[1]Department of Statistics and Data Science, University of Florence
[2]Department of Informatics and Statistics, Prague University of Economics and Business
[3]Department of Economics Science, University of Bergamo
[4]National Institute of Statistics, Istat

Greenness and Graph Analysis: A study on pollution produced by International Trade

Main Authors
Fabbrico Davide[1], Khateeb Ahmad Arsalan[2] and Pagnoncelli Diego[3]

Istat Reference Tutors
Amato Francesco[4], Bruno Mauro[4], Causo Maria Serena[4] and De Fausti Fabrizio[4]

Keywords Graph Analysis, Regression, International Trade, Fuel consumption, Greenness.

1 Abstract

In this paper we talk about our analysis regarding International Trade. We use graph theory for data analysis with particular interest in the Environment topic.
We are interested in the graph analysis of the International trading network between 2019 and 2021.
To compute our analysis we exploited the *Comext Data*, in particular, our data relate to monthly international transport. In the Eurostat database these data are found under folders *Transport HS* and *Transport NSTR*.
In order to better understand the commercial connection we decided to create an index which tells us how several means of transportation affects pollution in different places in the world throughout the year. The name of our index is *Greenness*.
To compute it, we start from the dependent variable which is *pollution* that is define as follows: fuel (fuel used per kilometers run), mean of transport (we created a categorical variable that contains the categories: Road, Air, Post, Sea, Rail), type of product traded, distance (km run) and the error term. Our aim is to analyze two type of *Greenness*, a local metric and a global metric. The global metric gives each single node (each country) a value, while the local metric gives the possibility to compare two countries based on the new *Greenness* index.
Our main function takes as input a set of data (dataframe), the month and year of interest for the analysis, the product involved, the type of flow (import or export), type of transport (Rail, Air, Road, etc...) and a threshold to control the percentage of imports/exports within the analysis, in particular the user can choose to manage euros or quantities in kilograms.
This function filters the dataset for each user choice and returns a direct graph with the previously desired filters.
The graph G, which represents an instance for each product traded, is a doublet (V, E) in which V is the set of nodes which represent each country and E is the set of edges

that connect two countries if an exchange has taken place between them. The edges also contain weights which indicate the distance between the two connected countries.
The data have been aggregated, that is, if we have two imports from Italy to the Czech Republic, always under the assumption of the same product and the same period, these two imports will be added in the quantity of kilograms and the quantity of euros.
After we have the data relating to the fuel consumption for each vehicle, we can calculate how many liters of fuel it has consumed and consequently the level of pollution for each transport.
Once the liters of fuel have been calculated, the nodes corresponding to each country will have another attribute called *Total Liter* (TL). This measure will be calculated as the sum of the weights (which will subsequently become liters of fuel and no longer distances) of the edges exiting the node (out-degree).