Trade Flow Analysis

Elia Cannas, Master's Degree in Computer Science Curriculum B: Computer Science for Management, ID number (0001097520)
Michele Abruzzese, Master's Degree in Computer Science Curriculum B: Computer Science for Management, ID number (0001097676)

1. Introduction

In the context of global economic evolution, trade data analysis plays a crucial role in understanding trade dynamics among nations. Globalization has expanded opportunities for interchange, bringing with it increasing complexity in international trade patterns. Against this backdrop, this study aims to explore and analyze the BACI dataset provided by the French research center CEPII (Centre d'Etudes Prospectives d'Informations Internationales) on global trade.

The **general context** concerns the global economy and geopolitics in fact, understanding how nations trade goods and services not only reflects economic dynamics, but also provides an analysis of how the balance of power and international relations develop in the current context.

The **specific application** reflects trade network analysis and temporal trends. Our investigation focuses on visualizing the trade network and evaluating its evolution over time. Through the application of network analysis techniques, we will explore the density of trade, highlighting how interconnections between nations have changed over the years. In addition, we will use measures of network centrality to identify countries that play a key role in the overall structure of international trade.

The work done is based on a study done previously *BACI: International Trade Database at the Product-level The 1994-2007 Version*, so we will use network analysis techniques that allow us to visualize the world trade network, its definition and description in topological terms. These analyses will allow us to produce and discuss some of the commonly used network statistics. We will refer to data ranging from 2017 to 2021, also examining the influence COVID-19 has had on them.

2. Problem and Motivation

Addressing problems related to the analysis of the trade network and its dynamics over time has several aspects of theoretical and practical relevance.

Network Dynamics over Time:

- *Problem*: To understand how the density of trade between countries changes over time and the consequences of these changes.
- *Importance*: This aspect is crucial for identifying temporal patterns, significant events (such as COVID-19) and structural adaptations in the network, providing a theoretical basis for predicting future trends and formulating adaptation strategies. We can then have a comparison with the study carried out in *BACI: International Trade Database at the Product-level The 1994-2007 Version.*

Measuring the Influence of Countries:

- *Problem*: Identify the most influential countries in the trade network through measures of centrality.

- *Importance*: Understanding which nations exert the greatest impact on the stability and evolution of the network is essential for holding economic policies accountable and anticipating how a country's decisions may propagate globally.

Practical Relevance of Network Analyses:

- *Problem*: Transform theoretical network analysis into practical and applicable results for decision makers and stakeholders.
- *Importance*: Project results must be communicated in a clear and accessible way so that strategic information from the analysis can be used by governments, companies, and international organizations to guide business policies and economic decisions.

Project Contributions:

- *Theoretical Significance*: The project will contribute to the development of international trade network theory by integrating temporal analysis and node influence assessment.
- *Practical Relevance:* It will provide practical tools to analyze the trade network and identify opportunities or risks related to global trade dynamics.
- *Socio-Economic Impact*: Information from the project can be used to improve understanding of global trade relations, guiding more informed and sustainable economic decisions.

In summary, the project aims to combine the theoretical aspect of network analysis with its practical applicability, offering a comprehensive picture of the dynamics of international trade over time and providing useful tools for interpreting, predicting and managing these dynamics.

3. Datasets

The datasets used during the study can be found online free of charge.

Data pertaining to **world trade** can be accessed at the following link:
http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=37 .
The folder downloaded to conduct this study is "H17"
(direct download: Dataset_HS17) which includes the BACI datasets from the year 2017 to the year 2021, updated on February 1, 2023, the dataset of trade-associated product codes and the dataset of trade-associated country codes.

A description of the dataset can be found at the link:
http://www.cepii.fr/DATA_DOWNLOAD/baci/doc/DescriptionBACI.html

Data regarding **information for the countries** within the study (country_name, country_code, latitude, longitude) can be downloaded at the following link:
https://gist.github.com/tadast/8827699

Data for **dividing countries into regions** and sub-regions were downloaded at the following link: https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv

To carry out the study more efficiently, a new dataset was created that can be downloaded at the following link:
https://github.com/elia99l/SocialAnalysis/blob/main/csv/country_code_coordinates_region_merged.csv

This dataset was created by us to have all the information inherent in the 226 countries that are part of the trade network in a single dataset. We then merged the dataset with country information (latitude and longitude) together with dataset that divides the countries by geographic regions.

The data is stored in the devices of the group components and then uploaded to the project's github repository.
Data manipulation is performed using PyCharm software and the Jupyter Notebook tool. With PyCharm it was possible to perform the various measurements and representations of them in the form of graphs and tables. With Jupyter Notebook we focused on the graphical representation of the exchange network. Thus, the Python programming language is used along with the NetworX library as far as the network is concerned, while for data manipulation and visualization we used: Pandas, Matplotlib and Numpy.

4. Validity and Reliability

Regarding the **validity** of our trade dataset we briefly focus on the history of the **BACI dataset**.
Bilateral trade flows are consistently collected at the commodity level in detail (6-digit Harmonized System) by the United Nations Department of Statistics, in the ComTrade database. Some research centers have recently begun to produce their own variants of the original database. The French research center CEPII (Centre d'Etudes Prospectives d'Informations Internationales) is one of the first that has begun to offer the results (or spin-offs) of its research to the public in terms of data. The BACI (Base pour l'Analyse du Commerce International) dataset is one such result. The original motivation for producing BACI is that, despite the richness of the data reported in ComTrade, the attempt to account for the largest number of countries, the largest time period, and the most disaggregated level of product is plagued by too many missing flows in the original UN database. BACI uses a mirror statistics strategy to impute the missing data. Using a "reconciliation" methodology, BACI substantially reduces the number of missing values. The BACI database is used overwhelmingly in applied trade analysis. It is spreading rapidly among scholars and is constantly updated by CEPII and is widely available.

The **validity** of **the data inherent in the information on countries** that are part of the trade was empirically confirmed by us. We checked the latitude and longitude for each country in the list. The same applies to the **dataset that contains information regarding the countries' geographic regions.**

The methodological approach followed in our study not only ensures the validity of the data used, but also the **reliability** of the results obtained. This reliability is supported by several key factors:

Analysis Scripts:

- The analysis scripts we have produced are thoroughly documented and are an integral part of our methodological approach. These scripts provide clear instructions on how the data were manipulated, processed, and analyzed. Such detailed documentation facilitates repeatability of the analysis.

Accessible Datasets:

- As mentioned earlier, datasets can be found easily at the links that are given in Section 3. This transparency in data access allows other researchers to examine the same information that we used to conduct our analysis. In addition, by making the original datasets available, it gives others the opportunity to reproduce the analysis and verify the results independently.

5. Measures and Results

Several measures of centrality have been calculated, which can be classified into three main groups:

- Degree centrality
- Closeness centrality
- Centrality of eigenvectors

What are they used for?

- **Degree centrality** allows us to measure the number of arcs in and out of a node, so we could observe the number of imports and exports for each node. Within our study, a country with high degree centrality is more directly connected to more other countries than those with lower degree centrality. Countries with high degree centrality may be considered key within the trade network. These countries could play a central role in trade facilitation and significantly influence the overall dynamics of the network.
- **Closeness centrality** allows us to measure the average distance of a node from other nodes, which is used to observe the ease with which a node can be reached by other nodes. referring to our study, A country with low closeness centrality is "closer" to other countries in terms of direct trade flows. This could indicate greater interconnectedness and trade accessibility.
- **Eigenvector centrality** allows us to measure the centrality of a node by the characteristics of its neighbors, referring directly to how important, central, influential, or closely clustered a node's neighbors are. A country with high Eigenvector centrality could be seen as a key player that is connected to other prominent countries in the trade network.

From the values obtained furthermore, comparisons were made to observe the change in trade flow as the years passed, thus also verifying the change that there was due to COVID-19.

Several files.csv containing the results of the calculations of the various measures and graphs were produced through PyCharm to have a more explicit representation of the observed phenomenon.

Link csv file and code:
https://github.com/elia99l/SocialAnalysis/tree/b558bf92bb60a2acf6668dd67b556cec4eb8e5da

(N.B. Keep in mind that the values shown in the Closeness Centrality column are calculated by the closeness centrality() function of Python's Networkx library, which associates a high value for nodes with a better closeness centrality value)

We are going to analyze and comment in more detail on the results obtained through a visual representation of graphs and charts created through the Python library Networkx.

The **graphs** produced by our study are as follows:

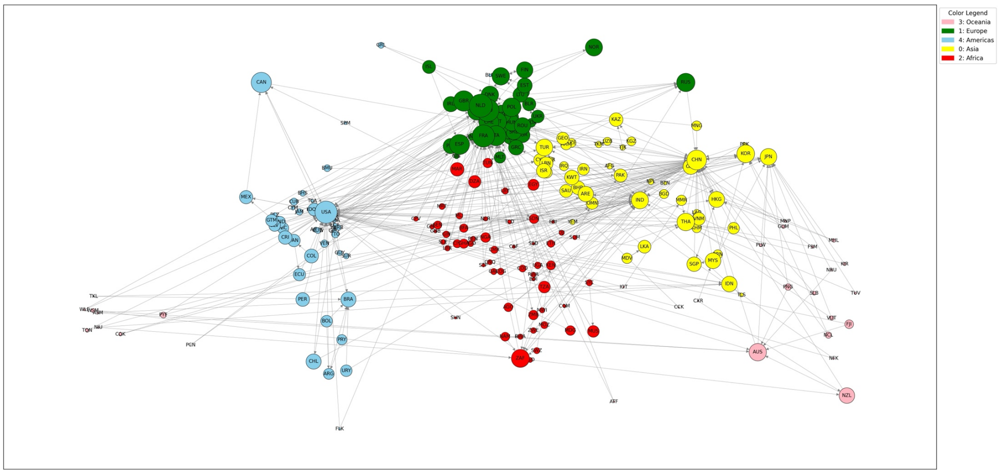

Fig.1 Graphic representation of trade flows for each country for the year 2017

Within *Fig.1*, the global trade network for the year 2017 has been represented through an oriented graph, in which we have:

- Nodes, representing the countries of the world identified by their ISO_3digit_alpha
- The oriented arcs, which represent an import or export of a given product from/to other countries (nodes)

For each node, only the first two outgoing arcs (exports) are displayed, calculated based on the value of trade (only the first two export flows were considered based on their value for reasons of graph explicability).
The location of the nodes is given by the latitude and longitude extracted from the *"country_code_and_coordinates_merged.csv"* dataset.
The nodes have different sizes that vary according to the number of inflows (imports); the size of the node area is proportional to the number of country import flows that the node represents. The size of the nodes has been normalized to best visualize the graph.
By means of a legend placed in the upper right-hand corner, we can see with which colors the various countries have been distinguished according to the region they belong to.

The code for creating the following figure can be accessed at the following link:
ViewMap.ipynb

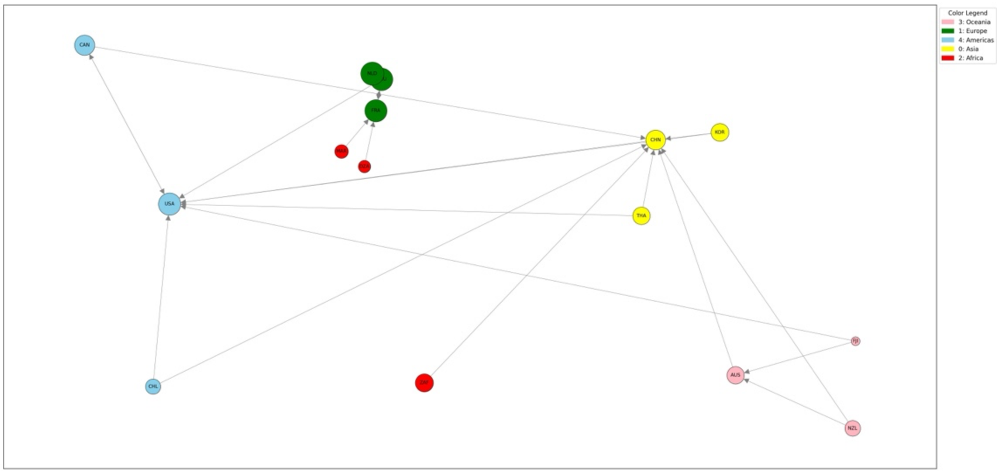

Fig. 2 Graphic representation of trade flows for the three largest importers for each region 2017

In *Fig.2* we have the representation through an oriented graph of the three largest importers for each region in the year 2017. This figure was constructed to focus on the world's largest importers. We can see from the graph that the largest importer globally is China, followed by the US.
These results are quite common with those obtained from the previous study for trade flows from 1995-2010, which also featured China and the U.S. as the main players in terms of imports in the year 2007.
The characteristics of the nodes are common to the graph shown in Fig. 1 with the addition of the filter of the three largest importers.

The code for creating the following figure can be accessed at the following link:
ViewMap.ipynb

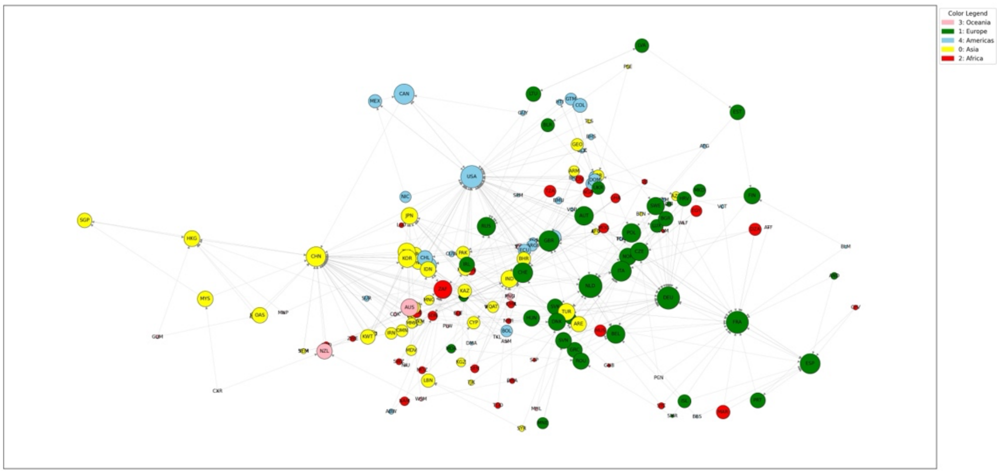

Fig. 3 Graphical representation of the network using the Kamada and Kawai force-directed algorithm.

For *Fig.3* unlike *fig.1* we see the placement in the center of the graph the highly connected nodes, unlike the less connected nodes that are placed at the edges of the figure.
For the representation we used the force-directed algorithm (Kamada and Kawai), which acts as a balanced spring system that minimizes the energy of the system, in other words it is as if the countries were connected by springs (connected countries tend to stay close together, while unconnected countries tend to be far apart). However, the position of each country depends not only on its bilateral ties, but also on the indirect effect of others; trading partners help determine the node's position in the network.

The graph helps to capture the multilateral effect on bilateral flows by assigning each country a position relative to all others as a function of the global trading system.

The code for creating the following figure can be accessed at the following link: [ViewMap.ipynb]

The **graphs** obtained from our study are as follows:

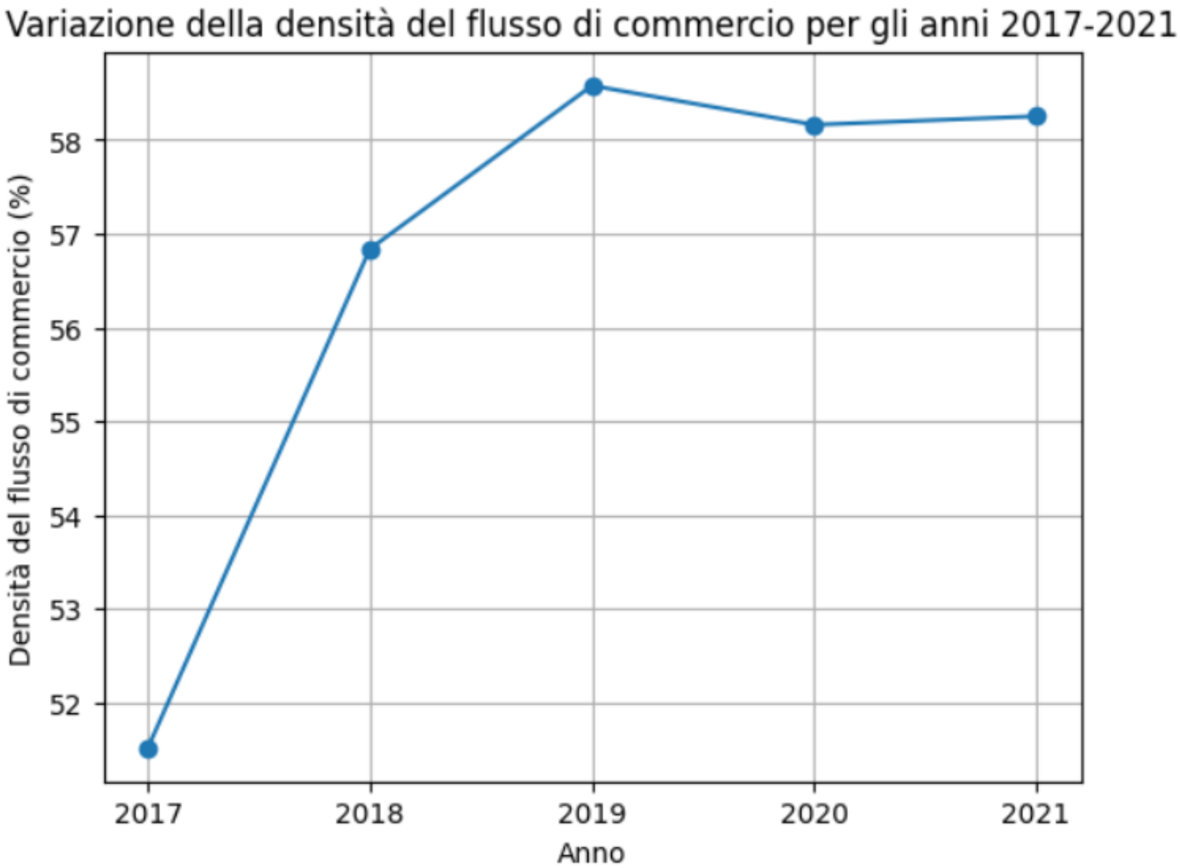

Fig.4 Change in trade flow density

Fig.4 represents the change in trade density flow for the years 2017, 2018, 2019, 2020, and 2021. The density values by year were calculated by counting the number of connections that came up in that year and dividing by the maximum number of connections that can be made for a given year.

$$Density = \frac{n_{link}}{n_{total}}$$

Where n_{link} represents the number of connections made and n_{total} represents the maximum number of connections that can be made.

Let's take an example: in 2017, 26194 links were formed, while the number of possible links with 226 countries is 50850 (number of possible pairs without repetition from 226). These data we can say that the density of links in 2017 is 51%, this means that taking two countries at random, the probability of an existing business link between them is 51%.
We can see through the graph, how this probability changes over the years, but more

importantly we can highlight how COVID-19 slightly affected the density by showing the data of 2019 with 58.57% and 2020 with 58.16%.

The code for creating the following figure can be accessed at the following link: [densityOfFlows.ipynb]

Country	Country Code	Exporter Degree	Importer Degree	Eigenvector Centrality In	Eigenvector Centrality Out	Closeness Centrality	Degree Centrality
China	156	495702	147070	0.0070623312976416	0.0066881774648029	0.933609958506224	642772
Germany	276	408040	191620	0.0071744341747524	0.0067669037579912	0.9533898305084746	599660
Puerto Rico and US Virgin Islands	842	380344	187982	0.0071467243528644	0.0067995326596715	0.9698275862068966	568326
Monaco	251	344174	186884	0.0071955179571442	0.0069087290465287	1.0	531058
Netherlands	528	315298	203204	0.0071973870322989	0.0067888540565503	0.9615384615384616	518502
Italy	380	345539	153062	0.0071823757318836	0.0066552422583461	0.9146341463414634	498601
...	...	..	..	..	..	..	..
British Indian Ocean Territories	86	238	609	0.0010085041963049	0.0020656146503564	0.5232558139534884	847
French South Antarctic Territories	260	111	691	0.0014710016892436	0.001059440338331	0.5344418052256532	802
Cocos Islands	166	368	425	0.0008090872569923	0.002440701256566	0.5184331797235023	793
Pitcairn	612	179	140	0.0008475528981628	0.0018825111464624	0.5196304849884527	319

Fig. 5 Table measures of centrality countries 2017

Within Fig. 5 we see a report of all the measures of centrality calculated for each country in the year 2017, these measurements were repeated for the years 2018, 2019, 2020, and 2021. The countries are sorted by "Degree Centrality," we notice China in first position, followed by Germany and Puerto Rico, while last, we have Pitcairn.

From the values obtained by calculating centralities, we then produced a.csv file containing for each year the countries that had the best centrality value.

Anno	Best Exporter Degree	Best Importer Degree	Best Exporter Eigenvector In	Best Exporter Eigenvector Out	Best Country Closeness Centrality
2017	China	Netherlands	United Kingdom	Monaco	Monaco
2018	China	Netherlands	Netherlands	South Africa	Monaco
2019	China	Netherlands	Italy	South Africa	Monaco
2020	China	Netherlands	Netherlands	Czechia	Czechia
2021	China	Netherlands	Spain	Czechia	Poland

Fig. 6 Table countries with best centrality values by year

Obtaining as a result the table depicted within *Fig. 6*.

We can see that for all years the best exporter of products according to degree centrality is China, while as the best importer always according to this metric we have the Netherlands, we see different countries differentiate over the years according to the centrality of eigenvectors and proximity.

In terms of the link between data collected, measures applied, and properties found, we can give our own explanation for the centrality measures we used. Degree centrality, in this study, is useful for us to determine the importance of a country only from the numerical perspective of imports and exports while, in order to determine the importance of a country based on the entire trade network, it is important to consider closeness centrality and eigenvector centrality. For these reasons, we have chosen to use these measures of centrality in order to address the problems highlighted at the beginning of the report from several different perspectives.

Links and codes:

- 2017, code: measure_centrality_17.py, csv: measure_centrality_2017.csv;
- 2018, code: measure_centrality_18.py, csv: measure_centrality_2018.csv;
- 2019, code: measure_centrality_19.py, csv: measure_centrality_2019.csv;
- 2020, code: measure_centrality_20.py, csv: measure_centrality_2020.csv;
- 2021, code: measure_centrality_21.py, csv: measure_centrality_2021.csv;
- Best exporters and importers by centrality, code: best_exp_imp.py, csv: best_exp_imp.csv;

6. Conclusion

In this study, we analyzed the world trade network through Social Analysis by first providing a topological representation of the network in which we infer: the main trade that occurred between countries, the difference in imports due to the size of vertices, and, in Fig. 3 using the Kamada and Kawai force-directed algorithm, the importance of countries derived from their placement in the network.
Next, we extrapolated important information such as major importers/exporters from the data, defining which are the world leaders in trade.
We based the concept of a country's "importance" not only on import/export data, but also on network topology, using various measures of centrality. Using various measures of centrality allowed us to make different considerations about the concept of "importance" of a country, in fact, we looked at the network from various points of view, and this is where Social Network Analysis makes a difference compared to other approaches.

In addition to this, from the graphs obtained, it was possible to observe that COVID-19 did not significantly affect the flow of trade.

In conclusion, from the results obtained we can observe that the largest exporter of products for our dataset is China, while the best importer is the Netherlands, according to degree centrality. France, with Monaco, has more interconnectedness and trade accessibility than the other countries, so it comes out the best with respect to proximity centrality. As for the other centralities, however, we refer to *Fig. 6* shown above.

7. Critique

The work done aimed to represent bilateral trade flows and the importance of countries involved in trade through Social Network Analysis. The objective was successfully achieved, as it was possible to represent the 2017 world trade network by comparing it with the 2007 world trade network derived from the study done earlier BACI: *International Trade Database at the Product-level The 1994-2007 Version*.
The goal of country importance is achieved through the use and explanation of the various measures of centrality that allowed us to define various concepts of a country's importance in the trade network.

A second objective we set for this work is to analyze the effects of COVID-19 on the flow of trade. From what the data showed, we found that for the years 2019-2020, COVID-19 slightly affected world trade, this can be observed from the network density in those years, noting that it did not drop by even 1 percentage point.

However, one downside to note is that the dataset we used in the analysis has a small number of features to describe each trade flow. It would have been useful to analyze trades according to a finer temporality than the one used in the dataset; in fact, for each trade we have reference only to the year but not to the precise date of the trade. Having a finer granularity on temporality would have led to a more accurate density study, and from the perspective of the impact of COVID-19, a study could have been done by merging the trade dataset with a COVID-19 trend dataset.

With this study we have brought to light the general characteristics of the trade network, but as a future development we could look at the characteristics of the network on a region or cluster basis, going to analyze in more detail the flows that occurred within it.