NOVA IMS Information Management School

MAA

Mestrado em Métodos Analíticos Avançados
Master Program in Advanced Analytics

Global cross-country investment relationships – using network analysis and interactive visualization techniques

Bojan Stavrikj

Dissertation presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, Major in Business Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

Global cross-country investment relationships – using network analysis and interactive visualization techniques

by

Bojan Stavrikj

Dissertation presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, Major in Business Analytics

Co-supervisors: Doutor Flávio Luís Portas Pinheiro & Doutor João Miguel Falcão Pinto da Silva

November 2021

Abstract. One important feature of the globalization process is the increase in the economic and financial interdependencies across counties. In this context, it is crucial to measure the connections between countries in order to identify where the financial centers are located, to characterize the foreign investments, and define the set of world countries that have stronger linkages, among others. In this context, foreign direct and portfolio investments play a crucial role in measuring international investments and understanding these dynamics. Such an interconnected web of foreign investment relationships is difficult to measure due to their complexity, but as well as the lack of unified data sources. This thesis aims to use network analysis to map the portfolio and foreign direct investment global relationships in order to identify patterns, preferential paths for investment, establish trends and describe the relations between countries over time. Secondly, it gathers the results of the network analysis and presents them in an intuitive web application, where the most important findings are highlighted allowing the users to interact with the data and extract insights over all the available years.

Keywords. Foreign direct investment, Portfolio investment, Network analysis, Interactive web application

Table of Contents

1 Introduction

With the ever-growing globalization, an increase in trade, and the possibility of investing in different countries at the tip of our fingers, it is becoming increasingly difficult to track the flow of money between countries. In such an interconnected web of relationships characterized by many players, markets, and investment opportunities, it is complex to map all the linkages between the origin and destination of each investment.

The external statistics compiled by central banks is an important tool to monitor the economic operations between one country and the rest of the world. This is crucial for policymakers when evaluating the existence of investment opportunities, financial/currency crises, the origins/destinations of external investments among other international relationships. In this context, the international accounts aim to track all international monetary and non-monetary transactions during a given period. The Balance of Payments Manual 6th edition (BPM6) from the International Monetary Fund (IMF) aims to establish a standard framework of statistics for recording transactions and positions between a country and the rest of the world.

According to the international standards, the Balance of Payments (BoP) is composed by three main components - the current account, the capital account, and the financial account. The current account deals with the trade of goods and services, primary income, and secondary income; the capital account deals with the flow of capital across borders. While the financial account is concerned with inter-country monetary flows related to financial investments from (to) the country to (from) the rest of the world, in the form of debt securities, equity, loans, amongst others. The financial account of the BoP is also an important factor in the International Investment Positions framework, which will be explained next.

The International Investment Positions (IIP) refer to the end-of-period stocks. It corresponds to the value, of all financial assets for residents of an economy that have claims on non-residents, as well as the liabilities of residents to non-residents at a specific time (IMF, 2009). The difference between the asset and liability sides gives rise to the net position of the IIP. The changes in the IIP are explained by the cross-border flows, including transactions, price changes, exchange rate changes, and other changes in volume. The amount of transactions corresponds to the financial flows recorded under the financial account in the BoP. While, the rest of the flows correspond to the change in the market valuation of financial assets/liabilities, appreciation and depreciation of currencies, and other changes not explained by the other flows. Therefore, the combined variation of the flow elements explain the difference in IIP between two periods (IMF, 2009). Notwithstanding, the standard measures used to record the external statistics are usually related between one country and the rest of the world, on an aggregated basis.

This thesis focuses on the bilateral analysis of one country vis-à-vis its main partners aiming to analyze the cross-border financial investments in the form of debt securities and equity. These financial instruments (debt securities and equity) are recorded in the financial account under the functional categories of the Foreign Direct Investment (FDI) and the Portfolio Investment (PI). FDI includes the initial investment and all the other financial linkages from residents of one country in an enterprise located in a foreign country, when the investor owns a minimum of 10 percent of the voting power (OECD, 2008). Such investments aim to establish a long-lasting interest in a foreign business. On the contrary, the Portfolio Investment corresponds to cross-border investments in the form of debt securities and equity that falls below the FDI threshold. This type of investment is oriented toward small investors that aim to get short-term returns from their investments (OECD, 2008). These two indicators are relevant as they correspond to foreign

investments. **Figure 1** represents the global foreign direct investment (assets) and the portfolio investment (assets) measured as % of the world GDP.

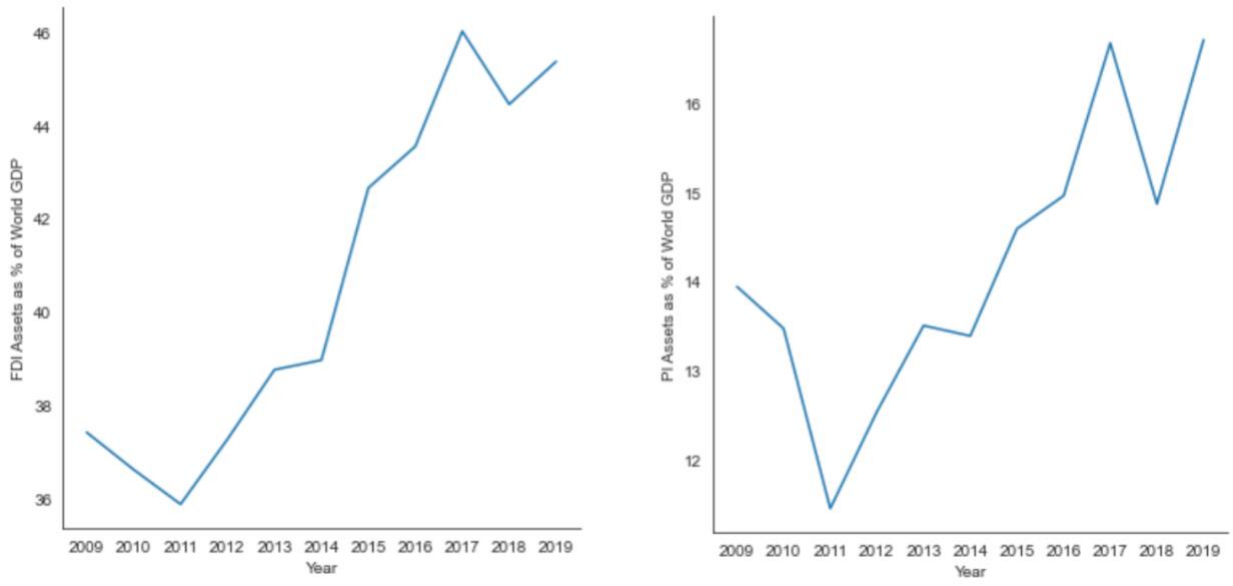


Figure 1 – Total FDI assets (left chart) and portfolio investments assets (right chart) as % of World GDP

As shown in the graphs of **Figure 1**, the FDI and PI represent a significant proportion of the world GDP. The global FDI assets and PI assets correspond to 45.36% and 16.71% of world GDP in 2019 respectively.

It is noticeable that as time passes and economies and markets reduce the barriers of entry (Sobel, J.R., & Dwigth, 2007), the significance of direct and portfolio investments as a proportion over GDP is growing. Some of these observations can be explained by the rapid expansion of multinational companies, international investments, as well as bilateral treaties between countries (Egger & Pfaffermayr, 2004). The growing importance of FDI and Portfolio investment is showcased in the graphs of **Figure 1**, making these complex relationships relevant and interesting to study.

With the ever-growing volume of investments and the tendency of corporations to funnel money through offshore centers, it is difficult to fully grasp international investment connections (Damgaard & Elkjaer, 2017). In the case of portfolio investment, many transactions occur in secondary markets, while in the case of direct investment according to the BPM6, the operations are recorded vis-á-vis the immediate counterpart and not the ultimate investor (IMF, 2009). In that sense, if an enterprise in economy A has a subsidiary[1] in economy B and B has a subsidiary in economy C, the ultimate investor of C is A, although only the direct investment relationship is recorded (ie. the FDI statistic would show a direct relationship between A and B and B and C separately). Therefore, making it difficult to trace the ultimate investor and, consequently, the "real" sources and destinations of the international investments.

According to the existing documentation, IMF provides additional information by counterpart country both for FDI and Portfolio Investment, consistent with the BPM6: the Coordinated Direct Investment Survey (CDIS) and Coordinated Portfolio Investment Survey (CPIS). The CDIS and the CPIS correspond to annual (CDIS) and semi-annual (CPIS) surveys where immediate counterpart relationships are observable. Since ultimate relationships are more complex, underlying structures are difficult to extract without applying advanced models that can imply these relationships. Hence this thesis aims to identify patterns, preferential paths for investment, establish trends and describe the relations between countries over time using the available information from CPIS and CDIS.

[1] A subsidiary is a company that is owned or controlled by another company, usually referred to as the parent company. (OECD, 2001)

We use network analysis to model the linkages between the sources and final destinations of the foreign direct and portfolio investments on a global level. This tool is commonly used in analyzing complex social interactions, where many actors relate through bilateral ties that create a relationship (Scott, Social Network Analysis, 1988). Networks have been used in analyzing corporate power (Scott, 2003), international ties (Hafner-Burton , Kahler, & Montgomery, 2009), class structure (Prota & Bersford, 2011), etc. Their complex systems roots, help understand how individual interactions lead to the creation of a macroscopic structure that exposes the fundamentals of the system behavior as a whole (Lima, Pinheiro, Silva, & Matos, 2020).

The importance of network analysis for this thesis lies in its potential to map the foreign investment linkages based on CDIS and CPIS data, both on an individual level against the rest of the world as well as globally. Applying network analysis to these datasets exposes the underlying structure of the networks. Therefore, this thesis will leverage on the properties of network analysis, allowing us to transform the bilateral connections to a closer to reality core structure of global investment interconnections. Ultimately, identifying the most important players in a global network of foreign direct investments and portfolio investments.

This thesis is structured in 8 main sections: after the introduction section, we present the literature review, then the data section, after which the network analysis section follows. Thereafter, the hypothesis formulation section will generate the main questions this thesis is aiming to answer. The results section will discuss the findings of the analysis and highlight the results and differences between networks. The platform section will justify the choices behind the web application visualizations, their functionality, purpose, and storyline. And lastly, the conclusion will summarize the findings, and point towards limitations but also improvements that can be done in the field.

2 Literature review on the Economic Perspective & Network Analysis

Direct and Portfolio Investment are two important measures of foreign international investments. According to the Balance of Payments Manual (BPM6) these categories are recorded on an asset/liability principle (IMF, 2009). An asset corresponds to something a person, company, or country owns that is expected to bring future economic gains to the holder. There are several types of assets, including current, fixed, financial, and intangible assets. Current assets are seen as short-term resources that can easily be liquidated and won't be held by the owner for a very long time, they include cash and cash equivalents. In general terms, financial assets are investments such as equity, bonds, deposits, loans and other equity, and their value is usually dependent on factors affecting the financial market. While intangible assets do not have a physical presence (OECD, 2001). A liability on the other hand is something a person, company, or a country owes to a counterparty. Liabilities include, for instance, short/long-term borrowing, which is mostly used for funding operations or expansions (OECD, 2001).

The asset side of the foreign direct investment positions represents the resident parents' equity in and lending to foreign affiliates plus the resident affiliates' equity in and lending to foreign parents. The liability direct investment positions represent the foreign parents' equity in and lending to resident affiliates plus the foreign affiliates' equity in and lending to resident parents (IMF, 2009).

In addition to the asset/liability principle defined in the BPM6, in the case of foreign direct investment, there is the directional principle to record the external operations. According to this principle, the external operations *vis-à-vis* the rest of the world through foreign direct investment

are recorded as inward/outward investments. The inward positions represent resident parents' equity in the foreign country and all lending to foreign affiliates minus foreign affiliates' equity in the resident country and lending to resident parents (IMF, 2009). The outward positions represent the foreign affiliates' equity in the resident country and lending to foreign affiliates minus the foreign affiliates' equity in the resident country and lending to resident parents (IMF, 2009). Although the two approaches differ conceptually, they are equal in net terms.

For an investment to be classified as a direct investment it needs to hold at least 10% of the voting power of a company residing in a foreign economy. Although, there are different direct and indirect combinations of flows between enterprises that are part of these flows. A foreign direct investor is an entity resident in one economy that owns directly or indirectly at least 10% of voting power in an enterprise resident in a foreign economy. It can be an individual, private entity, government body, etc… Furthermore, a foreign direct enterprise is an enterprise resident in one economy in which an investor resident in a foreign economy owns a minimum of 10% of its voting power. The above definitions give rise to the immediate relationships between two parties. Although, there is a much more complex structure of ownership chains due to the non-immediate relationships that may exist between the enterprises. The Framework of Direct Investment Relationships (FDIR) is a general methodology that helps identify non-immediate relationships of direct investors and direct investment enterprises (OECD, 2008). There are rather simple situations in which it is easy to identify the direct relationships. In a relationship where company A is linked to B and B to C, any transactions between A and C need to be considered in the FDI statistics collection. Therefore, A has indirect control over C. Although, there are also some more complex relationships where the FDIR helps identify the ultimate investor. Based on FDIR (OECD, 2008), a subsidiary is a foreign enterprise that owns more than 50% of the voting power, or an investor

and its subsidiaries that own at least 50% of a foreign company's voting power. Alternatively, an associate is considered when with percentages between 10% and 50%. The subsidiary and associate types of relationships are considered to be part of the FDI statistic, while ownerships lower than 10% of a foreign company's voting power are not.

Following this discussion, it is important to include not only the immediate counterpart country but also the ultimate and intermediate countries to understand how the foreign investments are being financed. In addition, for the case of portfolio investment as this type of investment is made through secondary markets it is important also to understand the following patterns of the immediate investment. These kinds of relationships are difficult to observe, especially due to the nature of the way these statistics are recorded. And that is why it is important to distinguish between immediate and ultimate investors.

According to the available data, it is possible to track the different foreign investment patterns through the CDIS and CPIS. When considering one such investment relationship it is possible to trace the patterns of the investments, although when considering a global interconnection of investments, it gets tricky. Thinking about these relationships in matrix form can help with beginning to understand the complexity of the connections. Taking two countries of the CDIS dataset as an example would generate a 2 by 2 matrix. This represents the direct connection between 2 countries, where each cell shows the strength of their connection except the diagonal (which is a self-relationship). It would be rather simple to understand the relationship between 4 countries, where a 16 cell matrix would be computed. Although, when considering 20 countries this would already imply a 400 cell matrix. This gets messy quickly, making it difficult to visualize the connections and understand the most influential ones and especially how one investment transcends through to other countries other than its reported destination. This gives rise the power

of network analysis in exposing the core structure of global investment ties. It takes the bilateral and immediate connections between participant countries in both the CDIS and CPIS and aims to identify the underlying investment paths, or in other words, it aims to identify the intermediate investors and the investment paths.

A network is a system made up of actors (individuals, organizations, countries, etc.) and sets of bilateral ties that define a relationship between them (Wasserman & Faust, 1994). This provides a structure for network analysis, allowing for the identification of central agents in complex local and global networks. Network science offers a unique set of tools and principles for studying complex relationships apparent in nature, technology, and society (Jackson, 2008). They help us understand how diseases spread, patterns in product purchases, languages spoken, voting, and educational decisions, to name a few (Jackson, 2008). Despite evident differences in various network domains, they emerge and evolve based on a set of fundamental mechanisms that govern network science (Barabási, 2013). Ter Wal & Boschama (2008) showcase the huge potential of network analysis to be incorporated in studying the structure and evolution of inter-organizational connection and knowledge sharing. They stress the importance of using high-quality data in building networks, and they identify primary data as the most statistically robust way of building networks (Ter Wal & Boschma, 2008).

Networks have a wide range of applications, including patterns in banking and cross-country financial investments. Hakeem and Suzuki (2016) take a network approach in foreign portfolio investment of the European Union and its main trading counterparts. They focus on the relationship between countries' centrality and their economic indicators, showing that the more connections an economy holds the higher the impact on economic growth patterns (Hakeem & Suzuki, 2016). In another study it is suggested that the more central a country is, the more embedded it is in a global

portfolio investment network, implying greater exposure to foreign financial markets. Ultimately, the more central a country is, the more its stock market comoves with the world stock market index (Chuluun, 2016). Norgren and Olsson (2021) in their study apply Stochastic Actor-Oriented network models to study the relationship between FDI and institutions. Where institutions are men made constraints that define general and economic interactions. They distinguish between formal and informal institutions, where informal institutions are culture and trends while formal institutions are the laws and rules of society. Ultimately concluding that FDI does not contribute to the spread of institutions to foreign countries, but the selection effect does play an important role between FDI and a subset of institutions (Norgren & Olsson, 2021). In another study Hafner-Burton, Kahler, & Montgomery (2009) use network analysis on key international outcomes and test network theory in the context of international relations.

Even though by definition FDI is meant to establish a long-lasting interest, lots of investments that are reported as FDI fall short of it. In many cases, the ultimate investor is difficult to trace, especially when investments involve offshore centers or Special Purpose Entities (SPE's). SPE's are legal entities that have little or no employment, operations, or physical presence in the jurisdiction in which they are created by their parent enterprises, which are typically located in other jurisdictions (economies) (OECD, 2008). They are often used as devices to raise capital or to hold assets and liabilities and usually do not undertake significant production (OECD, 2008). Small economies with inexplicably large FDI inflows are one of the clear signs that a country's total FDI value is inflated and its counterparts are not necessarily trying to establish long-lasting relations but instead use it for different financial planning goals (Damgaard & Elkjaer, 2017). Luxembourg and Netherlands are two such economies that host many foreign-owned multinational enterprises or SPE's. Damgaard and Elkjaer (2017) show the difference in those networks by

combining CDIS with the OECD data. Using regression analysis, they estimate the amount of "real" FDI each of these relationships holds. Once the transformations are applied there is a 34% decrease in total inward FDI. Representing the "real" and "phantom" FDI in a network shows some differences. Smaller economies, such as Netherlands and Luxembourg weakened their intermediating power in the network although remain one of the most important global intermediators. It is also important to note that outward and inward reported numbers for most country pairs do not match more often than they do. For example, in the CDIS data for 2015, one country's published number is at least twice as higher as their counterpart for 44% of the economy pairs (Damgaard & Elkjaer 2018). To further expand on that another study aims to map the global FDI offshore network. They conclude that the offshore network is extremely globalized with a centralized core in Northwest Europe and the Caribbean (Haberly & Wójcik, 2015).

In another study, Bolívar *et al*. (2019) find that there is a strong relationship between economy size (measured through GDP) and the centrality of an economy in a global network. Furthermore, a country's commercial openness also has a strong positive relationship with centrality in the network. The more open an economy is the more FDI it attracts and inversely the greater the involvement in the outward global FDI network (Bolívar, Casanueva, & Castro, 2019). On the other hand, political stability and average years of schooling have a moderate effect on inward FDI investment for that economy. They also show that economies in the same development state tend to have stronger ties. As evidenced, developed-to-developed connections represent (66%) of the weighted global FDI networks (Bolívar, Casanueva, & Castro, 2019).

As showcased in this section there is a wide range of current literature on the application of network analysis to FDI and portfolio investment relationships. This thesis adds to the currently existing literature by taking the bilateral immediate investment connections between countries and

inferring the path of foreign investment both to foreign direct and portfolio investments by applying network analysis without burdening countries to report supplementary data on ultimate investment targets. As well as communicating the results through an interactive web application.

3 Data

We use data sourced from IMF's database on Coordinated Direct Investment Survey (CDIS) and Coordinated Portfolio Investment Survey (CPIS). They are surveys that collect data with regards to FDI and portfolio investment respectively, are available on an annual basis (in the case of FDI) and semi-annual basis in the case of Portfolio Investment after 2013.

Each dataset follows a similar structure. The CDIS and CPIS datasets contain 6 main variables and are summarized in Table 1.

Table 1 — Summary of the main variables from the CDIS and CPIS datasets that were used in this thesis

Variable Name	Description	Data Type
Country Name	The country of origin, or the investor	Categorical
Country Code	The country of origin's ISO code	Categorical
Counterpart Country Name	The counterpart, or the recipient of the investment	Categorical
Counterpart Country Code	The counterpart country's ISO code	Categorical
Time Period	The period as a year	Continuous
Value	The investment amount	Continuous

The indicator names, codes, their naming convention in the web application we have developed are summarized in Table 2. Onwards, we use the terms CDIS Inward, CDIS Outward, CPIS Assets, and CPIS Liabilities to refer to each of these datasets.

Table 2 — Summary of indicators used in this thesis**.** The first four indicators are sourced directly from the IMF database, while the last two are a pairing of the other four indicators.

Web App	Indicator description	Indicator Code	Source
CDIS Inward	Inward Direct Investment Positions, US Dollars	IIW_BP6_USD	IMF
CDIS Outward	Outward Direct Investment Positions, US Dollars	IOW_BP6_USD	IMF
CPIS Assets	Assets, Total Investment, BPM6, US Dollars	I_A_T_T_T_BP6_USD	IMF
CPIS Liabilities	Liabilities, Total Investment, BPM6, US Dollars	I_L_T_T_T_BP6_USD	IMF
Assets Combined	CDIS Outward + CPIS Assets	IOW_BP6_USD + I_A_T_T_T_BP6_USD	Authors' calculations
Liabilities Combined	CDIS Inward + CPIS Liabilities	IIW_BP6_USD + I_L_T_T_T_BP6_USD	Authors' calculations

Since both CDIS and CPIS correspond to total amounts in US dollars and not proximities, which is the desirable metric for our analysis, it is therefore necessary to transform their values to proximities.

Let us start by defining f_{ij}as the total investment of country i in country j, which can in general terms concern any of the indicators shown in Table 2. The investments between two countries can be asymmetrical, that is, $f_{ij} \neq f_{ji}$ implying the directionality of the investments. While, in most cases, the value of f_{ij} is positive, in certain situations it can also be negative. For instance, suppose country i (the parent) invested and is holding a position in country j (the affiliate); the parent can use the affiliate for funding operations back at home. When the total amount of funding that flows back to the parent, exceeds the total amount of investment done in the affiliate it is standard to report it as a negative flow. As such, we shall consider the absolute value of f_{ij}, that is, $|f_{ij}|$. The

main reason for taking the absolute value instead of removing such observations is to prevent the loss of information about important investment partners.

In order, to obtain a proxy for proximity we consider the reciprocal of the absolute value of the directional investment. Therefore, the proximity calculation for each investment is as follows:

$$\phi_{ij} = \frac{1}{|f_{ij}|} \tag{1}$$

where $\phi_{ij} \neq \phi_{ji}$. In that sense, we say that the larger the investment amount between two countries, the closer they are to each other. Ultimately a proximity matrix is obtained, which forms the basis for building a directed weighted network that represents global trade flows in foreign investment.

Additionally, we combine both Inward and Outward flows from CDIS and CPIS datasets. It is interesting to understand how the networks change once the main cross-country investment datasets are combined, to map possible international investment linkages across countries. These datasets do not refer to the same statistical concepts because FDI is recorded in the directional principle, whereas the CPIS is recorded on the assets/liabilities perspective. Although there is no available information on the FDI assets/liabilities by counterpart country, therefore, the best proxy that can be used to understand the international investment in the form of debt/equity is to aggregate CDIS Outward with CPIS Assets as Combined Assets, and CDIS Inward with CPIS Liabilities as Combined Liabilities.

4 Network Analysis

Network analysis has found applications in many fields of study (Wasserman & Faust, 1994). In particular, it has been useful for the mapping and analysis of the web of interconnected relationships between actors that relate through bilateral ties (Scott, Social Network Analysis, 1988). Networks have been used in analyzing corporate power (Scott, 2003), international ties (Hafner-Burton , Kahler, & Montgomery, 2009), class structure (Prota & Bersford, 2011), etc. Network analysis's roots in the study of complex systems have helped us to unveil the link between individual interactions and the emergence of seemingly disconnected macroscopic patterns and behaviors (Lima, Pinheiro, Silva, & Matos, 2020).

A network, *G*, is composed of two different but complementary elements: a set of N vertices/nodes and a set of links/edges. Edges connect a pair of nodes and identify the existence of a relationship between them. Vertices represent the unit of analysis, while the edges represent the relationship between them. In a social network framework, a node would be an individual and an edge could represent the existence of friendship, co-work, or trust tie between pairs of individuals. Or a node could represent a firm and the edge, a financial relationship between them. In this thesis, we shall use nodes of a network as representations of countries, while edges represent the existence of an investment between a pair of countries.

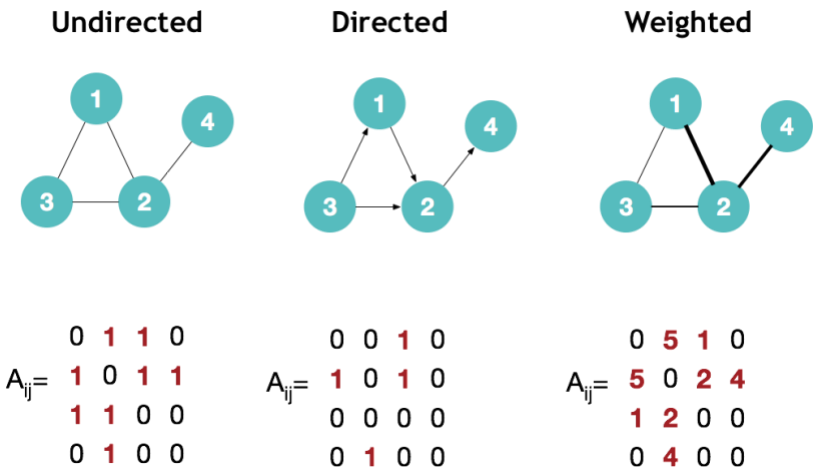


Figure 2 – Graphical (top) and matrix (bottom) representation of networks representing structures that have edges of different nature. Source: (Lima, Pinheiro, Silva, & Matos, 2020).

Depending on the nature of the relationships being modeled, there can be three main types of networks, namely: undirected; directed; and weighted. One way to represent such networks is through the adjacency matrix, see bottom panel of **Figure 2**. The adjacency matrix, A, of a network informs on the existing relationships between nodes/vertices. In that sense, the entry a_{ij} of A is zero if there is no relationship between nodes *i* and *j*, being non-zero if a relationship exists between such a pair of nodes. In a weighted graph, a_{ij} represents the weight (strength) of the relationship between the nodes[4]. In the case of an undirected network $a_{ij} = a_{ji}$. For the directed network case the matrix is not symmetric along the diagonal and indicates a relationship and its direction, therefore $a_{ij} \neq a_{ji}$. The diagonal entries of each matrix A indicate self-relationships, and as general practice are set to zero. In the case of all network types, the degree of a node is defined as the number of edges it has in common with all other nodes. While, the average degree refers to the network as a whole and aims to measure the relationship between the number of nodes and the

[4] In an unweighted graph this value is one. In other words, either the relationship/link exists or not.

number of edges. This measure is obtained by dividing the total number of edges by the number of nodes.

When building a graphical representation of a network, each edge represents a relationship between two nodes (i.e., person, country, institution). In a directed network, edges are represented with arrows to indicate the direction of the relationship (from source to destination). In the case of directed networks, two nodes can have two links (arrows) between each other. Additionally, in a weighted graph, the thickness of the edge can be used to represent the strength of the relationship.

Additional attributes can be associated with each relationship and each actor, which can be associated with the edges and nodes of the network (e.g., we might want to consider the gender or age of individuals). However, these additional attributes do not affect the core structure of the networks, but they add a dimension that allows to classify relationships and profile explanatory factors for the creation of relationships.

Weights can be interpreted either as proximities/similarities or distances. It is important to establish which measure is being used in a network, as they have opposite interpretations. However, the choice hinges on a balance between the available data and the analytical purpose of the network structure under study.

Lastly, it is also common to study a simplified projection of the network, for instance by using an unweighted projection of a weighted network by applying a threshold to edges. In the case of this thesis, a directed-weighted network will be constructed for the analytical part. This implies that each edge/link carries information on the strength of the relationship, as well as its directionality. For the case of analyzing global FDI and portfolio investment patterns, both directionality and strength matter. Therefore, directed-weighted networks are constructed to define and explain the underlying structure and identify the most central countries and their characteristics.

Often, it is the case that we want to identify the role of each actor in the overall system through its position in the network. In network analysis, this is done by estimating the centrality of nodes. Several measures exist for that purpose. For instance, one can argue that the most central/important node is the one with the highest degree/connectivity, which is the number of links that are connected/connect to a node. In the case of a weighted network, we often measure the strength, instead of the degree, as the total weight of incoming/outgoing edges. Moreover, in a directed network the measure can be analyzed separately into incoming and outgoing connections. Hence, allowing us to define three measures: degree centrality, in-degree centrality**,** and out-degree centrality. The standardized formula for degree centrality is:

$$C_D(a) = \frac{v_a}{n-1} \tag{2}$$

Where v_a is the number of nodes a is connected to, and n is the total number of nodes in the network. In terms of in/out-degree centrality, the formula follows the same logic, except taking into consideration only incoming or outgoing edges in the numerator.

However, the number of connections an edge holds can tell little about the role of a node in mediating information between different parts of the network. To that end, betweenness centrality assumes that a node is more important the more shortest paths (paths connecting pairs of nodes in the network) it mediates. The higher the betweenness centrality of a node the more central/relevant it is. The formula for betweenness centrality is:

$$C_B(a) = \sum_{i,j \in A} \frac{\sigma(i,j|a)}{\sigma(i,j)} \tag{3}$$

Where a is the node (country), $\sigma(i,j)$ represents the number of shortest (I, j)-paths, and $\sigma(i,j|a)$ is the number of shortest paths passing through node a, other than i, j. If $i = j$ then $\sigma(i,j) = 1$, and if $a \in i,j$ then $\sigma(i,j|a) = 0$.

Considering the distance between nodes in a network, closeness centrality measures the importance of a node depending on how close it is to the other nodes in the network. The closest it is, on average, the more central it is. Its formula is:

$$C_C(a) = \sum_{v=1}^{n-1} \frac{n-1}{d(v,a)} \quad (4)$$

Where $d(v,a)$ is the shortest path distance between v and a, and n is the number of nodes that can reach a.

Lastly, eigenvector centrality measures the centrality of a node by the probability of a random-walker spending time in that node if the walker is allowed indefinitely to move through the network by traversing its edges. It offers a dynamical measure of centrality, in that we are saying that a node is more relevant the more time a random flow of information transverses it.

$$Ax = \lambda x \quad (5)$$

Where A is an adjacency network and λ the eigenvalue.

Figure 3, visually shows how these different measures of centrality can classify different nodes as the most central, thus highlighting that each one plays a different role on different structural dimensions.

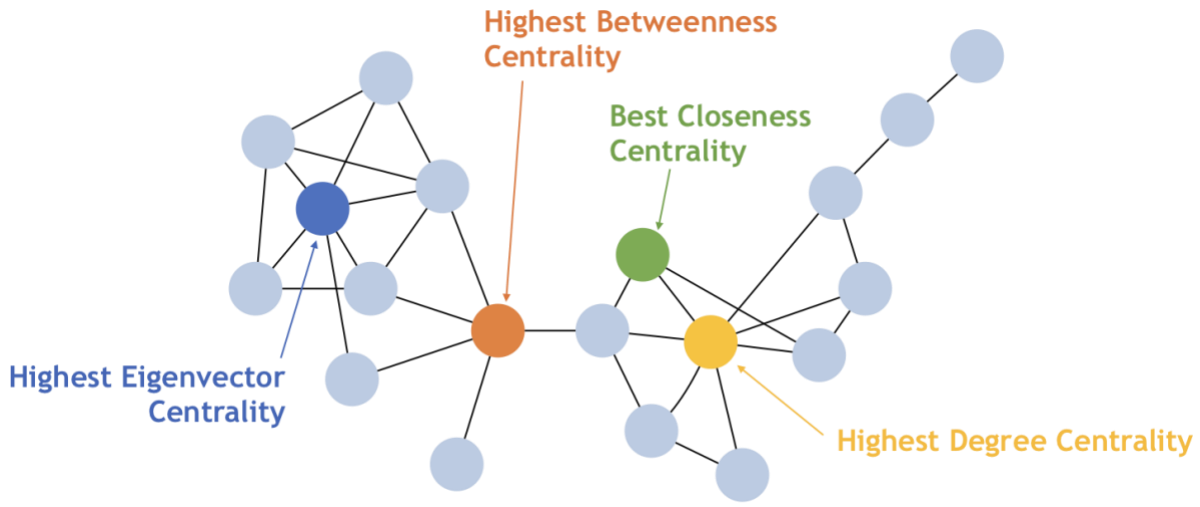


Figure 3 – Simple network with the most important centrality measures Source: (Lima, Pinheiro, Silva, & Matos, 2020)

In this thesis, we will use nodes to abstract countries and edges to identify financial relationships between pairs of countries. Moreover, edges will represent revealed proximities between countries. We will perform all computations in the entirety of the network (e.g., node centrality and shortest paths), with all its links, however, for visualization purposes (because the networks are very dense) we will represent only the most relevant edges.

To that end we shall follow the following steps: 1) we identify the Minimum Spanning Tree, which is a set of edges that ensures all nodes are interconnected while minimizing the sum of proximities between the selected edges; 2) then we enrich the Spanning Tree with the edges that identify the closest relationships until we reach a minimum average degree of 3.5 links, which we take as a thumb rule for a network density that would allow for interpretable network visualization.

5 Hypothesis Formulation

The discussion so far gives rise to the formulation of the main questions this thesis aims to answer. Can we use centrality measures generated by applying network analysis to track the positions of the world countries over time? More specifically, can the results of network analysis applied on each indicator of the CDIS and CPIS datasets be used to answer questions such as:

- Which countries have intermediated the most inward direct investment and portfolio investment liabilities paths for 2009 and 2019?
- Which countries experienced the highest rank variance over the last 10 years as top intermediators?
- What is the shortest investment path between, for instance, China and the U.S. or any other country pair?
- What is the breakdown of inward FDI investment per country, grouped by continent?
- What countries are the top direct investment and portfolio investment intermediators globally?

The following sections aim to extract the results of the analysis and answer the formulated question and in the discussion on how the web app can provide a tool for regulators and central bank officials to quickly obtain answers to those questions.

6 Results

In this section, we use network analysis to present and discuss the results that answer the questions formulated in the previous section. Although, there are many different combinations of networks, years, countries, and measurements all results can be found in the web app developed during this thesis (see https://zealous-noether-a3ce47.netlify.app/).

Figure 4A and 4B show the spearman correlation between the betweenness and closeness centrality values of countries in the year 2019. We show that country rankings are highly correlated across networks, a result that was expected. **Figure 4A** shows, for instance, that CDIS Inward has an almost perfect correlation (0.99) with Combined Liabilities, while CPIS Liabilities is substantially lower (0.28). Since CDIS Inward and CPIS Liabilities are the building blocks of Combined Liabilities, their respective correlations indicate that the CDIS Inward is much more influential in the aggregation.

Looking at the Combined Assets, the observed correlations are stronger. In this case, both CDIS Outward and CPIS Assets have the same weight towards the Combined Assets network. Another observation is the strong correlation between both betweenness and closeness centralities in the CDIS Inward and Outward datasets. Hence, it allows us to conclude that the same countries intermediate the most investment paths in the outgoing and incoming investments while being the closest to the other countries in their respective networks.

Another conclusion that can be taken out of the correlation analysis is that betweenness and closeness are very similar across the 6 different networks under study. The closeness rank correlations are slightly stronger. This shows that there are a few most central countries that hold the highest rank positions for these two centralities, and that play a more key role in foreign

investment. In other words, a few countries including the United States, Netherlands, Luxembourg, China, Hong Kong, and the United Kingdom are the main global intermediators, which will be justified throughout this section. These countries can usually be characterized by being a global economy such as the U.S. and China or a tax haven such as Luxembourg, the Netherlands, and the Cayman Islands.

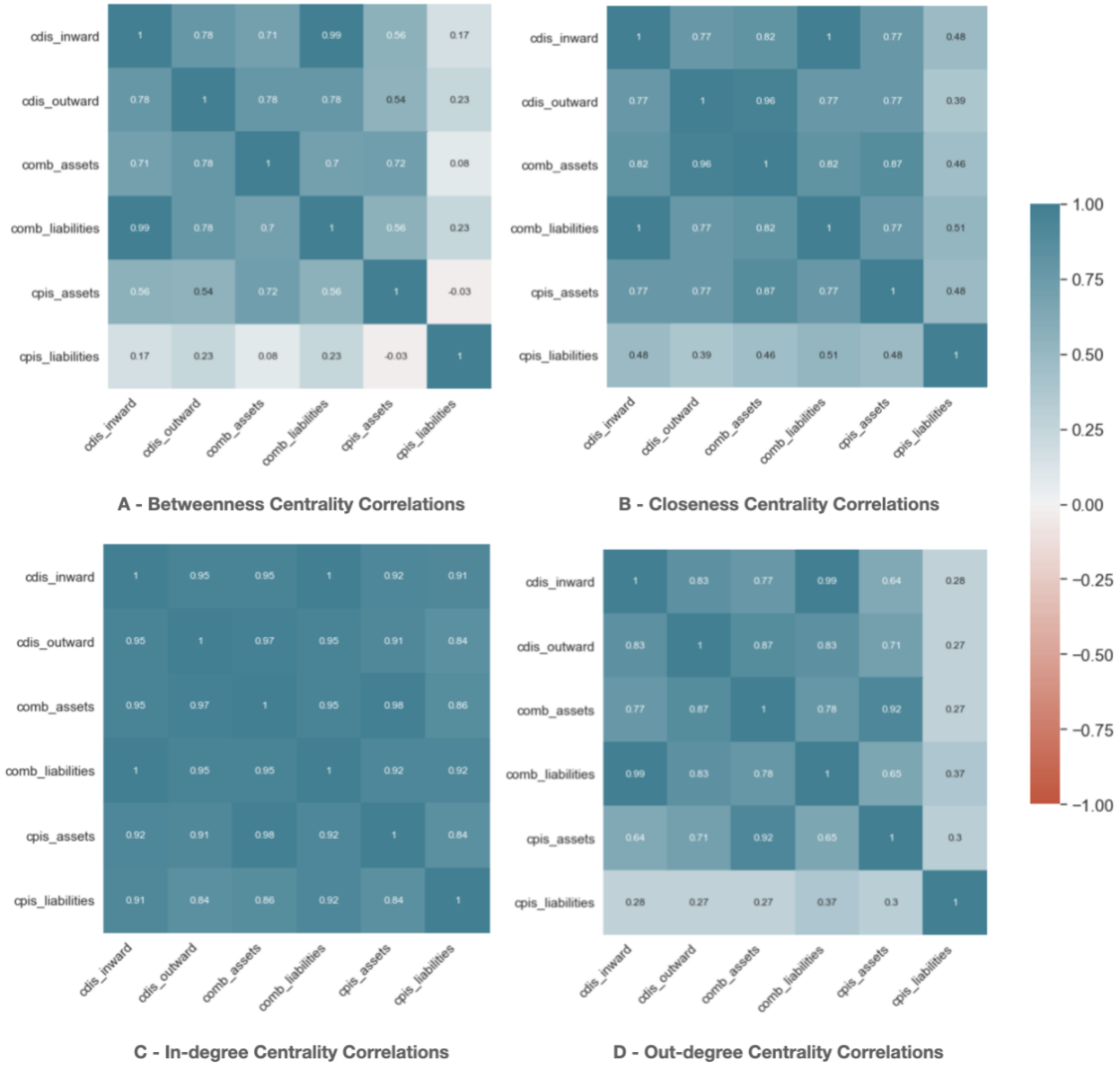


Figure 4 – Spearman Correlation Matrices for Betweenness (A), Closeness (B), In-degree (C), and Out-degree (D) centralities in the Networks estimated for the year of 2019.

In addition to the spearman correlations based on the betweenness and closeness centralities, we also present the spearman correlations based on the in and out-degree centrality values across all networks. **Figure 4C** highlights the correlations based on in-degree centralities, showing a nearly perfect correlation between all networks. This implies that no matter what network we look at, the ranks based on incoming investments are mostly the same. **Figure 4D** on the other hand follows a similar pattern to the betweenness centrality correlations across networks.

Leading to the first question of this thesis, is it is possible to use centrality measures to track the positions of the world countries on specific datasets over time? Our results show that by applying network analysis and using the centrality measures, it is possible to track the position of countries and their importance in a global investment network. Additionally, it is possible to understand the most important countries in terms of facilitation of investments, their importance at a specific point in time but as well as their evolution over time. Onwards, we will use these network centrality measures to track the position of the most influential countries between 2009 and 2019.

In this section, we will focus on the inward FDI, liabilities PI and the combination of both, to show the patterns of the international investments and to identify the international sources of investments (a similar analysis could be performed on the assets side). It is important to mention that the results obtained through the FDI inward and PI liabilities are not the same as FDI outward and PI assets due to lack of data and data imperfections.

CDIS Inward Network

We start by focusing on the CDIS Inward, which is shown in **Figure 5**. It is noticeable that some of the biggest global economies are present in the top ranks for the centralities throughout

the years. In 2009 the top intermediators are the Netherlands, the U.S., China, Canada, and Russia in that order. Similarly, in the 2019 network, the Netherlands and the U.S. hold the top two positions, while third-placed Luxembourg is followed by Hong Kong and China.

In the case of the Netherlands, it is interesting to point out that it has substantial inward direct investment flows relative to its GDP, which can be explained by its lenient corporate tax laws. The position of the U.S. on the other hand can be explained by its global economic strength, as well as it being an innovation hub that has attracted and produced some of the most successful multinational corporations. The position of Hong Kong, and its rise as the fourth most intermediator of FDI, implies that it overtook Singapore's position to become an influential player in East and South-East Asia as well as globally. Nonetheless, Singapore continues to play an important role in the region too.

Shifting focus on the highest-ranked countries based on closeness centrality, the results are somewhat different. The highest-ranked countries for 2009 include Bermuda, Netherlands Antilles, Samoa, Cayman Islands, and the U.S. Similarly, the top five ranked countries for 2019 are Samoa, Cayman Islands, Bermuda, Jersey, and the British Virgin Islands. Without surprise, most of these countries are well-known tax havens and practice lenient tax laws. In that sense, considering that closeness centrality highlights nodes that are on average closer to all the other nodes in the network, it is in line with our expectation as such countries should offer an efficient route to spread investment from a source investor that is looking to spread its investments to many destination countries.

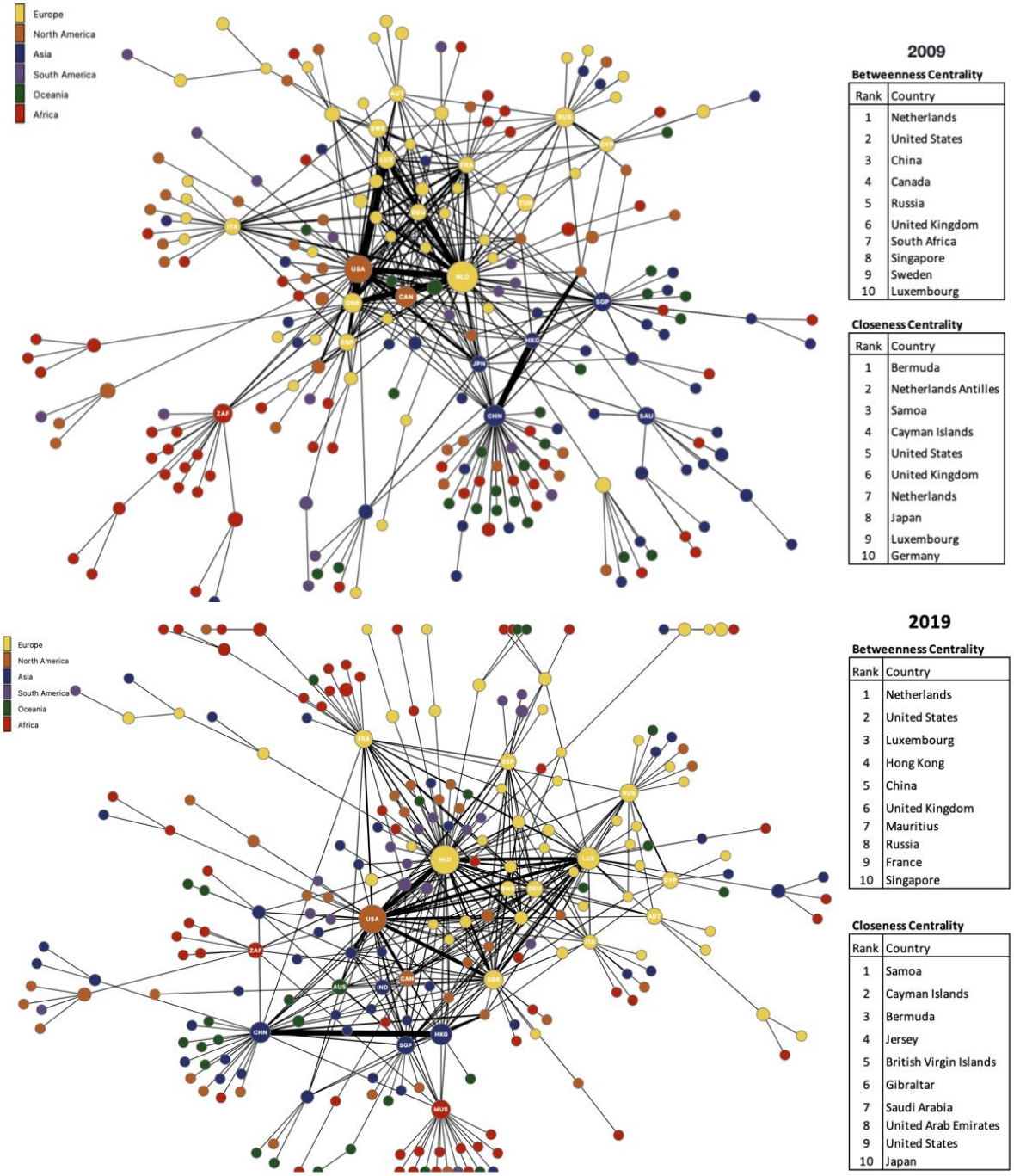


Figure 5 – CDIS Inward Network with node size representing relative amounts of Betweenness Centrality for 2009 (top) and 2019 (bottom). Source: web app

Nonetheless, the Netherlands and Luxembourg, the top intermediators, rank, respectively, 12th and 13th in closeness centrality, while the U.S. is in the 9th position. Additionally, based on the out-

degree centrality the same countries tend to stay in the top list for both 2009 and 2019, a ranking that is occupied by Italy in the top position followed by China, Thailand, and Bulgaria. In contrast, the in-degree centrality ranking lists the U.S. in the top position followed by the U.K., France, Switzerland, and the Netherlands. In both in- and out-degree rankings the top countries are generally not very volatile in their rank.

The treemap in **Figure 6** further reinforces the finding that several countries are the top facilitators of global trading roots in terms of FDI, by continents. The U.S., Netherlands, China, and Luxembourg combined make up 42.57% of global FDI inward investments for 2019.

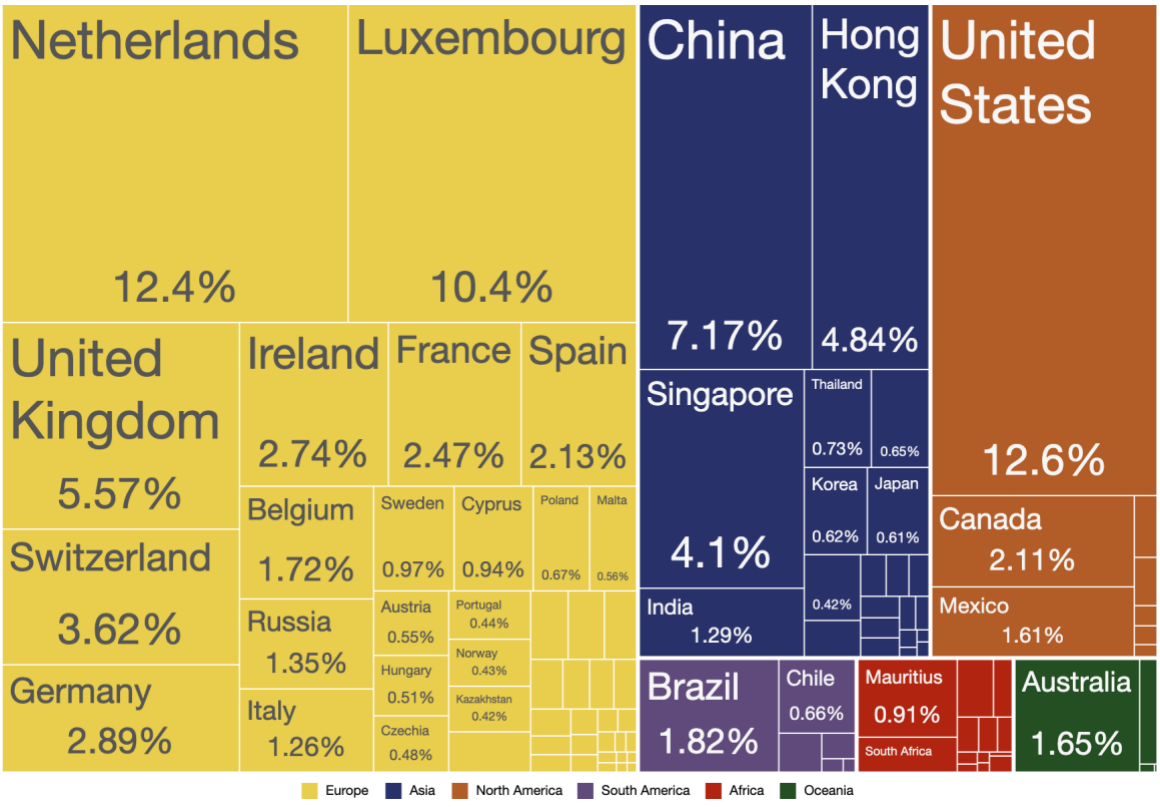


Figure 6 – CDIS Inward Treemap with block size representing the relative amount of FDI investment per country for 2019, with the colours representing the different continents. Source: web app

With China as a global manufacturing powerhouse and the U.S. as an innovation hub and both as one of the biggest world economies, they form one of the most important bilateral relationships

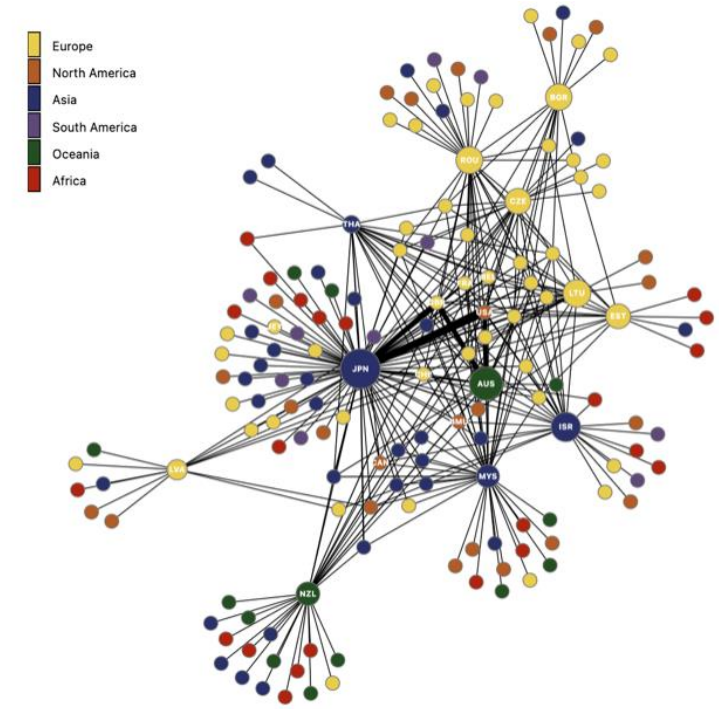
in the world. Their political and economic ties can be studied through the shortest paths of investment generated by the CDIS Inward Network. **Figure 7** represents the evolution of this relationship between 2009 and 2019. From 2009 up until 2017 Japan was the only intermediator in this relationship. While from 2018 to 2019 the U.K. and Hong Kong became intermediators between China and the U.S.

Figure 7 – CDIS Inward Shortest Path Visualization, where each node represents a country as part of the shortest paths of investment between the U.S. and China.

Portfolio Liabilities Investment Networks

Let us now focus on the portfolio liabilities investment networks. In 2009 the most important intermediators were: Japan, Australia, Israel, Lithuania and Romania. The visualizations in **Figure 8** represent the CPIS Liabilities network in the years of 2009 and 2019 where the top intermediators are shown according to their node size, based on betweenness centrality. For 2019 these are Japan, Australia, Finland, Estonia and Bulgaria. The structure of the CPIS Liabilities network is significantly different to the CDIS Inward network. Here Japan and Australia are the most dominant in terms of intermediary power based on betweenness centrality. Japan is consistently ranked 1st throughout the entire year range, while Australia has never dropped below 5th position, as shown in the rank evolution chart in the web app. Australia's rank falls to 5th place only in 2013 and is outranked by Israel, Romania and Lithuania. Furthermore, they are, respectively, part of 74% and 46% of all shortest investment paths in 2019. This showcases their ability to attract short-term

investments aimed at financial gain, by retail or institutional investors. Excluding, Finland the top four intermediators for 2019 have never ranked below 9th between 2009 and 2019.



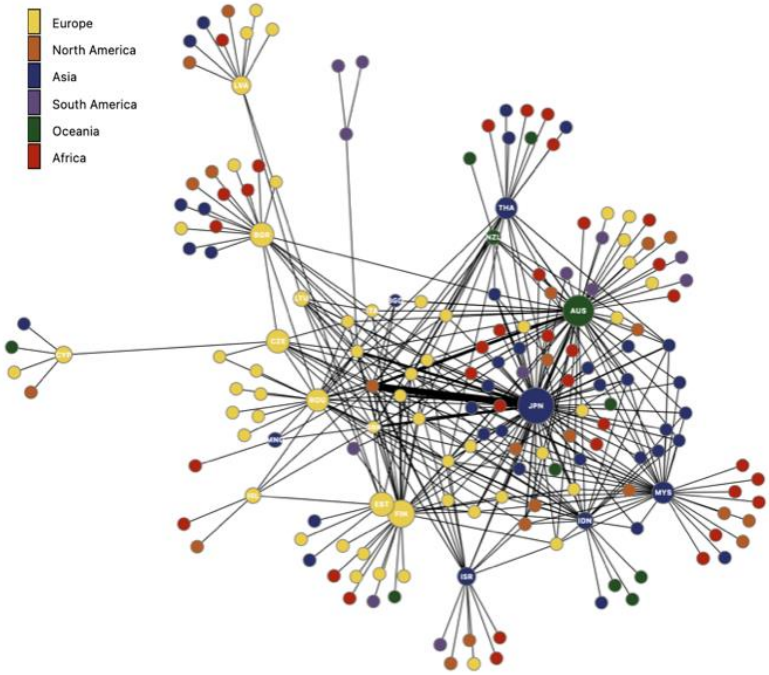
2009

Betweenness Centrality

Rank	Country
1	Japan
2	Australia
3	Israel
4	Lithuania
5	Romania
6	Bulgaria
7	Czechia
8	Estonia
9	New Zealand
10	Malaysia

Closeness Centrality

Rank	Country
1	United Kingdom
2	United States
3	Germany
4	Luxembourg
5	Austria
6	Netherlands
7	France
8	Sweden
9	Switzerland
10	Italy



2019

Betweenness Centrality

Rank	Country
1	Japan
2	Australia
3	Finland
4	Estonia
5	Bulgaria
6	Czechia
7	Romania
8	Malaysia
9	Thailand
10	Latvia

Closeness Centrality

Rank	Country
1	United States
2	Luxembourg
3	Ireland
4	Netherlands
5	Singapore
6	Saudi Arabia
7	British Virgin Islands
8	Taiwan
9	Kuwait
10	Mauritius

Figure 8 – CPIS Liabilities Networks with node size representing relative amounts of Betweenness Centrality for 2009 (top) and 2019 (bottom). Source: web app.

It is worth mentioning that the U.K., the Netherlands, and the U.S. are top ranked 19th, 21st and 23rd in terms of betweenness centrality rank. Their rank in terms of betweenness centrality has not been below 31 in the past 10 years. While Finland who ranks 3rd in 2019 became an important portfolio liabilities investment intermediator only in 2015.

The strength of the top 3 countries in terms of portfolio liabilities investments can further be reinforced by looking at the proportion of their portfolio liabilities investment relative to the total for 2019. **Figure 9** shows that the Japan is the most dominant, with a 63.5% of the total portfolio liabilities investments for 2019. It is followed by Australia and Finland. These countries make up 87.76% of total portfolio liabilities investment for 2019.

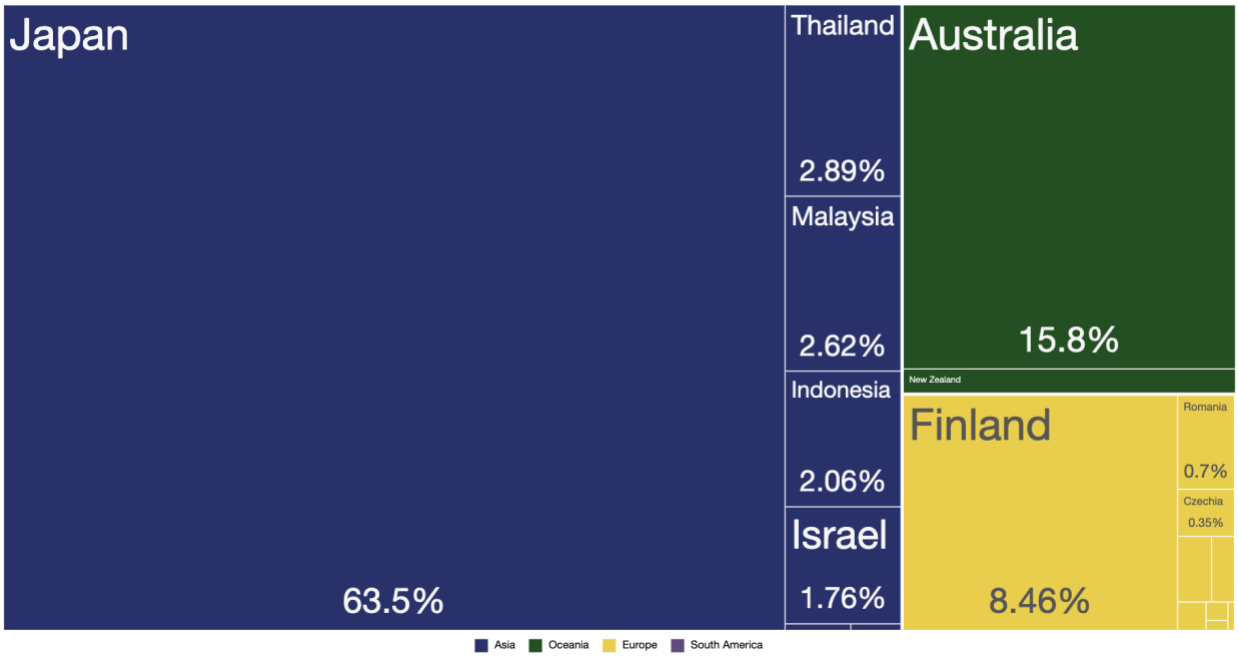


Figure 9 – Treemap representation of the total CPIS Liabilities broken down by country and continent for 2019. Source: web app.

In terms of closeness centrality, there are some more differences as compared to the CDIS Inward network. The top countries for the CPIS Liabilities in 2019 include the U.S., Luxembourg,

Ireland, the Netherlands, and Singapore. Although, both the CDIS Inward network ranks predominantly offshore centres the highest based on closeness centrality measures. The CPIS Liabilities network ranks the U.S. the highest, but also includes tax heavens like Luxembourg, Ireland, the Netherlands and Singapore. The only common countries in the top 10 are the British Virgin Islands and Saudi Arabia. In both cases, some of the main world economies including China, Germany, France, and the U.K. rank outside of the top 90.

Comparing the shortest paths of investment between the U.S. and China for CDIS Inward and CPIS Liabilities, there is a slight difference. As **Figure 10** indicates, Japan acts as the intermediator between the two major economies throughout the entire year range.

Figure 10 – The shortest path between the U.S. and China for CPIS Liabilities over between 2009 and 2019. Source: web app.

Concerning the in and out-degree centralities, there are some differences. The out-degree centrality identifies Bulgaria, Malaysia, Japan, Mongolia and Latvia as the 5 countries with the least amount of variance in their rank between 2009 and 2019. They have never been below the 15th rank. In terms of in-degree centrality the countries with the least variance include Germany, the Netherlands, the U.S., Luxembourg and Switzerland.

Combined Liabilities Network

In the combined liabilities network we obtain a similar network structure to the examples discussed in Section 5.1. Both in 2009 and 2019, the United States, the Netherlands, Luxembourg and China are ranked the highest as intermediators based on betweenness centrality. In fact, the

top 10 countries are the same for both the CDIS Inward and Combined Liabilities networks, except for Hong Kong and Japan. In comparison, the CDIS Inward network ranks Hong Kong as a more important intermediator; while the Combined Liabilities network ranks Japan higher. Nevertheless, its structure is very similar with the CDIS Inward network, which is to be expected since it is one of the main building blocks of Combined Liabilities. The top countries in terms of betweenness centrality, are stable in the highest ranks, with Luxembourg being the only one that ranks lowest at the 15th position between 2009 and 2019.

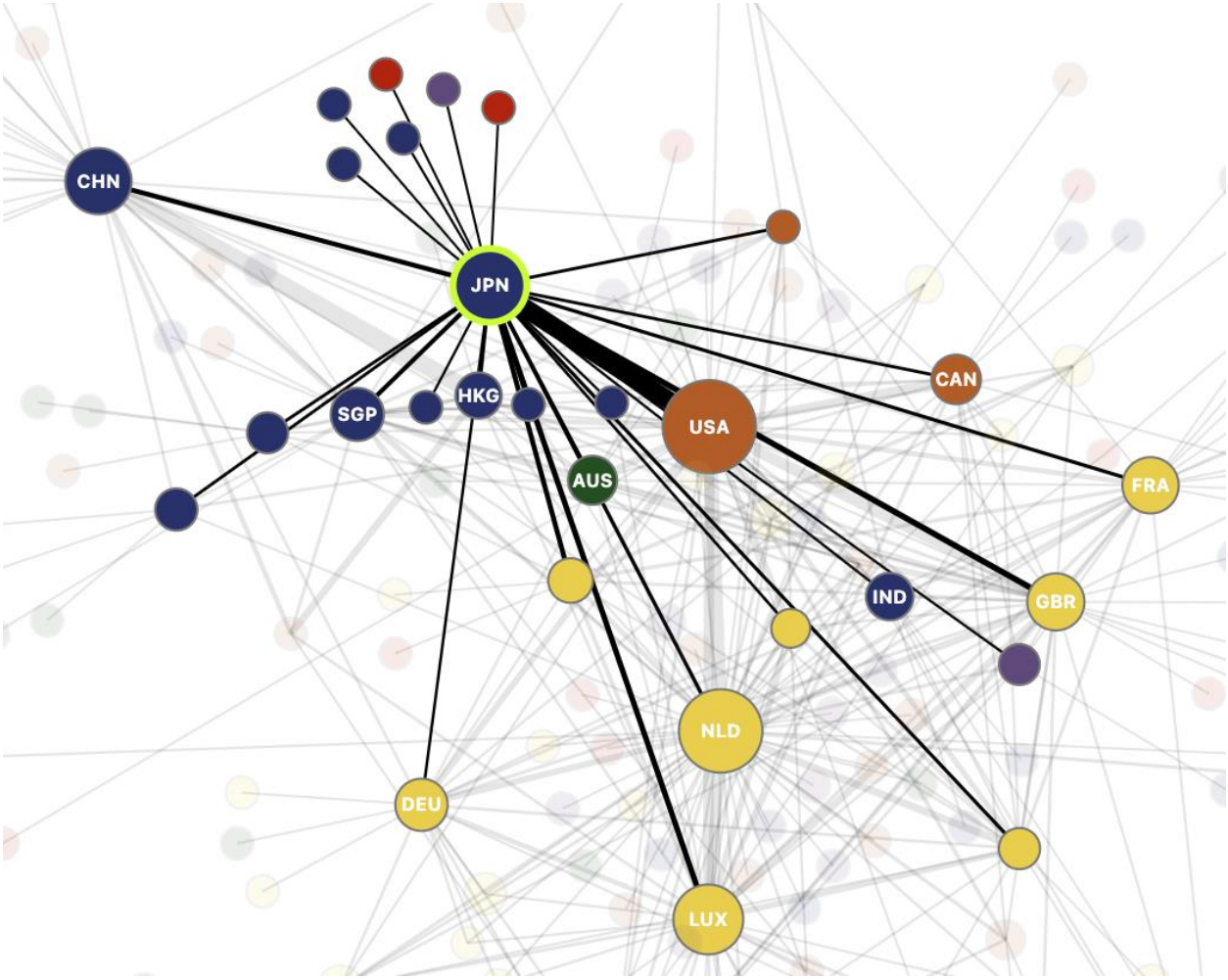


Figure 13 – Combined Assets Network with node size representing relative amounts of Betweenness Centrality for 2019, where the Japan and its connections are highlighted. Source: web app

Japan's position is not surprising since it is known to be one of the most indebted countries in the world, hence its high rank in a network related to liabilities. **Figure 13** highlights Japan's position in the Combined Liabilities network and shows its role as a connector between Asia, Europe and North America. Furthermore, the shortest path graph in the web app, showcases that Japan is the main intermediator between the U.S. and China between 2009-2019. Its dominant position in the CPIS Liabilities network further reinforces Japan's influence here, since CPIS Liabilities is the other building block of the Combined Liabilities network. This is also reinforced by Japan's proportion of total Combined Liabilities, adding up to 9.18% as shown in **Figure 14**.

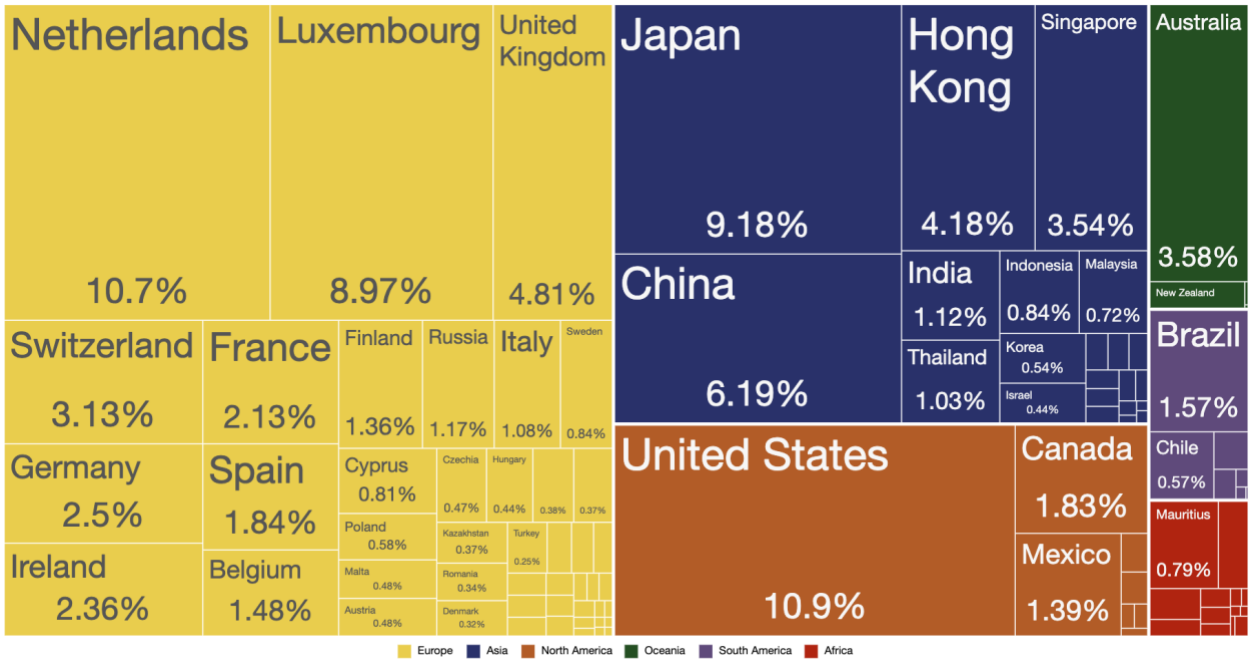


Figure 14 – Treemap representation of the total Combined Liabilities broken down by country and continent for 2019. Source: web app.

In terms of intermediation, the Netherlands 66% of the shortest paths of investment, while the U.S. 56% of the time and Japan 43% for 2009. Compared to 2019 the U.S. and the Netherlands switch positions, where the U.S. is part of 67%, the Netherlands is part of 47% of the shortest paths of investment; while Japan is thirds with 31%.

With regards to the closest countries within the network, the countries with lenient tax laws are dominant in 2019. These include Samoa, Cayman Islands, Bermuda, Jersey, and the British Virgin Islands. In 2009 similarly, Samoa, Bermuda, Netherlands Antilles, Cayman Islands and British Virgin Islands are ranked highest. These results are closer to the outputs of the CDIS Inward network, although all three networks offshore centers are the closest to the rest of the countries. This is expected as they are able to efficiently spread investments through the network.

The most important countries in terms of out-degree centralities for 2019 are Italy, China, Thailand, the Netherlands, and Bulgaria. That is the same top 5 as the CDIS Inward network. Similarly for in-degree centrality, the U.S., U.K., France, Switzerland, and the Netherlands rank the highest in both networks and have a stable rank never dropping below 6th position over the 2009 to 2019 period.

Summary

Throughout the results section we have shown the findings and comparisons of three different networks, namely CDIS Inward, CPIS Liabilities and Combined Liabilities networks. This discussion has proven the possibility of applying network analysis and using the centrality measures, to track the position of countries and their importance in a global investment network. Additionally, we have shown that it is possible to understand the most important countries in terms of facilitation of investments, their importance at a specific point in time but as well as their evolution over time.

7 The Platform

On top of the analytical perspective, the main product of this thesis was the development of a web application that allows to 1) communicate these findings in an interactive data visualization platform; and 2) allow for the quick exploration and discovery of relevant partnerships and investment paths.

In this section, we detail the steps conducted and their rationale for building interactive and visualization-driven web-applications for exploration of the networks of foreign investment. In that sense, we start by explaining the technological stack used, the narrative employed, the choice of particular visualization to support the narrative, the final product, and future steps.

The Stack

Traditionally data has been communicated and shared in a tabular format. Indeed, many open data platforms these days still rely on that format and thus make complex data exploration difficult for the user. Moreover, tabular data representations are not only difficult to interpret but also difficult to compare across years. In contrast, information about size, magnitude, proximity, comparability, and temporal evolution are easier to grasp when data is represented in the appropriated visualization format. In that sense, a common solution comes in the format of a Business Intelligence dashboard, which offers a layout to quickly present multiple relevant visualizations and key indicators to the user and communicate insights to a specialized audience.

Several choices were considered for the visual communication of the results. One option would have been to use an existing framework, such as Power BI or Plotly Dash. The advantages of using such frameworks are the ease of use and the short production times. However, these options usually

consider a rigid canvas for development and offer a limited range of visualization possibilities and customizability. Another significant drawback in using these options is the cost of licensing the software (e.g. Power BI) and hosting requirements (Dash).

The other choice involves the full-stack development of a web app. The advantage of this line of work is that it offers full flexibility in the development of the application to meet the development goals. Additionally, since it is fully developed by the team it creates an asset that can then be further developed and thus bring greater added value. The tools used in this line of development are free and open-source, allowing the final product to be replicated and shared with a wider audience. One of the disadvantages is the time it takes to develop such an application, as well as its complexity, and a steeper learning curve for beginners.

In the context of the master thesis, the choice ultimately fell into the development of a web application from scratch using free and open-source tools.

The exploration analysis of the data was initially conducted in python. All the calculations were performed in Jupyter notebook using common data science libraries, such as Pandas and NetworkX. In these early steps work consisted of data cleaning and export in json or csv formats, which would feed the visualizations to be developed ahead. The specific aspects of each visualization will be discussed in their sub-section.

The web application was built in a stack that comprised aspects of HTML 5, CSS, and javascript. In particular, we used several handy javascript libraries, such as jQuery, D3.js, and d3Plus. For the styling of the HTML elements, we resorted to the bootstrap library, a well-known framework developed by Twitter (see https://getbootstrap.com/). Additionally, we used Font Awesome icons throughout the website to style dynamic actions (e.g., mouse hovering) elements to buttons.

Concerning javascript libraries — D3.js and d3Plus — they operate as DOM manipulators while taking leverage on the use of *Scalable Vector Graphics* (SVG's) that allows drawing shapes in a browser window. The main advantage of using SVG's is that they don't lose quality when they are scaled, and they have low memory requirements. Moreover, SVG elements can be manipulated similarly to other DOM elements.

With regards to the data loading and storage, two options were considered. One was to load the data from a folder structure through json and csv files. The other option was to build a database and fetch the data through a database connection, which would add complexity to the development. The first option was chosen because the data does not evolve and does not need to be constantly updated. Initially, there were some concerns over the loading speed due to large files which would take a long time to render every time there is new user input. The workaround to this issue was to organize the data files in separate years or countries and load a different file for specific selections, which significantly improved the loading speed. For academic purposes, the current choice is sufficient, although if the web app gets developed in a fully-fledged product, an important improvement would be to build a database.

All the technologies combined, allowed us to build a web app and communicate the results across five pages, including a landing, network, methodology, about, and downloads page.

Narrative

This sub-section will outline the most important visualizations per page and justify the choices behind the types of visualizations and their purpose.

7.1.1 Landing Page

The goal of the landing page is to grab the attention of the user and get them engaged. The top of the landing page contains a few statistics cards about the datasets. Their goal is to give the user an overview of the magnitude of the data being used. Below that, 5 teaser questions are formulated that a user might already be thinking about in the context of this framework. Each teaser corresponds to one visualization on the network page. They are all clickable and linked to their respective visualization, which would ultimately answer the question. Taking the first question as an example, "*What is the top direct investment intermediator for 2019?"* would lead to the network visualization and highlight the country that has the top rank for betweenness centrality for 2019. The same logic can be applied to the rest of the teasers, except they take the user to a different visualization that would answer the respective question.

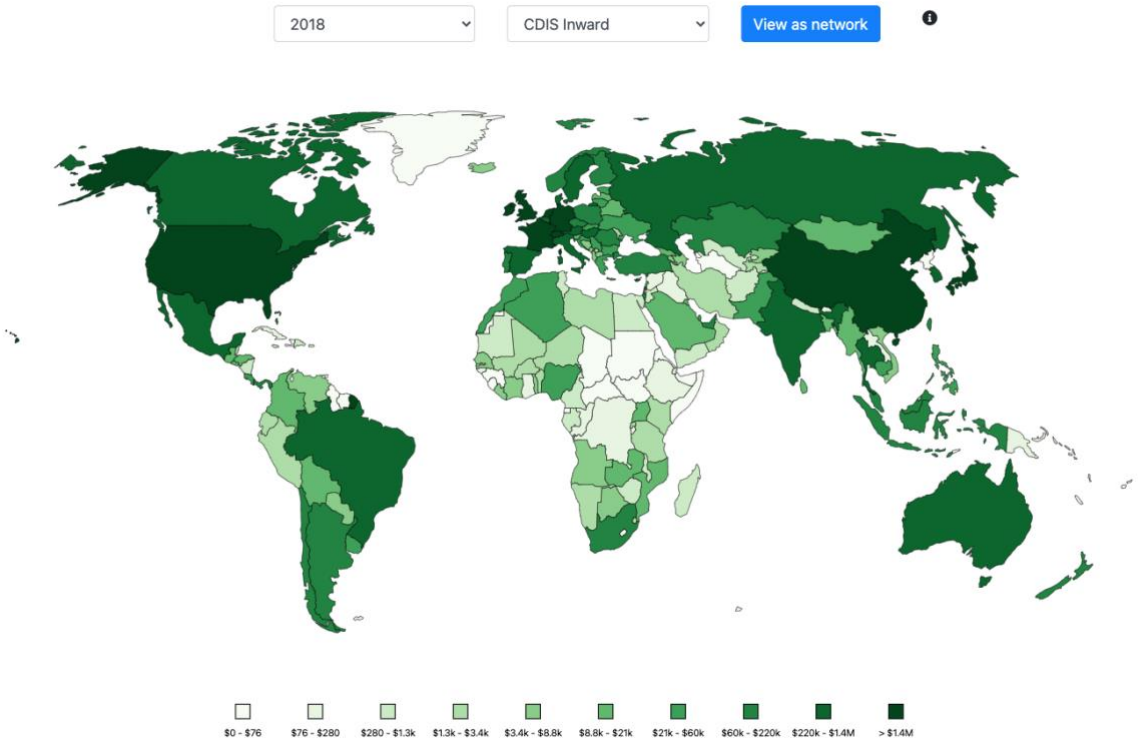


Figure 15 – World map visualization in the landing page shown for 2019 CDIS Inward. Source: web app.

Below the teasers, we present a world map highlighting the total FDI/Portfolio Investment amounts per country using different shades of green. As shown in **Figure 15**, the goal is to give a breakdown based on the total amount of investment. It shows the relative size of investments per country through the shades of green, the darker the color, the higher these amounts. The user has the option to switch between all the datasets and years which change the respective values on the map. The "*View as network*" button takes the user to the network page for the latest selection of year and dataset in the dropdown menus. This adds more interactivity and another direct link to the main visualization of the web application.

7.1.2 Network Page

The network page is the most important one of the web application. Each visualization in this page is broken down into two sections. The selection (left) side contains the selection options which allow users to interact with the visualization, by selecting different networks, years, countries, etc... The options change depending on visualization, although all of them present an option to select the network and year. There are also information icons, which contain information necessary for the user to understand what is presented in the respective visualizations. They also carry a direct link to the methodology page where everything is explained in more detail. The last element of the selection pane has its own set of teaser questions; these are clickable and highlight the answer in the respective visualization. To the right of the selection pane, the visualizations are presented.

To explain the storyline of the web app, the United States will be used as an example to show their position, evolution, and conclusions that can be drawn based on the information presented.

The network visualization is the most complex one and shows the network structures for the different datasets. Each node represents a country, while its size represents the value of the measure

that is selected (betweenness, closeness, in-degree or out-degree centrality). The colors are dependent on the continent. The edges connecting the countries represent an investment relationship between those counterparties, while their thickness is the strength of that investment. Thickness instead of color scale was chosen to represent the strength of the relationships because it makes it easier to observe the relative magnitudes. Since color is already used to distinguish the continent of each country, an additional color scheme would make it more difficult to interpret. It is important to note that the networks presented are not fully spanned. For the purpose of the visualization, a Minimum Spanning Tree (MST) of the network is computed. The main reason for this is that the fully spanned network is extremely dense, hindering the ability to identify the most important connections in that structure. The MST computes a network in which all available countries are connected to at least one other country. However, once computed, the MST itself was too sparse, therefore additional links with the lowest weights (identifying the strongest investments) were added. The goal is to achieve a dense enough structure that includes the most important relationships. The minimum average degree[5] changes case by case and can be adjusted by user input. The calculations are performed on fully spanned directed-weighted networks, even though the visualized networks are weighted MST's with the additional strongest links that reach the 3.5 minimum average degree threshold, and can be increased up to an average degree of 9. The most central countries per selection are highlighted with their ISO3 codes printed inside the nodes. Although, the user can manually search or click a country to highlight it in the network. The construction of the network is done using d3js and its inbuilt functions that allow to apply gravity force algorithms and push the edges around to keep them from overlapping. Clicking the teaser for

[5] Average degree is the average number of connections per node in the network.

the CDIS Inward network in 2019 would highlight the U.S. as the most central country based on betweenness centrality. In case a country is difficult to identify by looking at the network structure, the search menu helps in finding any country in the network.

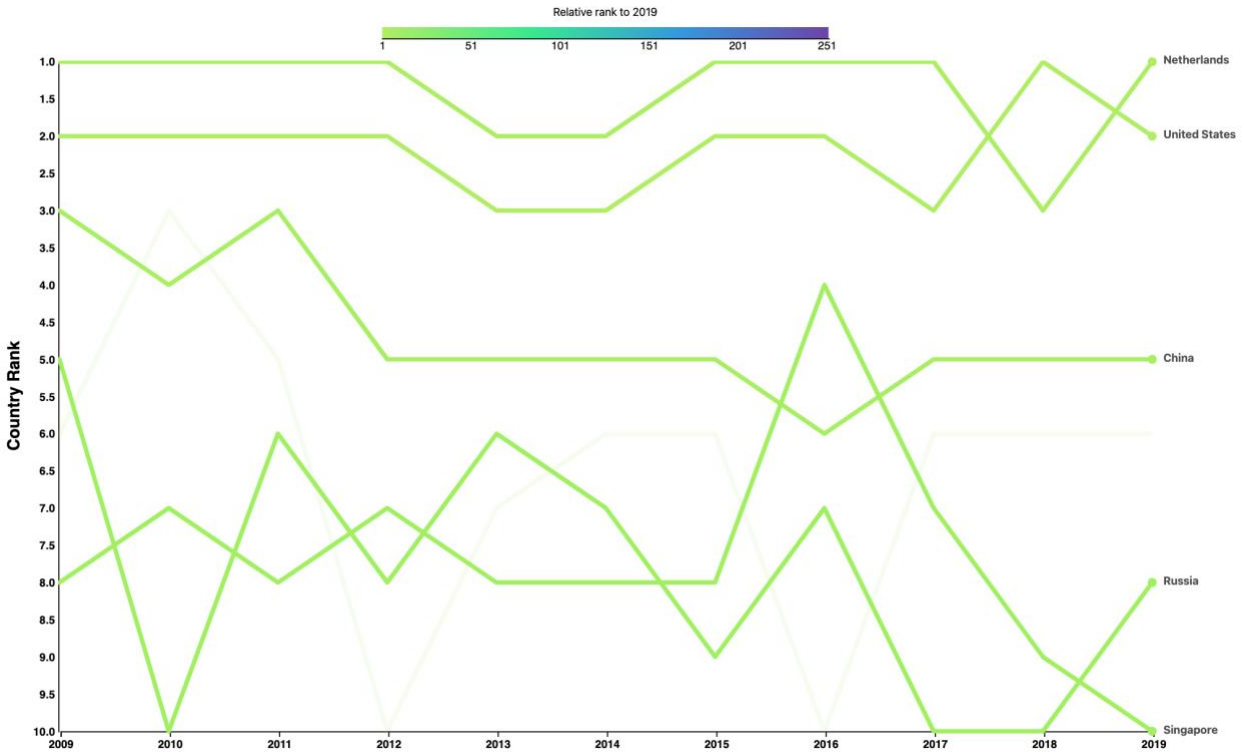


Figure 16 – Rank Evolution Chart between 2009 and 2019 based on Betweenness Centrality, showing the evolution of Netherlands, United States, China, Russia, and Singapore. Source: web app.

The second visualization is the rank evolution chart. Its purpose is to show the selected countries' variability in rank over the available time period per centrality measure. It shows how stable or volatile countries' position is based on the different centrality measures. In **Figure 16**, the 5 least volatile countries are highlighted in terms of betweenness centrality rank for the CDIS Inward network. The color is fixed at the respective country's rank in 2019. This gives the user a reference to the 2019 rank throughout the years. There is an option to select the network, centrality, and countries displayed in the chart. The number of countries that can be highlighted at a time is limited

to 6, in order to avoid congestion and to make it easy to distinguish between the countries. In the case of the U.S. **Figure 16** shows its low variability in terms of betweenness centrality, never dropping below the third position between 2009 and 2019.

Followed by the rank evolution chart is the shortest path chart. It shows the intermediators that make up the shortest path between any two countries dependent on network and year. This graph aims to depict the ultimate investor between a pair of selected countries. It also presents the evolution of the intermediating countries that an investment goes through before reaching the ultimate investor. It is important to note that the paths in this visualization are computed on a weighted graph, and not a weighted directed graph. The main reason is that with the directed weighted networks it was not possible to compute a shortest path between any two countries. **Figure 17**, shows the evolution of the shortest path between the U.S. and China based on the CDIS Inward network. Between 2009 and 2017 Japan was the sole intermediator, while in the last 2 years it has been the U.K. and Hong Kong.

Figure 17 – Shortest path chart between the United States and China between 2009 and 2019. Source: web app.

The purpose of the treemap chart is to show the relative magnitude and breakdown of the total amounts per country and continent dependent on year and network. It is useful for comparing the magnitude of a country's total investments relative to other countries and to its continent for different selections of network and year. It is possible to filter based on the continent, by clicking on it in the legend below the treemap. The 2019 treemap with CDIS Inward data shows that the

U.S. attracts the most investment, followed by the Netherlands and Luxembourg, indicated by the tile size based on total FDI for that year.

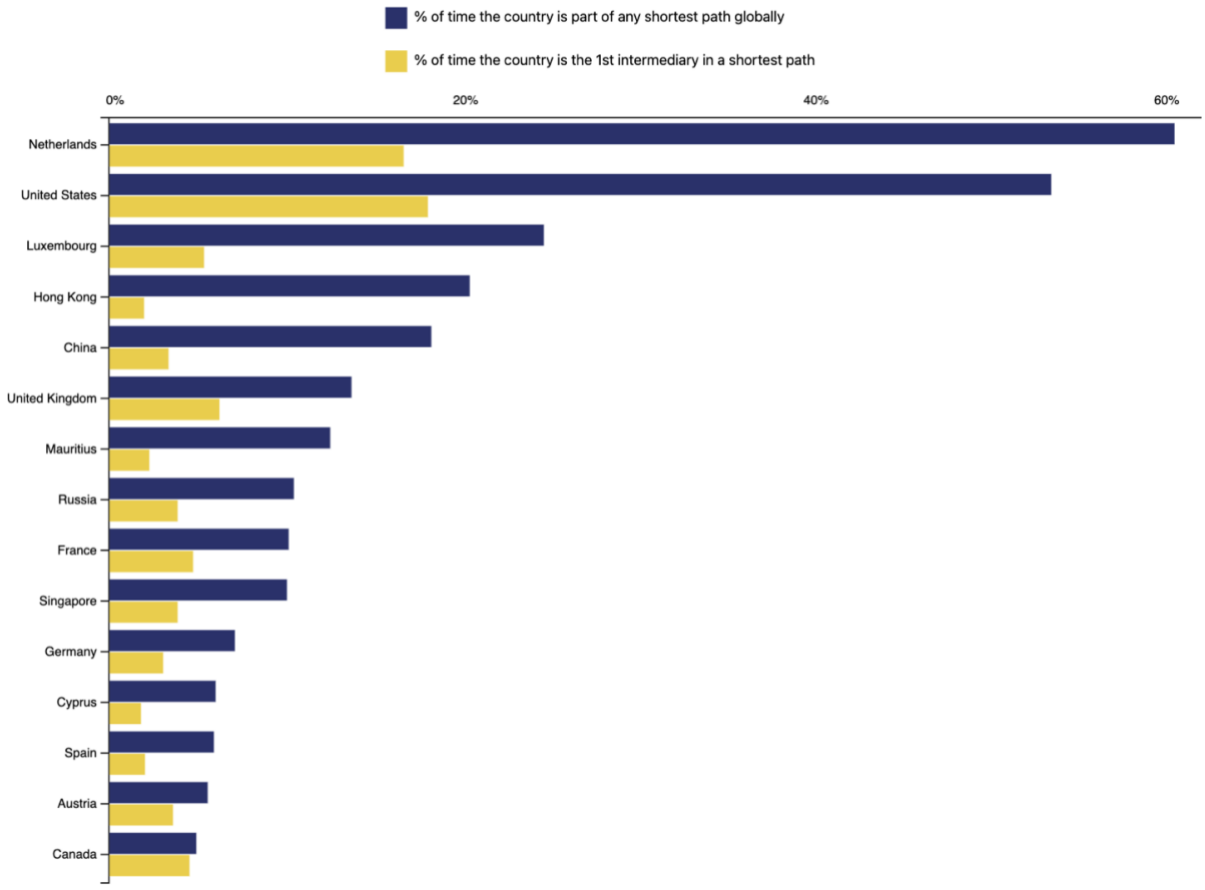


Figure 18 – Horizontal bar-chart chart showing the top intermediators based on the percentage of times they are part of a shortest path (blue bar) and percentage of time they are the first intermediator (yellow bar). Source: web app.

The top intermediators chart shown in **Figure 18** aims to identify the countries that most often facilitate investment paths globally. It ranks the top 15 countries by the proportion of paths they intermediate, based on two measures. One is the percentage of times a country is part of a shortest path of investment globally. While the other identifies the percentage of times a country is the first intermediator in the shortest path of investment. The users can explore this information based on all networks and years. It is also possible to isolate and sort by only one of the measures, or both.

The U.S. in the selection for CDIS Inward in 2019 is in the second position as top intermediator, being part of 53.73% of shortest investment paths (blue bar). While being the first intermediary 18.17% of the time (yellow bar).

The last section of the network page presents all the networks results in tabular form. It presents each centrality measure, their rank, and values, for all available networks and years. Here, the users have the option to download the data for further exploration.

7.1.3 Methodology

The methodology section outlines all the information a user needs in order to understand everything that is presented throughout the web application. It is broken down into several sections, mainly one per visualization. Information such as definitions of terms and data, as well as an explanation of what network analysis is and the reasoning behind data aggregations and transformations, is also presented on this page.

7.1.4 About and Downloads

The about section gives the users a summary of the web application, its purpose, and motivation. It also gives an introduction to the authors and developers.

The downloads page has several links that lead to the source code, datasets, and related institutions.

8 Conclusions

The results obtained in this thesis allow us to answer the main research question. The simple answer is, it is possible to use the centrality measures obtained by applying network analysis to track the position of countries over time.

More specifically, global economies and strategic partners such as the United States, China, Hong Kong, Netherlands, and Luxembourg rank the highest in the different networks in terms of intermediary power based on betweenness centrality for 2019. The results also show that the variation in these ranks is not the same for all the listed countries. Hong Kong and China, for instance, have become some of the most influential intermediators only in the last several years. While the U.S., Netherlands, and Luxembourg are in the top 20 ranks for most of the years between 2009 and 2019. In terms of closeness centrality, countries characterized by lenient corporate tax laws such as the Cayman Islands, Bermuda, Jersey, the and British Virgin Islands are among the top ranks in the last few years. Lastly, the in/out-degree centralities rank the biggest global economies the highest without significant volatility between 2009 and 2019.

In addition to the analytical perspective, the web application has proven to ease the tracking of individual countries' positions in global investment networks over time. It gives central bankers and policymakers an easy way to identify underlying investment paths from a global investment network that is built based on CDIS and CPIS. One of the advantages is that the networks and measures are pre-computed, and their results can be quickly rendered on the page. Additionally, the visualization of the results makes it easier to understand and extract valuable information from it.

Therefore, this thesis showcases the possibility of applying network analysis to analyze large datasets and produce significant results. More specifically, the power of network analysis to use immediate counterpart CDIS and CPIS data to trace the ultimate investors, as well as analyze any country's rank, influence, and connectedness in a global network. Lastly, the combination of the network analysis results with complex visualization techniques proves their ability to provide useful insights on economic analysis of international investment relationships such as those based on Foreign Direct Investment and Portfolio Investment.

Even though, this thesis has been a success it also has some limitations. From an analytical point of view, some new insights could be gained by analyzing a larger timeframe. Interesting results could be obtained when looking at the differences in the networks between certain events such as the financial crisis of 2008/2009 or very recently the structure of the networks before and after the Covid-19 pandemic. As a future improvement, it could also be good to collect the data on an ultimate counterpart basis and compare the models generated from this thesis to see how well it identifies the underlying investment paths. Currently, the analysis is focused on FDI and portfolio investment relationships between countries as a whole, although zooming in and collecting data based on types of investments differentiated by the sector in which the investment is part of would add another dimension. Another drawback of the thesis is in the way Combined Liabilities and Combined Assets values are created. They are a combination of CDIS and CPIS indicators, while the indicators used for their creation are recorded on different principles – namely directional and asset/liability principles respectively. Therefore, for future expansions as an improvement, it would be better to collect the CDIS data on the asset/liability principle for the combination of the indicators to be more significant. The data exhibits also some limitations as the total amount of the assets/liabilities is not equal due to the lack of information and not full coverage.

In terms of the platform, it is currently developed as an academic project although it has the potential to become a fully-fledged product. The data for this thesis is static and is not evolving, hence there was no necessity to store and fetch the data from a database. Currently, the data is stored in a folder structure and is fetched as such in the web application. For further improvement and as a best practice it would be advisable to create a database. It also has the potential to include even more visualizations that would add value to the users.

Lastly, user experience surveys and interviews have not been performed due to time limitations. Such exercises would be very useful in better understanding the positive and negative aspects of the design and communication in the web app. It would be interesting to gather this information from experts in the fields, such as central bankers and policymakers.

9 Bibliography

Barabási, A.-L. (2013). *Network science.* Philosophical Trasactions of the Royal Society A.

Bolívar, L. M., Casanueva, C., & Castro, I. (2019). Global Foreign Direct Investment: A network perspective. *International Business Review*.

Branco, J. (2017). *Shedding light on a liberalized market.* Lisbon: Universidade Católica Portuguesa.

Chuluun, T. (2016). Global portfolio investment network and stock market comovement. *Elsevier*.

Damgaard, J., & Elkjaer, T. (2017). *The Global FDI Network: Searching for Ultimate Investors.* International Monetary Fund.

E., B., De Gregorio, J., & Lee, J.-W. (1997). How does foreign direct investment affect economic growth? *Journal of International Economics*.

Egger, P., & Pfaffermayr, M. (2004). The impact of bilateral investment treaties on foreign direct investment.

Haberly, D., & Wójcik, D. (2015). *Regional Blocks and Imperial Legacies: Mapping the Global Offshore FDI Network.*

Hafner-Burton , E. M., Kahler, M., & Montgomery, A. H. (2009). Network Analysis for International Relations. *Cambridge University Press*, 559 - 592.

Hakeem, M. M., & Suzuki, K.-i. (2016). Foreign Portfolio Investment and Economy: The Network Perspective. *International Conference on Business, Economics, Management and Marketing.*

IMF. (2009). *BPM6 - Balance of Payments and International Investment Position Manual 6th Edition.* Washington, D.C: International Monetary Fund.

Jackson, M. O. (2008). *Social and Economic Networks.* Princeton University Press.

Lima, F., Pinheiro, F., Silva, J. F., & Matos, P. (2020). Foreign direct investment – using network analysis to understand the position of Portugal in a global FDI network.

Norgren, A., & Olsson, M. (2021). *Institutional Dynamics in the Global FDI Network.* Linköping.

OECD. (2001, September 25). *Glossary of Statistical Terms*. Retrieved from OECD: https://stats.oecd.org/glossary/detail.asp?ID=2092

OECD. (2008). *OECD Benchmark Definition of Foreign Direct Investment.* OECD.

Prota, L., & Bersford, M. (2011). Emerging Class Relations in the Mekong River Delta of Vietnam: A Network Analysis. *Journal of Agrarian Change*, 60-80.

Scott, J. (1988). Social Network Analysis.

Scott, J. (2003). Networks of Corporate Power: A Comparative Assessment. *Annual Review of Sociology*, 181-203.

Sobel, R. S., J.R., C., & Dwigth, L. R. (2007). Freedom, barriers to entry, entrepreneurship, and economic progress.

Ter Wal, A. L., & Boschma, R. A. (2008). Applying social network analysis in economic geography: framing some key analytic issues.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.