IFC-Bank of Italy Workshop on "Data Science in Central Banking: Applications and tools"

14-17 February 2022

Data science and statistics: a network analysis to understand the foreign investment[1]

João Falcão Silva, Banco de Portugal,
Flávio Pinheiro and Bojan Stavrik, NOVA Information Management School

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

Data science and Statistics: a network analysis to understand the foreign investment

Bojan Stavrik, Flávio Pinheiro, João Falcão Silva[1]

Abstract

One important feature of the globalization process is the increase in the economic and financial interdependencies across countries. 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, as well as the lack of unified data sources. This article aims to use network analysis to map both the foreign direct investment and portfolio 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

JEL classification: C02, C63, F21

1. Introduction

With the ever-growing globalization, an increase in trade, and the possibility of easily investing abroad, 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 and address the ultimate investors.

In this context, the external statistics play an important role to analyse the cross-border financial investments between one country and its main investors. The financial account under the Foreign Direct Investment (FDI)[2] and the Portfolio Investment (PI)[3] records the international investment. In the case of the FDI, it aims to establish a long-lasting interest in a foreign business, while the PI is oriented toward

[1] Bojan Stavrik (bojanstavrik@me.com) and Flávio Pinheiro (fpinheiro@novaims.unl.pt), NOVA Information Management School, João Falcão Silva (jmfsilva@bportugal.pt), Statistics Department, Banco de Portugal. The views expressed are those of the authors and not those of the NOVA Information Management School or Banco de Portugal.

[2] 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.

[3] corresponds to cross-border investments in the form of debt securities and equity that falls below the FDI threshold.

small investors that aim to get short-term returns. In this latter case, most transactions occur in secondary markets, thus not between the original issuer and the final investor.

Traditionally data has been communicated and shared in structured tabular formats. Indeed, many open data platforms these days still rely on such traditional format that makes it difficult for users to explore the underlying complex structures that often characterize the relationships between the elements that are being reported. Moreover, tabular data representations are not only difficult to interpret but also difficult to compare across time and space. In fact, information about size, magnitude, proximity, comparability, and temporal evolution are easier to grasp when data is represented and reported through appropriated visualization formats. 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.

Here, we focus on the analysis of FDI and PI bilateral flows of one country vis-à-vis its main investor partners aiming to analyze the cross-border financial investment linkages in the form of debt instruments and equity in the case of FDI, and in the case of the PI debt securities and equity. To that end we use data provided by the International Monetary Fund (IMF) - Coordinated Direct Investment Survey (CDIS) and Coordinated Portfolio Investment Survey (CPIS), for all the available world countries. Using network analysis to model the linkages between the sources and final destinations of the FDI and PI, we show that the countries with more FDI and PI interconnections usually correspond to advanced economies, financial centres, or countries that offer tax benefits to investments.

Moreover, we explore the development of a web application to report the analytical results to a wider audience. The web application was built in a stack of free and open-source tools comprised of HTML 5, CSS, and javascript (jQuery, D3.js, and d3Plus). The presented web application offers a medium to i) communicate our findings through a personalized and interactive visualization-rich data-driven platform; and ii) allows for users to quickly explore and discovery of relevant partnerships and investment paths, facilitating the comparison between different countries and in different periods in time..

The article is organised as follows: after the introduction, a literature review on the network analysis and its linkage with economic variables is presented in section 2. The data sources and variables are described in section 3. Section 4 presents the methodology and section 5 shows some results. The fi-networks.com portal is described in section 6 and section 7 concludes.

2. Network analysis and economic variables

A network is a system made up of actors (individuals, organizations, countries, etc.) and sets of bilateral ties that represent relationships 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).

Network analysis has a wide range of current literature on the application of network analysis to FDI and portfolio investment relationships, including patterns in banking and cross-country financial investments, both portfolio and direct investment. On the top of some examples, Hafner-Burton, Kahler, & Montgomery (2009) use network analysis on key international outcomes and test network theory in the context of international relations. Focusing on the portfolio investment linkages, Hakeem and Suzuki (2016) took a network approach in foreign portfolio investment of the European Union and its main 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. Moreover, the literature suggests 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.

On the FDI perspective, Bolívar et al. (2019) find 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. On the other hand, political stability and average years of schooling have a moderate effect on inward FDI investment for that economy. As evidenced, developed-to-developed connections represent (66%) of the weighted global FDI networks. More recently, Norgren and Olsson (2021) apply Stochastic Actor-Oriented network models to study the relationship between FDI and institutions. They distinguish between formal and informal institutions, where informal institutions are culture and trends while formal institutions are the laws and rules of society. Additionally, (Lima, Pinheiro, Silva, & Matos, 2020) analysed the use of the network analysis for FDI relationships. The authors highlighted the visualisation capabilities of the network analysis methodology and also its ability to apply metrics that provide useful information about economic relations.

By definition FDI is meant to establish a long-lasting interest and, in many cases, the ultimate investor is difficult to trace, especially when investments involve offshore centers or Special Purpose Entities (SPE's)[4]. 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.

3. Variables description and data source

According to the Balance of Payments Manual, in its 6th edition (BPM6), direct investment includes the cross-border investments where there is a control or a significant degree of influence on the management of an enterprise that is resident in another economy[5]. It captures the immediate direct

[4] 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).

[5] The significant degree of influence is determined to exist if the direct investor owns from 10 to 50 percent of the voting power in the direct investment enterprise. Control is determined to exist if the direct investor owns more than 50 percent of the voting power in the direct investment enterprise.

investment relationships, i.e., when a direct investor directly owns equity that entitles it to 10 percent or more of the voting power in the direct investment enterprise. On the contrary, the portfolio investment is defined as cross-border transactions and positions involving debt or equity securities, other than those included in direct investment or reserve assets, meaning that there is no control or significant degree of influence on the non-resident enterprise.

The direct investment is usually presented in two alternative perspectives – following the asset/liability principle (as introduced in BPM6) or directional principle (requested in previous editions), whereas in the case of portfolio investment only the asset and liability principle is presented.

Under the directional principle, direct investment is shown as either direct investment abroad (outward investment[6]) or direct investment in the reporting economy (inward investment[7]). The asset (liability) principle of the portfolio investment, represents the amount invested by resident (non-resident) entities in the form of equity/debt securities, on non-resident (resident) entities.

In this paper, the implementation of the network estimation uses statistical information on the FDI directional principle and asset-liability principle for the portfolio investment. According to the available information, the primary presentation of international accounts shows positions with all non-residents as a total. Although, we follow a network perspective aiming to map the foreign investment linkages between partner economies. In this regard, we obtained information from the Coordinated Direct Investment Survey (CDIS) and Coordinated Portfolio Investment Survey (CPIS) provided by the International Monetary Fund. The selected data contains annual information from 2009 until 2019 on the total inward direct investment (stocks), inward equity direct investment (stocks), portfolio investment assets and liabilities.

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[8]. 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.

To obtain a proxy of the foreign investment, we combine both Inward from CDIS and the liabilities from CPIS datasets (combined liabilities) and outward from CDIS with assets from CPIS (combined

[6] Investments by resident direct investors in their direct investment enterprises abroad deducted from the reverse investments by direct investment enterprises abroad in their resident direct investors.

[7] Investments in resident direct investment enterprises by direct investors abroad minus Reverse investments by resident direct investment enterprises in their direct investors abroad.

[8] 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 that were used. 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.

assets). 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[9].

4. Methodology

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 the context of this work, we shall use nodes of a network as representations of countries, while edges represent the existence of an investment between a pair of countries.

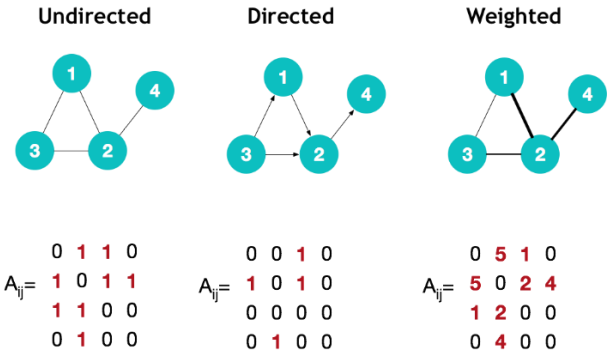


Figure 2 – Graphical (top) and matrix (bottom) representation of networks representing structures that have edges of different nature. Source: (Lima, Pinheiro, Silva, & Matos, 2020).

Depending on the nature of the relationships being modelled, 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, where $a_{ij} = a_{ji}$. 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.

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), with the possibility of two links (arrows) between two nodes. Additionally, in a weighted graph, the thickness of the edge represents the strength of the relationship.

Additional attributes can be associated with each relationship and each actor (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.

[9] 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.

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. 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. 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, 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. 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:

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

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

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.

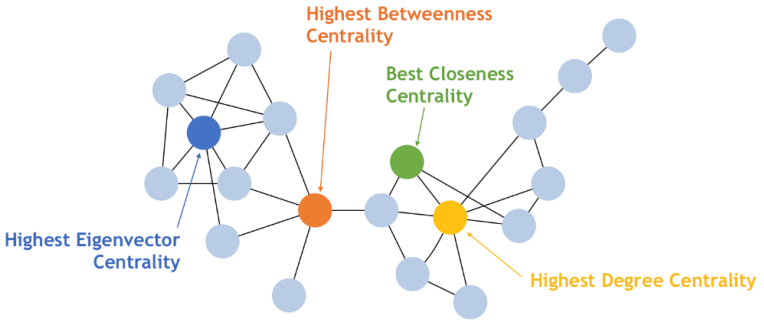


Figure 3 – Simple network with the most important centrality measures Source: (Lima, Pinheiro, Silva, & Matos, 2020)

In this article, 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. 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 application (see **https://fi-networks.com**).

Correlation

Figures 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, thus 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. 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.

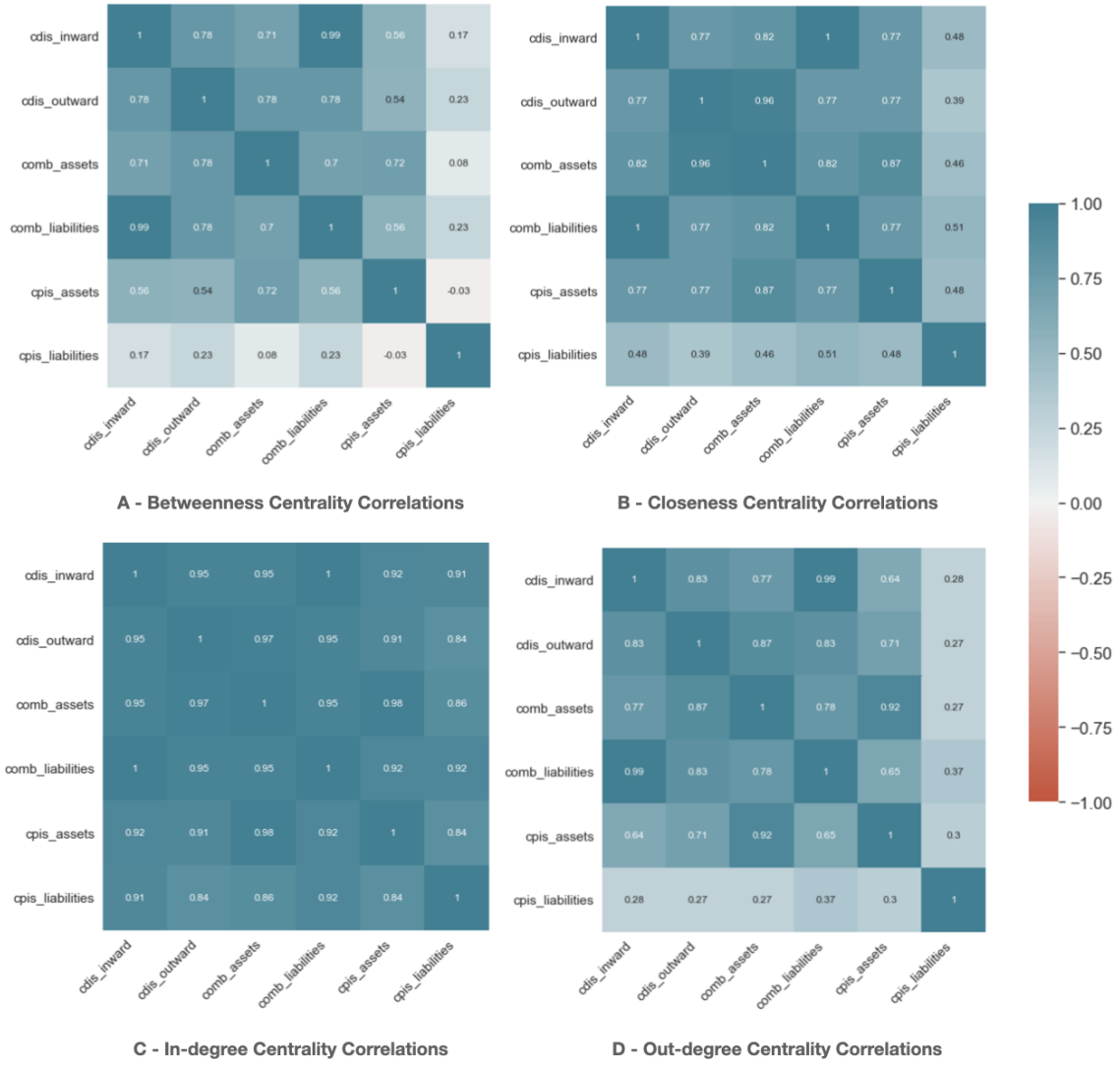


Figure 4 – Spearman Correlation Matrices for Betweenness (A), Closeness (B), In-degree (C), and Out-degree (D) centralities in the Networks estimated for the year of 2019.

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.

Foreign Direct Investment–- Inward Network

As previously mentioned, six data sets were considered in our analysis[10]. In this section, we illustrate the results for one specific data set: FDI inward. For the remaining data sets, the fi-networks portal will have all the information. Although, the interpretation of the remain data sets follows closely the one that is described below.

FDI Inward 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., the China, the Canada, and the 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. Most of these countries are tax havens. In that sense, considering that closeness centrality highlights nodes that are on average closer to all the other nodes in the network, 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.

Nonetheless, the Netherlands and the Luxembourg rank, respectively, 12th and 13th in closeness centrality, while the U.S. is 9th. 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 the 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.

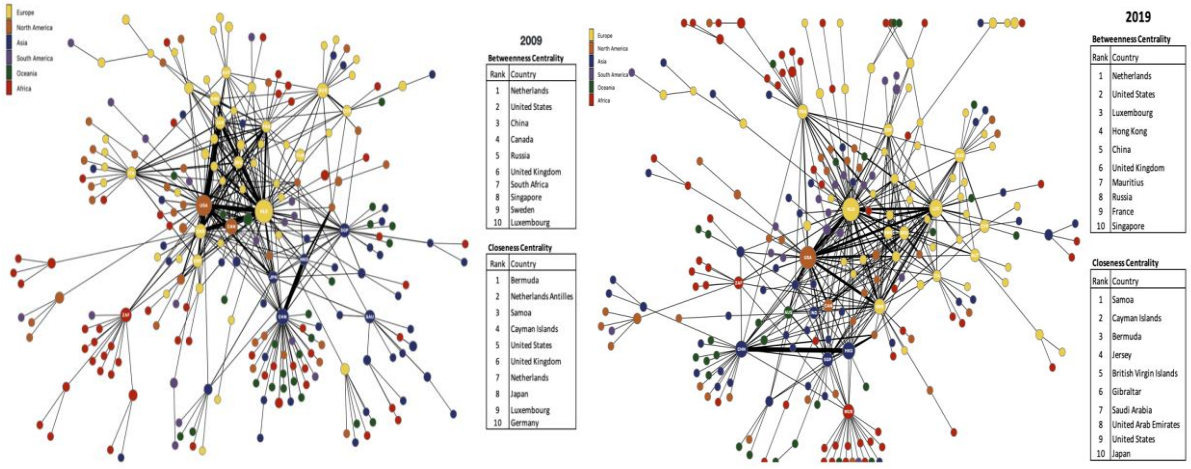


Figure 5 – CDIS Inward Network with node size representing relative amounts of Betweenness Centrality for 2009 (top) and 2019 (bottom). Source: web app

[10] inward/outward of the FDI, assets/liabilities of the portfolio investment and the combined assets and liabilities.

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 continent. The U.S., Netherlands, China, and Luxembourg combined make up 42.57% of global FDI inward investments for 2019.

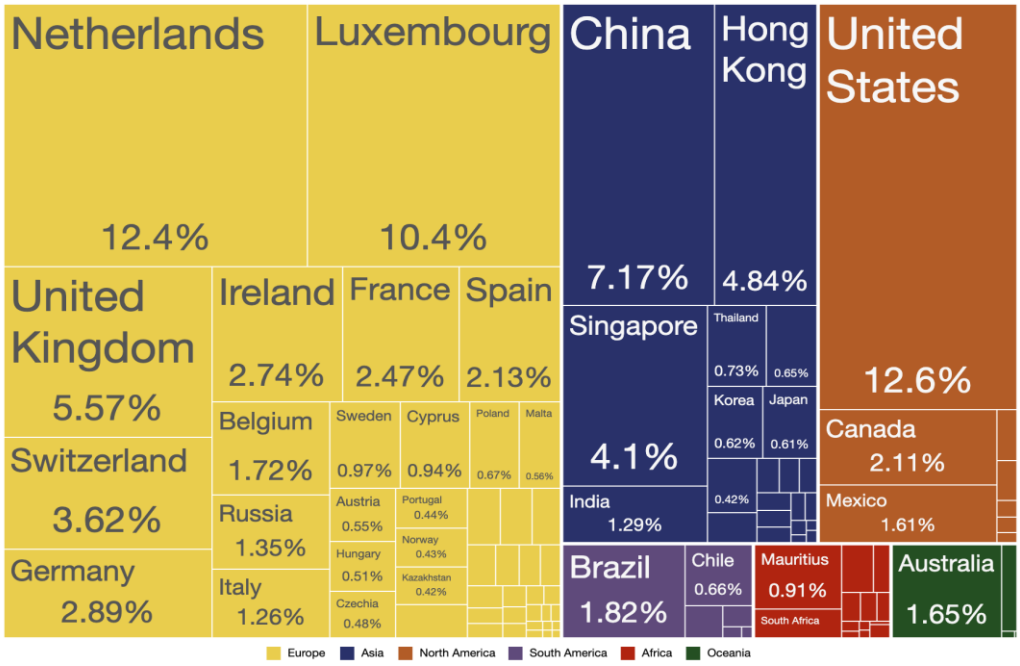


Figure 6 – CDIS Inward Treemap with block size representing the relative amount of FDI investment per country for 2019, with the colours representing the different continents. Source: web app

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. In addition, 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.

6. The fi-networks.com Portal

We chose to develop the web application from scratch using free and open-source tools. The exploration analysis and all computations were done using python and resorting to libraries common in the data science stack, such as Pandas and NetworkX. The early steps of the project consisted of data cleaning and pre-processing, stored in json or csv files. Those documents are the basis for the visualizations displayed in the portal.

The portal, as a web application, was built in a stack of free tools that comprised HTML 5, CSS, and javascript. In particular, several handy javascript libraries — such as jQuery, D3.js, and d3Plus — where used for the development of interactive elments and build the visualizations. Moreover, the bootstrap

CSS framework developed by Twitter (see https://getbootstrap.com/) was used for quick prototyping of html elements. Finally, 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), which allows drawing shapes in a browser window. The main advantage of using SVG's is that these objects do not lose quality when rescaled and have low memory requirements.

Regarding to the data loading and storage, data is loaded directly from csv files due to its lower development complexity and requirements[11].

1.1.1 Landing Page

The goal of the landing page is to grab the attention of the user while offering quick access to the different datasets available.

The top of the landing page contains a several relevant statistics cards about the used datasets. Their goal is to give the user an overview of the magnitude of the data being used. Below that, we present five 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 was applied to the rest of the teasers, except they take the user to a different section.

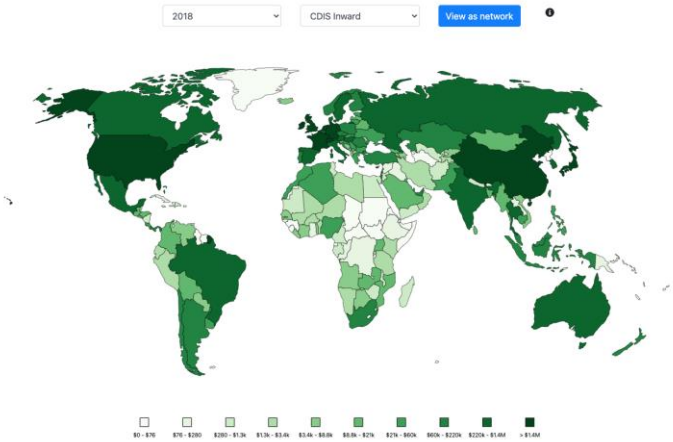


Figure 8 – World map visualization in the landing page shown for 2019 CDIS Inward. Source: web app.

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 8, 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[12].

[11] Another option was to build a database that would feed data to the portal through a RESTful API.

[12] 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.

1.1.2 Network Page

The network page is the core element of the portal. Each visualization in this page is broken down into two sections. The selection (left bar) side allows the user to select the visualization options, by selecting different networks, years, countries, and dataset. The options change depending on visualization, although all of them present an option to select the dataset and year. There are also information icons, which explain the contents of the visualizations, with a 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 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, putting into evidence the web of tradeflows that can be estimated from 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 represent the continent of each country. The link that connects countries represents 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 is a more natural degree of freedom to communicate intensity/magnitude.

The displayed networks represent a subset of the most relevant relationships on top of the Minimum Spanning Tree (MST). By doing this we overcome the issue of representing a very dense network, as most pairwise relationships exist, though they might be irrelevant in most cases. Hence, for visualization we enrich the MST with the most relevant links until a network with average degree of 3.5, which we use as a thumb rule. However, such filters are only done for visualization purposes, all calculations are performed on the fully spanned directed-weighted networks.

The most central countries per selection are highlighted with their ISO3 codes printed inside the nodes. Although, the user can manually click a country to highlight it in the network. Clicking the teaser for the CDIS Inward network in 2019 highlights the U.S.A. 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.

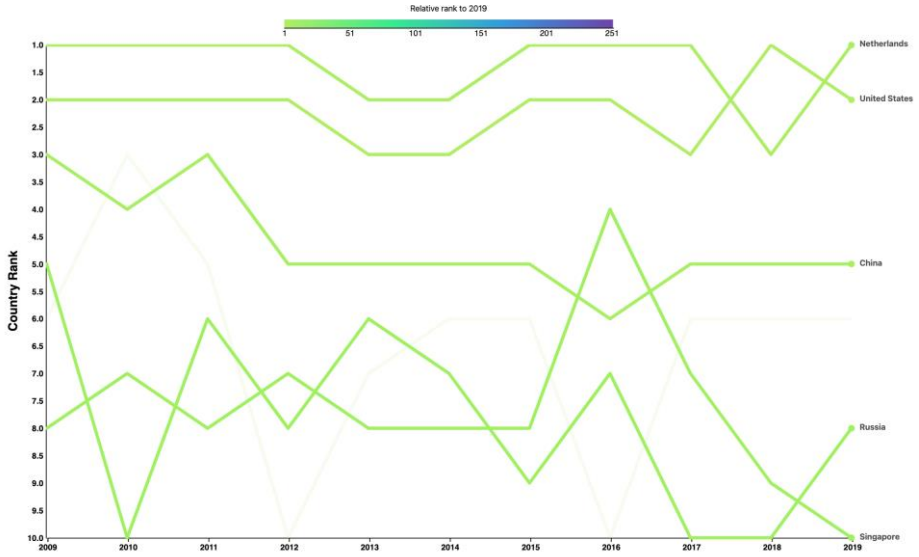


Figure 9 – Rank Evolution Chart between 2009 and 2019 based on Betweenness Centrality, showing the evolution of Netherlands, United States, China, Russia, and Singapore. Source: web app.

The second visualization presented is the rank evolution chart. The purpose of this visualization is to show the evolution of the relative importance of countries through their 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 9, 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, serving as a reference to the 2019 rank throughout the years. There is an option to select the network, centrality, and countries displayed in the chart. In the case of the U.S. Figure 9 shows its low variability in terms of betweenness centrality, never dropping below the third position between 2009 and 2019.

Next we show the time evolution of the shortest path charts. This particular visualization highlights the most likely intermediators if the flow of investments would follow a path of "minimum effort". More importantly, it shows how such paths changed from year to year. Figure 10, 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. In addition, the platform includes also a treemap chart to show the relative magnitude and breakdown of the total amounts per country and continent dependent on year and network.

Figure 10 – Shortest path chart between the United States and China between 2009 and 2019. Source: web app.

Lastly, the top intermediators chart (Figure 11) aims to identify the countries that more often participate as intermediators in the shortest 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; and secondly the other identifies the percentage of times a country is the first intermediator in the shortest path of investment. 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).

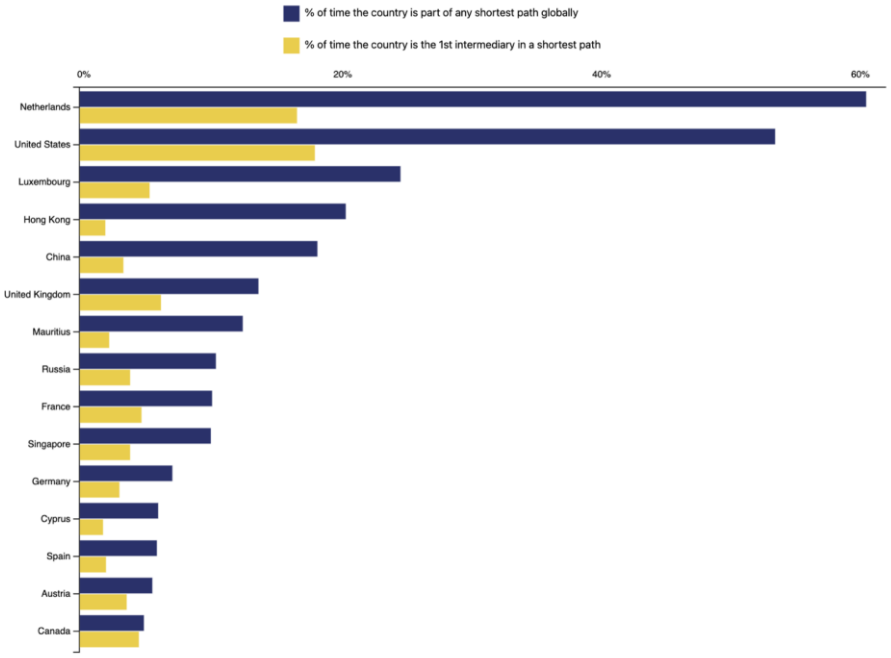


Figure 10 – Horizontal bar-chart chart showing the top intermediators based on the percentage of times they are part of a shortest path (blue bar) and percentage of time they are the first intermediator (yellow bar). Source: web app.

7. Conclusions

This article focuses on 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.

Our results show that 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. Notwithstanding, the variation in these ranks is not the same for all the listed countries. Hong Kong and China, have become some of the most influential intermediators only in the last several years. The United States, Netherlands, and Luxembourg are in the top 20 ranks for most of the years between 2009 and 2019. In terms of closeness centrality, off-shore countries such as the Cayman Islands, Bermuda, Jersey, the and British Virgin Islands are among the top ranks in the last few years, whilethe 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, policymakers and all the other users, 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.

Even though, this analysis 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. It is also important to mention that there exists some reporting data gaps as the information regarding to the assets/inwards is not symmetric to the liabilities/outwards.

User experience surveys and interviews have not been performed due to time limitations. They are integral part of building web applications in order to better understand the improvement aspects of the design and communication in the web app. Such information should be gathered from experts in the field, such as central bankers and policymakers.

Finally, the adoption of portals as the case of fi-networks can be a very comprehensive tool for users, and a more visualized way for the organizations to show/highlight the main messages that arise from the data.

References

Balance of Payments and International Investment Position Manual, 6th edition, International Monetary Fund.

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.

Damgaard, J and T Elkjaer (2017): "The Global FDI Network: Searching for Ultimate Investors", IMF Working Paper, November 2017.

Jackson, M (2008): "Social and Economic Networks". Princeton: Princeton University Press.

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.

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 Benchmark Definition of Foreign Direct Investment - 4th Edition.

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.

Data science and Statistics: a network analysis to understand the foreign investment

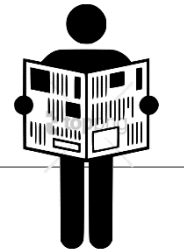
Bojan Stavrik| Flávio Pinheiro | João Falcão Silva
bojanstavrik@me.com | ***fpinheiro@isegi.unl.pt*** | **jmfsilva@bportugal.pt** |

Data science in central banking

14 February 2022

BANCO DE PORTUGAL
EUROSISTEMA

1 Applying network analysis to the foreign investment - WHY?



China Three Gorges buys EDP stake for 2.7 billion euros

Axel Bugge

LISBON (Reuters) - China Three Gorges won the competition to buy Portugal's stake in utility EDP (EDP.LS), paying 2.7 billion euros ($3.5 billion), in a privatization seen key to the indebted euro zone country's ability to sell state assets.

The deal, which also includes Chinese investment in the wider economy, is the brightest news for Portugal since it was forced to seek a 78 billion euro bailout from th European Union and International Monetary Fund in the spring after its financing cos soared.

State holding company Parpublica said on Thursday that China Three Gorges' offer fo the 21 percent stake in EDP, Portugal's largest company, was at a 53 percent premium to its share price.

The Chinese energy giant beat Germany's E.ON (EONGn.DE) and Brazil's Eletrobras (ELET6.SA) after a tough competition in which Three Gorges had promised to sharpl

Budget 2022: Investors brace for another year of high bond supply

The government will need to borrow a minimum Rs 4.5 lakh crore to repay past loans, apart from continued spending to support an economic recovery

APARNA IYER
JANUARY 24, 2022

Find a buyer

Bond supply jumped the most in FY21 due to the pandemic. However, bond yields fell more than 100 basis points during the year. Yields move inversely to bond prices.

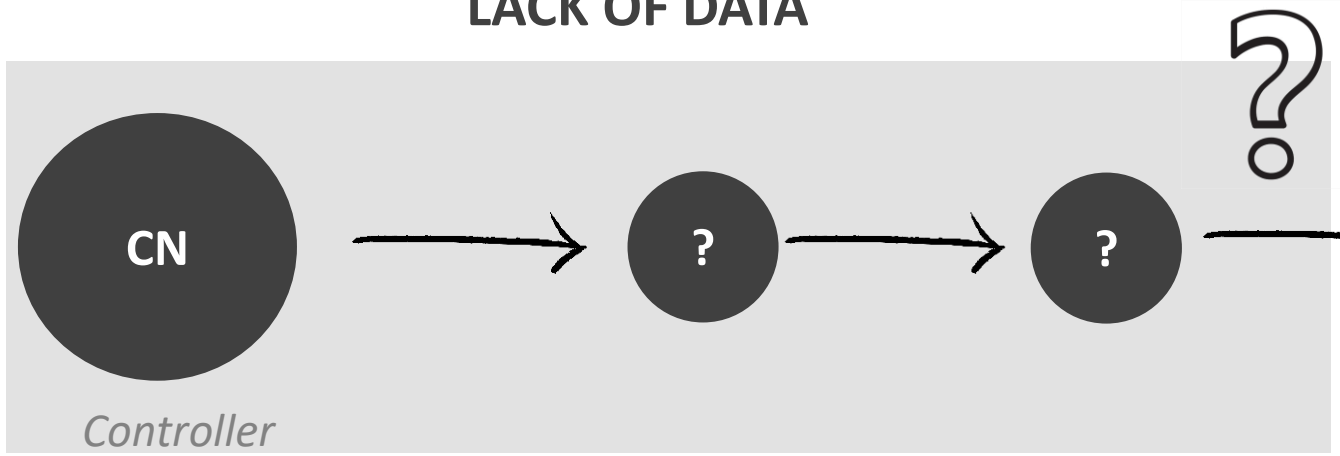
The key reason was that the RBI infused a historic amount of liquidity over various tenures through an array of instruments. The central bank also stood in the market as a constant buyer of government bonds.

In FY22, the central bank has been a net buyer, but a more reluctant one. Its purchases dropped to Rs 1.4 lakh crore this year from Rs 3.1 lakh crore in FY21. ***In recent weeks, the central bank has been a big seller of bonds in the secondary market.***

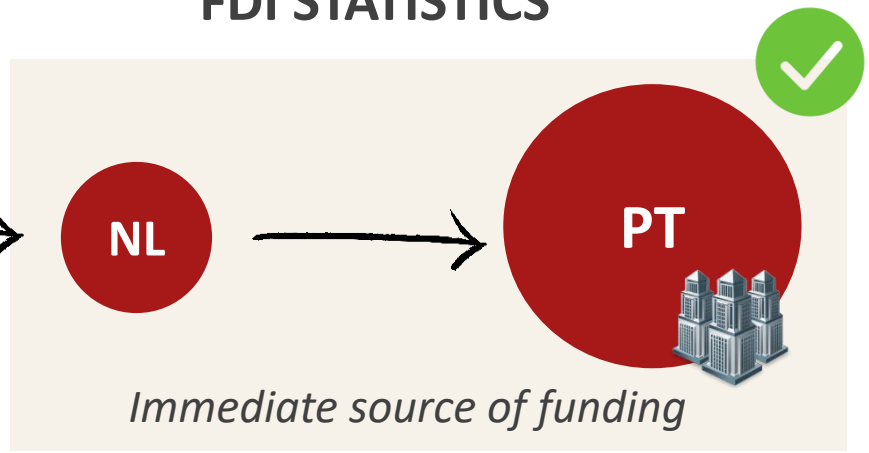
1 Applying network analysis to the foreign investment - WHY?

01

LACK OF DATA

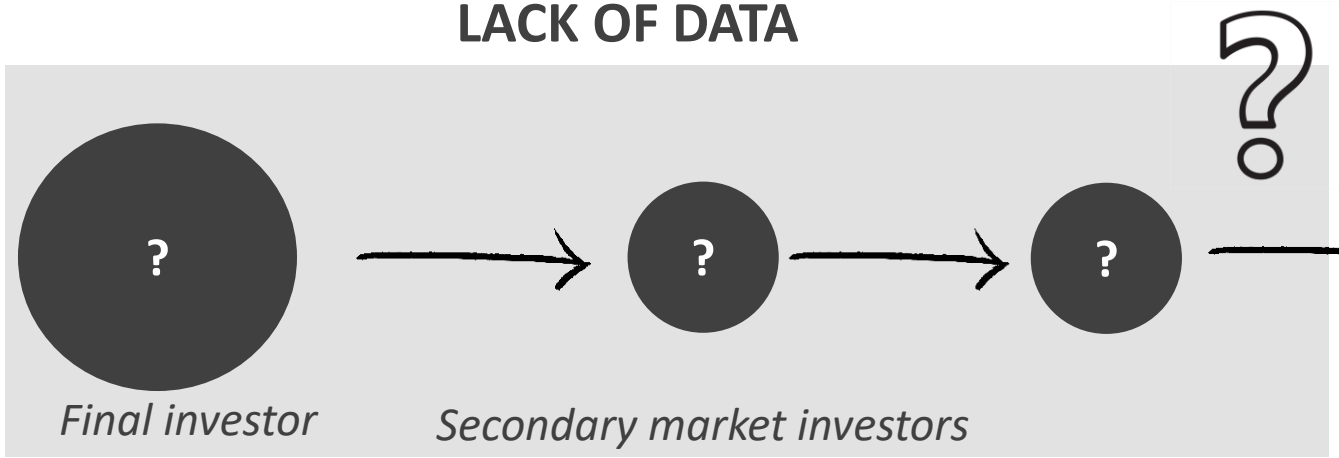


FDI STATISTICS

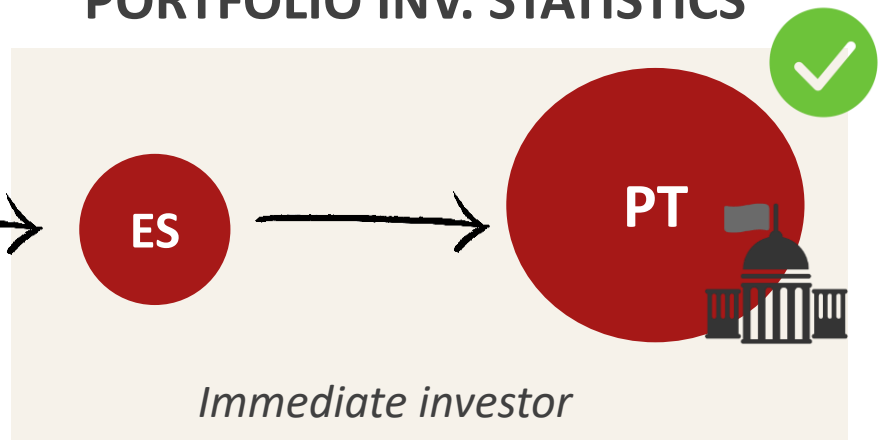


02

LACK OF DATA



PORTFOLIO INV. STATISTICS



2 Applying network analysis to the foreign investment - HOW?

Use the network analysis to map foreign investments:

FOREIGN DIRECT INVESTMENT

AND

PORTFOLIO INVESTMENT

01 IDENTIFICATION OF PATTERNS

Identify the **patterns** and **shortest paths** between the immediate and ultimate investors

02 ESTABLISHING TRENDS

Establish **trends** and describe the relations between countries over time

03 RESULTS ILLUSTRATION

Illustrate the results of the network in an intuitive **web application**

04 PREDICTIONS

Use the network analysis to **predict** the ultimate **investor** and **intermediaries**

2 Applying network analysis to the foreign investment - HOW?

DATA SOURCES

IMF Coordinated Portfolio Investment Survey - annual data from 2009-2019

IMF Coordinated Direct Investment Survey - annual data from 2009-2019

DISTANCE

Countries with larger bilateral stock are closer, thus are at a shorter distance from each other (the weight of the link is lighter):

$$\phi_{ij} = \frac{1}{|f_{ij}| + |f_{ji}|}$$

STOCKS

DIRECTIONAL PRINCIPLE – FDI INWARD AND FDI OUTWARD

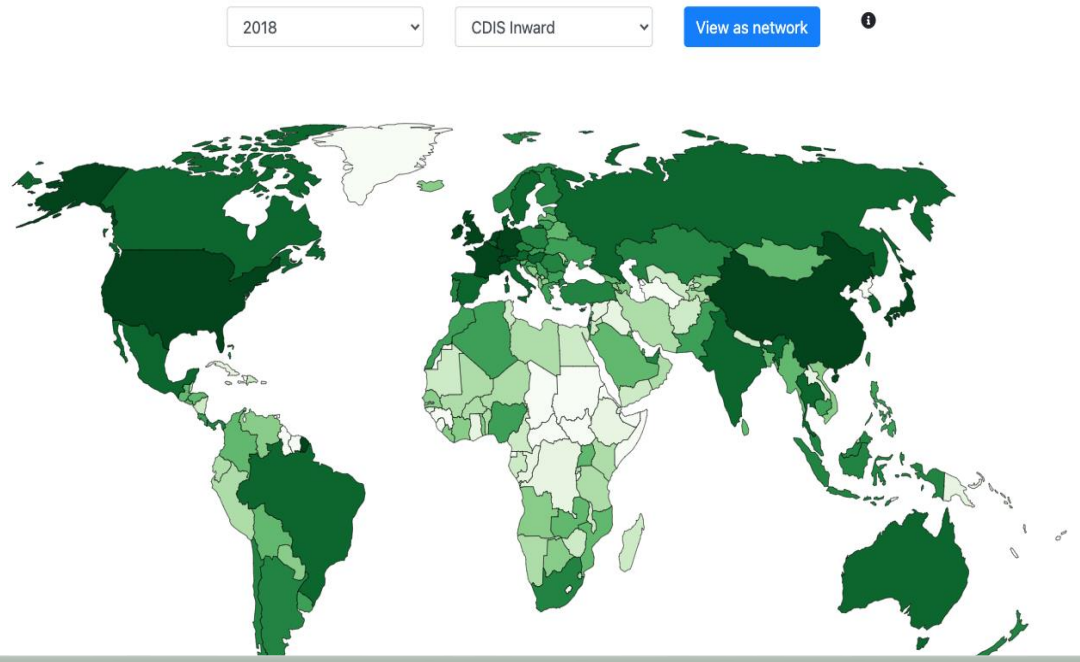
PI ASSETS AND PI LIABILITIES

COMBINED ASSETS AND LIABILITIES *

* **Assets** = FDI outward + PI assets

Liabilities = FDI inward + PI liabilities

3 THE PLATFORM: FI-NETWORKS



THE WEBSITE IS HOSTED ON THE NETLIFY PLATFORM!

HTML5 AND CSS

D3.JS AND D3PLUS

- For building the **skeleton of the website**
- Placing the **selection menus** and **descriptions**

- **D3.js** - JavaScript library for developing **dynamic, interactive data visualizations.** It uses Scalable Vector Graphics **(SVG's)**
- It builds interactive visualizations, such as the **network visualization, world map, treemap and bar** and **line charts**

3 THE PLATFORM: FI-NETWORKS

F I - N E T W O R K S

THE DATA BEHIND THE VISUALIZATIONS IS COMPUTED IN PYTHON

OUTPUTS STORED IN CSV/JSON FILES, ORGANIZED IN A FOLDER STRUCTURE

VISUALIZATIONS BUILT WITH JAVASCRIPT, LEVERAGING ON SVG PROPERTIES

SVG'S HAVE LOW MEMORY REQUIREMENT AND MAINTAIN QUALITY WHEN SCALED

HOSTED ON NETLIFY, A SERVERLESS HOSTING SERVICE PULLING SOURCE CODE FROM GITHUB

5 PAGES: HOME NETWORK METHODOLOGY ABOUT DOWNLOADS

F I - N E T W O R K S

THE PLATFORM: Fi-networks:

https://fi-networks.com

3 Conclusions and future work

01 Network science illustrates the enormous analytical power to **predict the ultimate investors for FDI and Portfolio Investment**

02 Few countries including the **US, NL, LU, CN, HK** and **the UK** are **the main global intermediators**

03 **FI-NETWORKS,** web application to:

i) communicate these findings in an **interactive data visualization platform**

ii) quick exploration and discovery of relevant **partnerships/investment paths**

WHAT'S NEXT?

EXPLORE HOW THE WEB APP CAN BE A FUNDAMENTAL TOOL FOR REGULATORS AND CENTRAL BANK OFFICIALS