Network Analysis Project Proposal

Francesco Pieroni, Vairo Di Pasquale, Artificial Intelligence Course

0.0.1 Context

This project is about the economy, in particular, trade flows between most of the countries in the world by means of international trade.
International trade consists of the purchase and sale of goods and services between companies in different countries, enabling them to expand their market share and gaining access to resources otherwise unavailable in their domestic markets. Therefore, the global marketplace is more competitive, resulting in more competitive prices and cheaper products for the everyday consumer.
In light of recent events and the current state of the global economy, our work will primarily focus on trades of energy-related commodities, such as gas, petroleum, and electrical energy.

0.0.2 Problem and Motivation

International trade was the key to the rise of the global economy, where supply and demand (and thus prices) both impact and are impacted by global events. Since the whole world is currently facing a massive energy crisis, we believe it's crucial to analyze the flows of those goods whose prices explosion lead to all-time-high energy quotations and thus have a concrete impact on people's everyday life, as inflation rose like in the '80s, which have been the last period of energy crisis.
The project is about finding imbalances in trade flows, identifying the dependencies between countries during the year 2020, which is the last year before the start of the crisis.

0.0.3 Datasets

The source of data we are going to use is of secondary type and is the 'BACI-CEPII' dataset, overwhelmingly used in applied trade analysis and freely available for the download. Since we want to focus on 2020 trades, we will use the Harmonized System (HS) revision 17, which includes all the flows of goods between 2017 and 2020. We used Python with the pandas and matplotlib libraries in order to study, inspect and filter the data to retrieve the needed records.
Furthermore, we will use Gephi and the NetworkX library to visualize the data, build the network and compute the needed measures.

0.0.4 Measures

For the purpose of the project, centrality measures play a key role, as they are able to capture the imbalances in Import/Export trades, highlighting the heterogeneity of the network, which is expected to be clear for the goods taken into consideration.
We will compute the following metrics:

- *Degree centrality*: In and out-degree of a node are the number of connections it has inward or toward other nodes, thus it tells us how many Import/Export connections every country has.

- *Strength centrality*: Since every trade is weighted by its volume in USD, degree centrality will be used also in its weighted version, also called strength centrality. This measure is useful to identify the most relevant commercial partnerships.
- *Closeness centrality*: It measures how close a node is with respect to all the others. The distance between two nodes is taken as the number of arches to traverse in order to go from one to the other. This measure is important because the more a country is "commercially close" to another one, the more likely it is they will intensify trade flows between each other.
- *Betweenness centrality*: Measures the extent to which a node lies on paths between other nodes. The idea is that if a node belongs to many shortest path, it means that it is likely to be one of the major player, as commercial routes will likely cross the country represented by that node.
- *PageRank*: Born to solve the zero-trailing of eigenvector centrality, which is determined by the eigenvector centrality of nodes' neighbors, relies on the idea that a node is central if it has connections with other popular nodes. This measure will make us able to see how the major players in the market are related.
- *Assortative Mixing*: Labelling each node with its region (e.g. Europe, North America), we want use this measure to see if countries tend to have more trades with others belonging to the same region.
- *Disassortativety by degree*: High-out-degree nodes are expected to be connected to low-out-degree ones, due to the fact that a great exporter of one good is likely to connect with countries which do not have that good and need to import it, rather than to trade with other exporters. Furthermore, a comparison between networks representing trade flows of different commodities will be carried out.
- *Centralisation and Core-periphery*: In order to better understand how much big exporters plays a key role in the whole network of the global sphere of trade, in the case of a few big exporters and many big importers we expect to have a network with very easily distinguishable dense core and a loosely interconnected periphery.
- *Groups of nodes*: To find groups of nodes we will consider a new network with the relationship "flow", in which the start node will be the exporter and the other one the importer. The weight will be the sum of the volumes of all the energy-related goods from the exporter to the importer. We are going to use 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 exists in the network, because they are considered to be more balanced then a dyad in which an actor can possibly impose itself on the other.
  - *Reciprocity*: To recognize how many pair of nodes have both a relationship of Import and Export.