	0.8
Estonia	0.8
Guatemala	0.8
El Salvador	0.8
Madagascar	0.8
Sao Tome and Principe	0.8
Portugal	0.733333
Philippines	0.733333

Table XII: Countries with the highest clustering coefficient calculated by the triangle method

7.3 Clustering effects-Leiden Algorithm

The Leiden algorithm finds well-connected communities in large scale networks. The algorithm, with the above configurations gave us a weird result. Only two clusters were found in our data.

After further analysis, we notice that Palau and Guam belong to one cluster, whilst all the other countries belong to the other larger cluster. This happened because Palau exports to Guam but Guam does not export any goods to any country.

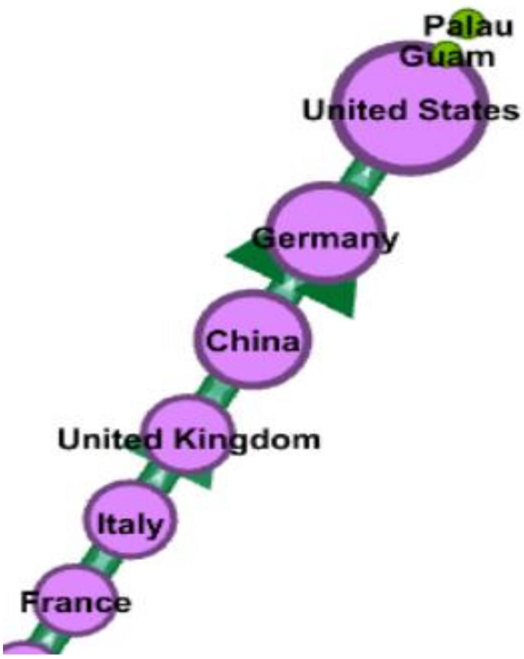
Algorithm	Leiden
Quality Function	Constant Potts Model (CPM)
Resolution	0.01
Number of iterations	100
Number of restarts	5
Random seed	0

Results

Quality	0.9999784763158178
Number of clusters	2

Id	Cluster
Palau	1
Guam	1

Using the Radial Axis Layout and coloring purple and green each cluster partition, we clearly see the 2 clusters. The rest of the image was cut off because the image was too large.



These results suggest that our nodes-countries are strongly connected.

However, by trying different configurations (varying the parameters of the program) within the Leiden Algorithm, we may obtain different results, but nevertheless the general picture is always the same. Specifically, no matter what configuration we try, we result in a big cluster containing the majority of the countries and a number of single-country clusters.

That shows, that practically through the years the countries have formed a complex web of intercountry relations, that is they have formed a big community involving practically all of them.

All the above configurations, in this section, were made using the CPM quality function. Please note that the other quality function, modularity (community structure) will be analyzed in the sections that follow.

8. Bridges

The Bridging Centrality is a node centrality index based on information flow and topological locality in networks. A bridging node is a node connecting densely connected components in a graph. It is the product of the betweenness centrality and the bridging coefficient, which measures the global and local features of a node respectively.

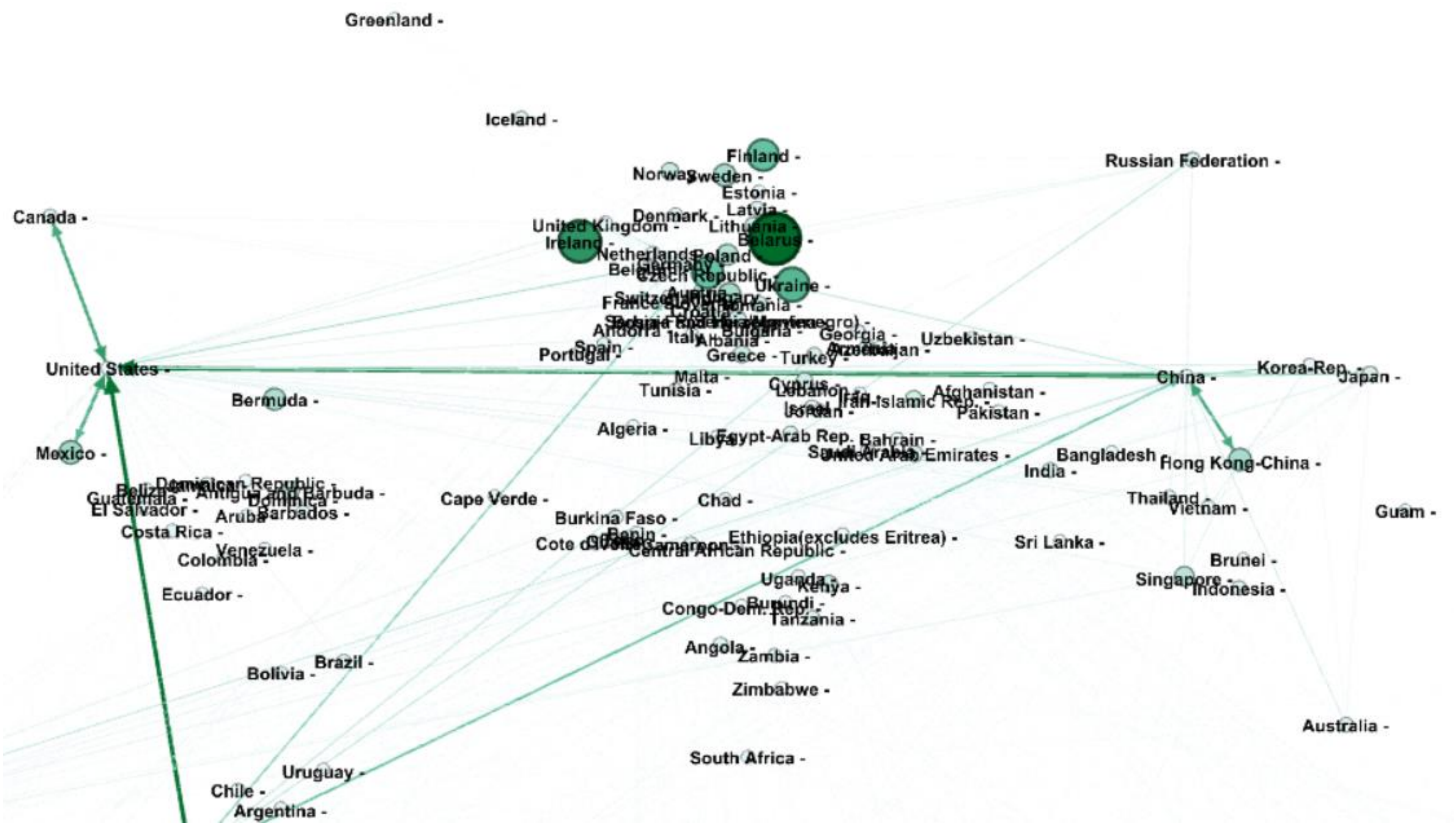
The only tool we have in our disposal concerning bridges, is Bridging centrality, which is measurable, but cannot be visualized. There is no plugin provided by Gephi that allows us to see global or local bridges.

With the help of Geo Layout which is referenced in chapter 13 we can easily visualize how bridges could be formed. For each continent there is at least one country with a relatively large bridging coefficient, for example: Europe (Ireland with a value of 0.0021), America (Mexico with a value of 0.0008), Asia (Hong Kong-China with a value of 0.0007), Africa (Kenya with a value of 0.0003). Therefore, these countries have a significant potential in future establishment of trade relations.

On the other hand, in Table XIII below, we consider the countries with the highest eigenvector centrality. We see that the bridging centrality for all these countries is relatively small, as anticipated. The reason being, these countries have exhausted any possible transactions with the rest of the countries and there is not much space left for advancing the intercountry trade relations.

Id	Eigenvector Centrality	Bridging Centrality
United States	1.0	0.000038871216727
China	0.725775	0.000054162025302
United Kingdom	0.548011	0.000078729517440
Germany	0.539908	0.000031481585828
Japan	0.500109	0.000186974285344
France	0.423363	0.000053079595729
Hong Kong-China	0.336726	0.000707018107744
Netherlands	0.292862	0.000036781738332
Korea-Rep.	0.238817	0.000035989083045
Italy	0.21433	0.000020353030284

Table XIII: Bridging centrality values for countries with highest eigenvector centrality



9.1 Modularity (community structure) / Girvan-Newman

The Girvan–Newman algorithm detects communities by progressively removing edges from the original network. The connected components of the remaining network are the communities. Instead of trying to construct a measure that tells us which edges are the most central between communities, the Girvan–Newman algorithm focuses on edges that are most likely "between" communities.

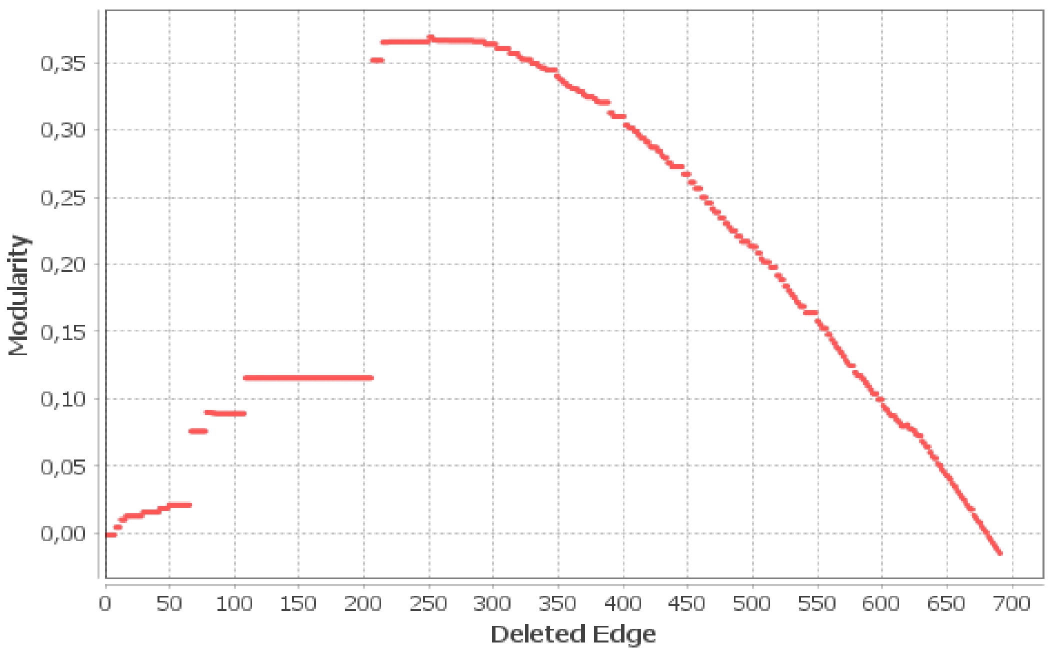
This alternative way of producing and searching for communities in our graph as we will see on the next section accumulates different results from those of Gephi’s algorithm.

The Girvan-Newman algorithm found 21 communities. In the graph below, we have an illustration of modularity’s variation with respect to the number of deleted edges. We clearly see that the maximum modularity is achieved when the number of deleted edges is around 250.

Communities

Number of communities: 21

Maximum found modularity: 0.37042916



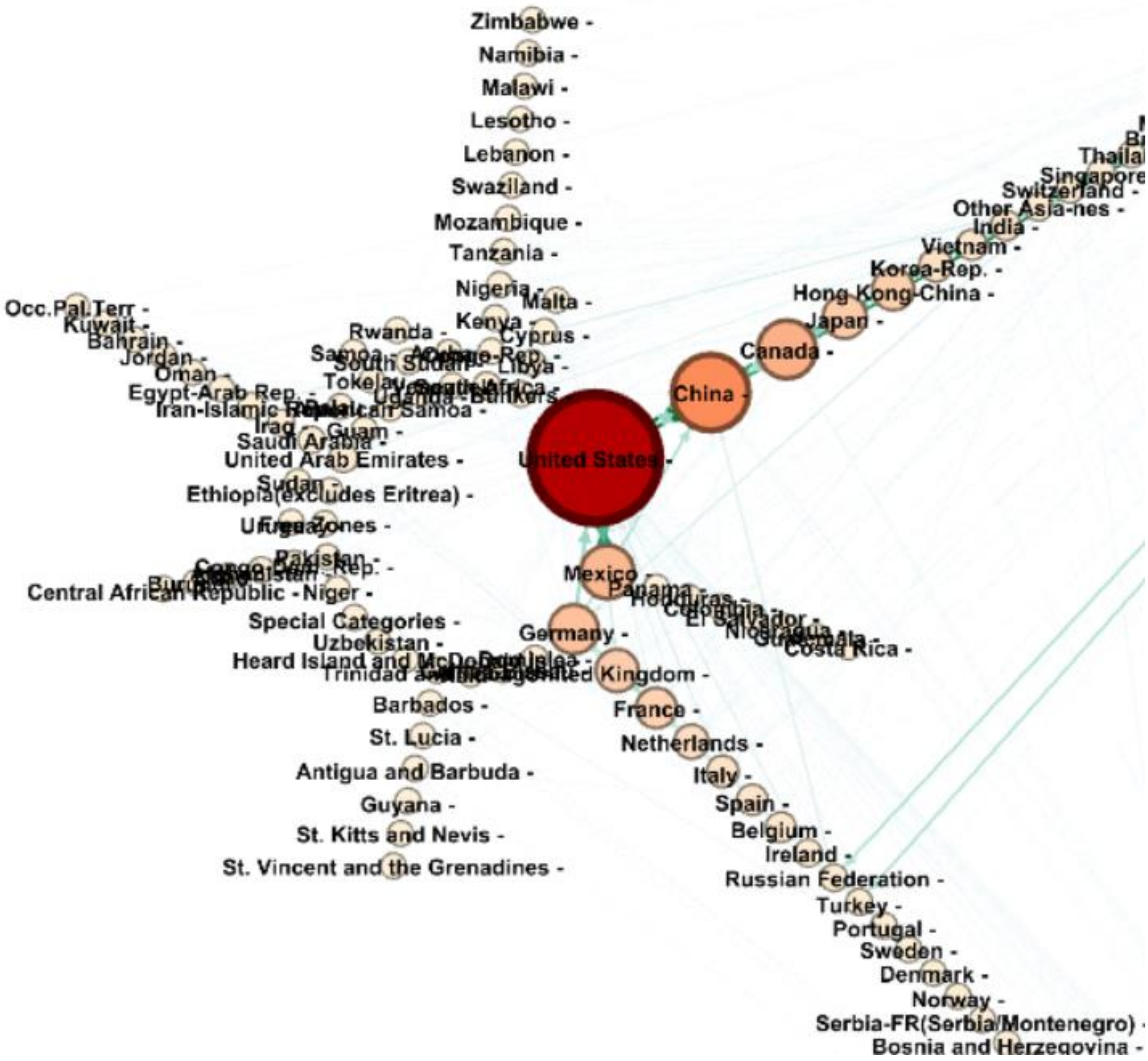
That explains the final result of 21 communities, which is achieved by this choice of number for the deleted edges. In other words, the topology of export business between our countries contains certain dynamics which boil down to no more than 20 communities around the world.

Looking up the 21 communities derived by the algorithm, we note the following:

- The communities with the larger number of member countries seem to have a geographical location: either they belong to the same continent or have proximity as countries.
- The community with the largest number of member states is community 18, which includes practically all the European countries. The community consists of 60 countries, out of which 49 are European countries and only 11 are non-European countries (see Table XIV below).
- In Table XIV below, we include a part of all these 60 countries but nevertheless all the 11 non-European countries which are highlighted in blue.
- These 11 non-European countries are: Greenland, Sri Lanka, Madagascar, Maldives, Mongolia etc. As we see, the 11 non-European countries are countries with extremely limited activities in the export business and therefore it is no surprise that they were selected by the algorithm in the strongest community.
- In a similar fashion, community #4 comprises 10 countries of the Arabic Peninsula (UAE, Iran, Iraq, Saudi Arabia, Bahrain etc.), community #17 comprises 8 countries of Latin America (Mexico, Colombia, Panama, Costa Rica, Nicaragua etc.) and community #20 comprises 12 countries of Africa (Kenya, Congo, Nigeria etc.).
- The only community that does not fit in the geographical analysis above is community #18 which includes USA, China, Canada, India, Brazil, Japan, Israel etc.). This community seems to have attracted the major players in the export business; hence the criterion here was the volume of business and not the geographical location of the countries.

Id	Cluter-ID	Eigenvector Centrality
Denmark	16	0.008163
Sweden	16	0.019661
Algeria	16	0.0
Estonia	16	0.003121
Finland	16	0.009032
Latvia	16	0.004967
Ireland	16	0.089779
Greenland	16	0.0
Iceland	16	0.004674
Hungary	16	0.010655
Kyrgyz Republic	16	0.0
Sri Lanka	16	0.0
Lithuania	16	0.003121
Luxembourg	16	0.0
Morocco	16	0.0
Moldova	16	0.0
Madagascar	16	0.0
Maldives	16	0.0
Macedonia-FYR	16	0.0
Mongolia	16	0.0
Montenegro	16	0.004682
Mauritius	16	0.0
Nepal	16	0.0
Sao Tome and Principe	16	0.0
Seychelles	16	0.0
Tunisia	16	0.0

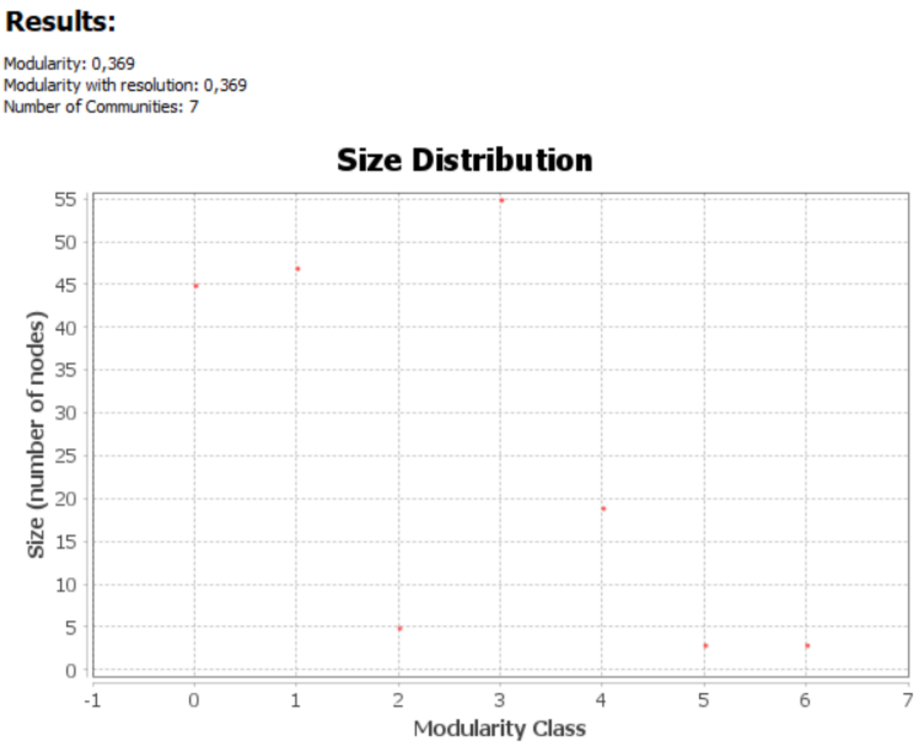
Table XIV: Community #16 comprising 60 countries mostly European. The only non-European countries are highlighted in blue.



Graph I: Illustration of 21 communities as a result of the Girvan-Newman Algorithm

9.2 Modularity (community structure) / Gephi

Gephi gives us the option to use its own modularity algorithm which is built to show us how easily communities are created within our graph.



The Gephi algorithm resulted in only 7 communities out of which 3 have a very limited number of member-countries (<5). There are 3 significant communities including 45-55 countries:

- Community #0 includes USA, Russia, Canada and a number European, African, American countries.
- Community #1 includes China, India, Australia, Japan and other countries from all continents.
- Community #3 includes most European countries and many African countries.

- In Table XV below we have included a small part of the countries comprising the large communities to show the diversity in the country inclusion.

Samoa	6	United Arab Emirates	4	Netherlands	3
American Samoa	6	Iraq	4	Albania	3
Tokelau	6	Saudi Arabia	4	Germany	3
Congo-Rep.	5	Burundi	4	Greece	3
Heard Island and McDonald Isla	5	Pakistan	4	Italy	3
Lebanon	5	Bahrain	4	Serbia-FR(Serbia/Montenegro)	3

Gambia-The	2	Vietnam	1	Aruba	0
Guinea-Bissau	2	Australia	1	Colombia	0
Mali	2	Japan	1	United States	0
Senegal	2	Korea-Rep.	1	Venezuela	0
Angola	1	Benin	1	Canada	0
China	1	Malaysia	1		

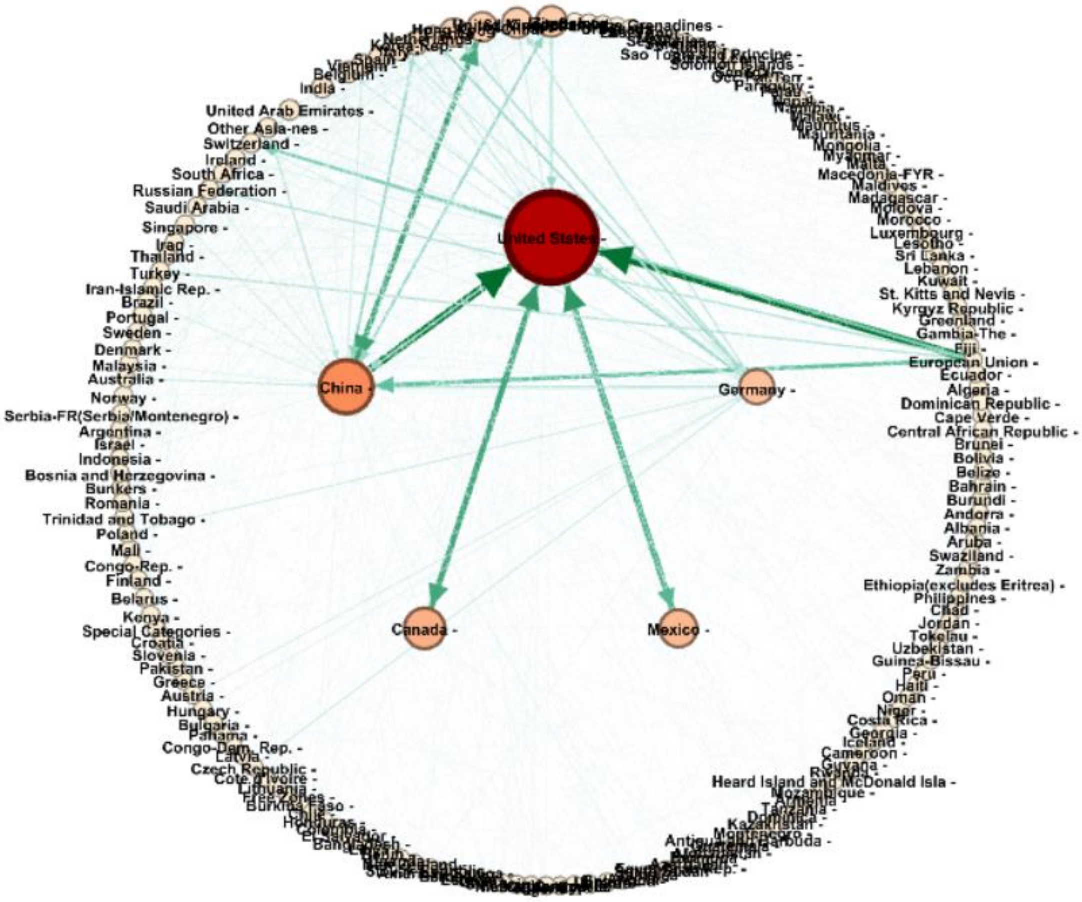
Table XV: Countries representing the 7 communities derived by the Gephi Algorithm.

Comparing the 2 methods: we clearly see that the Girvan-Newman algorithm is more suitable for our study.

The Gephi algorithm did not manage to categorize the dynamics of the commercial work. On the other hand, the Girvan-Newman algorithm managed to:

- Identify the dynamics of the power countries and group them in a strong community.
- Identify the geographical characteristics of each country and group them based on their location.

10. PageRank



The PageRank concept is a way in which a web page or social network node can be given an “importance score”. The PageRank is a variant of the Eigenvector centrality score, but because it uses backlinks/in-degrees it is used in directed networks.

We return to our favourite layout, the dual circle layout and apply a color and size ranking on our countries to make the top 5 stand out. We also see how much higher the USA’s PageRank value is compared to China (which is second) and even more higher than the other 3 countries: Canada, Mexico, Germany.

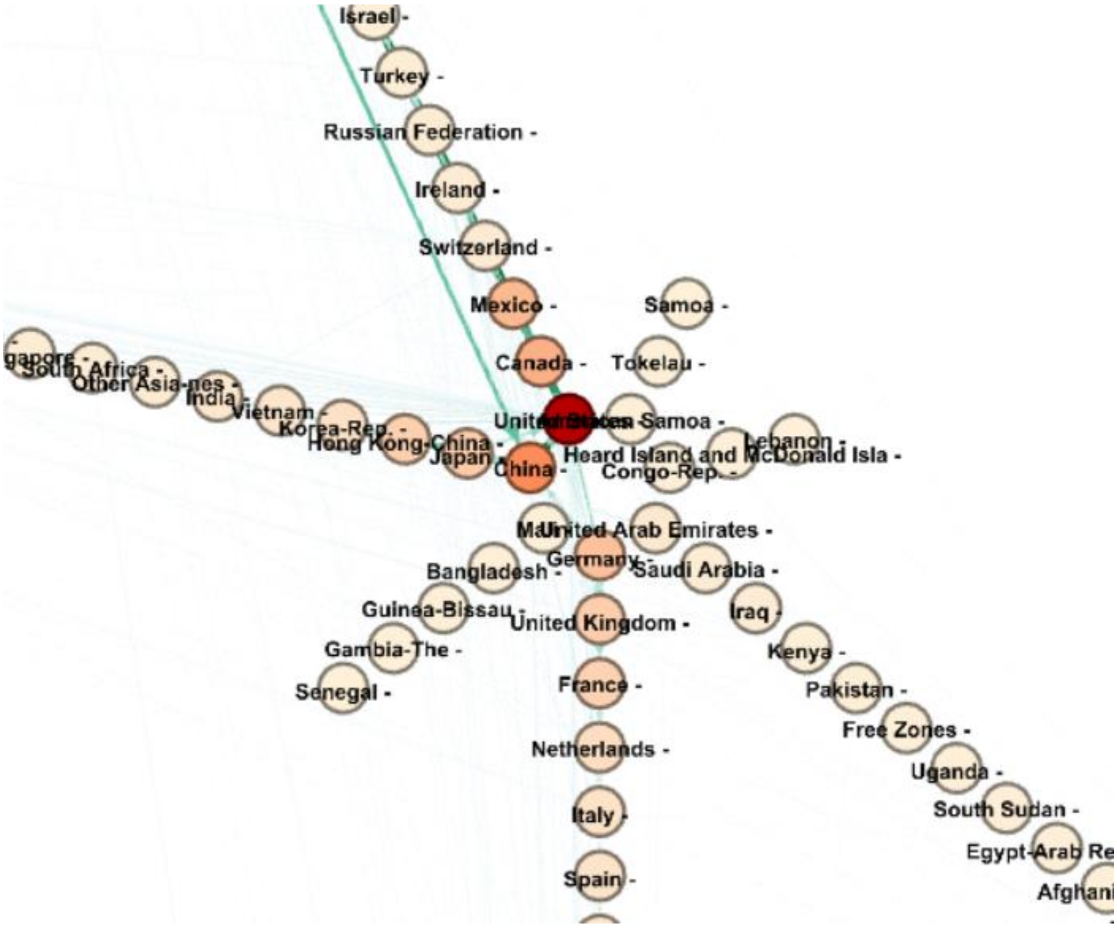
11. Homophily

To quote Barabassi: **homophily, a well documented social phenomena , indicates that individuals tend to associate with other individuals of similar background and characteristics, hence individuals with comparable degree tend to know each other.**

By transferring the way Barabassi defines the graph's homophily on to our dataset, we will begin to visualize the homophily and how countries of similar characteristics (ie sharing geographical borders) begin to form a cluster.

This can be seen by using the Radial Axis Layout and ordering them by their PageRank (importance).

Below is a snapshot of the communities that have been formed in our graph.



12. Graph density

Our graph density is not expected to be a big number since each country has a maximun out degree of 5, restricting it to connect to many countries.

Graph Density Report

Parameters:

Network Interpretation: directed

Results:

Density: 0,024

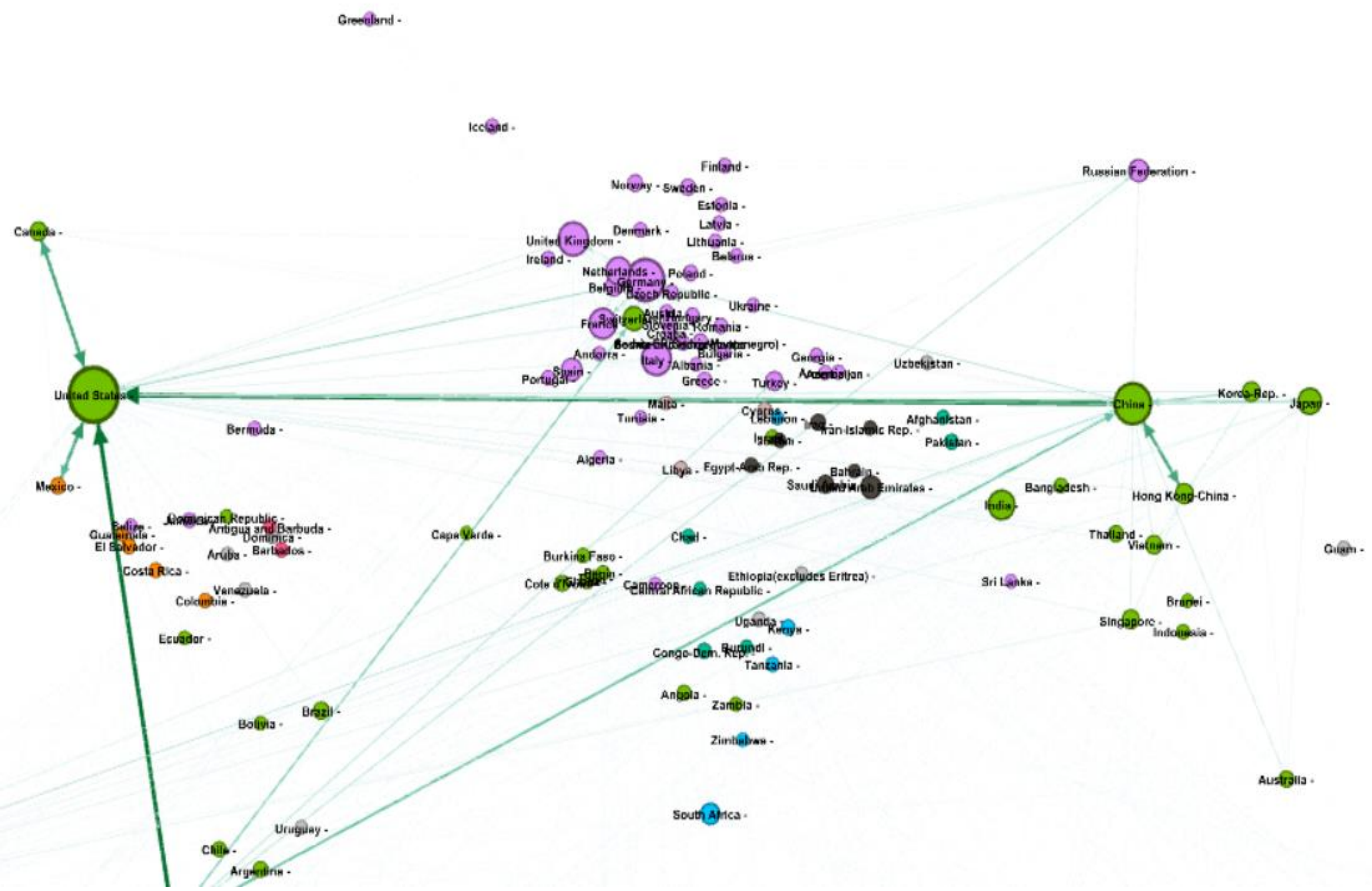
By observing the number of nodes (177) and edges (744) it is quite clear that our graph will not be that dense because 744/177 is about 4.2, which means that on average there are 4.2 edges that connect each country. In that respect, the obtained density value of 0.024 is something that we would anticipate to see in our study.

13. Extras

After looking through the available plugins that Gephi has, I found one named Geo Layout, which, given pairs of latitude-longitude coordinates is capable of mapping the countries in our graph and allow us to visualize the globe as it really is.

After extracting data from https://www.kaggle.com/parulpandey/world-coordinates and merging it with our current dataset with the help of python's library Pandas, we are able to use the layout.

In the picture below, the countries have been resized based on their PageRank (importance) and coloured based on their cluster ID, proving that Gephi's algorithm did a decent job creating communities, that are in fact, the continents.



14. Concluding remarks

In our study, we analyzed the global export trade values. We used the data extracted from the official webpage of World Integrated Trade Solution.

We analyzed the data via the Gephi software. We explored all the capabilities of the Gephi software:

- Layouts
- Plugins
- Algorithms
- Data Laboratory

Our analysis confirmed the following:

- USA, China, are indeed the major powerhouses in the export world. These countries influence and regulate the economies of the remaining countries.
- The other countries follow in some significant distance. UK, Germany, Japan, Canada seem to form the second group with international influence.
- The geographical location appears to be a significant factor for each country in selecting the countries to which this country exports its goods. In other words, distance and geographical proximity constitute significant factors in shaping the map of world's export business.

References

https://faculty.nps.edu/rgera/MA4404/Winter2018/16-Homophily.pdf

http://networksciencebook.com/chapter/7

https://github.com/jaroslav-kuchar/GirmanNewmanClustering

https://www.strategic-planet.com/2019/12/how-a-page-rank-is-calculated-in-gephi/

https://en.wikipedia.org/wiki/Centrality#Closeness_centrality

https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/

https://www.sciencedirect.com/topics/computer-science/closeness-centrality