6-2022

Analysis of Global Banking Network

Sonali Kannan
sk8359@rit.edu

ANALYSIS OF GLOBAL BANKING NETWORK

by

Sonali Kannan

A Capstone Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies:

Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology, Dubai

JUNE 2022

RIT

Master of Science in Professional Studies:

Data Analytics

Graduate Capstone Approval

Student Name**: Sonali Kannan**

Graduate Capstone Title**: Analysis of Global Banking Network**

Graduate Capstone Committee:

Name:	**Dr. Sanjay Modak** **Chair of committee**	**Date: 16-12-2021**
Name:	**Dr. Giulia De Masi** **Member of committee**	**Date: 16-12-2021**

1 Acknowledgments

I would like to extend my sincere gratitude to our president, Dr. Yousef Al Assaf for providing us with this opportunity to develop and complete our project in our field of study.

I wish to extend sincere thanks to Dr. Sanjay Modak , Assistant Professor and Chair, Graduate Programs and Research Department for his moral support and valuable advice.

This capstone could not have been possible without my project mentor, Dr. Giulia De Masi for having given me an opportunity to pursue my academic interests and for her constant support throughout the project. I am grateful to her inspiring guidance which motivated me, right from the inception of the project.

I would also like to thank my supporting mentors Dr. Gabriele Tedeschi and Dr. Giacomo Livan for their extended insights and contribution to this capstone.

2 Abstract

The outbreak of the Global Pandemic Covid-19 that spread terribly across various countries from the end of 2019, has severely altered people's life and economy. Various reports across papers and news articles on how each government was managing the costs of vaccines, medical equipment, and necessities. The world saw shifts in stock markets, unemployment, the tourism industry completely coming to a standstill, and more.

Has this Covid Pandemic which played a crucial role within geographical boundaries altered the financial transactions across countries on a higher level?

With the help of the statistics available with the Bank of International Settlements, this project aims to analyze the cross-border lending pattern across countries. This can be analyzed with the help of Complex Network analysis. The network reflects the data where the nodes are the countries and bilateral links correspond to credit linkages. Using various topological network measures such as Degree, Strength, Clustering coefficient, and Polya Filter, we can analyze the financial interconnectedness and the possibility of change in network patterns during times of crisis such as Covid-19. This will help to find a correlation between this sudden worldwide crisis and the lending market among banks.

Keywords: Bank of International settlements, Network Analysis, Polya Filter

Table of Contents

3 List of Figures

4 List of Tables

Chapter 1 – Introduction

1.1. Background

1.1.1. Bank of International Settlements

The BIS Group is owned by 63 central banks and monetary authorities. It is an international organization that works towards building a greater understanding of the world economy by promoting international cooperation among various authorities, conducting economic research and analysis on policy issues, and providing banking services to central banks communities. "BIS statistics, compiled in cooperation with central banks and other national authorities, are designed to inform analysis of financial stability, international monetary spillovers, and global liquidity." This organization is publishing data on lending among banks quarterly. It is this BIS Dataset that is going to be used to analyze the lending and borrowing pattern between central banks across countries.

1.1.2. Graph Networks

1.1.2.1. Konigsberg Problem

The Konigsberg Bridge Problem is an ancient puzzle that led to the discovery of a new branch of mathematics, Graph Theory. Back in the 18^{th} Century, Konigsberg had 7 bridges across which their citizens used to walk. There was a question on whether there is a way to cross these bridges exactly once, whilst walking through the town. This led to the discovery of Topology and Graphs. Euler, a mathematician proved that it is not possible to traverse such a path by representing the bridges in the form of a network.

1.1.2.2. Network Analysis in Finance

A Network is a graph consisting of a set of nodes and links. Network theory is a way of representing and analyzing a problem in the form of graphs. Each system of interacting agents can be represented as a network, where each agent is represented by a node, and the interaction between each possible pair is represented by a link. In the case of finance, there are many applications: credit networks where the nodes are banks and firms and links are debit/credit relationship [10,12], an interbank market where nodes are banks and links are their mutual lending [11,13], the stock market where the nodes are single stocks and links are the correlation of their price [14], foreign direct investments where nodes are countries of 4 investors and investment. In this project, nodes represent countries and links represent cross-border lending between banks of one country to the banks of the other country. Network metrics are mathematical measures that help to identify the network topology, in particular the structure of the interconnections. Such a structure plays a crucial role in the study of dynamic phenomena, those arising in a crisis event. This instrument helps in developing specific policies to prevent bad consequences during exogenous crisis events.

1.2. Statement of problem

The main objective of this project is to study whether the Global Pandemic Covid-19 has affected the lending and borrowing patterns of banks across countries.

1.3. Project goals

This project focuses on analyzing the patterns between banks during the last years including the Covid crisis. It includes constructing a network with the BIS data that includes countries across the globe. With the help of different network indicator measures such as (but not limited to) Degree, Strength, Clustering coefficient, and Polya Filter, we identify the interconnections across networks and how the patterns change during the financial downfall periods.

1.4. Methodology

The methodology includes the various phases to be carried out in this capstone project to arrive at the results. Following the crisp-DM outline model,

Phase 1:

Business Objective: This study provides an instrument to identify the changes in trends and patterns in the global banking network through various financial periods to see how a crisis can affect the interconnectedness of the network, thereby coming up with potential reasons for the same. This tool can be relevant for policymakers to design the best action to prevent bad consequences during the crisis.

Phase 2:

Data Understanding: This phase will include an initial collection of data, description of data, exploration of data, and verification. After performing considerable literature reviews, we have concluded that the BIS Dataset is the most appropriate data to be dealt with for this problem statement. This dataset will have the lending, borrowing, and other financial parameters across banks from various countries. Data are aggregated by country and released on a quarterly basis. Some exploratory statistics will provide a first view of the statistical properties of the dataset.

Phase 3:

Data Preparation: Data available after the collection phase will not be ready to be fed into a machine-learning algorithm and thus needs to undergo the following steps:

1. Clean the data – Remove null values, detect outliers, and treat them if necessary.
2. Transform the data – Changing the data type and normalizing the numerical attributes.
3. Integrate data – join different tables if there are more than one dataset to study on

Phase 4:

Modelling Network theory is going to be used for the analysis of global banking. After some exploratory statistics, the complex network representation will be built, starting from BIS data. Nodes are the countries present in the dataset and Links are the aggregated banks' cross-border bilateral lending. Statistical measures of the network topology will be used such as (but not limited to) Degree, Strength, Clustering coefficient, and Polya Filter. Most relevantly we use the Polya urn approach to filter the initial network to the most relevant links. The basic analysis will be done using Python packages as Networkx, visualization tools as Gephi, and in-house developed scripts.

Phase 5:

Evaluation The significance of the topological measures will be evaluated by comparing the results from real data with the same measures from a randomized network, where the nodes are the same as the initial one and the links are shuffled among countries.

1.5. Limitations of the Study

The initial major challenge is the huge dataset with a large volume of transactions. Hence, it must be cleaned and filtered properly for exploring and analyzing the data. The BIS Website updates the current year data after a long time and thus, we were unable to include the current year,2021. Also, they only update quarter by quarter, so we need to wait until they update all four quarters of a year to get the complete year's transaction data.

Chapter 2 – Literature Review

The application of complex networks theory to BIS data has already proven useful in a few state-of-art works. Clemente, G. P., et al. (2020) in their paper[1] focused on performing network analysis of systemic risks across the European Banks throughout 2003-2017, a period that covers sub-prime crisis and subsequent sovereign debt crisis. The network constructed is a directed weighted graph where nodes represent various bank entities and links are weighted quantified with market-based measures. The study used two network indicators namely ‘clustering’ and ‘Hub’ and authority centrality. Clustering indicates the tendency to form intricately connected groups within a network. It is observed that the clustering coefficient is related to the financial condition of the market, it is lower during calm periods and higher during peaks. Hub and Authority centrality address two issues regarding the banks in the sample. Firstly, it provides a dual centrality score that is based on each bank's role within the network, where its collector and/or originator roles are measured. Secondly, as it captures the authority centrality and hub centrality of the banks that are connected to each bank, it provides a dual centrality weighted measure of their importance within the network. There is a high correlation between the authorities and hubs that concludes banks that are important for spreading risk are also affected by other banks, namely, risk drivers are also risk-takers. The authors see this method as a complement to the traditional balance sheets and regulatory data.

Bongini, P. et al. (2018) in their paper[2] dealt with examining the level of interconnectedness within a network of global banks during the financial crisis to tackle the issue of systemic risk. With the BIS consolidated data, the authors built a network wherein nodes represent the countries and weighted directed edges represent cross-border exposures. With the help of various topological indicators such as degree, strength, clustering coefficient, and assortivity, the network is analyzed to see that globally systemic important banks(G-SIBs) and core countries exhibit different behaviors. Since 2013, it is evident that the G-SIBs have had a decline in lending/ borrowing practices and have moved their focus towards international banking activities. This shift is seen because of the systemic regulations introduced in 2011 by the Basel Committee on Banking Supervision (BCBS).

Cerqueti, R. et al. (2020) in their paper[3] mainly focused on how the clustering coefficient is used to assess systemic risk in financial networks. Systemic risk is a possibility of a negative trend or a pattern at a local level playing a huge impact on a global level. The more interconnectedness in the network of nodes constructed, the more probability of a systemic risk that may be caused due to diffusion of nodes at a local level. With the help of a weighted average of high-order clustering coefficients, the degree of flexibility in a network and state of stress can be assessed.

Vidal-Tomás, D. et al. (2020) in their paper[4] focused on identifying the credit relationships between banks during various crisis periods in the Euro countries. The study is conducted on the BIS Dataset wherein the data is divided into two categories, the rich countries and the poor countries known as PIGS countries. It is found that the PIGS Countries help each other without the help of NON-PIGS during economic turbulence. With the help of a measure known as Lender's Preference Method, defined as "the ratio of total claims that country "I" has lent to country "J" during a given quarter, over the total amount of claims that "I" has lent in the interbank market (J) during the same quarter.", it is found that the PIGS countries invest less in the NON-PIGS countries during the debt sovereign crisis.

Minoiu, C. et al. (2013) in their paper[5] focused on cross-border lending and borrowing across 184 countries during the time of 1978-2009. With the help of network analysis, each of the 184 countries is a node within a network constructed and directed links represent the financial flow. Various network metrics have been used such as centrality and network density to study the interconnectedness in the global banking network. The results suggest that the network has been relatively unstable, especially during the financial crisis,2008- 2009. Also, from 2002 to 2008, the high-growth European countries mark as highly interconnected borrowers along with expansions and contractions in the network.

Serena, J. M. et al. (2021) in their paper[6] summarized results obtained from the survey conducted by the Irving Fisher Committee in 2020 to understand the use of big data sources and applications at central banks. Big Data for central banks is known to have covered all the datasets that require non-standard technologies to be analyzed. The most important data that we can deal with for network analysis of global banking is the structured cross-relational data that collects multiple attributes at a single time. Central Banks, despite various challenges, are ready to reap the benefits of big data for three reasons mainly, knowledge sharing, use of big data on global issues, and developing joint exploratory projects. International financial institutions support to fulfill these cooperative methods with the help of BIS Innovation Hub to develop and identify insights into trends in fintech relevant to central banks.

Bank of international settlements provides local and consolidated banking statistics each having its aim of data collection. McGuire, P et. Al in their paper[7] threw light on how consolidated banking statistics were helpful to learn the foreign assets. Consolidated Banking Statistics is collected across the global portion concerning the location of the reporting Banks. This set of data also shows the debt between the creditor who is on the financial asset side and the borrower who is on the liability.

A large network with dense nodes and links is of high complexity which is a hindrance to visualizations and analysis done on the network. Radicchi, F. et. Al in their paper[8] explained the need for filtering and proposed a filtering technique, Global Null Model. It is important to preserve the network vertices as they are not independent of the structure of the network, meanwhile perform reduction based on edges. If the edges are filtered based on a threshold value, then the network will have only largely weighted edges disrupting the network properties.

In this project, we want to apply a new technique that has never previously been used for BIS data analysis: the Polya urn filter. Marcaccioli, R. et al. (2019) in their paper[9] proposed a completely new technique to filter the important links of a network based on the Polya urn probability problem. The Polya filter is filtering methodology based on a null hypothesis designed to respond to the specific heterogeneity of a network done with the help of a statistical test based on the Pólya urn, a well-known self-reinforcement mechanism according to which the observation of a certain event increases the probability of further observing it.

Chapter 3- Project Description

3.1. Overview of the project

This project involves steps from collecting the data, exploring, and understanding the data. The data must be preprocessed and required attributes for our study needs to be chosen. After which we will proceed with network analysis to get further conclusions on the Global Banking network

3.2. Data Collection

The data used in this project for analysis is taken from the BIS Statistics in bis.org Website that includes the consolidated banking statistics Data that can be downloaded in the form of a CSV file. BIS Statistics Warehouse tab in the BIS website consists of an interactive query option where users can download the data based on the filters we apply.

Why consolidated Banking Statistics?

- The consolidated banking statistics measures international banking activity from a nationality perspective, focusing on the country where the banking group's parent is headquartered.
- While residence-based data such as the locational banking statistics indicate where positions are booked, they do not always identify where underlying decisions are made. This is because banking offices in one country may operate within a business model decided by the group's controlling parent, which may be headquartered in another country.
- The CBS capture the worldwide claims of banking groups based in reporting countries and exclude intragroup positions, like the consolidation approach followed by banking supervisors.

3.3. Data Understanding

The dataset contains records observed from 1983 up to 2020. Until 2000, amounts were recorded twice a year. From 2000 to 2020, data is recorded every three months in a year as four quarters.

Major Attributes of the Data Set:

- Frequency – Data recorded at 4 quarters of a year.
- Measure – Amount outstanding which is the value of an asset or liability at a point in time.
- Reporting Country – Nationality of the banks that lend.
- Counterparty Country – Nationality of the banks that borrow.
- CBS Bank Type – Type of Banks based on whether they fall under the Domestic Group or Foreign Group.
- CBS Reporting Basis – On what basis are the country's claims are allocated.
- Balance sheet position – It states whether the claims are international, local, or total claims.

Chapter 4- Project Analysis

4.1. Data Preparation

Data in the form as acquired cannot be used for our study and analysis and thus needs to undergo a preparation and cleaning phase**.**

Filtered rows that include the following attribute values:

- Credit Transactions take place between a reporting country and a counterparty country thereby removing rows with Reporting or Counterparty Country stated as 'Residents', 'Unlocated', 'All reporting Countries' and 'All Countries'.
- CBS Bank type is chosen as Domestic Banks focusing only on banks whose controlling parent is in the respective BIS reporting country.
- CBS Reporting basis considered as Immediate Counterparty Basis where claims are allocated to the country and sector of entity to which has borrowed the funds.
- Balance sheet position is chosen as Total claims that are international claims comprising of cross-border claims and local claims of the domestic banks within the reporting countries.

Attributes required for our study:

- Reporting Country, Counterparty Country, and the years of focus (2017, 2018, and 2020)

Removed Rows:

- Rows for which all columns are Empty.
- Rows whose Reporting and Counterparty Country is the same thereby removing self-links.
- Rows whose Reporting and Counterparty Country belongs to Category other than countries as our study is strictly confined to Countries such as,
    1. Island country – Tuvalu
    2. Territory – Greenland, Wallus and Futuna, Turks and caicos (British) / Palestine territory / New Caledonia / cayman islands / Gibraltor / faeroe islands /

3. Islands - Bonaire, Sint Eustatius and Saba / St. Helena and dependencies / Falkland Islands / britsh overseas territory / US Pacific Islands / guernsey
4. Chinese Taipei – Taiwan International Organizations as one
5. Macao - Chinese special region
6. Isle of man - state
7. Congo democratic republic

Initially, the dataset contained 169,878 rows and 141 attributes which is reduced to a specific dataset with 3241 rows and 4 main attributes.

The cleaned dataset is focused on the problem statement. Since the study is to observe the transaction patterns before and after Covid, 2018 and 2020 are chosen. There was no strain of covid in 2018. 2019 had seen covid only in China at the year-end and hence, will not be suitable for the study. 2020, a well-known year for the spread of this pandemic. Hence, the transactions from Reporting country to Counterparty country in the year 2018 and 2020 are considered.

4.2. Data Exploration

The dataset is loaded as a Data frame with the help of Pandas, Python library for Data Analysis. Python comes with a set of inbuilt functions for exploring our data.

4.2.1. Summary of Central Tendency

- On average, 29,838 US Million dollars is the amount of transaction in 2018 and 32,426 US Million Dollars in 2020.
- The Standard Deviation for 2018 and 2020 is large, which tells that the data is quite vastly dispersed over the mean.

	2018	2020
count	3.241000e+03	3.241000e+03
mean	2.983846e+04	3.242687e+04
std	1.890882e+05	2.196310e+05
min	-1.900000e+01	-2.300000e+01
25%	2.986000e+00	2.574000e+00
50%	1.470000e+02	1.580000e+02
75%	3.640000e+03	3.650320e+03
max	5.912814e+06	7.242702e+06

Fig 1: Summary Description

4.2.2. Basic Plots

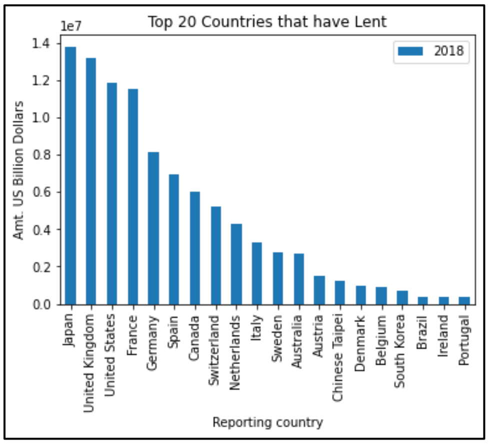

Fig 2: Lender’s plot for 2018

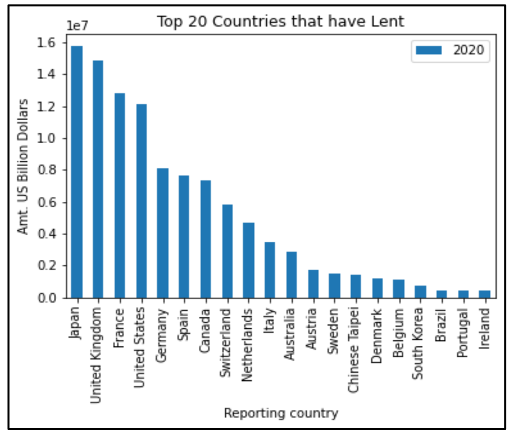

Fig 3: Lender’s plot for 2020

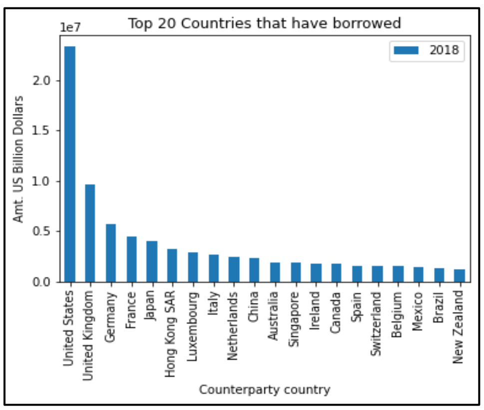

Fig 4: Borrower’s plot for 2018

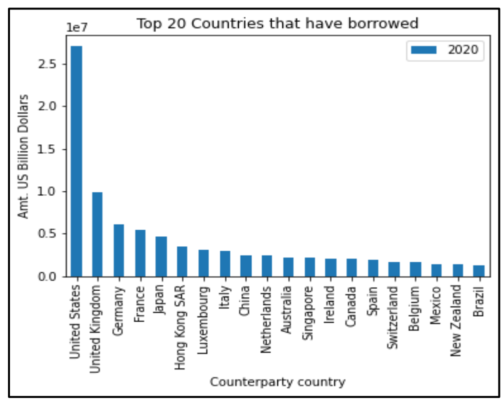

Fig 5: Borrower’s plot for 2020

4.2.3. Paired T-Test

Paired T-test is used as a measure to check the statistical independence between 'Two related' samples i.e., different variables but belonging to a similar group. In our case, the transaction attribute for 2018 and 2020 is checked for statistical independence.

For performing paired t-test, a null hypothesis and an alternate hypothesis are stated to see which hypothesis is rejected.

H0:Two attributes are not statistically independent

H1: Two attributes are statistically independent

In this case, the result is as follows:

Ttest_relResult(statistic=-3.8319624226547564, pvalue=0.00012953756175388475)

- the p-value is much lesser than the significance level 0.05 concluding that '2018' and '2020' are statistically different as a lower p-value rejects the Null Hypothesis of the two samples being equal.

4.3. Network Building

Network analysis is chosen as a method to analyze this field because organizing such huge data in the form of nodes and edges will facilitate us to understand the relationships in the social network formed.

Graph networks are built on empirical data, data obtained from observations. It can be a description of a real-world problem depicting the patterns in the relationships among actors[19]. By presenting the data in the form of graphs, computational efficiency is improved.

4.3.1. NetworkX

NetworkX is an inbuilt python library that is used to understand the dynamics, topology, and measures of a problem represented in the form of a network or graph.

Steps to construct a graph:

- A graph is a mathematical data type for networkx that holds the nodes or entities and the relation between them as edges.
- Initially, an empty graph is created with nx.graph() command to which we can further add the nodes and edges. In this case, the nodes are the countries, and the edges will be the money flow across them.
- Edgelists for 2018 and 2020 are created using nx.from_pandas_edgelist() command which creates a text file consisting of the edge list in the form Reporting Country, Counterparty Country, Weight (amount lent) is generated.

Three main steps follow the network edge list definition:

- Various Network Measures
- Results from Polya Filter Algorithm
- Visualization of the network

4.3.1.2. Network Measures

Various metrics are designed to understand the behavior of the network. Most of the measures stated below deal with the Node Activity and how a node (Country In this case) influences the network.

The degree of a node represents the number of neighbors it has or the number of links entering the node(in-degree) or leaving the node(out-degree). Since, in this case, the network is a directed network, the out-degree represents the lending links flowing from the lender to the borrower, and in-degree is borrowing links flowing from borrowers to lenders.

The clustering coefficient is a property of a node that indicates how well its neighbors are connected. If the value of the Clustering coefficient is closer to 1, neighbors are highly connected or knitted amongst them, if the value is closer to 0, the neighbors don’t have a connection.

In social networks, “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 nodes” [15,16]

Centrality indicates how important or central a particular node is in the network that is built.

- Degree centrality includes the in-degree and out-degree measures. In simple terms, nodes with more connections are more important or dominant in the network.
- Betweenness Centrality is measured based on the shortest path between the nodes. It indicates how important a node is in the flow of information across the nodes in the network. A node with high betweenness centrality means a large number of shortest paths include this node thereby giving information about various social circles[17].
- Closeness centrality based on the distance between the nodes indicates how fast the information can spread across the network with nodes being closely placed[18]
- Eigenvector centrality measures the influence of a node in the network, a higher eigenvector centrality value indicates that the node has a huge influence on the network. In

contrast to degree centrality, a node with a high degree does not mean that it can have a high eigenvector centrality value, as eigenvector value is based on neighbors' connections.

Table 1: Network Measure values for 2018 and 2020 network

Network Indicator / Measure	**2018**	**2020**
Average Degree	27.5707	26.7804
Mean of In- degree	13.7854	13.3902
Mean of out-degree	13.7854	13.3902
Std. of In-degree	6.5356	6.7998
Std. of Out-degree	41.2210	39.9327
Average Clustering Coefficient	0.8589	0.8491
Top 10 Nodes based on Degree (Lender)	('Switzerland', 197) ('France', 185) ('Australia', 181) ('United Kingdom', 177) ('Austria', 175) ('United States', 168) ('Sweden', 167) ('Spain', 164) ('Belgium', 137) ('Italy', 127)	('Switzerland', 195) ('France', 186) ('Australia', 175) ('United States', 172) ('Austria', 170) ('United Kingdom', 169) ('Spain', 167) ('Sweden', 144) ('Belgium', 128) ('Turkey', 120)
Top 10 Nodes based on Degree (Borrowers)	('Luxembourg', 26) ('Singapore', 26) ('Canada', 25) ('Switzerland', 25) ('Germany', 25)	('Luxembourg', 26) ('Canada', 25) ('Switzerland', 25) ('Germany', 25) ('Spain', 25)

	('Spain', 25) ('France', 25) ('United Kingdom', 25) ('Hong Kong SAR', 25) ('Japan', 25)	('France', 25) ('United Kingdom', 25) ('Hong Kong SAR', 25) ('Japan', 25) ('Netherlands', 25)

DIRECTED GRAPH (REPORTING COUNTRY -> COUNTERPARTY COUNTRY (2018):

```
Based on Betweeness Centrality
 ['Switzerland', 'Australia', 'France', 'United Kingdom', 'United States', 'Austria', 'Sweden', 'Spain', 'Belgium', 'Italy']
Based on Closeness Centrality
 ['Luxembourg', 'Singapore', 'Hong Kong SAR', 'Norway', 'Canada', 'Switzerland', 'Germany', 'Spain', 'France', 'United Kingdo
m']
Based on Eigenvector Centrality
 ['Luxembourg', 'Singapore', 'Hong Kong SAR', 'Norway', 'Canada', 'Switzerland', 'Germany', 'Spain', 'France', 'United Kingdo
m']
```

Fig 6: Central Nodes for 2018

DIRECTED GRAPH (REPORTING COUNTRY -> COUNTERPARTY COUNTRY (2020):

```
Based on Betweeness Centrality
 ['Switzerland', 'Australia', 'France', 'United States', 'Spain', 'United Kingdom', 'Austria', 'Sweden', 'Turkey', 'Belgium']
Based on Closeness Centrality
 ['Luxembourg', 'Hong Kong SAR', 'Singapore', 'Canada', 'Switzerland', 'Germany', 'Spain', 'France', 'United Kingdom', 'Japan']
Based on Eigenvector Centrality
 ['Luxembourg', 'Hong Kong SAR', 'Singapore', 'Canada', 'Switzerland', 'Germany', 'Spain', 'France', 'United Kingdom', 'Japan']
```

Fig 7: Central Nodes for 2020

Thus, from the basic network measures, we can observe that there are not many significant differences between the 2018 and 2020 results.

4.3.1.3. Polya Filter Algorithm

The Polya filter algorithm proposed in the paper focuses on a null hypothesis with respect to a particular heterogeneity of the network. This filtering methodology is based on the Polya urn problem.

Polya urn problem:

In this problem, the urn contains red and blue balls. A ball of color x is drawn at random and replaced with the urn and an additional ball of the same color is added to the urn. This process is repetitive leading to a "self-reinforcing property". This problem is compared to the analogy of "The Rich get richer" because there is a more likely probability that the same color ball could be drawn at subsequent draws.

How is Polya Urn Problem related to network?

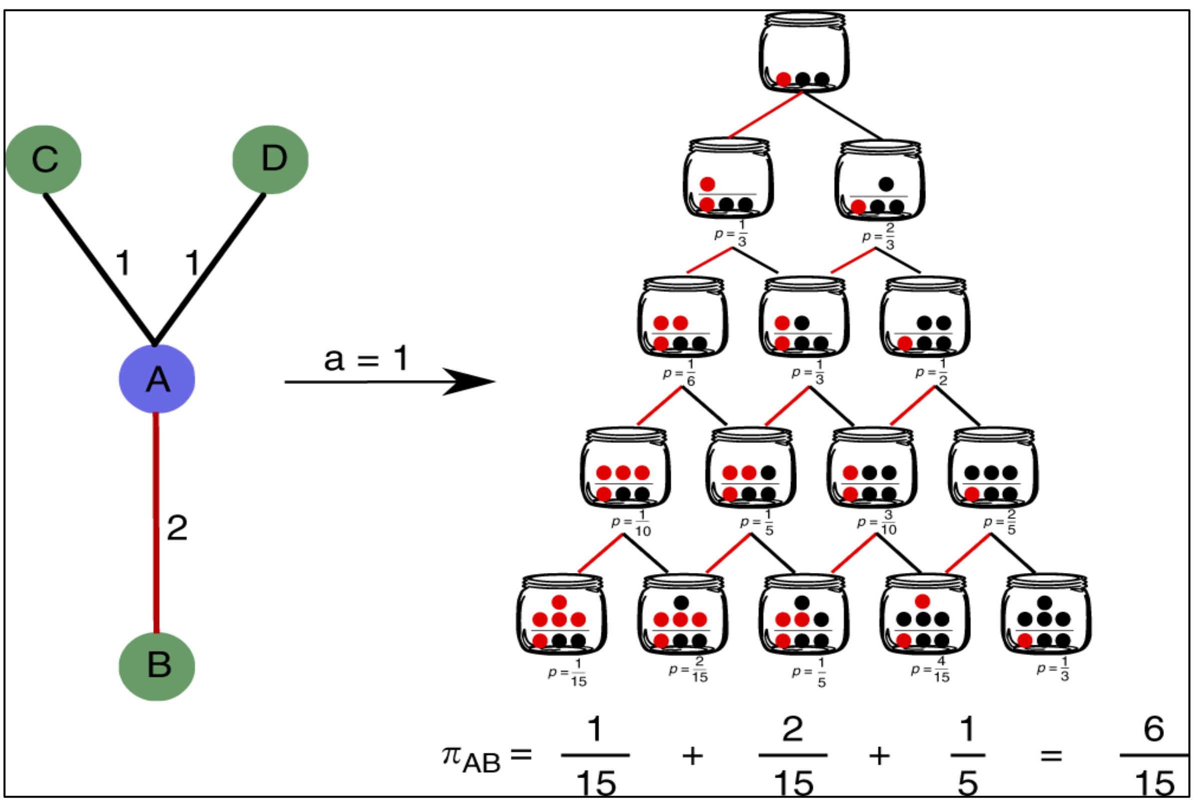

Fig 8: Depiction between Network and Polya urn problem[9]

This sketch depicts the relationship between the Polya urn problem and the network. The aim is to find the statistical significance of the edge connecting nodes A-B with weight 2. This is calculated from Node A's viewpoint whose degree, k=3 and strength, s=4. If we convert the same network in terms of a Polya urn problem, the problem starts with an urn consisting of one red ball and (k-1), 2 black balls. The right picture shows the combination for s draws and their corresponding probabilities keeping a=1, that is replace one ball of the last ball drawn. The ultimate results to find the p-value which is the sum of all outcomes with at least w red balls in addition to the initial red ball which is 2 red balls as per the network.

Filtering Based on Polya Urn:

The Polya urn problem is converted into network terms which form the Null hypothesis of our problem. This filter is going to assess the statistical significance of a link whose weight is 'w' with a node degree 'k' and strength 's'. A governing parameter 'a' is used for the reinforcement mechanism, ie., adding a ball of the same color to the urn.

With these values, a p-value for each link is generated that indicates how favorable that link is.

On applying the filter to the network, we get a set of p-values for each link in the network.

The P value is assigned as per the following equation

$$P(w|k,s,a) = \binom{s}{w} \frac{\left\{B\left(\left\{\frac{1}{a} + w, \frac{\{k-1\}}{a} + s - w\right\}\right)\right\}}{\left\{B\left(\left\{\frac{1}{a}, \frac{\{k-1\}}{a}\right\}\right)\right\}}.$$

The process of drawing a putting back 'a' new balls of the same color follows the Beta-Binomial distribution B

This is compared against a significance value to find the most significant links from the network.

Significance Level:

Since the process involves testing all links, there is a need to fix a "univariate significance level".

Bonferroni correction is used in our case which strictly rejects the false positives. This correction fixes a significance level by dividing the alpha value by the total number of tests.

In this case, the total number of tests is the total number of links with 5% significance level. This correction is used as a threshold to eliminate the insignificant links from the network.

4.3.2. Inferences drawn from the results

- From the degree, the largest nodes also known as Hubs can be identified. As per the Table, Nodes when organized based on In-Degree and Out-Degree have similar countries in order with Switzerland having large outward links and Luxembourg having a large number of inward links.
- The clustering coefficient for 2018 and 2020 is closer to 1 stating the network is highly connected with lesser or no presence weaker ties.
- The figure above shows the most central nodes based on centrality measures. The results prove that 2018 and 2020 share similar central nodes.
- Polya filter yielded 127 significant links out of 2826 links for the year 2018
- Polya filter yielded 117 significant links out of 2796 links for the year 2020.
- Based on the links generated, there is no similarity between the links that have the same lender and borrower.
- Thus, to check whether preferential relationships amongst networks do not exist generally, or is it only during covid, 2017 network is also constructed and filtered.
- It is shown that there are no similar links between 2017 and 2018 as well that proves the non-existence of preferential relationship.
- Jaccard similarity, an optimal way to find the similarity between the links in the filtered network and original network helps to check whether the backbone similarity is maintained.
    - JI (Original 2018, Original 2020) – 88.6 %
    - JI (Filtered 2018, 2020) – 7.5 %
- For 2018, Netherlands and Finland are not among the significant link lenders.
- For 2020, the Netherlands and Chinese Tapai are not among the significant link lenders.

4.3.3. Analysis of results from filtered links

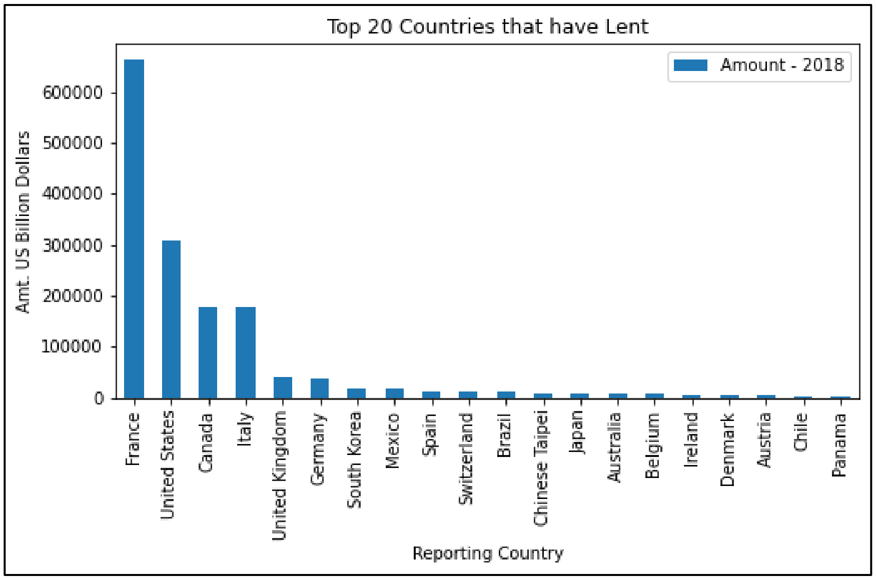

Fig 9: Lender's plot from filtered 2018 values

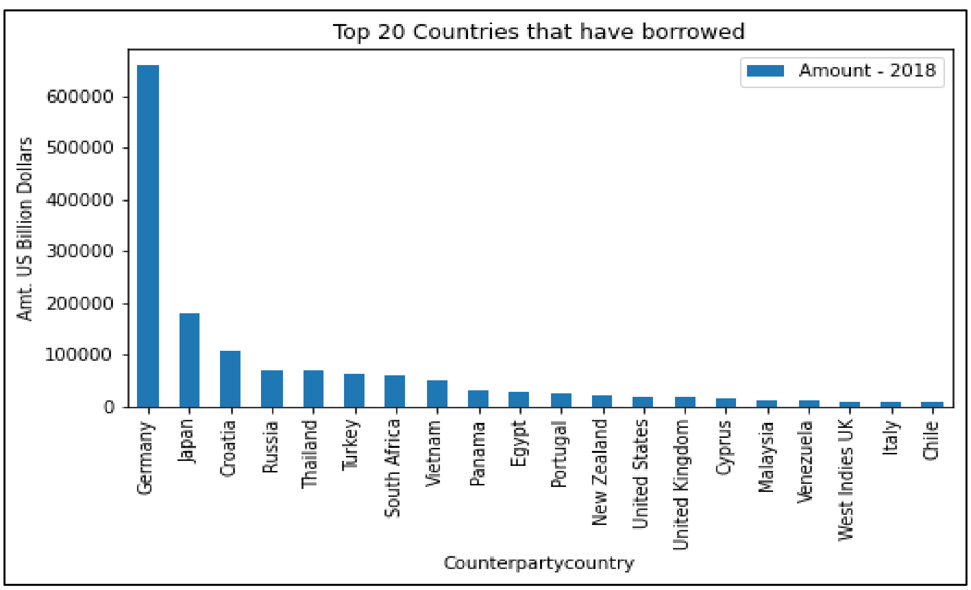

Fig 10: Borrower's plot from filtered 2018 values

In 2018, according to the filtered results, we can see that France has been the dominant lender as it has lent more amount of money and Germany has been the dominant borrower.

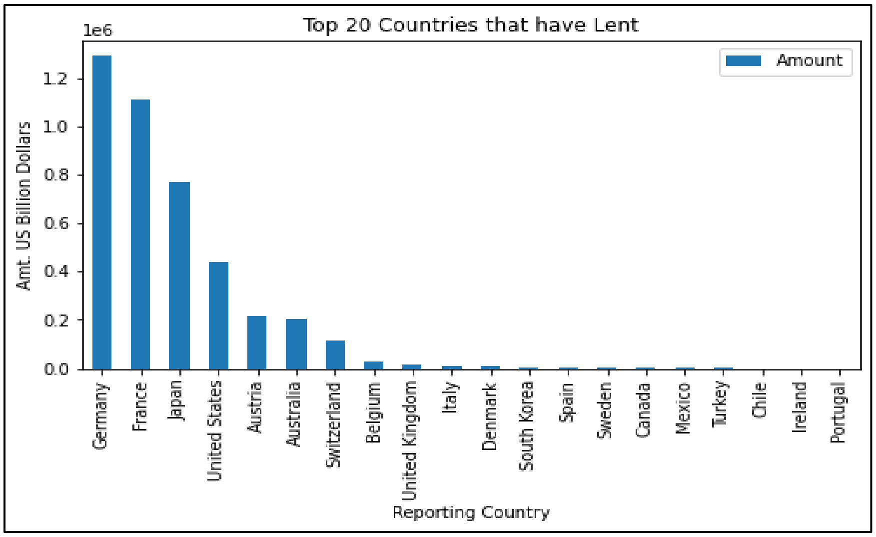

Fig 11: Lender's plot from filtered 2020 values

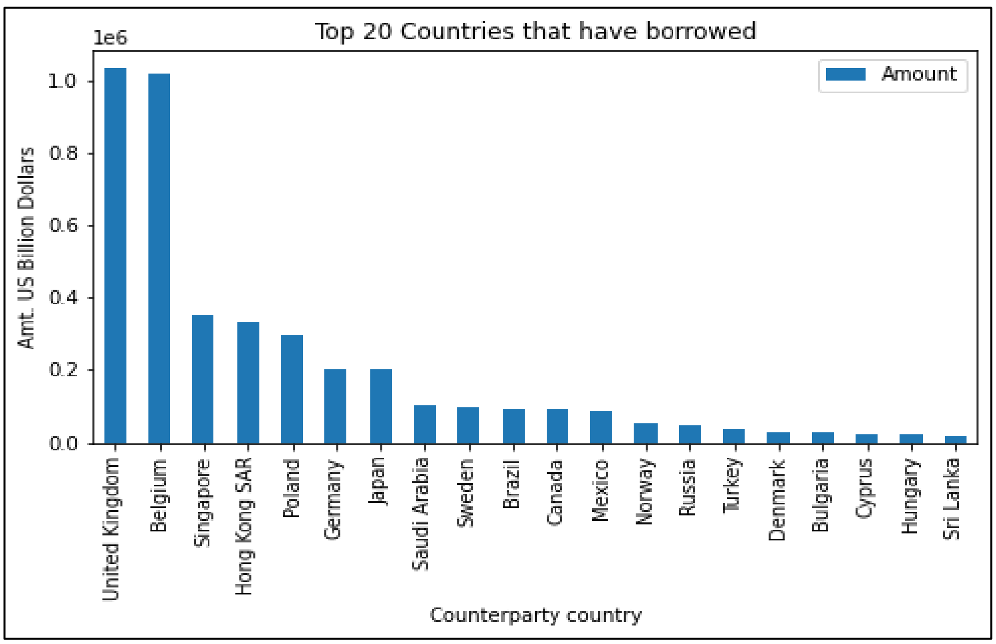

Fig 12: Borrower's plot from filtered 2020 values

In 2020, the scenarios have changed. Germany is the dominant lender and the United Kingdom is the dominant borrower.

Chapter 5 - Conclusion

5.1 Conclusion

The whole study revolved around understanding the bilateral capital flows across countries based on reporting status, claims nature, location of the banks, and outstanding liabilities. Representing the scenario of transactions in the form of a network or a mathematical graph allowed us to understand the problem and inferences better. 2018, a year of no covid, and 2020, a prominent pandemic year was taken for analysis. With the network measures, such as degree, clustering coefficient, centrality measures, and strength, we find that 2018 and 2020 represent similar numbers stating that there are no significant changes in the funds lent or borrowed because of Covid.

Thus, we moved towards the Polya filter algorithm to see whether there are new inferences drawn. The filtered algorithm stated a very strong observation that there are no preferential relationships in 2018 or 2020, i.e., the links between the same lender and borrower are not maintained in either of the 2 years. This states that there are no similar transactions before and after Covid. Before filtering, Japan was the dominant lender in 2018 but after filtering, we see that France is the dominant lender. Similarly, for Borrower, Germany is the dominant borrower in 2018 in contrast to the United States from unfiltered value. When we observe the Covid Year 2020, Germany which was a dominant borrower has become the dominant lender while the UK is the dominant borrower in the filtered results.

Thus, we can say from the statistically significant links, the UK has been in short of funds and thus was pushed to the situation of liability followed by Belgium and Singapore. The countries which were able to lend even in times of Covid have had surplus funds saved for emergencies such as an outbreak or a natural disaster.

5.2 Future Work

This project can be further extended by applying different corrections to fix the significance level to find the filtered links. All the smaller territories and islands should also be included in the study to understand their influence on this study. Further, more network concepts such as community detection and graph-based algorithms can be applied to see if their results add more to the inferences drawn here. We can also study and compare more years in our network analysis.

6. BIBLIOGRAPHY

1. Clemente, G. P., Grassi, R., & Pederzoli, C. (2020). Networks and market-based measures of systemic risk: the European banking system in the aftermath of the financial crisis. Journal of Economic Interaction and Coordination, 15(1), 159-181.
2. Bongini, P., Clemente, G. P., & Grassi, R. (2018). Interconnectedness, G-SIBs and network dynamics of global banking. Finance Research Letters, 27, 185-192.
3. Cerqueti, R., Clemente, G. P., & Grassi, R. (2020). Systemic risk assessment through high order clustering coefficient. Annals of Operations Research, 1-23.
4. Vidal-Tomás, D., Tedeschi, G., & Ripollés, J. (2020). The desertion of rich countries and the mutual support of the poor ones: Preferential lending agreements among the PIGS. Finance Research Letters, 34, 101255.
5. Minoiu, C., & Reyes, J. A. (2013). A network analysis of global banking: 1978– 2010. Journal of Financial Stability, 9(2), 168-184.
6. Serena, J. M., Tissot, B., Doerr, S., & Gambacorta, L. (2021). Use of big data sources and applications at central banks (No. 13). Bank for International Settlements.
7. McGuire, P., & Wooldridge, P. D. (2005). The BIS consolidated banking statistics: structure, uses and recent enhancements. *BIS Quarterly Review, September*.
8. Radicchi, F., Ramasco, J. J., & Fortunato, S. (2011). Information filtering in complex weighted networks. *Physical Review E*, *83*(4), 046101.
9. Marcaccioli, R., & Livan, G. (2019). A Pólya urn approach to information filtering in complex networks. Nature communications, 10(1), 1-10.
10. Battiston, S., Puliga, M., Kaushik, R., Tasca, P., & Caldarelli, G. (2012). Debtrank: Too central to fail? financial networks, the fed and systemic risk. Scientific reports, 2(1), 1- 6.
11. Iori, G., De Masi, G., Precup, O. V., Gabbi, G., & Caldarelli, G. (2008). A network analysis of the Italian overnight money market. Journal of Economic Dynamics and Control, 32(1), 259-278.
12. De Masi, G., Fujiwara, Y., Gallegati, M., Greenwald, B., & Stiglitz, J. E. (2011). Japanese financial credit networks. Evolutionary and Institutional Economics Review, forthcoming.

13. De Masi, G. (2009). Empirical analysis of the architecture of the interbank market and credit market using network theory. In Networks, Topology and Dynamics (pp. 241-256). Springer, Berlin, Heidelberg.
14. Garlaschelli, D., Battiston, S., Castri, M., Servedio, V. D., & Caldarelli, G. (2005). The scale-free topology of market investments. Physica A: Statistical Mechanics and its Applications, 350(2-4), 491-499. 13. Newman, M. E. (2003). The structure and function of complex networks. SIAM review, 45(2), 167-256.
15. Holland, P. W., & Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative group studies*, *2*(2), 107-124.
16. Watts, D., & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440-442.
17. Golbeck, J. (2015). *Introduction to social media investigation: A hands-on approach*. Syngress.
18. Zhang, J., & Luo, Y. (2017, March). Degree centrality, betweenness centrality, and closeness centrality in social network. In *Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017)* (Vol. 132, pp. 300-303).
19. Luke, D. A., & Harris, J. K. (2007). Network analysis in public health: history, methods, and applications. *Annu. Rev. Public Health*, *28*, 69-93.

7. APPENDIX

Robustness Analysis:

Kolmogorov-Smirnov Test:

The Kolmogorov-Smirnov test is used to test the null hypothesis that a set of data comes from a normal distribution. The 2-sample KM test checks whether the two samples come from the same distribution.

The KM Test for the initial 2018 and 2020 gives the result of the p-value as 0 which is less than the significant level. Thus, we can reject the null hypothesis that the distribution is not normal in nature.

Best Distribution that fits the data from KM Test against various distributions:

For 2018:

The distribution with least p-value :

```
betaprime: statistic=0.10491025713288443, pvalue=3.974682880616789e-29
mielke: statistic=0.11213710479220179, pvalue=3.1586735870759676e-33
lognorm: statistic=0.13676879447695475, pvalue=2.8054924027100014e-49
```

For 2020:

The distribution with least p-value:

```
johnsonsu: statistic=0.08246712430580636, pvalue=3.780073884000574e-18
betaprime: statistic=0.11696350294461366, pvalue=4.0529309208822264e-36
mielke: statistic=0.12390823274710763, pvalue=1.6998301305167225e-40
```

Thus, Betaprime distribution commonly seems to fit both the data columns.

Network Measure plots:

Degree Distribution for 2018:

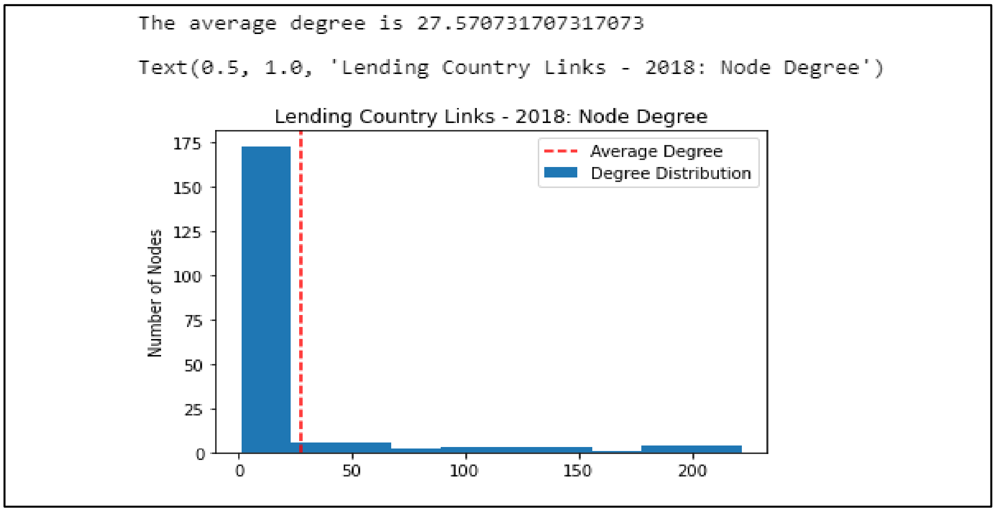

Fig 13: Node degree for unfiltered 2018 network

Local Clustering Coefficient:

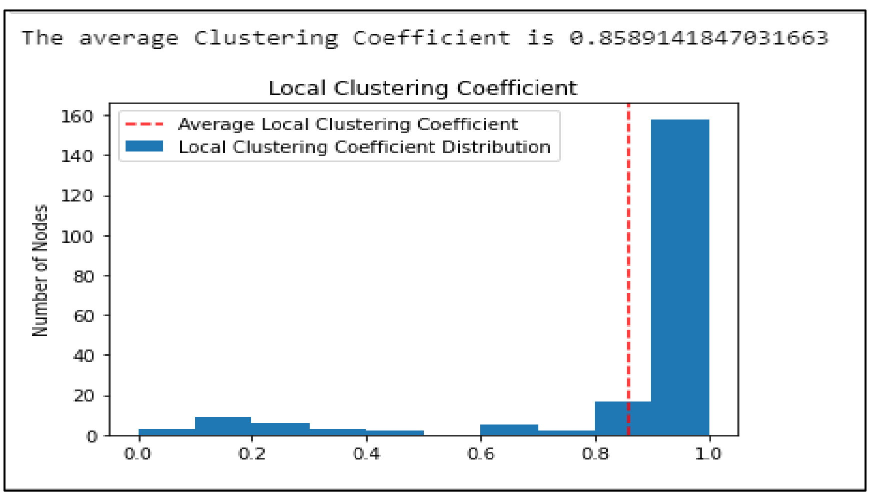

Fig 14: Clustering Coefficient plot for unfiltered 2018 network

GLOBAL BANKING NETWORK BETWEEN ONLY THE CORE LENDERS – 2018:

This network is constructed with values greater than the 75th Percentile of the data and between only the Core Reporting Countries.

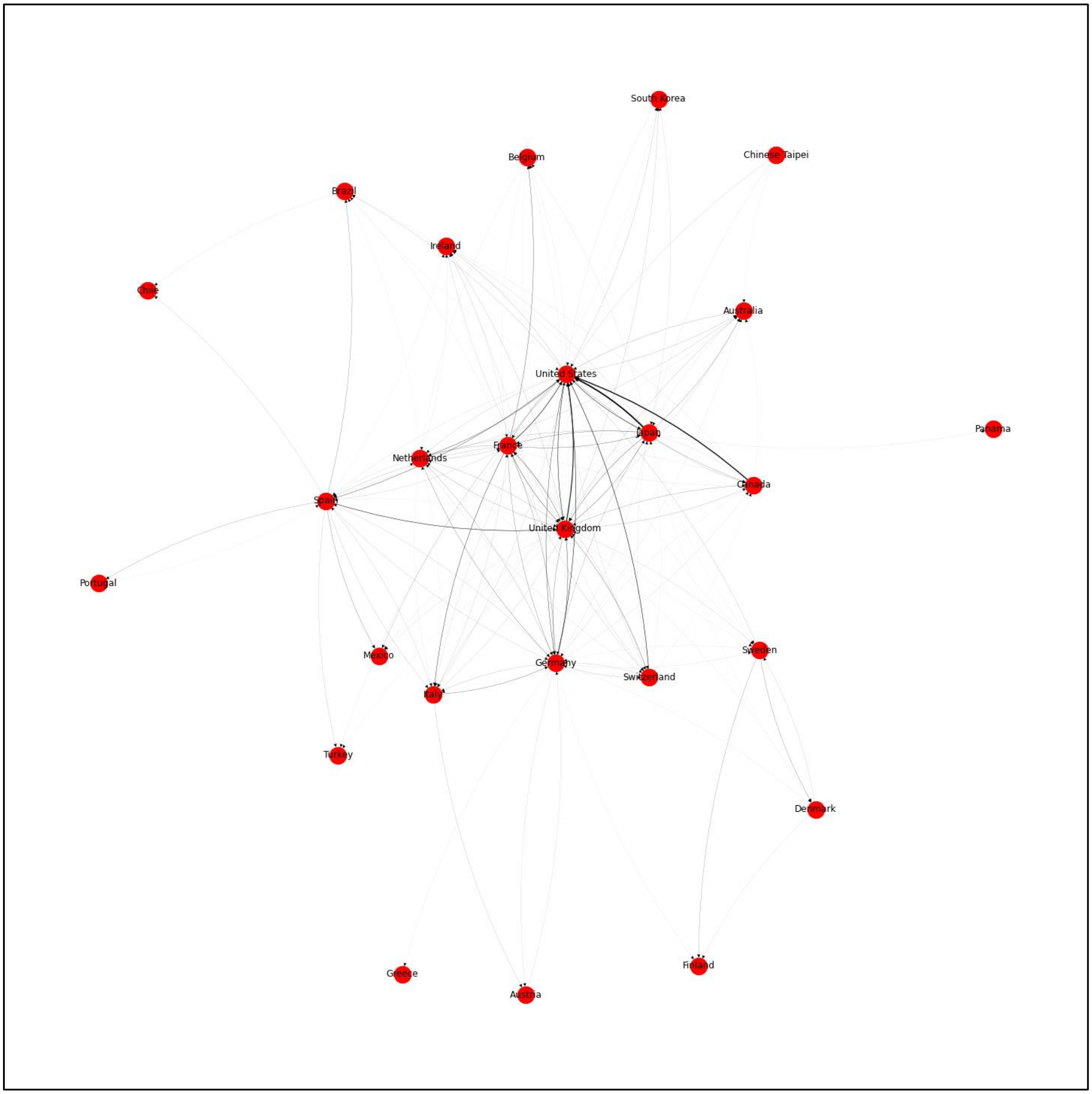

Fig 15: Overview of unfiltered 2018 network

Degree Distribution for 2020:

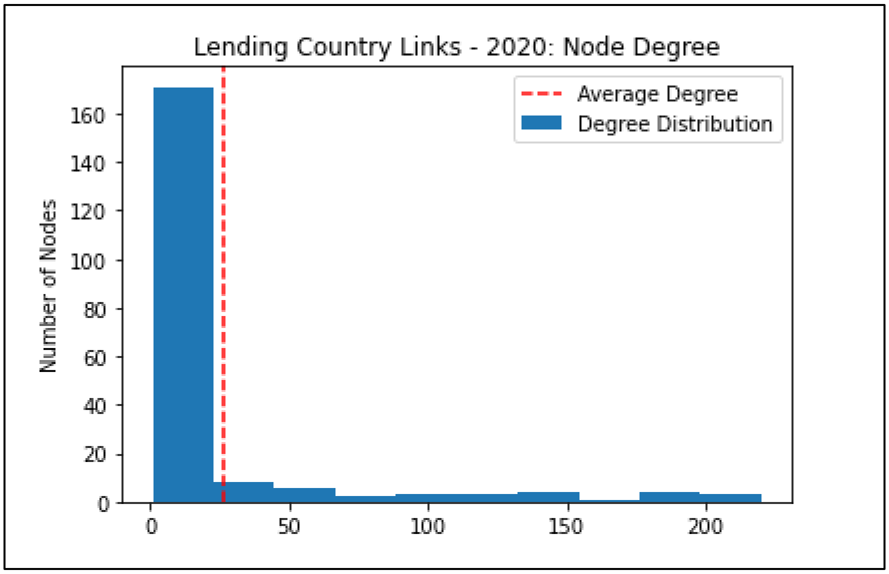

Fig 16: Node degree for unfiltered 2020 network

Local Clustering Coefficient 2020:

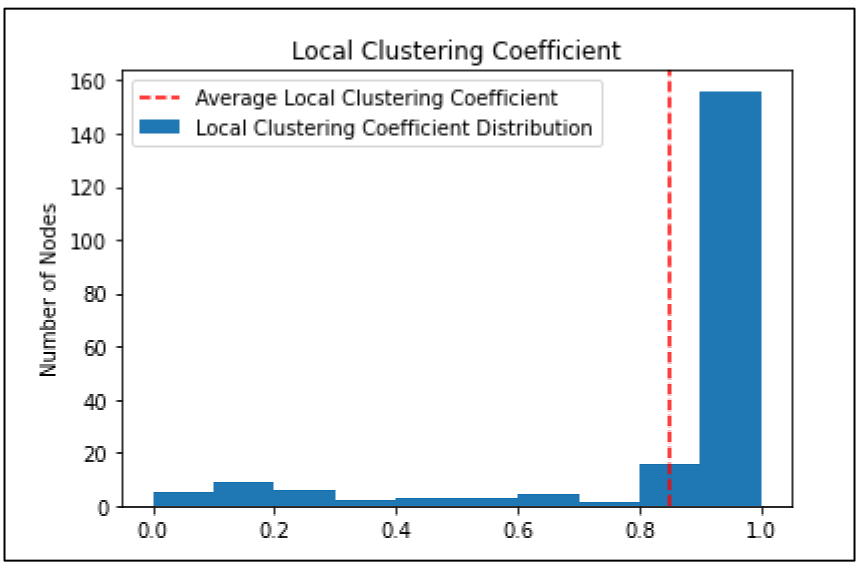

Fig 17: Clustering Coefficient plot for unfiltered 2020 network

GLOBAL BANKING NETWORK BETWEEN ONLY THE CORE LENDERS – 2020:

This network is constructed with values greater than the 75th Percentile of the data and between only the Core Reporting Countries.

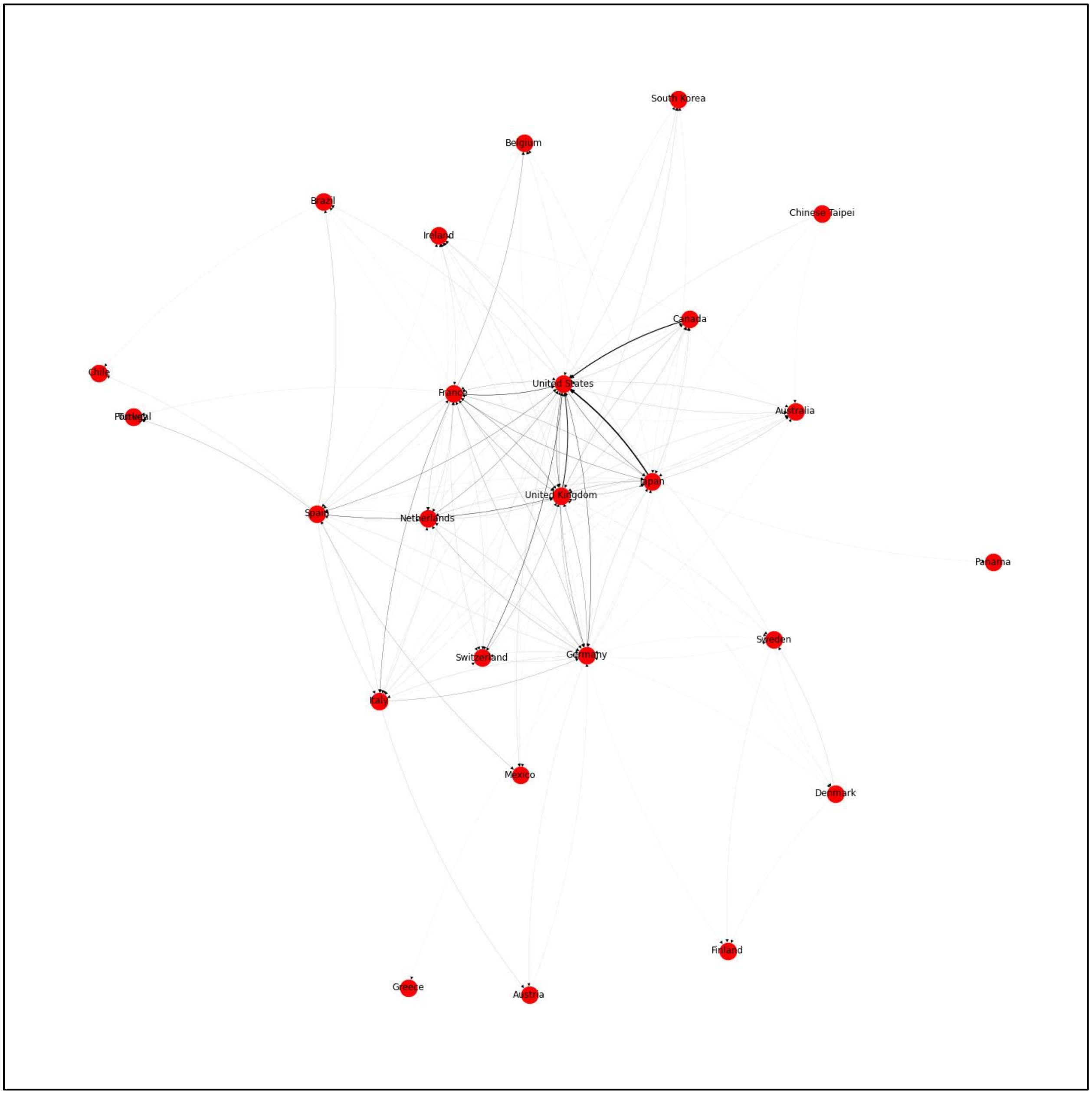

Fig 18: Overview of unfiltered 2020 network

Filtered Network Plotted over World Map using Gephi Tool:

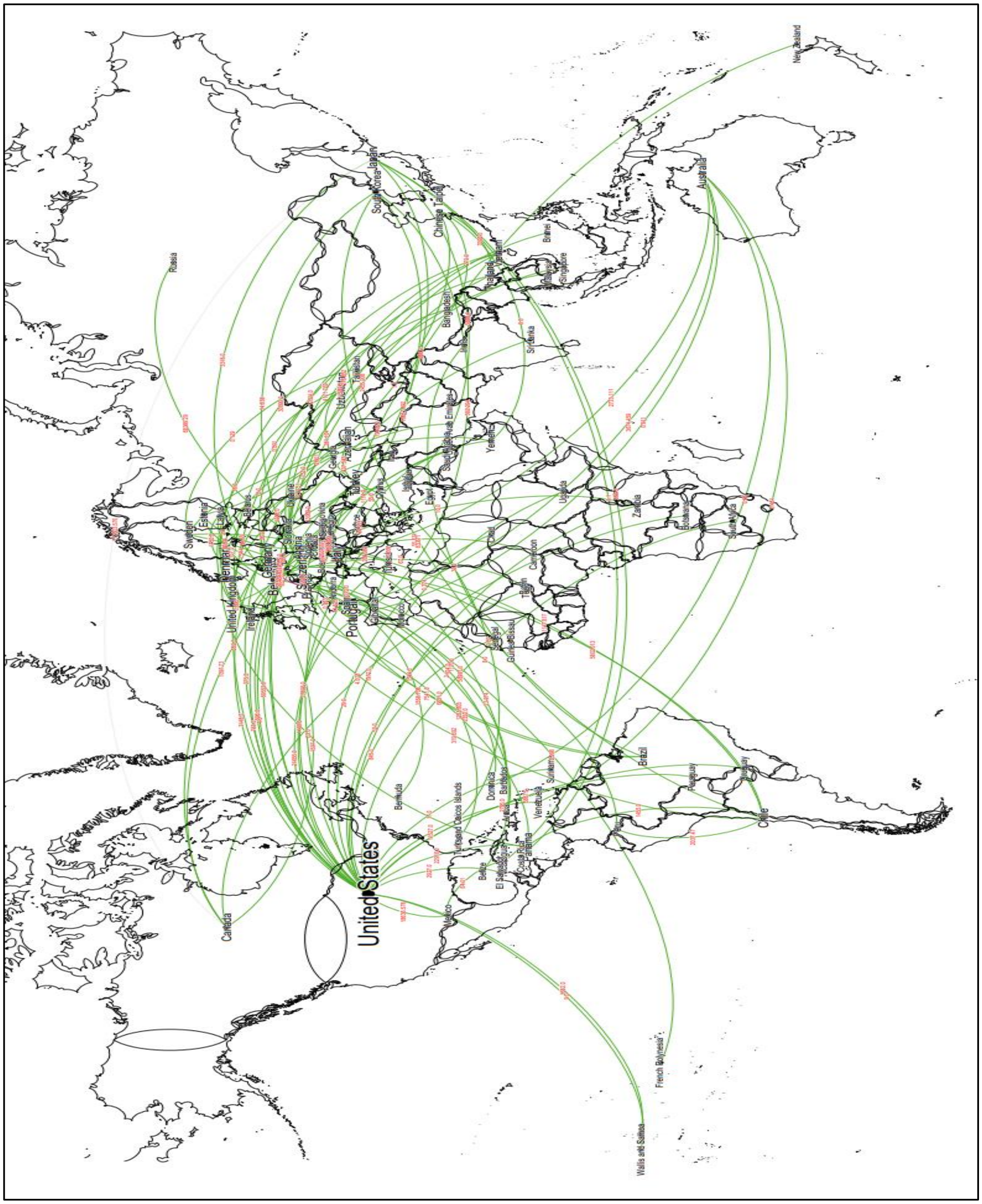

Fig 19: Filtered 2018 Network

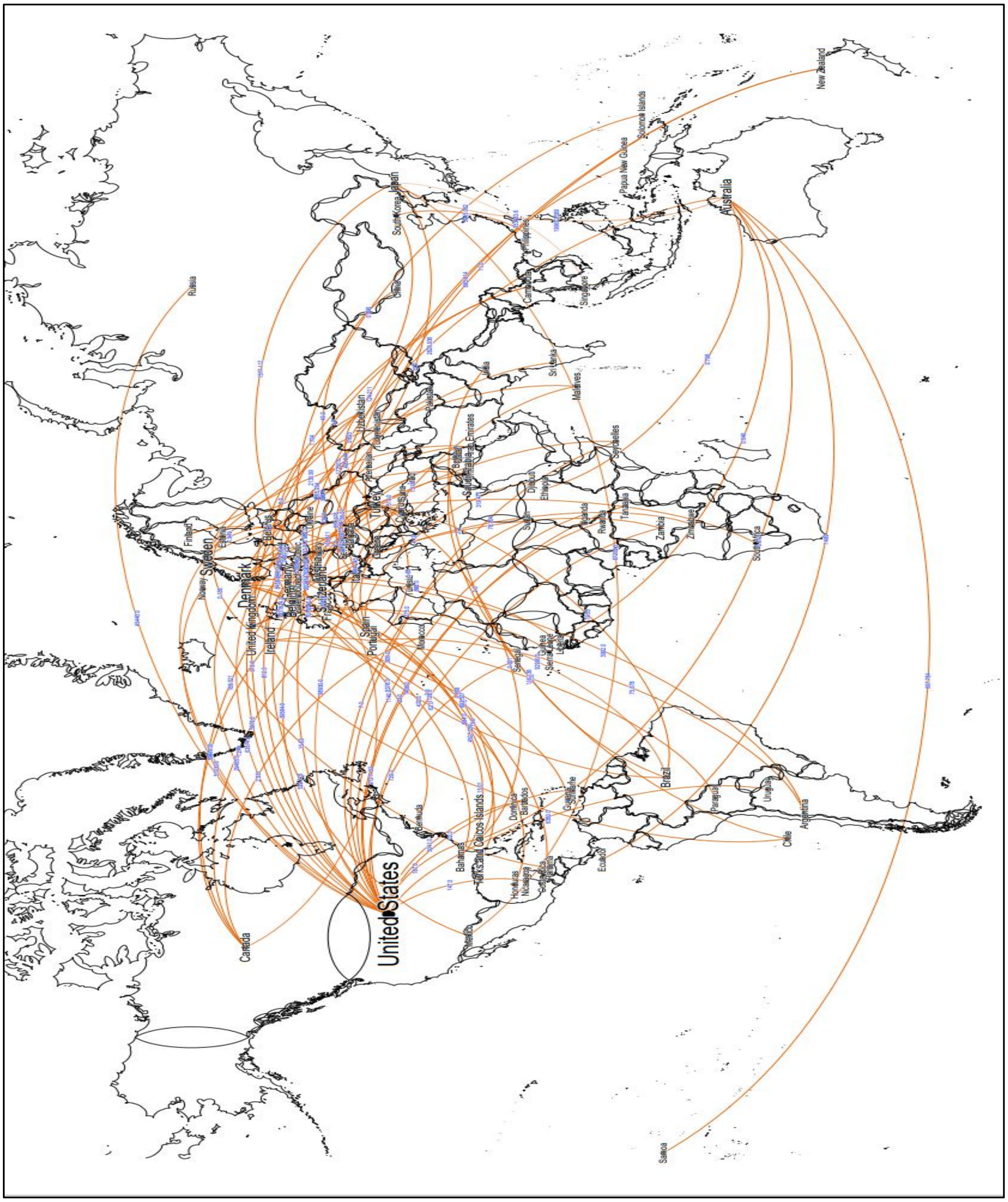

Fig 20: Filtered 2020 Network