Network Analysis Project Report

Francesco Pieroni, Vairo Di Pasquale, Artificial Intelligence Course

1 Abstract

In this study, we have examined the BACI-CEPII [2] database by means of Network Analysis tools. Building the 2020 World Trade Network of energy-related goods, such as Coal, Oil, Natural Gas, Liquefied Natural Gas, and Electrical Energy, we have been able to highlight the imbalances in their Import/Export trades. We found out that the networks are very fragile, possibly not able to resist extraordinary events. Indeed, the unexpected war between Russia and Ukraine has been able to cause 2022's energy crisis and the most affected countries are the ones which, according to our study, appeared to be in a more critical situation.

2 Context

The subject of this project is the economy, specifically, trade flows between countries that engage in international trade. The term *international trade* refers to the exchange of goods and services between companies located in different countries. By doing so, they can expand their market shares and gain access to all the resources they need. As a result, the global marketplace is more competitive, resulting in cheaper products and lower prices for the average consumer. Considering recent events and the current state of the global economy, our work will mainly focus on energy-related commodities, such as coal, natural gas, oil, and electricity.

3 Problem and Motivation

Throughout history, international trade has played a significant role in the rise of the global economy, as both supply and demand (and hence prices) are influenced by global events. Since the world is currently experiencing a huge energy crisis, we believe it is critical to examine the flows of goods whose prices explosion has led to all-time high energy quotations, and thus a concrete impact on people's lives, as inflation rose like in the '80s, the last time in which there was an energy crisis. A key component of the project is finding imbalances in trade flows and identifying dependencies between countries during 2020, which is the year before the start of the current crisis.

4 Datasets

In this study we used data from the 'BACI-CEPII' [2] dataset, a secondary type dataset widely used in applied trade analysis and freely available for download. The dataset includes 5000 products, which provides information on bilateral trade flows for 200 countries.

Exporter	Importer	Product	Value (thousands USD)	ContinentExporter	ContinentImporter
SGP	JPN	GNL	19660.154	AS	AS
CHN	BRA	Coal	78068.777	AS	SA
ESP	THA	Coal	433.824	EU	AS
PAK	PNG	Oil	191.818	AS	OC
NLD	BEL	GNL	2073.201	EU	EU
HRV	SVN	Eng	18066.778	EU	EU
KAZ	CZE	Oil	221445.829	AS	EU
CHN	EGY	Coal	3243.840	AS	AF
CHN	SLV	Coal	45.928	AS	NA
ITA	ISR	Oil	3712.467	EU	AS

Table 1: Final Dataset Sample

4.1 Data Preprocessing

In order to focus on 2020 trades, we used the Harmonized System (HS) revision 17, which covers all goods flows between 2017 and 2020. For the sake of keeping only the trades that occurred in 2020, we used Python with the pandas [3] library in order to filter the relevant information. As mentioned before, we have focused only on energy-related goods, namely "Coal", "Oil", "GNL" (Liquefied Natural Gas), "Eng" (Electrical energy), and "GAS" (Natural Gas at gaseous state). Additionally, we removed entries with missing data and values less than 1000 USD, as they are not interesting. That leaves us with a dataset of 2849 trades for the year 2020. A sample of the final dataset is shown in Table 1. Also, note that **ContinentExporter** and **ContinentImporter** features have been added. These columns show the continent code of the countries involved in the trade, which is useful for an in-depth understanding of intercontinental relationships.

5 Validity and Reliability

The source data we used for our study allows the model to be extremely valid and reliable: as mentioned before, we take the raw data from the "BACI dataset", a detailed international trade database overwhelmingly used in applied trade analysis [1]. Its massive popularity comes from the fact that it takes direct data from the "United Nations Comtrade database" [4], which aggregates detailed global annual and monthly trade statistics by product and trading partners to be used by governments, academia, research institutes, and enterprises. In addition to covering approximately 200 countries, the United Nations Statistics Division compiles data on more than 99% of the world's merchandise trade. The CEPII team, as reported in their official site and working paper [2], performed various operations to improve the data to create the BACI dataset, like mirror statistics strategies to impute missing data.

6 Measures

A key purpose of the project is to capture the imbalances in Import/Export trades using centrality measures. These measures highlight that the network is heterogeneous, as expected for the goods under consideration. Therefore, the metrics that we calculated in this study are:

- *Degree centrality*: A node's in and out-degrees tell us how many Import/Export connections it has inward and towards other nodes.
- *Strength centrality*: Due to the fact that every trade is weighted by its volume in USD, its degree centrality will also be used in its weighted version, also called strength centrality. It is useful to identify the most relevant commercial partnerships.
- *Closeness centrality*: This measure is used to determine how close a node is to the others. A node's distance is measured by the number of arches traversed in order to get between two nodes. As a country becomes more "commercially close" to another one, its trade flows are more likely to intensify.
- *Betweenness centrality*: Computes the extent to which a node lies on paths between other nodes. If a node belongs to many shortest paths, then it is likely to be a major player since commercial routes are likely to cross that country.
- *PageRank*: In this approach, a node is central if it is connected to other popular nodes. As a result, we are able to determine the relationships between the major players in the market.
- *Assortative Mixing*: Labelling each node with its continent (e.g. Europe, North America), we used this measure to see if countries tend to have more trade with others belonging to the same region.
- *Modularity - Disassortativety by degree*: Nodes with high out-degrees are likely to be connected to those with low out-degrees, since a leading exporter is more likely to trade with countries that lack of resources and must import them, rather than trade with other exporters. Furthermore, we also compared networks representing trade flows for different commodities, to see if they follow the same trend or not.
- *Centralisation and Core-periphery*: Used to understand how much big exporters play a key role in the whole network of the global sphere of trade. In particular, in the case of a few big exporters and many big importers, it is expected to have a network with a very easily distinguishable dense core and a loosely interconnected periphery.
- *Groups of nodes*: We used 4 types of metrics to identify interesting groups of nodes:
  - *N-cliques*: To identify groups of strongly interconnected nodes.
  - *K-components*: To point out different commercial areas in which goods flow mostly between nodes of those regions.
  - *Clustering Coefficients*: Useful to find how many commercial triangles exist in the network, because they are considered to be more balanced than a dyad in which an actor can possibly impose itself on the other.
  - *Reciprocity*: To recognize how many pairs of nodes have both a relationship of Import and Export.

  The grouping of nodes will be determined by a redesigned network with the relationship "flow", where the start node is the exporter and the end one is the importer. Its weight is the sum of the values of all energy-related trades from the exporter to the importer. In addition, for what concerns cliques and components, we applied two symmetrization techniques to make the graph undirected.

7 Results

In order to see the connection between the gathered data and the applied measures, it is important to see the structure of the graph we have built. For every commodity, we have created a graph in which for every relationship exporter-importer there is an arch from the exporter to the importer, with weight equal to the yearly volume of trade in USD. For instance, in 2020, Albania exported electrical energy for 42mln USD to Greece, so the graph about electrical energy will have an arch going from Albania to Greece with a weight equal to 42mln. In the following sections, we will inspect the results on all the obtained graphs.

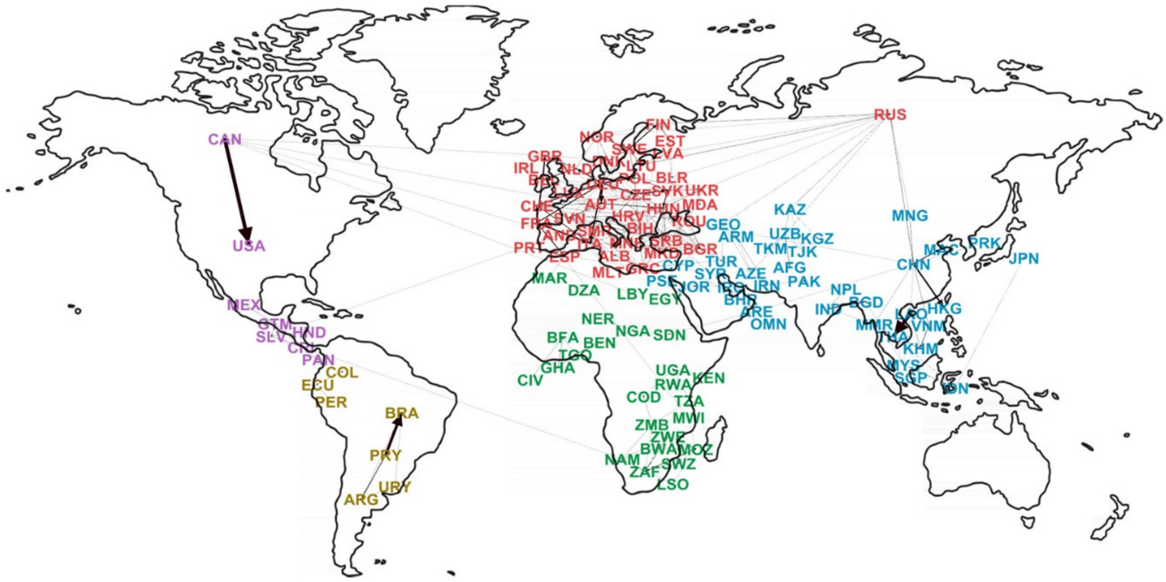

Figure 1: International World Trade Network Map

7.1 Degree and strength centrality

To start this analysis, we inspected In/Out degree and strength centralities, which are useful to identify the major players in the trade of the selected goods. Note that in order to fit with a reasonable number of pages for this report, only the measures about Electrical Energy trades will be shown while for the other goods only the main results will be mentioned.

From the charts below in Figure 2 and Figure 3, we can see the distribution of In/out-degree and strength centralities. It is clear that the higher the degree centrality, the higher number of trades of the specified good the country will have. Since these connections might have small weights (which means they represent small USD volumes of trades) it is more relevant to analyze strength centrality, in which edges' weights are taken into consideration. Indeed, the In/Out strength of a node is the sum of In/Out edges' weights, normalized by the sum of In/Out edges' weights. As a result, it is able to capture the importance of a country in the global trade of a specific good. Taking a look at the charts in Figure 2 we can see that the Czech Republic (CZE) has a high In/Out degree centrality, which means that it both imports and exports Electrical Energy from

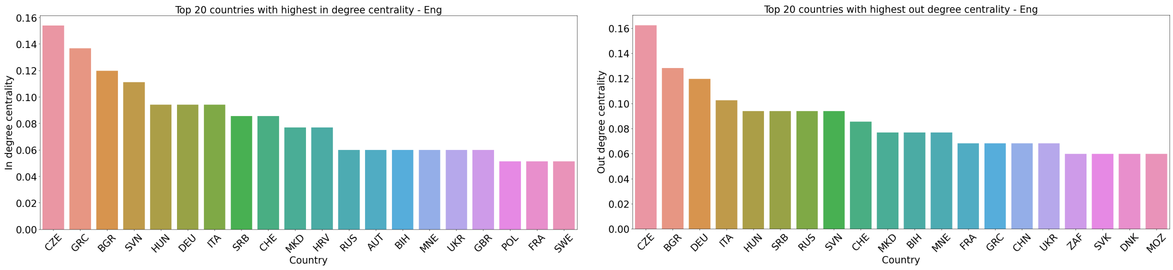

Figure 2: In/Out degree centralities for energy network

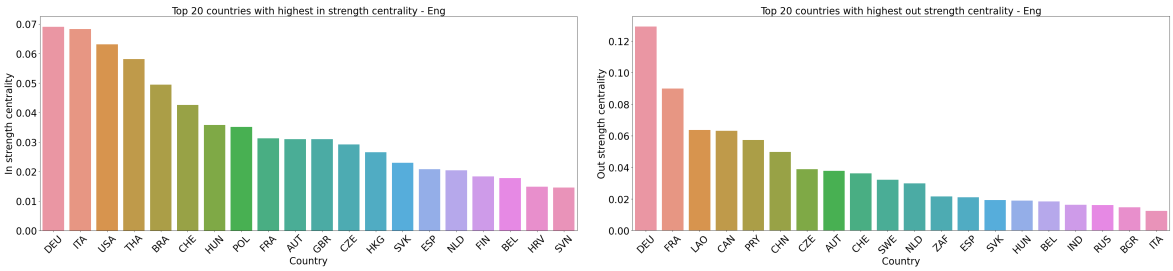

Figure 3: In/Out strength centralities for energy network

many countries, thus a high degree of diversification among its buyers and suppliers. Although it might seem strange for a country to both import and export Electrical Energy, it is perfectly normal, as it often happens that during the day, when more energy is needed, a country needs to import it, while during the night it has a surplus, which can be exported. Another possible option is that the country imports electrical energy to sell to other states. Moving to strength centrality, we can see the pitfalls of degree centrality. Indeed, the Czech Republic (CZE) is only the 12th importer and 7th exporter, as its trade volumes are smaller than the ones of other countries. On the other hand, we can see that Germany (DEU) is the major importer and exporter of Electrical Energy. Furthermore, an alarming situation can be noticed: Italy (ITA) is the second major importer of electrical energy (in-strength-centrality$\sim$0.069), while it is ranked only 20th when it comes to export (out-strength-centrality$\sim$0.018). This situation is heavily unbalanced, as it means that in case of an increase in the quotation of electrical energy, it would have to face huge increases in the import prices, which would not be balanced by export revenues, which are way smaller. Indeed, this strong dependence on external suppliers brought Italy to the energy crisis of 2022. For the results observed on the other commodities, similar observations can be made:

- **Coal**: India (IND) and China (CHN) are the major importers of coal, with high values ($\sim$0.11) of in-strength-centrality. This is alarming as coal's combustion creates a lot of CO_2 and toxic gases, and these results show that these two countries make larger use of coal than the other ones in the world. On the hand of exports, Poland (POL) is the biggest one by USD volumes. Its out-strength-centrality is around 0.25, which is extremely high and thus alarming, as it can lead to market manipulation. Indeed, this is probably already happening: Poland is not even in the top ten coal for energy purposes exporters if we analyze the metric tons of coal exported by this country, while it is the most central one if we consider the volume in USD. This means that Poland is probably taking advantage of its position to sell coal at higher prices to neighboring countries.

- **GAS**: as expected, the most central countries are Norway (NOR) and Russia (RUS), which have out-strength-centrality values of about 0.17 each. Together with the fact that the third highest value is about 0.08, this makes them absolute market dominators. In addition, taking a look at the in-strength-centrality, we can see that Italy (ITA), Germany (DEU), and France (FRA), which are suffering from the energy crisis, are indeed the countries with the highest in-strength-centrality values, which makes them highly dependent on the others, and thus, extremely exposed to price's volatility.

- **GNL**: observations similar to the one made for natural gas. There are two major exporters, Australia (AUS) and Qatar (QAT), which dominate the market (strength values: 0.26 and 0.18), and three major importers: Japan (JPN), China (CHN), and Korea (KOR).

- **Oil**: Saudi Arabia (SAU) and Russia (RUS) have out-strength values of 0.15 and 0.12, which are not enough to make them control the entire market as there are four other countries (USA; CAN; IRQ; ARE) which have a value of about 0.09 each. This means that the major players can be contrasted by the union of all the others, which makes it a more balanced situation. Indeed, during 2022 oil faced an increase in quotations but they did not explode like energy and gas ones.

7.2 Closeness centrality

In the world trade scenario, being commercially close to other countries is very useful, as it can lead to an intensification of trade along the commercial routes that a country has. Ideally, if a country is directly, or at most with few edges between, connected to many others, in situations like the current energy crisis it would be capable of getting new suppliers faster than a more isolated one.

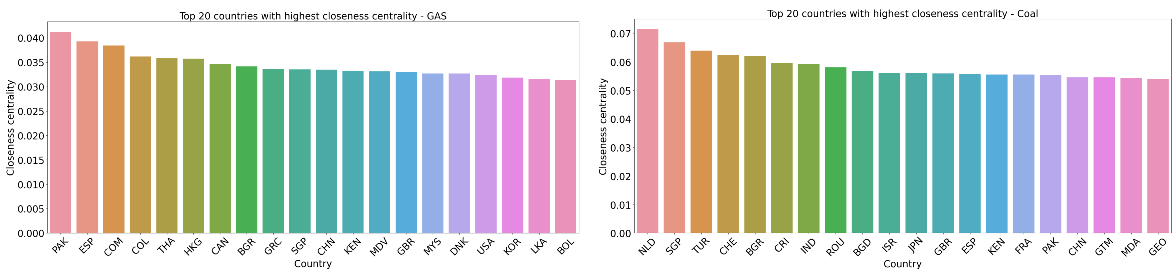

Figure 4: Closeness centralities for Natural Gas and Coal networks

From the above charts in Figure 4, it is clear that the values of closeness are very similar inside the two networks. This ensures that, at least among the major traders, there is no country completely isolated. For instance, Italy, which we have already seen suffers from high dependence on other countries for Natural Gas and Electrical Energy, still has a closeness-centrality value of around 0.029 in the gas network and 0.045 in Coal's one, which is comparable to the other countries' values. Since these values are about 2020 trades, at that time Italy could be considered capable of changing suppliers in a not-so-hard way. Indeed, after the start of the crisis, Italy did it, reducing its dependence on Russian gas. For what concerns Electrical energy, Oil, and GNL, the exact same observations can be made. All the values of closeness centralities are perfectly comparable and, excluding marginal traders with very limited in/out flows of goods, only a few countries can be considered too isolated in the network. Furthermore, this can happen also because the country has a production of a good which is very close to its yearly need, thus it has no relevant import or export trades. An example of this situation is Tunisia, which has a closeness-centrality value equal to 0 in the gas network.

7.3 Betweenness centrality

This measure is useful to identify the nodes that lie on commercial paths between other nodes. Their position is strategic, as they can act as a bridge and an intermediary between different parts of the network. In the following, we report the charts below in Figure 5about betweenness centrality values for Oil and GNL networks.

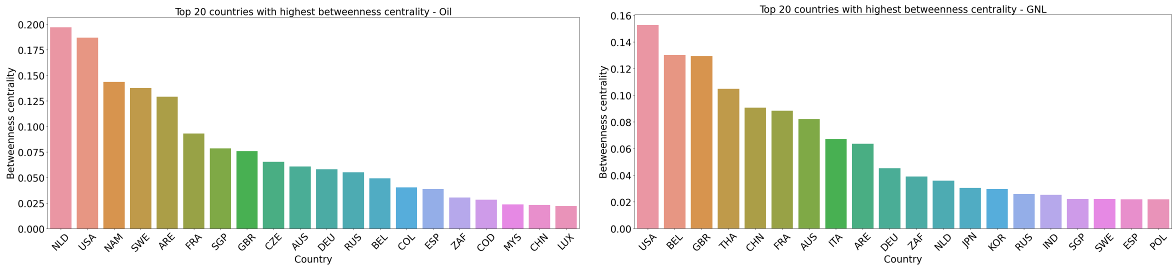

Figure 5: Betweenness centralities for Oil and GNL networks

For what concerns oil, it can be seen that the Netherlands (NLD) and the United States (USA) play a key role. Indeed the Netherlands, with its ports, represents the connection between Central and Southern Europe with Northern Europe and America. Especially, there is a strong commercial route involving USA and Netherlands, as the former is among the biggest exporters of oil. All in all, countries that lack infrastructure such as ports, have to rely on others in order to trade voluminous goods like oil barrels. This gives intermediary countries a strong power, as no other feasible option is available to establish the commercial route.
Similarly, there is a kind of gateway for GNL that is represented by the USA, Belgium, Great Britain, Thailand, and China. These countries have the highest values of betweenness centrality in the GNL network, which is due to the fact that they connect different regions of the world, mainly North America with Europe and Asia with Europe. An important observation is that there is a strong unbalance in both networks, as there are countries with betweenness centrality values that are extremely higher than the others. This configuration is potentially dangerous, as the intermediary countries have a monopoly on commercial routes. In addition, imbalance brings a lack of robustness. Indeed, if for some reason a crucial node dramatically reduces the flow of goods through itself, the network can become paralyzed.
For the other commodities, we see that Russia, China, Great Britain, Germany, United States, Netherlands, and France have the highest values of betweenness centrality, thus the countries that have been mentioned in oil and GNL analysis are confirmed to be part of the most developed commercial routes across the whole world.

7.4 PageRank

PageRank is an interesting measure that helps us understand which are the popular nodes, and which are the ones that have connections with other popular nodes. The results of the application of this measure on gas and electrical energy network are shown below in Figure 6

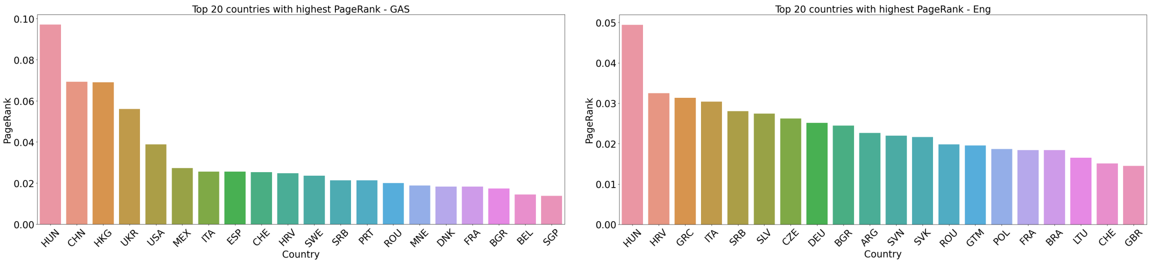

Figure 6: PageRank for natural gas and electrical energy networks

Unexpectedly, Hungary (HUN) is by far the country with the highest PageRank values. In addition, while for electrical energy the nodes tend to have PageRank values that have a steady trend, for gas only a few nodes have high values, suggesting that they probably interact a lot one with the other. The case of Hungary is very peculiar, as this country trades gas and electrical energy nearly with all the European states, which of course have many commercial connections with other important countries. For what concerns the other goods, it is worth mentioning that Italy has by far the highest PageRank value in the GNL network, which means that it trades it with the major market players. In addition, it is the second in the coal network. The main reason is that, due to its central position, Italy has many relevant trades with many important commercial partners.

7.5 Assortative Mixing

In order to check if countries tend to have trade with other close ones, namely belonging to the same continent, we labeled every state with its continent and computed the assortativity coefficient based on this new node attribute for every commodity's network. The results are shown below in Figure 7.

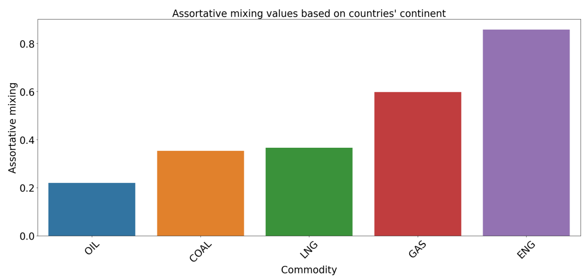

Figure 7: Assortativity coefficients based on countries' continent

The chart shows that assortativity varies a lot depending on the good taken into consideration. In detail, the electrical energy network has a coefficient of 0.86, which is extremely high, meaning that most of the trades of electrical energy happen between countries of the same continent. Considering the nature of electrical energy and its intrinsic difficulty in being stored, it perfectly agrees with the expectations, as intercontinental trade can happen only between two countries at the border of two continents (e.g. Turkey and Greece). Furthermore, a similar trend has been highlighted for trades of natural gas, as being a gaseous state, methane pipelines are needed. Since this type of infrastructure gets more and more expensive with the increase in the distance to cover, it is generally used to connect close countries, which tend to belong to the same continent.

Indeed gas coefficient is around 0.6. On the other hand, GNL, Coal, and Oil networks show the opposite trend. Their coefficients are 0.36, 0.35, and 0.22, which means that most of the trades are intercontinental. The main reason is that all these goods can be easily stored and shipped, so they can reach any part of the world, contrary to electrical energy and natural gas a gaseous state.

7.6 Modularity

While five disassortative networks were expected, the results were slightly different, as shown in Figure 8. Indeed electrical energy network has an assortativity value of 0.35, which means that nodes with high degrees tend to connect with others with high degrees. This is a bit strange, but we think that the main reason is that the electrical energy market is very peculiar, as it is differentiated between day and night. The other networks tend to be disassortative, but not as much as expected. Basically, values of assortativity around 0 mean that there is nearly no difference in trades between high and low-degree nodes.

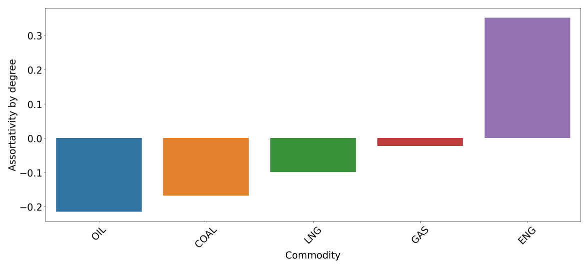

Figure 8: Assortativity by degree coefficients

7.7 Core-periphery

We decided that the most suitable approach to visualize a structure of core-periphery in our network would be to use the data for the strength centrality described in Section 7.1. The use of strength centrality over degree centrality is critical because the latter makes no sense as we have seen that some nodes have many edges but their weights are too small. Ideally, a single edge with a weight of 1 Billion USD is more important than 10 edges with a weight of 10000 USD each. The results for Coal and GAS trades can be seen in Figure 9 and Figure 10 respectively. Europe is the leading region with the most core nodes for the Gas trade, with Norway (NOR) renowned for its Gas exports, and Germany (DEU), Belgium (BEL), France (FRA), and Italy (ITA) importing a great deal of Gas from Russia (RUS), regarded as the world's leading exporter of Gas. Referring instead to the Coal trades in Figure 9, again Poland (POL) emerges in terms of USD volumes: with this visualization is even clearer the strategic position it makes advantage of. Indeed, nearly all of Europe imports from it, and thus its coreness is very high. It follows China (CHN) which is the second biggest core node, mainly because similarly to Poland in Europe, it has connections with many Asian countries. One of them is Turkmenistan (TKM), which in spite of the fact that it exports huge amounts of coal to China (note the big arrow from TKM to CHN), is marked as a periphery node. This happens because TKM has nearly no other connection with other nodes, so its strength is only due to one edge.

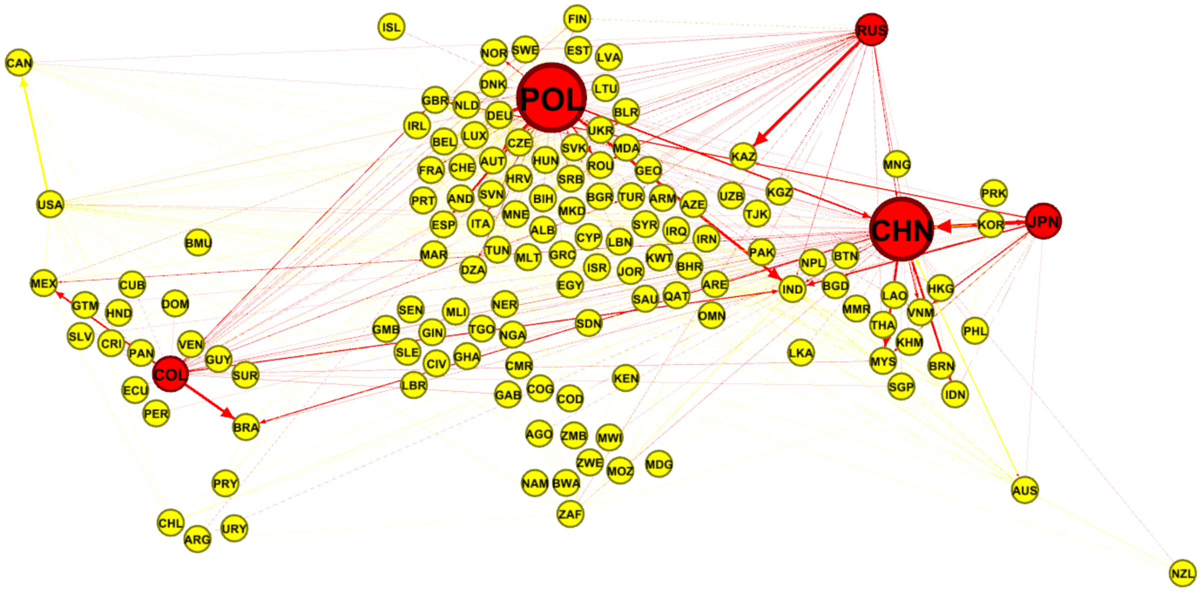

Figure 9: Core-periphery structure - Coal

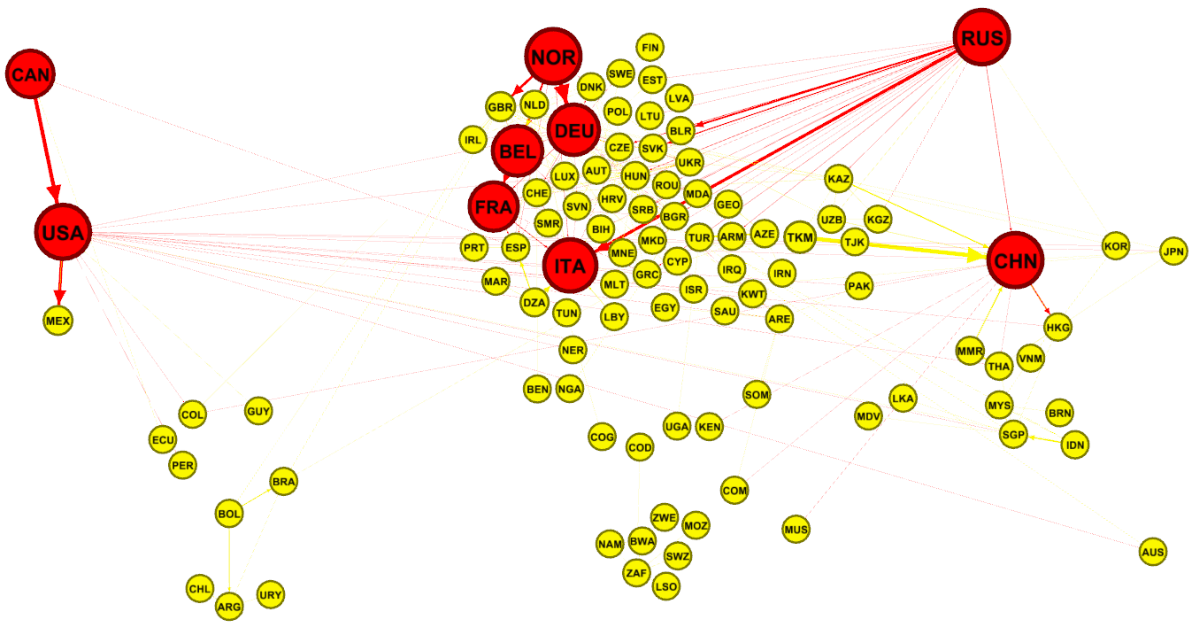

Figure 10: Core-periphery structure - Gas

7.8 Groups of nodes

In order to study how countries are grouped, we created a new graph in which all the trade volumes are put together. This means that an arch from node A to node B exists if A exports at least one of the selected goods to B. The weight will be the sum of all the trade volume in USD from A to B. Using this network we computed the following measures:

- *Local clustering*: it gives us information about the way nodes are connected. From the chart below it is clear that if country A does business with B and C, then it is very unlikely for B and C to have relevant trades. This result tells us that a country that needs to import one of the goods taken into consideration most probably will lack the other commodities too. As a result, it is very unlikely that it exports goods to other states.

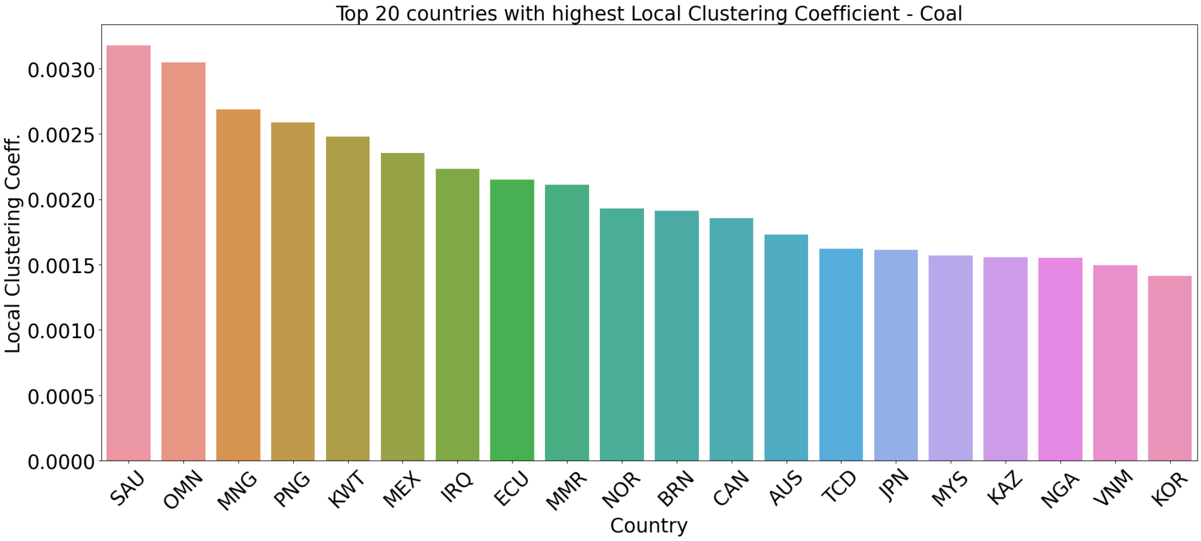

Figure 11: Local clustering

- *Clustering*: Moving to another clustering measure, we computed clustering on the whole network in order to find commercial triangles, which are more balanced than dyads and open triads. The main reason is that triangles reduce the dependency of one country on another because they bring competition between suppliers so that importers can diversify their sources of goods. In this network, the clustering coefficient is 0.28, which means that many triangles do exist in the network. This result is not in contrast to the local clustering one because it is computed without considering edge weights (volume of trades). Indeed, it highlights the fact that triangles exist but tend to have at least one edge with a small weight. All in all, this clustering measure only gives us an idea of the network's structure, but taking into account edges' weights as we did for local clustering, we can see that the network is extremely unbalanced.

- *Reciprocity*: It is also interesting to see how many nodes have a relation of both import and export between them. The reciprocity value is 0.42, which is considerably high and most probably it is due to the fact that trades of only a few thousand dollars are present in the networks. As a matter of fact, we dug deeper and inspected the network without edges representing trades of less than 1 million USD, obtaining another value of reciprocity that is 0.34. This value is still high, meaning that it is quite common to have both import and export between two countries, which can not be divided into exporters or importers. Although, it has to be said that in this market trades of billion of USD are very common, so the ones of millions of dollars can still be considered as marginal.

7.8.1 Clique an Components

Finally, we created an undirected version of the graph to complete the analysis of groups of nodes. Starting from the total graph mentioned above, we applied two symmetrization techniques, one conservative and another one more permissive.

- *Conservative*: two countries A and B will have a link if and only if the flow from A to B is at least a third of the flow between B and A (or vice versa). This technique ensures that two nodes are connected if and only if their Import/Export trades are comparable, so the undirected edge simply replaces two directed ones with comparable trade volumes. This technique creates fewer edges than the permissive one, giving more importance to trade volumes.

- *Permissive*: two countries A and B will have an edge if at least one exports to the other. This means that only one directed edge between two nodes is needed to have them connected by an undirected one. This technique creates more edges than the conservative one, giving more importance to the existence of an edge in the starting graph.

Using the conservative technique, the largest cliques we can find are of size three. For example, Japan (JPN), China (CHN), and South Korea (KOR) form one of them. These cliques are very important because they represent strong commercial triangles, as they have been found in the network generated with the conservative approach. As expected, larger cliques do not exist, because it is very unlikely to satisfy the conservative condition. Furthermore, we found out that in this network the maximal component has size 57, which is extremely high, proving that the goods are taken into consideration flow through the whole world.
On the other hand, using the permissive technique, we found cliques of size 13. Even though they are useful to see the parts of the network that are more interconnected, they can include importer-only or exporter-only countries, which means that a connection exists only in one direction. However, if we are interested in the areas where the trade of energy-related goods is more developed, without considering the role of each player, this is an interesting measure. As expected, one of the largest cliques is made off of most European countries: AUT, HUN, CZE, KAZ, DEU, ITA, ESP, FRA, BEL, POL, RUS, BGR, and NLD. For what concerns components, the maximal one is of size 185, which means that almost every country of the world is inside the maximal component. The ones outside the maximal component are not present in the data set as their trade volumes are close to 0.

8 Conclusion

All the results described up to now made it clear that the energy-related trade network was heavily unbalanced and thus very exposed to extraordinary events. We have seen that there are some countries such as Italy and Germany which are dramatically dependent on a few major exporters, which is alarming because every country should follow the basic principle of diversification. These results are clear also from the core-periphery study, which made us able to understand the major actors in the networks. On the other hand, we have been able to see that many countries are commercially close to others, so we can expect them to be able to easily intensify trade with one another and decrease trade with another one. This is an interesting result because it means that the network has the potential to absorb drastic changes. Indeed, we have been able to predict Italy's capability to change Natural Gas suppliers. Furthermore, we dug deeper inspecting also trade intermediaries, and pointing out the problem of commercial route monopolies. For example, according to betweenness centrality values, the Netherlands is in control of the oil commercial route to Europe. This situation is critical because it means that every country must trust the intermediary, making the network centralized. Finally, in a healthy trade network, there should be a lot of competition between the components both in import and in export, because in this way prices are better regulated and no country can impose itself on the others. Unfortunately, we have seen that in most of the cases the major actors are just a few nodes and the balanced situation can be found only in very small subgraphs, as shown in the

Cliques' analysis section using the conservative approach. All in all, the final conclusion is that the networks we have analyzed are too unbalanced to be considered healthy, as we have many measures that are pointing out critical situations.

8.1 Critique

It has to be said that we just scratched the surface of what can be done using network analysis and highly reliable data such as the ones available inside the BACI-CEPII dataset. Our work is a starting point that is still able to point out huge problems and make economy experts aware of them, in order to make some decisions to reduce the impact on the exposed countries. For instance, if all these considerations would have been made at the beginning of 2021, when there was still time to increase the diversification of Italy's natural gas suppliers, then something toward this direction could have been made and the impact of the current energy crisis could have been reduced. Although, this dataset contains data about two years before the current one, so it is not appropriate to analyze trades about the past year.

References

[1] BACI Cepii entries in Google Scholar. https://scholar.google.com/scholar?start=20&q=BACI+Cepii&hl=en&as_sdt=0,5. [Online; accessed 24-November-2022].

[2] Guillaume Gaulier and Soledad Zignago. *BACI: International Trade Database at the Product-Level. The 1994-2007 Version*. Working Papers. 2010-23. 2010. URL: http://www.cepii.fr/CEPII/fr/publications/wp/abstract.asp?NoDoc=2726.

[3] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

[4] UN Comtrade Database Website. https://comtrade.un.org/.