Network analysis of cross-border banking exposures

ZHENYU ZHANG

Master of Science
School of Informatics
University of Edinburgh
2023

Abstract

The cross-border banking exposures are double-edged swords, which can improve the efficiency of global capital and resource allocation, but may also cause cascading contagions leading to a global financial crisis. Different from many traditional data analysis methods based on mathematical statistics, this project aims to analyze cross-border banking exposures by performing network analysis. In general, to investigate research questions like the evolution of network properties, the risk level of each country, and the community structures in financial networks, this project conducts applied and exploratory research on global banking exposure data by applying network science methodology. To be specific, this project first performs inspections on the raw exposure data, which are then converted into temporal networks with banks as nodes and exposures as links. Next, it adds new features on these networks to better represent the inherent and propagated risks in each country. Finally, it performs comprehensive static and dynamic analysis, including basic statistics, homogeneous statistics, heterogeneous statistics, community detection by stochastic block models and even the simulation of exogenous shocks on the networks. As a result, throughout the implementation of this network analysis, a total of 8 insights or methodologies have been summarized and have finally been used to adequately answer the research question. In the future, there are still more details to be analyzed, more insights to be found, and more research strategies to be attempted in this field.

Research Ethics Approval

This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*ZHENYU ZHANG*)

Acknowledgements

I would like to express my gratitude to my supervisor, Valerio Restocchi for his guidance and support throughout this project. In addition, I want to show my thanks to project partner FNA for providing data, which gives me the opportunity to make an attempt to apply theoretical knowledge into practice. Finally, I would like to thank my family and friends.

Table of Contents

Chapter 1

Introduction

1.1 Motivation

According to the Cambridge Dictionary, the business meaning of exposure is, the risk of losing money, for example through a loan when a borrower defaults or an investment when it fails [11]. So cross-border banking exposures can also be expressed as the financial risks arising from the cross-border borrowing and lending business of international banks. While cross-border capital flows can facilitate a more efficient and sustainable allocation of resources around the world [16], it would also cause systemic risks [26]. For example, since the financial crisis in 2008, Europe as a whole had been falling into a sovereign debt crisis for quite a long time, which was only caused by the sovereign crisis of some countries [8].

Currently, there are many attempts to measure and regulate the potential systemic risk in the financial system. For measurement, indicators like value at risk (VaR) and expected shortfall (ES) are widely practised to quantify the potential loss risk of a company or portfolio [38]. Further, systemic expected shortfall (SES), conditional value at risk (CoVaR), and Granger causality are improved to measure systemic risk [1, 2, 7]. In terms of regulations, for example, according to the requirements of the Basel Accords, each bank must maintain a total risk-weighted capital ratio of at least 8% [12].

In order to focus more on analysis and pursue stronger interpretability, network analysis models have begun to be applied. The cross-border risk exposure of banks can be intuitively considered as a financial network, with banking institutions as nodes, and the risk exposures between banks as edges. There are usually two main methods: the first one is static network indicator analysis, and the second one is dynamic network stress testing [22].

1.2 Problem Statement

This project seeks to investigate questions such as:

- How do the cross-border banking network properties / statistics evolve over time?
- What are the individual and global systemic risk levels of each country?
- Are there any non-obvious community structures in the network?

1.3 Research Objectives

For research objectives [34, 36], this project aims to conduct applied and exploratory research on global banking exposure data by applying network science methodology. Specifically, there are several sub-goals:

- By conducting static network analysis, we will analyze things like which countries could be at risk of defaulting, or at risk of being defaulted.
- By conducting dynamic network analysis, we can do a stress test on the system, and then analyze how the network could change after external shocks.
- Combining the results of network analysis, we will try to answer our research questions.
- We also want to help banks optimize cross-border investment decisions. For example, the banks may realize which other countries they need to pay attention to when they have invested in some countries.

1.4 Timeliness and Novelty

Although graph theory has been established for a long time, the development of network science started relatively late, at around 2000 [6]. Therefore, from the perspective of timeliness, it would be very valuable to study the application of network science in the financial field.

As to novelty, there have been some studies exploring the use of network science and technology to solve similar problems [19, 29]. Based on these, this project will attempt to apply their methodology to a new data set to solve real business problems faced by enterprises. In addition, it will also explore combining community detection and other methods to conduct a more in-depth analysis.

1.5 Outline

The following is the structure of the remaining part:

- **Chapter 2 Background and Related Work** introduces the basic knowledge of network science and some application cases for static and dynamic analysis in the cross-border banking business.
- **Chapter 3 Methodology** demonstrates the methodology in the project design process, including research approaches and strategies, as well as the selection of specific techniques, and the overview of the entire analysis.
- **Chapter 4 Implementation, Results and Analysis** implements the project and analyzes the results. Besides, it continuously summarizes relevant insights during the process.
- **Chapter 5 Evaluation** evaluates the results of the project by answering the research question with all the insights.
- **Chapter 6 Conclusions** concludes the current work, and illustrates the limitations on current project and suggestions for future work.

Chapter 2

Background and Related Work

2.1 Network Science

Network Science can be used to describe the system in a top-down analysis, where we represent a system as a network / graph with distinct elements / entities as nodes, and the connections / interactions between nodes as links. This can include social networks, financial networks, etc [28].

Figure 2.1.1 is an example of a directed and weighted graph.

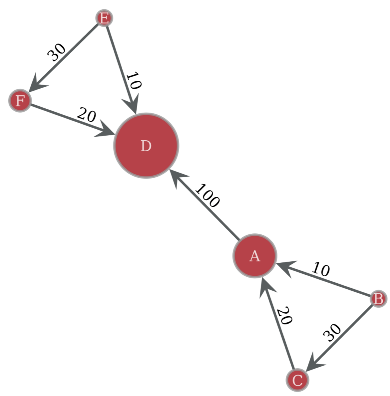


Figure 2.1.1: An example of a directed and weighted graph

2.1.1 Basic Statistics

Some main basic network properties are as follows:

- N represents the total number of nodes in a network, while L represents the total number of links in a network.

- According to whether the edge has direction and whether it has weight, the graph can be classified into 4 types (i.e., undirected and unweighted graph, undirected and weighted graph, directed and unweighted graph, directed and weighted graph).
- L_{max} tells the potential maximum number of links in a network with N nodes.
  - For an undirected graph, $L_{max} = \frac{N(N-1)}{2}$.
  - For a directed graph, $L_{max} = N(N-1)$.
- Density $d = \frac{L}{L_{max}}$, and it is common that density is quite low in social networks.
- The degree k of a node is the number of links the node has. This is equivalent to saying it is the number of neighbours a node has.
  - For a directed graph, there are three types of degree, which are in-degree k_{in}, out-degree k_{out} and total-degree k_{total}.
  - For a weighted graph, a node has a strength $s_i = \sum_{ij} w_{ij}$. And if it is also a directed graph, there are three types of strength, which are in-strength s_{in}, out-strength s_{out} and total-strength s_{total}.

For example, Figure 2.1.1 is a directed and weighted graph, with $N = 6$, $L = 7$, $L_{max} = N(N-1) = 30$, $d = \frac{L}{L_{max}} = \frac{7}{30}$. For node A, $k_{in} = 2$, $k_{out} = 1$, $k_{total} = 3$ and $s_{in} = 30$, $s_{out} = 100$, $s_{total} = 130$.

2.1.2 Homogeneous Statistics

Homogeneity, or homophily, can be considered as the tendency for nodes to be connected with others which are similar to themselves [23].

For homogeneity, we may usually focus on several properties:

- Connectivity, shows the likelihood of a connection between any nodes.
  - For an undirected graph, it is connected if there is at least a path between any pair of nodes.
  - For a directed graph, it is weakly connected if it is connected only disregarding the direction of links, and strongly connected if it is connected also when considering the direction of links.

- Average Shortest Path $< l >$, means average distance from one node to another.
- Diameter l_{max}, is the largest one within all the shortest paths between any 2 nodes.
- Clustering Coefficient, shows the probability that a node's two neighbours can have a link.

For example, Figure 2.1.1 is weakly connected, but not strongly. In directed graph, only the strongly connected one can have the average shortest path and the diameter. If we consider it as an undirected graph, $< l >= \frac{10+20+100+...+10+20+30}{6*5/2} = 80$ and $l_{max} = l_{C,F} = 140$. For node A, its clustering coefficient is $\frac{1}{3*2} = 0.16$.

2.1.3 Heterogeneous Statistics

Unlike homogeneity, which focuses on the common characteristics of the whole network, heterogeneity pays more attention to the differences between nodes, especially hubs, which can be considered as significantly large / important nodes [15].

For heterogeneity, we may use different centrality measurements to find hubs [25].

- Degree Centrality, ranks the nodes by their degrees.
- Closeness Centrality, measures how close a node is to the other nodes by the average path length from the given node to others.
- Betweenness Centrality, calculates how many shortest paths need to go through a given node.

2.1.4 Community Detection

Financial networks in the real world often have a large number of nodes and complex structures. Therefore, classification is often required before in-depth analysis, and community detection is a common method.

The most widespread method for community detection is Modularity Maximization, which selects the classification with maximum modularity value [17]. However, subsequent studies have demonstrated some of its flaws, one of which is that it is in fact merely descriptive, which means it only tries to describe the network, not to explain it. The difference between Description and Explanation can be seen in the example in Figure 2.1.2, what we really need is for the model to recognize a mountain on Mars, not a mountain that looks like a human face (i.e., over-fitting). To obtain the explanation, the non-parametric stochastic block models (SBMs) can be used, and have been shown to outperform many other community detection methods [33, 37].

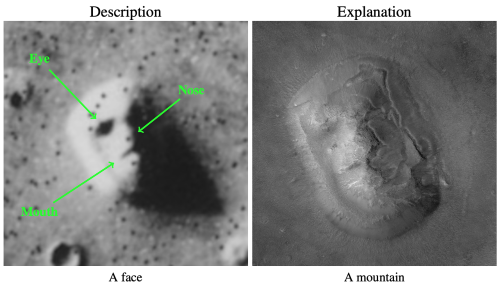


Figure 2.1.2: An example for description and explanation [33]

The idea of SBMs is a generative model based on the Bayesian formula [32].

To be specific, suppose a network is divided into B groups, and let $b_i \in [0, B-1]$ be the group index for node i. Then we define this partition as $\boldsymbol{b} = \{b_1, b_2, ..., b_i, ..., b_N\}$, and the probability that a model can generate a network A with this partition as $P(\boldsymbol{A}|\boldsymbol{b})$. Thus, according to the Bayesian formula, we can obtain the posterior probability:

$$P(\boldsymbol{b}|\boldsymbol{A}) = \frac{P(\boldsymbol{A}|\boldsymbol{b})P(\boldsymbol{b})}{P(\boldsymbol{A})} \tag{2.1}$$

Further, if we use $\boldsymbol{\theta}$ to represent additional model parameters that control how the node partition affects the structure of the network, the Equation 2.1 will transform like:

$$P(\boldsymbol{b}|\boldsymbol{A}) = \frac{P(\boldsymbol{A}|\boldsymbol{\theta}, \boldsymbol{b})P(\boldsymbol{\theta}, \boldsymbol{b})}{P(\boldsymbol{A})} \tag{2.2}$$

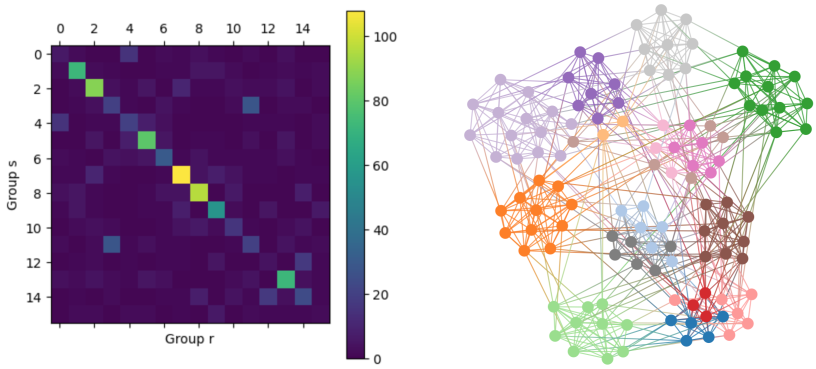
So far, the task of community detection is equivalent to finding a partition b that can maximize Equation 2.2.

In addition, we can transform the Equation 2.2 into the following form:

$$\begin{aligned} P(\boldsymbol{b}|\boldsymbol{A}) =& \frac{exp(-\Sigma)}{P(\boldsymbol{A})}, \\ \Sigma = -ln^{P(\boldsymbol{A}|\boldsymbol{\theta}, \boldsymbol{b})} &- ln^{P(\boldsymbol{\theta}, \boldsymbol{b})} \end{aligned} \tag{2.3}$$

In Equation 2.3, Σ is called the description length of the network A. Therefore, the task can also be considered to find a partition b that can obtain the minimum description length.

In the stochastic block model, which is arguably the simplest generative process [20], the additional parameters $\boldsymbol{\theta}$ can be considered as edge counts $\boldsymbol{e}$. That means, the model gives a partition of the nodes into groups $\boldsymbol{b}$ and a $B \times B$ matrix of edge counts $\boldsymbol{e}$, where e_{rs} is the number of edges between group r and s. Given these constraints, the edges are then placed randomly until it obtains the minimum description length. Figure 2.1.3 is an example of stochastic block model.



(a) Matrix of edge counts between groups (b) Generated network

Figure 2.1.3: An example for stochastic block model

Based on this model, there had been many further optimization ideas. For example, through the idea of recursion and hierarchy, the nested stochastic block model was proposed later [31].

2.1.5 Network Robustness

Usually, the robustness of a network can be tested by continuously deleting nodes in the network through a certain probability algorithm and observing some properties of the remaining nodes in the network [9].

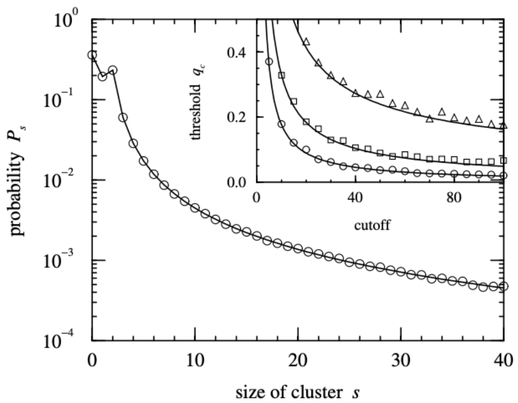


Figure 2.1.4: An example for network robustness analysis [9]

Figure 2.1.4 shows an example of network robustness analysis. The vertical axis represents the deletion of nodes according to the probability value, and the horizontal axis represents the cluster size value of the remaining network under this probability.

2.2 Related Network Analysis for Banking Exposures

The applications for cross-border banking exposures include static analysis and dynamic analysis. The former analyzes the network by calculating and interpreting different indicators. While the latter contains network diffusion simulation or the stress test.

2.2.1 Static Network Analysis

In the research explored by Masazumi and Yuko, they treated the webs of the cross-border bank exposures as networks which contained 16 countries and the time series between 1985 and 2006. After that, they investigated the characteristics of the network to draw a conclusion. Since the network had higher connectivity, a shorter average path length, a higher average degree and a higher clustering coefficient in 2006 than in the past, they concluded that the systemic risk in international financial markets was likely to increase [19]. Similarly, Camelia and Javier A did the research with more countries (184) and a longer time period (1978-2010) [29]. In general, their selection of indicators and interpretation of trends were helpful to this project. However, there could be one limitation that their analysis was not microscopic enough, such as a wide time range and too many nodes, which might be optimized by focusing on a short time period of specific economic events and community exploration.

2.2.2 Dynamic Network Analysis

For dynamic analysis, it mainly relies on the different strategies for simulation. One example of the simulation in the financial system is that first there is a default or liquidity shortage problem in one or several countries, and then it goes through a contagion mechanism, such as the issuing mortgage loans between associated banks, finally the risk spreads throughout the system [4]. By setting different contagion mechanisms, that is, whether affiliated banks choose to share risks and in what way, a large number of studies have proposed different analysis ideas [14, 10, 18, 13]. These studies used quite detailed strategies to simulate real scenarios. Thus, they could obtain more quantitative simulation effects. But they also increased the complexity of the entire project, which might reduce the interpretability.

Chapter 3

Methodology

This part will make a discussion and analysis to choose the proper research strategies, designs and methods for the project.

3.1 Reasoning Approaches

From a high level, among the three major reasoning approaches (i.e., deduction, induction and abduction) [34, 27], this project is more suitable for abduction and induction.

First of all, this project is more about the application of network science theory in cross-border banking business analysis scenarios. It does not involve the verification and development of original theories, so it is not suitable for deductive methods.

In general, starting from observation, the project analyzes the bank transaction network and explores its potential risks based on some existing network science theories. Therefore, it is a process of exploring the evidence and giving explanations, which is more in line with the principle of abductive approach.

In addition, the entire analysis process of the project may become a conceptual framework for a similar business, or draw some new insights, so it may also be applicable to the inductive method.

In summary, the overall flow of the project at the abstract level will be like:

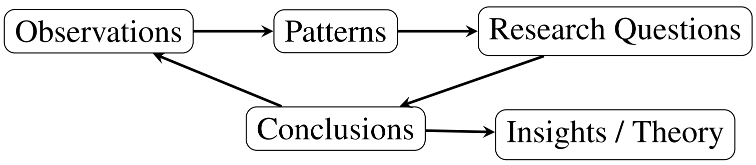


Figure 3.1.1: The flow of the project at the abstract level

3.2 Research Strategies

For research strategies [34], this project could be a mixed research, which collects quantitative data, but analyzes based on objective data / network and subjective perspective to make a qualitative conclusion.

3.3 Design and Methods

The overview for the flow of the project at the design level is showing in Figure 3.3.1.

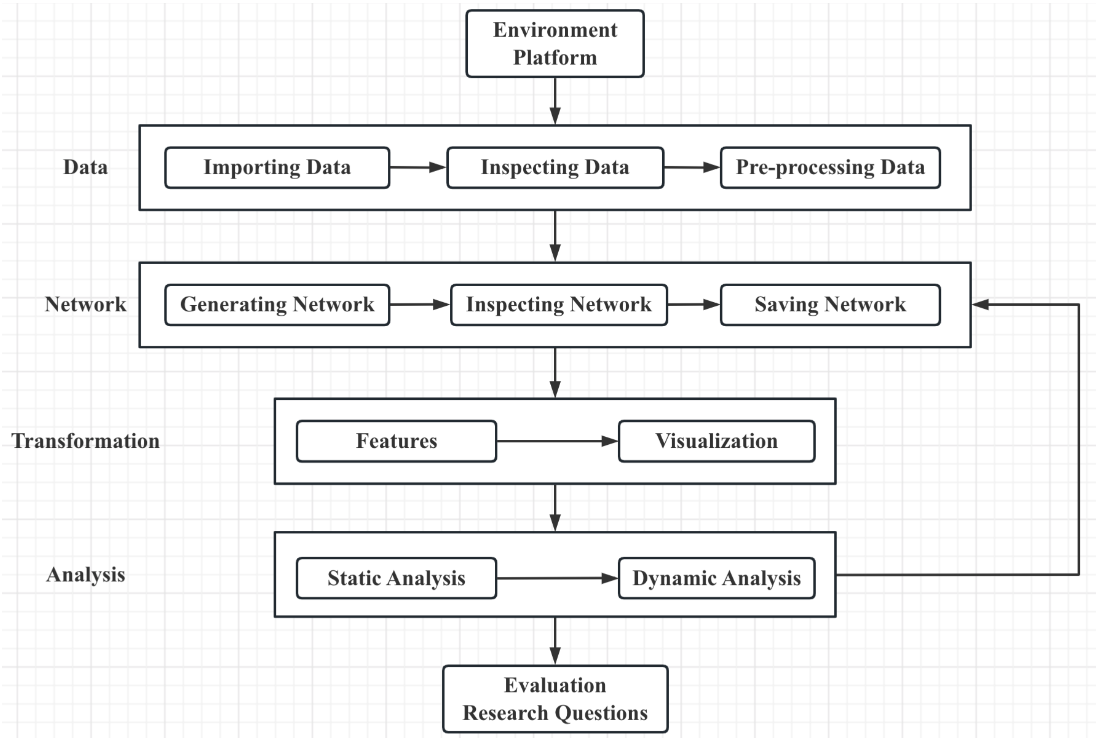


Figure 3.3.1: The flow of the project at the design level

3.3.1 Data and Pre-processing

The international banking data will come from project partner FNA.

According to the objectives the project pursues, The data should contain the basic information of the national bank, risk exposure or related characteristic data like the borrowing and lending amounts. In addition, the data needs to span a certain period of time for more detailed analysis like the evolution over time.

Before data pre-processing, we need to make an inspection, such as:

- Fields and meanings, data units, time spans.

- Data range, like mean, median, minimum and maximum values.
- The general trend of the data over time.

As to pre-processing [3], it may include:

- Removing outliers.
- Dealing with potential skewed distributions problem with log or $atan$ function.
- Normalization of data.

Many practices have proved that *Python*'s toolkits like *numpy* and *pandas* can handle data processing operations very well. In addition, *Jupyter Notebook* can provide a good experience for interactive programming and analysis, and its B / S (Browser / Server) architecture supports cross-platform online programming without encountering many basic environmental problems. Therefore, the project can run on the *Noteable* platform within the intranet of the school, so that on the one hand, the advantages of the above tools can be used, and on the other hand, for the concern of data security, it is also helpful for protecting the data privacy and reducing the risk of data theft.

3.3.2 Network Model

Among different types of networks, the temporal network can meet our needs, where each layer represents the data of a year, and within the same layer, it should be a directed and weighted graph, where nodes represent banks and links represent the borrowing and lending values between each bank [21].

Similarly, we will conduct a network inspection first. This may include:

- Displaying the basic distribution of the network through the drawing function.
- Calculating some basic properties, such as whether it is a strongly connected graph and the information for degrees.
- Calculating some homogeneous properties, like clustering statistics and paths.
- Exploring some heterogeneous properties, namely, the central nodes.
- Detecting the communities with stochastic block models.

Sometimes the dynamic generation of the network may be very time-consuming, so it is also necessary to store the network as a common file such as ".graphml" for direct importing next time.

Python has many network operation toolkits to choose from, such as *NetworkX*, *igraph* and *graph-tool*. In this project, we will mainly use *NetworkX* for static analysis and *graph-tool* for dynamic analysis because the former provides more comprehensive network operations, while *graph-tool* has stronger performance, better drawing ability and most critically, enough support for SBMs operations.

3.3.3 Features and Visualization

So far, we should have a basic network. But one of the biggest challenges is how to represent the abstract concept of risk exposure relationship between banks with concrete data. This involves the need to transform the data into features and display it.

Features

As mentioned above, there are actually many data indicators such as VaR to quantify the degree of risk. However, under this specific network, our target features need to meet the following conditions:

- The value can be truly comparable, namely, the higher (or lower) value means higher risk. For example, the borrowing and lending amount itself may not meet this requirement, because the high value does not necessarily mean that the risk is high since it ignores the country's own economic strength.
- Calculations are simple and interpretable. As it has been pointed out in the section on research strategies, the project would have some qualitative analysis. Therefore, reducing the complexity of features is conducive to improving the reliability of the entire project.

Specifically, in this project, the features mainly hope to quantify two kinds of risks: inherent risk and propagated risk [35]. For the former, we will use the ratio of borrowing to lending to estimate the ability of a country to repay its liabilities. For another, it may be related to the degree of centralization in the network, so we can use the PageRank value on the loan fraction of a lender to represent its ability of propagation.

Visualization

In addition to features, another important task in the pre-processing stage of the network is the visualization ability. In this project, we would mainly consider three ways:

- Node Filtering: When the amount of nodes is too large, the visualization of the network tends to ignore the details. Therefore, it is necessary to filter the nodes that do not care according to the requirements, so as to achieve a better visualization.
- Node Size: For the nodes that are more concerned, they are often the nodes with large values, so we can set the larger size related with the value to highlight the details.
- Node Color: Similar to node size, details can be highlighted through color contrast, and colors can often highlight multiple details such as small values, medium values, and large values at the same time.

Matplotlib provides strong drawing capabilities, and *NetworkX* and *graph-tool* also have built-in network drawing capabilities, so the current selection of technology can well support this task.

3.3.4 Analysis

In general, according to whether external shock simulation is performed, the project can be divided into static analysis and dynamic analysis.

Static Analysis

For static analysis, similar to the network inspection, we focus on a series of characteristics of the network itself, including the following indicators:

- Basic Statistics
- Homogeneous Statistics
- Heterogeneous Statistics
- Community Detection

On the basis of these data indicators, usually, the static analysis is then mainly divided into cross-sectional analysis and time series analysis. The former refers to selecting the data of one or more representative time points, and then drawing some conclusions by comparing the indicators of each node [30]; while the latter is to observe the trend of the indicators over time under the condition of fixing them, so as to analyze and summarize. However, in this project, there is also a comparison between the indicators after feature processing and the previous ones, which is also a way to verify whether feature processing is effective.

Dynamic Analysis

Dynamic analysis mainly studies the changes in the basic statistics and community structure of the remaining network after randomly deleting nodes and strategically deleting nodes. This can also be considered as repeated static analysis while performing external simulations.

3.3.5 Evaluation

As mentioned above, this is not a deductive project, and the research strategy of this project also includes subjective analysis, so it can be relatively difficult to conduct a quantitative evaluation.

One of the effective qualitative evaluation strategies is the expert scoring method, which evaluates the results by setting scoring items and calculating the average scores of multiple experts [24].

In order to simplify the task, this project adopts special treatment. In the evaluation part, we will use existing insights and network models to answer research questions. After that, we may summarize current limitations and point out future work.

Chapter 4

Implementation, Results and Analysis

4.1 Environment

According to the design, based on *Python* and *Noteable* platform, we will import *numpy* and *pandas* for data processing, *Matplotlib* for data visualization, *NetworkX* and *graph-tool* for network operation.

4.2 Data Inspection

First, after importing the data, we can get 27 files ranging from 20150331.csv to 20210930.csv. This means that the time span is from 2015/01/01 to 2021/09/30, and it is recorded every 3 months.

Fields

Then, we can focus on the first file (i.e., 20150331.csv) to study the columns / fields. The results are shown like:

net_id	arc_id	from_id	to_id	borrower	claims_held	claims_held_10_change	...	lender	share_of_claims_held	...
2015/3/31	AT-AU-0	AT	AU	Austria	5.220850e+08	NaN	...	Australia	0.005037	...
2015/3/31	AT-BE-0	AT	BE	Austria	1.750940e+09	NaN	...	Belgium	0.016893	...
...	...	...	...	...	...	...	...	...	...	...

Table 4.1: The sample for data fields

The shape of data in this file is $(406, 21)$, which means there are 406 records and 21 fields. The unique number of borrowers is 21, and that of lenders is 22. Furthermore, the data is not large (about $406 \times 21 \times 27$), but there are many original features. In order to simplify the problem, we only care about and explain some of the fields:

- "net_id": The value is the same as the date shown in the file name, so we can use the file name instead.
- "arc_id", "from_id", "to_id": They mark the direction of the borrowing and lending relationship, and the "id" is the abbreviation of the country.
- "borrower", "lender": Corresponding to "from_id" and "to_id", they are the full names of the countries.
- "claims_held": It is all the borrowing which hasn't been repaid at that particular point in time. In addition, the unit is USD and the values are kept at a large magnitude.

Fortunately, the values of these fields we are concerned about are not missing or abnormal, so we don't need to do special processing.

Merged Information

Next, we use the filtered columns (i.e., "arc_id", "from_id", "to_id", "borrower", "lender", "claims_held") and merge all the files together, then we get the following table:

arc_id	from_id	to_id	borrower	lender	20150331	20150630	...	20210930
AT-AU-0	AT	AU	Austria	Australia	5.220850e+08	5.084990e+08	...	4.686770e+08
AT-BE-0	AT	BE	Austria	Belgium	1.750940e+09	1.849885e+09	...	2.712196e+09
...	...	...	...	...	...	...	...	...

Table 4.2: The sample for merged information

Now the shape of data is $(417, 32)$. Similarly, the unique number of borrowers is still 21, and that of lenders is 22. The difference in the "arc_id" may indicate that at different times, the total links between countries may be different.

Through the *describe()* function of *pandas*, we can get the indicators of the data range:

	count	mean	std	min	25%	50%	75%	max
20150331	406	3.896289e+10	1.150953e+11	3.400000e+04	4.332500e+08	3.905123e+09	2.505605e+10	1.402600e+12
20150630	404	3.901076e+10	1.145297e+11	3.500000e+04	4.781960e+08	4.170236e+09	2.512682e+10	1.367110e+12
...	...	...	...	...	...	...	...	...
20210930	407	4.882246e+10	1.614671e+11	4.000000e+03	5.060000e+08	4.564000e+09	3.294450e+10	2.062660e+12

Table 4.3: The description of merged information

In the table 4.3, the mean and the median show that the average magnitude of the data is around 10^9 and 10^{10}, which means that we can use this as a unit to reduce the absolute value of the data. In addition, the standard deviation is relatively large, and the data magnitudes of the minimum and maximum values are also very different, which means that we may need to filter the data or use the *log()* function to deal with skewed distributions.

Time-trend graphs

Finally, we select the mean, median and max values to draw time-trend graphs, which are shown in the following figure:

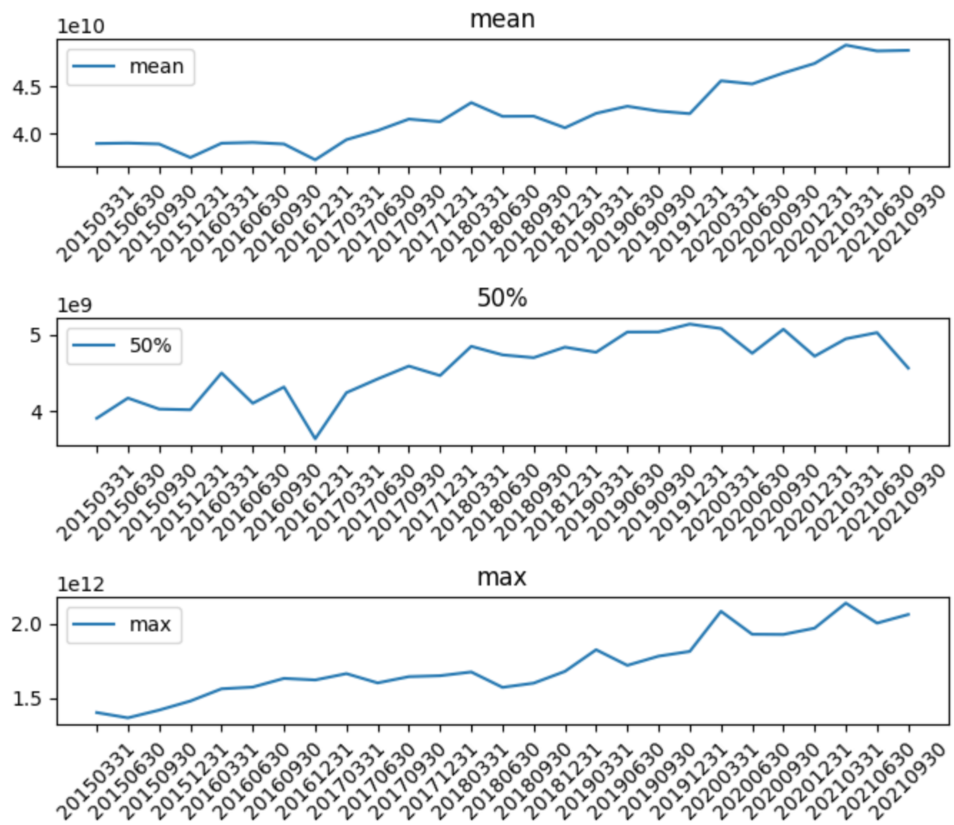


Figure 4.2.1: The time-trend graphs for data inspection

On the whole, the figure shows an upward trend, which means that the external debts of all countries are increasing. However, the current data alone is not enough to judge whether it is a natural increase in economic development or it represents an increase in potential risks. When we take a closer look, the mean starts to show an obvious upward trend in early 2020, while the median and maximum values remained almost at the peak. All three figures are partially declining in 2021. In fact, this is the range of COVID-19, so our future research can pay special attention to this stage.

4.3 Network Generation and Inspection

4.3.1 Network Generation

Based on the information obtained from the data inspection, we use the following strategies to generate the networks:

- We generate an array of 27 graphs.
- For each directed and weighted graph, nodes represent different countries, and the link represents there is a debt between two countries with the direction from the lender to the borrower. As for the weight, we divide the original ”claims_held” value by 10^9 to facilitate subsequent analysis and calculation, namely, the unit of weight is 1 billion dollars. Unless otherwise specified, this is the default weight.
- For subsequent needs, here we also add the reciprocal of ”claims_held” as another weight of the edge.

Next, we select the first network (i.e., 2015/03/31) for a preliminary exploration:

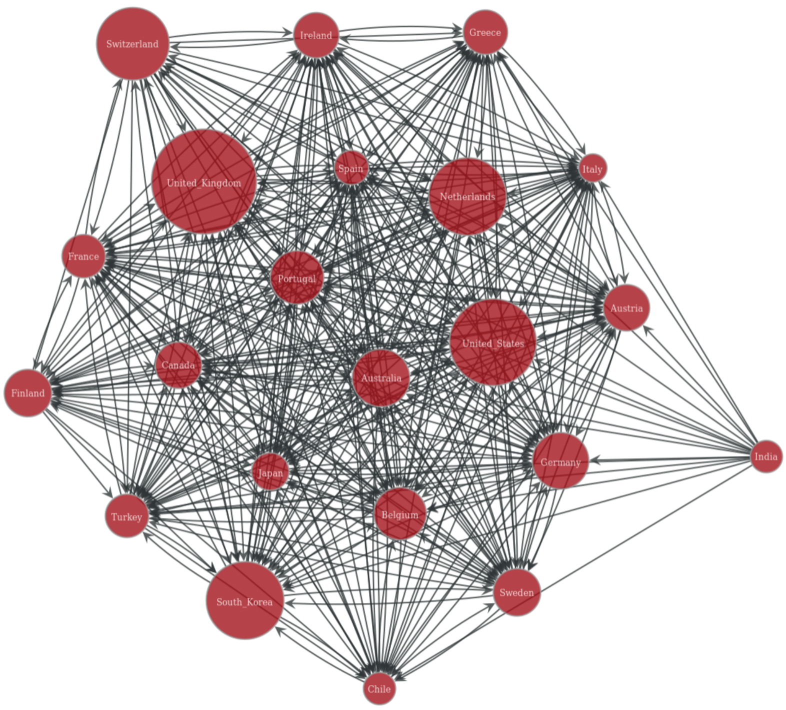


Figure 4.3.1: The sample network

The number of nodes in this graph is 22 and that of edges is 406. This can also be reflected in Figure 4.3.1, the density of the graph is relatively large. As to connectivity, it is a weakly connected network but not a strongly connected one. When we continue to explore, we will find that there are actually 2 strongly connected components, one of which is a single node, India. In Figure 4.3.1, we can indeed find that India only has outward arrows, but no incoming arrows, which means that India only lends money to other countries, but does not introduce funds.

4.3.2 Network Inspection by Basic and Homogeneous Statistics

To have a deep inspection, we will continue to calculate some properties, like basic statistics and homogeneous statistics for all networks and obtain the description of the statistics.

	count	mean	std	min	25%	50%	75%	max
nodes	27.0	22.000000	0.000000	22.000000	22.000000	22.000000	22.000000	22.000000
links	27.0	404.000000	2.935198	399.000000	402.000000	404.000000	406.500000	409.000000
density	27.0	0.874459	0.006353	0.863636	0.870130	0.874459	0.879870	0.885281
avg_in_degree	27.0	18.363636	0.133418	18.136364	18.272727	18.363636	18.477273	18.590909
avg_out_degree	27.0	18.363636	0.133418	18.136364	18.272727	18.363636	18.477273	18.590909
avg_total_degree	27.0	36.727273	0.266836	36.272727	36.545455	36.727273	36.954545	37.181818
avg_in_strength	27.0	775.640146	69.507916	680.491755	716.108719	766.387224	814.280785	909.058090
avg_out_strength	27.0	775.640146	69.507916	680.491755	716.108719	766.387224	814.280785	909.058090
avg_total_strength	27.0	1551.280292	139.015833	1360.983510	1432.217438	1532.774448	1628.561570	1818.116180
max_in_strength	27.0	2837.377944	308.274066	2326.251300	2648.198600	2736.680000	3068.365750	3437.434000
max_out_strength	27.0	6045.563033	654.391205	5164.569164	5484.810986	5878.186743	6505.231790	7294.908210
max_total_strength	27.0	8161.270737	874.440626	6936.898164	7421.274986	7994.035743	8708.479297	9908.429210
avg_clus_coef	27.0	0.917879	0.003762	0.910329	0.915375	0.918491	0.920792	0.922678
avg_clus_coef_weight	27.0	0.009799	0.000672	0.008604	0.009299	0.009724	0.010014	0.011400
avg_path_length_for_strong	27.0	1.075397	0.005485	1.066667	1.070238	1.076190	1.078571	1.085714
avg_path_length_weight_for_strong	27.0	1.310632	0.086950	1.145918	1.268590	1.307801	1.351592	1.525382
diameter_for_strong	27.0	2.000000	0.000000	2.000000	2.000000	2.000000	2.000000	2.000000
diameter_weight_for_strong	27.0	12.194128	1.782026	9.102623	10.907103	12.290150	12.891098	16.712108

Figure 4.3.2: The description of properties of all networks

In Figure 4.3.2, we can try to make an understanding.

- All networks are very dense, which is consistent with our observations.
- Considering the degree, the average in-degree and the average out-degree are equal to about 18, which is actually reasonable, because overall, the in-degree of a node is actually the out-degree of another node, and the data value may mean that on average, each country has a loan relationship with 18 other countries.

- Since they are also weighted graphs, so we will take a look at the strength. The maximum out-strength is twice as large as the maximum in-strength, indicating that overall, the lending amount is greater than the borrowing amount, which may indicate that the risk of default is controllable.

- The network has a high clustering coefficient, that is, there are trading links between almost any three countries. And this can also be seen in the shortest path and diameter data of the largest connected component (i.e., according to the previous analysis, that is, the network after removing the node Indian). The shortest path is 1, which means that almost any two countries have transactions, and the diameter is 2, which indicates that any three countries should have a transaction link.

Further, we can analyze from the perspective of time series by drawing line charts.

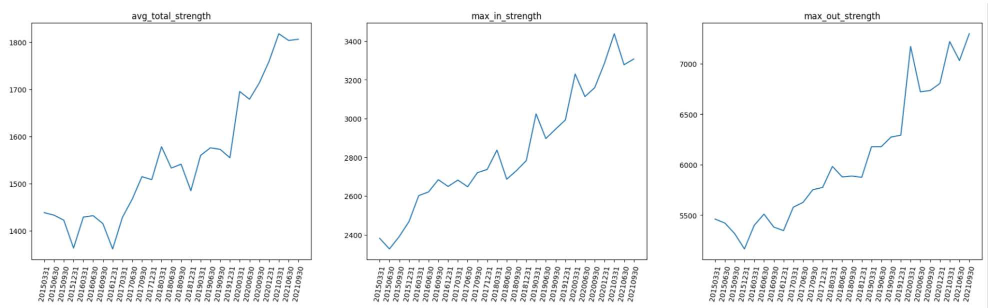


Figure 4.3.3: The time-trend graphs for network inspection

From Figure 4.3.3, the average strength, the maximum in-strength and the maximum out-strength all show an upward trend. This shows that, similar to the conclusion of the previous data inspection, the country's overall borrowing and lending amounts have increased, but there is no way to intuitively analyze whether it indicates an increase in risk, because the natural growth of the economy will also bring some inflation.

4.3.3 Network Inspection by Heterogeneous Statistics

For heterogeneous statistics, we focus on the difference between nodes / edges. Therefore, firstly we try to filter the edges according to the weight in the network of 2015/03/31.

```
Filter loan amount >= 1e+00:
    270 of 406 edges remain.
```

```
    Density: 0.584
    Weakly connected: True
Filter loan amount >= 1e+01:
    160 of 406 edges remain.
    Density: 0.346
    Weakly connected: True
Filter loan amount >= 1e+02:
    40 of 406 edges remain.
    Density: 0.0866
    Weakly connected: False
Filter loan amount >= 1e+03:
    1 of 406 edges remain.
    Density: 0.00216
    Weakly connected: False
```

We use 1, 10, 100, and 1000 as thresholds, and find that even with a limit of 1 billion, nearly half of the edges were filtered out. When the limit was increased to 100, the graph was no longer weakly connected. And only 1 edge's weight is greater than 1000.

This shows the existence of heterogeneity in the network, so we consider simply using the histogram to observe the distribution of node strength.

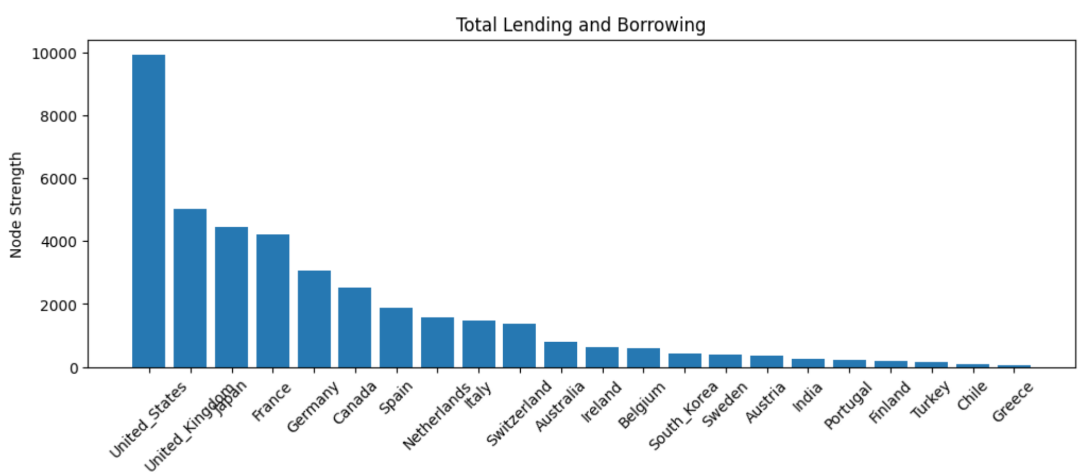


Figure 4.3.4: The distribution for total lending and borrowing amounts

As shown in the figure, the distribution reflects an obvious long-tail effect [5], and the amount in the United States is much higher than that in other countries. Next, we will calculate the strength centrality, closeness centrality and betweenness centrality separately and obtain the top 10 countries. It should be noted that we are relatively more concerned about nodes with relatively large absolute values (although as mentioned before, absolute values and risks may not necessarily have a positive relationship), so strength centrality will continue to use ”claims_held” as the weight, while closeness and betweenness will use the reciprocal of ”claims_held” as the weight, because the core of these two algorithms is to calculate the shortest path, and we hope that the node with a larger value of ”claims_held” appears on the shortest path.

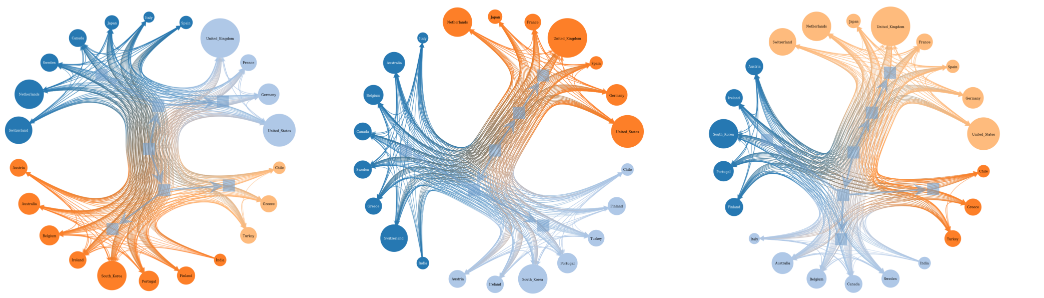
	0	1	2	3	4	5	6	7	8	9
In–strength	Japan	United_Kingdom	France	United_States	Germany	Switzerland	Canada	Spain	Netherlands	Italy
Out–strength	United_States	United_Kingdom	Germany	France	Japan	Italy	Netherlands	Spain	Australia	Canada
Total–strength	United_States	United_Kingdom	France	Japan	Germany	Switzerland	Canada	Spain	Netherlands	Italy
Closeness	France	Germany	United_States	Japan	United_Kingdom	Canada	Switzerland	Spain	Sweden	Netherlands
Betweenness	United_States	United_Kingdom	France	Germany	Spain	Sweden	Netherlands	Italy	Australia	Austria

Figure 4.3.5: The top 10 central nodes

The result is shown in Figure 4.3.5. Combined with Figure 4.3.4, the total amount of borrowing and lending in the United States is greater than that of all other countries, and it also appears in the top five in all centrality data, so it is a well-deserved hub. The United Kingdom, France, Germany and Japan followed closely behind. What needs special attention is that Japan ranks first in the in-strength centrality, indicating that it has a large amount of borrowing, and it may not maintain a reasonable ratio to the amount of lending. In closeness, France and Germany surpassed the United States. Closeness describes the overall short distance between this node and other surrounding nodes. In this example, it shows that the transaction volume between these two countries and other countries is generally large. In short, these countries deserve special attention in the subsequent analysis.

4.3.4 Network Inspection by Community Detection

We can use *graph-tool* for community detection, or simply put, to classify nodes. The simplest function it provides to implement the stochastic block model is *minimize_blockmodel_dl()*. Here we use an optimized version, *minimize_nested_blockmodel_dl()*, to obtain multi-level community recognition effects. In addition, when we explore the community, we need to add weight factors (i.e., ”claims_held”), because it can be seen from the previous analysis that each edge is not equivalent, and in fact, there is a lot of difference. When we run it many times, we may get the following distribution:



(a) community detection result 1 (b) community detection result 2 (c) community detection result 3

Figure 4.3.6: The unstable results of community detection

Obviously, there is a problem that the result is not stable. This may be due to the fact that there are alternative partitions with similar probabilities, or that the optimum is difficult to find. To solve this problem, we can use the idea of obtaining the minimum description length in Equation 2.3. In practice, we can run it multiple times and get the description length each time through the *entropy()* function. Finally we select the best classification.

After running 10 times, we can obtain the minimum description length of 1941 (and the maximum is 2051), and draw the distribution shown in the figure below:

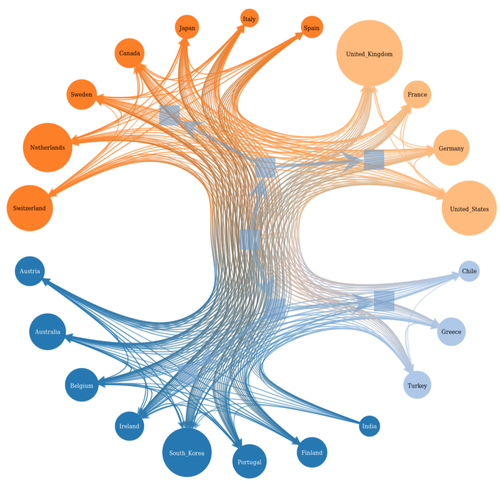


Figure 4.3.7: The stable result of community detection

In addition, we can get the summary for this partition:

```
<BlockState object with 22 blocks (4 nonempty), degree-corrected,
    with 1 edge covariate, for graph <Graph object, directed,
    with 22 vertices and 406 edges, 1 internal vertex property, 3
```

```
    internal edge properties>>
<BlockState object with 4 blocks (2 nonempty), with 1 edge
    covariate, for graph <Graph object, directed, with 22
    vertices and 16 edges, 2 internal vertex properties, 1
    internal edge property>>
<BlockState object with 2 blocks (1 nonempty), with 1 edge
    covariate, for graph <Graph object, directed, with 4 vertices
    and 4 edges, 2 internal vertex properties, 1 internal edge
    property>>
<BlockState object with 1 blocks (1 nonempty), with 1 edge
    covariate, for graph <Graph object, directed, with 2 vertices
    and 1 edge, 2 internal vertex properties, 1 internal edge
    property>>
```

In this partition, the total number of levels is 4. Excluding the 22 blocks of the first layer, in general, the network is divided into 4 main communities. We can find that the United Kingdom, France, Germany and the United States are in the same community, in consistent with what we find in the centrality analysis. In addition, Chile, Greece and Turkey are in the same small communities, which can be reasonable since in Figure 4.3.4, they are all in the last position.

4.3.5 Summary

So far, before the next feature processing, based on current data and networks, we can give a summary of our insights:

- **Insight 1**: The data range from 2015/01/01 to 2021/09/30, and they are recorded every 3 months. Within this range, the borrowing and lending amounts generally keep increasing, and special attention is required for the period of COVID-19 (from 2020 to 2021).

- **Insight 2**: The banking network is quite dense since the numbers of nodes and edges are 22 and around 400 separately. India is a special node which only has outward arrows (i.e., only lends money to others). In addition, when excluding India, for any 3 countries, there is at least 1 transaction path that can link to each other.

- **Insight 3**: In the network of 2015/03/31, there are significant differences in the size of transactions between countries, and there is an obvious long tail in this distribution. Japan is a special central node which has the highest in-strength value. Besides, the United States, the United Kingdom, France and Germany are the central nodes of the network, and they are also in the same community, one of the 4 communities we find so far.

As mentioned before, all the above conclusions are based on the unprocessed borrowing and lending amount, which is not equivalent to the level of risk. For example, it is quite reasonable for countries with more developed economies to have higher borrowing or lending amounts.

4.4 Feature: Inherent and Propagated Risks

Now, we need to perform feature processing, aiming at finding a reasonable indicator of a country's exposure risk.

We suppose that there are mainly 2 things to consider when estimating the risks in these networks. First is each country's own risk of default, namely, inherent risk. Second is the propagated risk, which means the risk of a country's debtors' defaulting. In addition, from the **Insight 3** in network inspection, we know that there are significant differences in the size of transactions, so we can normalize it before feature transformation.

4.4.1 Normalization

To normalize the edge weight, we will use the total lending amount of the node as the denominator to obtain the "claims_held_norm". For example, if a node has 3 outward arrows with "claims_held" value $[1,2,3]$, then the "claims_held_norm" value should be $[\frac{1}{6},\frac{2}{6},\frac{3}{6}]$. Similarly, this time we also store the reciprocal value for later calculation. The core pseudocode for this part of the logic should be like:

```
in_strength = G.in_degree(weight = 'claims_held')
out_strength = G.out_degree(weight = 'claims_held')
# add norm
for src, tar, wgt in G.edges.data('claims_held'):
    G.edges[src, tar]['claims_held_norm'] = wgt/out_strength[src]
    G.edges[src, tar]['claims_held_norm_reciprocal'] =
        1/G.edges[src, tar]['claims_held_norm']
```

4.4.2 Inherent Risk

Generally speaking, we believe that a country's own ability to prevent default is related to its own economic strength. So, in the absence of each country's GDP data, we try to estimate the inherent risk by using the ratio of borrowing to lending. As a result, we add an attribute "inherent_risk" to the node. The core pseudocode for this part of the logic should be like:

```
# add inherent risk
for node in list(G.nodes(data=False)):
    G.nodes[node]['inherent_risk'] =
        in_strength[node]/out_strength[node]
```

4.4.3 Propagated Risk

To measure the propagated risk, we have to consider the node and its edges. That means, if a country lends a lot of money to others, the default of the debtor countries may reduce the country's ability to solve its debts, and the degree of this weakening is also related to the size of the country's own inherent risks. Therefore, similar to the idea of the PageRank algorithm, we can use the number of links connected to this node to measure its importance or propagated risk. Besides, in the algorithm, inherent risk can be used as the initial state of each country, and "claims_held_norm" can represent the normalized importance of each edge. The core pseudocode for this part of the logic should be like:

```
# add propagated risk
propagated_risk = nx.pagerank(
    # The lender has higher risk, so change the arrow direction
    G = G.reverse(),
    nstart = dict(G.nodes.data('inherent_risk')),
    weight = 'claims_held_norm')
for node in list(G.nodes(data=False)):
    G.nodes[node]['propagated_risk'] = propagated_risk[node]
```

4.4.4 Summary and Visualization

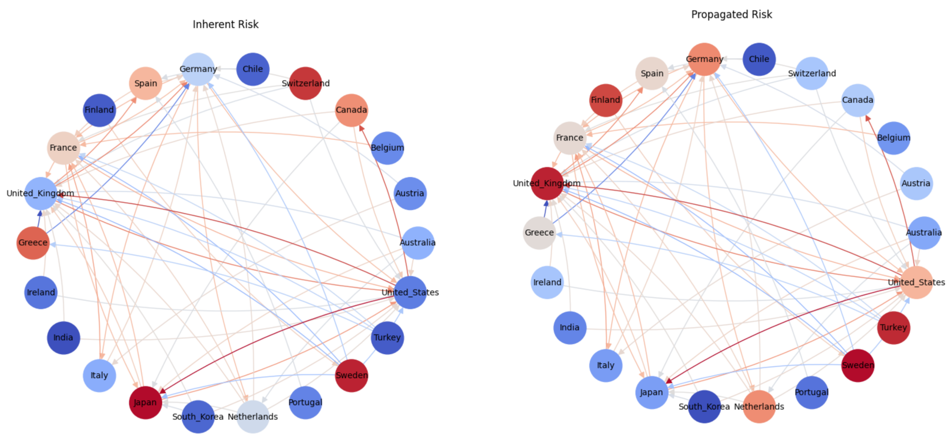
After feature processing, there are now 2 node attributes (”inherent_risk”, ”propagated_risk”) and 4 edge attributes (”claims_held”, ”claims_held_reciprocal”, ”claims_held_norm”, ”claims_held_norm_reciprocal”) in a network.

For visualization, we can use disparity filtering to filter edges rather than a single threshold. Then, use the color from blue to red to represent the increasing value for log(”claims_held”), ”inherent_risk” and ”propagated_risk”.

The core pseudocode for disparity filtering should be like:

```
def disparity_filter(G, alpha):
    s = G.out_degree(weight = 'claims_held')
    k = G.out_degree(weight = None)
    def pij(i, j):
        w = G.edges[i, j]['claims_held']
        return ((1 - (w / s[i])) ** (k[i] - 1)) < alpha
    return nx.subgraph_view(G, filter_edge = pij)
```

After filtering, there are only 60 edges left. And the result of the visualization is shown in the below figure:



(a) Sample for inherent risk (b) Sample for propagated risk

Figure 4.4.1: The sample for inherent and propagated risks

Next, we will explore the new network through static analysis and dynamic analysis.

4.5 Static Analysis

4.5.1 Different insights from previous networks

Compared with previous networks, the only difference is the weight of nodes and edges. Therefore, the new network only needs to recalculated in some parts, with the related analysis. For the analysis of a single network, we would still use the network of 2015/03/31 as a sample.

Distribution for total normalized lending and borrowing amount

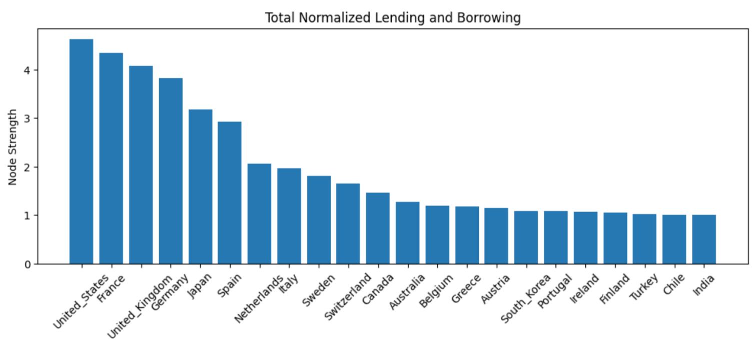


Figure 4.5.1: The distribution for total normalized lending and borrowing amount

Comparing Figure 4.5.1 and Figure 4.3.4, we will find that the ranking of countries is almost the same, but the differences between countries have been significantly reduced, especially that the United States is no longer much higher than other countries, and the long tail phenomenon has also improved. This is because, after normalization, all countries' total lending is 100%. Besides, normalized borrowing can be considered as this country's aggregated importance for its creditor countries. In other words, if a high-ranking country has high inherent risks, it will have a stronger propagating effect. For example, combined with Figure 4.4.1, Japan has relatively high normalized borrowing value, and its inherent risk is relatively high, so the United States, the Netherlands, and Sweden that lend money to it are affected to varying degrees, resulting in increasing the propagated risk. Although Switzerland also has a relatively high inherent risk, compared to Japan, the impact on the entire economic system is limited.

Central nodes

Similar to before, we use "claims_held_norm" for strength centrality, but for closeness and betweenness, we will use "claims_held_norm_reciprocal".

	0	1	2	3	4	5	6	7	8	9
In–strength	United_States	France	United_Kingdom	Germany	Japan	Spain	Netherlands	Italy	Sweden	Switzerland
Out–strength	Spain	Australia	Belgium	Switzerland	Germany	France	Finland	United_Kingdom	India	South_Korea
Total–strength	United_States	France	United_Kingdom	Germany	Japan	Spain	Netherlands	Italy	Sweden	Switzerland
Closeness	United_States	France	United_Kingdom	Germany	Japan	Spain	Canada	Netherlands	Italy	Switzerland
Betweenness	United_Kingdom	Germany	France	United_States	Spain	Netherlands	Sweden	Japan	Italy	Greece

Figure 4.5.2: The top 10 central nodes

According to Figure 4.5.2, it can be found that the ranking is not much different from before. Comparing it with the nodes in Figure 4.4.1, central nodes usually have high propagated risks. This also shows that a country with a large absolute value of transaction volume usually has a large risk of being propagated.

Communities

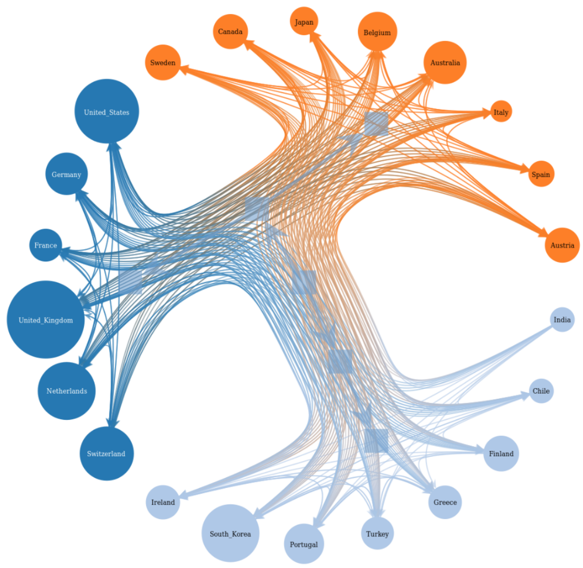


Figure 4.5.3: The result of community detection

Compared with the previous results, there are still some differences in the results of the community. The community where the United Kingdom is located has added the Netherlands and Switzerland now. In Figure 4.4.1, if we look at the inherent or propagated risks alone, these nodes are not prominent at the same time, but if we consider the two together, it can be approximately considered to be similar to the distribution of community exploration. So the current community can represent the classification of the financial system under the comprehensive consideration of countries' inherent risks and propagated risks.

Summary

To sum up, there are several different insights:

- After feature processing, **Insight 1** and **Insight 2** are still retained, while **Insight 3** has some changes.
- **Insight 4**: After feature processing, the normalized borrowing can be considered as this country's aggregated importance for its creditor countries. In the network of 2015/03/31, from the node strength distribution and central node ranking, countries with high importance (i.e., United States, United Kingdom, France, Germany, Spain...) usually have high propagated risks. Also, if a country with high importance has a high inherent risk (i.e., Japan), it can cause quite large effects on the whole financial system. In addition, a country with a large absolute value of transaction volume usually has a large risk of propagation.
- **Insight 5**: After feature processing, in the network of 2015/03/31, the number of communities decreases from 4 to 3. Under the comprehensive consideration of countries' own risks and propagated risks, the community where the United States, the United Kingdom, France and Germany are located has added Netherlands and Switzerland.

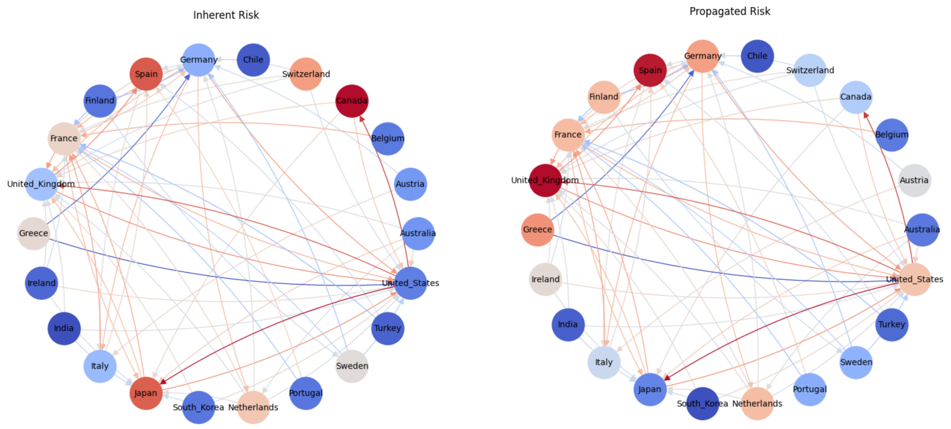
4.5.2 Cross-Sectional Analysis

In this part, we will select a network at another point in time, and focus on the inherent and propagated risks. From **Insight 1**, we know that the period of COVID-19 is a special time span, so we select the network of 2021/03/31 for analysis.

After disparity filtering, there are only 64 edges left. Then, we calculate the top 10 risk nodes and visualize inherent and propagated risks in the following figures:

	0	1	2	3	4	5	6	7	8	9
inherent_risk	Canada	Spain	Japan	Switzerland	Netherlands	France	Greece	Sweden	United_Kingdom	Italy
propagated	United_Kingdom	Spain	Greece	Germany	France	Finland	Netherlands	United_States	Ireland	Austria

Figure 4.5.4: The top 10 risk nodes of 2021/03/31



(a) Inherent risk of 2021/03/31 (b) Propagated risk of 2021/03/31

Figure 4.5.5: The inherent and propagated risks of 2021/03/31

First, we try to do a statistical summary of the risk (i.e., red/orange) nodes:

- There are 5 inherent risk nodes: Canada, Spain, Japan, Switzerland and Netherlands.
- There are 8 propagated risk nodes: the United kingdom, Spain, Greece, Germany, France, Finland, Netherlands and the United States.

Second, we can explain the connection between inherent and propagated risks. For example, Spain has a high inherent risk while this value of the United Kingdom is quite low, but the United Kingdom has a quite high investment on it because the color of its edges to Spain is quite warm, so when we explore the propagated risk, the United Kingdom can have quite a large value.

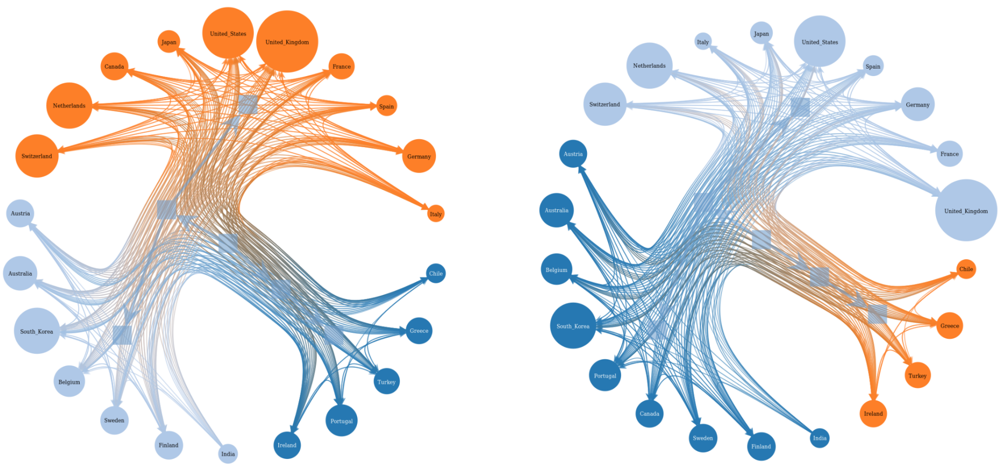
Finally, we can also use this theory to help banks to optimize cross-border investment decisions. For example, Portugal will need to pay special attention to the economic situation of Spain, because almost 70% of its investment is in Spain, and Spain itself has high internal risks. Therefore, we will find that although Portugal itself has a low risk, it has a higher propagated risk. Although the color of Portugal is not the most obvious in the map of propagated risks, this is partly due to the long-tail effect where countries like UK, France and Germany, as major economic participants, have a very high propagated risk value.

To sum up, we now have a new insight:

- **Insight 6**: Based on the inherent and propagated risk networks, combined with the analysis of network connection information, we can obtain more insights to help banks make investment decisions.

4.5.3 Time Series Analysis

In the previous section, we selected the network during COVID-19 for analysis. Next, we will select one before the epidemic (i.e., 2019/03/31, we choose the same quarter to reduce the impact of differences between quarters), and then compare them to gain new insights. This time, we mainly want to know the changes in the communities.



(a) Communities of 2019/03/31 (b) Communities of 2021/03/31

Figure 4.5.6: The communities before and within the COVID-19

In Figure 4.5.6, it can be seen that the overall layers and structure of the community have not changed, but the composition of each community has changed. We can use the table to show the comparison:

	same	2019/03/31	2021/03/31
Community 1	United_States, United_Kingdom, France, Spain, Germany, Italy, Japan, Netherlands, Switzerland	Canada	-
Community 2	Austria, Australia, South_Korea, Belgium, Sweden, Finland, India	-	Portugal, Canada
Community 3	Ireland, Turkey, Greece, Chile	Portugal	-

Table 4.4: The comparison for two communities

So we can see the main difference is between Portugal and Canada. More specifically, Canada has been relegated from Community 1 to Community 2 in 2021, while Portugal has been promoted from Community 3 to Community 2. Next, we will analyze the investment information of their debtor countries and creditor countries to try to explain this phenomenon.

For Canada, its related information is shown in the figure below:

	out country	20190331 claims_held	20210331 claims_held	20190331 claims_held_norm	20210331 claims_held_norm	claims_diff
0	Austria	1.326616	1.791729	0.002989	0.003174	0.465113
1	Australia	11.312021	13.157539	0.025485	0.023306	1.845518
2	Belgium	3.325000	4.830000	0.007491	0.008555	1.505000
3	Switzerland	23.986149	29.811022	0.054038	0.052805	5.824873
4	Chile	0.059385	0.029754	0.000134	0.000053	−0.029631
5	Germany	34.167000	38.868000	0.076975	0.068847	4.701000
6	Spain	4.715695	5.201965	0.010624	0.009214	0.486270
7	Finland	3.066000	3.532000	0.006907	0.006256	0.466000
8	France	30.368000	36.636000	0.068416	0.064894	6.268000
9	United_Kingdom	99.638000	127.598000	0.224474	0.226015	27.960000
10	Greece	0.048000	0.020000	0.000108	0.000035	−0.028000
11	Ireland	1.625071	2.011818	0.003661	0.003564	0.386747
12	Italy	2.479197	2.956778	0.005585	0.005237	0.477581
13	Japan	76.283200	101.892000	0.171858	0.180482	25.608800
14	South_Korea	3.016028	4.496539	0.006795	0.007965	1.480511
15	Netherlands	14.591781	18.667296	0.032874	0.033066	4.075515
16	Portugal	0.124766	0.132552	0.000281	0.000235	0.007786
17	Sweden	2.432880	0.792652	0.005481	0.001404	−1.640228
18	Turkey	0.014394	0.031848	0.000032	0.000056	0.017454
19	United_States	131.295000	172.097000	0.295793	0.304837	40.802000

(a) Information for Canada's lending

	in country	20190331 claims_held	20210331 claims_held	20190331 claims_held_norm	20210331 claims_held_norm	claims_diff
0	Australia	21.147346	22.575065	0.048379	0.041820	1.427719
1	Austria	1.919362	2.327722	0.010432	0.010121	0.408360
2	Belgium	2.263015	1.546472	0.005943	0.003430	−0.716543
3	Switzerland	8.984105	15.117394	0.024047	0.036245	6.133289
4	Germany	32.363637	25.810481	0.023281	0.014946	−6.553156
5	Spain	1.732413	2.006919	0.004242	0.004009	0.274506
6	Finland	2.367553	2.398883	0.017422	0.015812	0.031330
7	France	25.279107	32.708234	0.020607	0.021706	7.429127
8	United_Kingdom	126.916000	182.298000	0.056010	0.072208	55.382000
9	Ireland	30.934869	39.829740	0.071160	0.073940	8.894871
10	Italy	1.574582	0.495262	0.002303	0.000637	−1.079320
11	Japan	40.974525	48.412514	0.039447	0.040436	7.437989
12	South_Korea	1.209863	1.195375	0.004339	0.003598	−0.014488
13	Netherlands	8.688485	8.950130	0.017713	0.017262	0.261645
14	Sweden	7.335088	6.347792	0.054644	0.043315	−0.987296
15	United_States	1168.040000	1459.400000	0.189107	0.202203	291.360000
16	Greece	NaN	0.000472	NaN	0.000027	NaN

(b) Information for Canada's borrowing

Figure 4.5.7: The information for Canada's lending and borrowing

From 2019 to 2021, Canada has increased its lending to the UK, Japan, and the US, and has also increased its borrowings from the UK and the US, especially the US, which is close to 300 billion, accounting for 20% of the US loans. Therefore, in general, Canada is a net inflow of funds. Similarly, when we carry out the corresponding operation on Portugal, we will find that its main capital change is the increase of lending to Spain, and finally, nearly 70% of the funds are invested in Spain.

In summary, we can speculate the following new insights:

- **Insight 7**: The detection of communities, in a physical sense, is the result of comprehensive consideration of the country's economic volume, its inherent risks and propagated risks. When a large proportion of capital flows into the country, it may cause it to drop to other communities (i.e., low total risks); conversely, if a large proportion of capital flows out, it may rise to other communities (i.e., high total risks).

4.6 Dynamic Analysis

In the dynamic analysis, we will select the network of 2021/03/31 for processing and compare it with the previous results. In the simulation strategy, we adopt the following scheme:

1. Selecting nodes to delete: in the previous analysis, we know Canada, Spain, Japan, Switzerland and Netherlands have high inherent risks, so we will delete these 5 nodes.

2. Recalculating attributions: we assume that after a country is deleted, the rest of the countries will lose their lending to this country, but still retain borrowings from this country. That is, ”in_strength” will use the data of the original network, while ”out_strength” will use the data of the current network. Then we recalculates ”claims_held_norm”, ”inherent_risk”, ”propagated_risk” and other values.

3. Displaying: we consider using tables to show the comparison of relevant data before and after deleting nodes. Besides, we use network diagrams to more conveniently display the current network distribution.

4. Detecting communities: similar to before, we will re-perform the community detection and compare the previous results for analysis.

5. Iterating: we will try more node selection and data recalculation strategies, and continue to repeat the above steps to get more insights.

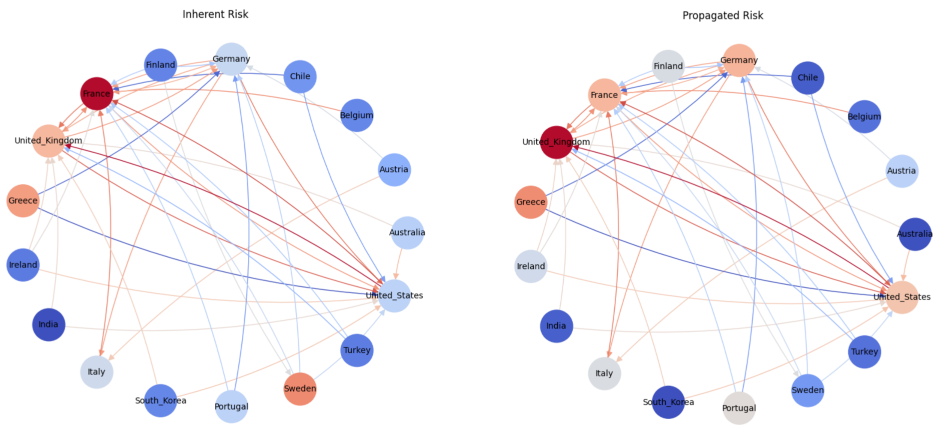
Firstly, after deleting nodes, the number of edges in the remaining network is 241, nearly half of before. And the comparison between risks is shown in the figure below:

	inherent_risk_before_deletion	inherent_risk_after_deletion	propagated_risk_before_deletion	propagated_risk_after_deletion
0	(Canada, 3.279436230223105)	(France, 2.819779165253216)	(United_Kingdom, 0.0692381146915677)	(United_Kingdom, 0.09543567721705071)
1	(Spain, 2.8982666368586854)	(Sweden, 2.220080819270625)	(Spain, 0.06850635252576369)	(Greece, 0.08241168605425063)
2	(Japan, 2.871092610464302)	(Greece, 2.082681121519881)	(Greece, 0.05982975073514005)	(Germany, 0.07645097147100623)
3	(Switzerland, 2.4237244451641877)	(United_Kingdom, 1.8783333937463362)	(Germany, 0.0584028473434621)	(France, 0.07594001298782484)
4	(Netherlands, 2.006200687678921)	(Italy, 1.2450947577997007)	(France, 0.055099320772883945)	(United_States, 0.07315915235061783)
5	(France, 1.8393602113378291)	(Germany, 1.1689268482463466)	(Finland, 0.05503877576059898)	(Portugal, 0.06666071637125452)
6	(Greece, 1.744085786388376)	(Portugal, 1.0881602905207552)	(Netherlands, 0.05492860343159948)	(Italy, 0.06503683206275343)
7	(Sweden, 1.6880519661675164)	(United_States, 1.081108193987226)	(United_States, 0.05387045039652726)	(Finland, 0.06452570694165613)
8	(United_Kingdom, 1.012396006676269)	(Australia, 1.0509549081343135)	(Ireland, 0.04995740986990863)	(Ireland, 0.06297460599915577)
9	(Italy, 0.9163581915535357)	(Austria, 0.7027345048263981)	(Austria, 0.048390898919203604)	(Austria, 0.05891658977795809)

Figure 4.6.1: The comparison for inherent and propagated risks

In Figure 4.6.1, we can see that although the rankings of the corresponding countries' own risks are nearly the same, the corresponding values have risen significantly, and the propagated risks also show a similar trend.

Next, we demonstrate these 2 risks in the network:



(a) Inherent risk of 2021/03/31 after deletion (b) Propagated risk of 2021/03/31 after deletion

Figure 4.6.2: The inherent and propagated risks of 2021/03/31 after deletion

From Figure 4.6.2, we can draw a similar conclusion that the ranking of the country's risk level has not changed.

Finally, we explore the changes in the community:

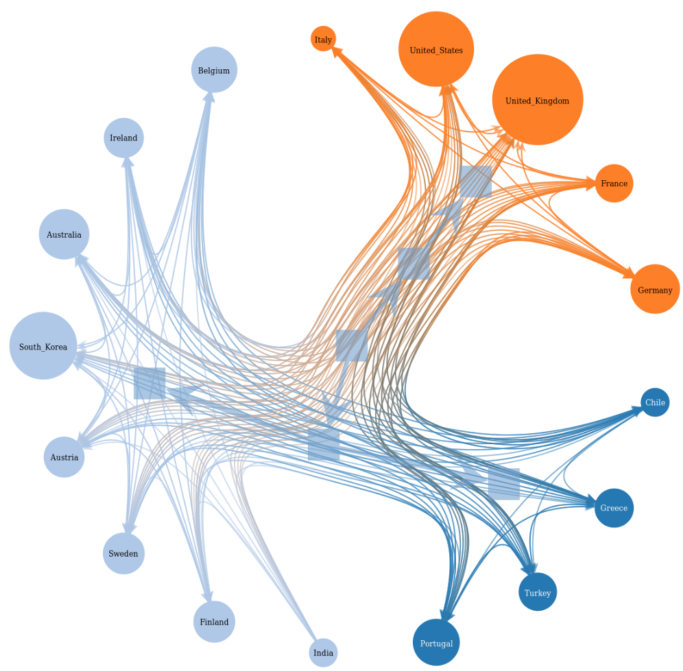


Figure 4.6.3: The communities after deletion

Also, we want to use a table to show the comparison:

	same	before deletion (ignore deletions)	after deletion
Community 1	United_States, United_Kingdom, France, Germany, Italy	-	-
Community 2	Austria, Australia, South_Korea, Belgium, Sweden, Finland, India	Portugal	Ireland
Community 3	Turkey, Greece, Chile	Ireland	Portugal

Table 4.5: The comparison for two communities before and after deletion in 2021/03/31

After ignoring the deleted nodes, the only difference is that in the new network, Portugal is further down to community 3, while Ireland rises to community 2. For Portugal, this may be because after deleting the node Spanish, its largest venture capital is no longer there, so compared to other countries, the risk of being propagated is relatively low. While France and the United Kingdom, where Ireland mainly invested, became new high inherent risk nodes after the node was deleted, so the risk of Ireland being propagated began to rise.

To sum up, our last insight in this project should be:

- **Insight 8**: When an external shock is added, that is, when high inherent risk nodes are unable to repay their debts, the entire network will be reorganized to generate a new round of risks, and the risk coefficient will increase compared to before. Besides, the community will also produce corresponding changes according to the differences in the relative rankings between countries.

Chapter 5

Evaluation

For evaluation, here we will use insights to answer our research questions.

How do the cross-border banking network properties/statistics evolve over time?

In Section 4.3 (Network Inspection), we studied the evolution of network properties over time. As concluded in **Insight 1** and **Insight 2**, and shown in Figure 4.3.3, most indicators showed an upward trend and fluctuated during the period of COVID-19. However, we explained that this indicator cannot determine whether it means risk. Next, we conducted a more detailed analysis through feature processing and visualization.

What are the individual and global systemic risk levels of each country?

After feature processing, we solved this question in Section 4.5 (Cross-Sectional Analysis) and concluded this methodology as **Insight 6**. Through visualization, we can intuitively see the inherent risk and propagated risk of each country. At the same time, we can obtain more insights to help banks make investment decisions.

Are there any non-obvious community structures in the network?

We have used stochastic block models to do many community detections during the analysis. For each detection, we ensured the stability of the result by performing multiple times and selecting the one with minimum description length. **Insight 3**, **Insight 5**, **Insight 7** and **Insight 8** show that usually, the number of communities in the network is kept at 3. With the evolution of time or stress tests from the external shock, changes in borrowing and lending amounts between countries will also affect the composition of members in each community.

Chapter 6

Conclusions

6.1 Conclusions

In conclusion, based on the knowledge of network science, this project conducted a network analysis to investigate cross-border banking exposures. This should be an applied and exploratory research, consisting of inductive and abductive methods.

In the design and implementation, we divided the process into 6 parts: environment, data, network, transformation, analysis and evaluation.

For environment, based on *Python* and *Noteable* platform, we imported *numpy* and *pandas* for data processing, *Matplotlib* for data visualization, *NetworkX* and *graph-tool* for network operation.

As to data, We gained and inspected the data from the aspects of fields, description of merged information, and time-trend analysis. In **Insight 1**, we pointed out that although the claim amounts kept increasing, special attention should be required for the period of COVID-19.

After generating the networks, we did a detailed inspection of them, including their basic statistics, homogeneous statistics, heterogeneous statistics and community structures. We concluded the findings into **Insight 2** and **Insight 3**, such as India is a special node which only has outward arrows while Japan is a node that has the highest in-strength value.

During transformation, to better understand the risk of each country, we made some assumptions like the lending amount can represent the economic strength of a country, and then we added 2 new features: inherent risk and propagated risk. Finally, we displayed these features in the network graph by setting different node colors.

Since analysis is the core part, we subdivided it into static analysis and dynamic analysis, according to whether the simulation operation is performed.

- Static Analysis: After feature processing, in the comparison with previous networks, we obtained **Insight 4** and **Insight 5**, which try to explain the meaning of normalized borrowing value and the meaning of different communities. Next, in the cross-sectional analysis, we concluded **Insight 6**, which is a methodology to explore the risk level of each country. For example, in 2021/03/31, we found 5 high inherent risk countries (e.g., Canada, Spain, Japan...) and 8 high propagated risk countries (e.g., United Kingdom, Spain, Greece...). As to time series analysis, we compared the distributions of communities in 2 networks before and within COVID-19, and got the **Insight 7**, linking the capital flows to the changes in the community structure.

- Dynamic Analysis: When we deleted the high inherent risk nodes and observed the consequence, we found the **Insight 8**, where most countries' inherent and propagated risks are increasing, and their risk rankings also have changed due to the changes in edges.

Finally, in evaluation, we simply tested our findings by using these 8 insights to answer our research questions.

6.2 Limitations and Future work

Limited by time and resources, the analysis in this project has not achieved complete coverage, and it can still be optimized in the future:

- We only focused on the network at 3 time points (2015/03/31, 2019/03/31, 2021/03/31) for analysis. In such a long time range, there must be other interesting points to obtain more meaningful insights.

- There is still room for optimization in the feature processing, especially in the transformation of propagated risk. We can try more ranking algorithms and hyperparameter adjustments in the future.

- In dynamic analysis, we can try more strategies for node deletions and consequence simulation. In addition to risk analysis and community analysis, we can also do a complete static analysis.

Bibliography

[1] Viral V Acharya, Lasse H Pedersen, Thomas Philippon, and Matthew Richardson. Measuring systemic risk. *The review of financial studies*, 30(1):2–47, 2017.

[2] Tobias Adrian and Markus K Brunnermeier. Covar: A method for macroprudential regulation. *Federal Reserve Bank of New York Staff Report*, 348, 2008.

[3] Suad A Alasadi and Wesam S Bhaya. Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16):4102–4107, 2017.

[4] Ivan Alves, Stijn Ferrari, Pietro Franchini, Jean-Cyprien Heam, Pavol Jurca, Sebastiano Laviola, Sam Langfield, Franka Liedorp, Antonio Sánchez, Santiago Tavolaro, et al. The structure and resilience of the european interbank market. *ESRB: Occasional Paper Series*, (2013/03), 2013.

[5] Chris Anderson, Christopher Nissley, and Chris Anderson. The long tail, 2006.

[6] Rodolfo Baggio, Noel Scott, and Chris Cooper. Network science: A review focused on tourism. *Annals of Tourism Research*, 37(3):802–827, 2010.

[7] Monica Billio, Mila Getmansky, Andrew W Lo, and Loriana Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics*, 104(3):535–559, 2012.

[8] Adrian Blundell-Wignall. Solving the financial and sovereign debt crisis in europe. *OECD Journal: Financial Market Trends*, 2011(2):201–224, 2012.

[9] Duncan S Callaway, Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. Network robustness and fragility: Percolation on random graphs. *Physical review letters*, 85(25):5468, 2000.

[10] Joao F Cocco, Francisco J Gomes, and Nuno C Martins. Lending relationships in the interbank market. *Journal of Financial Intermediation*, 18(1):24–48, 2009.

[11] Roz Combley. *Cambridge business English dictionary*. Cambridge University Press, 2011.

[12] Ines Drumond. Bank capital requirements, business cycle fluctuations and the basel accords: a synthesis. *Journal of Economic Surveys*, 23(5):798–830, 2009.

[13] Jin-Chuan Duan and Changhao Zhang. Cascading defaults and systemic risk of a banking network. *Available at SSRN 2278168*, 2013.

[14] Helmut Elsinger, Alfred Lehar, and Martin Summer. Using market information for banking system risk assessment. *Available at SSRN 787929*, 2005.

[15] Ernesto Estrada. Quantifying network heterogeneity. *Physical Review E*, 82(6):066102, 2010.

[16] Katrin Forster, Melina A Vasardani, and Michele Ca'Zorzi. Euro area cross-border financial flows and the global financial crisis. *ECB occasional paper*, (126), 2011.

[17] Benjamin H Good, Yves-Alexandre De Montjoye, and Aaron Clauset. Performance of modularity maximization in practical contexts. *Physical review E*, 81(4):046106, 2010.

[18] Andrew G Haldane and Robert M May. Systemic risk in banking ecosystems. *Nature*, 469(7330):351–355, 2011.

[19] Masazumi Hattori and Yuko Suda. Developments in a cross-border bank exposure "network". *IDEAS Working Paper Series from RePEc*, 2007.

[20] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[21] Petter Holme. Modern temporal network theory: a colloquium. *The European Physical Journal B*, 88:1–30, 2015.

[22] Masayasu Kanno. Assessing systemic risk using interbank exposures in the global banking system. *Journal of Financial Stability*, 20:105–130, 2015.

[23] Kazi Zainab Khanam, Gautam Srivastava, and Vijay Mago. The homophily principle in social network analysis: A survey. *Multimedia Tools and Applications*, 82(6):8811–8854, 2023.

[24] Chun Kit Kwong, WH Ip, and JWK Chan. Combining scoring method and fuzzy expert systems approach to supplier assessment: a case study. *Integrated manufacturing systems*, 13(7):512–519, 2002.

[25] Andrea Landherr, Bettina Friedl, and Julia Heidemann. A critical review of centrality measures in social networks. *Wirtschaftsinformatik*, 52:367–382, 2010.

[26] Priscilla Yin-Pheng Lim, Roy WL Khong, and Hui-Boon Tan. Bank integration and systemic risk: Panacea or pandemic? *The Journal of Developing Areas*, pages 301–312, 2015.

[27] Saku Mantere and Mikko Ketokivi. Reasoning in organization science. *Academy of management review*, 38(1):70–89, 2013.

[28] Filippo Menczer, Santo Fortunato, and Clayton A Davis. *A first course in network science*. Cambridge University Press, 2020.

[29] Camelia Minoiu and Javier A Reyes. A network analysis of global banking: 1978–2010. *Journal of Financial Stability*, 9(2):168–184, 2013.

[30] Chris Olsen and DMM St George. Cross-sectional study design and data analysis. *College entrance examination board*, 26(03):2006, 2004.

[31] Tiago P Peixoto. Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X*, 4(1):011047, 2014.

[32] Tiago P Peixoto. Bayesian stochastic blockmodeling. *Advances in network clustering and blockmodeling*, pages 289–332, 2019.

[33] Tiago P Peixoto. Descriptive vs. inferential community detection in networks: pitfalls, myths, and half-truths. *arXiv preprint arXiv:2112.00183*, 2021.

[34] Mohamed AF Ragab and Amr Arisha. Research methodology in business: A starter's guide. 2018.

[35] Thiago Christiano Silva, Sergio Rubens Stancato Souza, and Benjamin Miranda Tabak. Monitoring vulnerability and impact diffusion in financial networks. *Journal of Economic Dynamics and Control*, 76:109–135, 2017.

[36] Herman Strydom. An evaluation of the purposes of research in social work. *Social Work/Maatskaplike Werk*, 49(2), 2013.

[37] Felipe Vaca-Ramírez and Tiago P Peixoto. Systematic assessment of the quality of fit of the stochastic block model for empirical networks. *Physical Review E*, 105(5):054311, 2022.

[38] Yasuhiro Yamai and Toshinao Yoshiba. Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance*, 29(4):997–1015, 2005.